RNA (2002), *8*:1308–1318. Cambridge University Press. Printed in the USA.

DOI: 10.1017/S1355838202020034

The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: Identification of SBP2 amino acids important to SECIS RNA binding

CHRISTINE ALLMANG, PHILIPPE CARBON, and ALAIN KROL

Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance,
Unité Propre de Recherche 9002 du Centre National de la Recherche Scientifique–Université Louis Pasteur,
Institut de Biologie Moléculaire et Cellulaire, 67084 Strasbourg Cedex, France

ABSTRACT

Selenoprotein synthesis in eukaryotes requires the selenocysteine insertion sequence (SECIS) RNA, a hairpin in the 3′ untranslated region of selenoprotein mRNAs. The SECIS RNA is recognized by the SECIS-binding protein 2 (SBP2), which is a key player in this specialized translation machinery. The objective of this work was to obtain structural insight into the SBP2-SECIS RNA complex. Multiple sequence alignment revealed that SBP2 and the U4 snRNA-binding protein 15.5 kD/Snu13p share the same RNA binding domain of the L7A/L30 family, also found in the box H/ACA snoRNP protein Nhp2p and several ribosomal proteins. In corollary, we have detected a similar secondary structure motif in the SECIS and U4 RNAs. Combining the data of the crystal structure of the 15.5 kD-U4 snRNA complex, and the SBP2/15.5 kD sequence similarities, we designed a structure-guided strategy predicting 12 SBP2 amino acids that should be critical for SECIS RNA binding. Alanine substitution of these amino acids followed by gel shift assays of the SBP2 mutant proteins identified four residues whose mutation severely diminished or abolished SECIS RNA binding, the other eight provoking intermediate down effects. In addition to identifying key amino acids for SECIS recognition by SBP2, our findings led to the proposal that some of the recognition principles governing the 15.5 kD-U4 snRNA interaction must be similar in the SBP2-SECIS RNA complex.

Keywords: L7A/L30 RNA binding domain; RNA–protein interactions; SECIS-binding protein 2; selenocysteine; U4 snRNA

INTRODUCTION

Selenium is mostly found in the active site of selenoproteins, in the form of the amino acid selenocysteine. In mammals, selenoproteins participate in several glutathione- or thioredoxin-dependent oxidation–reduction reactions, or in the maturation of the thyroid hormone (reviewed in Köhrle et al., 2000; Gladyshev & Kryukov, 2001). The importance of selenium and selenoproteins was further underscored by two recent discoveries. The first one refers to the capital roles for sperm maturation of the phospholipid hydroperoxide glutathione peroxidase (Ursini et al., 1999) and protamine thiol crosslinking glutathione peroxidase (Pfeifer et al., 2001), two splice variants of the same pre-mRNA. It is remarkable that these findings provided the molecular basis for earlier observations linking selenium deficiencies and male infertility. The second discovery is that patients developing a form of congenital muscular dystrophy carry mutations in the gene encoding selenoprotein SePN1 (Moghadaszadeh et al., 2001). This finding constituted the first report establishing a direct correlation between the occurrence of a genetic disease and mutations in a selenoprotein gene. Eukaryotic selenocysteine biosynthesis and cotranslational incorporation in response to a redefined UGA Sec codon are achieved by a complex molecular machinery containing RNA and protein partners (reviewed in Fagegaltier et al., 2001; Lescure et al., 2002b). This amino acid is synthesized from the seryl residue of the Ser-tRNA^{Sec}, generating the Sec-tRNA^{Sec} that is loaded onto the selenocysteine-specialized translation elongation factor mSelB/eEFsec (Fagegaltier et al., 2000a; Tujebajeva et al., 2000). Decoding of UGA Sec codons necessitates not only the presence of this elongation factor but also the SECIS element, an RNA hairpin in

Reprint requests to: Alain Krol, Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, Unité Propre de Recherche 9002 du Centre National de la Recherche Scientifique-Université Louis Pasteur, Institut de Biologie Moléculaire et Cellulaire, 15 Rue René Descartes, 67084 Strasbourg Cedex, France; e-mail: A.Krol@ibmc.u-strasbg.fr.

the 3′ UTR of selenoprotein mRNAs (Berry et al., 1991). Structure–function studies proposed secondary and three-dimensional structure models for the SECIS element (Walczak et al., 1996, 1998; Martin et al., 1998; Grundner-Culemann et al., 1999; Fagegaltier et al., 2000b). The core of the hairpin consists of a quartet of non-Watson–Crick base pairs containing a tandem of sheared G-A base pairs that are pivotal for mediating UGA Sec decoding (Walczak et al., 1996, 1998). SBP2, the SECIS binding protein 2, interacts with the SECIS element (Copeland et al., 2000; Lescure et al., 2002a) and most likely with mSelB/eEFsec (Fagegaltier et al., 2000a; Tujebajeva et al., 2000). From these and other functional data (Low et al., 2000), it is obvious that SBP2 is a key player in the machinery. Two major studies were previously undertaken to delineate the SECIS RNA and SBP2 domains important for the interaction. In the first one, structural investigations of the SECIS RNA-SBP2 complex revealed that the phosphate backbone and the non-Watson–Crick base pairs at the core of the SECIS RNA are important features governing the interaction (Fletcher et al., 2001). The other study dealt with the functional dissection of SBP2. It was discovered that it belongs to the family of proteins containing the L7A/L30 RNA-binding domain (Copeland et al., 2001). This domain comprises several ribosomal proteins of the large and small subunits, Nhp2p that is the core component of the yeast H/ACA family of small nucleolar ribonucleoprotein particles (Henras et al., 1998), and the eRF1 subunit of the translation termination release factor. Interestingly, the existence of such an RNA-binding domain was hypothesized several years ago, based on amino acid sequence comparisons of the limited number of proteins available at the time (Koonin et al., 1994).

An extensive study of the amino acids required for the binding of SBP2 to the SECIS RNA has not been published yet. The issue is especially crucial because the various proteins of the L7A/L30 family can specifically recognize their cognate RNA yet share identical or similar sequences in their homologous RNA-binding domains. The objective of the work reported here was precisely to identify amino acids in the RNA-binding domain of SBP2 that are important for recognition of the SECIS RNA. The strategy that was taken stemmed from our two initial findings described in this report: (1) the RNA-binding domain of SBP2 displays amino acid sequence identity to another member of the L7A/L30 family, the human 15.5 kD protein (ortholog of the yeast Snu13p) that binds the 5′ stem-loop of spliceosomal U4 snRNAs but also box C/D snoRNAs (Nottrott et al., 1999; Gottschalk et al., 1999; Stevens & Abelson, 1999; Watkins et al., 2000); (2) the SECIS RNA and the 5′ stem-loop of U4 snRNA possess common structural features. Combining the data of the crystal structure of the 15.5 kD-U4 snRNA complex (Vidovic et al., 2000) and the sequence alignment between the 15.5 kD and SBP2 proteins, we designed a structure-guided strategy to identify SBP2 amino acids that should be important for the SECIS RNA interaction. The prediction was tested in the human SBP2 by alanine substitution of the relevant amino acids followed by RNA binding assays of the SBP2 mutant proteins. This enabled the identification of amino acids critical for the SBP2-SECIS RNA interaction.

