



**HAL**  
open science

# Assemblage et fonction de complexes ARN-protéines

Christine Allmang

► **To cite this version:**

Christine Allmang. Assemblage et fonction de complexes ARN-protéines. Sciences du Vivant [q-bio].  
Université Louis Pasteur - Strasbourg I, 2007. tel-00265610

**HAL Id: tel-00265610**

**<https://theses.hal.science/tel-00265610>**

Submitted on 19 Mar 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Université Louis Pasteur**  
Faculté des Sciences de la vie  
STRASBOURG

*Habilitation à diriger des Recherches*

*Assemblage et fonction de complexes  
ARN-protéines*

Présentée par

**Christine Allmang-Cura**

Chargée de Recherche au CNRS  
Architecture et Réactivité de l'ARN, UPR 9002 du CNRS

Soutenue le 26 novembre 2007

**Membres du Jury :**

Dr. Michèle Caizergues-Ferrer

Pr. Jean-Pierre Rousset

Dr. James Stevenin

Pr. David Tollervey

Pr. Eric Westhof

Dr. Alain Krol

rapporteur externe

rapporteur externe

rapporteur interne

examineur

examineur

garant d'habilitation

# Sommaire

<i>Sommaire</i>	1
<i>Curriculum vitae</i>	2
<i>Production scientifique</i>	5
A- Publications	5
B- Communications	7
<i>Synopsis</i>	10
<i>Introduction générale</i>	12
<i>Projet de thèse : Le site de fixation de la protéine ribosomique S8 sur l'ARNr 16S</i>	15
<i>Stage post-doctoral (1994-1996) : Mécanismes de maturation des pré-ARNr</i>	17
1- Mise en évidence d'une coordination de la maturation des pré-ARNr	19
2- Démonstration de l'activité endonucléolytique de la RNase MRP <i>in vitro</i>	19
<i>Poste de chargée de recherche l'Université d'Edimbourg (1996-2001) : L'exosome et la synthèse des ARN stables</i>	21
1- L'exosome et le complexe PM-Scl humain	21
2- Les fonctions de l'exosome	23
3- Mécanismes de synthèse des ARN stables	25
<i>Activité de recherche et projets scientifiques actuels : Le mécanisme de synthèse des sélénoprotéines</i>	27
A- Introduction	27
B- Les interactions autour de l'ARN SECIS	30
1- La protéine SBP2 humaine et son mode d'interaction avec l'élément SECIS	31
2- Principes de reconnaissance entre protéines de la famille L7Ae et les ARN en K-turn	33
3- Objectif : Résolution de la structure cristallographique des complexes SBP2-SECIS, L30-SECIS	34
4- La protéine SBP2 de drosophile	35
C- Les complexes supramoléculaires impliqués dans la synthèse des sélénoprotéines	37
1- Un mécanisme commun pour l'assemblage des RNP L7Ae (manuscrit soumis)	39
2- Projets à court terme: l'assemblage de la mRNP SECIS	43
3- Projets à plus long terme : Purification des complexes associés à SBP2	46
<i>Références bibliographiques</i>	48
<i>Principales publications</i>	53

## *Curriculum vitae*

### **Dr. ALLMANG-CURA Christine**

UPR 9002 du CNRS  
Institut de Biologie Moléculaire et Cellulaire  
15, rue René Descartes  
67084 Strasbourg Cedex, FRANCE  
Tel: + 33 3 88 41 70 80  
Fax: + 33 3 88 60 22 18  
e-mail: C.Allmang@ibmc.u-strasbg.fr

Née le 26 Août 1967 à Sarreguemines (Moselle)  
Mariée, 1 enfant  
Nationalité française

### **Formation universitaire à l'Université Louis Pasteur de Strasbourg :**

- Mars 1994 : **Doctorat de l'Université Louis Pasteur de Strasbourg en Biologie Moléculaire** (mention Très Honorable)
- Juin 1990 : **DEA de Biologie Cellulaire et Moléculaire** (mention B)
- 1988-1990 : **Magistère de Chimie Biologie** (mention B)

### **Position actuelle :**

Depuis Octobre 2001 : **Chargée de Recherche CR1 au CNRS**. UPR 9002 du CNRS. IBMC, Strasbourg. Directeur : Professeur Eric Westhof. Equipe du Docteur Alain Krol.

### **Poste à l'Université d'Edimbourg :**

Octobre 1996 – mars 2001 : “**Research Fellow**”. Wellcome Trust Centre for Cell Biology, Université d'Edimbourg, Grande Bretagne. Equipe du Professeur David Tollervey.

### **Stage Post-doctoral :**

Mai 1994- septembre 1996 : Département d'Expression Génétique. EMBL, Heidelberg, Allemagne. Directeur : Iain Mattaj. Equipe du Docteur David Tollervey.

### **Thèse :**

1990-1994 : DEA et thèse en Biologie Cellulaire et Moléculaire à l'Université Louis Pasteur de Strasbourg I. IBMC, Strasbourg. Equipe du Professeur Bernard Ehresmann. Directeur de thèse: Docteur Chantal Ehresmann. Thèse soutenue le 23 mars 1994.

Titre : *Ingénierie et études structurales d'ARN en solution. Application à trois systèmes: Le site de fixation de la protéine S8 sur l'ARN ribosomique 16S d'Escherichia coli, l'ARN ribosomique 5S de Xenopus laevis et de l'ARN 3 du virus des nervures jaunes et nécrotiques de la betterave.*

### **Séjours scientifiques :**

- Université libre d'Amsterdam, Pays-Bas. Septembre 1996. Equipe du Professeur R.J. Planta. *Analyse de protéines ribosomiques sur gel bidimensionnels.*
- Université de Bayreuth, Allemagne. Mai 1992. Equipe du Professeur M. Sprinzl *Synthèse chimique d'ARN par la voie des H- phosphonates.*
- Université de Victoria, Colombie Britannique, Canada. Mai - juin 1991. Equipe du Professeur P. Romaniuk. *Synthèse chimique d'ARN par la voie des phosphoramidates.*

### **Encadrement et responsabilités collectives :**

#### ***Encadrement, formation :***

**Akiko Takeuchi**, Doctorante (Aspects moléculaires et cellulaires de la Biologie) depuis septembre 2006. Directeur de thèse : Docteur Alain Krol. Sujet : *Caractérisation, purification et cristallisation de la protéine SBP2 de Drosophila melanogaster*

**Laurence Wurth**, étudiante en DEA de Biologie Moléculaire et Cellulaire puis Doctorante (allocataire de recherche MREST), depuis septembre 2005. Directeur de thèse : Docteur Alain Krol. Sujet : *Identification de nouveaux facteurs moléculaires impliqués dans la synthèse des sélénoprotéines*

**Vincent Olieric**, étudiant en DEA puis Doctorant en cristallographie biologique, de septembre 2003 à novembre 2006. Directeur de thèse : Docteur Philippe Dumas (UPR 9002 du CNRS). Sujet: *Cristallisation du complexe SBP2/ARN SECIS au cœur du mécanisme de synthèse des sélénoprotéines*

**David Schmitt**, étudiant en DEA de Biologie Moléculaire et Cellulaire de septembre 2002-juin 2003. Sujet : *Dissection fonctionnelle de SBP2, protéine impliquée dans la synthèse des sélénoprotéines.*

**Emmanuelle Kiefer**, étudiante en BTS Biotechnologie de janvier- février 2002. Sujet : *Clonage de l'ADNc de la protéine hSBP2 entière et tronquée dans des vecteurs d'expression en vue d'études cristallographiques.*

**Thomas Khalish**, étudiant en Licence L3. Stage de Biologie Moléculaire juillet 2007. Sujet : *Clonage de l'ADNc de facteurs d'assemblages impliqués dans le mécanisme de synthèse des sélénoprotéines.*

***Participation à des contrats de recherche (Partenariat et valorisation) :***

J'ai participé à la rédaction des demandes de financement du laboratoire ci-dessous, relatives aux sujets de ma thématique. J'ai plus particulièrement pris en charge la rédaction de la demande de contrat ANR concernant notre équipe.

**ANR blanc (depuis octobre 2006)**

Titre du projet : *Nufip/Rsa1: a common assembly machine for snoRNPs, telomerase, and selenoprotein mRNPs.*

Rôle : Responsable de thématique et rédaction du projet strasbourgeois. Obtention d'un financement pour un post-doctorant qui sera recruté en 2008 et que je superviserai.

Coordinateur du projet : Edouard Bertrand (UMR 5535 CNRS-IGM, Montpellier)

Partenaires : Bruno Charpentier et Christiane Branlant (UMR 7567 CNRS-UHP, Nancy), Alain Krol (UPR 9002 du CNRS, Strasbourg), Barbara Bardoni (Université Nice, Sophia Antipolis).

**Action Concertée Incitative Biologie Cellulaire Moléculaire et Structurale-Ministère de la Recherche (Octobre 2004-Septembre 2007)**

Titre du projet : *Comment les structures formées entre ARN en K-turn et protéines de la famille L7Ae initient-elles la formation de particules ribonucléoprotéiques ayant des fonctions cellulaires variées ?*

Coordinateur du projet : Alain Krol

Partenaires du projet : Christiane Branlant (UMR 7567 CNRS-UHP, Nancy), André Aubry (UMR 7086 CNRS-UHP, Nancy), Philippe Dumas (UPR 9002 du CNRS, Strasbourg), Alain Krol (UPR 9002 du CNRS).

**Programme Toxicologie Nucléaire Environnementale CEA-CNRS-INSERM-INRA-Ministère de la Recherche**

**Projet Signalisation, Détection, Détoxication & Contrôle Redox en Réponse aux Métaux Lourds (SIDDERE) (Octobre 2004-Septembre 2007)**

Titre du projet : *Fonction biologique du sélénium. Incorporation ciblée dans les protéines à sélénocystéine chez les eucaryotes supérieurs*

Coordinateurs du projet : Michel Toledano, Jean-Luc Montillet et Alain Vavasseur

***Animation et administration de la recherche :***

Membre élu du conseil de laboratoire de l'UPR 9002 du CNRS depuis septembre 2004

Membre de la Société Française de Biochimie et Biologie Moléculaire

## *Production scientifique*

### **A- Publications**

- 1- Cléry A, Bourguignon-Igel V, Allmang C, Krol A, and Branlant C. (2007) An improved definition of the RNA binding specificity of SECIS binding protein 2, an essential component of the selenocysteine incorporation machinery. *Nucleic Acids Res.* **35**,1868-84.
- 2- Allmang C and Krol A. (2006) SECIS RNAs and K-turn binding proteins. A survey of evolutionary conserved RNA and protein motifs. In Selenium, its molecular Biology and role in human Health 2<sup>nd</sup> edition. DL Hatfield (ed) Kluwer Academic Publishers. **5**, 51-61.
- 3- Allmang C and Krol A. (2006). Selenoprotein synthesis: UGA does not end the story. *Biochimie.* **88**, 1561-1571.
- 4- Milligan L, Torchet C, Allmang C, Shipman T and Tollervey D. (2005) A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Mol. Cell. Biol.* **25**, 9996-10004.
- 5- Kufel J, Allmang C, Verdone L, Beggs J and Tollervey D. (2003). A complex pathway for 3'-processing of the yeast U3 snoRNA. *Nucleic Acids Res.* **31**, 6788-6797.
- 6- Kufel J, Allmang C, Petfalski E, Beggs J and Tollervey D. (2003). Lsm Proteins Are Required for Normal Processing and Stability of Ribosomal RNAs. *J. Biol. Chem.* **278**, 2147-2156.
- 7- Allmang C, Carbon P and Krol A. (2002). The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding *RNA* **8**, 1308-1318.
- 8- Lescure A, Allmang C, Yamada K, Carbon P and Krol A. (2002). cDNA cloning, expression pattern and RNA binding analysis of human SECIS binding protein 2. *Gene* **291**, 279-285.
- 9- Kufel J, Allmang C, Verdone L, Beggs JD and Tollervey D. (2002). Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. *Mol. Cell. Biol.* **22**, 5248-5256.
- 10- Brouwer R, Allmang C, Rajmakers R, van Aarssen Y, Vree Egberts W, Petfalski E, van Venrooij WJ, Tollervey D and Pruijn GJM. (2001). Three novel components of the human exosome. *J. Biol. Chem.* **276**, 6177-6184.
- 11- Kufel J, Allmang C, Chanfreau G, Petfalski E, Lafontaine DL and Tollervey D. (2000). Precursors to the U3 snoRNA lack snoRNP proteins but are stabilized by La binding. *Mol. Cell. Biol.* **20**, 5415-5424.
- 12- Allmang C, Mitchell P, Petfalski E and Tollervey D. (2000) Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res.* **28**, 1684-91.

- 13- Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E and Tollervey D. (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**, 5399-5410.
- 14- Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D and Mitchell P. (1999) The yeast exosome and human PM-Scl are related complexes of 3' → 5' exonucleases. *Genes and Dev.* **13**, 2148-2158.
- 15- Allmang C and Tollervey D. (1998) The role of the 3' external transcribed spacer in yeast pre-rRNA processing. *J. Mol. Biol.* **278**, 67-78.
- 16- Lygerou Z, Allmang C, Tollervey D and Séraphin B. (1996) Accurate processing of a eukaryotic precursor ribosomal RNA by Ribonuclease MRP *in vitro*. *Science* **272**, 268-270.
- 17- Allmang C, Henry Y, Morrissey J.P, Wood H, Petfalski E and Tollervey D. (1996) Processing of the yeast pre-rRNA at sites A<sub>2</sub> and A<sub>3</sub> is linked. *RNA* **2**, 63-73.
- 18- Allmang C, Henry Y, Wood H, Morrissey J.P, Petfalski E and Tollervey D. (1996) Recognition of cleavage site A<sub>2</sub> in the yeast pre-rRNA. *RNA* **2**, 51-62.
- 19- Allmang C, Mougél M, Westhof E, Ehresmann B and Ehresmann C. (1994) Role of conserved nucleotides in building the 16S rRNA binding site of *E. coli* ribosomal protein S8. *Nucleic Acids Res.* **22**, 3708-3714.
- 20- Mougél M, Allmang C, Eyermann F, Cachia C, Ehresmann B and Ehresmann C. (1993) Minimal 16S rRNA binding site and role of conserved nucleotides in *E. coli* ribosomal protein recognition. *Eur. J. Biochem.* **215**, 787-792.
- 21- Gilmer D, Allmang C, Ehresmann C, Guilley H, Richards K, Jonard G and Ehresmann B. (1993) The secondary structure of the 5'-noncoding region of beet necrotic yellow vein virus RNA 3: evidence for a role in viral RNA replication. *Nucleic Acids Res.* **21**, 1389-1395.



## B- Communications

### **Présentations orales :**

2<sup>ème</sup> rencontre entre l'Unité 'Architecture et Réactivité des ARN' (IBMC) et le département de Biologie et Génomique Structurale (IGBMC) (Mont Sainte Odile, France, 25-26 janvier 2007). A common mechanism for the assembly of nuclear and SECIS RNPs. Allmang C.

Rencontre de l'Unité 'Architecture et Réactivité des ARN' (IBMC) (CIARUS, Strasbourg, 14 décembre 2006). A common mechanism for the assembly of sno, sn and SECIS mRNPs. Allmang C, Wurth L et Krol A.

Séminaire ToxNuc-E (Dijon, France, juin 2005). Un mécanisme original pour la synthèse des sélénoprotéines. Allmang C et Krol A.

4<sup>ème</sup> Rencontre SIFRA<sup>RN</sup>. Structure, intégration, fonction et réactivité des ARN. (Nancy, France, 14-17 octobre 2002). Un même domaine de liaison à l'ARN pour les protéines SBP2 et 15.5 kD/Snu13p. Allmang C et Krol A.

3<sup>ème</sup> Rencontre SIFRA<sup>RN</sup>. Le monde des ARN et ses nouvelles frontières. (Toulouse, France, 19-21 janvier 2000). L'Exosome: un complexe multifonctionnel d'exonucléases 3' ->5'. Allmang C, Mitchell P, Kufel J, Brouwer R, Petfalski L, van Venrooij W et Tollervey D.

RNA' 99 (The fourth annual meeting of the RNA Society, University of Edinburgh 23-27 juin 1999). The yeast exosome and human PM-Scl are related complexes of 3' ->5' exonucleases. Allmang C, Mitchell P, Brouwer R, Petfalski E. van Venrooij W et Tollervey D.

RNA' 98 (The third annual meeting of the RNA Society, University of Wisconsin-Madison, 26-31 mai 1998). The exosome: A surprisingly large complex of 3' ->5' exonucleases. Allmang C, Mitchell P, Petfalski E et Tollervey D.

2<sup>ème</sup> Rencontre SIFRA<sup>RN</sup>. Le monde des ARN et leurs nouvelles fonctions: réalités et perspectives. (Strasbourg, France, 28-30 avril 1998). Coordination de la maturation des pré-ARN. Allmang C et Tollervey D.

5 th UK RNA Processing Workshop (Rydal Hall, Ambleside, Cumbria, 23-25 janvier 1998). The coordination of pre-rRNA processing. Allmang C et Tollervey D.

Young Scientist's View of Molecular Biotechnology (Mt. Ste. Odile, France, 28 février- 6 mars 1993). The binding site of *E. coli* ribosomal protein S8 on 16S rRNA: Minimal RNA structural requirement and role of conserved nucleotides in protein S8 recognition. Allmang C, Mougél M, Eyermann F, Ehresmann B et Ehresmann C.

XVIIIe Forum des Jeunes Chercheurs (Tours, France 3-6 septembre 1991). Le site de reconnaissance de la protéine ribosomique S8 sur l'ARN ribosomique 16S d'*E. coli* : Relations structure/reconnaissance par mutagenèse dirigée. Allmang C, Mougél M, Ehresmann B et Ehresmann C.

**Posters :**

6ième rencontre SifrARN. Structure, integration, fonction et réactivité des ARN (Rennes, France, 3-6 juillet 2006). Interactions ARN/protéines au niveau de l'ARN SECIS: variations autour du motif L7A/L30 et des ARN en K-turn. Wurth L, Cléry A, Branlant C, Krol A et Allmang C.

Séminaire de Toxicologie Nucléaire Environnementale.(Auteuil, France, 5-7 décembre 2005). Un mécanisme original pour la synthèse des protéines à selenium. Allmang C., Beniaminov A., Wurth L., et Krol A.

EMBO Conference on Protein Synthesis and Translational Control (EMBL, Heidelberg, Allemagne, 14-18 Septembre 2005). Principles of RNA-protein recognition between proteins of the L7A/L30 family and K-turn RNA motifs. Cléry A, Schmitt D, Wurth L., Bourguignon-Igel V, Branlant C, Krol A et Allmang C.

5ième rencontre SifrARN. ARN, le nouveau monde (Arcachon, France, 10-13 Octobre 2004). Dimérisation de SBP2, une protéine de liaison à l'ARN impliquée dans le mécanisme de synthèse des sélénoprotéines. Schmitt D, Krol A et Allmang C.

Symposium. Structure, function and dynamics of RNA-protein complexes. (Göttingen, Allemagne, 17-20 septembre 2003). SBP2, a multifunctional protein involved in selenoprotein synthesis. Schmitt D, Krol A et Allmang C.

1<sup>ier</sup> congrès de Traduction Francophone (Institut Pasteur, Paris, 12-13 décembre 2002). Un même domaine de liaison à l'ARN pour les protéines SBP2 et 15.5 kD/Snu13p. Allmang C et Krol A.

RNA 2002 (The Seventh Annual meeting of the RNA Society, University of Wisconsin-Madison, 28 mai- 2 juin 2002). Similar protein contacts for SBP2/SECIS RNA and 15.5 kD/U4 snRNA complexes. Allmang C et Krol A.

The ribosome: Its (nucleolar) synthesis and structure (Amsterdam, 16-20 Août 1997). Interactions between yeast pre-rRNA processing complexes. Allmang C et Tollervey D.

RNA' 96 (The first annual meeting of the RNA Society, University of Wisconsin-Madison, 28 mai - 02 juin 1996). Interactions between yeast pre-rRNA processing complexes. Allmang C, Henry Y, Morrissey JP, Wood H, Petfalski E et Tollervey D.

Molecular Biology of RNA: Splicing and 3'-end formation of RNA (Mont Ste Odile, France, 13-17 septembre 1995). Pre-rRNA processing in ITS1 of *Saccharomyces cerevisiae*. Allmang C, Lygerou Z, Henry Y, Morrissey JP, Wood H, Séraphin B et Tollervey D.

Ribosome synthesis and nucleolar function (Cold Spring Harbor, New York, 28 septembre- 02 mars 1994). Role of the conserved nucleotides of S8 16S rRNA binding site in *E. coli*. Moine H, Allmang C, Mougé M, Westhof E, Ehresmann B et Ehresmann C.

Symposium on structural Tools for the Analysis of Protein- Nucleic Acid Complexes (Wildbad Kreuth, Allemagne, 3-7 mai 1992). Three-dimensional structure of the binding site of *E. coli*

ribosomal protein S8 on 16S rRNA by structure probing, site directed mutagenesis and graphic modelling. Allmang C, Mougel M, Ehresmann B, Eyermann F, Westhof E et Ehresmann C.

**Autre participations :**

8<sup>th</sup> Symposium on Selenium in Biology and Medecine (Madison, Etats-Unis, 25-30 juillet 2006). RNA-recognition at the SECIS RNA. Allmang C, Cléry A, Bourguignon V, Allamand V, Richard P, Lescure A, Guicheney P, Branlant C and Krol A.

Second JSPS (Japan Society for the Promotion of Science) Forum in France (Strasbourg, 28 novembre 2003).

International Conference on the Translational Apparatus, (Berlin, Allemagne, 31 octobre- 5 novembre 1992). *E.coli* ribosomal protein S8 recognizes specific three-dimensional features of 16S RNA. Mougel M, Allmang C, Westhof E, Ehresmann B et Ehresmann C.

## *Synopsis*

Les particules ribonucléoprotéiques (ou RNP) sont à la base de nombreuses fonctions cellulaires fondamentales. La formation de ces particules RNP est un processus très complexe qui nécessite de nombreuses étapes de maturation et de multiples facteurs d'assemblage. Par ailleurs, une structure correcte des particules RNP est essentielle à leur fonction. Il est donc critique de comprendre comment ces particules sont formées dans la cellule. Au cours de ma carrière, je me suis intéressée à plusieurs aspects de ces mécanismes.

Au cours de ma thèse dirigée par Chantal Ehresmann (1990-1994), dans l'équipe de Bernard Ehresmann (UPR 9002 du CNRS) j'ai étudié le mode d'interaction de la protéine ribosomique S8 sur l'ARNr 16S d'*E. coli*.

Mon travail post-doctoral dans l'équipe de David Tollervey (EMBL, Heidelberg (1994-1996) et Université d'Edimbourg (1996-2001) a porté sur l'étude des mécanismes de maturation, d'assemblage et de dégradation de diverses RNP. J'ai notamment contribué à la caractérisation de l'exosome, un complexe d'exonucléases 3'-> 5' impliqué dans la maturation et la dégradation de divers ARN chez la levure. J'ai également étudié le rôle de protéines chaperons dans la biogenèse des snoARN (biogenèse des ribosomes), des ARN ribosomiques et des ARNt.

En 2001, j'ai été recrutée au grade de chargée de recherche au CNRS dans l'équipe d'Alain Krol où nous étudions les mécanismes de synthèse des sélénoprotéines. L'incorporation de sélénocystéine dans les sélénoprotéines fait appel au recodage co-translationnel d'un codon UGA<sub>sec</sub> en phase. Chez les eucaryotes, ce mécanisme implique l'assemblage d'un complexe ARN-protéine au niveau d'une structure en tige-boucle ou ARN SECIS (Selenocysteine Insertion Sequence) située dans la région 3' non codante de l'ARNm des sélénoprotéines. La protéine SBP2 se fixe spécifiquement à l'ARN SECIS et recrute les facteurs de la machinerie de biosynthèse. Elle fait également partie de complexes supramoléculaires dans le cytoplasme et le noyau, suggérant un possible assemblage nucléaire de la mRNP SECIS. Nous avons montré que la protéine SBP2 présentait une origine évolutive commune avec des protéines de la famille L7Ae. Ces protéines partagent un domaine de liaison à l'ARN similaire et participent à la construction de plusieurs RNP essentielles telles les sous-unités ribosomiques, les snoRNP (biogenèse des ribosomes), les snRNP (épissage), et les mRNP codant pour les sélénoprotéines. Nos objectifs sont d'élucider les principes d'interaction SBP2/SECIS, d'identifier les composants moléculaires des complexes qui se forment autour du SECIS et de comprendre leur assemblage.

En collaboration avec Edouard Bertrand (Montpellier) et Bruno Charpentier et Christiane Branlant (Nancy) nous avons identifié une machinerie d'assemblage des RNP L7Ae conservée de la levure à l'homme et d'importance fondamentale pour la cellule. Elle est constituée d'une protéine adaptatrice et d'un complexe de protéines chaperons. Notre objectif est de comprendre son rôle dans l'assemblage des mRNP de sélénoprotéines.



## ***Introduction générale***

Les particules ribonucléoprotéiques (ou RNP) sont à la base de nombreuses fonctions cellulaires fondamentales chez les eucaryotes. Au sein de ces particules, des ARN non codants participent à des mécanismes aussi variés que la traduction (ARNr, ARNt), l'épissage des ARN pré-messagers (UsnARN), la biogenèse des ribosomes et d'ARN non codants, la modification de bases des ARN (snARN, snoARN, scaARN), la réplication des télomères (télomérase) et la sécrétion des protéines (SRP). Enfin, ces dernières années ont vu l'émergence de microRNPs, contenant des ARN non codants (ARNsi, ARNmi) capables de moduler l'efficacité de la transcription, la stabilité des ARNm et vraisemblablement la structure de la chromatine.

La plupart des petits ARN non codants accomplissent leurs fonctions en association avec des protéines sous la forme de ribonucléoparticules (ou RNP). La formation des particules RNP est un processus très complexe qui nécessite de nombreuses étapes de maturation et de multiples facteurs d'assemblage (Fatica & Tollervey, 2002; Matera *et al.*, 2007 ; Yong *et al.*, 2004). En effet, le nombre de facteurs requis pour l'assemblage de RNP fonctionnelles dépasse souvent le nombre de protéines présentes au sein de la particule mature. Ainsi, les ribosomes sont constitués d'environ 80 protéines, mais n'utilisent pas moins de 140 facteurs pour leur assemblage (Fatica & Tollervey, 2002). Ces facteurs semblent non seulement importants pour faciliter l'assemblage de la particule, mais aussi pour exercer un contrôle strict sur la qualité des particules produites. Au niveau cellulaire, l'assemblage des RNP est également synonyme de mécanismes complexes de trafic intracellulaire car il peut avoir lieu dans des compartiments cellulaires différents du site fonctionnel. C'est le cas des particules UsnRNP impliquées dans les mécanismes d'épissage (Bertrand E & R., 2004; Carmo-Fonseca *et al.*, 2002; Yong *et al.*, 2004) : elles sont tout d'abord exportées dans le cytoplasme, où leur assemblage fait appel au complexe SMN (Yong *et al.*, 2004) ; puis réimportées dans le noyau vers leur site final de maturation et enfin leur site fonctionnel. En plus des RNP non codantes qui agissent en *trans*, la régulation de l'expression des gènes chez les eucaryotes est souvent dépendante de la formation de complexes RNP directement sur l'ARNm au niveau d'éléments structuraux régulateurs capables d'agir en *cis*.

Dans chacun des cas, la structure correcte des particules RNP est essentielle à leur fonction. Il est donc critique de comprendre comment ces particules sont formées dans la cellule. Au cours de ma carrière, je me suis intéressée à plusieurs aspects de ces mécanismes.

Durant ma thèse dirigée par Chantal Ehresmann (1990-1994), j'ai étudié le mode d'interaction de la protéine ribosomique S8 sur l'ARNr 16S d'*E. coli*. Cette protéine primaire joue un rôle central dans l'assemblage coordonné de la petite sous-unité ribosomique ainsi que dans la régulation de son propre opéron chez les procaryotes. Nous avons proposé un modèle de repliement tridimensionnel de son site ARN par analyse structurale en solution et modélisation.

Mon travail dans l'équipe de David Tollervey (1994-2001) a plus directement porté sur l'étude des mécanismes de maturation, d'assemblage et de dégradation de diverses RNP. J'ai notamment contribué à la caractérisation de l'exosome, un complexe d'exonucléases 3'→5' impliqué dans la maturation et la dégradation de divers ARN chez la levure. Notre travail a permis d'élucider les fonctions de l'exosome, notamment dans le noyau où ce complexe participe à la synthèse des ARN ribosomiques, des petits ARN nucléolaires (snoARN) et nucléaires (snARN) mais joue également un rôle dans les mécanismes de dégradation et de surveillance des ARN. La majorité des ARN stables qui constituent les RNP non codantes sont maturés à partir de précurseurs. Leur maturation ou dégradation implique en plus des exonucléases et endonucléases, une série de cofacteurs, d'hélicases et de protéines chaperons. Nous avons montré qu'un jeu limité de ces facteurs pouvait être recruté sous différentes combinaisons pour la synthèse de presque tous les ARN cellulaires. Deux types de protéines chaperons (Lhp1 et Lsm) ont été plus particulièrement analysées pour leur rôle dans la biogenèse du snoARN U3, des ARN ribosomiques et des ARNt. Ces protéines contribuent à faciliter les interactions ARN/protéines ainsi que les réarrangements structuraux lors de l'assemblage des particules ribonucléoprotéiques.

En 2001, j'ai été recrutée au grade de chargée de recherche au CNRS dans l'équipe d'Alain Krol qui étudie les mécanismes de synthèse des sélénoprotéines. L'incorporation de sélénocystéine dans les sélénoprotéines fait appel au recodage co-translationnel d'un codon UGA<sub>Sec</sub> en phase. Chez les eucaryotes, ce mécanisme implique l'assemblage d'un complexe ARN-protéine au niveau d'une structure en tige-boucle ou ARN SECIS (Selenocysteine Insertion Sequence) située dans la région 3'UTR de l'ARNm des sélénoprotéines (Allmang & Krol, 2006b). La protéine SBP2 se fixe spécifiquement à l'ARN SECIS et recrute les facteurs de la machinerie de biosynthèse. Elle fait également partie d'un complexe supramoléculaire et est soumise au transport nucléocytoplasmique suggérant un possible assemblage nucléaire de la mRNP SECIS (de Jesus *et al.*, 2006 ; Small-Howard *et al.*, 2006). Nous avons montré que la protéine SBP2 présentait une origine évolutive commune avec des protéines de la famille

L7Ae. Ces protéines partagent un domaine de liaison à l'ARN similaire et participent à la construction de plusieurs RNP essentielles telles les sous-unités ribosomiques, les snoRNP à boîtes C/D et H/ACA (biogenèse des ribosomes), la snRNP U4 (épissage), et comme nous l'avons montré dans les mRNP codant pour les sélénoprotéines. Nous étudions les principes d'interaction SBP2/SECIS et tentons d'identifier les composants moléculaires du complexe qui se forme autour du SECIS afin de comprendre quelle est l'origine de la diversité de fonctions apparues au cours de l'évolution pour les RNP L7Ae. Des déterminants de spécificité pour le SECIS et les ARN cibles des protéines L7Ae ont ainsi été identifiés en collaboration avec l'équipe de Christiane Branlant (Nancy).

Enfin, notre objectif est d'élucider le mécanisme d'assemblage de la mRNP SECIS et de comprendre comment il s'intègre dans le schéma d'assemblage général des RNP de la famille L7Ae mais aussi dans le mécanisme traductionnel des sélénoprotéines. En collaboration avec Edouard Bertrand (Montpellier) et Bruno Charpentier (Nancy) nous avons identifié une machinerie conservée destinée à l'assemblage des particules RNP stables de la famille L7Ae ainsi que de la RNP SECIS que nous caractérisons.

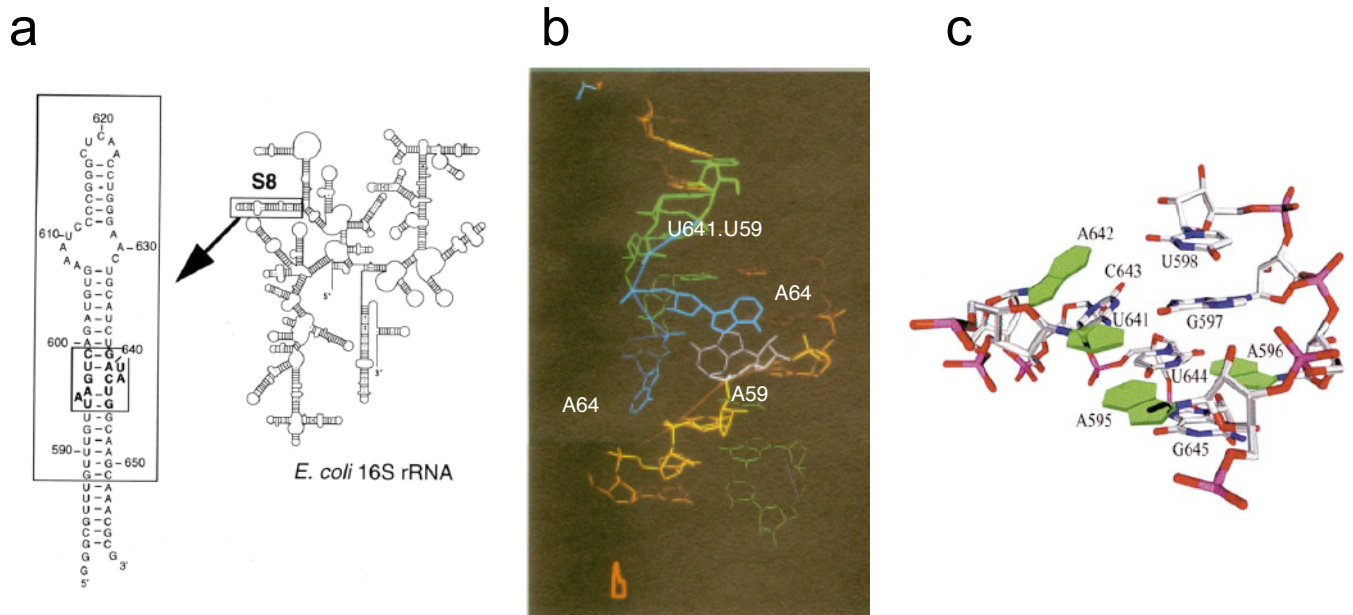


## ***Projet de thèse : Le site de fixation de la protéine ribosomique S8 sur l'ARNr 16S***

*Equipe du Pr. Bernard et du Dr. Chantal Ehresmann. IBMC, UPR 9002 du CNRS, Université Louis Pasteur, Strasbourg*

Mon travail de thèse a été effectué avec Marylène Mougel dans l'équipe de Bernard et Chantal Ehresmann. Il a consisté en l'ingénierie et l'étude structurale d'ARN en solution. En particulier, nous avons étudié le site de fixation de la protéine ribosomique S8 sur l'ARNr 16S d'*E. coli*. Cette protéine primaire se fixe dans le domaine central de l'ARNr 16S et permet la fixation coopérative de protéines secondaires par induction d'une modification conformationnelle de leur site ARN. La région de fixation de S8 représente l'un des sites de nucléation lors de l'assemblage de la sous-unité ribosomique 30S (Held *et al.*, 1974; Mizushima *et al.*, 1970). La protéine S8 intervient également dans la régulation de son propre opéron (spc). En se liant sur son ARNm, elle en inhibe la traduction et régule ainsi sa propre synthèse et celle des autres protéines ribosomiques de son opéron. En 1990, lorsque j'ai débuté ma thèse, il était admis que le site de fixation de S8 situé au sein d'une structure en tige boucle de l'ARNr 16S, était centré sur une région hélicoïdale irrégulière à la base de l'hélice et dont un nombre limité de nucléotides dictait la conformation. Le repliement de cette région était cependant très controversé et plusieurs modèles étaient en vigueur, dont l'un proposé par notre laboratoire (Mougel *et al.*, 1987). Nos travaux ont contribué à affiner la connaissance de ce site et à définir le site minimum d'ARN reconnu par S8 (Mougel *et al.*, 1993). Par la combinaison de techniques de mutagenèse dirigée, d'analyses structurales en solution et de modélisation graphique nous avons construit un nouveau modèle tridimensionnel de ce site avec Eric Westhof (Allmang *et al.*, 1994). Cette étude a révélé la présence de contraintes structurales importantes, conférant une géométrie et accessibilité particulière à certains résidus spécifiques du site ainsi qu'au squelette sucre-phosphate de l'ARN. Nous avons notamment proposé l'existence d'interactions non canoniques et d'un réseau important de liaisons hydrogènes. Depuis, ce modèle s'est avéré incorrect quant à la nature exacte des connections prédites, mais il avait mis en valeur la complexité du site. Prédire de telles interactions reste un défi dans l'étude du repliement des ARN et ce système modèle nous a mené aux limites des méthodologies employées. Après ma thèse, l'étude du site s'est poursuivie à l'aide de techniques plus adaptées à la détection d'interactions non canoniques multiples telle que le SELEX (Moine *et al.*, 1997) et enfin la résolution de la structure cristallographique du complexe S8-ARNr chez *Methanococcus jannashii* (Tishchenko *et al.*, 2001). La résolution de la structure du ribosome de *Thermus thermophilus*

a quant à elle permis de replacer ces interactions dans le contexte de la particule RNP du ribosome (Brodersen *et al.*, 2002 ; Yusupov *et al.*, 2001).



