

Crystal Structure of the Spliceosomal 15.5kD Protein Bound to a U4 snRNA Fragment

Ivan Vidovic,* Stephanie Nottrott,*†
Klaus Hartmuth,*† Reinhard Lührmann,*†
and Ralf Ficner*†‡

*Institut für Molekularbiologie und Tumorforschung
Philipps-Universität Marburg
Emil-Mannkopff-Strasse 2
35037 Marburg

†Max-Planck-Institut für Biophysikalische Chemie
Abteilung Zelluläre Biochemie
Am Fassberg 11
37077 Göttingen
Germany

Summary

We have determined the crystal structure of a spliceosomal RNP complex comprising the 15.5kD protein of the human U4/U6.U5 tri-snRNP and the 5' stem-loop of U4 snRNA. The protein interacts almost exclusively with a purine-rich (5+2) internal loop within the 5' stem-loop, giving an unusual RNA fold characterized by two tandem sheared G-A base pairs, a high degree of purine stacking, and the accommodation of a single RNA base, rotated out of the RNA chain, in a pocket of the protein. Apart from yielding the structure of an important entity in the pre-mRNA splicing apparatus, this work also implies a model for the complex of the 15.5kD protein with box C/D snoRNAs. It additionally suggests a general recognition principle in a novel family of RNA binding proteins.

Introduction

The splicing of nuclear pre-mRNA to remove the introns and thus produce translatable mRNA is catalyzed by the spliceosome; this is a large ribonucleoprotein complex formed by the ordered interaction of the four small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U5, and U4/U6, appropriate regions of the pre-mRNA, and numerous splicing factors (reviewed in Krämer, 1996; Will and Lührmann, 1997; Burge et al., 1999). During the splicing sequence, the U4 and U6 snRNAs undergo dramatic conformational rearrangements (Brow and Guthrie, 1988); before association, these are joined by base pairing (Figure 1A; Bringmann et al., 1984; Hashimoto and Steitz, 1984; Rinke et al., 1985), whereas after integration into the spliceosome, the double helix joining the two snRNAs is completely unwound (reviewed in Nilsen, 1998; Staley and Guthrie, 1998). Whereas the U6 snRNA remains in the spliceosome, interacting there with the U2 snRNA and with intron sequences at the 5' splice site (reviewed in Staley and Guthrie, 1998), the U4 snRNA is released from the spliceosome or remains only loosely attached to it (Lamond et al., 1988; Yean

and Lin, 1991). It is thought that the U4/U6 interaction is made and broken in each cycle of splicing. The structural rearrangements of the U4 and U6 snRNAs are evolutionarily conserved between yeast and man, and they are also observed with the recently discovered U4atac and U6atac snRNAs of the so-called minor or U12-type spliceosome, which are functional analogs of U4 and U6 (Tam and Steitz, 1996; Yu and Steitz, 1997).

U4 snRNA possesses a stem-loop near its 5' end (Figure 1A) that has been found to be essential for pre-mRNA splicing and spliceosome assembly in vivo and in vitro (Vankan et al., 1992; Wersig and Bindereif, 1992). A major role of this 5' stem-loop appears to be to affix snRNP proteins to the spliceosome. Initial evidence for this idea came from studies in yeast where it was shown that antibodies directed against the protein Prp4p (60kD in human) precipitate the 5' portion of U4 snRNA including the 5' stem-loop (Banroques and Abelson, 1989; Bjorn et al., 1989; Bordonné et al., 1990; Xu et al., 1990). Additional U4 snRNP proteins are Prp3p (90kD in human), which directly interacts with Prp4p (Anthony et al., 1997; Horowitz et al., 1997; Lauber et al., 1997; Ayadi et al., 1998), and Prp31p (Weidenhammer et al., 1997). However, for none of these proteins a direct interaction with U4 snRNA could be demonstrated, and none of these proteins are found in the free U6 particle.

We recently identified the 15.5kD protein in human U4/U6.U5 tri-snRNPs (Snu13p in yeast) that directly and specifically interacts with the 5' stem-loop of U4 snRNA. This protein was shown to play an important part in the late stage of spliceosome assembly before splicing catalysis (Nottrott et al., 1999), consistent with an earlier finding that the 5' stem-loop of U4 RNA is itself required for the late stage of spliceosome assembly (Wersig and Bindereif, 1992; Hu et al., 1995). There is the further possibility that this protein may be involved in binding one or more of the aforementioned U4 snRNP proteins.

Apart from its involvement in spliceosome assembly, the 5' stem-loop of U4 snRNA has a number of other interesting features. There is strong biochemical and phylogenetic evidence (Nottrott et al., 1999) for the existence of the (5+2) internal loop indicated in Figure 1A. Thus, the integrity of the base pairs in the two double-helical regions (one consisting of only two base pairs), but not their identity, is decisive for binding the 15.5kD protein. In contrast, this protein binding is strictly dependent on the precise identity of the bases in the internal loop: five of the seven positions (U31, G32, A33, G43, and A44) cannot be changed or deleted without abolishing protein binding activity in vitro, while the remaining positions 29 and 30 must be occupied by purines for efficient protein binding. The nucleotides of the internal loop are evolutionarily highly conserved among U4 snRNAs of different species, including the human U4atac snRNAs found in minor spliceosomes (Figure 1B), to which the 15.5kD protein also binds via the stem-loop.

Like the internal loop of U4 snRNA, the 15.5kD protein is highly conserved, with 71% sequence identity between man and yeast (Nottrott et al., 1999). It belongs to a family of homologous proteins that share a novel RNA recognition motif and which are all components

‡To whom correspondence should be addressed (e-mail: ralf.ficner@mpibpc.mpg.de).

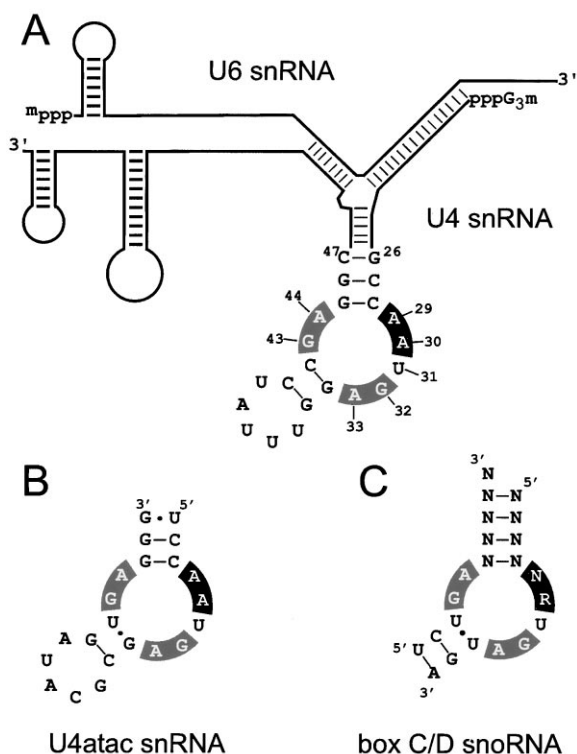


Figure 1. Secondary Structures of RNAs Binding the 15.5kD Protein (A) Sequence and secondary structure of the human U4/U6 snRNA duplex. The portion of the 5' stem-loop of the U4 snRNA used in this work is indicated by its nucleotide sequence. It comprises an asymmetric (5+2) internal loop formed by five nucleotides of the 5' strand and two nucleotides of the 3' strand. (B) The portion of the U4atac RNA from the minor spliceosome corresponding to the U4 snRNA fragment used in this work. (C) The corresponding portion of the box C/D snoRNA consensus sequence.

of ribonucleoprotein particles, such as ribosomes and small nuclear RNPs (Koonin et al., 1994; Nottrott et al., 1999 and references therein). The 15.5kD protein was also recently found to be a component of the box C/D small nucleolar RNPs (snoRNPs), whose function is the processing or 2'-O-methylation of pre-rRNA (Watkins et al., 2000). The latter work showed that this protein binds directly the snoRNA C/D motif, a structure that closely resembles the U4 snRNA 15.5kD binding site (Figure 1C), possessing the essential nucleotides of the internal (5+2) loop of U4 snRNA. Furthermore, the conserved box B/C motif of U3 snRNA (not shown), which can fold into a structure reminiscent of the internal loops of U4 snRNA and box C/D, also binds the 15.5kD protein in vitro (Watkins et al., 2000).