RESULTS

The RNA-binding domain of SBP2 and spliceosomal 15.5 kD/Snu13p proteins exhibits striking sequence similarities

To identify amino acids conserved in the RNA-binding domain (RBD) of various SBP2 and that could be involved in SECIS RNA interaction, databases were searched for SBP2 sequences from distantly related species. Various attempts were carried out to minimize the many hits engendered by ribosomal proteins possessing the L7A/L30 RBD. The best procedure for discarding ribosomal protein sequences was to perform Blastp searches of the nonredundant database with a 84-amino-acid-long subdomain of the human SBP2 RBD encompassing residues 673–756, and not with the entire domain. This 84-amino-acid sequence was obtained after proceeding by trial and error with several overlapping sequences of the hSBP2 RBD, seeking the largest sequence that did not match ribosomal proteins. Two hits, which were not included in a previously reported sequence alignment (Copeland et al., 2001), drew our attention: They corresponded to the human spliceosomal 15.5 kD protein (Nottrott et al., 1999) and its Snu13p ortholog in yeast (Gottschalk et al., 1999; Stevens & Abelson, 1999). This incited us to obtain more information on the degree of sequence similarity between SBP2, the 15.5 kD protein, and other members of the L7A/L30 family. A multiple sequence alignment was performed between the human SBP2 (hSBP2), 15.5 kD, Snu13p, Nhp2p, yeast ribosomal protein L30 (yRPL30), and human ribosomal protein L7A (hRPL7A). Figure 1 shows the region of maximum homology that was obtained between 79 amino acids of the hSBP2 RBD (positions 672–750) and the RBDs of the other proteins. From the alignment, we found that hSBP2 and 15.5 kD/Snu13p possess 47% amino acid similarity (26% identity) over the homologous sections. The similarity between hSBP2 and Nhp2p is 43% (20% identity), the value dropping to 30% (16% identity) with yRPL30 and hRPL7A. Identical results were obtained when the sequence of the rat SBP2 RBD was used in the alignment (data not shown). Thus, the RBD sequences in the mammalian SBP2 and 15.5 kD/Snu13p are closer to each other than to other members of the L7A/L30 family.

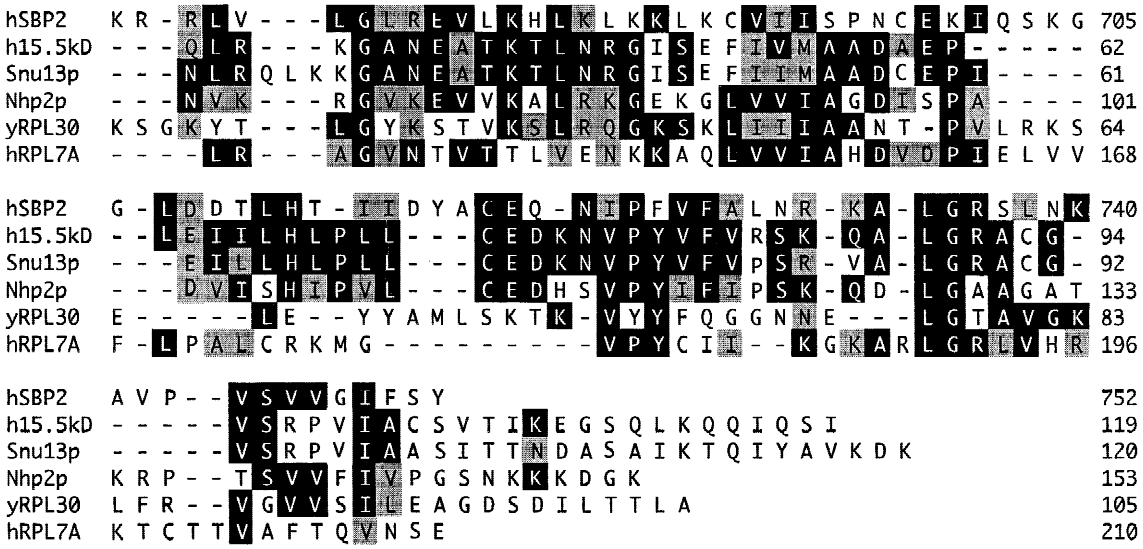

FIGURE 1. Multiple sequence alignment of the RNA binding domain of human SBP2 (hSBP2), human 15.5 kD (h15.5 kD), yeast Snu13p (Snu13p), yeast Nhp2p (Nhp2p), yeast ribosomal protein L30 (yRPL30), and human ribosomal protein L7A (hRPL7A). The alignment was made with ClustalW and manually refined with MegAlign (DNASTAR). Identical amino acids are shown in reverse, similar residues are shaded in gray. The sequences are from: hSBP2 (Lescure et al., 2002a); Snu13p, accession number NP010888; Nhp2p (Henras et al., 1998); yRPL30 (Mao et al., 1999); hRPL7A, accession number AAH05128.

Similar structural features in the SECIS RNA and 5′ stem-loop of spliceosomal U4 snRNA

We next asked whether the sequence conservation of the hSBP2 and 15.5 kD/Snu13p RBD correlates with structural features that could be shared by the SECIS and U4 RNA targets. Experimental secondary structure models for a variety of SECIS RNAs (Walczak et al., 1996, 1998; Fagegaltier et al., 2000b; reviewed in Krol, 2002) proposed that the core of the SECIS RNA is formed by four consecutive non-Watson–Crick base pairs containing the invariant tandem of G3-A8/G7-A4 sheared base pairs (Fig. 2). Indeed, this quartet of base pairs represents an important functional motif for selenoprotein synthesis and a critical recognition site for SBP2 (Walczak et al., 1998; Fletcher et al., 2001). Striking similarities were detected in the core structures of the SECIS RNA and the 5′ stem-loop of U4 snRNA (Fig. 2): Helices 1 and 2 are separated by an asymmetrical internal loop; helix 2 contains a tandem of sheared G-A base pairs shown to be the major functional motif of the SECIS RNA (Walczak et al., 1998). Whereas the size of the internal loop is invariant in U4 snRNA, it is variable in the different SECIS RNAs. However, despite this difference, it is remarkable that the similar sequences R1U2 (SECIS RNA) and A30U31 (U4 snRNA) reside 5′ to the G3-A8 and G32-A44 base pairs in the SECIS and U4 RNAs, respectively (Fig. 2). In the crystal structure of the 15.5 kD-U4 snRNA complex, U31 is flipped out (Vidovic et al., 2000), whereas our structure probing experiments favored the U2-N9 base pairing in the SECIS RNA (Walczak et al., 1996). Worth noting were the findings that substitutions of U2 in the SECIS RNA and U31 in U4 snRNA, or those aiming at debilitating the sheared G-A base pairs in both RNAs, compromised the in vitro binding of SBP2 and 15.5 kD to their cognate RNAs (Nottrott et al., 1999; Fletcher et al., 2001). As reported by Watkins et al. (2000), Vidovic et al. (2000), and Klein et al. (2001), it is very likely that the internal loop of box C/D snoRNAs adopts the same asymmetrical structure as in U4 snRNA (Fig. 2). Thus, the U4 snRNA/box C/D snoRNAs and the SECIS RNA possess similarities in their core structures interacting with the 15.5 kD/Snu13p and SBP2 proteins, respectively.