**Figure 1: Le site de fixation de la protéine S8 sur l'ARNr 16S. a.** Structure secondaire du site sur l'ARNr 16S **b.** Modèle de structure tridimensionnelle proposé en 1994 (Allmang *et al.*, 1994). Les trois adénines A595, A640 et A642 sont proposées en bulge et U641.U598 sont en interaction **c.** Structure cristallographique du site de fixation chez *Methanococcus jannashii* d'après Tishchenko *et al.* (2001). Deux plateformes nucléotidiques se font face. La première est constituée de l'appariement A595-A596, elle joue un rôle important dans l'empilement de bases. La seconde plateforme (U641-A642) est spécifiquement reconnue par S8. A642 joue un rôle essentiel dans la cohésion du complexe car elle est impliquée dans un réseau de liaisons hydrogènes important notamment avec l'interaction triple G597-C643.U641.

Un autre volet de ma thèse a constitué à mettre au point les techniques de synthèse chimique d'ARN à grande échelle, alors en plein développement, en vue d'études structurales par RMN ou cristallographie. L'ARN ribosomique 5S de *Xenopus laevis* possède plusieurs boucles de structure intrinsèque particulièrement stable. Elles ont été choisies comme système modèle. Nous avons synthétisé chimiquement deux tige-boucles de cet ARN et réalisé leur étude par RMN. Il est apparu que celles-ci établissaient des interactions intermoléculaires pour former un duplex.

Enfin, les techniques de cartographie m'ont permis, en collaboration avec David Gilmer dans l'équipe de Gérard Jonard (IBMP, Strasbourg), de proposer un modèle de structure secondaire de la région 5' non codante de l'ARN du virus des nervures jaunes et nécrotiques de la betterave (BNYVV). Ces résultats ont confirmé l'existence de trois domaines appariés dans l'ARN, importants pour sa réplication.

## ***Stage post-doctoral (1994-1996) : Mécanismes de maturation des pré-ARNr***

*Equipe de David Tollervey. Département d'Expression Génétique, EMBL, Heidelberg, Allemagne*

La biogenèse des ribosomes eucaryotes est un mécanisme complexe et dynamique qui a lieu successivement dans le nucléole, le nucléoplasme et le cytoplasme. Ce processus englobe les étapes de transcription, maturation, modifications des précurseurs d'ARN ribosomiques (pré-ARNr), leur assemblage en sous-unités ribosomiques et enfin leur transport. Plus d'une centaine de petits ARN nucléolaires (snoARN) servent de guides pour les modifications de l'ARN et environ 140 protéines non-ribosomiques sont impliquées dans les étapes de maturation (Fatica & Tollervey, 2002).

L'ARNr 18S de la petite sous-unité et les deux ARNr 5,8S et 25S de la grande sous-unité ribosomique sont transcrits par l'ARN polymérase I sous la forme d'un long précurseur unique: le pré-ARNr 35S. Les régions correspondant aux ARN matures du pré-ARNr 35S sont flanquées d'espaceurs externes en 5' et en 3' (5' et 3' ETS) et séparées par des espaceurs internes (ITS1 et ITS2). Les différentes étapes de maturation des pré-ARNr ainsi que les enzymes impliquées sont représentées Figure 2 (pour une revue voir Venema & Tollervey, 1999). Il existe une séparation entre la voie de synthèse de l'ARN 18S, qui implique quatre coupures successives par des endonucléases, et la voie de synthèse des ARN 5,8S et 25S, plus complexe et qui fait appel à une coupure endonucléolitique suivie par des étapes multiples de digestion par des exonucléases. Une fois modifiés les ARNr matures s'assemblent avec les 80 protéines ribosomiques et l'ARNr 5S qui est transcrit indépendamment.

Le laboratoire du Pr. David Tollervey, dans lequel j'ai effectué mon stage post-doctoral, a largement contribué à l'élucidation des étapes de maturation des pré-ARNr chez la levure et à l'identification d'un nombre important de facteurs impliqués dans le mécanisme. Mon travail post-doctoral dans cette équipe a contribué à identifier certains des signaux de coupures sur le pré-ARNr mais aussi à caractériser plusieurs facteurs de maturation et à analyser leur fonction.



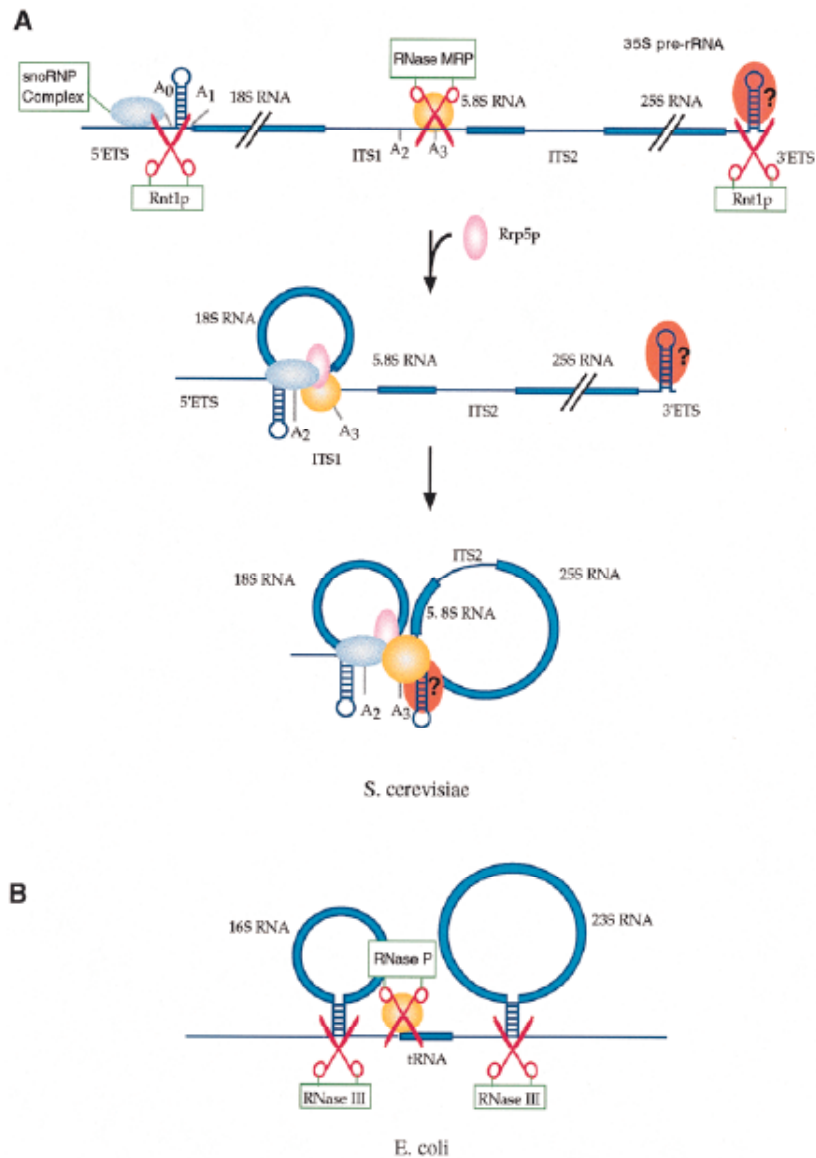
## 1- Mise en évidence d'une coordination de la maturation des pré-ARNr

Je me suis particulièrement intéressée aux mécanismes de maturation au sein de deux espaceurs (ITS1 et 3'ETS). ITS1 contient deux sites de clivage importants : A2 et A3. Le clivage au niveau du site A2 est une étape cruciale dans la maturation du pré-ARNr 35S qui permet de scinder le pré-ARNr en deux fragments 5' et 3' destinés respectivement à former les ARN de la petite et de la grande sous-unité ribosomique. Il est dépendant de signaux en *cis* à proximité de A2 mais également dans la région 5'ETS, tel que le site de fixation du snoARN U3 (Beltrame *et al.*, 1994; Venema *et al.*, 1995). Le clivage du site A2 dépend d'un complexe de maturation composé de snoRNP. Le clivage au niveau du site A3, assuré par la RNase MRP, permet quant à lui de générer l'extrémité 5' de la forme majeure de l'ARN 5,8S. La région 3'ETS est coupée par Rnt1 (homologue de la RNase III), ce clivage initie l'ensemble du mécanisme de maturation du pré-ARNr 35S. Mon travail a contribué à identifier les signaux requis en *cis* pour le clivage des régions ITS1 et 3'ETS (Allmang *et al.*, 1996b; Allmang & Tollervey, 1998). Cette analyse a permis de révéler un lien tout à fait inattendu entre les sites de clivage A2 et A3 (Allmang *et al.*, 1996a), suggérant que le complexe de maturation du site A2 et 5'ETS (snoRNP) interagissait avec celui du site A3 (RNase MRP). Par ailleurs, nous avons montré qu'une structure en tige-boucle en 3'ETS est nécessaire et suffisante pour la coupure par Rnt1p, mais également pour le clivage à distance du site A3 par la RNase MRP. La maturation des régions 3'ETS et ITS1 apparaît donc, elle aussi, couplée (Allmang & Tollervey, 1998). L'ensemble de nos résultats a suggéré une coordination de la maturation des pré-ARN qui peut être mise en parallèle avec les mécanismes de maturation bactériens (voir Figure 3). Chez les eubactéries, les extrémités non codantes situées en 5' et 3' des ARNr matures s'apparient et sont coupées par la RNase III. Un ARNt présent dans l'espaceur fournit un site de clivage pour la RNase P. Chez la levure, des tiges-boucles en 5' et 3'ETS constituent les sites de clivage de Rnt1p. Des complexes de snoRNP remplacent les appariements intramoléculaires afin d'assurer la coordination de la maturation en 5' et 3'. La RNase MRP assure un rôle comparable à celui de la RNase P dans l'espaceur ITS1, mais interagit également avec les complexes de maturation en 5' et 3' ETS.

## 2- Démonstration de l'activité endonucléolytique de la RNase MRP *in vitro*

En accord avec sa localisation nucléolaire, de nombreux arguments génétiques suggéraient que la RNase MRP intervenait dans la maturation du pré-ARNr *in vivo*. Le rôle de la RNase MRP était cependant controversé car elle fut initialement identifiée comme une endonucléase mitochondriale (Chang & Clayton, 1987). En collaboration avec Bertrand Séraphin (alors à l'EMBL), un système capable de reproduire avec précision la coupure au site A3 par la RNase MRP *in vitro* a été mis au point. Ces expériences nous ont permis de démontrer que la RNase MRP était bien directement responsable de la coupure du site A3 et de

confirmer son rôle dans le nucléole. C'est la première endonucléase pour laquelle nous avons pu démontrer sa capacité à couper un précurseur d'ARNr *in vitro* et *in vivo* (Lygerou *et al.*, 1996).



**Figure 3 : Coordination de la maturation des pré-ARNr** d'après Allmang et Tollervey (1998). Modèle comparant les mécanismes de maturation chez la levure (A) et *E. coli* (B).

La vision du mécanisme de maturation des pré-ARNr a beaucoup évolué depuis le travail réalisé au cours de mon stage post-doctoral. Pour autant, tous les enzymes de maturation et de modification des pré-ARNr ne sont toujours pas identifiés mais ces dernières années ont vu des avancées spectaculaires dans la compréhension des mécanismes d'assemblage des pré-ARN et des ARNr. En effet, la combinaison des méthodes de purification de complexes protéiques en tandem et d'analyses protéomiques à haut débit a permis l'analyse de la composition des particules pré-ribosomiques (Gavin *et al.*, 2006; Gavin *et al.*, 2002 ; Ho *et al.*, 2002). Ces études ont permis de dresser une carte d'assemblage des particules pré-ribosomiques qui révèle un processus dynamique et d'une complexité inattendue.

## ***Poste de chargée de recherche l'Université d'Edimbourg (1996-2001) : L'exosome et la synthèse des ARN stables***

*Research Fellow dans l'équipe du Pr. David Tollervey. Wellcome Trust Centre for Cell Biology, Université d'Edimbourg, Ecosse.*

Dans la continuité du travail amorcé à l'EMBL (Heidelberg), notre équipe, une fois installée au «Wellcome Trust Center for Cell Biology» de l'Université d'Edimbourg (Ecosse) a continué à s'intéresser aux mécanismes de maturation des ARN. La recherche d'enzymes impliquées dans la maturation des pré-ARNr a conduit à l'identification d'un complexe d'exonucléases multifonctionnel : l'exosome. J'ai participé à sa caractérisation et à son analyse fonctionnelle.

### **1- L'exosome et le complexe PM-Scl humain**

L'exosome est un complexe d'exonucléases 3'->5' impliqué dans la maturation et la dégradation de divers ARN chez la levure. Il fut initialement identifié par Phil Mitchell dans notre équipe comme un complexe de 5 exonucléases essentielles, toutes impliquées dans la maturation de l'extrémité 3' de l'ARNr 5,8S (Mitchell *et al.*, 1997). La combinaison d'analyses biochimiques et génétiques nous a conduit à identifier six nouveaux composants du complexe décrits Tableau I (Allmang *et al.*, 1999b). A notre grande surprise, à l'exception de Rrp6p, tous étaient essentiels et participaient à la maturation de l'ARN 5,8S. La majorité des composants identifiés sont des homologues d'exonucléases 3'->5' bactériennes. L'activité d'un certain nombre d'entre eux a été démontrée *in vitro* (voir Tableau I).

Nous avons par ailleurs identifié les homologues humains de 9 des composants de l'exosome. Deux des exonucléases identifiées chez la levure sont homologues de protéines du complexe PM-Scl humain (voir Tableau I). Le complexe PM-Scl comporte onze à seize polypeptides reconnus par les anticorps de malades souffrant de la maladie auto-immune de polymyosite (Polymyositis-scleroderma overlap syndrome). Des sérums de patients atteints de polymyosite, fournis par le Pr. van Venrooij (Université de Nimègue, Pays-Bas), m'ont permis d'établir le lien entre le complexe PM-Scl humain et l'exosome. Nous avons montré que l'homologue du composant Rrp4p de l'exosome faisait partie du complexe PM-Scl humain et démontré que ce dernier était bien l'homologue fonctionnel de l'exosome (Allmang *et al.*, 1999b). Nous l'avons confirmé en clonant et identifiant trois autres composants du complexe humain (Brouwer *et al.*, 2001). Ces expériences ont, pour la première fois, conduit à l'identification de la cible des auto-anticorps de patients atteints de polymyosite dont la nature était inconnue.

En examinant la distribution cellulaire des composants de l'exosome et du complexe PM-Scl par immunolocalisation et fractionnement biochimique nous avons mis en évidence deux formes du complexe, l'une dans le noyau, l'autre dans le cytoplasme. Ces complexes partagent 10 composants communs, mais diffèrent par la présence de Rrp6p/PM-Scl-100 (Allmang *et al.*, 1999b) et Rrp47 (travaux ultérieurs Mitchell *et al.*, 2003) dans le complexe nucléaire ; et de la GTPase Ski7 dans le complexe cytoplasmique (Araki *et al.*, 2001). Ces deux complexes assurent des fonctions différentes dans les deux compartiments cellulaires (voir ci-dessous).

Exosome de levure	Exosome humain	Domaine conservés / Commentaires
Rrp41p*	<b>hRrp41p</b> 35% (55%)	RNase PH. Homologue de la PNPase d'E. coli
Rrp42p	hRrp42p 25% (51%)	RNase PH.
Rrp43p		RNase PH.
<b>Rrp45p</b>	PM-Scl 75 38% (64%)	RNase PH.
<b>Rrp46p</b>	<b>hRrp46p</b> 35% (48%)	RNase PH.
<b>Mtr3p*</b>		RNase PH.
Rrp4p*	hRrp4p 43% (52%)	domaine S1; domaine KH. Présent dans la PNPase d'E. coli
<b>Rrp40p</b>	<b>hRrp40p</b> 35% (48%)	domaine S1: domaine KH
<b>Csl4p</b>	hCsl4p 48% (56%)	domaine S1
Rrp44p/Dis3p*	hDis3p 45%	RNase R (RNase II)
<b>Rrp6p*</b>	PM-Scl 100 32% (52%)	RNase D. Composant exclusivement nucléaire.
Rrp47		Protéine de liaison à l'ARN Composant exclusivement nucléaire
Ski7		GTPase putative. Composant exclusivement cytoplasmique.

**Tableau 1 : Les composants de l'exosome.** Les protéines dont l'activité catalytique a été démontrée *in vitro* sont marquées par un astérisque. Pour les homologues humains, les pourcentages d'identité (de similarité) sont indiqués. Les composants communs aux complexes nucléaires et cytoplasmiques sont surlignés en gris. Les exonucléases que j'ai contribué à identifier sont indiquées en rouge.

Des composants de l'exosome ont maintenant été trouvés chez la drosophile, les plantes, le trypanosome et les archae (Raijmakers *et al.*, 2004). Par ailleurs, la structure cristallographique de l'exosome de l'archae *Sulfolobus solfataricus* a été résolue. Il s'agit d'un

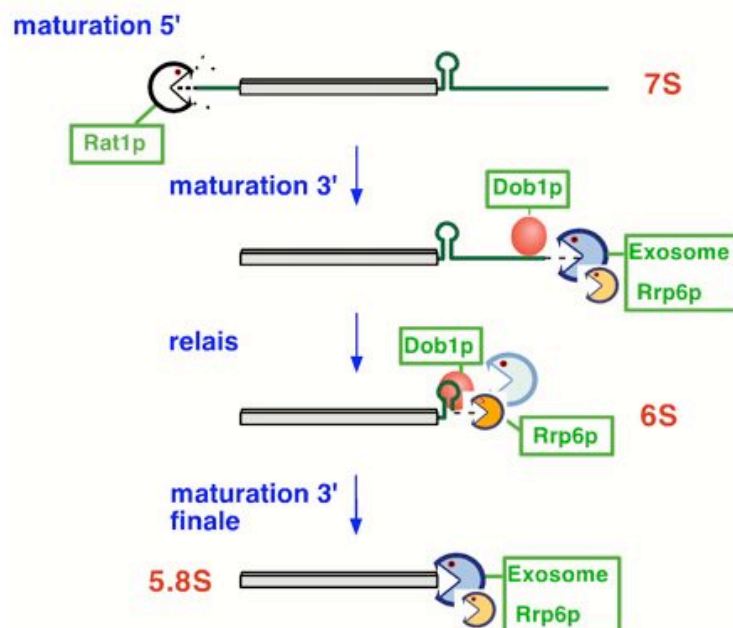


anneau hexamérique composé de 3 RNases PH actives et de trois RNases PH inactives surmonté d'un trimère de protéines de liaison à l'ARN (Buttner *et al.*, 2005; Lorentzen *et al.*, 2007; Lorentzen *et al.*, 2005). Les importantes similitudes de séquence avec les composants de l'exosome eucaryote et les similarités structurales avec les exonucléases bactériennes permettent de proposer une origine commune pour les machineries de dégradation des ARN dans les trois domaines du vivant.

## 2- Les fonctions de l'exosome

L'existence d'un complexe composé d'un si grand nombre d'exonucléases, n'avait pas manqué de soulever de nombreuses questions. Pourquoi de si nombreuses activités sont-elles présentes au sein du complexe ? Différentes exonucléases ont-elles différentes fonctions ou sont-elles impliquées dans la maturation d'un même substrat ? Existe-t-il plusieurs substrats? Nous avons pu apporter une série de réponses afin d'étayer chacune de ces hypothèses.

Dans le noyau, il apparaît que tous les composants de l'exosome sont requis pour la synthèse de l'extrémité 3' d'un même substrat: l'ARNr 5,8S. Mais, au cours de ce processus, trois étapes peuvent être résolues, impliquant deux changements d'exonucléases. Différentes exonucléases ont donc différentes fonctions au sein du complexe. L'hélicase putative Dob1p (Mtr4) (de la Cruz *et al.*, 1998) fonctionne avec l'exosome dans chacune de ces étapes. Dans le cytoplasme, nous avons montré que l'exosome co-sédimentait avec l'hélicase Ski2p impliquée dans les mécanismes de dégradation des ARNm (Anderson & Parker, 1998).



**Figure 4 : Modèle de mécanisme de maturation de l'ARNr 5,8S par l'exosome.** La première étape de maturation en 3' du pré-ARN 7S nécessite tous les composants de l'exosome. Rrp6p prend le relais et est

spécifiquement requise pour l'obtention du pré-ARNr 6S. Enfin, la maturation finale en ARNr 5,8S implique à nouveau l'ensemble des exonucléases. Chacune des étapes est dépendante de l'hélicase putative Dob1p/Mtr4p.

Nous avons également identifié de nouveaux substrats nucléaires de l'exosome, en démontrant qu'il était impliqué dans la synthèse des snoARN et des snARN (Allmang *et al.*, 1999a) ainsi que dans diverses étapes de la maturation du pré-ARNr (Allmang *et al.*, 2000). Chez les eucaryotes, les petits ARN nucléolaires (ou snoARN) jouent un rôle majeur dans la maturation et la modification des pré-ARNr. La plupart des snoARN sont codés par des introns ou synthétisés sous la forme d'un précurseur polycistronique. Dans chacun des cas, leur excision requiert des mécanismes de maturation par des endonucléases et exonucléases. Rnt1p clive les précurseurs polycistroniques; l'épissage initie la synthèse des snoARN introniques. J'ai montré que la synthèse de leur extrémité 3' était alors dépendante de l'action de l'exosome. L'extrémité 5' est générée par l'exonucléase 5' → 3' Rat1p (Petfalski *et al.*, 1998). Ce processus est multiphasique et plusieurs exonucléases du complexe ont des fonctions distinctes. Le composant Rrp6p de l'exosome est spécifiquement impliqué dans l'étape finale de maturation en 3' (Allmang *et al.*, 1999a). Le pré-snoARN U3 et les pré-snARN U1, U4 et U5, bien que synthétisés à partir de leurs propres promoteurs, sont maturés en 3' par les mêmes enzymes.

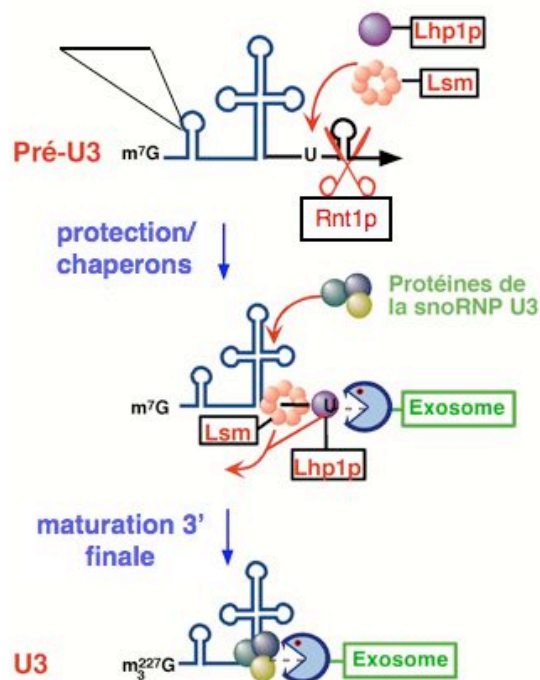
Un équilibre entre maturation et dégradation a pu être mis en évidence pour tous les substrats de l'exosome. L'exosome joue en effet un rôle important dans les mécanismes de dégradation des ARN, comme celle des précurseurs d'ARNr aberrants (Allmang *et al.*, 2000).

D'autres travaux ont révélé que l'exosome fonctionnait également dans la dégradation nucléaire de pré-ARNm (Bousquet-Antonelli *et al.*, 2000) suggérant un rôle potentiel dans la régulation de l'expression des gènes. Plus récemment, une partie du travail que j'avais initié a été poursuivie et complétée par Laura Milligan et Claire Torchet. Elle a conduit à démontrer que le composant exclusivement nucléaire de l'exosome (Rrp6) était impliqué dans les mécanismes de surveillance des ARN. En effet, dans une souche mutée au niveau la poly(A) polymérase et qui conduit à un ralentissement de polyadénylation des ARNm, les ARNm sont détectés par l'activité de surveillance de Rrp6p, déadénylés et rapidement dégradés par l'exosome (Milligan *et al.*, 2005).

De nombreux autres travaux réalisés par la suite dans l'équipe de David Tollervey et ailleurs ont imposé l'exosome comme un acteur clé de la machinerie de surveillance des ARN. L'exosome intervient dans tous les types mécanismes de surveillance cytoplasmique des ARNm, tels la dégradation des ARNm sans codon de terminaison (non-stop decay), des ARNm à codon non-sens (non-sense-mediated decay) et des ARNm sujets à des arrêts prématurés de traduction (no-go decay) (pour une revue voir Houseley *et al.*, 2006).

### 3- Mécanismes de synthèse des ARN stables

La quasi-totalité des ARN de la cellule est synthétisée à partir de précurseurs. L'une des observations les plus frappantes de notre étude est que la maturation ou dégradation de ces ARN implique un jeu bien défini mais limité d'enzymes et de cofacteurs. Les enzymes majeures sont le complexe de l'exosome, les exonucléases 5'→3' Rat1p, Xrn1p et les endonucléases RNaseP/MRP et Rnt1p. Les cofacteurs incluent les hélicases Dob1p et Ski2p ainsi que des protéines chaperons telles que Lhp1p et le complexe Lsm. Nous avons analysé systématiquement le rôle de chacun de ces facteurs dans la synthèse des ARN stables (snoARN, snARN, ARNr et ARNt). Notre objectif était de comprendre le rôle relatif de ces facteurs, notamment dans la détermination de l'équilibre entre maturation et dégradation des ARN par l'exosome.



**Figure 5 : Modèle simplifié du mécanisme de synthèse du snoARN U3.** Le précurseur de U3 se caractérise par la présence de séquences poly(U), d'une structure en tige boucle en 3' et d'un intron. Le précurseur est initialement clivé par Rnt1p. Les protéines Lhp1p et Lsm se lient aux séquences poly(U), protégeant le précurseur de la dégradation par l'exosome. Elles sont vraisemblablement déplacées au moment de la fixation des protéines de la snoRNP U3.

Avec Joanna Kufel, nous nous sommes notamment intéressés à la synthèse du snoARN U3. Le snoARN U3 est transcrit par l'ARN polymérase II; son précurseur est alors clivé par Rnt1p et mûri en 3' par l'exosome (Kufel *et al.*, 2000). Nous avons établi que d'autres facteurs étaient impliqués, notamment Lhp1p et le complexe Lsm. Lhp1p est une protéine chaperon qui se fixe sur la région poly (U) en 3' des transcrits de l'ARN polymérase III (Pannone *et al.*, 1998; Rinke & Steitz, 1982; Yoo & Wolin, 1997). Nous avons montré que Lhp1p stabilisait l'extrémité 3' du pré-ARN U3 en s'y fixant. Le complexe Lsm, composé d'un anneau de 7 protéines (Achsel *et al.*, 1999; Bouveret *et al.*, 2000; Mayes *et al.*, 1999) joue lui aussi ce rôle en coordination avec Lhp1 (Kufel *et al.*, 2003b). Ces deux facteurs favorisent la maturation de l'extrémité 3' en protégeant le précurseur de la dégradation par l'exosome tant que les protéines de la snoRNP mature ne sont pas fixées. Les protéines de la particule U3 déplaceraient les protéines chaperons, permettant la maturation finale de l'ARN (Kufel *et al.*, 2000). Ce type de mécanisme est vraisemblablement ubiquitaire, car nous avons montré que les protéines chaperons Lhp1p et Lsm fonctionnaient également dans la synthèse de nombreux autres ARN, tels les ARNt et pré-ARNr (Kufel *et al.*, 2003a; Kufel *et al.*, 2002). Il en va de même pour l'exosome, les exonucleases 5'→3' et les hélicases. Ces différents facteurs semblent pouvoir être recrutés sous différentes combinaisons et vers différents substrats pour en assurer la maturation et la dégradation.

Une question restait en suspens, à savoir comment se fait la discrimination des différents substrats et leur orientation vers les voies de maturation ou de dégradation. Ces dernières années ont vu d'importants développements dans la compréhension de ces mécanismes, en particulier en ce qui concerne la connaissance des signaux et mécanismes d'activation et de régulation de l'exosome. Plusieurs cofacteurs ont été identifiés, le plus impressionnant est certainement le complexe TRAMP dont la fonction semble être de diriger les précurseurs d'ARN défectueux vers l'exosome en les polyadénylant (Houseley *et al.*, 2006; LaCava *et al.*, 2005).

# ***Activité de recherche et projets scientifiques actuels : Le mécanisme de synthèse des sélénoprotéines***

*Equipe du Dr. Alain Krol. Unité Architecture et Réactivité de l'ARN, CNRS, Université Louis Pasteur, IBMC, Strasbourg.*

## **A- Introduction**

Le sélénium est un oligo-élément essentiel. Son importance physiologique n'a été appréciée à sa juste valeur que depuis les années 1970 avec l'identification de la forme biologique majeure du sélénium, l'acide aminé sélénocystéine qui est incorporé dans les sélénoprotéines par une machinerie de traduction spécialisée (Flohe *et al.*, 2000).

Les premières sélénoprotéines identifiées étaient essentiellement des enzymes utilisant le potentiel d'oxydoréduction du sélénium dans leur site actif pour la lutte contre les radicaux libres oxygénés, telles les glutathion peroxydases (Flohe *et al.*, 1973 ; Rotruck *et al.*, 1973). L'identification récente de nouvelles sélénoprotéines montre qu'elles sont impliquées dans une grande variété de fonctions telles que le transport, la signalisation, la structure ou le développement musculaire (pour une revue voir Moghadaszadeh & Beggs, 2006). Celles-ci peuvent être intracellulaires, transmembranaires ou sécrétées et leur expression est tantôt ubiquitaire, tantôt tissu spécifique. On trouve des sélénoprotéines chez les archae, les bactéries et les eucaryotes mais elles ne sont pas représentées de façon égale dans ces trois règnes (Castellano *et al.*, 2004; Kryukov & Gladyshev, 2004). Des mécanismes de biosynthèse différents sont mis en jeu chez les bactéries, les archae et les eucaryotes (pour une revue voir Allmang & Krol, 2006b).

L'équipe d'Alain Krol, que j'ai rejointe en 2001, a contribué à l'identification de nouvelles sélénoprotéines mais également à l'élucidation du mécanisme de synthèse des sélénoprotéines chez les eucaryotes. J'ai intégré cette dernière thématique et développé un nouveau sujet en étudiant le rôle de facteurs d'assemblage dans le mécanisme.

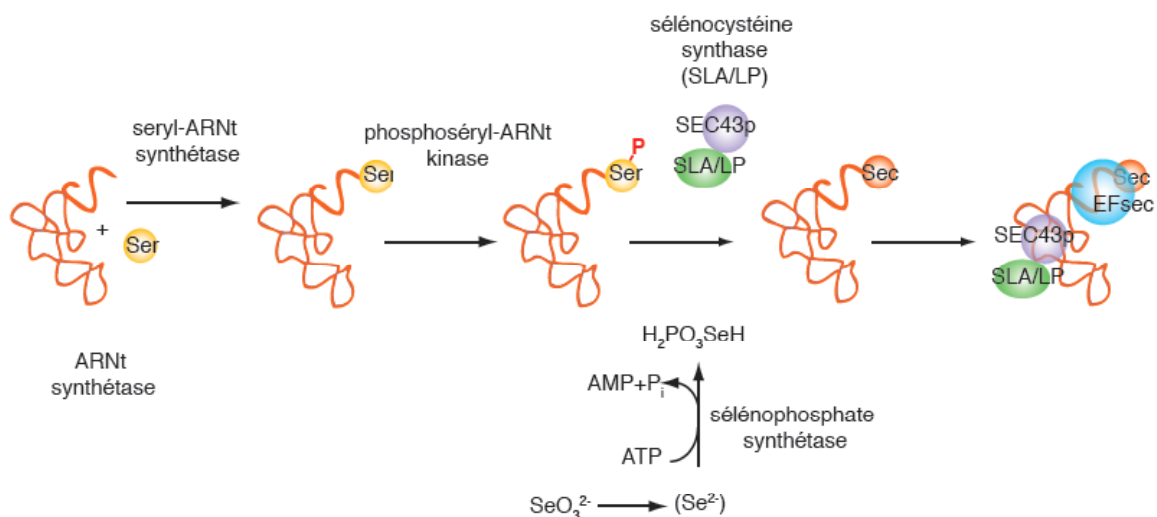
La sélénocystéine (Sec) est considérée comme le 21<sup>e</sup> acide aminé. Cet analogue de la cystéine dont le groupement thiol est remplacé par un groupement séléinol est incorporé dans la chaîne peptidique de façon co-translationnelle en réponse à un codon UGA habituellement reconnu comme l'un des trois codons de terminaison. Deux étapes majeures peuvent être résolues : la biosynthèse de la sélénocystéine et son incorporation par recodage du codon UGA<sub>Sec</sub>. Chez les eucaryotes, ces mécanismes sont particulièrement complexes et coordonnés par des facteurs capables de s'organiser en complexes supramoléculaires (Figures 6 et 7).

### *La biosynthèse de la sélénocystéine*

La sélénocystéine n'existe pas en tant qu'acide aminé libre et c'est la sérine qui est, dans un premier temps, chargée sur l'ARNt<sup>Sec</sup> par la sérine-ARNt synthétase conventionnelle avant

d'être convertie en sélénocystéine directement sur l'ARNt<sup>Sec</sup> par une sélénocystéine synthase (Hendrickson, 2007). Deux équipes viennent d'élucider ce mécanisme longtemps controversé (voir Figure 6). Les travaux de (Xu *et al.*, 2007b) et de (Yuan *et al.*, 2006) ont démontré que la sélénocystéine synthase (SecS ou SepSecS) est la protéine SLA/LP (Soluble Liver Antigen/Liver Pancreas) identifiée précédemment en complexe avec l'ARNt<sup>Sec</sup> chez des patients souffrant d'hépatite chronique autoimmune (Costa *et al.*, 2000; Kernebeck *et al.*, 2001). Cette protéine, de la famille des transférase à phosphate de pyridoxal, utilise du sélénophosphate et un intermédiaire O-phosphoséryl-ARNt<sup>Sec</sup> pour générer le Sec-tRNA<sup>Sec</sup>. L'intervention d'une O-phosphoséryl-tRNA(Sec) kinase (PSTK) dans la production de cet intermédiaire a également été démontrée (Carlson *et al.*, 2004 ; Xu *et al.*, 2007b). Enfin, la synthèse du sélénophosphate est assurée par la sélénophosphate synthétase (ou SPS2) à partir de sélénite et d'ATP (Xu *et al.*, 2007a). SPS2 est elle-même une sélénoprotéine, suggérant l'existence d'une régulation du mécanisme de biosynthèse en fonction de la biodisponibilité du sélénium. Une autre sélénophosphate synthétase (SPS1) identifiée précédemment (Low *et al.*, 1995) aurait pour fonction de générer un niveau basal de sélénocystéine nécessaire à la synthèse de SPS2.

Plusieurs enzymes du mécanisme de biosynthèse ont été trouvées associées sous forme de complexe. De façon étonnante, la protéine SECp43 impliquée dans la 2'O-méthylation de la base mcm5U34 de l'ARNt<sup>Sec</sup> (Ding & Grabowski, 1999), a été trouvée associée à SLA/LP et à l'ARNt<sup>Sec</sup> *in vivo* (Xu *et al.*, 2005). SLA/LP et SPS1 sont également en interaction et SEC43p servirait de chaperon pour localiser SLA/LP et SPS1 dans le noyau (Small-Howard *et al.*, 2006). Le rôle de cette redistribution cellulaire reste à être élucidé.



**Figure 6 : Mécanisme de biosynthèse de la sélénocystéine chez les eucaryotes.** L'ARNt<sup>Sec</sup> est aminoacylé avec de la sérine par la séryl-ARNt synthétase. Le Ser-ARNt<sup>Sec</sup> est phosphorylé par la phosphoséryl-ARNt kinase. La sélénocystéine synthase (SecS ou SLA/LP) assure la conversion de la sérine phosphorylée en sélénocystéine. La

synthèse du sélénophosphate nécessaire à cette étape est catalysée par une sélénophosphate synthétase. L'ARNt<sup>Sec</sup> est pris en charge par le facteur d'élongation spécialisé eEFsec.