There is thus a seeming generality about the properties of both the 5' stem-loop of U4 snRNA and the 15.5kD protein that makes a knowledge of their three-dimensional structures and their mode of interaction of particular interest.

Results

Crystallographic Structure Determination

We cocrystallized the full-length human 15.5kD protein and nucleotides 26–47 of the human U4 snRNA 5' stem-

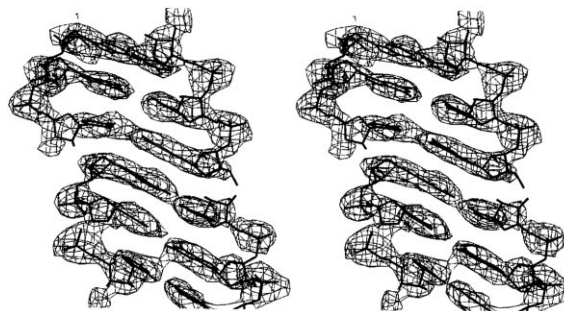


Figure 2. Excerpt from the Electron Density Map Obtained by MAD Phasing

The refined model (thick line) is superimposed upon the density nets. The figure shows the crystal contact of the two RNA molecules in the asymmetric unit. The blunt ends of the two RNA molecules are stacked one upon the other.

loop (Figure 1A), obtaining good quality crystals of their 1:1 complex. The structure was solved by multiwavelength anomalous diffraction (MAD), even though this 22 kDa RNP complex contained only one selenomethionine residue contributing to the anomalous diffraction; a detail from the experimental electron density map is shown in Figure 2. The crystal structure was refined to a resolution of 2.9 Å. The orthorhombic crystals contain two protein–RNA complexes per asymmetric unit. The two crystallographically independent protein molecules have almost identical structures, while the corresponding two RNA structures show minor differences, as discussed below, apparently caused by packing forces. The overall fold and structural features—patterns of base pairing and stacking—are the same in the two RNA molecules, while a few hydrogen bonds within the RNA and between the RNA and the protein differ.

Overall Structure

The structure of the 15.5kD protein–U4 snRNA complex is shown schematically in Figure 3A. The 15.5kD protein folds into a single globular domain with a familiar alternating α - β pattern. Contact is made with the oligonucleotide on one side of the protein only. A single nucleotide base (U31) from the internal loop of the oligonucleotide fits into a pocket in the protein, while the rest of the oligonucleotide folds into a structure approximating to two double helices converging at an angle of some 30°. One double helix contains two tandem sheared G-A base pairs, and both double helices are capped by a purine base at the junction; it appears that the unusual RNA fold is stabilized by a high degree of base stacking and a network of hydrogen bonds. The relatively small contact area of the two components of the complex mean that a large proportion of their surface is exposed.

RNA Structure

The RNA molecule folds into a compact structure that consists of two double-helical stems (stems 1 and 2) bridged by the (5+2) internal loop, which is asymmetric and highly structured (Figure 3). The two stems exhibit a slightly distorted A-form RNA double helix. A striking feature of the RNA structure is the strong bend at the

internal loop, which adopts a complex fold: of its seven nucleotides, four form tandem G-A base pairs extending stem 2, while the remaining three are unpaired. Of these, one (U31) is flipped out: it protrudes away from the rest of the oligonucleotide chain and toward the protein. The sugar-phosphate backbone forms a sharp hairpin-like turn at this point. The other two unpaired bases, A29 and A30, are stacked onto the ends of stem 1 and stem 2: A29 is stacked onto the base pair G45-C28 and thus caps stem 1, while A30, which has a syn and a 2'-endo conformation, is stacked onto A44 of the opposite strand and thus caps stem 2.

Previous chemical and phylogenetic analysis of the secondary structure of the 5' stem-loop of U4 snRNA bound to the 15.5kD protein had revealed four structural elements. These were the asymmetric (5+2) internal loop, the two double-helical stems flanking the internal loop, and the five-nucleotide hairpin loop. It was shown that stem 1 ends in three G-C base pairs (G26-C47, C27-G46, and C28-G45), while stem 2 consists of the two G-C base pairs G34-C42 and G35-C41 (Nottrott et al., 1999; see Figure 1A).

The crystal structure confirms the existence of both stems (Figure 3). However, the crystal structure also shows that stem 2 is extended by two G-A pairs formed by the internal loop nucleotides (G32-A44 and A33-G43). The two G-A base pairs are of the sheared type characterized by hydrogen bonds from the 2-amino group of guanine to N7 of adenine and from the 6-amino group of adenine to N3 of guanine. Between these two tandem G-A base pairs, the helix is strongly overwound, with a twist angle of 81°. This extreme helix twist leads to cross-strand stacking of the two adenines, seen clearly in Figure 3. Both the high twist and a cross-strand stacking of the adenines and guanines are common in tandem sheared G-A base pairs (SantaLucia and Turner, 1993; Heus et al., 1997). In the present RNP complex, only adenines A33 and A44 exhibit perfect cross-strand stacking, while the corresponding guanines G43 and G32 are displaced. Interestingly, the cross-strand stacking of A33 and A44 is continued by a third adenine, A30, which is one of the unpaired internal loop nucleotides; thus, the structure exhibits a three-adenine cross-strand stack, with A44 from the one strand sandwiched between A33 and A30 from the other. In this way stem 2, which has only two Watson-Crick base pairs, is extended on one side by the two sheared G-A base pairs and a further stacked purine, giving an energetically favorable structure that includes a five-purine stack.

Besides the base pairing and base stacking interactions, a network of hydrogen bonds involving several ribose 2'-OH groups further stabilizes the fold of the internal loop (Figure 3C). The 2'-OH group of A44, which has a 2'-endo conformation, is within hydrogen-bonding distance of the N6 of A30, while the N1 of A44 forms a hydrogen bond with the 2'-OH group of A29. Furthermore, the 2'-OH group of A33 forms hydrogen bonds to N3 of G45, and the 2'-OH of the flipped-out U31, which has also a 2'-endo conformation, forms a hydrogen bond with the phosphate group of A30. The 2'-OH groups of both guanosine nucleotides of the tandem G-A base pairs contribute to the hydrogen bond network as well. The G32 exhibits 2'-endo conformation and its 2'-OH group forms a hydrogen bond with N2 of G43. The conformation of the G43 ribose of one of the crystal-

lographically independent RNA molecules is 2'-endo, and its 2'-OH group forms a hydrogen bond with the phosphate of A44. The conformation of the G43 of the other RNA could not be determined unambiguously, since its 2'-OH group is not well defined in the electron density map. However, modeling of the two ribose conformations showed that both conformers could form favorable hydrogen bonds. The 2'-exo conformation of G43 would lead to hydrogen bonds with the N6 of the base pair mate A33 and the N7 of A44, while the 2'-endo conformation would lead to a hydrogen bond with the protein, namely with the carboxylate group of Glu-41 (see below).

A striking feature of the RNA structure is the sharp bend between the two double-stranded stems. An angle between the helical axes of about 65° was calculated. However, this calculation is erroneous, as the stems contain only two and three base pairs respectively (for further details, see Lavery and Sklenar, 1988). Modeling an elongation of both stems with a 10 base pair RNA double helix (A form) gave an angle between the axes of about 30° with some difference between the two RNA molecules of the asymmetric unit. This difference is not reflected in notable changes of the base pairing or base stacking. The superposition of the two crystallographically independent RNA molecules (Figure 3B) reveals that their conformations are most similar close to the flipped-out U31 and in stem 2. In stem 1, there are differences that increase with increasing distance from U31. However, there are differences in several lengths of RNA-RNA and RNA-protein hydrogen bonds, and the pattern of hydrogen bonds around the 2'-OH groups is also affected.