Structure-guided prediction of SBP2 amino acids involved in the interaction with the SECIS RNA

In a further step, we reasoned that the sequence similarities between the hSBP2 and 15.5 kD RBDs and the common structural features in the SECIS RNA and U4 snRNA could be exploited to identify hSBP2 amino acids contacting the SECIS RNA. We first tested the ability of the hSBP2 RBD to fold into a similar domain structure as the 15.5 kD protein by secondary structure predictions using the PHDSec program (Rost & Sanders, 1993). Predictions schematized in Figure 3A reveal striking similarities with the secondary structure of the 15.5 kD. Differences occur at the edges of the SBP2 RBD, which is not surprising, as the RBD only represents one domain of the 854-amino-acid full-length

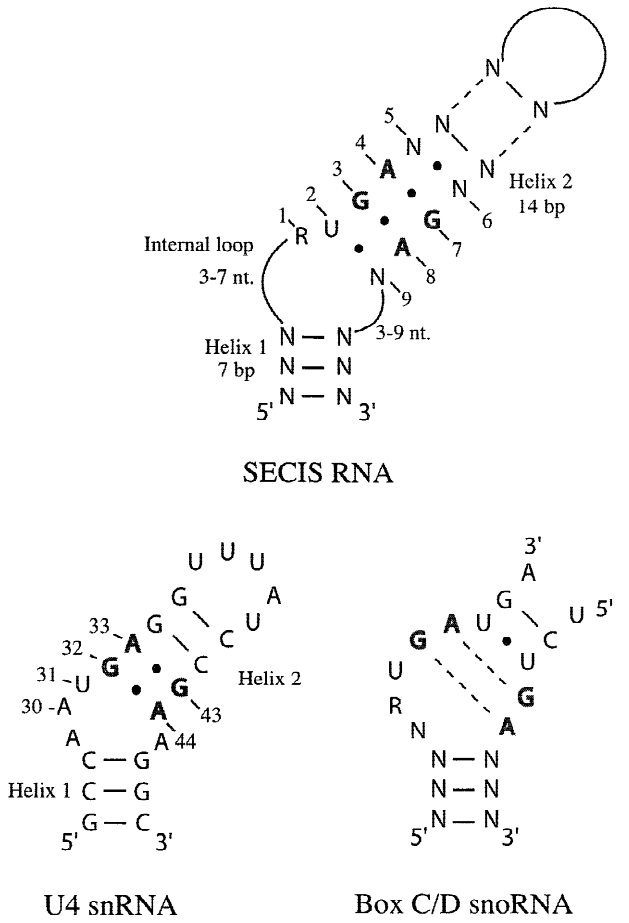

FIGURE 2. Secondary structure models displaying the similar features between the consensus SECIS RNA, the 5′ stem-loop of the human U4 small nuclear RNA, and the consensus Box C/D small nucleolar RNAs. The structures were adapted from Walczak et al. (1996), Vidovic et al. (2000), Klein et al. (2001), and Krol (2002). Numbering of the consensus SECIS RNA sequence started arbitrarily at R1 to position the base pairing partners at the non-Watson–Crick quartet; only a portion of the SECIS helices 1 and 2 is depicted. Sheared G-A base pairs are in bold; the putative G–A base pairs in Box C/D snoRNA are represented by dashed lines. R stands for A or G, N for any nucleotide.

protein. Another difference concerns the β2-α3 junction (see Fig. 3A) where helix α3 is predicted to be slightly extended in hSBP2, resembling more the ribosomal protein L30 in this respect (Mao et al., 1999). The good overall conservation of the secondary structure elements between the 15.5 kD and hSBP2 proteins suggests that the three-dimensional folding and the positioning of amino acids involved in RNA binding are likely to be similar in the two proteins. Having established this, we examined the crystal structure of the 15.5 kD-U4 snRNA complex. It revealed that 14 amino acids in the 15.5 kD RBD participate in the interaction with U4 snRNA (Vidovic et al., 2000). They are marked by dots above the 15.5 kD sequence (Fig. 3A) and the 15.5 kD-U4 snRNA interactions are represented in Figures 3B and 4A. We hypothesized that the homologous hSBP2 residues (Fig. 3A) could fulfill similar roles in the hSBP2-SECIS RNA complex. We therefore proposed the putative interaction scheme (Fig. 4B) in which: Gly676^{SBP2}, Arg678^{SBP2}, Glu679^{SBP2}, and Lys682^{SBP2} could contact the bases or the phosphodiester backbone of the SECIS RNA at G3 and/or G7; Leu677^{SBP2}, Glu699^{SBP2}, Asp709^{SBP2}, Arg731^{SBP2}, and Ile749^{SBP2} could interact with U2; and Val744^{SBP2} could interact with A1. To test the hypothesis, we made the corresponding alanine replacements and assayed the abilities of the mutant proteins to bind the SECIS RNA. Additionally, Lys732^{SBP2} was substituted to determine whether Arg731^{SBP2} or Lys732^{SBP2} is homologous to Lys86$^{15.5}$. Ser745^{SBP2} was mutated because it resides within a block of conserved sequences found only in nonribosomal proteins (see Fig. 1). In the 15.5 kD-U4 snRNA complex, Arg36$^{15.5}$, Lys37$^{15.5}$, and Arg48$^{15.5}$ contribute essentially to electrostatic interactions with the phosphates at positions 41–44 in U4 snRNA. Because the corresponding residues Val674^{SBP2}, Leu675^{SBP2}, and Leu686^{SBP2} are hydrophobic, their interaction with the SECIS RNA was hardly predictable and they were not mutated. Likewise, Val746^{SBP2} was not substituted because its Arg97$^{15.5}$ counterpart interacts with A29 in U4 snRNA, a nucleotide that has no identified homolog in the internal loop of the SECIS RNA. In summary, 12 amino acids were substituted and are represented in Figure 3A. The mutations were engineered in the hSBP2/512 cDNA, a construct that encodes the C-terminal 512 amino acids containing the RBD of the protein and that was shown previously to display SECIS RNA binding activity in vitro (Lescure et al., 2002a). This protein will be considered as the wild-type (wt) hSBP2.

Identification of SBP2 residues important for the interaction with the SECIS RNA

The [^{35}S]-methionine-labeled hSBP2 proteins used in this study were generated by in vitro coupled transcription/translation in rabbit reticulocyte lysates. This system offers the advantage of containing limiting amounts of endogenous SBP2 (Copeland et al., 2000) that will not interfere with the assay, rendering it suitable for studying the effects of the hSBP2 mutations. The translation efficiencies of the wild-type and mutant hSBP2 proteins were verified and quantitated by gel electrophoresis (data not shown) and their abilities to bind the [^{32}P]-labeled human SePN1 SECIS RNA (Fagegaltier et al., 2000b) were assessed by electrophoretic mobility shift assays (Fig. 5A, B). As anticipated, no SBP2-SECIS RNA complex could form with the unprogrammed reticulocyte lysate in which SBP2 is limiting (Fig. 5A, B, lanes 2). The band marked by an asterisk, appearing also in the other lanes, corresponds to another SECIS RNA–protein complex that we previously characterized (Hubert et al., 1996). It contains a SECIS-binding protein that differs from SBP2 and does not