### Le mécanisme d'incorporation de sélénocystéine

Chez les eucaryotes, le recodage du codon UGA<sub>Sec</sub> dépend de l'interaction de plusieurs complexes ARN-protéine (voir Figure 7). Le premier est constitué par l'ARNt<sup>Sec</sup> et le facteur d'élongation spécialisé mSelB/eEFsec (Fagegaltier *et al.*, 2001). Un second complexe est formé directement sur l'ARNm des sélénoprotéines au niveau d'une tige-boucle ou élément SECIS pour **S**Elenocysteine **I**nsertion **S**equences situées dans la région 3' non traduite (3' UTR) (Berry *et al.*, 2001). La protéine SBP2 (Secis Binding Protein 2) se lie à l'élément SECIS (Copeland *et al.*, 2000 ; Lescure *et al.*, 2002) et recrute les facteurs de la machinerie d'incorporation. SBP2 interagit notamment avec eEFsec lorsque celui-ci est lié à l'ARNt<sup>Sec</sup> (Zavacki *et al.*, 2003) pour le canaliser vers le codon UGA<sub>Sec</sub>. SBP2 est également capable d'interagir avec le ribosome (Copeland *et al.*, 2001). La protéine ribosomique L30 est, quant à elle, capable de se lier à l'élément SECIS (Chavatte *et al.*, 2005). L30 entre en compétition avec SBP2 pour la liaison au SECIS et stimule le recodage du codon UGA<sub>Sec</sub> *in vivo*, constituant de ce fait un composant de la machinerie de recodage. Il est probable qu'en se liant au SECIS, L30 déplace SBP2 lui permettant de délivrer le complexe eEFsec/Sec-ARNt<sup>Sec</sup> près du site de décodage du ribosome.

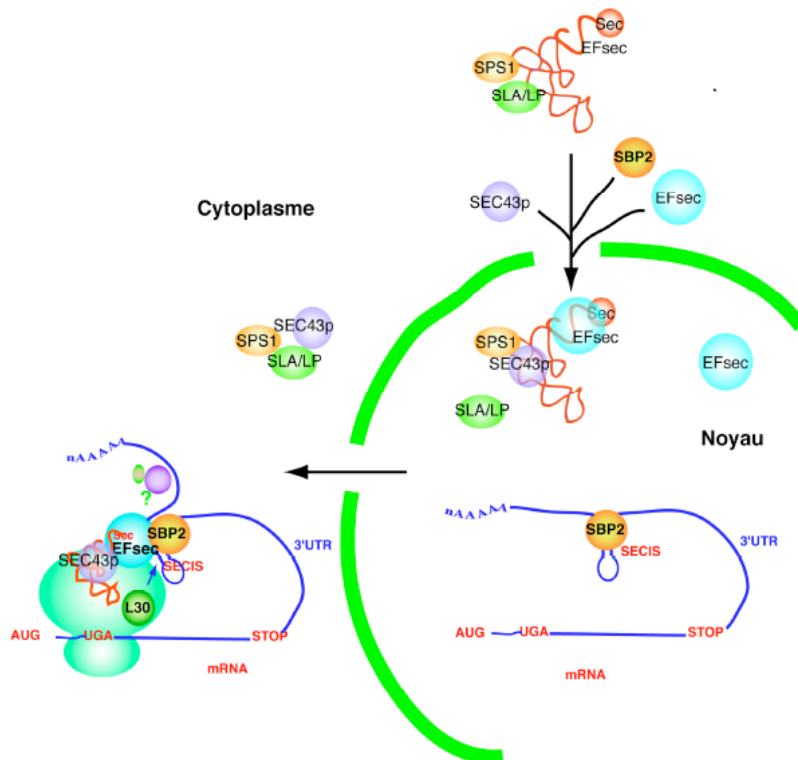


Figure 7 : Mécanisme postulé pour la synthèse des sélénoprotéines chez les eucaryotes

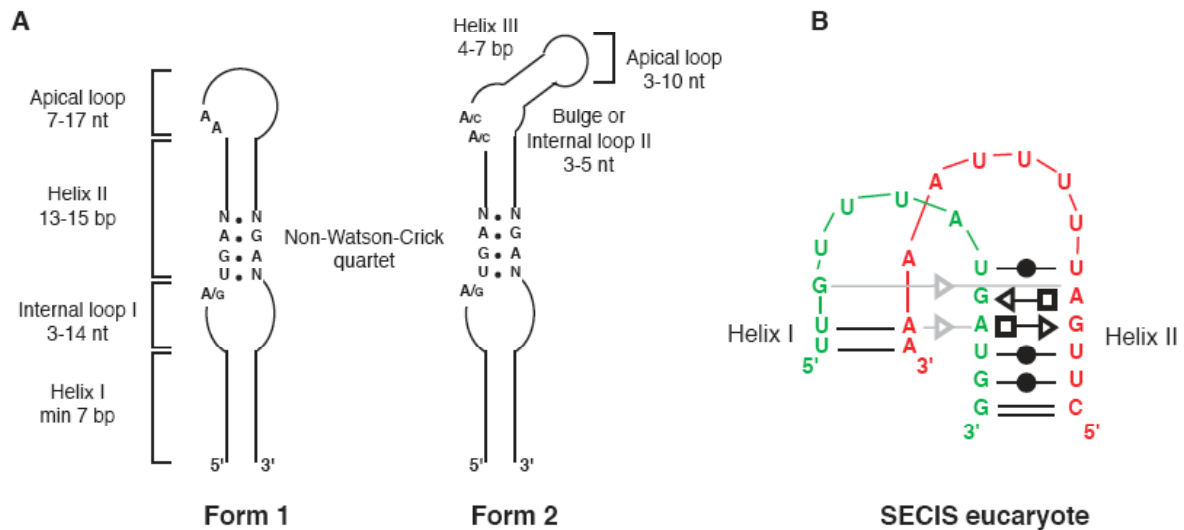
Plusieurs modèles ont été proposés pour expliquer ce mécanisme selon que SBP2 est initialement associée au ribosome ou à l'ARN SECIS (Chavatte *et al.*, 2005; Kinzy *et al.*, 2005).

Des résultats récents ont montré que SBP2 faisait partie de complexes supramoléculaires et est présente à la fois dans le cytoplasme et le noyau. En effet, il apparaît que la protéine SECp43 trouvée associée à l'ARNt<sup>Sec</sup> et aux facteurs de la biogenèse de la sélénocystéine (voir paragraphe précédent) est également capable de promouvoir l'interaction entre eEFsec et SBP2 *in vivo* (Small-Howard *et al.*, 2006). SECp43 influence par ailleurs la localisation nucléaire de ces protéines. Des signaux de localisation et d'export nucléaire ont pu être identifiés pour eEFsec et SBP2 et un assemblage nucléaire précoce des facteurs du mécanisme d'incorporation de sélénocystéine sur l'ARN SECIS a été proposé (de Jesus *et al.*, 2006). La séquestration nucléaire de SBP2 peut être induite par un stress oxydatif et l'oxydation de cystéines essentielles qui la rendent incapable d'interagir avec le facteur d'export nucléaire CRM1 (Papp *et al.*, 2006). Ceci a pour conséquence une diminution de l'incorporation de sélénocystéine et pourrait également représenter un moyen de régulation de l'expression des sélénoprotéines en fonction du statut redox de la cellule. Il est également vraisemblable que l'assemblage du complexe dans le noyau permet d'éviter que les ARNm de sélénoprotéines ne soient soumis aux mécanismes de dégradation des ARN à codon non-sens ou nonsense-mediated decay (NMD) (de Jesus *et al.*, 2006).

## **B- Les interactions autour de l'ARN SECIS**

L'interaction entre SBP2 et l'ARN SECIS est au cœur du processus de synthèse des sélénoprotéines. La combinaison d'analyses structurales en solution et d'analyses bioinformatiques a permis de proposer un modèle de structure secondaire de l'élément SECIS (Fagegaltier *et al.*, 2000b ; Fletcher *et al.*, 2001; Walczak *et al.*, 1997). Il s'agit d'une hélice-bulle interne - hélice surmontée d'une boucle apicale de taille variable (voir Figure 8). A l'exception d'une succession d'adénines/cytosines dans la boucle apicale, tous les nucléotides conservés se situent dans l'hélice supérieure. Celle-ci comporte quatre paires de bases non Watson-Crick, dont des appariements en tandem G.A/A.G de type sheared. Un coude important au niveau de l'axe de l'hélice dû à la présence de ces appariements a pu être proposé par modélisation graphique par Eric Westhof à partir des résultats d'analyses structurales en solution (Walczak *et al.*, 1996). Les appariements G.A conservés sont essentiels à la fonction de SBP2 *in vivo* (Fagegaltier *et al.*, 2000b; Walczak *et al.*, 1996). Des expériences d'empreintes chimiques et enzymatiques ont montré que SBP2 reconnaissait précisément l'ARN SECIS au niveau de ces appariements (Fletcher *et al.*, 2001).





**Figure 8** : L'ARN SECIS d'après Allmang et Krol (2006). **A.** Modèles de structure secondaire des ARN SECIS eucaryotes de forme 1 et 2. Les séquences et caractéristiques structurales conservées sont indiquées. N, n'importe quel nucléotide ; A/g et A/c indique que A est le nucléotide majoritaire. **B.** Représentation de la structure en K-turn potentielle de l'ARN SECIS de l'iodotyrosine désiodase de rat de type 1 d'après la nomenclature de Leontis et Westhof (2001).

Notre objectif a été d'analyser plus précisément les détails de cette interaction en identifiant les acides aminés de SBP2 importants pour la liaison à l'ARN SECIS par dissection fonctionnelle, sélection d'ARN *in vitro* et résolution de la structure cristallographique du complexe SBP2/SECIS. De façon plus générale, nous avons également tenté de dégager les principes d'interaction des protéines de la famille L7Ae avec leurs ARN cibles.

## 1- La protéine SBP2 humaine et son mode d'interaction avec l'élément SECIS

Lorsque j'ai rejoint l'équipe d'Alain Krol en octobre 2001, l'ADNc de la protéine SBP2 humaine venait d'être cloné au laboratoire par Alain Lescure (Lescure *et al.*, 2002). J'ai été associée à ce travail en montrant que la liaison de SBP2 à l'ARN SECIS était stimulée par le facteur d'élongation spécialisé eEFsec sans qu'il ne s'associe pour autant au complexe. Ces résultats suggéraient que eEFsec était capable d'induire une conformation de SBP2 plus propice à la reconnaissance de l'ARN SECIS. Ceci nous a conduit à entamer une dissection fonctionnelle de SBP2 afin d'affiner la compréhension de ses différents domaines et plus particulièrement son domaine de liaison à l'ARN.

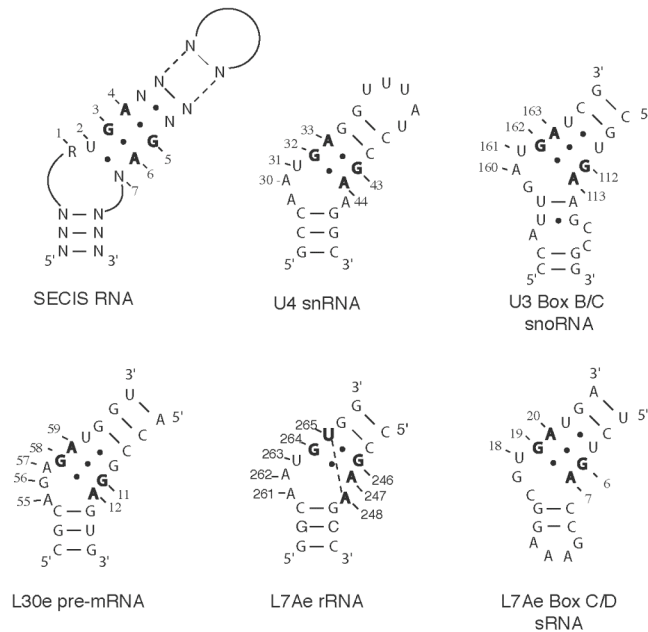
Le domaine de liaison à l'ARN de SBP2 se situe entre les acides aminés 500 et 750 (Allmang *et al.*, 2002 ; Copeland *et al.*, 2000 ; Lescure *et al.*, 2002). Nous avons montré qu'il s'agissait d'un domaine bipartite constitué de séquences spécifiques à SBP2 et d'un module structural conservé. Par alignement de séquences, nous avons découvert au sein du domaine de liaison à l'ARN un module appartenant à la famille des protéines ribosomiques L7Ae (Allmang

*et al.*, 2002). Cette famille comprend, en plus de nombreuses protéines ribosomiques, la protéine Nhp2p des snoRNP à boîte H/ACA, la protéine 15.5 kD (ou Snu13p chez la levure) des snoRNP à boîte C/D et son orthologue archaebactérien L7Ae (voir Figure 9A). Ces protéines reconnaissent toutes des ARN cibles capables de se structurer en « K-turn ». Ce motif fut identifié initialement lors de la résolution de la structure du snARN U4 lié à la protéine 15.5 kD (Vidovic *et al.*, 2000b) et celle des sous-unités ribosomiques de *H. marismortui* et *T. thermophilus* (Klein *et al.*, 2001). Il s'agit d'une hélice-bulle interne qui se caractérise par la présence de deux paires de bases consécutives G.A/G.A et d'un résidu protubérant (voir Figure 9B). La structure locale du squelette sucre-phosphate se caractérise par la présence d'un coude important résultant en une différence d'orientation de 120° entre les axes des hélices adjacentes. L'élément SECIS présente un repliement secondaire très similaire et nous proposons qu'il s'agit vraisemblablement d'un variant de K-turn (voir la revue Allmang & Krol, 2006a).

## A

hSBP2	K R - R L V - - - L G L R E V L K H L K L K K L K C V I T S P N C E K I T Q S K G	705
h15.5kD	- - - O L R - - - K G A N E A T K T L N R G I S E F I V M A A D A E P L - - - -	63
Snu13p	- - - N L R Q L K K G A N E A T K T L N R G I S E F I T M A A D C E P I - - - -	61
Nhp2p	- - - N V K - - - R G V K E V V K A L R K G E K G L V V I A G D I S P A - - - -	101
yRPL30	K S G K Y T - - - L G Y K S T V K S L R Q G K S K L I T I A A N T - P V L R K S	54
hRPL7A	- - - - L R - - - A G V N T V T T L V E N K K A Q L V V I A H D V D P I E L V V	168
hSBP2	G L D D T L H T - I I D Y A C E Q - N I P F V F A L N R - K A - L G R S L N K	740
h15.5kD	- - E I T L H L P L L - - - C E D K N V P Y V F V R S K - Q A - L G R A C G -	94
Snu13p	- - E I T L H L P L L - - - C E D K N V P Y V F V P S R - V A - L G R A C G -	93
Nhp2p	- - D V I S H I P V L - - - C E D H S V P Y T I F I P S K - Q D - L G A A G A T	133
yRPL30	- - E - L E - - Y Y A M L S K T K - V Y Y F Q G G N E - - - L G T A V G K	83
hRPL7A	F L P A L C R K M G - - - - - - - - V P Y C I I - - - K G K A R L G R L V H R	196
hSBP2	A V P - - V S V V G T F S Y	752
h15.5kD	- - - - V S R P V I A C S V T I K E G S Q L K Q Q I Q S I	119
Snu13p	- - - - V S R P V I A A S I T T N D A S A I K T Q I Y A V K D K	121
Nhp2p	K R P - - T S V V F I V P G S N K K K D G K	153
yRPL30	L F R - - V G V V S I L E A G D S D I L T T L A	105
hRPL7A	K T C T T V A F T Q V N S E	210

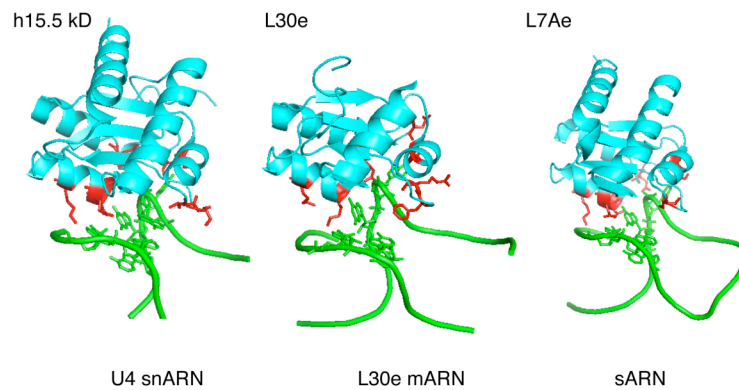
## B



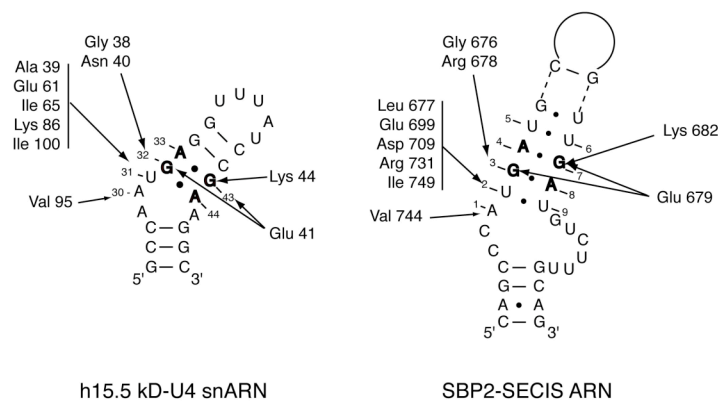
**Figure 9 :** A- Alignements de séquences entre le domaine de liaison à l'ARN de SBP2 et les protéines de la famille L7Ae. B- Structure secondaire des ARN cibles des protéines de la famille L7Ae.

Les structures cristallographiques des complexes 15.5kD-U4, L30e-pre-ARNm et L7Ae-sRNA à boîte C/D, représentées Figure 8A, ont révélé une interface ARN-protéines commune (Chao & Williamson, 2004 ; Moore *et al.*, 2004; Vidovic *et al.*, 2000a). Elle se caractérise par l'interaction d'un nucléotide en bulge U (snARN U4, ARNr, sARN) ou A (pré-ARNm de L30e) avec une poche d'acides aminés hydrophobes au sein de la protéine, et par les contacts spécifiques de quelques acides aminés avec les deux paires de bases G.A. Des prédictions structurales basées sur les homologues de SBP2 avec les protéines de la famille L7Ae nous ont permis de prédire et d'identifier les acides aminés conservés de SBP2 impliqués dans l'interaction avec l'ARN SECIS (Allmang *et al.*, 2002). Ces résultats ont mis en évidence l'existence de modes d'interactions vraisemblablement très similaires pour les complexes 15.5kD/U4 et SBP2/SECIS (voir Figure 10B). Ceci suggère l'existence d'une origine commune pour les domaines de liaison à l'ARN des protéines SBP2/15.5kD ainsi que pour les structures des ARN SECIS et U4.

**A**



**B**



**Figure 10 : Mode d'interaction entre les protéines de la famille L7Ae et leur ARN cible en K-turn. A-** Structures cristallographiques des complexes 15.5kD-U4, L30e-pre-ARNm et L7Ae-sRNA à boîte C/D (Chao & Williamson, 2004 ; Moore *et al.*, 2004; Vidovic *et al.*, 2000a). Les figures ont été générées par PyMol à partir des coordonnées respectives de PDB 1E7K, 1T0K et 1RLG. **B-** Schémas d'interaction similaires pour les complexes

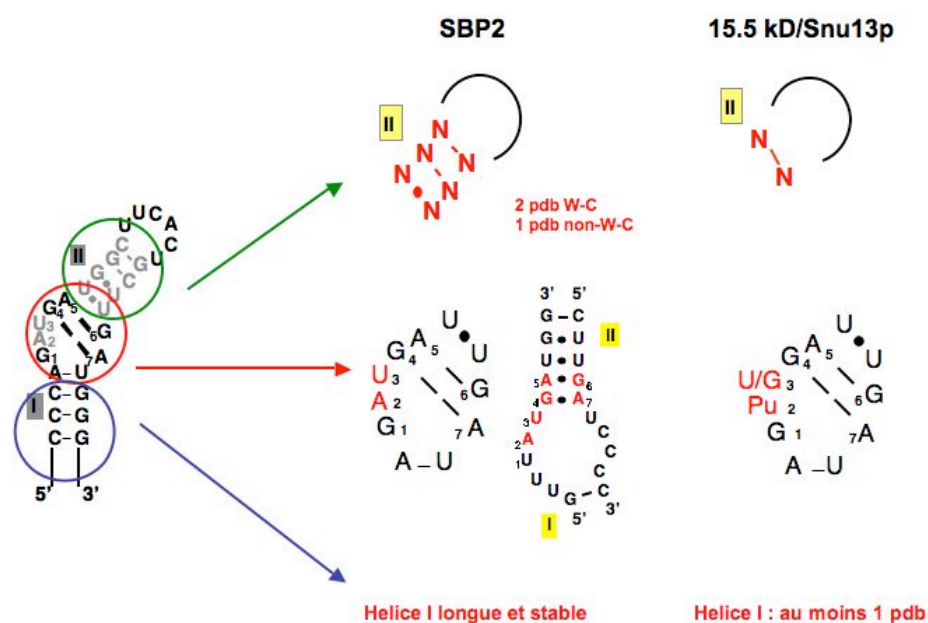
15.5kD-snARN U4 et SBP2-SECIS. 4 acides aminés de SBP2 sont essentiels à l'interaction : Gly676 et Glu679 postulés en contact avec les guanines des paires G.A ; Glu699 et Arg731 postulés en contact avec le U en bulge.

Par ailleurs, avec David Schmitt (étudiant en DEA de Biologie Moléculaire en 2002) nous avons délimité plus précisément le domaine de liaison à l'ARN et montré qu'il s'étendait bien au-delà du module conservé L7Ae et comprenait une séquence conservée riche en lysines. Ces résultats suggèrent que des contacts additionnels existent entre SBP2 et l'ARN SECIS par rapport aux autres protéines de la famille L7Ae.

## **2- Principes de reconnaissance entre protéines de la famille L7Ae et les ARN en K-turn**

*En collaboration avec Antoine Cléry et Christiane Branlant (UMR 7567 CNRS-UHP, Nancy)*

Au vu du degré de similitude élevé entre les modes d'interaction des protéines de la famille L7Ae avec leurs ARN cibles, se posait la question de l'existence de déterminants de spécificité pour la discrimination des cibles. En collaboration avec Antoine Cléry de l'équipe de Christiane Branlant (UMR 7567, Nancy) nous avons analysé les principes de reconnaissance entre les protéines de la famille L7Ae et les ARN en K-turn (Cléry *et al.*, 2007). De façon surprenante, nous avons montré que 15.5kD/Snu13p et L7Ae étaient capables de reconnaître l'ARN SECIS *in vitro*. En revanche, SBP2 est incapable de reconnaître les motifs en K-turn des ARN U4 et U3B/C. La reconnaissance des cibles par SBP2 répond donc à des critères de spécificité plus stricte. Pour identifier les déterminants requis au niveau de l'ARN pour la reconnaissance par SBP2, nous avons utilisé la méthode de SELEX combinée à la mutagenèse dirigée. A notre grande surprise, tous les ARN sélectionnés par SBP2 ont la capacité de se replier en K-turn canoniques avec un nucléotide U en bulge et répondent à des contraintes structurales fortes (Cléry *et al.*, 2007). Nous avons comparé les propriétés de liaison à l'ARN de SBP2 et de la protéine Snu13p de *S. cerevisiae* qui se fixe à la fois sur le K-turn du snARN U4 et du snoARN U3 (voir Figure 11). Il apparaît que, contrairement à Snu13p, SBP2 reconnaît préférentiellement des K-turn à grande boucle interne. L'identité des nucléotides 2 et 3 de la boucle est importante pour la reconnaissance par SBP2. Par ailleurs, de nouveaux déterminants de spécificité, uniques à SBP2, ont été mis à jour au sein de l'hélice II (voir Figure 11). Snu13p a montré une capacité à s'adapter à une plus grande variété d'ARN cibles. L'ensemble de ces résultats est en accord avec nos données de dissection fonctionnelle qui montrent que des contacts ARN-protéine additionnels sont mis en jeu dans le complexe SBP2-SECIS (voir paragraphe B-1).



**Figure 11 : Déterminants de spécificité reconnus par SBP2 et Snu13p/15.5kD au niveau des ARN en K-turn.**

A gauche : Structure secondaire de l'ARN obtenu par SELEX et reconnu par SBP2 avec la meilleure affinité. Des variants de cet ARN (mutations au sein des hélices I, II et de la boucle interne) ont été testées pour leur capacité à être reconnus par SBP2 et Snu13p. Les déterminants de spécificité identifiés pour chacune des protéines sont représentés en rouge.

L'assemblage de la machinerie de synthèse des sélénoprotéines est vraisemblablement initiée par la fixation de SBP2 aux éléments SECIS dans le noyau (voire le nucléole) où la protéine 15.5kD/Snu13p est très abondante. Les différences importantes au niveau des déterminants de spécificité des ARN cibles de SBP2 et 15.5kD/Snu13p contribuent vraisemblablement à la spécificité d'association des complexes SBP2-SECIS dans ce compartiment cellulaire.

### 3- Objectif : Résolution de la structure cristallographique des complexes SBP2-SECIS, L30-SECIS

*Travail de thèse de Vincent Olieric dirigé par Philippe Dumas (UPR 9002 du CNRS) et d'Akiko Takeuchi (Doctorante au laboratoire depuis 2006).*

Afin d'établir définitivement si l'ARN SECIS possède un repliement en K-turn nous avons entrepris de résoudre la structure aux rayons X du complexe SBP2/ARN SECIS en collaboration avec l'équipe de P. Dumas (UPR 9002 du CNRS). La résolution de cette structure permettrait également de valider nos prédictions structurales (Allmang *et al.*, 2002) ainsi que de comprendre le rôle du motif de liaison à l'ARN additionnel par rapport aux autres protéines de la famille L7Ae.

Ce travail a fait l'objet de la thèse de Vincent Olieric dans l'équipe de Philippe Dumas. Des protocoles d'expression et de purification de SBP2 ont été optimisés et une grande variété d'ARN SECIS a été synthétisée. Il n'a pas été possible d'obtenir de cristaux du complexe, ni de la protéine isolée. En revanche, la caractérisation biophysique de SBP2 par RMN et ultracentrifugation analytique a révélé une absence de structuration. Ceci est en accord avec des analyses bioinformatiques prédisant une prédominance de zones non repliées, en dehors du module L7Ae. SBP2 semble répondre à plusieurs critères caractéristiques des protéines intrinsèquement non structurées ou « IUP » (Intrinsically Unstructured Proteins) (Dosztanyi *et al.*, 2005). Chez les eucaryotes supérieurs, bon nombre de protéines impliquées dans des mécanismes de régulation ou de transduction des signaux ne se replient de façon stable qu'en présence de leurs partenaires moléculaires (pour des revues voir Dunker *et al.*, 2005; Tompa, 2005). SBP2 sert de plateforme pour le recrutement des autres partenaires de la machinerie de biosynthèse des sélénoprotéines. Des résultats récents obtenus au laboratoire montrent que SBP2 interagit avec un complexe de protéines chaperons lié à la protéine HSP90, et que cette association joue un rôle fonctionnel important dans le repliement de SBP2 et son interaction avec ses cibles (voir paragraphe C). La possibilité que SBP2 soit partiellement non structurée est donc compatible avec les interactions multiples qu'elle doit assurer et nos nouvelles données. Il n'est cependant pas possible d'exclure que l'absence de structuration résulte de l'expression de SBP2 dans *E.coli* qui ne permet pas d'assurer les modifications post-traductionnelles. Afin de vérifier cette hypothèse, la protéine SBP2 sera produite dans des cellules eucaryotes à partir de vecteurs de type baculovirus. Ce travail a été initié par Akiko Takeuchi étudiante en thèse dans notre laboratoire depuis septembre 2006 avec l'aide de la plateforme de biologie génomique et structurales (CEBGS-Illkirch) et du service baculovirus de l'IGBMC (Illkirch). Nous tenterons également de co-cristalliser SBP2 en présence de HSP90, car il est vraisemblable que l'interaction SBP2-HSP90 facilite la structuration de SBP2 afin de la rendre apte à interagir avec ses partenaires finaux.

Enfin, nous tenterons de surproduire et de cristalliser la protéine SBP2 issue d'un autre organisme, en particulier celle de *Drosophila melanogaster* qui présente la particularité d'être dépourvue du domaine N-terminal présent dans la protéine humaine (voir paragraphe B-4). Nous évaluerons sa capacité à se lier spécifiquement à l'élément SECIS et à se replier de façon stable.

La fixation de la protéine ribosomique L30 sur l'ARN SECIS a également été identifiée comme l'une des étapes du mécanisme de recodage des sélénoprotéines (Chavatte *et al.*, 2005). Nous tenterons de cristalliser le complexe L30/SECIS. La comparaison des structures des deux complexes SBP2/SECIS, L30/SECIS devrait nous donner de précieux renseignements quant au rôle central de l'ARN SECIS dans l'étape de translecture, c'est-à-dire s'il existe une compétition des protéines pour les mêmes sites au niveau du SECIS, ou des changements allostériques permettant interaction avec le ribosome.

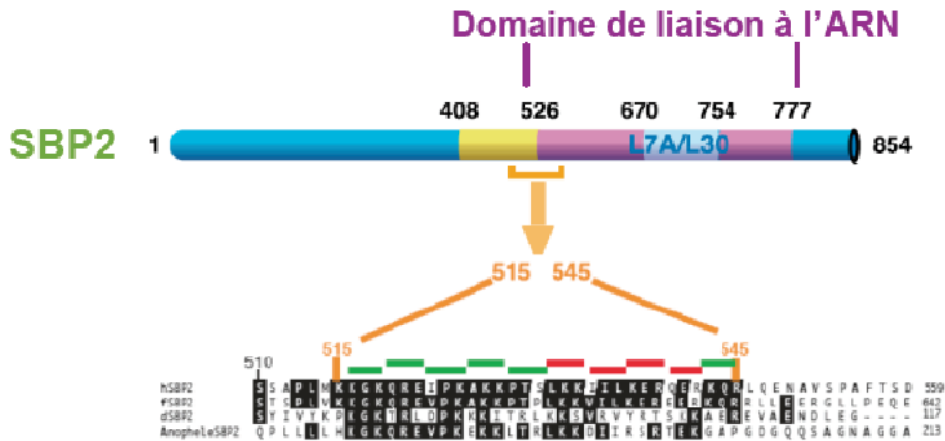
#### 4- La protéine SBP2 de drosophile

*Avec Akiko Takeuchi (Doctorante au laboratoire depuis 2006)*

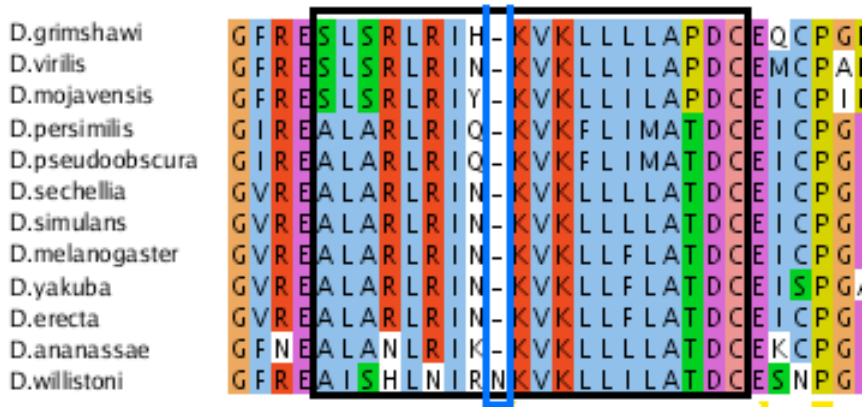
L'analyse de banques de données nous a permis d'identifier et de cloner l'ADNc de la protéine SBP2 de *Drosophila melanogaster*. Cette protéine de 314 acides aminés est dépourvue de la région N-terminale de fonction inconnue présente chez la protéine humaine. Elle possède un motif de liaison à l'ARN de type L7Ae, mais le domaine de liaison à l'ARN additionnel riche en lysine que nous avons identifié chez la protéine humaine n'est pas présent chez la drosophile. De façon surprenante, nos résultats préliminaires semblent indiquer que la protéine de drosophile ne soit capable de reconnaître qu'une seule forme d'ARN SECIS : les ARN SECIS à boucle apicale structurée (ou type 2, voir Figure 8). C'est la seule conformation d'ARN SECIS trouvée chez *Drosophila melanogaster* (Castellano *et al.*, 2001). Par ailleurs, des mutations au sein du motif riche en lysine abolissent la liaison de SBP2 humaine aux ARN SECIS de type 1 mais pas de type 2. La fonction du motif riche en lysines est vraisemblablement de permettre d'accommoder des ARN SECIS de type 1 qui sont apparus plus tard au cours de l'évolution. Nous testerons cette hypothèse en échangeant les motifs des deux protéines et en évaluant leur capacité à reconnaître les différents types d'ARN SECIS. Une collaboration a été engagée avec l'équipe de bioinformatique de Roderic Guigo (Barcelone) pour vérifier s'il existe une corrélation entre la présence du motif lysine riche et la nature des ARN SECIS présents chez divers organismes par analyse comparative de génomes entiers. Enfin, nous déterminerons si la protéine SBP2 de *Drosophila melanogaster* est capable de stimuler à elle seule l'incorporation de sélénocystéine au sein d'une sélénoprotéine rapporteur dans un système de traduction *in vitro*, comme c'est le cas pour SBP2 humaine, ou s'il faut envisager la participation d'un facteur additionnel.

La séquence de toutes les protéines de *D. melanogaster* impliquées dans la synthèse des sélénoprotéines a été recherchée par analyse comparative dans les génomes de drosophiles par C. Chapple dans l'équipe de R. Guigo. Cette étude a révélé que l'essentiel de ces facteurs était absent chez *D. Willistoni*. C'est le cas par exemple de eEFsec et de SPS2. La machinerie de synthèse des sélénoprotéines semble être absente chez cet organisme. La protéine SBP2 de *D. Willistoni* présente quant à elle la particularité d'avoir un acide aminé additionnel au sein de son domaine de liaison à l'ARN (Figure 12). Par ailleurs, la présence de reliques d'ARN SECIS (R. Guigo, communication personnelle) suggère que SBP2 a perdu sa capacité de liaison à l'ARN SECIS et a été maintenue pour assurer une autre fonction. Nous testerons cette hypothèse en évaluant l'impact de l'insertion d'un acide aminé dans la protéine humaine sur la reconnaissance du SECIS et la synthèse des sélénoprotéines.

A



B



**Figure 12 : A. La protéine SBP2 humaine.** Les différents domaines sont représentés: le domaine de liaison à l'ARN SECIS (violet, acides aminés 526-777), le site putatif d'interaction au ribosome (jaune) et le domaine N-terminal de fonction inconnue (bleu). En plus du module L7Ae (bleu), les acides aminés conservés 515-545 sont impliqués dans la liaison à l'ARN SECIS. Un alignement de cette région est représenté pour les protéines humaines (hSBP2), de rat (rSBP2) de drosophile (dSBP2) et d'anophèle. Les mutations par alanine scanning dans cette région de hSBP2, qui abolissent la liaison au SECIS de type 1 (531-542), sont surlignées en rouge. Cette région n'est pas conservée chez *D. melanogaster*. Les mutations d'acides aminés surlignés en vert sont sans effet.

**B. Alignements du domaine de liaison à l'ARN de SBP2 chez les drosophiles** d'après C. Chapple et R. Guigo (communication personnelle). Une caractéristique générale du site de liaison à l'ARN des protéines de la famille L7Ae est la conservation d'une distance constante de 19 acides aminés entre les résidus strictement conservés E/D (région encadrée). Chez *D.willistoni*, cette distance est de 20 acides amines.



## **C- Les complexes supramoléculaires impliqués dans la synthèse des sélénoprotéines**

*Avec Laurence Wurth (Doctorante depuis octobre 2005).*

La famille des protéines L7Ae, à laquelle appartient SBP2, participe à la construction de plusieurs RNP essentielles et leur liaison à l'ARN conditionne le recrutement des autres protéines. Nos objectifs sont, d'une part de comprendre les mécanismes d'assemblages généraux mis en jeu pour la construction de ces RNP et d'autre part, d'identifier les composants encore inconnus du complexe moléculaire qui est recruté spécifiquement par SBP2 sur l'ARN SECIS.

### **1- Un mécanisme commun pour l'assemblage des RNP L7Ae (manuscrit soumis)**

*En collaboration avec Edouard Bertrand (CNRS UMR 5535, Montpellier), Bruno Charpentier et Christiane Branlant (UMR 7567 CNRS-UHP, Nancy), Tamas Kiss (LBME, Toulouse) et Barbara Bardoni (Université de Nice-Sophia antipolis) dans le cadre d'un contrat de l'Agence Nationale pour la Recherche.*

Nous avons identifié une machinerie d'assemblage des RNP L7Ae conservée de la levure à l'homme et d'importance fondamentale pour la cellule. Elle est constituée d'une protéine adaptatrice Nufip (Rsa1 chez la levure) et d'un complexe de protéines chaperons.