Protein Structure

The 128 amino acid residues of the 15.5kD protein fold into a single, compact globular domain of alternating α helices and β strands, forming an α - β - α sandwich structure (Figure 3A). The central β sheet consists of three antiparallel and one parallel β strand positioned in the order β 1, β 4, β 2, β 3. Helices α 1, α 4, and α 5 pack against one side of the β sheet, while helices α 2 and α 3 are located on the other side. The α - β - α sandwich fold of the 15.5kD protein belongs to the most frequently occurring family of protein folds found in the database SCOP (Murzin et al., 1995), and it is similar to the recently published structure of the ribosomal protein L30 from *Saccharomyces cerevisiae* (Mao and Williamson, 1999), so it is not discussed in detail here. The 15.5kD and L30 proteins share a sequence identity of only 18%, but, as highlighted by the superposition of their structures, their tertiary structures exhibit remarkable similarities and also significant differences (Figure 4). Most of the additional 23 residues extend the N and C termini of the 15.5kD protein. Another striking difference concerns the loop β 2- α 3 and the adjacent helix α 3. In the 15.5kD protein, helix α 3 is much shorter than in L30, because in the former protein Pro-70 interrupts the α -helical pattern. Instead, residues 63-66 in 15.5 kD protein, located where helix α 3 is foreshortened, form a 3_{10} helix (see Figure 4). These changes lead to a displacement of loop β 2- α 3 relative to its position in L30, which is interesting, as precisely these residues are crucial for the recogni-

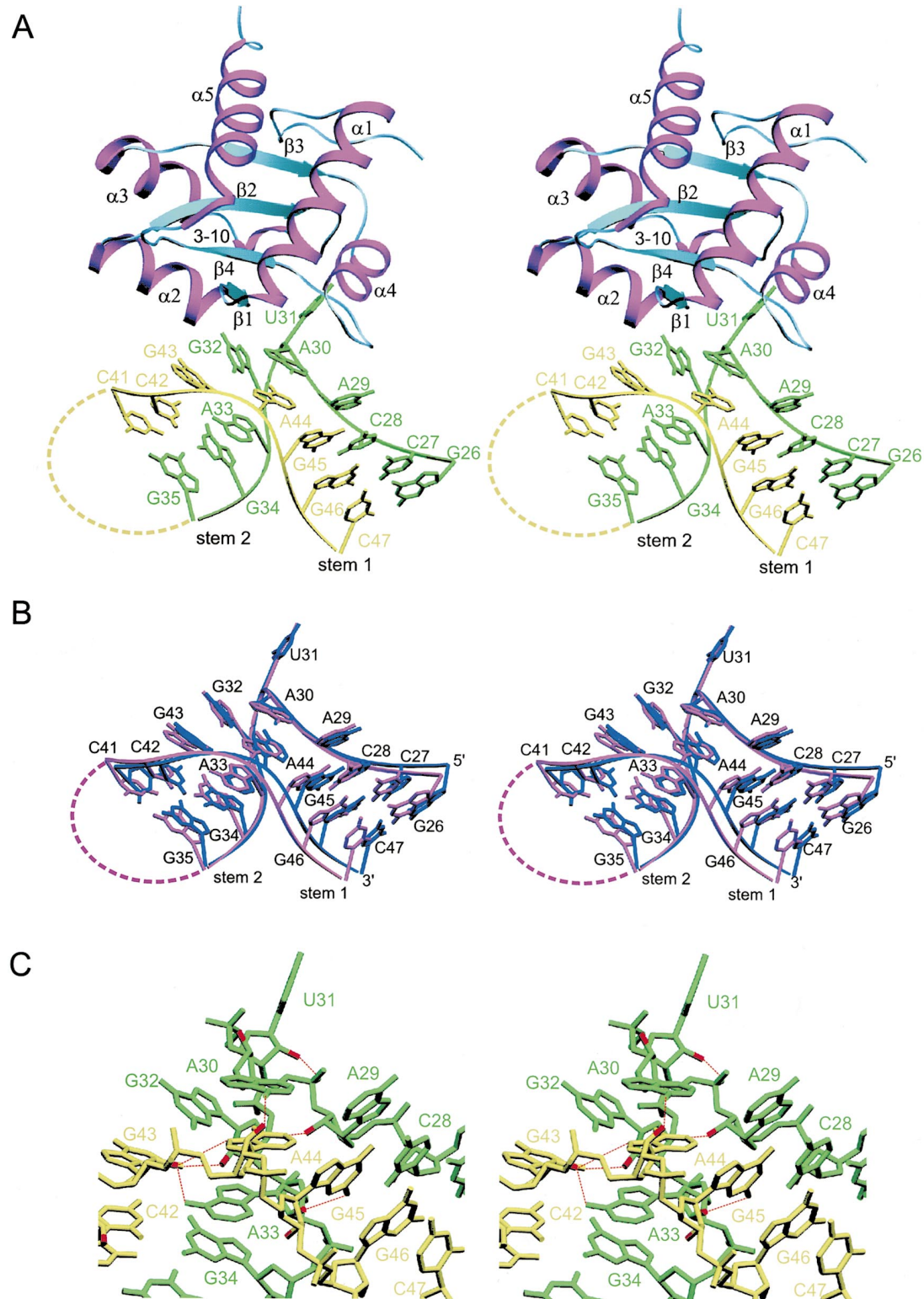


Figure 3. Three-Dimensional Structure of the Spliceosomal 15.5kD-U4 snRNA Complex

(A) Ribbon plot of the 15.5kD protein with the bound fragment of the U4 snRNA 5' stem-loop. The loop-closing nucleotides U36-U40 are not shown, since they were not well defined in the electron density map. The two strands of the two RNA stems are distinguished by the use of yellow and green.

(B) Superposition of the two RNA structures observed in the asymmetric unit of the crystal.

(C) The network of hydrogen bonds (red dotted lines) formed by ribose 2'-OH groups, stabilizing the RNA structure.