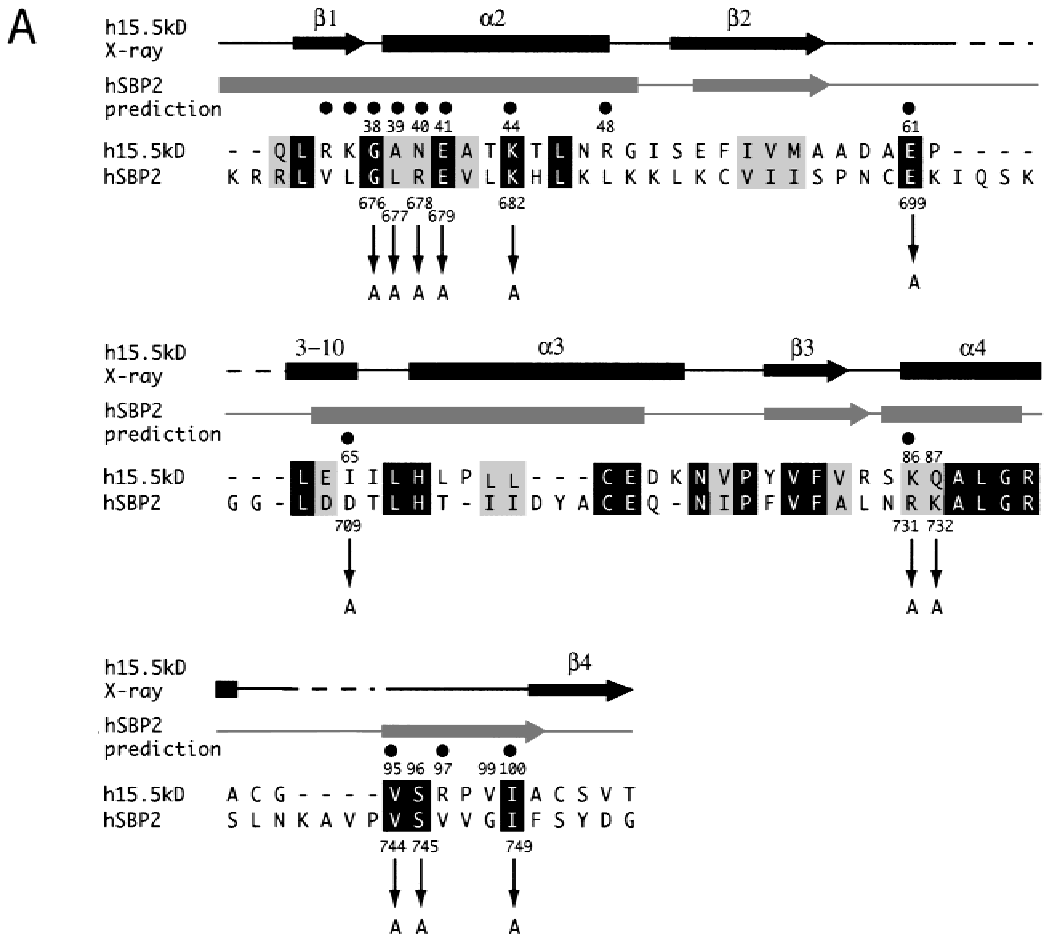

FIGURE 3. *See caption on facing page.*

share the same binding site on the SECIS RNA. Addition of the in vitro translated wild-type hSBP2 to the SECIS RNA led to the formation of two retarded complexes containing monomeric and homodimeric forms of hSBP2 (Fig. 5A, B, lanes 3 and 4), as previously reported for the recombinant hSBP2 produced in *Escherichia coli* (Lescure et al., 2002a). The yield of the monomeric and homodimeric forms of complexes was 24% and 9%, respectively.

The RNA binding activities of the hSBP2 mutants are shown in Figure 5A,B, lanes 5–12 and 5–20, respectively, and quantitated in Table 1. All the mutants affected hSBP2 binding to various extents, strongly suggesting that the residues designed by the structure-guided strategy contribute to SECIS RNA binding. Identical results were obtained with the SECIS RNA of the rat glutathione peroxidase mRNA (data not shown). The most drastic effects were produced by E699A and R731A and led to a complete or almost complete (E699A) loss of RNA recognition (Fig. 5B, lanes 7, 8 and 11, 12). Interestingly, the homologous amino acids Glu61$^{15.5}$ and Lys86$^{15.5}$ are the only two residues establishing hydrogen bonds with the bulged U31 base in U4 snRNA (see Figs. 3B and 4A). The G676A and E679A mutations were severely deleterious to SECIS RNA binding, entailing 19–28% of residual binding activity (Fig. 5A, lanes 5, 6 and 9, 10). The homologous residues Gly38$^{15.5}$ and Glu41$^{15.5}$ contact the sheared G-A base pairs of U4 snRNA at G32 and G43, respectively (Figs. 3B and 4A). The deleterious effects of E699A, R731A, G676A, and E679A did not originate from a subsequent loss of protein solubility because we could establish that the four mutant proteins are still soluble when expressed in *E. coli* BL21 (DE3) RIL (data not shown). Moderate effects for the other eight substitutions were observed. In this regard, the result of the K732A mutation strengthens the prediction that Arg731^{SBP2}, rather than Lys732^{SBP2}, is the homolog of Lys86$^{15.5}$. S745A provoked a drop of about 50% in the RNA binding activity. Surprisingly, the R678A mutation had a rather benign effect, whereas we anticipated it to be more harmful as the homologous Asn40$^{15.5}$ residue establishes hydrophobic and hydrogen bond contacts in U4 snRNA with G32 at the sheared G32-A44 (Figs. 3B

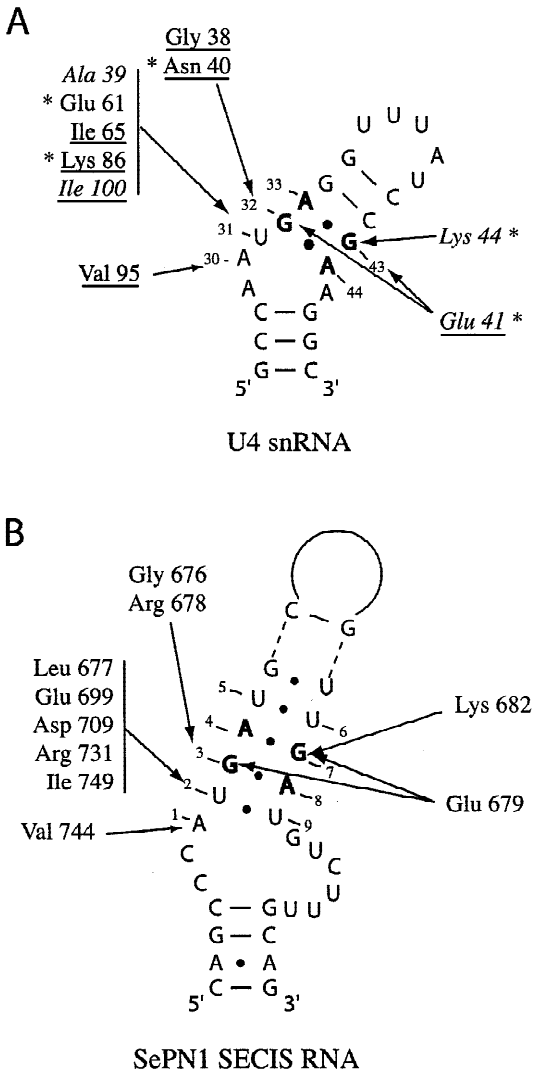

FIGURE 4. Scheme of RNA–protein interactions in the h15.5 kD-U4 snRNA and hSBP2-SECIS RNA complexes. **A:** Contacts between 15.5 kD amino acids and the human U4snRNA, adapted from data of the crystal structure of the complex (Vidovic et al., 2000). Only the contacts at A30, U31, and G32-A44/G43-A33 base pairs are represented. Underlined amino acids: hydrophobic interactions with bases; residue with an asterisk: hydrogen bonds involving bases; italicized residues: hydrogen bond involving phosphates or ribose. **B:** Putative contacts between hSBP2 amino acids and the SECIS RNA. The amino acid residues are homologous to those in **A**. The SePN1 SECIS RNA positions were arbitrarily numbered, as in Figure 2. Only a portion of helices 1 and 2 is displayed. Sheared G-A base pairs are in bold.