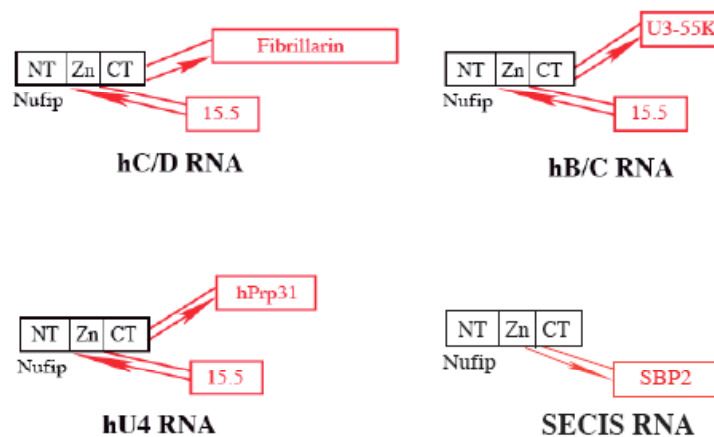
#### *Nufip, un facteur d'assemblage des particules de la famille L7A*

Nous avons utilisé des cribles double et triple hybrides pour caractériser de nouveaux facteurs impliqués dans la biogenèse des snoARN à boîte C/D. Chez *Saccharomyces cerevisiae*, ceux-ci ont permis de détecter la protéine Rsa1 (Kressler *et al.*, 1999) et de montrer qu'elle interagissait avec Snu13p *in vitro* (Bruno Charpentier, Nancy). Parallèlement, la protéine humaine Nufip a été trouvée en interaction avec 15.5kD (Edouard Bertrand, Montpellier). Cette interaction a été confirmée *in vivo*. Nufip est une protéine nucléaire qui se lie à l'ARN. Elle fut initialement identifiée en interaction avec la protéine FRMP qui est impliquée dans le transport et la localisation d'ARNm (Bardoni *et al.*, 1999). Des comparaisons de séquences ont révélé que Nufip et Rsa1 présentent un motif conservé de 32 acides aminés (ou motif PEP). Nous avons démontré que Nufip et Rsa1 étaient capables d'interagir respectivement avec 15.5kD et Snu13p, par l'intermédiaire du motif PEP et que Nufip était l'homologue fonctionnel de Rsa1.

De façon intéressante, Nufip reconnaît deux autres membres de la famille L7Ae, hNhp2p qui fait partie des snoRNP à boîte H/ACA et SBP2. Nous avons notamment pu confirmer l'interaction entre Nufip et SBP2 *in vivo* en co-purifiant les protéines endogènes à

partir d'extraits nucléaires de cellules HeLa. Cependant, le domaine PEP seul de Nufip n'interagit que faiblement avec SBP2 et pas du tout avec hNhp2p. Les déterminants de l'interaction semblent par conséquent différents de ceux qui sont utilisés pour la reconnaissance de 15.5kD.

Afin de déterminer si Nufip était associée *in vivo* aux RNP L7Ae, nous avons réalisé une série d'immunoprécipitations après co-transfection de Nufip et des divers ARN cibles des protéines L7Ae dans des cellules eucaryotes. Nos résultats ont démontré que Nufip était capable de s'associer aux snoARN à boîte C/D et H/ACA, au snARN U4 et même aux ARNm de sélénoprotéines. Nous avons par ailleurs établi que Nufip était capable d'interagir avec d'autres protéines core des RNP telles que hPRP31 (protéine de la snRNP U4), U3-55K (composant de la particule U3) et la fibrillarine (composant des snoARN à boîte C/D) ; il en va de même pour Rsa1. Nufip est également capable de stimuler l'interaction entre 15.5kD et ces protéines. En effet, dans des tests d'interaction double-hybride où aucune de ces protéines isolée n'interagit directement avec 15.5kD, l'ajout d'un vecteur codant pour Nufip permet la formation d'un complexe ternaire. Nufip semble donc jouer le rôle d'adaptateur pour recruter les autres protéines core vers les complexes 15.5kD/ARN des RNP U4, à boîte C/D et B/C (voir Figure 13).



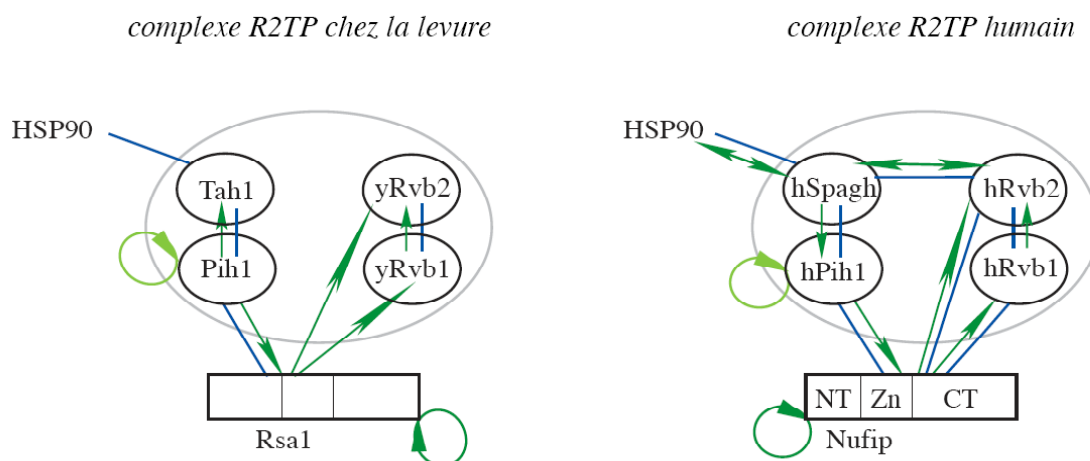
**Figure 13 : Résumé des interactions entre Nufip et les protéines core spécifiques des snoARN à boîte C/D, de U3 (B/C RNA), U4 et de l'ARN SECIS.** Les interactions double-hybride sont représentées par des flèches. Les interactions directes obtenues par GST pull-down sont représentées par des traits.

*HSP90 et le complexe de co-chaperons R2TP participent à l'assemblage des RNP U3, U4 et de la mRNP SECIS*

Bien que l'assemblage des RNP nécessite le repliement correct de l'ARN et des protéines, aucun chaperon protéique n'a été impliqué jusqu'à présent dans ce mécanisme. Chez

la levure, certains facteurs requis pour la biogenèse des snoARN ont cependant été identifiés par ailleurs comme co-chaperons d'HSP90. Cette protéine conservée à travers l'évolution est un chaperon important de la cellule, impliquée notamment dans le contrôle des récepteurs nucléaires et des protéines kinase (pour revues voir Caplan *et al.*, 2007 ; Pearl & Prodromou, 2006).

HSP90 est associée à un complexe de co-chaperons appelé R2TP (Zhao *et al.*, 2005). Celui-ci se compose de deux ATPase AAA+ (Rvb1 et Rvb2) et des protéines Pih1 et Tah1. De façon intrigante, l'accumulation des snoRNP à boîtes C/D chez la levure requiert l'ATPase essentielle AAA+ Rvb2 (King *et al.*, 2001) et la protéine Pih1 (ou Nop17 ; Gonzales *et al.*, 2005). Ceci suggère qu'HSP90 et le complexe R2TP sont impliqués dans la biogenèse des snoRNP à boîte C/D. Nous avons testé cette hypothèse. Une analyse protéomique récente de cellules humaines a révélé que des homologues de Rvb1, Rvb2 et de Pih1 étaient également associés à HSP90 (Te *et al.*, 2007). Par analyse systématique de banques de données, nous avons identifié l'homologue de Tah1 que nous avons appelé hSpagh (par référence à son homologue Spaghetti chez la drosophile). Nous avons démontré, par immunoprécipitation, que hSpagh était associée à hRvb1, hRvb2 et hPih1, suggérant que le complexe R2TP est également présent et conservé chez l'homme. Pour compléter cette analyse, nous avons vérifié le réseau d'interaction protéique au sein du complexe R2TP chez la levure et l'homme par des tests d'interaction double hybride systématiques. Nos équipes ont testé collectivement 471 interactions, les résultats sont représentés Figure 14. Enfin, nous avons démontré qu'HSP90 et les protéines du R2TP étaient capables de s'associer aux ARN U3 et U4 ainsi qu'à SBP2.

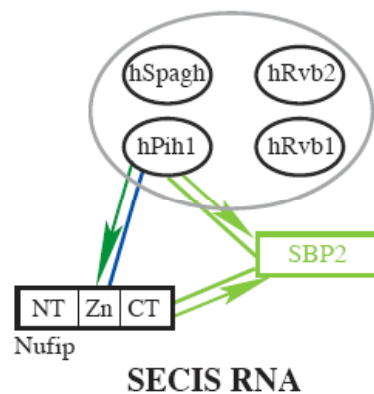


**Figure 14 : Résumé des interactions entre Rsa1, Nufip et les protéines du R2TP chez la levure et l'homme.** Les interactions double-hybride sont représentées par des flèches vertes, les interactions directes par GST pull-down par des traits bleus.

L'ensemble de nos résultats suggère qu'il existe un lien entre la machinerie de repliement des protéines et l'assemblage des RNP de la famille L7Ae. Les homologues humains des protéines du R2TP semblent jouer le rôle de co-chaperons d'HSP90 pendant l'assemblage de ces RNP.

*Nufip sert d'adaptateur entre les protéines L7Ae et le complexe R2TP*

Nous avons pu montrer qu'en présence de Nufip, les protéines 15.5kD et hNhp2p étaient capables d'interagir avec le composant hPih1 du complexe R2TP. Nufip joue donc le rôle d'adaptateur pour amener le complexe de chaperons vers les complexes L7Ae/ARN. De façon surprenante, SBP2 est capable d'interagir directement avec ce composant du R2TP (Figure 15).



**Figure 15 : Interactions de Nufip et du complexe R2TP avec SBP2 sur l'ARN SECIS.** Les interactions sont représentées comme sur la Figure 14.

Nous avons étudié le rôle fonctionnel de ces interactions, pour déterminer si elles permettaient de connecter les protéines L7Ae à HSP90 afin de participer à leur repliement. Pour cela, nous avons traité des cellules eucaryotes en culture par la geldanamycine. Cette molécule est capable de se fixer au site ATPase d'HSP90 et d'inhiber le repliement de ses protéines cibles qui deviennent instables (Stebbins *et al.*, 1997). En effet, nous avons pu montrer que 15.5kD, hNhp2p et SBP2 devenaient instables après inhibition d'HSP90 alors que Nufip et les protéines du R2TP n'étaient pas affectées. Ceci a également pour conséquence de déstabiliser les ARN U3, U4 et de la télomérase (autre cible de hNhp2p).

Nous avons donc confirmé le rôle essentiel d'HSP90 dans la biogenèse des RNP L7Ae. Il semble que l'assemblage des RNP soit plus fortement dépendant du repliement des protéines que ce qui avait été initialement imaginé. Il est vraisemblable que les protéines core des RNP sont instables à l'état isolé et ne sont stabilisées qu'en présence de leurs partenaires dans le

complexe final. Dans le cas de SBP2, nos résultats d'analyse structurale prédisent en effet qu'elle présente toutes les caractéristiques d'une protéine intrinsèquement non repliée (voir paragraphe B- 3).

HSP90 est directement impliqué dans les mécanismes du cancer chez l'homme et apparaît comme un acteur clé de la régulation de la prolifération cellulaire car elle contrôle plusieurs cascades de signalisation cellulaire. Nos travaux enrichissent cette vision, puisqu'ils établissent qu'HSP90 peut également contrôler la croissance cellulaire en influant sur la biosynthèse des ribosomes, la réplication via la production de l'ARN de la télomérase et la lutte contre les radicaux libres via la synthèse des sélénoprotéines.

## **2- Projets à court terme: l'assemblage de la mRNP SECIS**

Notre objectif est maintenant de mieux comprendre comment est assuré le contrôle de l'assemblage des RNP L7Ae par HSP90, le complexe R2TP et l'adaptateur Nufip en disséquant ces mécanismes *in vivo et in vitro*. Les équipes d'Edouard Bertrand, de Bruno Charpentier et Christiane Branlant étudieront ces mécanismes dans le cas des sn et snoRNP. Notre équipe s'attachera à élucider ces aspects dans le cas de la mRNP SECIS.

### *Localisation subcellulaire de Nufip et SBP2*

Bien que majoritairement cytoplasmique, SBP2 est capable de transiter entre le noyau et le cytoplasme et sa localisation conditionne celle du facteur d'élongation eEFsec (Small-Howard *et al.*, 2006). L'existence d'une interaction entre SBP2 et Nufip renforce l'hypothèse d'un mécanisme d'assemblage nucléaire des facteurs sur l'ARN SECIS. Il est de ce fait crucial d'examiner les distributions cellulaires respectives de Nufip et SBP2 par immunolocalisation et celle d'un ARNm de sélénoprotéine tel que celui de la glutathion peroxydase (GPx) par hybridation *in situ*. Ces expériences seront également réalisées dans différentes conditions notamment en présence de geldanamycine (inhibiteur d'HSP90) qui devrait affecter la formation du complexe, ou de stress oxydant qui conduit à l'accumulation de SBP2 dans le noyau (Papp *et al.*, 2006).

Un assemblage nucléaire précoce pourrait également servir à diriger les ARNm de sélénoprotéines vers une voie d'export spécialisée. L'export nucléaire de SBP2 semble être dépendant du facteur CRM1 (Papp *et al.*, 2006). De façon intéressante, CRM1 et une ATPase du complexe R2TP (Rvb2) ont été trouvés associés au snoARN U3 mature (Watkins *et al.*, 2004). CRM1 est impliqué dans le transport nucléolaire d'U3, mais il a été suggéré qu'il

pourrait également jouer un rôle dans l'assemblage de la RNP U3 (Boulon *et al.*, 2004). Au vu de similitudes entre les mécanismes d'assemblage des RNP U3 et des mRNP SECIS, nous vérifierons le rôle exact de CRM1 dans le mécanisme de synthèse des sélénoprotéines. Nous testerons si CRM1 est associé aux mRNP de sélénoprotéines et à SBP2 par des expériences de co-transfections et immunoprécipitations en présence ou en absence de leptomycine B (inhibiteur de CRM1).

#### *Influence de Nufip et HSP90 sur stabilité des ARNm de sélénoprotéines et la synthèse des sélénoprotéines*

Nous avons montré que Nufip et HSP90 étaient associés aux mRNP SECIS. En favorisant l'assemblage de protéines sur l'ARN SECIS, il est vraisemblable que ces facteurs permettent d'éviter la dégradation des ARNm de sélénoprotéines par les mécanismes du NMD. Nous testerons l'effet de l'inactivation du gène de Nufip par RNA interférence, et celui de l'inhibition HSP90 par la geldanamycine, sur la stabilité des ARNm de sélénoprotéines endogènes *in vivo* par RT-PCR quantitative. Par transfections transitoires d'ARNm de sélénoprotéines rapporteurs, nous évaluerons l'impact de l'inhibition de Nufip et HSP90 sur la synthèse de la sélénoprotéine correspondante. Un défaut d'association de SBP2 devrait inhiber la synthèse de sélénoprotéines et conduire à l'arrêt prématuré de la traduction au niveau du codon UGA<sub>Sec</sub> et la production d'une protéine tronquée. Ceci devrait nous aider à mieux comprendre le rôle fonctionnel de Nufip et HSP90 dans le mécanisme de synthèse des sélénoprotéines.

#### *Etude in vitro de l'interaction Nufip/SBP2*

L'interaction de Nufip avec la protéine 15.5kD se fait par l'intermédiaire du motif PEP mais nous avons montré que celui-ci ne semblait pas suffisant pour interagir efficacement avec SBP2. Nous affinerons l'analyse du domaine d'interaction de Nufip par mutagenèse. Nous tenterons également d'identifier le domaine de SBP2 impliqué dans l'interaction avec Nufip. Il est vraisemblable que cette interaction ait lieu par l'intermédiaire du domaine conservé L7Ae. Nous le vérifierons à l'aide de protéine SBP2 tronquées, puis déterminerons la nature des acides aminés impliqués par mutagenèse, GST-pull down et co-immunoprécipitations. Des études similaires seront réalisées par nos collaborateurs en ce qui concerne Rsa1 et Snu13p (Bruno Charpentier, Nancy), Nufip et 15.5kD (Montpellier). L'ensemble de ces travaux devrait nous permettre d'établir si les modes d'interaction entre Nufip (Rsa1) et de ses différentes

cibles font appel aux mêmes surfaces d'interactions protéine-protéine. A plus long terme, nous pourrions également envisager la cristallisation de ces complexes.

### *Quelles protéines core pour l'ARN SECIS ?*

Si l'assemblage des facteurs sur l'ARN SECIS est au cœur du mécanisme de synthèse des sélénoprotéines, la composition de la mRNP SECIS reste mal connue. Certaines pistes directes s'offrent cependant à nous. En effet, des expériences d'immunopurification de SBP2 nous ont permis d'isoler la protéine NSEP1 et de démontrer qu'elle interagissait *in vivo* avec SBP2. NSEP1 est une protéine capable de se lier à l'ARN SECIS *in vitro* (Fagegaltier *et al.*, 2000a). Son rôle dans la synthèse des sélénoprotéines, longtemps controversé, vient récemment d'être démontré (Qichang Shen, 2006). Nos résultats préliminaires montrent que NSEP1 interagit également avec Nufip *in vivo*. Par des tests double-hybride et GST-pull down *in vitro*, nous vérifierons si ces interactions sont directes ou si Nufip sert d'adaptateur entre SBP2 et NSEP1.

De façon surprenante, nous avons détecté une interaction entre SBP2, Nop58 et Nop56 par co-transfection et co-immunoprécipitation. Ceci pose la question de savoir si certaines des protéines core des snoRNP à boîte C/D sont également recrutées vers l'ARN SECIS. Nous examinerons cette possibilité.

### *Assemblage in vitro des mRNP de sélénoprotéines*

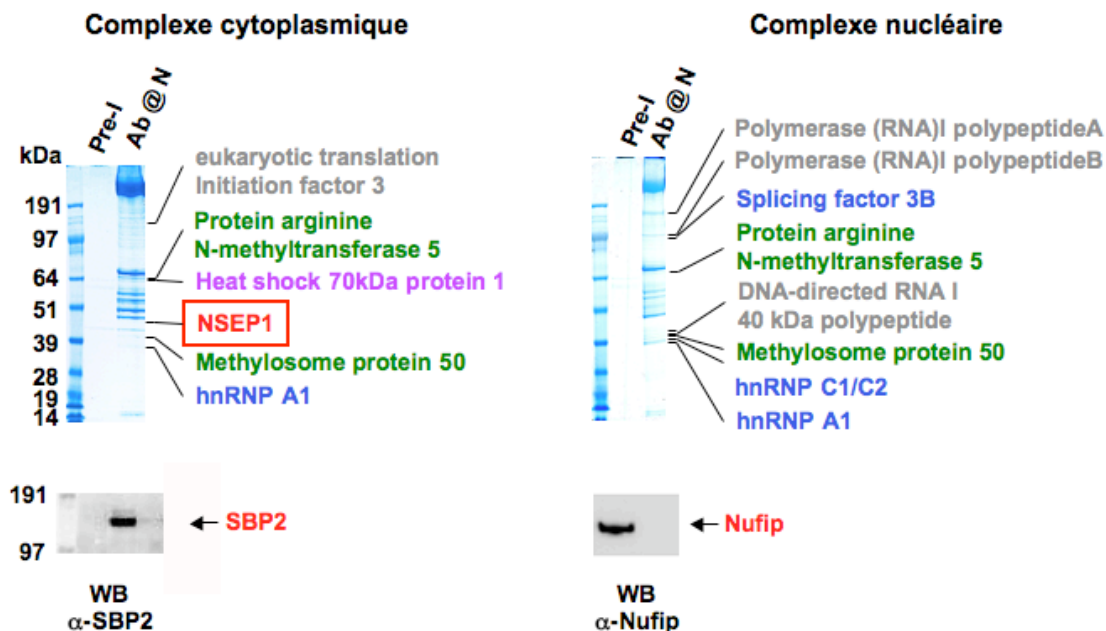
Notre étude *in vitro* sera complétée par des essais de reconstitution *in vitro* de la RNP SECIS. Dans un premier temps, nous déterminerons si Nufip est capable de se lier directement à l'ARN SECIS par des expériences de retard sur gel, ou s'il stimule l'assemblage des autres facteurs sur l'ARN SECIS comme dans le cas des autres RNP (voir paragraphe C-1). Nous testerons la capacité de Nufip à favoriser la reconstitution d'un complexe entre les protéines recombinantes SBP2, NSEP1 et l'ARN SECIS. Le facteur d'élongation eEF<sup>sec</sup> sera également inclus dans les expériences de reconstitution ; nous avons montré qu'il est capable de stimuler la liaison de SBP2 au SECIS sans pour autant s'associer au complexe (Lescure *et al.*, 2002). Il sera intéressant de vérifier si Nufip stabilise ou non ce complexe ou s'il s'agit d'une interaction transitoire qui n'est stimulée que par la présence de l'ARN<sup>Sec</sup> (Zavacki *et al.*, 2003).

Le rôle des protéines du complexe R2TP sur l'assemblage, et notamment des deux ATPases Rvb1 et Rvb2, sera également testé en présence ou en absence d'ATP. Il est vraisemblable que ces protéines participent au remodelage de la RNP, voire au transfert des protéines core recrutées par Nufip vers l'ARN, et à la dissociation de Nufip du complexe. Des

expériences de reconstitution *in vitro* similaires seront réalisées pour la snRNP U4. Pour tester le rôle des ATPases, une collaboration sera engagée entre notre réseau de laboratoires et l'équipe de M. Grigoriev (LBME, Toulouse) qui s'intéresse au rôle de ces protéines dans la translocation des jonctions Holliday le long de l'ADN et possède l'expertise nécessaire à cette analyse.

### 3- Projets à plus long terme : Purification des complexes associés à SBP2

La protéine SBP2 est présente au sein de complexes supramoléculaires dans le noyau et le cytoplasme (Small-Howard *et al.*, 2006). A l'aide d'anticorps anti-peptide de SBP2, Laurence Wurth, étudiante en thèse dans notre équipe, a mis au point les conditions d'immunopurification des complexes endogènes nucléaires et cytoplasmiques associés à SBP2 dans des cellules HeLa (voir Figure 16). Des résultats préliminaires ont conduit à l'identification d'interactants potentiels par spectrométrie de masse (Philippe Wolff et Plateforme protéomique de l'Esplanade, Strasbourg). Une stratégie complémentaire sera développée, basée sur la surexpression de SBP2 fusionnée à une double étiquette structurale TAP dans cellules HeLa suivie de la purification du complexe en tandem par des méthodes biochimiques.



**Figure 16 : Immunopurification des complexes cytoplasmiques et nucléaires associés à SBP2.** Les protéines purifiées, isolées et identifiées par spectrométrie de masse sont indiquées. En gris : facteurs de traduction (dans le cytoplasme) et de transcription (dans le noyau). En vert : protéines du méthylosome, complexe d'assemblage des snRNP, en bleu : facteurs d'épissage et d'assemblage des hnRNP. Des interactants connus de SBP2 ont été



déTECTÉS, notamment la protéine NSEPI, le chaperon HSP70 (voir paragraphes C-1 et C-2). SBP2 et Nufip ont été détectées par Western blot (panneaux inférieurs).

Notre objectif sera d'évaluer la validité des interactions mises au jour et d'établir si ces protéines font partie de façon stable ou transitoire des complexes supramoléculaires de la machinerie de synthèse des sélénoprotéines. Si tel est le cas, ces facteurs seront caractérisés fonctionnellement, et leur mode de liaison avec l'ARN SECIS et leurs protéines cibles seront analysés. Notre objectif sera de comprendre comment ceux-ci s'insèrent dans le contexte global du mécanisme traductionnel des sélénoprotéines.

#### *Plusieurs complexes d'assemblage pour la RNP SECIS?*

De façon surprenante, deux protéines du complexe du méthylosome ont été co-purifiées avec SBP2 (voir Figure 16). Le méthylosome est impliqué avec le complexe SMN dans les mécanismes d'assemblage des snRNP (Yong *et al.*, 2004). En effet, il permet la méthylation de protéines Sm qui sont prises en charge par le complexe SMN et fonctionnent comme chaperon d'assemblage des snARN. Une hypothèse possible stipule qu'il existerait un lien entre le complexe d'assemblage que nous avons caractérisé (Nufip, HSP90 et le complexe R2TP) et l'autre grande machinerie d'assemblage de la cellule, le complexe SMN. Nous examinerons dans un premier temps si les protéines du méthylosome et du complexe SMN forment un complexe avec SBP2 ou Nufip mais également si les protéines Sm sont capables de se lier à l'ARN SECIS et participent à son assemblage. En fonction des résultats de ces expériences préliminaires, nous vérifierons si le complexe R2TP et le complexe SMN collaborent pour l'assemblage de la RNP SECIS.

L'ensemble de ces expériences devrait nous apporter une vision globale du mécanisme de synthèse des sélénoprotéines et nous aider à comprendre la composition mais également la dynamique d'assemblage et de désassemblage des complexes multi protéiques impliqués dans le processus.

## Références bibliographiques

- Achsel, T., Brahm, H., Kastner, B., Bachi, A., Wilm, M. and Luhrmann, R. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J.*, **18**, 5789-5802.
- Allmang, C., Carbon, P. and Krol, A. (2002) The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *RNA*, **8**, 1308-1318.
- Allmang, C., Henry, Y., Morrissey, J.P., Wood, H., Petfalski, E. and Tollervey, D. (1996a) Processing of the yeast pre-rRNA at sites A(2) and A(3) is linked. *RNA*, **2**, 63-73.
- Allmang, C., Henry, Y., Wood, H., Morrissey, J.P., Petfalski, E. and Tollervey, D. (1996b) Recognition of cleavage site A(2) in the yeast pre-rRNA. *RNA*, **2**, 51-62.
- Allmang, C. and Krol, A. (2006a) SECIS RNAs and K-turn binding proteins. A survey of evolutionary conserved RNA and protein motifs. *Selenium, its molecular Biology and role in human Health 2nd edition*. DL Hatfield (ed) Kluwer Academic Publishers. , **5**, 51-61.
- Allmang, C. and Krol, A. (2006b) Selenoprotein synthesis: UGA does not end the story. *Biochimie*.
- Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. and Tollervey, D. (1999a) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J*, **18**, 5399-5410.
- Allmang, C., Mitchell, P., Petfalski, E. and Tollervey, D. (2000) Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res*, **28**, 1684-1691.
- Allmang, C., Mougél, M., Westhof, E., Ehresmann, B. and Ehresmann, C. (1994) Role of conserved nucleotides in building the 16S rRNA binding site of E. coli ribosomal protein S8. *Nucleic Acids Res*, **22**, 3708-3714.
- Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D. and Mitchell, P. (1999b) The yeast exosome and human PM-Scl are related complexes of 3' → 5' exonucleases. *Genes Dev*, **13**, 2148-2158.
- Allmang, C. and Tollervey, D. (1998) The role of the 3' external transcribed spacer in yeast pre-rRNA processing. *J Mol Biol*, **278**, 67-78.
- Anderson, J. and Parker, R. (1998) The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J.*, **17**, 1497-1506.
- Araki, Y., Takahashi, S., Kobayashi, T., Kajiho, H., Hoshino, S.-i. and Katada, T. (2001) Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast. *EMBO J.*, **20**, 4684-4693.
- Bardoni, B., Schenck, A. and Mandel, J.L. (1999) A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. *Hum Mol Genet*, **8**, 2557-2566.
- Beltrame, M., Henry, Y. and Tollervey, D. (1994) Mutational analysis of an essential binding site for the U3 snoRNA in the 5' external transcribed spacer of yeast pre-rRNA. *Nucl. Acids Res.*, **22**, 5139-5147.
- Berry, M.J., Tujebajeva, R.M., Copeland, P.R., Xu, X.M., Carlson, B.A., Martin, G.W., 3rd, Low, S.C., Mansell, J.B., Grundner-Culemann, E., Harney, J.W., Driscoll, D.M. and Hatfield, D.L. (2001) Selenocysteine incorporation directed from the 3'UTR: characterization of eukaryotic EFsec and mechanistic implications. *Biofactors*, **14**, 17-24.
- Bertrand E and R., B. (2004) Assembly and traffic of small nuclear RNPs. *Prog Mol Subcell Biol.* , **35**, 79-97.
- Boulon, S., Verheggen, C., Jady, B.E., Girard, C., Pescia, C., Paul, C., Ospina, J.K., Kiss, T., Matera, A.G., Bordonne, R. and Bertrand, E. (2004) PHAX and CRM1 Are Required Sequentially to Transport U3 snoRNA to Nucleoli. *Molecular Cell*, **16**, 777-787.
- Bousquet-Antonelli, C., Presutti, C. and Tollervey, D. (2000) Identification of a Regulated Pathway for Nuclear Pre-mRNA Turnover. *Cell*, **102**, 765-775.
- Bouveret, E., Rigaut, G., Shevchenko, A., M, W. and B, S. (2000) A Sm-like protein complex that participates in mRNA degradation. *EMBO J*, **19**, 1661-1671.
- Brodersen, D.E., Clemons, J.W.M., Carter, A.P., Wimberly, B.T. and Ramakrishnan, V. (2002) Crystal structure of the 30 s ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 s RNA. *Journal of Molecular Biology*, **316**, 725-768.
- Brouwer, R., Allmang, C., Rajmakers, R., van Aarssen, Y., Egberts, W.V., Petfalski, E., van Venrooij, W.J., Tollervey, D. and Pruijn, G.J. (2001) Three novel components of the human exosome. *J Biol Chem*, **276**, 6177-6184.
- Buttner, K., Wenig, K. and Hopfner, K.-P. (2005) Structural Framework for the Mechanism of Archaeal Exosomes in RNA Processing. *Molecular Cell*, **20**, 461-471.
- Caplan, A.J., Mandal, A.K. and Theodoraki, M.A. (2007) Molecular chaperones and protein kinase quality control. *Trends in Cell Biology*, **17**, 87-92.

- Carlson, B.A., Xu, X.-M., Kryukov, G.V., Rao, M., Berry, M.J., Gladyshev, V.N. and Hatfield, D.L. (2004) Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. *PNAS* **101**, 12848-12853.
- Carmo-Fonseca, M., Platani, M. and Swedlow, J.R. (2002) Macromolecular mobility inside the cell nucleus. *Trends in Cell Biology*, **12**, 491-495.
- Castellano, S., Morozova, N., Morey, M., Berry, M.J., Serras, F., Corominas, M. and Guigó, R. (2001) In silico identification of novel selenoproteins in the *Drosophila melanogaster* genome *EMBO reports*, **2**, 697-702.
- Castellano, S., Novoselov, S., Kryukov, G., Lescure, A., Blanco, E., Krol, A., Gladyshev, V. and Guigó, R. (2004) Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution. *EMBO Rep.* , **5**, 71-77.
- Chang, D.D. and Clayton, D.A. (1987) A novel endoribonuclease cleaves at a priming site of mouse mitochondrial DNA replication. *EMBO J.*, **6**, 409-417.
- Chao, J.A. and Williamson, J.R. (2004) Joint X-ray and NMR refinement of the yeast L30e-mRNA complex. *Structure (Camb)*, **12**, 1165-1176.
- Chavatte, L., Brown, B.A. and Driscoll, D.M. (2005) Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat Struct Mol Biol.*
- Cléry, A., Bourguignon-Igel, V., Allmang, C., Krol, A. and Branlant, C. (2007) An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery. *Nucleic Acids Res.* , **35**, 1868-1884.
- Costa, M., Rodríguez-Sánchez, J., Czaja, A. and Gelpí, C. (2000) Isolation and characterization of cDNA encoding the antigenic protein of the human tRNP(Ser)Sec complex recognized by autoantibodies from patients with type-1 autoimmune hepatitis. *Clin Exp Immunol.* , **121**, 364-374.
- Copeland, P.R., Fletcher, J.E., Carlson, B.A., Hatfield, D.L. and Driscoll, D.M. (2000) A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J*, **19**, 306-314.
- Copeland, P.R., Stepanik, V.A. and Driscoll, D.M. (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.
- de Jesus, L.A., Hoffmann, P.R., Michaud, T., Forry, E.P., Small-Howard, A., Stillwell, R.J., Morozova, N., Harney, J.W. and Berry, M.J. (2006) Nuclear Assembly of UGA Decoding Complexes on Selenoprotein mRNAs: a Mechanism for Eluding Nonsense-Mediated Decay? *Mol Cell Biol*, **26**, 1795-1805.
- de la Cruz, J., Kressler, D., Tollervey, D. and Linder, P. (1998) Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in *Saccharomyces cerevisiae*. *EMBO J.* , **17**, 1128-1140.
- Ding, F. and Grabowski, P.J. (1999) Identification of a protein component of a mammalian tRNA(Sec) complex implicated in the decoding of UGA as selenocysteine. *RNA*, **5**, 1561-1569.
- Dosztanyi, Z., Csizmek, V., Tompa, P. and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434.
- Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M. and Uversky, V.N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS Journal*, **272**, 5129-5148.
- Fagegaltier, D., Carbon, P. and 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., Carbon, P. and Krol, A. (2000a) The selenocysteine insertion sequence binding protein SBP is different from the Y-box protein dbpB. *Biochimie*, **82**, 117-122.
- Fagegaltier, D., Lescure, A., Walczak, R., Carbon, P. and Krol, A. (2000b) Structural analysis of new local features in SECIS RNA hairpins. *Nucleic Acids Res*, **28**, 2679-2689.
- Fatica, A. and Tollervey, D. (2002) Making ribosomes. *Current Opinion in Cell Biology*, **14**, 313-318.
- Fletcher, J.E., Copeland, P.R., Driscoll, D.M. and Krol, A. (2001) The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA*, **7**, 1442-1453.
- Flohe, L., Andreesen, J.R., Brigelius-Flohe, R., Maiorino, M. and Ursini, F. (2000) Selenium, the element of the moon, in life on earth. *IUBMB Life*, **49**, 411-420.
- Flohe, L., Günzler, W. and Schock, H. (1973) Glutathione peroxidase: a selenoenzyme. *FEBS Lett.*, **32**, 132-134.
- Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M.-A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.-M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B. and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature*, **440**, 631-636.
- Gavin, A.-C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K.,

- Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141-147.
- Gonzales, F.A., Zanchin, N.I.T., Luz, J.S. and Oliveira, C.C. (2005) Characterization of *Saccharomyces cerevisiae* Nop17p, a Novel Nop58p-Interacting Protein that is Involved in Pre-rRNA Processing. *Journal of Molecular Biology*, **346**, 437-455.
- Held, W.A., Ballou, B., Mizushima, S. and Nomura, M. (1974) Assembly Mapping of 30 S Ribosomal Proteins from *Escherichia coli*. Further studies. *J. Biol. Chem.*, **249**, 3103-3111.
- Hendrickson, T.L. (2007) Easing selenocysteine into proteins. *Nat Struct Mol Biol*, **14**, 100-101.
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutillier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskata, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W.V., Figeys, D. and Tyers, M. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, **415**, 180-183.
- Houseley, J., LaCava, J. and Tollervey, D. (2006) RNA-quality control by the exosome. *Nat Rev Mol Cell Biol*, **7**, 529-539.
- Kernebeck, T., Lohse, A. and Grötzinger, J. (2001) A bioinformatical approach suggests the function of the autoimmune hepatitis target antigen soluble liver antigen/liver pancreas. *Hepatology*, **34**, 230-233.
- King, T.H., Decatur, W.A., Bertrand, E., Maxwell, E.S. and Fournier, M.J. (2001) A Well-Connected and Conserved Nucleoplasmic Helicase Is Required for Production of Box C/D and H/ACA snoRNAs and Localization of snoRNP Proteins. *Mol. Cell. Biol.*, **21**, 7731-7746.
- Kinzy, S.A., Caban, K. and Copeland, P.R. (2005) Characterization of the SECIS binding protein 2 complex required for the co-translational insertion of selenocysteine in mammals. *Nucleic Acids Res*, **33**, 5172-5180.
- Klein, D.J., Schmeing, T.M., Moore, P.B. and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J*, **20**, 4214-4221.
- Kressler, D., Doere, M., Rojo, M. and Linder, P. (1999) Synthetic Lethality with Conditional dbp6 Alleles Identifies Rsa1p, a Nucleoplasmic Protein Involved in the Assembly of 60S Ribosomal Subunits. *Mol. Cell. Biol.*, **19**, 8633-8645.
- Kryukov, G. and Gladyshev, V. (2004) The prokaryotic selenoproteome. *EMBO Rep.*, **5**, 538-543.
- Kufel, J., Allmang, C., Chanfreau, G., Petfalski, E., Lafontaine, D.L. and Tollervey, D. (2000) Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Mol Cell Biol*, **20**, 5415-5424.
- Kufel, J., Allmang, C., Petfalski, E., Beggs, J. and Tollervey, D. (2003a) Lsm Proteins are required for normal processing and stability of ribosomal RNAs. *J Biol Chem*, **278**, 2147-2156.
- Kufel, J., Allmang, C., Verdone, L., Beggs, J. and Tollervey, D. (2003b) A complex pathway for 3' processing of the yeast U3 snoRNA. *Nucleic Acids Res*, **31**, 6788-6797.
- Kufel, J., Allmang, C., Verdone, L., Beggs, J.D. and Tollervey, D. (2002) Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. *Mol Cell Biol*, **22**, 5248-5256.
- LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A. and Tollervey, D. (2005) RNA Degradation by the Exosome Is Promoted by a Nuclear Polyadenylation Complex. *Cell*, **121**, 713-724.
- Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499-512.
- Lescure, A., Allmang, C., Yamada, K., Carbon, P. and Krol, A. (2002) cDNA cloning, expression pattern and RNA binding analysis of human selenocysteine insertion sequence (SECIS) binding protein 2. *Gene*, **291**, 279-285.
- Lorentzen, E., Dziembowski, A., Lindner, D., Seraphin, B. and Conti, E. (2007) RNA channelling by the archaeal exosome. *EMBO reports*, **8**, 470-476.
- Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G. and Conti, E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat Struct Mol Biol*, **12**, 575-581.
- Low, S.C., Harney, J.W. and Berry, M.J. (1995) Cloning and Functional Characterization of Human Selenophosphate Synthetase, an Essential Component of Selenoprotein Synthesis. *J. Biol. Chem.*, **270**, 21659-21664.