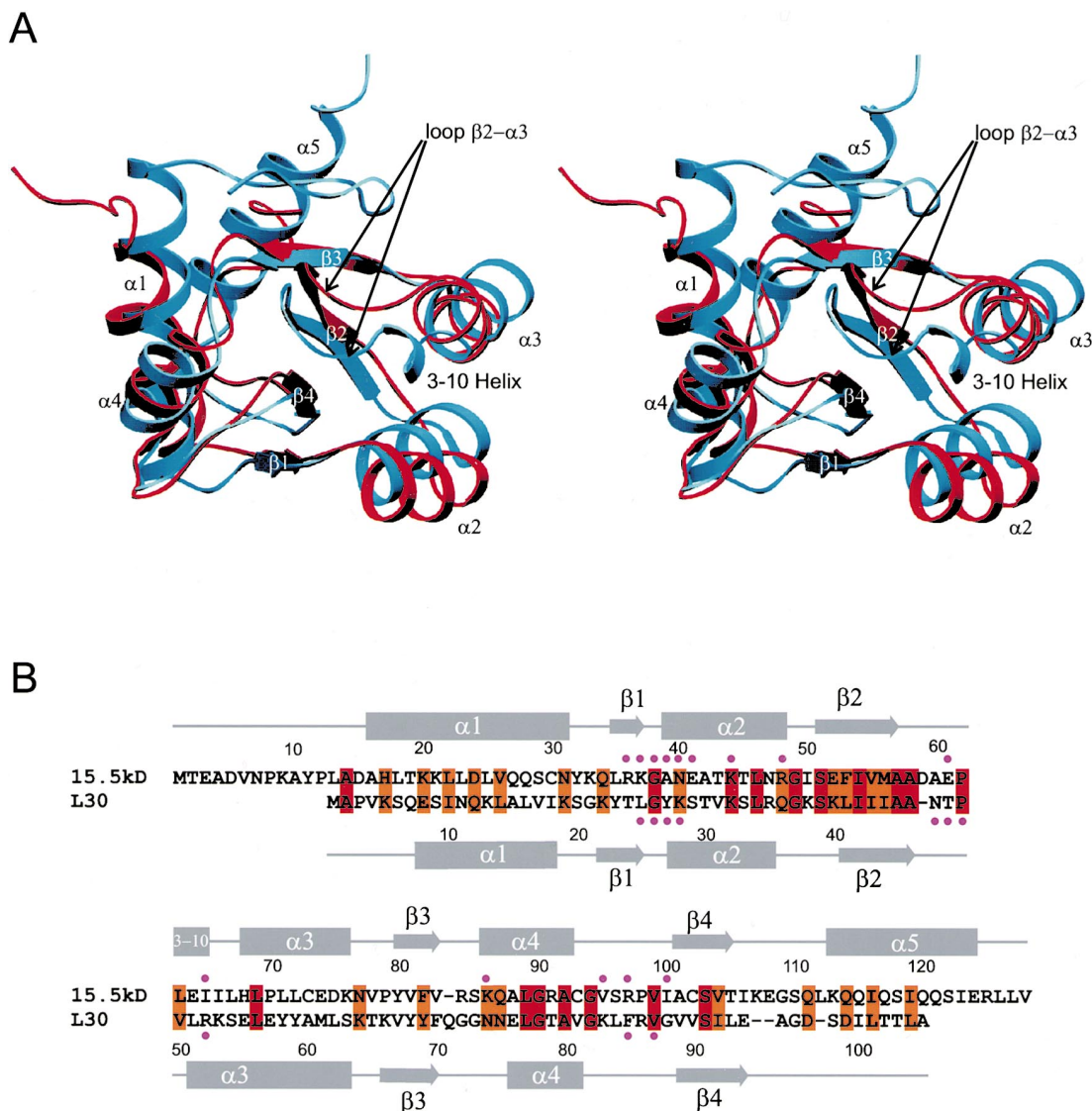


Figure 4. Comparison of the Spliceosomal 15.5kD and the Ribosomal L30 Proteins

(A) Superposition of the 15.5kD protein (blue) and ribosomal L30 protein (red) structures.

(B) Alignment of the amino acid sequences of 15.5kD and L30. Identical residues are highlighted in red, homologous residues in orange. Residues involved in protein-RNA interaction are marked by blue dots. The secondary structures of 15.5kD and L30 were assigned according to PROCHECK (Laskowski et al., 1993).

tion and specific binding of the RNA (see below). Likewise, the position of helix α_2 , which also contains several residues in contact with RNA, differs significantly between the L30 and 15.5kD structures.

Protein-RNA Interaction

In the 15.5kD protein, the RNA binding surface consists of amino acid residues located in two α helices (α_2 and α_4), one β strand (β_1), and three different loops (β_1 - α_2 , β_2 - α_3 , and α_4 - β_4), as indicated in Figure 4B. These residues interact predominantly with the nucleotides of the (5+2) internal loop, and there are also contacts with the sugar-phosphate backbone of stem 2 (Figure 6A). We distinguish four major kinds of interaction between the oligonucleotide and the 15.5kD protein, as follows.

The Flipped-Out Nucleotide U31

U31 is tightly bound in a pocket of the protein formed by Glu-61 and Ile-65, Lys-86, and Ile-100 (Figure 5B). The O4 of U31 forms two hydrogen bonds, with the amino group of Lys-86 and the main chain amide of Glu-61, respectively, and the 3-imino group of U31 forms a hydrogen bond with the main chain oxygen of Glu-61. Furthermore, hydrogen bonds are present between the phosphate group of U31 and the main chain amides of Ala-39 and Ile-100. In addition to these hydrogen bonds, the base of U31 is in van der Waals contact with the hydrophobic side chains of Ile-65 and Ile-100 and the hydrophobic part of the Lys-86 side chain. These multiple contacts obviously explain the earlier observation (Nottrott et al., 1999) that the replacement of U31 with

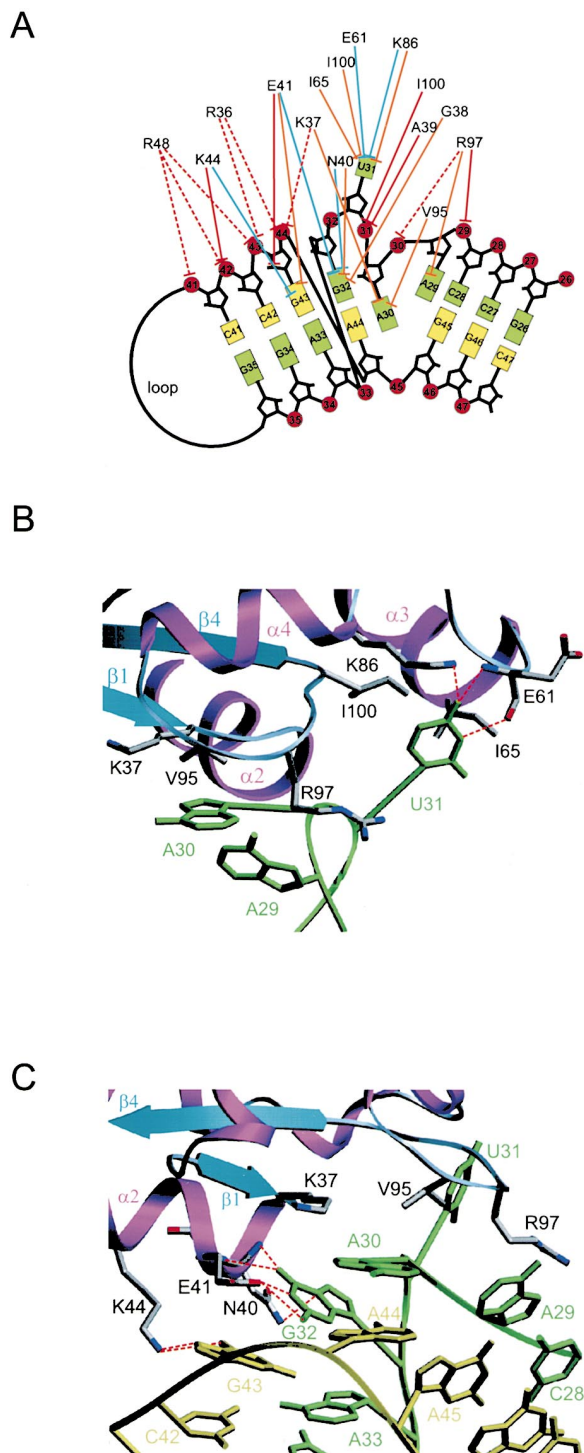


Figure 5. Protein-RNA Interaction

(A) Schematic drawing of the U4 snRNA fragment secondary structure and of the protein-RNA contacts. Bases are colored according to Figure 3. Hydrogen bonds involving bases are colored blue and hydrophobic interactions are orange; hydrogen bonds and electrostatic interactions with phosphates are represented by red and red dashed lines, respectively.

(B) This detail shows the recognition of flipped-out U31 and the hydrophobic interactions of the unpaired nucleotides A29 and A30.

(C) Interaction of the tandem G-A base pairs with the protein. The two guanines stack on each other and form several hydrogen bonds with amino acid residues located in helix α_2 .

any other nucleotide abolishes the 15.5kD-U4 snRNA complex formation in vitro.

The Tandem G-A Base Pairs

Not only are these an important feature of the internal loop structure as described above, but they are also involved in the specific protein-RNA recognition. Both guanines (G32 and G43), which almost stack on each other because of the extreme helical twist, interact with the protein through their exposed atoms in the major groove; in contrast, the two adenines involved in cross-strand stacking (A33 and A44) are averted from the protein. The amino acid residues involved are Asn-40, Glu-41, and Lys-44 (in loop β_1 - α_2 and helix α_2), which bind to the G32 Watson-Crick edge and the G43 Hoogsteen edge (Figure 5C). The carboxylate group and the main chain amide of Glu-41 are within hydrogen bonding distance of N1, N2, and O6 of G32; the ϵ -amino group of Lys-44 is in hydrogen bond distance to N7 and O6 of G43; and the ND2 of Asn-40, which forms a hydrogen bond with the N7 of G32 and the main chain amide of Asn-40, is in hydrogen bond distance with the O6 of G32.

Tandem G-A base pairs are present in many RNAs, and because of their unique structural features and electrostatic properties, they have been anticipated to play a functional part in these RNAs (Gautheret et al., 1994; Chin et al., 1999). The 15.5kD protein-U4 RNA complex represents an example of a sheared tandem G-A mismatch being a pivotal element in specific RNA-protein recognition. Indeed, this tandem G-A mismatch is an absolute prerequisite for 15.5kD protein binding in vitro, as any change in these two bases abolishes the binding in vitro (mutagenesis experiments by Nottrott et al., 1999).