FIGURE 3. Structure-guided mutagenesis. **A:** Folding predictions and positions of the alanine-substituted amino acids in the hSBP2 sequence. Substitutions are positioned by the arrows below the hSBP2 sequence. The secondary structure elements of the h15.5 kD protein (shown in black) and the residues involved in the 15.5 kD-U4 snRNA interaction (marked by dots above the sequence) are from Vidovic et al. (2000). The secondary structure prediction of hSBP2 shown in gray was generated with the Predict Protein program PHDSec (Rost & Sander, 1993). The sequence alignment is from Figure 1. **B**: Sketch of the three-dimensional structure of the 15.5kD-U4 snRNA complex solved by Vidovic et al. (2000). Only the regions of the protein and the RNA predicted to be conserved between 15.5 kD/U4 and hSBP2/SECIS are represented. Ribbon plot of the 15.5 kD residues 38 to 105 is shown in blue. The amino acids involved in RNA recognition and targeted for mutagenesis are highlighted: strictly conserved residues are in red, others are in pink. The U4 snRNA backbone, U31, and A30 are in green, the sheared G-A pairs are in yellow. The graphic representation was generated with the program SETOR (Evans, 1993) using the Protein Data Bank coordinates 1E7K.

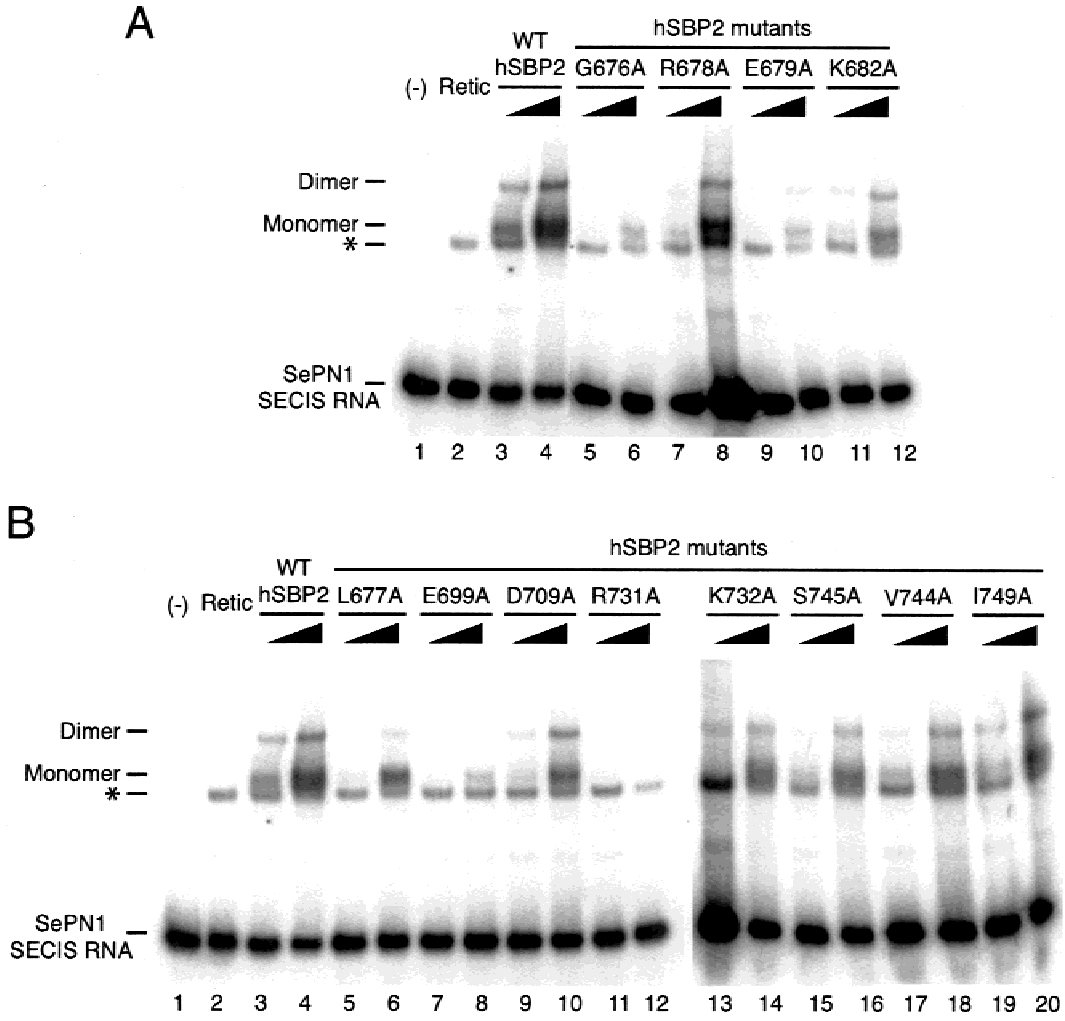

FIGURE 5. Gel retardation assays of the hSBP2 mutant proteins with the human SePN1 SECIS RNA. In each lane, the [^{35}S]-methionine-labeled hSBP2 protein obtained by in vitro translation in rabbit reticulocyte lysates was added to 150,000 cpm of [^{32}P]-labeled human SePN1 SECIS RNA and the complexes were resolved on 4% nondenaturing polyacrylamide gels. **A:** Effects of alanine substitutions at positions predicted to interact with the G3-A8/G7-A4 base pairs (lanes 5–12). **B:** Effects of alanine substitutions at positions predicted to interact with U2 (lanes 5–12, 19, and 20), A1 (lanes 17 and 18), or nonpredicted positions (lanes 13–16). Lanes 1: SePN1 SECIS RNA alone (−); lanes 2: unprogrammed rabbit reticulocyte lysate (Retic); odd and even lanes contained 20 and 120 fmol of in vitro translated wild-type (wt hSBP2) or mutant hSBP2 protein, respectively. The asterisk indicates the position of the complex formed between the SECIS RNA and another SECIS-binding protein (Hubert et al., 1996).

and 4A). Among the remaining mutants, V744A and I749A affected more the monomer formation whereas L677A reduced more the homodimeric than the monomeric complexes (Table 1). At this stage of the work, it is difficult to rationalize this finding, but we can hypothesize that the RNA binding and dimerization domains of hSBP2 partially overlap. Assuming that homodimerization stabilizes the binding to the SECIS RNA, the mutations would be less detrimental, and the homodimer would attenuate the down effects of V744A and I749A that are otherwise more harmful to monomer formation.

We conclude from these experiments that we have identified 12 hSBP2 amino acids important to SECIS RNA binding. This includes those derived directly from the sequence comparison and the structure-guided strategy, as well as Lys732^{SBP2} and Ser745^{SBP2}, which could not be predicted. The four amino acids Gly676^{SBP2}, Glu679^{SBP2}, Glu699^{SBP2}, and Arg731^{SBP2} appear to be crucial; the other eight residues contribute to the SECIS RNA binding activity but to a lower extent.