- Lygerou, Z., Allmang, C., Tollervey, D. and Seraphin, B. (1996) Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science*, **272**, 268-270.
- Matera, A.G., Terns, R.M. and Terns, M.P. (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol*, **8**, 209-220.
- Mayes, A., Verdone, L., Legrain, P. and Beggs, J. (1999) Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J*, **18**, 4321-4331.
- Milligan, L., Torchet, C., Allmang, C., Shipman, T. and Tollervey, D. (2005) A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Mol Cell Biol*, **25**, 9996-10004.
- Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M. and Tollervey, D. (2003) Rrp47p Is an Exosome-Associated Protein Required for the 3' Processing of Stable RNAs. *Mol. Cell. Biol.*, **23**, 6982-6992.
- Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. and Tollervey, D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell*, **91**, 457-466.
- Mizushima S and M, N. (1970) Assembly mapping of 30S ribosomal proteins from E. coli. *Nature*, **226**, 1214.
- Moghadaszadeh, B. and Beggs, A.H. (2006) Selenoproteins and Their Impact on Human Health Through Diverse Physiological Pathways. *Physiology 2006*, **21**, 307-315.
- Moine, H., Cachia, C., Westhof, E., Ehresmann, B. and Ehresmann, C. (1997) The RNA binding site of S8 ribosomal protein of Escherichia coli: Selex and hydroxyl radical probing studies. *RNA*, **3**, 255-268.
- Moore, T., Zhang, Y., Fenley, M.O. and Li, H. (2004) Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure (Camb)*, **12**, 807-818.
- Mougel, M., Allmang, C., Eyermann, F., Cachia, C., Ehresmann, B. and Ehresmann, C. (1993) Minimal 16S rRNA binding site and role of conserved nucleotides in Escherichia coli ribosomal protein S8 recognition. *Eur J Biochem*, **215**, 787-792.
- Mougel, M., Eyermann, F., Westhof, E., Romby, P., Expert-Bezancon, A., Ebel, J.P., Ehresmann, B. and Ehresmann, C. (1987) Binding of Escherichia coli ribosomal protein S8 to 16 S rRNA. A model for the interaction and the tertiary structure of the RNA binding site. *J Mol Biol.*, **198**, 91-107.
- Pannone, B., Xue, D. and Wolin, S. (1998) A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J.*, **17**, 7442-7453.
- Papp, L.V., Lu, J., Striebel, F., Kennedy, D., Holmgren, A. and Khanna, K.K. (2006) The redox state of SECIS binding protein 2 controls its localization and selenocysteine incorporation function. *Mol Cell Biol*, **26**, 4895-4910.
- Pearl, L.H. and Prodromou, C. (2006) Structure and mechanism of the HSP90 molecular chaperone machinery. *Annual Review of Biochemistry*, **75**, 271-294.
- Petfalski, E., Dandekar, T., Henry, Y. and Tollervey, D. (1998) Processing of the Precursors to Small Nucleolar RNAs and rRNAs Requires Common Components. *Mol. Cell. Biol.*, **18**, 1181-1189.
- Qichang Shen, L.F.P.E.N. (2006) Nuclease sensitive element binding protein 1 associates with the selenocysteine insertion sequence and functions in mammalian selenoprotein translation. *Journal of Cellular Physiology*, **207**, 775-783.
- Raijmakers, R., Schilders, G. and Pruijn, G. (2004) The exosome, a molecular machine for controlled RNA degradation in both nucleus and cytoplasm. *Eur J Cell Biol.*, **83**, 75-83.
- Rinke, J. and Steitz, J.A. (1982) Precursor molecules of both human 5S ribosomal RNA and transfer RNAs are bound by a cellular protein reactive with anti-La Lupus antibodies. *Cell*, **29**, 149-159.
- Rotruck, J., Pope, A., Ganther, H., Swanson, A., Hafeman, D. and Hoekstra, W. (1973) Selenium: biochemical role as a component of glutathione peroxidase. *Science*, **179**, 588-590.
- Small-Howard, A., Morozova, N., Stoytcheva, Z., Forry, E.P., Mansell, J.B., Harney, J.W., Carlson, B.A., Xu, X.M., Hatfield, D.L. and Berry, M.J. (2006) Supramolecular complexes mediate selenocysteine incorporation in vivo. *Mol Cell Biol*, **26**, 2337-2346.
- Stebbins, C.E., Russo, A.A., Schneider, C., Rosen, N., Hartl, F.U. and Pavletich, N.P. (1997) Crystal Structure of an Hsp90-Geldanamycin Complex: Targeting of a Protein Chaperone by an Antitumor Agent. *Cell*, **89**, 239-250.
- Te, J., Jia, L., Rogers, J., Miller, A. and Hartson, S.D. (2007) Novel Subunits of the Mammalian Hsp90 Signal Transduction Chaperone. *J. Proteome Res.*, **6**, 1963-1973.
- Tishchenko, S., Nikulin, A., Fomenkova, N., Nevskaya, N., Nikonov, O., Dumas, P., Moine, H., Ehresmann, B., Ehresmann, C., Piendl, W., Lamzin, V., Garber, M. and Nikonov, S. (2001) Detailed analysis of RNA-protein interactions within the ribosomal protein S8-rRNA complex from the archaeon Methanococcus jannaschii. *Journal of Molecular Biology*, **311**, 311-324.
- Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*, **579**, 3346-3354.

- Venema, J., Henry, Y. and Tollervey, D. (1995) Two distinct recognition signals define the site of endonucleolytic cleavage at the 5'-end of yeast 18S rRNA. *EMBO J.* , **14**, 1995
- Venema, J. and Tollervey, D. (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet.* , **33**, 261-311.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. and Ficner, R. (2000a) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell*, **6**, 1331-1342.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. and Ficner, R. (2000b) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell*, **6**, 1331-1342.
- Walczak, R., Hubert, N., Carbon, P. and Krol, A. (1997) Solution structure of SECIS, the mRNA element required for eukaryotic selenocysteine insertion-interaction studies with the SECIS-Binding Protein SBP. *Biomed. and Environ. Sci.*, **10**, 177-181.
- Walczak, R., Westhof, E., Carbon, P. and Krol, A. (1996) A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA*, **2**, 367-379.
- Watkins, N.J., Lemm, I., Ingelfinger, D., Schneider, C., Hobach, M., Urlaub, H. and Luhrmann, R. (2004) Assembly and Maturation of the U3 snoRNP in the Nucleoplasm in a Large Dynamic Multiprotein Complex. *Mol Cell*, **16**, 789-798.
- Xu, X., Carlson, B., Irons, R., Mix, H., Zhong, N., Gladyshev, V. and DL, H. (2007a) Selenophosphate synthetase 2 is essential for selenoprotein biosynthesis. *Biochem J.* , **404**, 115-120.
- Xu, X.-M., Carlson, B.A., Mix, H., Zhang, Y., Saira, K., Glass, R.S., Berry, M.J., Gladyshev, V.N. and Hatfield, D.L. (2007b) Biosynthesis of Selenocysteine on Its tRNA in Eukaryotes. *PLoS Biology*, **5**, e4.
- Xu, X.-M., Mix, H., Carlson, B.A., Grabowski, P.J., Gladyshev, V.N., Berry, M.J. and Hatfield, D.L. (2005) Evidence for Direct Roles of Two Additional Factors, SECp43 and Soluble Liver Antigen, in the Selenoprotein Synthesis Machinery. *J. Biol. Chem.*, **280**, 41568-41575.
- Yong, J., Wan, L. and Dreyfuss, G. (2004) Why do cells need an assembly machine for RNA-protein complexes? *Trends in Cell Biology*, **14**, 226-232.
- Yoo, C.J. and Wolin, S.L. (1997) The Yeast La Protein Is Required for the 3' Endonucleolytic Cleavage That Matures tRNA Precursors. *Cell*, **89**, 393-402.
- Yuan, J., Palioura, S., Salazar, J.C., Su, D., O'Donoghue, P., Hohn, M.J., Cardoso, A.M., Whitman, W.B. and Soll, D. (2006) RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *PNAS*, **103**, 18923-18927.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D. and Noller, H.F. (2001) Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science*, **292**, 883-896.
- Zavacki, A.M., Mansell, J.B., Chung, M., Klimovitsky, B., Harney, J.W. and Berry, M.J. (2003) Coupled tRNA(Sec)-dependent assembly of the selenocysteine decoding apparatus. *Mol Cell*, **11**, 773-781.
- Zhao, R., Davey, M., Hsu, Y.-C., Kaplanek, P., Tong, A., Parsons, A.B., Krogan, N., Cagney, G., Mai, D., Greenblatt, J., Boone, C., Emili, A. and Houry, W.A. (2005) Navigating the Chaperone Network: An Integrative Map of Physical and Genetic Interactions Mediated by the Hsp90 Chaperone. *Cell*, **120**, 715-727.



*Principales publications*





# Selenoprotein synthesis: UGA does not end the story

C. Allmang, A. Krol \*

*Institut de Biologie Moléculaire et Cellulaire, UPR 9002 du CNRS Architecture et Réactivité de l'ARN,  
Université Louis-Pasteur, 15, rue René-Descartes, 67084 Strasbourg Cedex, France*

Received 1 February 2006; accepted 24 April 2006  
Available online 19 May 2006

## Abstract

It is well established that the beneficial effects of the trace element selenium are mediated by its major biological product, the amino acid selenocysteine, present in the active site of selenoproteins. These fulfill different functions, as varied as oxidation-reduction of metabolites in bacteria, reduction of reactive oxygen species, control of the redox status of the cell or thyroid hormone maturation. This review will focus on the singularities of the selenocysteine biosynthesis pathway and its unique incorporation mechanism into eukaryal selenoproteins. Selenocysteine biosynthesis from serine is achieved on tRNA<sup>Sec</sup> and requires four proteins. As this amino acid is encoded by an in-frame UGA codon, otherwise signaling termination of translation, ribosomes must be told not to stop at this position in the mRNA. Several molecular partners acting in *cis* or in *trans* have been identified, but their knowledge has not enabled yet to firmly establish the molecular events underlying this mechanism. Data suggest that other, so far uncharacterized factors might exist. In this survey, we attempted to compile all the data available in the literature and to describe the latest developments in the field.

© 2006 Elsevier Masson SAS. All rights reserved.

**Keywords:** Selenium; Selenocysteine; Selenoproteins; tRNA<sup>Sec</sup>; SECIS RNA-protein interactions; SECp43; SLA/LP

## 1. Introduction

The element selenium was discovered by the Swedish chemist Berzelius in 1817 and named after Sélênê, the goddess of moon. This non-metal was long considered as a potent toxic substance, especially to grazing animals that would eat selenium accumulator plants of the genus *Astragalus* during periods of drought in arid or desert regions of western USA and China. Between 1930 and the mid-1950s, selenium attracted the attention of animal nutritionists who eventually defined it as an essential micronutrient endowed with a number of significant health benefits [reviewed in 1,2]. In the 1970s, the biological activity of selenium could be attributed to selenocysteine, a then novel amino acid found in selenoproteins. The majority of selenoproteins whose function is known are oxidation-reduction enzymes using selenocysteine in the active site. The chemical structure of selenocysteine differs from cysteine only by the selenium instead of the sulfur atom; however, the electronic structure of the selenium atom renders the selenolate an-

ion, the conjugated base of selenocysteine, more stable than the corresponding cysteine thiolate. The selenol proton is thus more acidic than in the cysteine thiol (pK<sub>a</sub> of 5.2 versus 8.5 for the thiol), hence ionization of selenocysteine at physiological pH.

A further breakthrough appeared in the mid-1980s with the discovery that selenocysteine is encoded by UGA, a codon otherwise specifying termination of protein synthesis. Immediately, this finding aroused the interest of the scientific community who aimed at challenging this novel alternate reading of the genetic code. It is largely the pioneering work in *E. coli*, by the group of August Böck, that helped solve how selenocysteine is biosynthesized and specifically incorporated into selenoproteins in response to UGA [reviewed in 3]. Selenoproteins have been found in the three kingdoms of life, but not in all species of bacteria, archaea and eukarya. For example, neither fungi nor higher plants can incorporate selenocysteine at specific locations. How ribosomes are told not to stop at UGA Sec codons results from the combined action of several partners, acting in *cis* or in *trans*. The underlying mechanisms in archaea and eukarya present similarities but also dissimilarities to bacteria that will be discussed in this review. Focus will be put

\* Corresponding author.

E-mail address: [a.krol@ibmc.u-strasbg.fr](mailto:a.krol@ibmc.u-strasbg.fr) (A. Krol).

primarily on eukarya with comparisons to the bacterial and archaeal systems wherever needed. Two aspects will be addressed, biosynthesis of selenocysteine in the first place, followed by its co-translational incorporation into selenoproteins.

## 2. Biosynthesis of selenocysteine

Selenocysteine does not occur as a free amino acid. Thus, the first step of its biosynthesis consists in the charge of serine on the specific tRNA<sup>Sec</sup> by the conventional seryl-tRNA synthetase. The Ser-tRNA<sup>Sec</sup> is next converted into Sec-tRNA<sup>Sec</sup> by selenocysteine synthase that utilizes monoselenophosphate as the substrate. This compound is produced from sodium selenite or more likely selenide by a reaction catalyzed by selenophosphate synthetase. We will describe in this paragraph the characteristic features of tRNA<sup>Sec</sup> and the selenocysteine biosynthesis pathway.

### 2.1. Structure-function of the tRNA<sup>Sec</sup>

Secondary structure models for tRNAs<sup>Sec</sup> are shown in Fig. 1, arising from experimental determination in bacteria and eukarya [4–6], or structure-based sequence alignments in archaea [6]. Two main characteristic features distinguish tRNAs<sup>Sec</sup> from canonical tRNAs. First, they share the hallmark of having a 6 bp D-stem, instead of 3–4 bp in other tRNAs. This extended D-stem was shown to be a major identity determinant for serine phosphorylation [7], a likely intermediate in selenocysteine biosynthesis in eukarya (see below). Second, the amino acid acceptor arm (A-T), resulting from coaxial stacking of the A and T-stems, is longer in tRNAs<sup>Sec</sup> (13 bp) than in canonical tRNAs where it is 12 bp long (7 + 5 bp). In bacteria, the 13 bp A-T arm is formed by coaxial stacking of the 8 bp A-stem and 5 bp T-stem whereas the same length is obtained in archaea and eukarya by stacking of the longer A-stem (9 bp) and shorter T-stem (4 bp) [5,6,8–14]. This evolutionary conservation is obviously a signal for one or more li-

gand(s). In bacteria, the extra length of the A-T arm is a determinant for binding to the specialized translation elongation factor SelB whereas it is required for serine to selenocysteine conversion in eukarya [15,16].

The position and nature of post-transcriptional modifications have been investigated in the vertebrate tRNA<sup>Sec</sup> [17, 18]. It contains only four modified bases, thus fewer than canonical tRNAs. Apart from pseudo-U55 and m<sup>1</sup>A58 in the T-loop, mass spectrometry identified 6-isopentenyl-A37 (i<sup>6</sup>A37) and mcm<sup>5</sup>Um34, the 5-methylcarboxymethyl-2'-O-methyluridine modification, in the anticodon loop. The 2' O-ribose modification, associated to mcm<sup>5</sup>U, has been found so far in tRNA<sup>Sec</sup> only and its yield is a function of the dietary selenium status [17]. Formation of mcm<sup>5</sup>U34 depends on the tRNA<sup>Sec</sup> tertiary structure and completion of all the other base modifications [19]. Interestingly, protein SECp43 identified earlier in a complex with the tRNA<sup>Sec</sup> [20], might be involved directly or indirectly in the 2'-O-methylation of mcm<sup>5</sup>U34 [21]. Modification of i<sup>6</sup>A37 has also a great importance as its absence produced a severe down effect on selenoprotein synthesis [22]. However, as conversion of A37 to i<sup>6</sup>A37 occurs before U34 is modified to mcm<sup>5</sup>Um and is indeed required for obtaining the latter, it was difficult to assign the observed effect to the lack of one or the other modification. To address the issue, knock-out transgenic mice were obtained wherein the tRNA<sup>Sec</sup> was replaced by the wt or a mutant transgene producing a tRNA that lacked both the U34 and A37 modified bases [23]. This study concluded that U34 modification has a greater influence than i<sup>6</sup>A37 in regulating the expression of various mammalian selenoproteins.

### 2.2. The Ser-tRNA<sup>Sec</sup> to Sec-tRNA<sup>Sec</sup> conversion step

Neither in eukarya nor in archaea has been isolated monoselenophosphate, the biological donor of selenium in bacteria. Two enzymes catalyzing formation of this compound have however been described. A human cDNA of selenophosphate

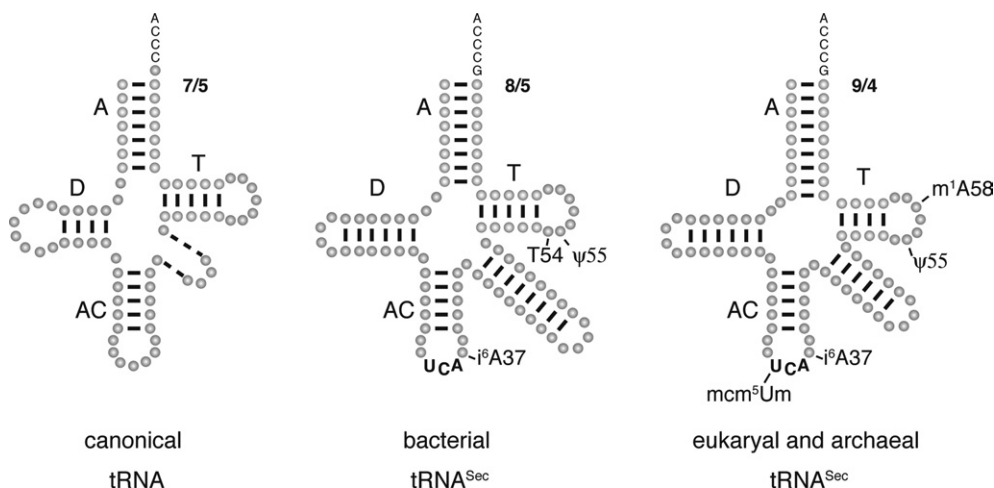


Fig. 1. Secondary structure comparisons of canonical tRNAs versus selenocysteine tRNAs<sup>Sec</sup>. The various secondary structure elements are indicated: A, D, AC, and T stand for the amino acid, D, anticodon and T stems, respectively. 7/5, 8/5, 9/4 indicate the number of base pairs forming the coaxial A-T arm in the tRNAs shown. Dashes in the canonical tRNA structure signify that the extra arm is of variable length in different tRNAs. Modified bases are indicated where identified in the bacterial and eukaryal tRNAs<sup>Sec</sup>. They were omitted in the canonical tRNA. The archaeal tRNA<sup>Sec</sup> was not investigated for its base modification content.

synthetase 1 (SPS1) was initially cloned, showing only 32% of amino acid sequence similarity with its bacterial homolog [24]. Both the bacterial and SPS1 enzymes are active in mammalian cells but SPS1 is unable to complement an inactive bacterial gene. Another selenophosphate synthetase cDNA was later cloned in mammals, called SPS2 to differentiate it from the former one [25]. A very interesting key feature of this enzyme is the presence of a selenocysteine residue, suggesting that it possesses a higher catalytic activity than SPS1. As selenocysteine, and thus monoselenophosphate, is needed prior to SPS2 synthesis, it has been proposed that SPS1 contributes to manufacture basal levels of this amino acid. SPS2 could then function as a privileged effector under stimulatory conditions.

Fig. 2 summarizes the essential steps leading to Sec-tRNA<sup>Sec</sup>, implying the identified factors and their established or putative function. No specific Sec-tRNA synthetase has been identified so far and it is likely that the Ser-tRNA synthetase serylates the tRNA<sup>Sec</sup> in vivo, as it does in vitro [26]. In *E. coli*, selenocysteine synthase, a pyridoxal phosphate enzyme, catalyzes the Ser-tRNA<sup>Sec</sup> to Sec-tRNA<sup>Sec</sup> conversion on the tRNA<sup>Sec</sup>. So far, no protein has been isolated in archaea or eukarya carrying a selenocysteine synthase activity. A candidate, showing blocks of amino acid sequence similarity to the

*E. coli* selenocysteine synthase, was isolated in the archeon *M. jannaschii* [27]. Its crystal structure revealed a multimeric organization reminiscent of the *E. coli* enzyme. However, this protein was unable to ensure the Ser to Sec conversion in vitro. On the eukaryal front, a protein identified more than twelve years ago is now attracting attention. It was discovered as part of a ribonucleoprotein particle containing a 48 kDa protein generating autoantibodies in a group of patients with a severe form of autoimmune chronic active hepatitis; the autoantibodies precipitated the tRNA<sup>Sec</sup> in human whole cell extracts [28]. In further investigations, cDNAs encoding this protein, now called SLA/LP for Soluble Liver Antigen/Liver Pancreas, were obtained and sequenced [29,30]. A theoretical study with a bioinformatic approach predicted that the SLA/LP sequence is compatible with the architecture of the superfamily of pyridoxal phosphate-dependent transferases [31], indicating that it might possess a selenocysteine synthase function. Recent data indicated that SLA/LP indeed participates in the pathway of selenoprotein synthesis. The same authors established that SECp43 and SLA/LP co-exist in a complex in vivo with the tRNA<sup>Sec</sup> and that the former protein may act as a chaperone to address SLA/LP to the nucleus [21]. Moreover, other investigators reported that SLA/LP and SPS1 interact in vitro and in

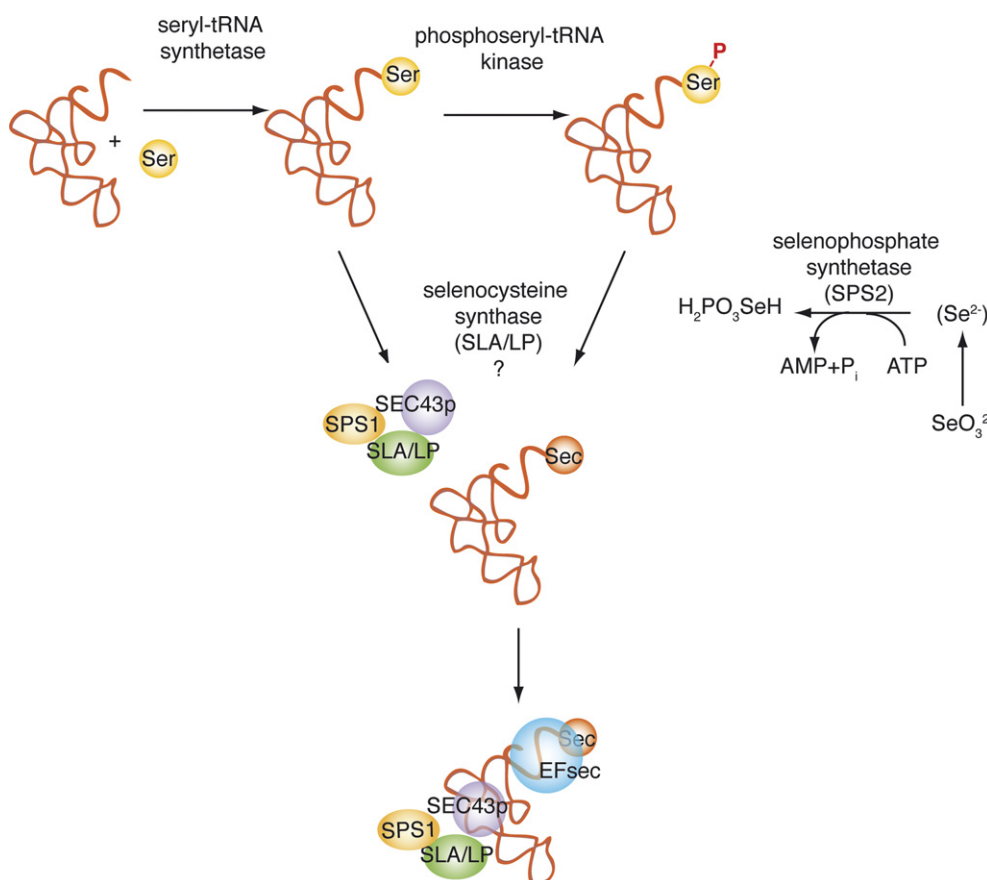


Fig. 2. Putative selenocysteine biosynthesis pathways in eukarya. The tRNA<sup>Sec</sup> is charged with serine by the conventional Seryl-tRNA synthetase and the seryl residue of Ser-tRNA<sup>Sec</sup> is phosphorylated by the phosphoseryl-tRNA kinase. The seryl to selenocysteyl conversion occurs either via the phosphoseryl intermediate or directly. The phosphoseryl residue could also harbor a regulatory function. The selenocysteine synthase activity could be borne by SLA/LP alone or by the SLA/LP-SECp43 complex. The specialized translation elongation factor EFsec binds the tRNA<sup>Sec</sup> alone or the SLA/LP-SECp43-tRNA<sup>Sec</sup> complex. Selenophosphate synthetase SPS2 catalyzes formation of monoselenophosphate from selenite (SeO<sub>3</sub><sup>2-</sup>) or more likely from an unstable selenide compound depicted as (Se<sup>2-</sup>). The role of SPS1 is still elusive.

vivo and that SECp43 indeed helps redistributing these proteins to the nucleus [32]. However, full characterization of SECp43 and SLA/LP must await further studies, in particular to ascertain whether SLA/LP does possess the selenocysteine synthase activity. In this regard, both the *M. jannaschii* and SLA/LP proteins exhibit local amino acid sequence similarity to the *E. coli* selenocysteine synthase, in particular in the vicinity of an essential active site lysine in the *E. coli* enzyme [33]. Indeed, the pathway of selenocysteine biosynthesis appears more sophisticated in archaea and eukarya than in bacteria. This is exemplified by the identification and characterization of a specific phosphoseryl-tRNA<sup>Sec</sup> kinase (PSTK) in archaea and mammals [27,34]. Such a kinase activity was detected more than 30 years ago [34 and references therein]. Interestingly, finding the PSTK gene only in those archaea and eukarya that possess the capacity of synthesizing selenoproteins, strongly argues in favor of the important role that this enzyme must play in selenocysteine synthesis.

Whether phosphoseryl-tRNA<sup>Sec</sup> is an obligatory intermediate in selenocysteine biosynthesis or participates in its regulation is still a matter of debate. In any event, a strikingly similar mechanism was discovered recently for cysteine biosynthesis in several methanogenic archaea, such as *M. jannaschii*, that lack cysteinyl-tRNA synthetase [35]. The alternative route that was described to provide Cys-tRNA<sup>Cys</sup> consists in aminoacylation of the tRNA<sup>Cys</sup> with O-phosphoserine by an O-phosphoseryl-tRNA synthetase (SepRS). The Sep-tRNA<sup>Cys</sup> is further converted to Cys-tRNA<sup>Cys</sup> by a Sep-tRNA:Cys-tRNA synthase. This puzzling similarity to selenocysteine biosynthesis suggests the interesting possibility that a common mechanism was shared for cysteine and selenocysteine biosynthesis in the primordial times.

### 3. Molecular partners for co-translational incorporation of selenocysteine into selenoproteins

In bacteria, the pathway is now well elucidated and proceeds as follows. Two molecular partners are involved. The *cis*-acting one is a stem-loop structure, called SECIS (SElenoCysteine Insertion Sequence), embedding the UGA codon and residing in the open reading frame of selenoprotein mRNAs. The factor acting in *trans* is protein SelB, a translation elongation factor dedicated to selenoprotein synthesis. As a matter of fact EF-Tu, the general translation elongation factor, is unable to recognize the Sec-tRNA<sup>Sec</sup> [3]. SelB is composed of two domains. The N-terminal one is highly sequence-similar and functionally homologous to EF-Tu; the smaller, additional C-terminal domain binds the SECIS stem-loop by recognizing a very limited number of nucleotides at its apex. The Sec-tRNA<sup>Sec</sup>, harbored by SelB, is thus conveyed to the A site of the ribosome to decode the UGA Sec codon. Eukarya, and to a lesser extent archaea, have been also investigated for their abilities to biosynthesize and incorporate selenocysteine. A higher degree of complexity arose in these two kingdoms as a consequence of the localization of the SECIS element outside of the coding region. In contrast to bacteria, not all the components

are identified and the major mechanistic steps of this process are still unclear [reviewed in 1,3,36–39].

#### 3.1. SECIS RNA structures in eukarya and archaea

Fig. 3A shows the secondary structure model of the eukaryal SECIS elements, derived from extensive structure probing studies and site-directed mutagenesis [40–42, reviewed in 43]. Only the conserved sequences are displayed. The foot of helix II is constituted by four consecutive non-Watson-Crick base pairs — the quartet — which is a motif essential to selenocysteine incorporation in vivo [40,41]. Within the quartet, the tandem of G•A base pairs with the sheared geometry is of prime importance [41]. The presence of such a tandem of G•A base pairs was detected earlier in other RNAs such as ribosomal and snRNAs, constituting a recurrent motif called the kink-turn, or K-turn motif, and we recently proposed that SECIS RNAs can also adopt a K-turn motif [43]. The predicted structure of the SECIS RNA K-turn is depicted in Fig. 3B, using the scheme proposed for K-turn RNAs in [44] and the graphical nomenclature of non-Watson-Crick base pairs described in [45].

A more detailed sequence and structure analysis established that there exists in fact two slightly different SECIS RNA secondary structure models, only varying at the apex, and giving rise to Forms 1 and 2 [46,47]. Form 2 SECIS possesses an additional helix III but a shorter apical loop, compared to Form 1 (Fig. 3A). As a consequence, the conserved run of As lies in an internal loop (Form 2) instead of the apical loop (Form 1). More systematic identification of a variety of novel selenoprotein mRNAs including vertebrates, invertebrates and green algae clearly indicated that Form 2 SECIS are more widespread than Form 1. However, swapping experiments could not assess that Form 2, although preponderant, provides a functional advantage to selenocysteine incorporation. It is even remarkable that mRNAs encoding the same selenoprotein can harbor either a Form 1 or a Form 2 SECIS, depending on the animal species [reviewed in 43]. NMR and UV melting data are consistent with the 2D models and the existence of Form 1 and Form 2 SECIS but the authors did not find evidence in favor of the existence of the sheared G•A base pairs [48]. One possibility to explain the absence of the sheared G•A base pair signature may reside in the choice of the investigators for short SECIS hairpins lacking helix I, thus less stable and prone to adopt a different fold.

Selenoprotein mRNAs in archaea also contain a functional SECIS element in the untranslated regions [49,50]. It resides in the 3' UTR in the vast majority of the cases, but was surprisingly found once in the 5' UTR. The 2D structure of the archaeal SECIS RNA was derived by structure probing and sequence comparisons [50,51], leading to the consensus structure [51] shown in Fig. 3C. The archaeal SECIS differs from eukarya by the remarkable absence of the non-Watson-Crick quartet. Given the prime importance of this motif in eukaryal SECIS, it is unlikely that the archaeal and eukaryal SECIS can functionally substitute for each other.



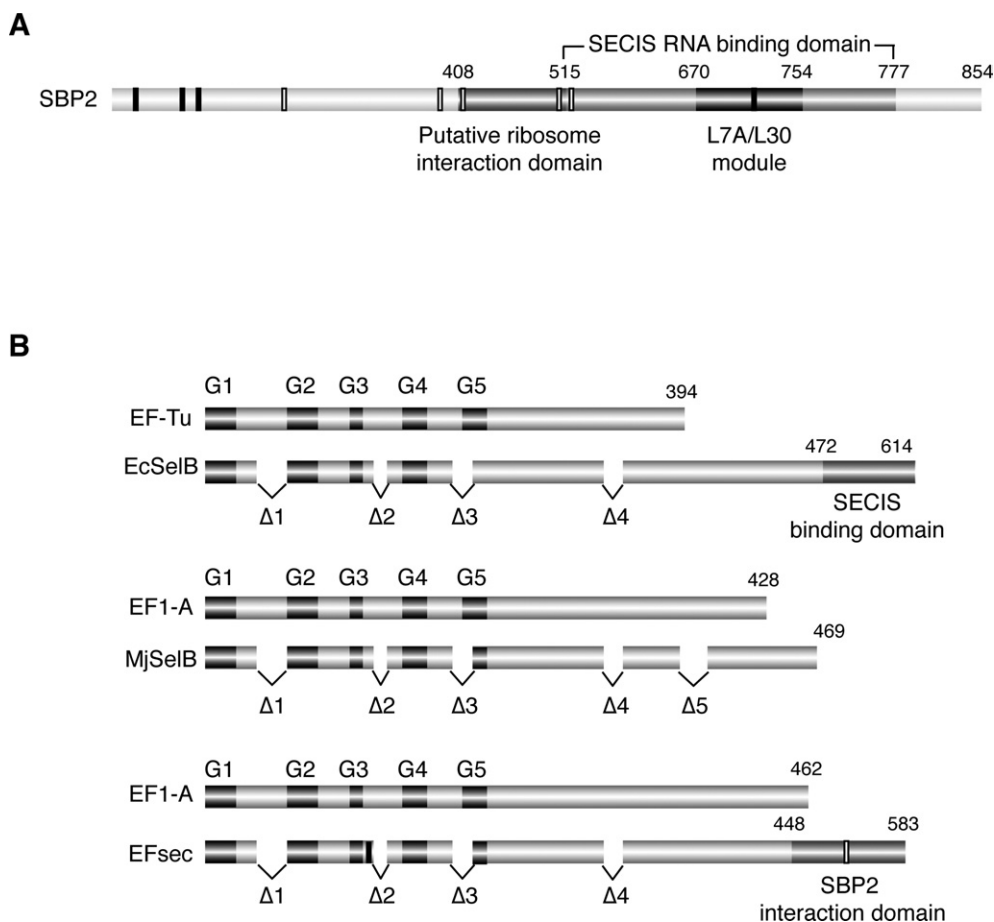


Fig. 4. Proteins involved in eukaryal selenoprotein synthesis. (A) Representation of the SECIS binding protein 2 with the SECIS RNA binding domain, the L7A/L30 module and the putative ribosome interaction domain. The N-terminal (1-408) and very C-terminal (777-854) portions have unknown function. (B) Schematic drawings of the specialized translation elongation factors in *E. coli* (EcSelB), *Methanococcus jannaschii* (MjSelB) and eukarya (EFsec), in comparison with the general elongation factors EF-Tu or EF1-A. The C-terminal extensions carry the SECIS binding activity in EcSelB and the SBP2 interaction domain in EFsec; the role of the MjSelB C-terminal extension has not been assigned yet. The GTP binding domains are depicted (G1-G5);  $\Delta$ 1- $\Delta$ 5 are the deletion regions relative to EF-Tu/EF1-A. The predicted nuclear export and localization signals in SBP2 and EFsec (filled and open rectangles, respectively) are from [55].