The Stacked, Unpaired Bases A29 and A30

Besides the base-specific contacts, which are mediated mainly by hydrogen bonds, some hydrophobic contacts of bases A29 and A30 with the 15.5kD protein appear to be crucial for the stability of the complex as well. The base of the unpaired nucleotide A29, which stacks on the base pair G45-C28 of stem 1, packs with its other side against the hydrophobic part of the side chain of Arg-97 in loop α_4 - β_4 . Likewise, the unpaired nucleotide A30, which extends the purine stacking of stem 2, packs on its opposite side against a hydrophobic protein surface provided by Lys-37 and Val-95 in loops β_1 - α_2 and α_4 - β_4 , respectively. Both A29 and A30 can be replaced by guanine without any great effect upon protein binding in vitro, and either (but not both at once) can be replaced by a pyrimidine (Nottrott et al., 1999). Since both A29 and A30 are stacked onto other purines and pack against a hydrophobic patch on the protein surface, their function could just as well be performed by a G at this position.

Electrostatic Interactions

The negative charge of the RNA phosphates is neutralized by several basic protein residues (Figure 5A). Lys-44 and Arg-97 are located within hydrogen bonding distance of the phosphates of C42 and A29, respectively. Additionally, Arg-36, Lys-37, and Arg-48 are within 7–8 Å of the RNA and contribute significantly to the overall electrostatic picture. The potentially crucial role of such distant basic residues in RNA-protein binding has been demonstrated for the Rev-RRE complex (Tan and Franke, 1994; Battiste et al., 1996; Ye et al., 1996) and for

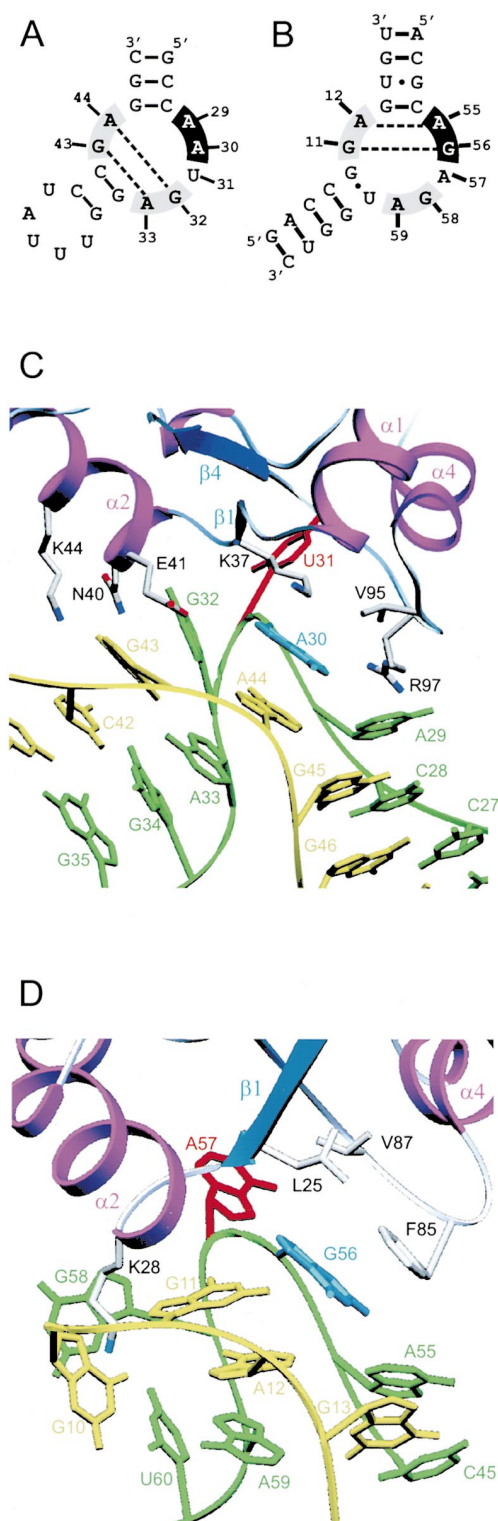


Figure 6. Comparison of the 15.5kD-U4 snRNA and the L30 Pre-mRNA Protein-RNA Complexes

(A and B) RNA sequences and secondary structures of the (5+2) internal loops of U4 snRNA and L30 pre-mRNA according to crystal and NMR structures, respectively.
(C and D) Comparison of the same internal loop 3D structures and their interactions with the protein. Nucleotides that correspond are highlighted in red and blue, respectively.

the complex of a peptide of the protein N of the bacteriophage λ and the box B RNA (Su et al., 1997; Legault et al., 1998).

Discussion

A General Principle of Protein-RNA Binding via an Asymmetric Internal Loop?

The 15.5kD protein belongs to a family of proteins that share a homologous region; this region corresponds to residues 35-90 of the 15.5kD protein and contains most of the residues involved in RNA binding (Koonin et al., 1994; Nottrott et al., 1999 and references therein). Remarkably, most of these proteins are components of large RNP particles, such as the ribosomal proteins L7a, S12, L30, and HS6 from various organisms and the NHP2 from yeast snoRNPs. Open reading frames for some other homologous sequences are found in the database, such as YBXF from *Bacillus subtilis*; however, the corresponding proteins have not yet been characterized.

So far, the only one of these proteins whose three-dimensional structure has been reported is the ribosomal protein L30 from *S. cerevisiae*. This structure has been determined by NMR in a complex between L30 and a fragment of the L30 pre-mRNA (Mao et al., 1999). As pointed out above, the tertiary structure of L30 closely resembles that of the 15.5kD protein (Figure 5A). Apart from this, the two RNP complexes show the following major similarities (Figure 6): (1) both RNAs contain an internal (5+2) loop; (2) both proteins use the same regions of their surfaces for the interaction with the cognate RNA; (3) both proteins bind to the internal loop in the respective RNA; (4) there is considerable similarity between the sequences of the two (5+2) loops, with 5/7 identical bases and only 1/7 purine-pyrimidine exchange, the latter being the base flipped out; (5) the base in the middle of the long strand of the internal loop of L30 pre-mRNA, A57, is flipped out and protrudes into the protein, just like its positional analog U31 in U4 snRNA; (6) the corresponding A30 of U4 snRNA and G56 of L30 pre-mRNA have both a syn and a 2'-endo conformation; (7) the sugar-phosphate backbone forms a hairpin-like turn at this position; (8) the double-helical stems are capped by unpaired bases where the RNA is kinked.

Despite this similarity, there are significant differences, all of which are related to the folding of the (5+2) internal loop: (1) the identities of the purine pairs formed are different, being G-A, G-A in U4 snRNA and G-G, A-A in the L30 pre-mRNA (Figures 6A and 6B); (2) in L30 pre-mRNA, base A59 forms a base triple together with the Watson-Crick base pair G13-C54, whereas the U4 structure has no base triple; (3) the two purine-purine base pairs in the L30 complex are stacked onto the A59-G13-C54 base triple; (4) G58 of L30, which corresponds to G32 in U4 snRNA, is bulged out; (5) the cross-strand adenine sandwich in the U4 complex (A30-A44-A33), absent in the L30 complex, with the top purine, here G56, rotated away from the others and stacks against the aromatic ring of F85; (6) the nucleotides at the bottom and the middle of the three-layer purine stack correspond (A33, A44 in U4 and A59, A12 in L30), but the nucleotide at the top differs (A30 in U4 and G11 in L30).

A further important consequence of the different internal loop fold in L30 pre-mRNA is the relative orientation of the RNA stems with respect to the protein: while stem 2 of U4 RNA makes extensive contacts with several residues of helix $\alpha 2$, the corresponding stem of L30 is turned away from helix $\alpha 2$ and contacts just one residue of helix $\alpha 2$ (Lys-28).