DISCUSSION

A previous report identified the structural determinants of the SECIS RNA necessary for the interaction with SBP2 (Fletcher et al., 2001). The objective of this study was to obtain a better understanding of the principles

TABLE 1. Quantitation of the binding activities of the hSBP2 mutant proteins from the gels shown in Figure 5.[a]

hSBP2 Mutants	Monomer	Dimer
Wild-type	100	100
E699A	13.5	0
R731A	0	0
G676A	19	16
E679A	19	28
L677A	57	27
R678A	64	67
K682A	56	78
D709A	46	76
K732A	57	46
S745A	56	51
V744A	51	100
I749A	47	100

[a]The percentage of monomeric and dimeric complexes was calculated as the ratio of the values obtained with the highest amount of hSBP2 (even lanes in Fig. 5A, B) to those of the wild-type monomeric and dimeric complexes taken as 100% (lanes 4).

governing the SBP2-SECIS RNA interaction, in particular the identification of the amino acids important for SECIS RNA recognition. Three main findings emerged from our investigations. The first one was the discovery that SBP2 shares the same RBD as the mammalian 15.5 kD protein (or the Snu13p ortholog in yeast) that binds the 5′ stem-loop of the spliceosomal U4 snRNA. In corollary, the second finding concerned the similarities detected in the core structures of the SECIS RNA and U4 snRNA bound by the SBP2 and 15.5 kD proteins, respectively. It is precisely these protein and RNA similar features, combined with secondary structure prediction of hSBP2 and the information from the crystal structure of the U4 snRNA-15.5 kD complex (Vidovic et al., 2000), that enabled the prediction of amino acids in the human SBP2 protein that should be critical to SECIS RNA binding. In the absence of a structural model for the SBP2-SECIS RNA complex, the structure-guided strategy offers the advantage of targeting amino acids that contribute to the interaction with the RNA, rather than those participating in the overall folding of the protein. This was verified by *in silico* investigation of the 15.5 kD three-dimensional structure; indeed the side chains of the amino acids corresponding to those mutated in hSBP2 do not establish intramolecular contacts important to the overall folding of the 15.5kD. Gly38$^{15.5}$ (corresponding to Gly676^{SBP2}) is particular in that it adopts a conformation that is not allowed to any other amino acid at the β1-α2 junction. This invariant amino acid is thus important for folding the RBD, but is also in close contact to the G-A pairs. Assays of the RNA-binding activities of the hSBP2 mutants allowed the identification of 12 amino acids whose substitution led to a complete or partial loss of RNA binding. From this data, we inferred that the four amino acids Gly676^{SBP2}, Glu679^{SBP2}, Glu699^{SBP2}, Arg731^{SBP2} are primordial to the interaction and that the other eight participate in SECIS RNA binding, constituting the third finding of this report.

Our structure-guided strategy allowed the identification of hSBP2 amino acids important for recognition of the SECIS RNA (Fig. 4B). Obviously, the detailed RNA–protein contacts cannot be predicted by this type of study. However, solution structure probing of the SECIS RNA and SECIS RNA-SBP2 complex (Walczak et al., 1996; Fletcher et al., 2001), combined with the work presented here, suggest the putative interaction scheme. Gly676^{SBP2} and Arg678^{SBP2} could interact with G3; Glu679^{SBP2} with G3 and G7; Lys682^{SBP2} with G7; Leu677^{SBP2}, Glu699^{SBP2}, Asp709^{SBP2}, Arg731^{SBP2}, and Ile749^{SBP2} with U2; and Val744^{SBP2} with A1. Interestingly, the solution structure of the complex between the yeast ribosomal protein L30 and its autoregulatory site in the L30 mRNA was solved by NMR spectroscopy (Mao et al., 1999). L30 binds to an internal loop constituted by a complex array of non-Watson–Crick base pairs whose three-dimensional structure differs from that of the U4 snRNA. In this RNA–protein complex, it is remarkable that Gly26^{L30}, which corresponds to Gly38$^{15.5}$ and Gly676^{SBP2} (see Fig. 1), is central to the interaction with the mRNA. Substitution of the Gly676^{SBP2} homologs in the rat SBP2 and 15.5 kD proteins led to detrimental effects as well (Nottrott et al., 1999; Copeland et al., 2001), in good agreement with the important role of this amino acid in the L7A/L30 family for both the structure of the RNA binding domain and recognition of the sheared G-A base pairs. Additionally, important roles were established in the L30-mRNA complex for Tyr27^{L30} (corresponding to Ala39$^{15.5}$ and Leu677^{SBP2}) and Lys28^{L30} (corresponding to Asn40$^{15.5}$ and Arg678^{SBP2}). This is consistent with our results for Leu677^{SBP2}. However, our data did not suggest a major role for Arg678^{SBP2}, highlighting subtle variations in the RNA–protein recognition schemes. We found that substitution of Lys732^{SBP2} and Ser745^{SBP2} affected the SECIS RNA binding, although the 15.5 kD homologous residues do not establish contacts with the U4 snRNA (Vidovic et al., 2000). It is unlikely that the mutations led to misfolding of hSBP2 because we could verify *in silico* that the corresponding amino acids Gln87$^{15.5}$ and Ser96$^{15.5}$ do not establish intramolecular contacts (Fig. 3B). One explanation could arise from the intimate RNA structure of the core and its vicinity that may not be strictly identical in both RNAs, particularly at the asymmetrical loop. Lys732^{SBP2} and Ser745^{SBP2} could thus be involved in the specialization for the SECIS RNA interaction.

The results of our mutagenesis study are in good agreement with a previous work assaying mutant SECIS RNAs for their ability to bind SBP2 (Fletcher et al., 2001). It was found that the U2C mutation, or the G3-A8/A4-G7 changes to A3-G8/G4-A7 or A3-A8/G4-A7, impaired formation of the SBP2-SECIS RNA complex.

This underscored the importance of U2 as well as of G3 and G7 in the sheared G3-A8/G7-A4 base pairs. Altogether, the data presented may suggest that SBP2 recognizes the SECIS RNA sheared G-A base pairs in a manner similar to the 15.5 kD protein in the 15.5 kD-U4 snRNA complex. It is worth mentioning that U31 is flipped out in the crystal structure (Vidovic et al., 2000; Fig. 3B) whereas structure probing of the SECIS RNA proposed that U2 is not bulged out but rather base paired with U9 (Walczak et al., 1996; see also Fig. 2). Indeed, one could envisage the unpairing of U2 by an induced fit of the SECIS RNA upon SBP2 binding.

It was recently reported that U4 snRNA, RNase MRP RNA, human SRP 7SL RNA, and several ribosomal RNA regions contain a new secondary structure motif called the kink-turn, or K-turn motif (Klein et al., 2001). These authors proposed that the L30 mRNA binding site can also adopt the K-turn motif. It is characterized by an asymmetrical internal loop flanked by a regular helix on one side and an irregular helix containing sheared G-A base pairs on the other (Fig. 2). A kink occurs at the internal loop, causing a sharp turn in the RNA helix. Interestingly, the K-turn in U4 snRNA and L30 mRNA is the binding site for the 15.5 kD/Snu13p and yeast L30 proteins, respectively. It was proposed that box C/D snoRNAs also contain a K-turn motif (Klein et al., 2001), in line with the binding of 15.5 kD/Snu13p to this type of snoRNA (Watkins et al., 2000). Structure probing data combined with computer modeling led to a three-dimensional structure model for the SECIS RNA where a kink occurs at the internal loop, showing the G-A base pairs well accessible to the solvent (Walczak et al., 1996). Considering this particular structural feature of the SECIS RNA and the binding of SBP2 at the G-A base pairs (Fletcher et al., 2001), we speculate that the SECIS RNA is another member of the RNA family containing a K-turn motif.