[reviewed in 39]. It is puzzling that a database search for invertebrate SBP2 sequences yielded only putative SBP2 lacking the N-terminal domain [A.K., unpublished data]. This could be connected to the finding that this domain was not essential for selenoprotein synthesis in rabbit reticulocyte lysates [52], pointing to a possible regulatory or fine-tuning function in vertebrates. Amino acid sequence comparisons showed that the SECIS RNA binding domain contains the L7A/L30 module, shared by other functionally unrelated proteins such as ribosomal proteins L7A(e) and L30, U4 small nuclear RNP protein 15.5 kD/Snu13p, small nucleolar RNP Nhp2p, all of which bind K-turn RNAs [52,56, reviewed in 43]. Two aspects of the SECIS RNA-SBP2 interactions were investigated. The first one aimed at delineating the SECIS RNA regions interacting with SBP2, the second one looking for SBP2 amino acids important for binding. Footprinting and site-directed mutagenesis experiments established that the non-Watson-Crick quartet of the SECIS RNA, as well as phosphates distributed along helix I, are important sequence and structural determinants for SBP2 binding [57]. The invariant U residing 5' to the G•A sheared base pairs (Fig. 3A,B) has very recently revealed its impor-

tance for SBP2 binding [58]. Indeed, patients carrying a homozygous point mutation in the gene encoding selenoprotein N (SEPN), converting this U to a C, developed congenital muscular dystrophies known as SEPN1-related myopathies. This leads to impairment of selenoprotein N synthesis, very likely caused by the inability of SBP2 to bind the SECIS mutant in vivo, as shown in vitro by a gel shift assay.

Taking advantage of the crystal structure of the U4 snRNA-15.5 kD complex, a structure-guided strategy followed by experimental validation proposed a biochemical model describing putative SECIS RNA-SBP2 contacts [56]. Although awaiting confirmation from the crystal structure of the complex, the model indicated that similar RNA-protein interaction principles exist between the U4 snRNA-15.5 kD and the SECIS RNA-SBP2 complexes. Up to now, a bunch of proteins of the L7A/L30 family and a number of diverse RNAs containing the K-turn motif have been identified, the majority of the K-turns being localized in the small and large ribosomal RNAs. Altogether, these findings suggest that the L7A/L30 fold and the K-turn are ancient structural motifs that have evolved specialized roles in many different biological processes.

Attempts to find a SECIS binding activity have not been successful so far in archaeal extracts. Given the absence of sheared base pairs in archaeal SECIS RNAs, one cannot expect proteins of the L7A/L30 family to bind, rendering difficult an *in silico* search.

### 3.2.2. Ribosomal protein L30

This protein is specific to the eukaryal and archaeal kingdoms, although not all archaeal ribosomes possess it. Its role in translation is still elusive. Interestingly, the rat L30 protein was reported to be a novel component of the selenoprotein synthesis machinery [59]. It binds the SECIS RNA *in vivo* and *in vitro*, and competes efficiently with SBP2 for the SECIS RNA *in vitro*. In addition, the ribosome-associated L30 interacts with a higher affinity to the SECIS RNA than the recombinant version. This observation prompted the authors to propose a model in which L30 displaces transiently SBP2 to bring the SECIS RNA to the vicinity of the ribosomal A site. L30, however, was localized by another group at the interface between the large and small subunits in the cryo-EM map of the 80S wheat germ ribosomes, in a region distant from the A site [60]. How to reconcile the two sets of data will undoubtedly emerge from further experiments.

The interaction of SBP2 and L30 at the SECIS RNA raised the question of whether other L7A/L30 proteins could recognize it as well. The answer was positive for L7Ae (the archaeal version of L7A) and 15.5 kD/Snu13p but SBP2 was unable to interact with U4 snRNA or an L7Ae RNA target [A.Cléry, C. A., A.K and C.Branlant, manuscript in preparation]. This experiment indicated that the SBP2 RNA binding domain is more complex than in the other proteins of the family, the SECIS RNA binding specificity being very likely provided by amino acids flanking the L7A/L30 module. In fact, our unpublished data support this hypothesis.

### 3.2.3. Other SECIS-binding proteins

The existence of SECIS-binding protein activities was reported before the discovery of SBP2 [61–63]. The same cDNA was obtained independently by two groups and by two different methods using either northwestern cloning or the three-hybrid system [64,65]. Surprisingly, it corresponded to the sequence of a cold-shock protein known in databases as dbpB or Y-box binding protein, a transcriptional activator in bacteria. The predicted amino acid sequences showed also perfect similarity with the eukaryotic p50. Protein p50 was detected in free and polysomal mRNPs and associates very tightly with all kinds of mRNA nucleotide sequences, very likely to ensure mRNA storage. Further experiments established that the recombinant dbpB was unable to bind the SECIS RNA, suggesting that it was not a bona fide SECIS-binding protein [65]. Very recent data, however, pointed to a possible role for dbpB in selenoprotein synthesis. Indeed, renamed as NSEP1 standing for nuclease sensitive element binding protein 1, it was found associated to the SECIS RNA *in vivo* and its knock-down by RNAi induced reduction of the activity of a chimeric reporter gene [66]. NSEP1 may therefore function either in direct sup-

port of the selenoprotein synthesis machinery or as a more general mRNA stabilizing element.

### 3.3. The specialized translation elongation factors

The archaeal *M. jannaschii* (MjSelB) and mouse selenocysteine-specialized elongation factors were characterized [67–69]. The mouse protein was called either EFsec [68] or mSelB [69] but, for reason of convenience, we will designate it hereafter as EFsec. Similarly to bacterial SelB, MjSelB and EFsec are composed of two domains (Fig. 4B), the N-terminal one being functionally homologous to the corresponding conventional elongation factor EF1-A. The bacterial SelB C-terminal extension possesses the SECIS RNA binding activity. In contrast, the C-terminal extensions in MjSelB and EFsec show no amino acid sequence similarity to SelB and are unable to bind specifically the cognate SECIS RNA, indicating another role than in bacteria [67–69]. Indeed, EFsec co-immunoprecipitated with SBP2 from mammalian cells overexpressing both proteins, in an RNA-dependent complex [68]. The RNA is in fact tRNA<sup>Sec</sup>, in the absence of which complex formation between both proteins is impaired [70]. EFsec-SBP2 interaction can occur *in vitro* independently of tRNA<sup>Sec</sup> only with shortened versions of the isolated SBP2 interaction domain of EFsec. In this way, EFsec amino acids involved in the SBP2 interaction could be mapped at the very C-terminal end. Thus, the C-terminal extension of EFsec, and very likely that of MjSelB, makes protein-protein and not RNA-protein interactions. Functional nuclear localization and export signals were mapped in both EFsec and SBP2 [55]. Besides, SBP2 levels and localization were shown to influence EFsec localization, suggesting that the fate of the two proteins could be linked.

An earlier communication reported that a conserved non-Watson-Crick base pair in the tRNA<sup>Sec</sup> amino acid acceptor arm is critical for binding to the specialized elongation factor [71]. However, the structural determinants required for EFsec binding to tRNA<sup>Sec</sup> have not been investigated in detail yet. In contrast, a larger body of structural studies were carried out with the bacterial and archaeal SelB. Determination of the crystal structure of the complex between the bacterial SECIS and SelB revealed the existence of a winged-helix (WH) domain in SelB, a motif usually found in DNA binding proteins and discovered recently in RNA binding proteins [72]. This structure is the first example of a complex between an RNA and a winged-helix domain. A new mode of RNA recognition enabling the complex to wrap around the small ribosomal subunit was proposed by the authors. Another group solved the crystal structure of the SelB factor from the archeon *M. maripaludis* [13]. The global shape of the protein resembles a chalice observed so far only for the translation initiation factors IF2/eIF5B. This raises the interesting issue that mechanistic similarities may exist between selenocysteine incorporation and initiation of translation. Besides this evolutionary aspect, knowledge of the protein structure allowed identification of important amino acids. In the aminoacyl-binding pocket, two positively charged amino acids, an arginine and an histidine, replace the EF1-A asparagine and aspartic acid residues, presumably to



compensate for the negatively charged selenium. In the same region, a phenylalanine (histidine in EF1-A) protrudes from another domain and it was suggested that this hydrophobic residue could serve as a lid to protect the highly reactive selenocysteine selenol from oxidation. Another feature of SelB factors is the necessity to contact the extended 13 bp long amino acid acceptor arm of tRNAs<sup>Sec</sup>. By docking the structure of a modeled tRNA, the authors concluded that an extended loop in the archaeal SelB is able to contact a large area of the tRNA<sup>Sec</sup> 13 bp amino acid acceptor arm. This loop is strictly conserved among archaea and is also present in eukarya and bacteria, but absent in EF-Tu and EF1-A, suggesting a unified tRNA<sup>Sec</sup>-SelB/EFsec recognition pattern. Once again, this interaction principle represents an appealing adaptive evolution of two ligands.

#### 4. How does the ribosome know that UGA is not the end?

This is obviously the burning question in the field. With the available set of data, two groups have recently come up with distinct models describing the steps prior to Sec-tRNA<sup>Sec</sup> delivery to the ribosomal A site [reviewed in 39]. Based on its finding that SBP2 sediments with ribosomes under low-salt conditions, but cannot bind simultaneously the SECIS RNA

[73], one group proposed that a subset of ribosomes with pre-bound SBP2 are somehow determined for selenoprotein mRNA translation (Fig. 5A). To interact with a distant SECIS RNA, SBP2 takes advantage of the ribosome stalling at the UGA codon, the close proximity of the 5' and 3' ends of the mRNA facilitating the folding back of the SECIS RNA in proximity to the UGA codon. The movement of the ribosome-bound SBP2 triggers a conformational change at the ribosomal A site, allowing delivery of the EFsec/Sec-tRNA<sup>Sec</sup>. Ribosomal protein L30 would displace SBP2 from the SECIS RNA to relocate it to its original position on the ribosome. In the other model, SBP2 does not travel with the ribosome [59]. Instead, it binds the SECIS RNA and serves as a platform to recruit the EFsec/Sec-tRNA<sup>Sec</sup> complex, prior to UGA decoding. An approaching ribosome will lead L30 to displace SBP2, the binding of L30 to the SECIS RNA inducing a more closed conformation of the SECIS K-turn. This movement triggers the release of the Sec-tRNA<sup>Sec</sup> and GTP hydrolysis.

How to distinguish the two possibilities? Experimental validation currently suffers from the bitter lack of knowledge whether all the factors of the system are identified, and of an in vitro reconstitution assay recapitulating selenoprotein synthesis. Whereas the precise function of L30 is still unknown and requires its location on the ribosome to be confirmed by higher

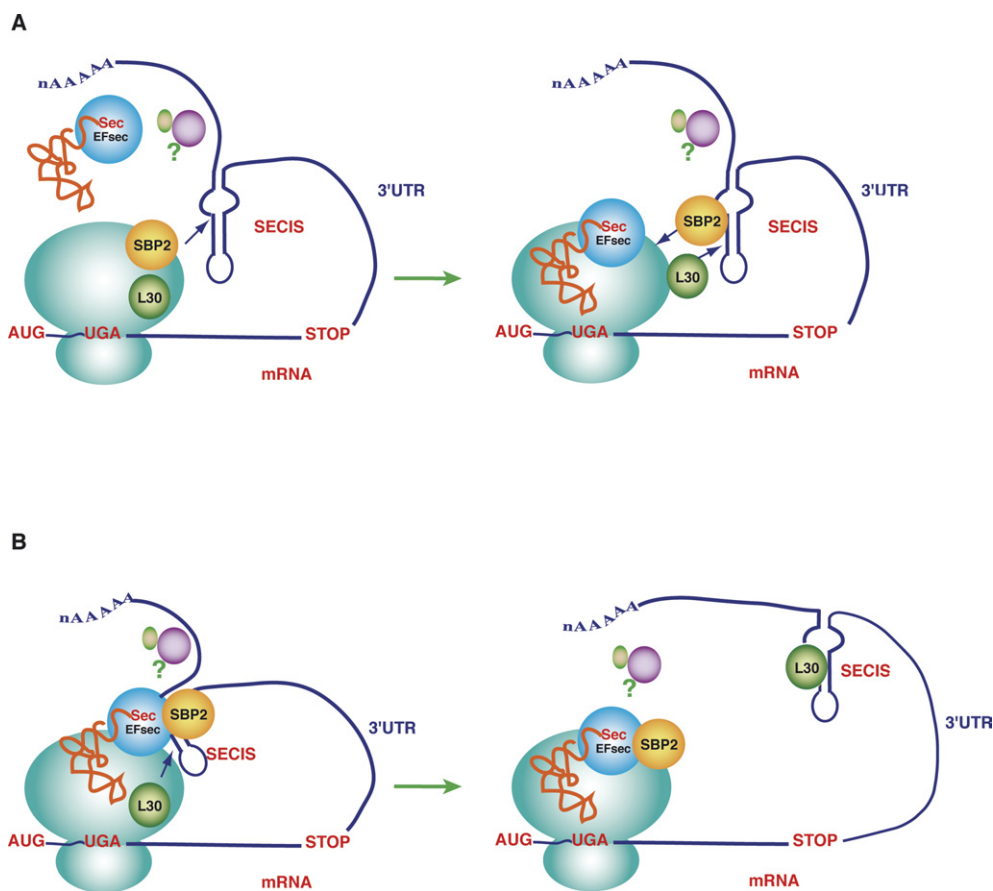


Fig. 5. Current models for selenocysteine incorporation. (A) SBP2 travels with ribosomes, interacts with the SECIS RNA and the EFsec/Sec-tRNA<sup>Sec</sup> to deliver this complex to the A site of the ribosome [73]. L30 displaces the SECIS-bound SBP2. (B) The EFsec/Sec-tRNA<sup>Sec</sup> complex is recruited at the SECIS RNA by SBP2. Ribosome-bound L30 displaces SBP2 [59]. In both models, L30 must leave the SECIS RNA to reset the system. Black arrows indicate factor reshuffling; as yet unidentified factors, possibly involved in the mechanism, are indicated with the question mark.

resolution data, both models converge to propose that this protein displaces SBP2 from the SECIS RNA. The L30-SBP2 reshuffling hypothesis, however, does not consider that SBP2 was found to dissociate very slowly or not at all *in vivo*, once binding to SECIS RNA has taken place [74]. Another central question asks how the SECIS RNA-bound complex folds back at an approaching ribosome and competes with the release factor: no need to say that mechanistic issues are here of prime importance. In this regard, the recent discovery of stem-loop structures different from SECIS RNAs, lying in the open reading frame at the 3' vicinity of the UGA selenocysteine codon in some eukaryal selenoprotein mRNAs, added a missing piece to our knowledge [75]. Such stem-loops were found in some but not all selenoprotein mRNAs and they do not share a conserved secondary structure in the different selenoprotein mRNAs. This observation may legitimately lead to ask whether the function of these stem-loops is pivotal to selenoprotein synthesis. One possibility is their requirement under certain circumstances to favor ribosome pausing in much the same way as stem-loops or pseudoknots contribute to frameshifting.

We are left with the take-home message that a clearer picture of selenoprotein synthesis is popping up but, at the same time, a number of questions remain unsolved. It looks as if a single supramolecular complex could achieve sequentially (or simultaneously) selenocysteine biosynthesis and its incorporation into selenoproteins [32; reviewed in 38]. The cornerstone in this complex is protein SECp43 that establishes RNA-protein and protein-protein contacts with several partners, notably facilitating the interaction between tRNA<sup>Sec</sup>, EFsec and SBP2 *in vivo*. SECp43 thus appears as a key player in orchestrating multiple interactions and redistributing the nucleocytoplasmic localization of other components involved. Taking into account that EFsec and SBP2 undergo nucleocytoplasmic shuttling [55], assembly of this supramolecular complex in the nucleus is an appealing model that could help circumvent nonsense-mediated decay at UGA Sec codons [55]. Another advantage would be to provide ribosomes with SBP2-pre-bound selenoprotein mRNAs, fueling the hypothesis of the existence of a fraction of ribosomes pre-determined for selenoprotein synthesis [37]. Extremely dynamic contacts must exist to enable multiple conformational changes to occur. There is obviously room for future directions of research in this actively evolving field.

## Acknowledgments

Philippe Carbon and Alain Lescure are thanked for careful reading of the manuscript.

## References

[1] D.L. Hatfield, V.N. Gladyshev, How selenium has altered our understanding of the genetic code, *Mol. Cell. Biol.* 22 (2002) 3565–3576.  
 [2] M. Rederstorff, A. Krol, A. Lescure, Understanding the importance of selenium and selenoproteins in muscle function, *Cell. Mol. Life Sci.* 63 (2006) 52–59.

[3] A. Böck, Selenium metabolism in bacteria, in: D.L. Hatfield (Ed.), *Selenium, its molecular biology and role in human health*, Kluwer Academic Publishers, Norwood, MA, 2001, pp. 7–22.  
 [4] C. Baron, E. Westhof, A. Böck, R. Giege, Solution structure of selenocysteine-inserting tRNA<sup>Sec</sup> from *E. coli*. Comparison with canonical tRNA<sup>Ser</sup>, *J. Mol. Biol.* 231 (1993) 274–292.  
 [5] C. Sturchler, E. Westhof, P. Carbon, A. Krol, Unique secondary and tertiary structural features of the eukaryotic selenocysteine tRNA<sup>Sec</sup>, *Nucleic Acids Res.* 21 (1993) 1073–1079.  
 [6] N. Hubert, C. Sturchler, E. Westhof, P. Carbon, A. Krol, The 9/4 secondary structure of eukaryotic selenocysteine tRNA: more pieces of evidence, *RNA* 4 (1998) 1029–1033.  
 [7] X.-Q. Wu, H.J. Gross, The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosphorylation, *EMBO J.* 13 (1994) 241–248.  
 [8] A. Böck, K. Forchhammer, J. Heider, C. Baron, Selenoprotein synthesis: an expansion of the genetic code, *Trends Biochem.* 16 (1991) 463–467.  
 [9] M. Rao, B.A. Carlson, S.V. Novoselov, D.P. Weeks, V.N. Gladyshev, D.L. Hatfield, *Chlamydomonas reinhardtii* selenocysteine tRNA<sup>[Ser]Sec</sup>, *RNA* 9 (2003) 923–930.  
 [10] A.G. Russell, M.N. Schnare, M.W. Gray, Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*, *RNA* 10 (2004) 1034–1046.  
 [11] R.K. Shrimali, A.V. Lobanov, X.-M. Xu, M. Rao, B.A. Carlson, D.C. Mahadeo, C.A. Parent, V.N. Gladyshev, D.L. Hatfield, Selenocysteine tRNA identification in the model organisms *Dictyostelium discoideum* and *Tetrahymena thermophila*, *Biochem. Biophys. Res. Com.* 329 (2005) 147–151.  
 [12] T. Mourier, A. Pain, B. Barrell, S. Griffiths-Jones, A selenocysteine tRNA and SECIS element in *Plasmodium falciparum*, *RNA* 11 (2005) 119–122.  
 [13] M. Leibundgut, C. Frick, M. Thanbichler, A. Böck, N. Ban, Selenocysteine tRNA-specific elongator factor SelB is a structural chimaera of elongation and initiation factors, *EMBO J.* 24 (2005) 11–22.  
 [14] A.V. Lobanov, C. Delgado, S. Rahlfs, S.V. Novoselov, G.V. Kryukov, S. Gromer, D.L. Hatfield, K. Becker, V.N. Gladyshev, The *Plasmodium* selenoproteome, *Nucleic Acids Res.* 34 (2006) 496–505.  
 [15] C. Baron, A. Böck, The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA<sup>Sec</sup> of *E. coli* is the determinant for binding to elongation factors SelB or Tu, *J. Biol. Chem.* 266 (1991) 20375–20379.  
 [16] C. Sturchler-Pierrat, N. Hubert, T. Totsuka, T. Mizutani, P. Carbon, A. Krol, Selenocysteinylation in eukaryotes necessitates the uniquely long aminoacyl acceptor stem of selenocysteine tRNA<sup>Sec</sup>, *J. Biol. Chem.* 270 (1995) 18570–18574.  
 [17] A.M. Diamond, I.-S. Choi, P.F. Crain, T. Hashizume, S.C. Pomerantz, R. Cruz, C.J. Steer, K.E. Hill, R.F. Burk, J.A. McCloskey, D.L. Hatfield, Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA<sup>[Ser]Sec</sup>, *J. Biol. Chem.* 268 (1993) 14215–14223.  
 [18] C. Sturchler, A. Lescure, G. Keith, P. Carbon, A. Krol, Base modification pattern at the wobble position of *Xenopus* selenocysteine tRNA<sup>Sec</sup>, *Nucleic Acids Res.* 22 (1994) 1354–1358.  
 [19] L.K. Kim, T. Matsufuji, S. Matsufuji, B.A. Carlson, S.S. Kim, D.L. Hatfield, B.J. Lee, Methylation of the ribosyl moiety at position 34 of selenocysteine tRNA<sup>[Ser]Sec</sup> is governed by both primary and tertiary structure, *RNA* 6 (2000) 1306–1315.  
 [20] F. Ding, P.J. Grabowski, Identification of a protein component of a mammalian tRNA<sup>Sec</sup> complex implicated in the decoding of UGA as selenocysteine, *RNA* 5 (1999) 1561–1569.  
 [21] X.-M. Xu, H. Mix, B.A. Carlson, P.J. Grabowski, V.N. Gladyshev, M.J. Berry, D.L. Hatfield, Evidence for direct roles of two additional factors, SECp43 and Soluble Liver Antigen, in the selenoprotein synthesis machinery, *J. Biol. Chem.* 280 (2005) 41568–41575.  
 [22] G.J. Warner, M.J. Berry, M.E. Moustafa, B.A. Carlson, D.L. Hatfield, J.R. Faust, Inhibition of selenoprotein synthesis by selenocysteine tRNA<sup>[Ser]Sec</sup> lacking isopentenyladenosine, *J. Biol. Chem.* 275 (2000) 28110–28119.

- [23] B.A. Carlson, X.-M. Xu, V.N. Gladyshev, D.L. Hatfield, Selective rescue of selenoprotein expression in mice lacking a highly specialized methyl group in selenocysteine tRNA, *J. Biol. Chem.* 280 (2005) 5542–5548.
- [24] S.C. Low, J.W. Harney, M.J. Berry, Cloning and functional characterization of human selenophosphate synthetase, an essential component of selenoprotein synthesis, *J. Biol. Chem.* 270 (1995) 21659–21664.
- [25] M.J. Guimaraes, D. Peterson, A. Vicari, B.G. Cocks, N.G. Copeland, D.J. Gilbert, N.A. Jenkins, D.A. Ferrick, R.A. Kastelein, J.F. Bazan, A. Zlotnik, Identification of a novel *selD* homolog from eukaryotes, bacteria and archaea: is there an autoregulatory mechanism in selenocysteine metabolism?, *Proc. Natl. Acad. Sci. USA* 93 (1996) 15086–15091.
- [26] T. Ohama, D.C.H. Yang, D.L. Hatfield, Selenocysteine tRNA and serine tRNA are aminoacylated by the same synthetase, but manifest different identities with respect to the long extra arm, *Arch. Biochem. Biophys.* 315 (1994) 293–301.
- [27] J.T. Kaiser, K. Gromadski, M. Rother, H. Engelhardt, M.V. Rodnina, M.C. Wahl, Structural and functional investigation of a putative archaeal selenocysteine synthase, *Biochem.* 44 (2005) 13315–13327.
- [28] C. Gelpi, E.J. Sontheimer, J.L. Rodriguez-Sanchez, Autoantibodies against a serine tRNA-protein complex implicated in cotranslational selenocysteine insertion, *Proc. Natl. Acad. Sci. USA* 89 (1992) 9739–9743.
- [29] I. Wies, S. Brunner, J. Henninger, J. Herkel, S. Kanzler, K.-H. Meyer zum Büschenfelde, A.W. Lohse, Identification of target antigen for SLA/LP autoantibodies in autoimmune hepatitis, *Lancet* 355 (2000) 1510–1515.
- [30] M. Costa, J.L. Rodriguez-Sanchez, A.J. Czaja, C. Gelpi, Isolation and characterization of cDNA encoding the antigenic protein of the human tRNP<sup>[Ser]<sup>Sec</sup></sup> complex recognized by autoantibodies from patients with type-1 autoimmune hepatitis, *Clin. Exp. Immunol.* 121 (2000) 364–374.
- [31] T. Kernebeck, A.W. Lohse, J. Grötzing, A bioinformatical approach suggests the function of the autoimmune hepatitis target antigen Soluble Liver Antigen/Liver Pancreas, *Hepatology* 34 (2001) 230–233.
- [32] A. Small-Howard, N. Morozova, Z. Stoytcheva, E.P. Forry, J.B. Mansell, J.W. Harney, B.A. Carlson, X.-M. Xu, D.L. Hatfield, M.J. Berry, Supramolecular complexes mediate selenocysteine incorporation in vivo, *Mol. Cell. Biol.* 26 (2006) 2337–2346.
- [33] P. Tormay, R. Wilting, F. Lottspeich, P.K. Mehta, P. Christen, A. Böck, Bacterial selenocysteine synthase, structural and functional properties, *Eur. J. Biochem.* 254 (1998) 655–661.
- [34] B.A. Carlson, X.-M. Xu, G.V. Kryukov, M. Rao, M.J. Berry, V.N. Gladyshev, D.L. Hatfield, Identification and characterization of phosphoserine-tRNA<sup>[Ser]<sup>Sec</sup></sup> kinase, *Proc. Natl. Acad. Sci. USA* 101 (2004) 12848–12853.
- [35] A. Sauerwald, W. Zhu, T.A. Major, H. Roy, S. Palioura, D. Jahn, W.B. Whitman, J.R. Yates 3rd, M. Ibba, D. Söll, RNA-dependent cysteine biosynthesis in archaea, *Science* 307 (2005) 1969–1972.
- [36] A. Krol, Evolutionary different RNA motifs and RNA-protein complexes to achieve selenoprotein synthesis, *Biochimie* 84 (2002) 765–774.
- [37] D.M. Driscoll, P.R. Copeland, Mechanism and regulation of selenoprotein synthesis, *Annu. Rev. Nutr.* 23 (2003) 17–40.
- [38] A.L. Small-Howard, M.J. Berry, Unique features of selenocysteine incorporation function within the context of general eukaryotic translational processes, *Biochem. Soc.* 33 (2005) 1493–1497.
- [39] K. Caban, P.R. Copeland, Size matters: a view of selenocysteine incorporation from the ribosome, *Cell. Mol. Life Sci.* 63 (2006) 73–81.
- [40] R. Walczak, E. Westhof, P. Carbon, A. Krol, A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs, *RNA* 2 (1996) 367–379.
- [41] R. Walczak, P. Carbon, A. Krol, An essential non-Watson-Crick base pair motif in 3'UTR to mediate selenoprotein translation, *RNA* 4 (1998) 74–84.
- [42] G.W. Martin IV, J.W. Harney, M.J. Berry, Functionality of mutations at conserved nucleotides in eukaryotic SECIS elements is determined by the identity of a single nonconserved nucleotide, *RNA* 4 (1998) 65–73.
- [43] C. Allmang, A. Krol, SECIS RNAs and K-turn binding proteins. A survey of evolutionary conserved RNA and protein motifs, in *Selenium, its molecular biology and role in human health* 2<sup>nd</sup> edition, M.J. Berry, V.N. Gladyshev, D.L. Hatfield (eds), Kluwer Academic Publishers, 2006, in press.
- [44] D.J. Klein, T.M. Schmeing, P.B. Moore, T.A. Steitz, The kink-turn, a new RNA secondary structure motif, *EMBO J.* 20 (2001) 4214–4221.
- [45] N.B. Leontis, E. Westhof, Geometric nomenclature and classification of RNA base pairs, *RNA* 7 (2001) 499–512.
- [46] E. Grundner-Culemann, G.W. Martin III, J.W. Harney, M.J. Berry, Two distinct SECIS RNA structures capable of directing selenocysteine incorporation in eukaryotes, *RNA* 5 (1999) 625–635.
- [47] D. Fagegaltier, A. Lescure, R. Walczak, P. Carbon, A. Krol, Structural analysis of new local features in SECIS RNA hairpins, *Nucleic Acids Res.* 28 (2000) 2679–2689.
- [48] A. Ramos, A.N. Lane, D. Hollingworth, T.W.-N. Fan, Secondary structure and stability of the selenocysteine insertion sequences (SECIS) for human thioredoxin reductase and glutathione peroxidase, *Nucleic Acids Res.* 32 (2004) 1746–1755.
- [49] R. Wilting, S. Schorling, B.C. Persson, A. Böck, Selenoprotein synthesis in archaea: identification of an mRNA element of *Methanococcus jannaschii* probably directing selenocysteine insertion, *J. Mol. Biol.* 266 (1997) 637–641.
- [50] M. Rother, A. Resch, W.L. Gardner, W.B. Whitman, A. Böck, Heterologous expression of archaeal selenoprotein genes directed by the SECIS element located in the 3' non-translated region, *Mol. Microbiol.* 40 (2001) 900–908.
- [51] G.V. Kryukov, V.N. Gladyshev, The prokaryotic selenoproteome, *EMBO Rep.* 5 (2004) 538–543.
- [52] P.R. Copeland, J.E. Fletcher, B.A. Carlson, D.L. Hatfield, D.M. Driscoll, A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs, *EMBO J.* 19 (2000) 306–314.
- [53] A. Lescure, C. Allmang, K. Yamada, P. Carbon, A. Krol, cDNA cloning, expression pattern and RNA binding analysis of human selenocysteine insertion sequence (SECIS) binding protein 2, *Gene* 291 (2002) 279–285.
- [54] A.M. Dumitrescu, X.-H. Liao, M.S.Y. Abdullah, J. Lado-Abeal, F. Abdul Majed, L.C. Moeller, G. Boran, L. Schomburg, R.E. Weiss, S. Refetoff, Mutations in SECISBP2 result in abnormal thyroid hormone metabolism, *Nat. Genet.* 37 (2005) 1247–1252.
- [55] L.A. de Jesus, P.R. Hoffmann, T. Michaud, E.P. Forry, A. Small-Howard, R.J. Stillwell, N. Morozova, J.W. Harney, M.J. Berry, Nuclear assembly of UGA decoding complexes on selenoprotein mRNAs: a mechanism for eluding nonsense-mediated decay?, *Mol. Cell. Biol.* 26 (2006) 1795–1805.
- [56] C. Allmang, P. Carbon, A. Krol, The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding, *RNA* 8 (2002) 1308–1318.
- [57] J.E. Fletcher, P.R. Copeland, D.M. Driscoll, A. Krol, The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2, *RNA* 7 (2001) 1442–1453.
- [58] V. Allamand, P. Richard, A. Lescure, C. Ledeuil, D. Desjardin, N. Petit, C. Gartioux, A. Ferreira, A. Krol, N. Pellegrini, J. Andoni Urtizberea, P. Guicheney, A single homozygous point mutation in a 3' untranslated region motif of selenoprotein N mRNA causes SEPNI-related myopathy, *EMBO Rep.* 7 (2006) 450–454.
- [59] L. Chavatte, B.A. Brown, D.M. Driscoll, Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes, *Nat. Struct. Mol. Biol.* 12 (2005) 408–416.
- [60] M. Halic, T. Becker, J. Frank, C.M.T. Spahn, R. Beckmann, Localization and dynamic behavior of ribosomal protein L30e, *Nat. Struct. Mol. Biol.* 12 (2005) 467–468.
- [61] Q. Shen, P.A. McQuikin, P.E. Newburger, RNA-binding proteins that specifically recognize the selenocysteine insertion sequence of human cellular glutathione peroxidase mRNA, *J. Biol. Chem.* 270 (1995) 30448–30452.
- [62] N. Hubert, R. Walczak, P. Carbon, A. Krol, A protein binds the selenocysteine insertion element in the 3'UTR of mammalian selenoprotein mRNAs, *Nucleic Acids Res.* 24 (1996) 464–469.

- [63] T. Fujiwara, K. Busch, H.J. Gross, T. Mizutani, A SECIS binding protein (SBP) is distinct from selenocysteyl-tRNA protecting factor (SePF), *Biochimie* 81 (2000) 213–218.
- [64] Q. Shen, R. Wu, J.L. Leonard, P.E. Newburger, Identification and molecular cloning of a human selenocysteine insertion sequence-binding protein. A bifunctional role for DNA-binding protein B, *J. Biol. Chem.* 273 (1998) 5443–5446.
- [65] D. Fagegaltier, N. Hubert, P. Carbon, A. Krol, The selenocysteine insertion sequence binding protein SBP is different from the Y-box protein dbpB, *Biochimie* 82 (2000) 117–122.
- [66] Q. Shen, L. Fan, P.E. Newburger, Nuclease sensitive element binding protein 1 associates with the selenocysteine insertion sequence and functions in mammalian selenoprotein translation, *J. Cell. Physiol.* 207 (2006) 775–783.
- [67] M. Rother, R. Wilting, S. Commans, A. Böck, Identification and characterization of the selenocysteine-specific translation factor SelB from the archeon *Methanococcus jannaschii*, *J. Mol. Biol.* 299 (2000) 351–358.
- [68] R.M. Tujebajeva, P.R. Copeland, X.-M. Xu, B.A. Carlson, J.W. Harney, D.M. Driscoll, D.L. Hatfield, M.J. Berry, Decoding apparatus for eukaryotic selenocysteine incorporation, *EMBO Rep.* 2 (2000) 158–163.
- [69] D. Fagegaltier, N. Hubert, K. Yamada, T. Mizutani, P. Carbon, A. Krol, Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation, *EMBO J.* 19 (2000) 4796–4805.
- [70] A.M. Zavacki, J.B. Mansell, M. Chung, B. Klimovitsky, J.W. Harney, M.J. Berry, Coupled tRNA<sup>Sec</sup>-dependent assembly of the selenocysteine decoding apparatus, *Mol. Cell* 11 (2003) 773–781.
- [71] T. Mizutani, K. Tanabe, K. Yamada, A G.U base pair in the eukaryotic selenocysteine tRNA is important for interaction with SePF, the putative selenocysteine-specific elongation factor, *FEBS Lett.* 429 (1998) 189–193.
- [72] S. Yoshizawa, L. Rasubala, T. Ose, D. Kohda, D. Fourmy, K. Maenaka, Structural basis for mRNA recognition by elongation factor SelB, *Nat. Struct. Mol. Biol.* 12 (2005) 198–203.
- [73] S.A. Kinzy, K. Caban, P.R. Copeland, Characterization of the SECIS binding protein 2 complex required for the co-translational insertion of selenocysteine in mammals, *Nucleic Acids Res.* 33 (2005) 5172–5180.
- [74] S.C. Low, E. Grundner-Culemann, J.W. Harney, M.J. Berry, SECIS-SBP2 interactions dictate selenocysteine incorporation efficiency and selenoprotein hierarchy, *EMBO J.* 19 (2000) 6882–6890.
- [75] M.T. Howard, G. Aggarwal, C.B. Anderson, S. Khatri, K.M. Flanigan, J.F. Atkins, Recoding elements located adjacent to a subset of eukaryal selenocysteine-specifying UGA codons, *EMBO J.* 24 (2005) 1596–1607.



## **Chapter 5. SECIS RNAs and K-turn binding proteins. A survey of evolutionary conserved RNA and protein motifs**

Christine Allmang and Alain Krol

*Architecture et Réactivité de l'arN. UPR 9002 du CNRS-Université Louis Pasteur. Institut de Biologie Moléculaire et Cellulaire. 67084 Strasbourg, France*

**Summary:** The SelenoCysteine Insertion Sequences (SECIS) is a stem-loop structure residing in the 3' untranslated region of all selenoprotein mRNAs. Its presence is mandatory to allow the ribosome to readthrough the UGA selenocysteine codon. The SECIS RNA possesses a well-defined secondary structure. Four consecutive non-Watson-Crick base pairs, with a central tandem of sheared G.A/A.G base pairs, constitute the functional motif of the SECIS RNA which is recognized by the SECIS binding protein SBP2. The tandem of sheared base pairs is part of a recurrent motif, the kink-turn (K-turn), occurring in a variety of different RNAs. The K-turn is a helix-internal loop-helix composed of a non-Watson-Crick stem containing the G.A base pairs and a canonical stem. The internal loop between the stems is always asymmetrical and usually contains three unpaired nucleotides on one strand and none on the other. We propose here that the SECIS RNA must represent a K-turn variant with regard to the limited structural differences that distinguish it from consensus K-turns. Work by others showed that ribosomal protein L30 also binds the SECIS RNA in a specific manner. L30 and SBP2 are members of a family of proteins sharing the same RNA-binding domain called L7A/L30. All proteins possessing this fold recognize K-turn RNAs. Three structures of RNA-protein complexes containing the L7A/L30 protein fold and cognate K-turn RNAs have been solved. In light of the interaction principles governing these RNA-protein complexes, we discuss how L30 can recognize the SECIS RNA. Collectively, all the findings suggest that the L7A/L30 protein fold and the K-turn are ancient structural motifs that have evolved various functions, from pre-mRNA splicing to protein synthesis.