The comparison of the two protein surfaces reveals significant differences concerning both the shape and the electrostatic properties of their RNA binding sites. The structural differences concern the shortened helix $\alpha 3$, the placement of helix $\alpha 2$ and loop $\beta 2$ - $\alpha 3$, and several electrostatically dissimilar amino acid residues at corresponding positions within the protein-RNA interface. One obvious difference concerns the hydrophobic cleft provided by the long side chains of Arg-52 and Lys-28 of the L30 protein that accommodates the flipped-out nucleotide G58, which corresponds to G32 of the 15.5kD protein. The structurally equivalent residues of the 15.5kD protein are Asn-40 and Glu-41, which cannot embrace a flipped-out guanine at that position.

The L30 and the U4 protein-RNA complexes both involve (1) proteins with a particular globular structure and (2) RNA molecules with a purine-rich (5+2) internal loop that (3) surrenders a single base on the longer arm of the internal loop into the protein pocket and then folds into an energetically favorable stacked and hydrogen-bonded structure that causes a bend in the RNA double helix. In spite of the many differences of detail between the structures of the two complexes, both appear to involve folding of RNA according to similar principles. We therefore advance the hypothesis that a general principle is at work, according to which each RNA adopts a stable, optimized folding pattern that is induced when its flipped-out base (U31 or A57) is specifically bound to the protein pocket, whereupon the rest of the internal loop adopts the most stable conformation possible in contact with the protein surface; the stabilization involves hydrogen bonding and purine stacking, and the result is a sharp bend in the double helices meeting at the loop. The hypothesis is supported by the results of NMR studies on the naked L30 pre-mRNA, which clearly demonstrated that the internal loop has no ordered 3D structure in the absence of the L30 protein (Mao et al., 1999). Furthermore, our own biochemical and mutagenesis studies have also indicated that the internal loop of U4 snRNA has no fixed structure in the absence of protein (S.N., K.H., and R.L., unpublished results). It remains to be seen, when further 3D structures of RNP complexes of this kind become known, whether the sequence similarities reflect a general principle of an internal loop binding protein family.

Implications for U4/U6 snRNP and Box C/D snoRNP Structures

An interesting feature of the 15.5kD protein is its presence in several multicomponent RNPs comprising different proteins and different RNAs. These include not only the human U4/U6 snRNP and its homologous variants such as human minor spliceosomes and yeast snRNPs (containing the 15.5kD protein ortholog Snu13p), but also the snoRNPs, which do not take part in splicing but instead participate in the processing and modification of rRNA.

The human U4/U6 snRNP contains, apart from the 15.5kD protein, four other particle-specific proteins: the 20kD, 60kD, 61kD, and 90kD proteins (Horowitz et al., 1997; Lauber et al., 1997; Teigelkamp et al., 1998; O. Makarov, E. Makarov, and R. L., unpublished data). For yeast, the particle-specific protein pattern is similar, except for the absence of an ortholog of the 20kD protein (Gottschalk et al., 1999; Stevens and Abelson, 1999). However, none of these U4/U6 snRNP-specific proteins appears to be capable of stable association with U4 and/or U6 snRNA on its own. Hence, the determinant of binding of these particle-specific proteins must be either the 15.5kD (Snu13p) protein or the common Sm proteins. Therefore, the 15.5kD (Snu13p) protein is a plausible candidate for the function of nucleating the assembly of proteins onto the U4/U6 snRNA. Since the bound RNA covers only approximately 10% of the total solvent exposed surface area of the 15.5kD protein, most of this surface remains accessible for interaction with at least one other protein. In addition, we suggest that one function of the 15.5kD protein is to induce and stabilize the sharp bend of the U4 snRNA 5' stem-loop, which could be required for the association of other proteins during the particle assembly and/or for the function of the U4 snRNA 5' stem-loop in the splicing process.

The 15.5kD protein has been shown to bind to the U4atac snRNA (Nottrott et al., 1999), which is found in minor spliceosomes. Its degree of identity shared with U4 snRNA is very low (~35%). The 5' stem-loop of the U4atac snRNA also contains a (5+2) asymmetric internal loop whose sequence is strictly conserved with respect to the U4 snRNA, although there is some variation in the sequence of both stems 1 and 2 (Figures 1A and 1B). This suggests that the 15.5kD protein-U4atac RNA complex adopts the same three-dimensional structure, since there are no base-specific contacts of the 15.5kD protein with stem 1 or stem 2 (Figures 3 and 6). Therefore, this structure represents a conserved RNP complex of both major and minor spliceosomes. However, it is not known whether this conservation includes the other particle-specific proteins of U4/U6 snRNP as well, since no other particle-specific proteins of U4atac have yet been identified.

The 15.5kD protein is capable of binding other RNAs in other RNPs, as recently demonstrated for the snoRNPs of the box C/D type (Watkins et al., 2000). It must therefore have a considerable versatility in recognition and binding. It was shown to bind specifically the box C/D motifs of various snoRNAs; a major feature of this motif is a (5+2) internal loop with a sequence very similar to that of U4 snRNA (Figures 1A and 1C). In this sequence, all the nucleotides involved in base-specific contacts with the 15.5kD protein are conserved, i.e., the two G-A base pairs and the unpaired flipped-out U base. The two remaining nucleotides of the box C/D internal loop correspond to the unpaired A29 and A30 of U4 snRNA, whose bases pack against hydrophobic patches on the protein surface without forming hydrogen bonds. This suggests that in the complex with the 15.5kD protein, the internal loop of box C/D may adopt a structure similar to that of the internal loop of U4 snRNA. The common RNP core shared between U4 snRNP and box C/D snoRNPs might even extend, as the Prp31p protein

Table 1. Crystallographic Statistics

Data collection				
Data Set	SeMet- λ 1	SeMet- λ 2	SeMet- λ 3	SeMet- λ 4
Resolution (Å)	30.0–2.9	30.0–2.9	30.0–2.9	30.0–2.9
Wavelength (Å)	0.9790	0.9791	0.9184	0.9919
Number of observed reflections	30,157	28,120	28,188	28,438
Number of unique reflections	8,309	8,033	8,039	8,125
Completeness				
Overall (%)	91.5	91.3	91.4	93.2
Last shell (%)	87.0	86.0	86.0	89.3
R_{sym}				
Overall (%)	6.5	5.5	5.8	5.5
Last shell (%)	14.2	12.0	11.8	11.8
Phasing				
PhP centric (iso)	1.13	1.64	0.53	0.0
PhP acentric (iso/ano)	0.17/2.11	2.33/1.74	0.71/1.41	0.0/0.74
Mean FOM (overall)		0.50		
Mean FOM (after solvent flat.)		0.89		
Refinement				
R-factor (%)		21.6		
R_{free} (%)		30.2		
Deviations from ideal geometry				
Bond lengths (Å)		0.007		
Bond angles (deg.)		1.21		

$R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the average intensity for multiple measurements.

PhP, Phasing power is the root-mean-square ($|F_h|/E$), where $|F_h|$ is the heavy atom structure factor amplitude and E is the residual lack of closure error.

FOM, figure of merit.

R-factor = $\sum |F_o - F_c| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

R-free is the cross-validation R-factor calculated for 10% of the reflections omitted in the refinement process.

appears to be homologous to Nop56p and Nop58 from snoRNPs (Watkins et al., 2000).

The three base pairs containing stem 2 (Figure 1C) is highly conserved within the box C/D snoRNA family but different from the corresponding sequence in U4 snRNA. It may thus be a binding site for another protein that binds box C/D snoRNAs specifically, which should be in close contact with the 15.5kD protein. A candidate could be fibrillarin, which was recently shown to bind the box C/D motif (Fatica et al., 2000). Fibrillarin is thought to embody the methyl transferase activity, as suggested by the three-dimensional structure of an archaeobacterial fibrillarin homolog (Wang et al., 2000). The site of modification appears to be defined by the formation of a duplex between snoRNA and pre-rRNA. This duplex starts just after the three base pair duplex of the snoRNA that corresponds to stem 2 of U4 snRNA 5' stem-loop structure (Figure 3). The site of methylation is another four base pairs distant; hence, there is the possibility that the 15.5kD protein is in contact with the snoRNP methyl transferase.