We observed the formation of hSBP2-SECIS RNA complexes containing monomeric and homodimeric forms of hSBP2. This observation is in line with our earlier report using the recombinant hSBP2 protein produced in *E. coli* (Lescure et al., 2002a). Using glycerol gradient centrifugation, Copeland et al. (2001) also observed homodimerization of the rat SBP2 protein. Taking into account that SBP2 binds not only the SECIS RNA but also the 28S ribosomal RNA via a ribosome-binding domain located N-ter to the RBD, these authors hypothesized that homodimers could represent the functional form of SBP2. It has not been reported yet that other members of the L7A/L30 family possess the capacity to homodimerize, but SBP2 could be unique in this respect: it is a rather large protein (854 amino acids for the full-length protein and 512 amino acids in the hSBP2 fragment used in this study) compared to the relatively small size of the 15.5 kD (128 amino acids) and other proteins of the L7A/L30 family. Actually, homodimerization of RNA-binding proteins is not unprecedented, and was already reported for RRM-containing proteins such as U1A, hnRNP A1, the La autoantigen and eIF4B (reviewed in Méthot et al., 1996; Craig et al., 1997; Puglisi, 2000).

There is a growing importance of functionally diverse eukaryotic proteins containing the L7A/L30 RBD. In fact, this domain was recently renamed Pelota (Anantharaman et al., 2002) after the name of a locus that encodes a protein required for meiotic cell division in *Drosophila* (Eberhart & Wasserman, 1995). However, our amino acid sequence alignment of the L7A/L30 family of proteins with Pelota orthologs indicated that the latter contain sequence similarity to only the first 28 amino acids (with respect to hSBP2) at the N-terminus of the L7A/L30 RBD (data not shown). Therefore, the blocks of homology in the C-terminal half of the L7A/L30 domain (Fig. 1), lacking in Pelota proteins, may provide different binding opportunities.

Some of the members of the L7A/L30 protein family bind, or potentially bind, RNAs with K-turn motifs. We have shown here that the 15.5 kD/Snu13p-U4 snRNA and SBP2-SECIS RNA complexes exhibit structural similarities, raising the question of how each protein can specifically identify its cognate RNA. This is especially crucial in light of the following recent reports adding evolutionary aspects to the issue. It was found that the archaeal ribosomal protein L7A possesses the other function of binding the archaeal box C/D small RNA and mammalian box C/D snoRNAs. The archaeal L7A protein is therefore the functional homolog to the eukaryotic 15.5 kD/Snu13p (Kuhn et al., 2002; Tang et al., 2002). Additionally, the Nhp2p protein that contains an L7A/L30 RBD and is a constituent of box H/ACA snoRNAs can also bind box C/D snoRNAs in vitro (Henras et al., 2001). Distinct features in the structures of each K-turn-containing RNA can account for the specificity of binding. Another straightforward and not mutually exclusive possibility is that every L7A/L30 RBD contains nonconserved amino acids, specifically dedicated to recognizing each individual RNA target. Elucidation of this question represents the route for future investigations.

MATERIALS AND METHODS

cDNA constructs and site-directed mutagenesis

To allow in vitro transcription/translation of hSBP2 wild-type and mutant constructs, the hSBP2 cDNA was subcloned downstream of the T7 promoter of pBluescript II KS (−). To do this, the 2.1 kb *Xba*I-*Hin*dIII fragment arising from phSBP2/512 (Lescure et al., 2002a), containing an N-terminal *Strep*-tag II (IBA, Germany) fused to the 512 C-terminal amino acids of hSBP2, was inserted into pBluescript II KS (−). The resulting plasmid was termed pKS-hSBP2/512. Alanine substitution

mutants were generated in pKS-hSBP2/512 by site-directed mutagenesis. Mutant constructs were entirely sequenced by automated DNA sequencing.

Oligonucleotides used for mutagenesis were as follows:

G676A: 5′-GAGAACCTCCCTCAAGGCCAACACAAGTCG ACG-3′;
L677A: 5′-TTTGAGAACCTCCCTGGCCCCCAACACAAG TCG-3′;
R678A: 5′-GTGTTTGAGAACCTCGGCCAACCCCAACAC AAG-3′;
E679A: 5′-CAGGTGTTTGAGAACGGCCCTCAACCCCAA CAC-3′;
K682A: 5′-TTTGAGCTTCAGGTGGGCGAGAACCTCCCT CAA-3′;
E699A: 5′-TTTTGACTGTATCTTGGCACAGTTGGGAGA AAT-3′;
D709A: 5′-TAATTGTGTGCAAAGTGGCATCCAGCCCACC TTT-3′;
R731A: 5′-GCGCCCCAGAGCTTTGGCGTTGAGAGCAAA CAC-3′;
K732A: 5′-ACTGCGCCCCAGAGCGGCGCGGTTGAGAG CAAA-3′;
V744A: 5′-GATCCCCACCACACTGGCAGGAACTGCCTT ATT-3′;
S745A: 5′-GAAGATCCCCACCACGGCGACAGGAACTGC CTT-3′;
I749A: 5′-CCCATCATAGCTGAAGGCCCCCACCACACTG AC-3′.

In vitro translation

Wild-type and mutant hSBP2 proteins were generated in vitro using TNT coupled Reticulocyte Lysate Systems (Promega). One microgram of each of the pKS-hSBP2/512 wild-type or mutant plasmid DNAs was used as the template in 50 μL in vitro transcription/translation reactions in the presence of 25 μL rabbit reticulocyte lysate and 20 μCi of [^{35}S]-methionine (1,175 Ci/mmol). The yield of [^{35}S]-methionine incorporation was determined by 5% TCA precipitation of 2-μL aliquots of the reactions, followed by scintillation counting and calculation with respect to the [^{35}S]-methionine input. Obtaining of the translation products was verified by electrophoresis on 12% SDS-PAGE. The amount of each hSBP2 protein was quantitated with a Fuji BioImage BAS2000 analyzer. To assay the solubility of the hSBP2 mutant proteins that affected SECIS RNA binding, constructs were transformed into *E. coli* BL21 (DE3) RIL. After induction of protein synthesis, the soluble and insoluble fractions were loaded on SDS-PAGE and analyzed by western blotting using an anti-*Strep*-tag II antibody (IBA, Germany).

Electrophoretic mobility shift assays

For in vitro transcription of the human SePN1 and rat GPx SECIS RNAs, plasmids pT7BcKSelN and pRGPxBcK were linearized by *Eco*RI (Walczak et al., 1998; Fagegaltier et al., 2000b). Internally labeled SePN1 and GPx SECIS RNAs were obtained by T7 transcription with [α-^{32}P]-ATP (3,000 Ci/mmol) according to Hubert et al. (1996). Formation of the SePN1 SECIS RNA-hSBP2 and GPx SECIS RNA-hSBP2 complexes were conducted as described in Copeland et al. (2001) and Fletcher et al. (2001). Routinely, 150,000 cpm (2.4 fmol) of ^{32}P-labeled SECIS RNA were incubated for 30 min at 30 °C with 20 or 120 fmol of in vitro translated wild-type or mutant hSBP2 protein, in 20 μL of phosphate buffer saline pH 7.4, 10 mM DTT. RNA–protein complexes were separated by 4% nondenaturing polyacrylamide gel electrophoresis in Tris-glycine, pH 8 (Fletcher et al., 2001). The intensities of free and bound RNAs were quantitated with a Fuji BioImage BAS2000 analyzer. Two independent experiments were performed for both SePN1 and GPx SECIS RNAs. Quantitation of the results varied within 15%.