### **Introduction**

The field of eukaryotic selenoprotein research is fascinating in several respects. First, the existence of taxa-specific selenoproteins altered the initial perception that mammals recapitulate the eukaryotic selenoproteome. Second, it becomes increasingly apparent that the number of molecular

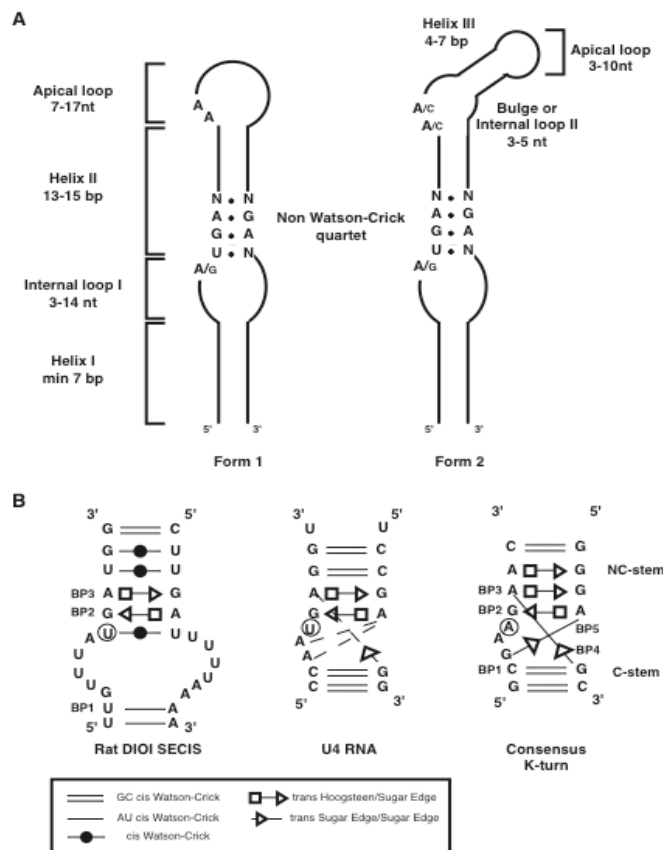
partners involved in selenoprotein synthesis is larger than previously thought. Relocation of the SECIS element, from the coding frame in bacteria to the 3'-untranslated region (3'UTR) of selenoprotein mRNAs in eukaryotes, may be responsible only in part for this complexification. Indeed, even selenocysteine biosynthesis itself seems to take a more sophisticated pathway in eukaryotes.

The SECIS stem-loop contains a well defined structural motif composed of four consecutive non-Watson-Crick base pairs, with a central tandem of sheared G:A base pairs [1]. This motif (Figure 1A) also ensures a functional role as it is essential to selenocysteine incorporation in vivo [2,3] and constitutes the binding site of SBP2, a protein binding specifically to the SECIS RNA [4,5]. The SBP2 RNA-binding domain contains a region sharing a high degree of amino acid sequence similarity to the L7A/L30 protein family containing ribosomal proteins L7Ae and L30, the 15.5kD/Snu13p spliceosomal protein and other functionally unrelated proteins [6,7]. Cocystal structures of the L7Ae, L30 and 15.5 kD proteins in complex with their cognate RNAs revealed that the proteins fold into a highly conserved compact globular domain, the L7A/L30 domain, that binds specifically to RNAs possessing a kink-turn (K-turn) motif. The canonical K-turn is a recurrent element, occurring notably in ribosomal RNAs, U4 snRNA, and box C/D regions of snoRNAs and archaeal sRNAs. It contains a tandem of sheared G:A/A:G base pairs that have an important structural role in forming and stabilizing the turn [8]. In earlier studies, we proposed a 3D model for the SECIS RNA where the phosphodiester backbone is bent at the non-Watson-Crick base pairs [1]. Combined with the presence of sheared base pairs, the proposed folding of the SECIS RNA suggests that it could be a canonical K-turn or a K-turn related RNA. From all these findings emerges the important issue of how different RNAs harboring K-turn motifs can selectively discriminate proteins sharing the same RNA-binding domain. This is a particularly burning question in light of the finding that ribosomal protein L30 is another SECIS-binding protein [9]. This chapter will describe the SECIS RNA structure with comparison to canonical K-turn RNAs, then highlight the similarities/differences between protein-RNA complexes formed with proteins of the L7A/L30 family and K-turn RNAs.

#### **The SECIS RNA: a K-turn variant**

An experimental secondary structure model for SECIS RNAs (Figure 1A) was proposed about ten years ago based on structure probing in solution [1]. It was next discovered that certain SECIS RNAs can adopt a slightly different 2D structure at their apex [10]. Called Form 2 SECIS (Figure 1A), they possess an additional helix III but a shorter apical loop, compared to Form 1 SECIS. Besides the non-Watson-Crick quartet, Form 2 shares with Form 1 the other conserved features characterizing SECIS RNAs, i.e. the run

of As in the apical (Form 1) or internal loop II (Form 2) and the 13-15 bp long helix II. More systematic identification of a variety of novel selenoprotein mRNAs including vertebrates, invertebrates and green algae [11-20] clearly established that Form 2 SECIS are more widespread than Form 1. However swapping experiments could not establish that Form 2, although preponderant, provides a functional advantage to selenocysteine incorporation [10].



**Figure 1.** Structure models for the SECIS RNA. (A) Secondary structure models of Forms 1 and 2 SECIS. The conserved sequence and structural features are indicated. N, any nucleotide; A/G and A/C indicate that A is the prevalent base. (B) Secondary structure diagrams of the U4 snRNA and consensus K-turns adapted from [22,26] and the putative SECIS K-turn of the rat type I iodothyronine deiodinase (DIOI). BP1 to BP5 stands for base pairs 1 to 5. Circled bases are discussed in the text. NC-stem: non-Watson-Crick stem; C-stem: Watson-Crick stem. Broken lines in U4 snRNA stand for hydrogen bonds between N6A30 and 2'OH of A44, N1A44 and 2'OH of A29 [22]; the latter interaction differs from that proposed at the homologous position in the consensus K-turn. The graphical conventions for displaying non-Watson-Crick base pairs are from [28].



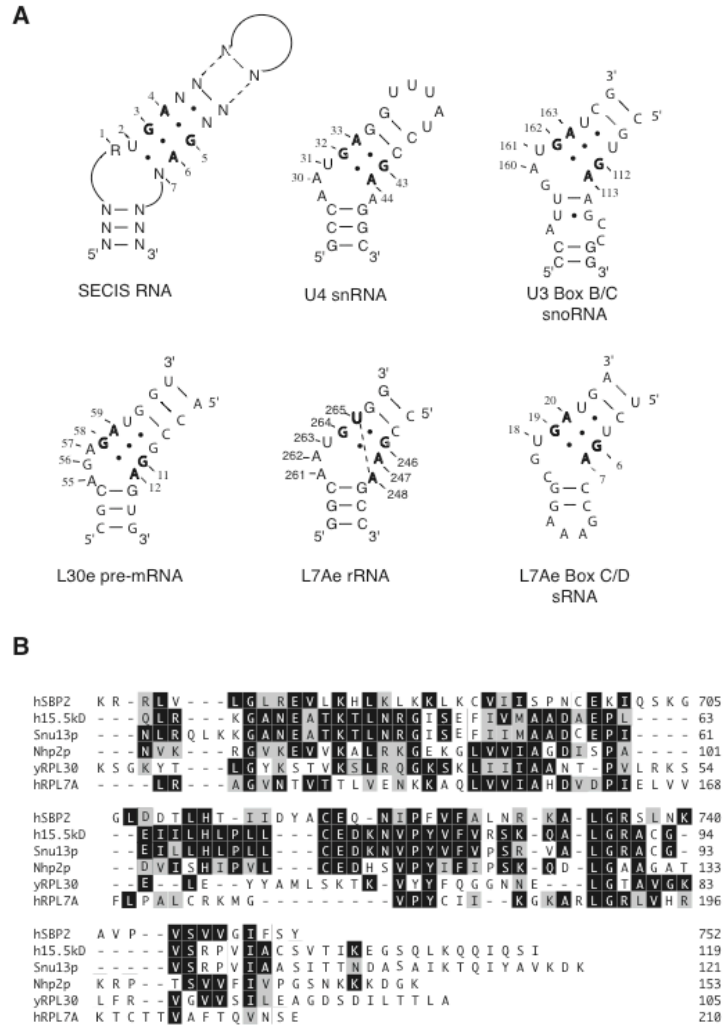
It is remarkable that mRNAs encoding the same selenoprotein can harbor either a Form 1 or a Form 2 SECIS, depending on the animal species. This is well exemplified by the SelM mRNA where Form 2 occurs in mammals while zebrafish harbors Form 1 SECIS [14].

Remarkably, chemical probing experiments indicated that the conserved As are single stranded and well accessible whatever the SECIS form [21]. In a few cases, especially in the mammalian SelM and some *Chlamydomonas* Form 2 SECIS, Cs are found instead of As without apparently altering the SECIS function [14,15]. Thus the universal conservation of the As, which was taken for granted at the time when the number of available SECIS sequences was too little to make statistically valid comparisons, is called into question. The mechanistic role of these unpaired A/Cs is still unknown, but their functional importance has been experimentally proven *in vivo* by site-directed mutagenesis. Along the same lines, the nucleotide 5' to the non-Watson-Crick quartet is A in the vast majority of SECIS RNAs. However, compilation of selenoprotein mRNA sequences in other organisms led to the conclusion that G can sometimes be found instead, an interesting example being provided by the single selenoprotein mRNA in nematodes [11,20,21]. This correlates with *in vivo* experiments and mobility shift assays with SBP2 and SECIS RNAs concluding that an A is preferred but not mandatory [5,21]. In conclusion, it emerges from phylogenetic studies that SECIS RNAs exhibit a remarkable conservation of the 2D structures but few invariant nucleotides. Clearly, elucidation of the function of the single stranded A/C and conserved length of helix II is a necessary step toward an in-depth understanding of the function of the SECIS RNA.

The non-Watson-Crick quartet at the foot of helix II is a characteristic feature recognized by SBP2 (Figure 1B). The central G.A tandem was shown by structure probing experiments and computer modeling to adopt the sheared geometry [1]. Tandem sheared base pairs were initially discovered in the crystal structure of the 5' stem-loop of U4 snRNA in complex with the 15.5 kD protein [22]. The prevalence of this RNA motif was in fact revealed by the analysis of the atomic structures of the large and small ribosomal subunits where it occurs six times in *H.marismortui* 23S rRNA and twice in *T.thermophilus* 16S rRNA [8]. Its presence was further identified in the crystal structures of three other RNA-protein complexes: the yeast ribosomal protein L30e with its pre-mRNA, and the archaeal ribosomal protein L7Ae in complex with box C/D or box H/ACA sRNAs [23-25]. A two-dimensional representation of the tertiary structure of a consensus K-turn is diagrammed in Figure 1B, which was adapted from [26]. In this publication, the authors derived the consensus from examination of K-turns in crystal structures and compared them with the sequence alignments of rRNAs from the three kingdoms of life. The K-turn is a two-stranded, helix-internal loop-helix

motif comprising about 15 nucleotides. The first stem (canonical or C-stem) ends at the internal loop with two Watson-Crick base pairs, mostly G-C. The non-canonical stem (NC-stem) starts typically with the sheared G.A base pairs. The internal loop is always asymmetrical with three unpaired nucleotides on one strand and none on the other. Because of the cross-strand stacking of the sheared base pairs, a sharp bend of the sugar-phosphate backbone occurs between the C and NC-stems. Five base pairs characterize the K-turn motif (Figure 1B): the Watson-Crick C-G base pair 1, the sheared G.A base pairs 2 and 3, the triple A.C-G base pair 4, and G.A base pair 5. The adenine of base pair 4 mediates the minor groove interaction with the C-stem (A-minor motif; see reference 27) and is crucial for K-turn folding. Figure 1B shows the 2D structure model of the non-Watson-Crick quartet of the rat type I iodothyronine deiodinase (DIO1) SECIS RNA [1] compared to the structure of the U4 snRNA K-turn motif adapted from [22]. Visual inspection of the SECIS 2D structure identified an important K-turn characteristic feature: the C-stem separated by an internal loop from the NC-stem comprising the invariant sheared base pairs. Despite the similarity, a few SECIS specific structural features led us to ask whether they form genuine K-turns. The non-Watson-Crick U.U base pair 3' to the sheared base pairs will not be discussed further because it displays sequence variation in SECIS and other RNAs and does not participate directly in the K-turn interactions [26]. The first question concerns the U residue (circled in Figure 1B) 5' to the sheared base pair 2. Chemical probing experiments detected that it forms a non-Watson-Crick U.U base pair in the naked SECIS RNA [1]. In contrast, the homologous position is unpaired in U4 snRNA and in the consensus K-turn (Figure 1B; see also Figure 2B). Moreover, data from crystal structures of RNA-protein complexes showed that the base at this position protrudes away from the RNA chain and is tightly bound in a pocket of the protein [22-25]. However, one cannot exclude the possibility that an SBP2-promoted induced fit leads to unpairing and flipping out of the U residue. It could thus be the positional analog of the protruding base in the other K-turns (Figures 1B and 2B). The second question asks whether the counterparts to base pairs 4 and 5 of the consensus K-turn also exist in the SECIS RNA as chemical probing cannot detect them. Formation of base pair 4 in the SECIS RNA will only depend on the sequence of base pair 1 since A is invariant in base pair 3 (Figure 1B). Base pair 1 is U-A in the SECIS RNA shown, very often C-G and G-C but rarely A-U or G.U in others [1,11-21]. Interestingly, tables of sequence variation in [26] show the prevalence of C-G, G-C or U-A at base pair 1 in K-turns, indicating that formation of base pair 4 is theoretically possible in SECIS RNAs. Likewise, base pair 5 could form in SECIS RNAs as tables in [26] established that base pairing is permitted in canonical K-turns between the invariant A of base pair 2 and

any nucleotide. Lastly, one could argue that the size of internal loop I of SECIS RNAs is larger than in canonical K-turns.



**Figure 2.** K-turn motifs and amino acid sequence alignments of L7A/L30 RNA-binding domains. (A) The secondary structures of the U4 snRNA, L30e pre-mRNA, L7Ae rRNA, L7Ae box C/D sRNA were taken from the crystal structures of the corresponding RNA-protein complexes, those of the SECIS RNA and U3 snoRNA result from structure probing (see text). In bold are the sheared G.A base pairs. Numbering is from the original publications, except that of the SECIS RNA which is arbitrary. The dotted line between A248 and U265 in L7Ae rRNA represents the hydrogen bond giving rise to the base triple A.U.G [30]. (B) Amino acid sequence alignment of L7A/L30 proteins. hSBP2, human SBP2; h15.5 kD, human 15.5 kD; Snu13p, the yeast 15.5 kD ortholog; Nhp2p, the yeast core protein of box H/ACA snoRNPs; yRPL30, yeast ribosomal protein L30; hRPL7A, human ribosomal protein L7A. Identical and similar amino acids are displayed in black and gray backgrounds, respectively.

However, the structures in [26] showed that the increased length of the 5' and 3' strands in loop I versus the consensus K-turn is not a major obstacle to K-turn formation as variable lengths do exist in the 5' strands of various K-turns. Regarding the 3' strand, examination of the K-turn crystal structures pointed to the possibility of accommodating its extra length.

In conclusion, we propose that the core of SECIS RNA is a K-turn like motif where the bend occurs at the internal loop I, providing less structural constraint than canonical K-turns. As a consequence, the SECIS RNA must be endowed with a greater flexibility enabling it to switch easily from an open to a closed kinked conformation, thus triggering a major conformational change of the SBP2 bound complex [9].

### **A phylogenetically conserved RNA-protein interface at work for selenoprotein synthesis**

Proteins containing the L7A/L30 RNA-binding domain include ribosomal proteins L7A (L7Ae in Archaea) and L30, human 15.5kD (Snu13p in yeast) in box H/ACA snoRNPs. Archaea contain neither 15.5 kD nor Nhp2p, L7Ae being the surrogate in box C/D and box H/ACA sRNPs. Crystal structures attested to the presence of a K-turn motif in the yeast L30e pre-mRNA, L7Ae rRNA and box C/D sRNA, in addition to U4 snRNA discussed above (Figure 2A).

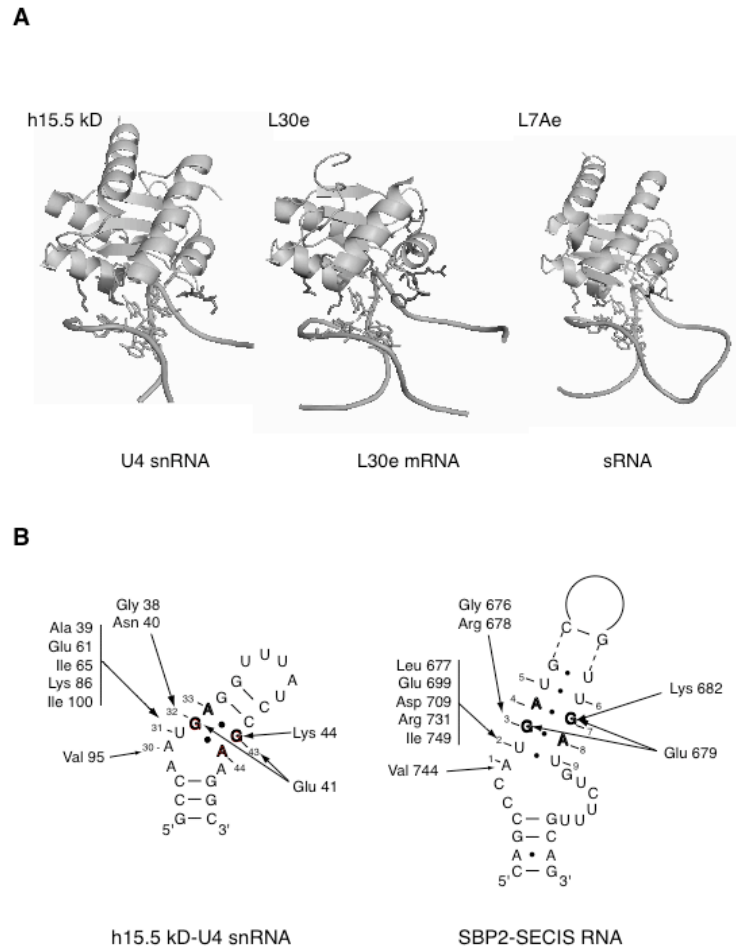
In this series, the only K-turn sequence variant is the L7Ae rRNA where U.G substitutes for the A.G (top) base pair. U.G can nevertheless form base pair 4 described in Figure 1B [26,30]. U3 snoRNA contains one B/C and one C'/D box instead of the classical box C/D, both recognized by 15.5 kD/Snu13p; the B/C box structure shown was derived from probing experiments [31]. Figure 3A shows views of the crystal structures of the 15.5 kD-U4 snRNA, L30e-pre-mRNA and L7Ae-box C/D sRNA complexes, adapted from [22-24]. A detailed description of the RNA-protein contacts fall beyond the scope of this review. Inspection of Figure 3A, however, reveals that the three structures form analogous protein-RNA interfaces despite the differential orientation of some helices (compare for example the bottom right helix in L30e with the proteins in the other two complexes).

The interface is provided by the flipped-out bases U31 (U4 snRNA), U263 and U18 (rRNA and sRNA), A57 (L30e pre-mRNA) protruding into an electrostatically neutral pocket of the cognate protein, and by a few amino acids that make base-specific contacts with the guanines of the sheared G.A base pairs. Yet differences can be found. For example, binding of 15.5 kD is highly susceptible to mutations of U31 while changing A57 and U18 is less deleterious to L30e and L7Ae interaction [22-24]; the angle of the kink shows subtle variations in each complex; finally, it is worth noting that L30e and the pre-mRNA interact through a mutually induced fit [32] whereas only

the RNA component (U4 snRNA or sRNA) undergoes an induced fit upon binding to 15.5 kD and L7Ae, respectively [31, 33-37].

Earlier work localized the SBP2 RNA-binding domain in a region lying approximately between positions 500 and 750 [4,7,29, and our unpublished results]. Within this area, database searches [4,6,7] identified a subdomain homologous to the L7A/L30 RNA-binding domain (Figure 2B), SBP2 and 15.5 kD/Snu13p sharing the highest amino acid sequence similarity [7]. The RNA-binding domain of SBP2 is thus bipartite, composed of the conserved L7A/L30 module flanked by SBP2-specific sequences. A structure-guided strategy, based on the similarities between SBP2/15.5 kD and SECIS RNA/U4 snRNA, and the crystal structure of the 15.5 kD-U4 snRNA complex, predicted SBP2 amino acids that should contact the SECIS RNA [7]. Changing them to alanines led to the identification of eight amino acids critical for SECIS binding, four of them being crucial: Gly676 and Glu679 are postulated to contact the guanines of the sheared base pairs, Glu699 and Arg731 being very likely part of the pocket accommodating the SECIS RNA U2 (Figure 3B). These findings established that the recognition principles governing the 15.5 kD-U4 snRNA interaction must be similar in the SBP2-SECIS RNA complex especially at the guanines of the G.A base pairs and at U2 (SECIS RNA) and U31 (U4 snRNA). Another member of the L7A/L30 family, the rat ribosomal protein L30, was recently shown to be a novel SECIS-binding protein [9]. As a follow-up, determination of the molecular basis underlying this interaction would be instructive in particular to understand how L30 can recognize the SECIS K-turn and compete with SBP2. In the absence of a structural model though, comparison of the structures of the L30e, 15.5 kD, L7Ae and SBP2 RNA-protein complexes [7, 22-25, 37] provided some clues that may explain the L30 versatility. In all of the complexes, mutations of the bases comprising the sheared G.A base pairs resulted in the complete loss of protein binding *in vitro*. Together with the high amino acid sequence similarity between yeast L30e and rat L30, these findings strongly suggest that rat L30 in complex with the SECIS RNA should also interact at the G.A tandem of the SECIS RNA, most likely at the guanines. An interesting difference between 15.5 kD and SBP2 on the one hand, and L7Ae and L30e on the other, occurs at the flipped-out base. In the 15.5 kD and SBP2 complexes, mutations of U31 and U2 to any nucleotide is detrimental to binding *in vitro* and function *in vivo* (for the SECIS RNA). In contrast, L7Ae can accommodate a C instead of U18 and there is little sequence preference at A57 for the binding *in vitro* of L30e which can tolerate G or even C [38]. Remarkably, a correlation can be made at the protein level. In the 15.5kD-U4 snRNA and SBP2-SECIS RNA complexes, five (almost) identical amino acids contact U31 and probably U2, respectively (Figure 3B). Instead, only two L7Ae amino acids make base specific contacts with U18, L30e showing the simplest interaction scheme

with one single contact between A57 and Asn47 (or Asn74, depending on whether the NMR or X-ray structures are considered). Given that G or C may substitute for A57, it is conceivable that U can also fit.



**Figure 3.** RNA-protein interfaces at various L7A/L30 protein-K-turn RNA complexes. (A) Overall crystal structures of the human h15.5 kD-U4 snRNA, L30e-mRNA and L7Ae-box C/D sRNA complexes adapted from [22-24]. The ribbon plots of the proteins with the bound RNA fragments are shown. Figures were generated with PyMOL in an orientation expliciting structural similarities. (B) Similar interaction principles govern the 15.5 kD-U4 snRNA and SBP2-SECIS RNA complexes [7].

Taking into account that a single contact forms between L30e and A57, we propose that the SECIS RNA U2 could also hydrogen bond with L30e Asn47 or Asn74 upon repositioning of the base to offer the appropriate hydrogen bond donor and acceptor groups.

As rat L30 binds the SECIS RNA, we assayed other L7A/L30 proteins for their abilities to recognize the SECIS RNA. Snu13p and L7Ae indeed bound the SECIS RNA but the reverse did not happen since SBP2 was unable to interact with U4 snRNA or an L7Ae RNA target (A.Cléry, C.A, A.K and C. Branlant, manuscript in preparation). We concluded from this experiment that the SBP2 RNA-binding domain is more complex than in the other proteins of the family, the SECIS RNA binding specificity being very likely provided by amino acids flanking the L7A/L30 subdomain. In fact, our unpublished data support this hypothesis.

The analogous protein-RNA interface formed between L7A/L30 proteins and various K-turn RNAs suggests a conformational adaptability of the RNA upon binding to its cognate protein. Such a dynamic process could potentially confer the binding specificity for different K-turns, as exemplified by rat L30. This adaptability could be facilitated by the dimorphism of K-turn RNAs that are in dynamic equilibrium between a tightly kinked-turn and a more open structure [39]. A bunch of proteins of the L7A/L30 family and a large number of diverse RNAs containing the K-turn motif have been identified, with the majority of the K-turns residing in the small and large ribosomal RNAs [8]. Altogether, these findings suggest that the L7A/L30 fold and the K-turn are ancient structural motifs that have evolved specialized roles in many different biological processes.

## References

1. R Walczak, E Westhof, P Carbon, A Krol 1996 *RNA* 2:367
2. R Walczak, P Carbon, A Krol 1998 *RNA* 4:74
3. GW Martin III, JW Harney, MJ Berry 1998 *RNA* 4:65
4. PR Copeland, JE Fletcher, BA Carlson, DL Hatfield, DM Driscoll 2000 *EMBO J* 19:306
5. JE Fletcher, PR Copeland, DM Driscoll, A Krol 2001 *RNA* 7:1442
6. PR Copeland, DM Driscoll 2001 *Selenium: Its Molecular Biology and Role in Human Health* DL Hatfield (Ed) Kluwer Academic Publishers Boston/Dordrecht/London pp 55
7. C Allmang, P Carbon, A Krol 2002 *RNA* 8:1308
8. DJ Klein, TM Schmeing, PB Moore, TA Steitz 2001 *EMBO J* 20:4214
9. L Chavatte, BA Brown, DM Driscoll 2005 *Nature Struct & Mol Biol* 12:408
10. E Grundner-Culemann, GW Martin III, JW Harney, MJ Berry 1999 *RNA* 5:625
11. C Buettner, JW Harney, MJ Berry 1999 *J Biol Chem* 274:21598
12. M Hirosawa-Takamori, H Jäckle, G Vorbrüggen 2000 *EMBO Rep* 1:441
13. S Castellano, N Morozova, M Morey, MJ Berry, F Serras, M Corominas, R Guigo 2001 *EMBO Rep* 2:697
14. KV Korotkov, SV Novoselov, DL Hatfield, VN Gladyshev 2002 *Mol Cell Biol* 22:1402
15. SV Novoselov, M Rao, NV Onoshko, H Zhi, GV Kryukov, Y Xiang, DP Weeks, DL Hatfield, VN Gladyshev 2002 *EMBO J* 21:3681

16. GV Kryukov, S Castellano, SV Novoselov, AV Lobanov, O Zehtab, R Guigo, VN Gladyshev 2003 *Science* 300:1439
17. C Thisse, A Degrave, GV Kryukov, VN Gladyshev, S Obrecht-Pflumio, A Krol, B Thisse, A Lescure 2003 *Gene Expression Patterns* 3:525
18. S Castellano, SV Novoselov, GV Kryukov, A Lescure, E Blanco, A Krol, VN Gladyshev, R Guigo 2004 *EMBO Rep* 5:71
19. T Mourier, A Pain, B Barrell, S Griffiths-Jones 2005 *RNA* 11:119
20. K Taskov, C Chapple, GV Kryukov, S Castellano, AV Lobanov, KV Korotkov, R Guigo, VN Gladyshev 2005 *Nucleic Acids Res* 33:2227
21. D Fagegaltier, A Lescure, R Walczak, P Carbon, A Krol 2000 *Nucleic Acids Res* 28:2679
22. I Vidovic, S Nottrott, K Hartmuth, R Lührmann, R Ficner 2000 *Mol Cell* 6:1331
23. JA Chao, JR Williamson 2004 *Structure* 12:1165
24. T Moore, Y Zhang, MO Fenley, H Li 2004 *Structure* 12:807
25. T Hamma, AR Ferré-D'Amaré 2004 *Structure* 12:893
26. A Lescoute, NB Leontis, C Massire, E Westhof 2005 *Nucleic Acids Res* 33:2395
27. P Nissen, JA Ippolito, N Ban, PB Moore, TA Steitz 2001 *Proc Natl Acad Sci USA* 98:4899
28. NB Leontis, E Westhof 2001 *RNA* 7:499
29. A Lescure, C Allmang, K Yamada, P Carbon, A Krol 2002 *Gene* 291:279
30. N Ban, P Nissen, J Hansen, PB Moore, TA Steitz 2000 *Science* 289:905
31. N Marmier-Gourrier, A Cléry, V Senty-Ségault, B Charpentier, F Schlotter, F Leclerc, R Fournier, C Branlant 2003 *RNA* 9:821
32. JA Chao, GS Prasad, SA White, C David Stout, JR Williamson 2003 *J Mol Biol* 326:999
33. V Cojocar, S Nottrott, R Klement, TM Jovin 2005 *RNA* 11:197
34. B Turner, SE Melcher, TJ Wilson, DG Norman, DMJ Lilley 2005 *RNA* 11:1192
35. AK Wozniak, S Nottrott, E Kühn-Hölsken, GF Schröder, H Grubmüller, R Lührmann, CAM Seidel, F Oesterhelt 2005 *RNA* 11:1545
36. C Charron, X Manival, A Cléry, V Senty-Ségault, B Charpentier, N Marmier-Gourrier, C Branlant, A Aubry 2004 *J Mol Biol* 342:757
37. J Suryadi, EJ Tran, E Stuart Maxwell, BA Brown 2005 *Biochem* 44:9657
38. SA White, M Hoeger, JJ Schweppe, A Shillingford, V Shipilov, J Zarutskie 2004 *RNA* 10:369
39. TA Goody, SE Melcher, DG Norman, DMJ Lilley 2004 *RNA* 10:254





# 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<sup>Sec</sup>, generating the Sec-tRNA<sup>Sec</sup> 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).

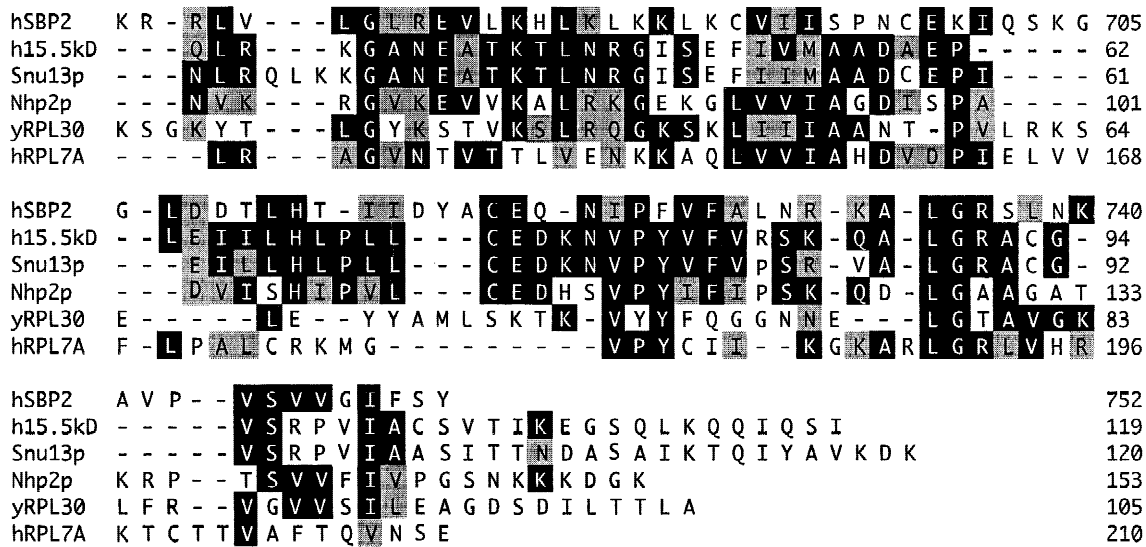
An extensive study of the amino acids required for the binding of SBP2 to the SECIS RNA has not been published yet. The issue is especially crucial because the various proteins of the L7A/L30 family can specifically recognize their cognate RNA yet share identical or similar sequences in their homologous 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.

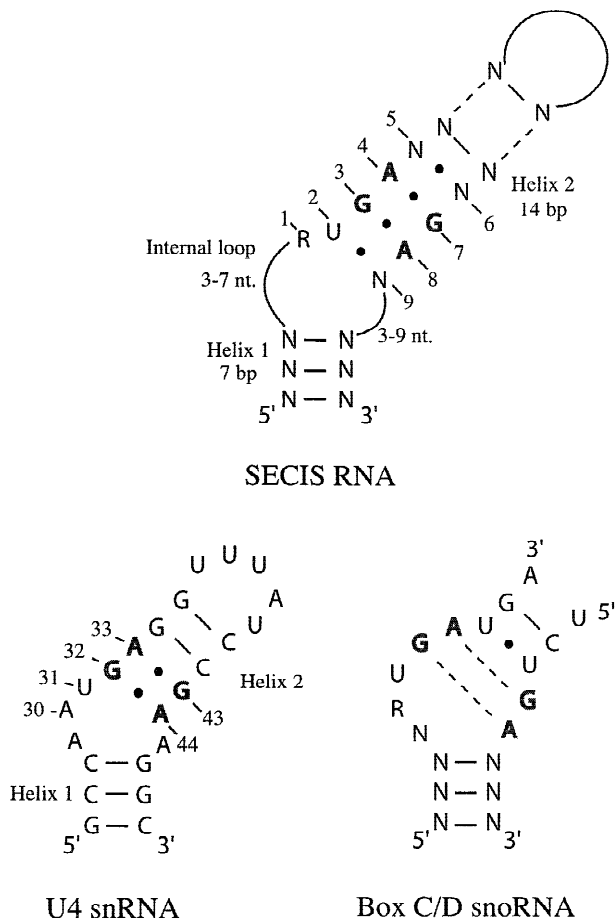
### Similar structural features in the SECIS RNA and 5' stem-loop of spliceosomal U4 snRNA

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  $\beta 2$ - $\alpha 3$  junction (see Fig. 3A) where helix  $\alpha 3$  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<sup>SBP2</sup>, Arg678<sup>SBP2</sup>, Glu679<sup>SBP2</sup>, and Lys682<sup>SBP2</sup> could contact the bases or the phosphodiester backbone of the SECIS RNA at G3 and/or G7; Leu677<sup>SBP2</sup>, Glu699<sup>SBP2</sup>, Asp709<sup>SBP2</sup>, Arg731<sup>SBP2</sup>, and Ile749<sup>SBP2</sup> could interact with U2; and Val744<sup>SBP2</sup> 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<sup>SBP2</sup> was substituted to determine whether Arg731<sup>SBP2</sup> or Lys732<sup>SBP2</sup> is homologous to Lys86<sup>15.5</sup>. Ser745<sup>SBP2</sup> 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<sup>15.5</sup>, Lys37<sup>15.5</sup>, and Arg48<sup>15.5</sup> contribute essentially to electrostatic interactions with the phosphates at positions 41–44 in U4 snRNA. Because the corresponding residues Val674<sup>SBP2</sup>, Leu675<sup>SBP2</sup>, and Leu686<sup>SBP2</sup> are hydrophobic, their interaction with the SECIS RNA was hardly predictable and they were not mutated. Likewise, Val746<sup>SBP2</sup> was not substituted because its Arg97<sup>15.5</sup> 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.