Conclusions

The complex formed between the 15.5kD protein and the fragment of U4 snRNA used in this work shows a protein-RNA interaction with the following features: a stably folded protein binding specifically to a base from a (5+2) internal loop of the snRNA, with consequent formation (or stabilization) of an RNA fold characterized by base stacking and hydrogen bonds (both RNA-RNA and RNA-protein); the RNA fold engenders a sharp bend between the two contiguous RNA stems. We show that the same principle appears to govern RNA-protein bind-

ing in the minor spliceosome and also the central protein-RNA binding at the RNA box C/D in the human snoRNP. A similar pattern has been described independently for a non-snoRNP complex (ribosomal protein L30 with L30 pre-mRNA; Mao et al., 1999). Therefore, we suggest that this structural binding pattern may have some generality for the interaction of other members of the 15.5kD/L30 protein family with their RNA counterparts.

In the assembly of the U4/U6 snoRNP, it seems probable that the binding of the 15.5kD protein to U4/U6 snRNA provides the structural prerequisite for the subsequent attachment of other particle-specific proteins, either through direct binding to the 15.5kD protein or by recognition of the bend it induces in the RNA. Further crystal structures of other, especially larger, complexes will be needed to allow an understanding of the details of the attachment of other particle-specific proteins to the 15.5kD protein-RNA complex.

Experimental Procedures

Protein Purification

Large quantities of the 15.5kD protein were prepared from a GST-15.5kD fusion protein as described by Nottrott et al. (1999). For MAD data collection, a derivative of the 15.5kD protein with the two methionine residues replaced by selenomethionines was prepared by using the methionine auxotrophic strain *Escherichia coli* B834(DE3). A 1 liter culture of B834(DE3)(pLys/pGEX4T2-15.5kD) was grown at 30°C with glucose (4 g l⁻¹) as the carbon source. The medium contained 100 mg l⁻¹ ampicillin and 10 mg l⁻¹ chloramphenicol. It was supplemented with biotin and thiamine (2 mg l⁻¹ each) and L-methionine (50 mg l⁻¹). After the culture had reached an OD₆₀₀ of 0.8, cells were centrifuged down and resuspended in 1 liter of the same medium without methionine. After 1 hr, 50 mg l⁻¹ of DL-

selenomethionine was added. After an additional 30 min, induction of the selenomethionine 15.5kD-GST fusion protein expression was achieved through addition of IPTG to a final concentration of 1 mM. Overexpression was allowed to continue for 4 hr. Purification of the selenomethionine 15.5kD protein was identical to that of the native protein, except that oxidation of selenomethionine was prevented by adding 5 mM DTT to all buffers. A difference in molecular weight of 94 Da between the native and selenomethionine proteins was measured by electrospray mass spectrometry. This corresponds to the incorporation of two Se atoms (data not shown).

RNA Synthesis and Purification

RNA oligonucleotides were obtained commercially from Eurogentec (Belgium) or from the microchemistry laboratory at the IMT (M. Krause, Marburg). The RNA was either purified by anion-exchange HPLC or by HPLC and PAGE. To determine the purity of the RNA the oligoribonucleotides were 5'-end-labeled using 30 pmol of RNA and 60 μ Ci [γ - 32 P]ATP (5000 Ci/mmol; Amersham) in a polynucleotide kinase reaction. Radioactive RNA oligoribonucleotides were analyzed on a 20% polyacrylamide/8 M urea sequence gel and visualized by autoradiography. The RNA oligoribonucleotides were dissolved in $0.5 \times$ TBE with 50% urea, 0.02% bromophenol blue, and 0.02% xylene cyanol, heated for 5 min at 95°C, put on ice, and then applied at 4 μ g RNA per mm² to a 21 cm \times 35 cm \times 0.1 cm, 22% (w/v) polyacrylamide/8 M urea gel in $0.5 \times$ TBE. The gel was run at 10 W for 10 hr. The RNA bands were visualized by ultraviolet shadowing at 312 nm, and the full-length RNA oligoribonucleotide was excised. RNA was eluted from the gel in buffer A (20 mM HEPES/NaOH [pH 7.6], 120 mM NaCl, 2 mM DTT, 0.5% SDS) at room temperature for 8 hr.

The eluted RNA was precipitated in ethanol by standard methods and resuspended in buffer A. The RNA was then extracted with phenol/chloroform and precipitated twice with ethanol. To remove contaminating polyacrylamide fragments, the RNA was redissolved in buffer B (50 mM NaAc [pH 7.0], 1 mM EDTA, 50 mM NaCl) and applied to a DE52 (Whatman) column preequilibrated with buffer B. After washing with buffer B, the RNA was eluted from the column with 50 mM NaAc (pH 7.0), 1 mM EDTA, 800 mM NaCl.

RNA concentration was determined by absorbance at 260 nm. The RNA was then reprecipitated and dissolved at a concentration of 25 μ g μ l⁻¹ in 20 mM HEPES/NaOH (pH 7.6), 120 mM NaCl, 2 mM DTT.

Crystallization

Crystallization experiments were performed at 21°C in Linbro plates by the hanging-drop vapor diffusion technique. For initial cocrystallization experiments, various RNA oligoribonucleotides of different lengths were tested. The best crystals were obtained after 5 days using a protein concentration of 5 mg ml⁻¹ and a protein:RNA molar ratio of 1:2. Protein and RNA were preincubated for 1 hr at 21°C before mixing with an equal volume of reservoir solution (buffer H) containing 0.1 M HEPES/NaOH (pH 7.6), 16.5% PEG-2000, 30 mM MgCl₂, 5 mM DTT, and 0.04% (w/v) sodium azide. The dimensions of the crystals were about 0.40 \times 0.10 \times 0.04 mm.

X-Ray Data Collection

In order to collect diffraction data under cryoconditions, crystals were flash frozen in buffer H containing 12.5% (v/v) ethylene glycol as a cryoprotectant. The crystals were very sensitive to handling and cryobuffer. Alternatively, the crystals were frozen in liquid ethane without the use of a cryoprotectant and subsequently transferred into liquid nitrogen. The crystals were tested on an R-AXIS IV image plate detector mounted on an RU300 rotating-anode generator (Rigaku/MSC) equipped with focusing mirrors and operating at 50 kV and 100 mA. The quality of the crystals turned out to be very heterogeneous. The best crystals diffracted to a resolution of 3.5 Å and were stored in liquid nitrogen for data collection at a synchrotron. All crystals showed a high mosaicity of 1.6°–4.0°.

Structure Determination and Refinement

X-ray data collection was carried out at beamline BM14 of the ESRF (Grenoble), and the data were processed and scaled with DENZO/

SCALEPACK (Otwinowski and Minor, 1997). The data statistics are summarized in Table 1. The crystals belong to space group P2₁2₁2₁, with unit cell dimensions a = 45.8 Å, b = 55.3 Å, c = 146.7 Å, and contain two protein-RNA complexes in the asymmetric unit. One selenomethionine site was found by SOLVE (Terwilliger and Berendzen, 1999), while the N-terminal selenomethionine turned out to be disordered. MAD phases were calculated by SHARP (De La Fortelle and Bricogne, 1997) and subsequently improved by solvent flattening using the implemented program SOLOMON (Abrahams and Leslie, 1996). The model was built using the program O (Jones and Kjeldgaard, 1997), and the structure was refined using X-PLOR (Brünger, 1993). The refinement included several cycles of simulated annealing refinement applying bulk solvent correction and noncrystallographic symmetry restraints. The crystal structure was refined against the λ 4 MAD data set until the working R factor converged at 21.6% and an R_{free} value of 30.2% for data from 30 Å to 2.9 Å. The high R_{free} appears to be a consequence of the crystal quality and resolution limit. The refined model comprises two protein molecules (residues 4–128/6–128) and two RNA molecules (both contain nucleotides 25–35 and 41–47). The quality of the refined crystal structure was analyzed with help of the programs PROCHECK (Laskowski et al., 1993) and X-PLOR. The angle between the double-helical stems of the RNA was calculated with the program CURVES (Lavery and Sklenar, 1988). Figures were generated with the programs O and SETOR (Evans, 1993).