ACKNOWLEDGMENTS

We are grateful to V. Cura for the graphical representation and help with *in silico* analysis. A. Lescure and E. Myslinski are thanked for helpful comments on the manuscript and C. Loegler for excellent technical assistance. C.A. was awarded a fellowship of the Fondation pour la Recherche Médicale. This work was supported by the Fondation pour la Recherche Médicale, the Ligue Régionale contre le Cancer and the Association pour la Recherche contre le Cancer.

Received June 14, 2002; returned for revision
June 28, 2002; revised manuscript received July 19, 2002

REFERENCES

Anantharaman V, Koonin EV, Aravind L. 2002. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res 30*:1427–1464.

Berry MJ, Banu L, Chen YY, Mandel SJ, Kieffer JD, Harney JW, Larsen PR. 1991. Recognition of UGA as a selenocysteine codon in type 1 deiodinase requires sequences in the 3′ terminal untranslated region. *Nature 353*:273–276.

Copeland PR, Fletcher JE, Carlson BA, Hatfield DL, Driscoll DM. 2000. A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J 19*: 306–314.

Copeland PR, Stepanik VA, Driscoll DM. 2001. Insight into mammalian selenocysteine insertion: Domain structure and ribosome binding properties of Sec insertion sequence Binding protein 2. *Mol Cell Biol 21*:1491–1498.

Craig AWB, Svitkin YV, Lee HS, Belsham GJ, Sonenberg N. 1997. The La autoantigen contains a dimerization domain that is essential for enhancing translation. *Mol Cell Biol 17*:163–169.

Eberhart CG, Wasserman SA. 1995. The *pelota* locus encodes a protein required for meiotic cell division: An analysis of G_2/M arrest in *Drosophila* spermatogenesis. *Development 121*:3477–3486.

Evans SV. 1993. SETOR: Hardware-lighted three-dimensional solid model representations of macromolecules. *J Mol Graph 134*: 127–128.

Fagegaltier D, Carbon P, Krol A. 2001. Distinctive features in the SelB family of elongation factors for selenoprotein synthesis. A glimpse of an evolutionary complexified translation apparatus. *BioFactors 14*:5–10.

Fagegaltier D, Hubert N, Yamada K, Mizutani T, Carbon P, Krol A. 2000a. Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J 19*:4796–4805.

Fagegaltier D, Lescure A, Walczak R, Carbon P, Krol A. 2000b. Structural analysis of new local features in SECIS RNA hairpins. *Nucleic Acids Res 28*:2679–2689.

Fletcher JE, Copeland PR, Driscoll DM, Krol A. 2001. The selenocysteine incorporation machinery: Interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA 7*:1442–1453.

Gladyshev VN, Kryukov GV. 2001. Evolution of selenocysteine-containing proteins: Significance of identification and functional characterization of selenoproteins. *BioFactors 14*:87–92.

Gottschalk A, Neubauer G, Banroques J, Mann M, Lührmann R, 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.

Grundner-Culemann E, Martin GW III, Harney JW, Berry MJ. 1999. Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. *RNA 5*:625–635.

Henras A, Dez C, Noaillac-Depeyre J, Henry Y, Caizergues-Ferrer M. 2001. Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic Acids Res 29*:2733–2746.

Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gélugne J-P, Caizergues-Ferrer M. 1998. Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J 17*:7078–7090.

Hubert N, Walczak R, Carbon P, Krol A. 1996. A protein binds the selenocysteine insertion element in the 3′ UTR of mammalian selenoprotein mRNAs. *Nucleic Acids Res 24*:464–469.

Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: A new RNA secondary structure motif. *EMBO J 20*:4214–4221.

Köhrle J, Brigelius-Flohé R, Böck A, Gärtner R, Meyer O, Flohé L. 2000. Selenium in biology: Facts and medical perspectives. *Biol Chem 381*:849–864.

Koonin EV, Bork P, 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.

Krol A. 2002. Evolutionarily different RNA motifs and RNA–protein complexes to achieve selenoprotein synthesis. *Biochimie*. In press.

Kuhn JF, Tran EJ, Maxwell ES. 2002. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5 kD/Snu13p snoRNP core protein. *Nucleic Acids Res 30*:931–941.

Lescure A, Allmang C, Yamada K, Carbon P, Krol A. 2002a. cDNA cloning, expression pattern and RNA binding analysis of human selenocysteine insertion sequence (SECIS) binding protein 2. *Gene 291*:279–285.

Lescure A, Fagegaltier D, Carbon P, Krol A. 2002b. Protein factors mediating selenoprotein synthesis. *Curr Prot Pept Sci 3*:143–151.

Low SC, Grundner-Culemann E, Harney JW, Berry MJ. 2000. SECIS-SBP2 interactions dictate selenocysteine incorporation efficiency and selenoprotein hierarchy. *EMBO J 19*:6882–6890.

Mao H, White SA, Williamson JR. 1999. A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nature Struct Biol 6*:1139–1147.

Martin GW III, Harney JW, Berry MJ. 1998. Functionality of mutations at conserved nucleotides in eukaryotic SECIS elements is determined by the identity of a single nonconserved nucleotide. *RNA 4*:65–73.

Méthot N, Song MS, Sonenberg N. 1996. A region rich in aspartic acid, arginine, tyrosine, and glycine (DRYG) mediates eukaryotic Initiation Factor 4B (eIF4B) self-association and interaction with eIF3. *Mol Cell Biol 16*:5328–5334.

Moghadaszadeh B, Petit N, Jaillard C, Brockington M, Quijano-Roy S, Merlini L, Romero N, Estournet B, Desguerre I, Chaigne D, Muntoni F, Topaloglu H, Guicheney P. 2001. Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome. *Nature Genet 29*:17–18.

Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, Lührmann R. 1999. Functional interaction of a novel 15.5 kD [U4/U6.U5] tri-snRNP protein with the 5′ stem-loop of U4 snRNA. *EMBO J 18*:6119–6133.

Pfeifer H, Conrad M, Roethlein D, Kyriakopoulos A, Brielmeier M, Bornkamm GW, Behne D. 2001. Identification of a specific sperm nuclei selenoenzyme necessary for protamine thiol cross-linking during sperm maturation. *FASEB J 15*:1236–1238.

Puglisi JD. 2000. mRNA processing: The 3′ end justifies the means. *Nature Struct Biol 7*:263–264.

Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol 232*:584–599.

Stevens SW, 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.

Tang TH, Rozhdestvensky TS, Clouet d'Orval B, Bortolin M-L, Huber H, Charpentier B, Branlant C, Bachellerie J-P, Brosius J, Hüttenhofer A. 2002. RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res 30*:921–930.

Tujebajeva RM, Copeland PR, Xu XM, Carlson BA, Harney JW, Driscoll DM, Hatfield DL, Berry MJ. 2000. Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep 1*:158–163.

Ursini F, Heim S, Kiess M, Maiorino M, Roveri A, Wissing J, Flohé L. 1999. Dual function of the selenoprotein PHGPx during sperm maturation. *Science 285*:1393–1396.

Vidovic I, Nottrott S, Hartmuth K, Lührmann R, Ficner R. 2000. Crystal structure of the spliceosomal 15.5 kD protein bound to a U4 snRNA fragment. *Mol Cell 6*:1331–1342.

Walczak R, Carbon P, Krol A. 1998. An essential non-Watson–Crick base pair motif in 3′ UTR to mediate selenoprotein translation. *RNA 4*:74–84.

Walczak R, Westhof E, Carbon P, Krol A. 1996. A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA 2*:367–379.

Watkins NJ, Ségault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Lührmann R. 2000. A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell 103*:457–466.