Acknowledgments

We are very grateful to V. Stojanoff (ESRF) and G. Leonard (ESRF) and to our colleague K. Reuter for the support during the MAD data collection at the ESRF beamline BM14. This work was supported by grants from the Deutsche Forschungsgemeinschaft to R. F. (SFB286/TPA11) and to R. L. (SFB397/TPA6 and Gottfried Wilhelm Leibniz Program).

Received September 18, 2000; revised October 17, 2000.

References

- Abrahams, J.P., and Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* 52, 30–42.
- Anthony, J.G., Weidenhammer, E.M., and Woolford, J.L., Jr. (1997). The yeast Prp3 protein is a U4/U6 snRNP protein necessary for integrity of the U4/U6 snRNP and the U4/U6.U5 tri-snRNP. *RNA* 3, 1143–1152.
- Ayadi, L., Callebaut, I., Saguez, C., Villa, T., Mornon, J.P., and Banroques, J. (1998). Functional and structural characterization of the prp3 binding domain of the yeast prp4 splicing factor. *J. Mol. Biol.* 284, 673–687.
- Banroques, J., and Abelson, J.N. (1989). PRP4: a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* 9, 3710–3719.
- Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D., and Williamson, J.R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* 273, 1547–1551.
- Bjorn, S.P., Soltys, A., Beggs, J.D., and Friesen, J.D. (1989). PRP4 (RNA4) from *Saccharomyces cerevisiae*: its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* 9, 3698–3709.
- Bordonné, R., Banroques, J., Abelson, J., and Guthrie, C. (1990). Domains of yeast U4 spliceosomal RNA required for PRP4 protein binding, snRNP-snRNP interactions, and pre-mRNA splicing in vivo. *Genes Dev.* 4, 1185–1196.
- Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lührmann, R. (1984). Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J.* 3, 1357–1363.

- Brow, D.A., and Guthrie, C. (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334, 213–218.
- Brünger, A.T. (1993). X-PLOR version 3.1 (New Haven, CT: Yale Univ. Press).
- Burge, C.B., Tuschl, T.H., and Sharp, P.A. (1999). Splicing of precursors to mRNAs by the spliceosome. In *The RNA World*, R.F. Gesteland, T. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 525–560.
- Chin, K., Sharp, K.A., Honig, B., and Pyle, A.M. (1999). Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nat. Struct. Biol.* 6, 1055–1161.
- De La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* 276, 472–494.
- Evans, S.V. (1993). SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* 134, 127–128.
- Fatica, A., Galardi, S., Altieri, F., and Bozzoni, I. (2000). Fibrillarin binds directly and specifically to U16 box C/D snoRNA. *RNA* 6, 88–95.
- Gautheret, D., Konings, D., and Gutell, R.R. (1994). A major family of motifs involving G-A mismatches in ribosomal RNA. *J. Mol. Biol.* 242, 1–8.
- Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Lührmann, R., and Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J.* 18, 4535–4548.
- Hashimoto, C., and Steitz, J.A. (1984). U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* 12, 3283–3293.
- Heus, H.A., Wijmenga, S.S., Hoppe, H., and Hilbers, C.W. (1997). The detailed structure of tandem G-A mismatched base-pair motifs in RNA duplexes is context dependent. *J. Mol. Biol.* 271, 147–158.
- Horowitz, D.S., Kobayashi, R., and Krainer, A.R. (1997). A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA* 3, 1374–1387.
- Hu, J., Xu, D., Schappert, K., Xu, Y., and Friesen, J.D. (1995). Mutational analysis of *Saccharomyces cerevisiae* U4 small nuclear RNA identifies functionally important domains. *Mol. Cell. Biol.* 15, 1274–1285.
- Jones, T.A., and Kjeldgaard, M. (1997). Electron-density map interpretation. *Methods Enzymol.* 277, 173–208.
- Koonin, E.V., Bork, P., and Sander, C. (1994). A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Res.* 22, 2166–2167.
- Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* 65, 367–409.
- Lamond, A.I., Konarska, M.M., Grabowski, P.J., and Sharp, P.A. (1988). Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proc. Natl. Acad. Sci. USA* 85, 411–415.
- Laskowski, R.A., McArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
- Laub, J., Plessel, G., Prehn, S., Will, C.L., Fabrizio, P., Groning, K., Lane, W.S., and Lührmann, R. (1997). The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA* 3, 926–941.
- Lavery, R., and Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6, 63–91.
- Legault, P., Li, J., Mogridge, J., Kay, L.E., and Greenblatt, J. (1998). NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* 93, 289–299.
- Mao, H., and Williamson, J.R. (1999). Local folding coupled to RNA binding in the yeast ribosomal protein L30. *J. Mol. Biol.* 292, 345–359.
- Mao, H., White, S.A., and Williamson, J.R. (1999). A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat. Struct. Biol.* 6, 1139–1147.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.
- Nilsen, T.W. (1998). RNA-RNA interactions in nuclear pre-mRNA splicing. In *RNA Structure and Function*, R.W. Simons and M. Grunberg-Manago, eds. (Cold Spring Harbor Press, Cold Spring Harbor, NY), pp. 279–307.
- Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Lührmann, R. (1999). Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J.* 18, 6119–6133.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- Rinke, J., Appel, B., Digweed, M., and Lührmann, R. (1985). Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. *J. Mol. Biol.* 185, 721–731.
- SantaLucia, J., Jr., and Turner, D.H. (1993). Structure of (rGCGAGCC)2 in solution from NMR and restrained molecular dynamics. *Biochemistry* 32, 12612–12623.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315–326.
- Stevens, S.W., and Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc. Natl. Acad. Sci. USA* 96, 7226–7231.
- Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeets, L.A., and Weiss, M.A. (1997). RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. *Biochemistry* 36, 12722–12732.
- Tan, R., and Frankel, A.D. (1994). Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry* 33, 14579–14585.
- Tarn, W.Y., and Steitz, J.A. (1996). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* 273, 1824–1832.
- Teigelkamp, S., Achsel, T., Mundt, C., Gotherl, S.F., Cronshagen, U., Lane, W.S., Marahiel, M., and Lührmann, R. (1998). The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA* 4, 127–141.
- Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. D* 55, 849–861.
- Vankan, P., McGuigan, C., and Mattaj, I.W. (1992). Roles of U4 and U6 snRNAs in the assembly of splicing complexes. *EMBO J.* 11, 335–343.
- Wang, H., Boisvert, D., Kim, K.K., Kim, R., and Kim, S.H. (2000). Crystal structure of a fibrillarin homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *EMBO J.* 19, 317–323.
- Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., and Lührmann, R. (2000). A common core RNP structure shared between the small nuclear box C/D RNPs and the spliceosomal U4 snRNP. *Cell* 103, 457–466.
- Weidenhammer, E.M., Ruiz-Noriega, M., and Woolford, J.L., Jr. (1997). Prp31p promotes the association of the U4/U6 × U5 tri-snRNP with prespliceosomes to form spliceosomes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 17, 3580–3588.
- Wersig, C., and Bindereif, A. (1992). Reconstitution of functional mammalian U4 small nuclear ribonucleoprotein: Sm protein binding is not essential for splicing in vitro. *Mol. Cell. Biol.* 12, 1460–1468.
- Will, C.L., and Lührmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 9, 320–328.
- Xu, Y., Petersen-Bjorn, S., and Friesen, J.D. (1990). The PRP4 (RNA4) protein of *Saccharomyces cerevisiae* is associated with the 5' portion of the U4 small nuclear RNA. *Mol. Cell. Biol.* 10, 1217–1225.

Ye, X., Gorin, A., Ellington, A.D., and Patel, D.J. (1996). Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nat. Struct. Biol.* **3**, 1026–1033.

Yean, S.L., and Lin, R.J. (1991). U4 small nuclear RNA dissociates from a yeast spliceosome and does not participate in the subsequent splicing reaction. *Mol. Cell. Biol.* **11**, 5571–5577.

Yu, Y.T., and Steitz, J.A. (1997). Site-specific crosslinking of mammalian U11 and U6atac to the 5' splice site of an AT-AC intron. *Proc. Natl. Acad. Sci. USA* **94**, 6030–6035.

Protein Data Bank ID Code

Coordinates have been deposited with the Protein Data Bank under ID code 1E7K.