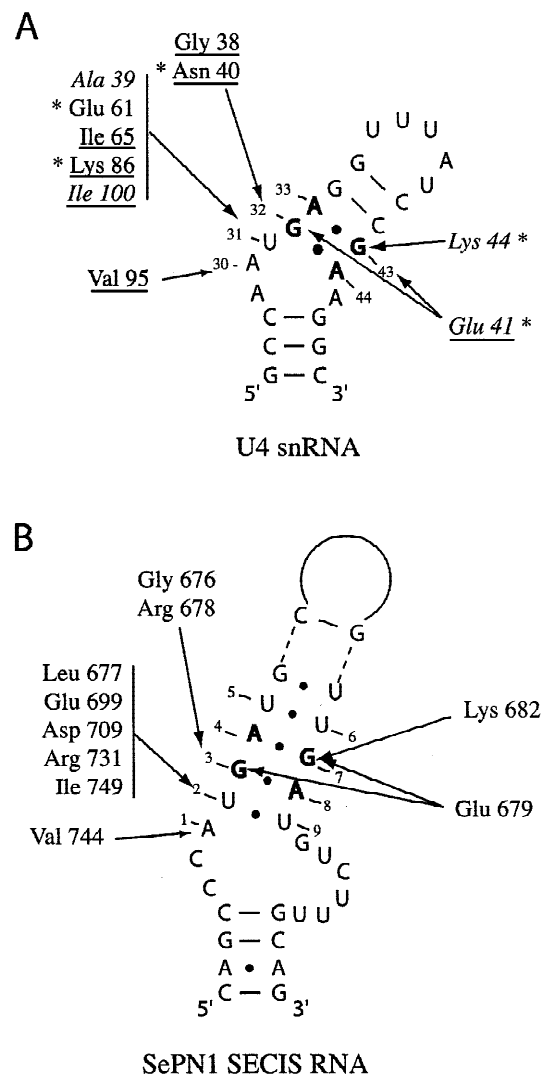
**Identification of SBP2 residues important for the interaction with the SECIS RNA**

The [<sup>35</sup>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 [<sup>32</sup>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



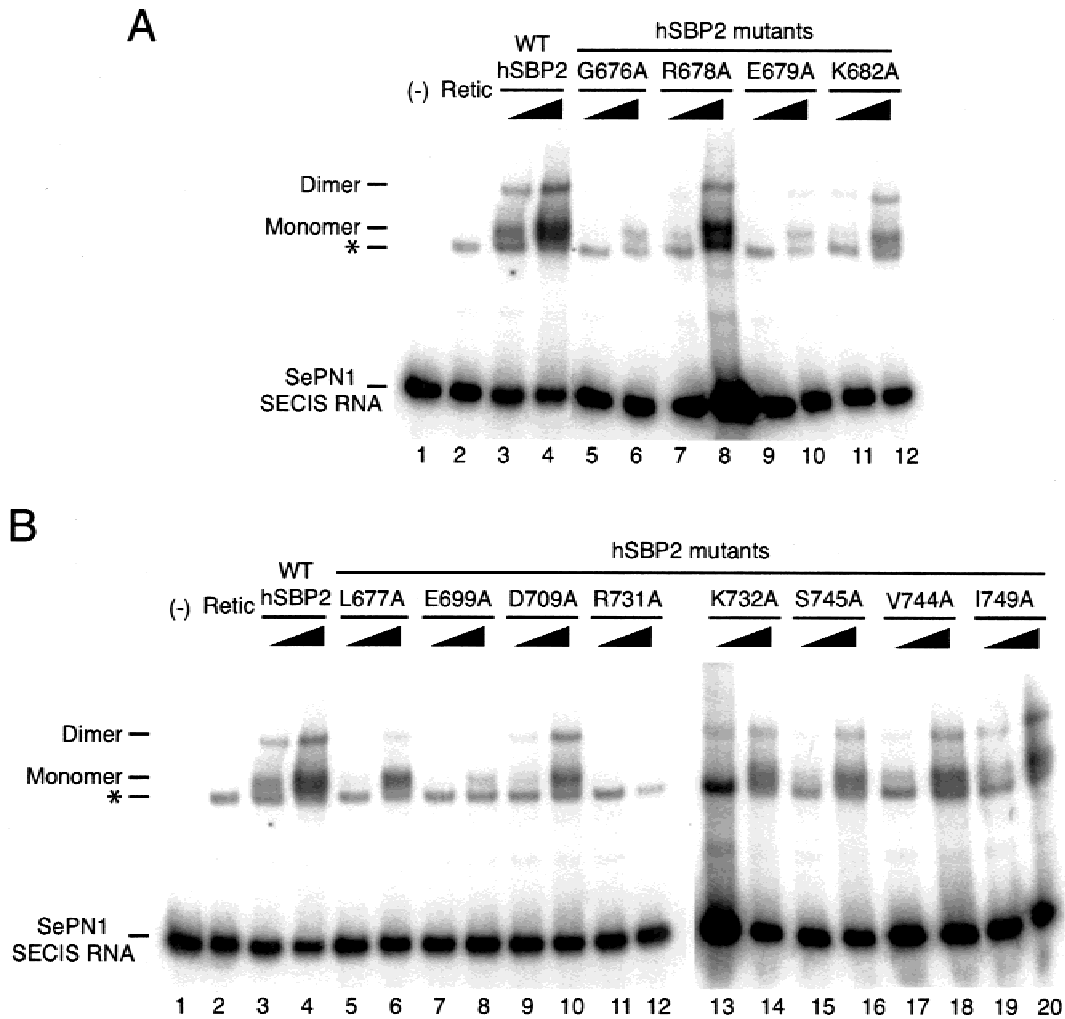
share the same binding site on the SECIS RNA. Addition of the in vitro translated wild-type hSBP2 to the SECIS RNA led to the formation of two retarded complexes containing monomeric and homodimeric forms of hSBP2 (Fig. 5A, B, lanes 3 and 4), as previously reported for the recombinant hSBP2 produced 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<sup>15.5</sup> and Lys86<sup>15.5</sup> 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<sup>15.5</sup> and Glu41<sup>15.5</sup> 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<sup>SBP2</sup>, rather than Lys732<sup>SBP2</sup>, is the homolog of Lys86<sup>15.5</sup>. 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<sup>15.5</sup> 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 [<sup>35</sup>S]-methionine-labeled hSBP2 protein obtained by in vitro translation in rabbit reticulocyte lysates was added to 150,000 cpm of [<sup>32</sup>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<sup>SBP2</sup> and Ser745<sup>SBP2</sup>, which could not be predicted. The four amino acids Gly676<sup>SBP2</sup>, Glu679<sup>SBP2</sup>, Glu699<sup>SBP2</sup>, and Arg731<sup>SBP2</sup> 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.<sup>a</sup>

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

<sup>a</sup>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<sup>15.5</sup> (corresponding to Gly676<sup>SBP2</sup>) is particular in that it adopts a conformation that is not allowed to any other amino acid at the  $\beta$ 1- $\alpha$ 2 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<sup>SBP2</sup>, Glu679<sup>SBP2</sup>, Glu699<sup>SBP2</sup>, Arg731<sup>SBP2</sup> 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<sup>SBP2</sup> and Arg678<sup>SBP2</sup> could interact with G3; Glu679<sup>SBP2</sup> with G3 and G7; Lys682<sup>SBP2</sup> with G7; Leu677<sup>SBP2</sup>, Glu699<sup>SBP2</sup>, Asp709<sup>SBP2</sup>, Arg731<sup>SBP2</sup>, and Ile749<sup>SBP2</sup> with U2; and Val744<sup>SBP2</sup> 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<sup>L30</sup>, which corresponds to Gly38<sup>15.5</sup> and Gly676<sup>SBP2</sup> (see Fig. 1), is central to the interaction with the mRNA. Substitution of the Gly676<sup>SBP2</sup> 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<sup>L30</sup> (corresponding to Ala39<sup>15.5</sup> and Leu677<sup>SBP2</sup>) and Lys28<sup>L30</sup> (corresponding to Asn40<sup>15.5</sup> and Arg678<sup>SBP2</sup>). This is consistent with our results for Leu677<sup>SBP2</sup>. However, our data did not suggest a major role for Arg678<sup>SBP2</sup>, highlighting subtle variations in the RNA-protein recognition schemes. We found that substitution of Lys732<sup>SBP2</sup> and Ser745<sup>SBP2</sup> 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<sup>15.5</sup> and Ser96<sup>15.5</sup> 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<sup>SBP2</sup> and Ser745<sup>SBP2</sup> 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-*Hind*III 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'-TTTGAGAACCTCCCTGGCCCCAACACAAG TCG-3';  
 R678A: 5'-GTGTTTGAAGAACCTCGGCCAACCCCAACAC AAG-3';  
 E679A: 5'-CAGGTGTTTGAAGAACGGCCCTCAACCCCAA CAC-3';  
 K682A: 5'-TTTGAGCTTCAGGTGGGCGAGAACCTCCCT CAA-3';  
 E699A: 5'-TTTTGACTGTATCTTGGCACAGTTGGGAGA AAT-3';  
 D709A: 5'-TAATTGTGTGCAAAGTGGCATCCAGCCCACC TTT-3';  
 R731A: 5'-GCGCCCCAGAGCTTTGGCGTTGAGAGCAA CAC-3';  
 K732A: 5'-ACTGCGCCCCAGAGCGGCGCGTTGAGAG CAAA-3';  
 V744A: 5'-GATCCCCACCACACTGGCAGGAAGTGCCTT ATT-3';  
 S745A: 5'-GAAGATCCCCACCACGGCGACAGGAAGTGC CTT-3';  
 I749A: 5'-CCCATCATAGCTGAAGGCCCCACCACACTG 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  $\mu$ L in vitro transcription/translation reactions in the presence of 25  $\mu$ L rabbit reticulocyte lysate and 20  $\mu$ Ci of [<sup>35</sup>S]-methionine (1,175 Ci/mmol). The yield of [<sup>35</sup>S]-methionine incorporation was determined by 5% TCA precipitation of 2- $\mu$ L aliquots of the reactions, followed by scintillation counting and calculation with respect to the [<sup>35</sup>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 pT7BcKSeIN 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 [ $\alpha$ -<sup>32</sup>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 <sup>32</sup>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  $\mu$ L 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. Mysliński 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<sub>2</sub>/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.
- Köhrlé 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.
- Méhot 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 SEPNI1 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, Bachelier 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.



# An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery

A. Cléry<sup>1</sup>, V. Bourguignon-Igel<sup>1</sup>, C. Allmang<sup>2</sup>, A. Krol<sup>2</sup> and C. Branlant<sup>1,\*</sup>

<sup>1</sup>Laboratoire de Maturation des ARN et Enzymologie Moléculaire – UMR 7567 CNRS-UHP, Nancy Université, Faculté des Sciences et Techniques – BP 239, 54506 Vandoeuvre-lès-Nancy Cedex, France and <sup>2</sup>Architecture et Réactivité de l'arN – CNRS-Université Louis Pasteur, Institut de Biologie Moléculaire et Cellulaire 15 Rue René Descartes, 67084 Strasbourg Cedex, France

Received September 27, 2006; Revised January 20, 2007; Accepted January 22, 2007

## ABSTRACT

By binding to SECIS elements located in the 3'-UTR of selenoprotein mRNAs, the protein SBP2 plays a key role in the assembly of the selenocysteine incorporation machinery. SBP2 contains an L7Ae/L30 RNA-binding domain similar to that of protein 15.5K/Snu13p, which binds K-turn motifs with a 3-nt bulge loop closed by a tandem of G.A and A.G pairs. Here, by SELEX experiments, we demonstrate the capacity of SBP2 to bind such K-turn motifs with a protruding U residue. However, we show that conversion of the bulge loop into an internal loop reinforces SBP2 affinity and to a greater extent RNP stability. Opposite variations were found for Snu13p. Accordingly, footprinting assays revealed strong contacts of SBP2 with helices I and II and the 5'-strand of the internal loop, as opposed to the loose interaction of Snu13p. Our data also identifies new determinants for SBP2 binding which are located in helix II. Among the L7Ae/L30 family members, these determinants are unique to SBP2. Finally, in accordance with functional data on SECIS elements, the identity of residues at positions 2 and 3 in the loop influences SBP2 affinity. Altogether, the data provide a very precise definition of the SBP2 RNA specificity.

## INTRODUCTION

Based on ribosomal subunit 3D-structure analysis, K-turn motifs were found to be frequent protein-recognition motifs in ribosomal RNAs (1). A total of

8 K-turn motifs were detected in the 23S rRNA from *Haloarcula marismortui* and the 16S rRNA from *Thermus thermophilus* (1–4). K-turn motifs are all characterized by a helix I-loop-helix II structure, and the formation of two non-Watson–Crick base pairs (most frequently G.A and A.G) within the internal loop extends helix II (1,5). Due to the stacking onto helix I or helix II of residues in the internal loop, one of the RNA strand forms a sharp angle (1,5). Only one of the residues in the loop is projected out of the K-turn structure and is located in a pocket of the protein in RNA–protein complexes. In addition to their presence in rRNAs, K-turn motifs are also found in the U4 and U4atac spliceosomal snRNAs (5,6) and in the numerous C/D box snoRNAs (7), that guide 2'-O-methylation and cleavages in pre-ribosomal RNA (for review, 8). K-turn motifs were also recently found in both C/D and H/ACA sRNAs, that guide rRNA modifications in archaea (9–11). They are thus very ancient RNA-binding motifs. Both in eukarya and in archaea, small RNAs containing K-turn motifs assemble into RNP particles and the K-turn motifs play a central role in protein assembly (7,9–15). More specifically, the ribosomal L7Ae protein in archaea or its eukaryal homolog, the Snu13p (yeast)/15.5K (human) protein, first recognizes the K-turn structure and the complex formed then serves as a platform for assembly of the other proteins (9,10,12–19).

The Snu13p/15.5K and L7Ae proteins belong to the L7Ae/L30 protein family, which is characterized by the presence of an L7Ae/L30 RNA-binding domain (6,20). The founding member of this protein family, the yeast L30 ribosomal protein recognizes a peculiar K-turn motif in its own pre-mRNA (21–23). One difference between the yeast L30 RNA–protein complex, and the Snu13p/15.5K or L7Ae RNA–protein complexes is the

\*To whom the correspondence should be addressed. Tel: 33 383684303; Fax: 33 383684307; Email: christian.branlant@maem.uhp-nancy.fr

identity of the nucleobase located in the protein pocket. Whereas, a strong preference for an U residue is observed for proteins Snu13p/15.5K and L7Ae (7,24–26), C and A residues are preferentially accommodated in the yeast L30 protein pocket (27). The possibility to bind a G residue was however recently observed (28).

In vertebrates, SECIS-binding protein 2, another member of the L7Ae/L30 protein family, binds SECIS elements in mRNAs (29,30). The SECIS elements contain determinants needed for selenocysteine incorporation into selenoproteins (31,32). Selenocysteine incorporation involves reprogramming of a nonsense UGA codon into a codon recognized by the selenocysteine specific tRNA<sup>Sec</sup>. Understanding the mechanism of selenocysteine incorporation into proteins is important as they are key players in the antioxidant defense system (for review, 33). They are also key participants in a variety of other systems including thyroid hormone metabolism, muscle function, transportation and distribution of selenium to remote tissues and can have roles as structural proteins (for reviews, 34–37). In eukarya, the SECIS elements and SBP2 are two essential components of the selenocysteine incorporation machinery. All SECIS elements consist of a hairpin structure composed of two helices I and II, separated by an internal loop. A highly conserved cluster of four non-Watson–Crick base pairs is located in helix II. It contains a tandem of G.A and A.G pairs, which is needed for SBP2 binding (29,30). This cluster of non-Watson–Crick pairs is an essential determinant for selenocysteine incorporation (31,32). A highly conserved AAR sequence present in a loop of all SECIS elements is also important for selenoprotein synthesis *in vivo*, but not for binding of SBP2 *in vitro* (30,38). As SBP2 also binds the specific mSelB/EFSec elongation factor, it is proposed to recruit this dedicated elongation factor in a complex formed with the selenocysteyl-tRNA<sup>Sec</sup> to the ribosomes (39–41). Additionally, according to a recent investigation on the selenocysteine incorporation machinery (42), the ribosomal protein L30 is able to bind the SECIS motif by displacing protein SBP2. This substitution would facilitate the interaction of the Sec-tRNA<sup>Sec</sup> with ribosomes.

A prerequisite to fully understand the SBP2 activity is thus to obtain a more complete picture of the RNA sequence and structural determinants required for SBP2 binding. To this end, we combined the SELEX approach and site-directed mutagenesis experiments. As the RNAs recovered after SELEX experiments could form canonical K-turn motifs with a protruding U residue, we compared the RNA-binding properties of the human SBP2 protein with those of a well-characterized member of the L7Ae/L30 protein family, the *S. cerevisiae* Snu13p protein. This protein recognizes K-turn motifs in U4 snRNA, the C/D box snoRNAs and U3 snoRNA. Altogether, we show here that in contrast to protein Snu13/15.5K, SBP2 preferentially binds RNA motifs with a large internal loop. In addition, we demonstrate the existence of previously undetected important determinants for RNA recognition by SBP2 that are located in helix II.

## MATERIALS AND METHODS

### Strains and growth conditions

The *Escherichia coli* TG1 strain was used as the host strain for plasmid construction. Growth was performed at 37°C in Luria Broth medium, complemented with 100 µg/ml of ampicillin when necessary. The *E. coli* strain BL21-CodonPlus (Stratagene) was the host strain for production of the recombinant GST/Snu13p, GST/L7Ae and GST/C-SBP2 proteins.

### Recombinant DNA

Plasmids pT7SelN (40), pUC18::U3AΔ2,3,4 (26) and pyU4 (43) were used for the production of matrices for *in vitro* transcription of the SelN, yU3B/C and yU4 RNAs, respectively. The yU3B/C and yU4 matrices were obtained by PCR amplification, under conditions previously described (26). Oligonucleotides yU3B/C-5', yU3B/C-3', yU4-5' and yU4-3', given in Table 1 of the Supplementary Data, were used as primers. Plasmids pGEX-6P-1::SNU13, pGEX-6P-1::L7AE (44) and pGEX-6-P1::C-SBP2 (this work) were used for production of the recombinant GST/Snu13p, GST/L7Ae and GST/C-SBP2 proteins, respectively. Plasmid pA11 was used for amplification of the PCR fragment coding for region 515–854 of human SBP2 protein (45). DNA fragments amplified by RT-PCR from RNAs obtained after the fourth cycle of the SELEX experiment were cloned into plasmid pCR2.1 (Invitrogen). Mutagenesis of the RNA Sel coding sequence was performed by the PCR-based site-directed strategy (primers are listed in Table 1 in the Supplementary Data).

### *In vitro* transcription

The EcoRI linearized pT7::SelN plasmid was used as the template for SelN RNA transcription. The yU3B/C, yU4, Sel1–Sel7 and Sel1 variant RNA-coding sequences were transcribed from PCR amplified DNA fragments obtained as described above. Transcriptions were carried out on 1 µg of plasmid DNA linearized with EcoRI or 500 ng of PCR product, in a 15 µl reaction as described in Marmier-Gourrier *et al.* (26).

RNAs were 5'-end labeled using 10 units of T4 polynucleotide kinase (MBI-Fermentas), 20 pmol of RNA, 5 pmol of [ $\gamma$ -<sup>32</sup>P] ATP, in a 10-µl reaction mixture containing 10 mM MgCl<sub>2</sub>; 5 mM DTT; 0.1 mM spermidine; 0.1 mM EDTA; 50 mM Tris-HCl pH 7.6 at 37°C. The 5'-end labeled RNAs were purified on a 10% denaturing polyacrylamide gel.

### Recombinant protein preparation

The recombinant GST/Snu13p and GST/L7Ae fusion proteins were produced in *E. coli* as described in Marmier-Gourrier *et al.* (26). The same procedure was used for the production of C-SBP2. For purification of untagged proteins, they were bound on glutathione-sepharose 4B as previously described (44) and cleaved on the beads using 80 U of PreScission protease (Pharmacia) per ml of glutathione-sepharose bead suspension, under published conditions (44). The purified proteins were dialyzed

against buffer D (150 mM KCl; 1.5 mM MgCl<sub>2</sub>; 0.2 mM EDTA; 20 mM HEPES, pH 7.9; 10% glycerol) and aliquots were stored at -80°C.

### SELEX experiment

The starting DNA matrix containing a 18-nt-long degenerated sequence was produced by PCR amplification, using two partially complementary oligonucleotides (Table 1 in Supplementary Data): SELEX N18 with a 18-nt-long degenerated sequence and SELEX-5', that generated a T7 RNA polymerase promoter. PCR amplification was as previously described (26), except that MgCl<sub>2</sub> was added at a 4 mM concentration in the incubation buffer. About 500 ng of amplified DNA was used for *in vitro* transcription with T7 RNA polymerase (26). Transcripts were purified by electrophoresis on a 6% denaturing polyacrylamide gel as in Mougin *et al.* (46). About 0.2 nmol of transcripts were used for the first round of selection. To eliminate RNA molecules having an affinity for the glutathione-sepharose beads, the RNA mixture was first incubated with 30 µl of beads in the absence of the GST/C-SBP2. For RNP complexes, 0.1 nmol of treated RNAs was incubated with 0.01 nmol of purified GST/C-SBP2 for 30 min at 4°C, in 20 µl of buffer D, in the presence of 2 µg of a yeast tRNA mixture (Roche). The mixture was then incubated with 15 µl of glutathione-sepharose beads (Amersham) equilibrated in buffer D. After extensive washing with buffer D, the selected RNAs were released by a 30-min incubation at 37°C, with 20 µg of proteinase K in buffer D. They were extracted with a phenol-chloroform mixture, ethanol precipitated, dissolved in sterile water, hybridized with 50 pmol of SELEX-3' primer, ethanol precipitated, and finally reverse-transcribed with 25 U of AMV Reverse transcriptase (Q.Biogene) for 30 min at 42°C. Next, 30 cycles of PCR amplification were performed in the presence of primers SELEX-5' and SELEX-3' (50 pmol each). The amplified DNA fragments were gel purified and used as the matrix for *in vitro* transcription. At each cycle of the SELEX experiment, a filter-binding assay was performed after incubation of the uniformly labeled transcripts produced from the DNA pool with the GST/C-SBP2 protein. At the fourth cycle of the amplification-selection experiment, DNA fragments were cloned into plasmid pCR2.1 (Invitrogen). Plasmids were prepared from 30 randomly selected clones and sequenced by the dideoxysequencing method.

### Electrophoretic mobility shift assay

About 5 fmol of *in vitro* transcribed 5'-end labeled RNAs, mixed with 2 µg of yeast tRNAs (Roche), were denatured during 10 min at 65°C in 15 µl of buffer D containing 1.5 mM of MgCl<sub>2</sub>, followed by a slow cooling to room temperature for renaturation. To test for the effect of Mg<sup>++</sup> on complex formation, the Mg<sup>++</sup> concentration was adjusted to 1.5, 5, 10, 15 or 20 mM by addition of MgCl<sub>2</sub>, without modification of the final volume of incubation and a control experiment was performed in the absence of Mg<sup>++</sup>. The Snu13p or C-SBP2 recombinant proteins were added at various concentrations

(from 0 to 4 µM) and the mixture was incubated for 30 min at 4°C. RNA-protein complexes were fractionated by electrophoresis on 6% non-denaturing polyacrylamide gel as in Marmier-Gourrier *et al.* (26). The amount of radioactivity in the bands, corresponding to the free and complexed RNA, was estimated using a PhosphorImager and the ImageQuant Software. Using these values, apparent *K*<sub>d</sub>s were determined with the SigmaPlot Software (SPSS Science Software). For competition assays with an excess of cold RNA or protein, protein-RNA complexes were preformed as mentioned above, and various amounts of cold competitor RNAs or competitor proteins were added, followed by a 30-min incubation at 4°C. The remaining complexes were subjected to gel electrophoresis.

### RNA secondary structure analysis

*In vitro* transcribed 5'-end labeled RNAs (25 fmol) were pre-incubated in buffer D for 5 min at 65°C, in the presence of 2 µg of tRNA followed by a slow cooling for renaturation. The renatured RNAs were then incubated for 30 min at 4°C in the absence or presence of C-SBP2 (100, 50 and 30 pmol, respectively), Snu13p (10, 100 and 30 pmol, respectively) or L7Ae (10 pmol), in 10 µl of buffer D. Digestion was for 6 min at 20°C, in the presence of 0.8 U of T1 RNase (Roche), 2.4 U of T2 RNase (Gibco) or 0.001 U of V1 RNase (Kemetex). V1 RNase reactions were stopped by addition of 100 mM EDTA, followed by phenol extraction. T1 and T2 RNase digestions were stopped by addition of 20 µg of tRNA, followed by phenol extraction and ethanol precipitation. For production of a ladder, an alkaline hydrolysis of the naked RNA was performed for 5 min at 96°C, using 10 fmol of RNA dissolved in 1 µl of 100 mM sodium bicarbonate. The cleavage products were fractionated by electrophoresis on a 10% polyacrylamide-8 M urea gel.

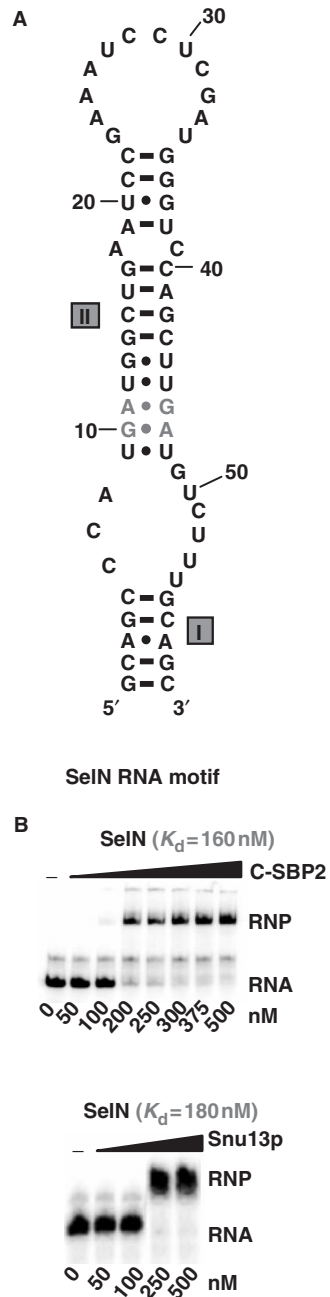
The free energies of the 2D structures of the selected RNAs were calculated at 37°C and in 1 M NaCl with the M-fold software (46).

## RESULTS

### Protein C-SBP2 does not interact with K-turn motifs recognized by Snu13p

As ribosomal protein L30 was shown to displace SBP2 from SECIS motifs, our first goal was to test whether SBP2 can bind RNA targets of members of the L7Ae/L30 protein family. The large human SBP2 protein (854 aa) has a low solubility. As we wanted to study the RNA-binding property of its L7Ae/L30 domain, we used a truncated version containing this domain. This human SBP2 fragment encompassing residues 515-854 was produced in a soluble form in *E. coli*. It will be hereafter designated as C-SBP2. To test its capacity to bind SECIS RNAs, we used the well-characterized SECIS RNA motif from the human selenoprotein N mRNA (SelN RNA) (Figure 1A) (30). RNP complexes were formed by incubation of uniformly labeled SelN RNA (5 fmol) with C-SBP2, at concentrations ranging from 50 to 500 nM, in the presence of 2 µg of tRNAs (see the Materials and





**Figure 1.** C-SBP2 and Snu13p interact with SelN RNA. (A) The secondary structure of the SelN RNA motif is according to Fagegaltier *et al.* (52). The G.A sheared base pairs are shown in gray and helices I and II are indicated. (B) The affinity of C-SBP2 and Snu13p for SelN RNA was tested by gel-shift assay using 5 fmol of labeled SelN RNA and protein concentrations ranging from 0 to 500 nM, as indicated below the lanes. Incubation conditions are described in the Materials and Methods section. RNP formation was revealed by electrophoresis on 6% non-denaturing polyacrylamide gels. The apparent  $K_d$  values (indicated above the autoradiograms) were calculated with the SigmaPlot Software (SPSS Science Software), by measuring the radioactivity signals corresponding to the free and bound RNAs.

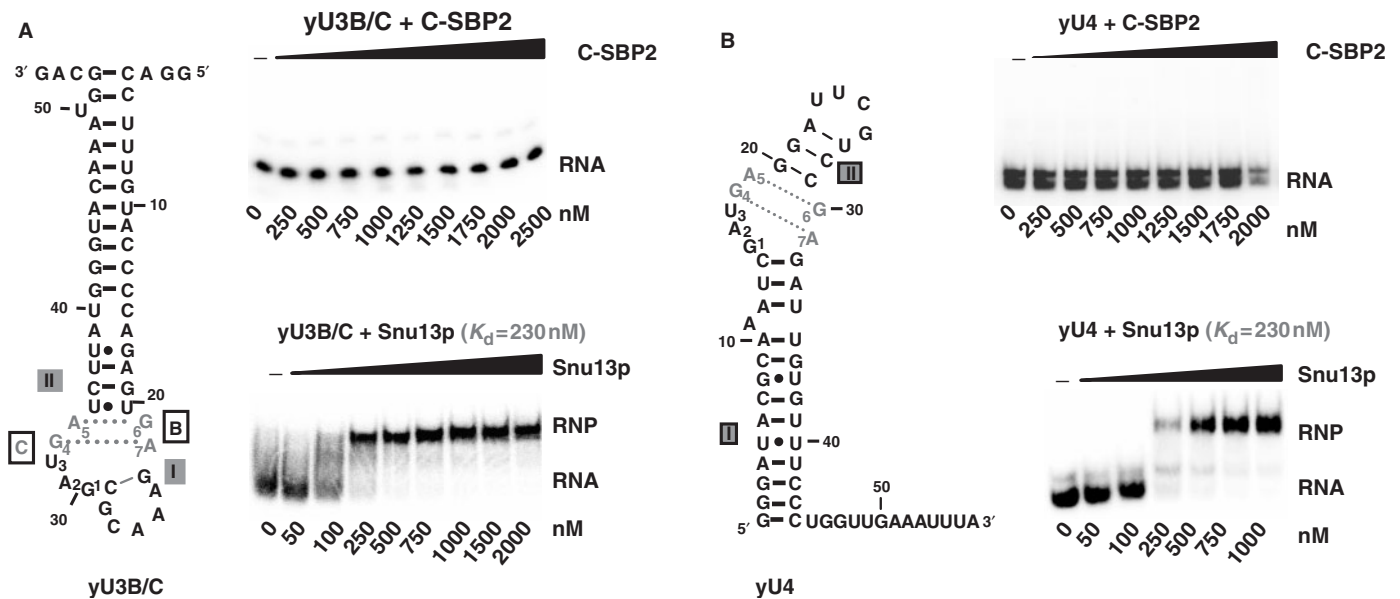
Methods section for the incubation conditions). As evidenced by gel electrophoresis performed under non-denaturing conditions (Figure 1B), C-SBP2 formed an RNP complex with the SelN RNA and the apparent

$K_d$  was of 160 nM. Next, we tested the capacity of this protein to bind K-turn motifs targeted by Snu13p. Two well-characterized RNAs were used: RNA yU3B/C containing the B/C motif of yeast U3 snoRNA (26), and RNA yU4 containing the K-turn motif of yeast U4 snRNA (see the Materials and Methods section for their production). Complexes were formed under the same conditions as for SelN RNA. As a control, the same experiment was performed with Snu13p. Gel electrophoresis revealed the absence of binding of C-SBP2 to both Snu13p RNA targets, even at a high protein concentration (Figure 2). As in contrast, Snu13p was found to bind the SelN RNA with an apparent  $K_d$  similar to that of C-SBP2 (Figure 1B), we concluded that to bind C-SBP2, the RNA should have sequence or structure peculiarities, which are not required for association with Snu13p.

### A limited diversity of RNAs selected by C-SBP2 in SELEX experiments

To progress in the understanding of how the SBP2 L7Ae/L30 domain recognizes RNA, we used the yU3B/C RNA, and tried to define by SELEX experiments which kinds of mutations can convert this RNA into a C-SBP2 target. To this end, we degenerated a 18-nt long fragment in the central part of the yU3B/C coding region. The transcripts produced from this degenerated matrix (N18 RNA) were subjected to selection with a GST/C-SBP2 protein fusion that was bound to glutathione-sepharose beads. In spite of the degenerated sequence, all the RNAs were expected to contain the long-terminal stem of RNA yU3B/C (Figure 2A). As the same kind of experiment has previously been performed with Snu13p (47), we also expected to compare the RNA motifs selected by C-SBP2 and by Snu13p. To initiate the selection cycles, we used 5  $\mu$ g of degenerated RNA mixture, so that each possible RNA sequence was expected to be present 2300 times (47). As a first step, RNAs that might have an affinity for the matrix were eliminated from the RNA pool by incubation with the glutathione-sepharose beads in the absence of protein. Following each selection cycle, the interaction of the pool of selected RNAs with the GST/C-SBP2 fusion was tested by gel-shift assays (conditions for the amplification–transcription–selection cycles are described in the Materials and Methods section). A strong increase of the amount of RNAs showing an affinity for the fusion protein was observed after the fourth cycle of selection. After this cycle, the totality of the selected RNAs was subjected to gel electrophoresis under non-denaturing conditions, and the RNA mixture contained in the slice of gel corresponding to RNPs was extracted, converted into cDNAs, and cloned into plasmid pCR2.1. After transformation of *E. coli* TG1 cells, thirty colonies were randomly selected among >100 colonies obtained.

Several of them contained plasmids with identical inserts (Figure 3A). Only seven distinct sequences were found (RNAs denoted Se1 to Se7) (Figure 3A). In addition, three of the sequences that corresponded to the most abundant clones were very similar, suggesting that their small differences were most probably generated



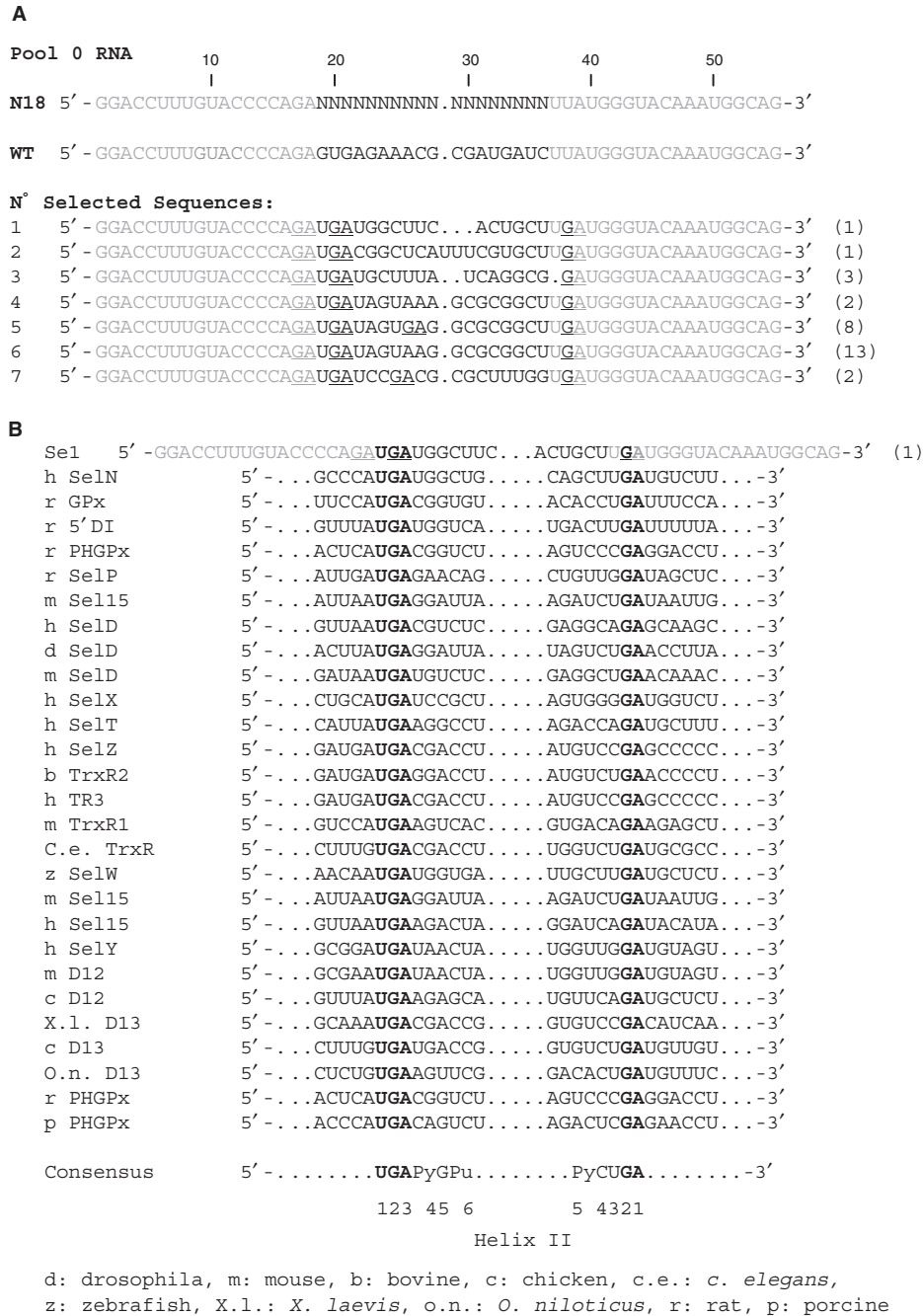
**Figure 2.** C-SBP2 does not interact with yU3B/C and yU4 RNAs. The binding of C-SBP2 and Snu13p on yU3B/C or yU4 RNAs was tested by gel-shift assays. The secondary structures of yU3B/C (Panel A) and yU4 (Panel B) are according to Marmier-Gourrier *et al.* (26) and Mougin *et al.* (43), respectively. The G.A sheared pairs are in gray and helices I and II are indicated. Nucleotides involved in the K-turn folding are numbered from 1 to 7. Complexes were formed between 5 fmol of uniformly labeled yU3B/C (Panel A) or yU4 (Panel B) RNAs, and C-SBP2 or Snu13p at concentrations ranging from 0 to 2500 nM, as indicated below the lanes. Incubation conditions were as described in the Materials and Methods section. The autoradiograms obtained after electrophoresis on 6% polyacrylamide gels are shown. For the Snu13p–RNP complexes shown as controls, the apparent  $K_d$  values were calculated as described in the Materials and Methods section.

by RT-PCR errors in the course of the amplification. Therefore, only four main classes of Se RNAs were selected in the experiment (Figure 3A). This limited diversity of the selected sequences was one major difference, compared to the SELEX experiment performed with Snu13p (31 very different sequences were obtained starting from the same initial RNA mixture). This suggested the existence of strong sequence and/or structural constraints for SBP2 recognition. Three of the less frequent Se RNAs had a different length as compared to the initial RNAs: RNA Se2 (Figure 3A) contained an additional residue in the degenerated sequence, whereas RNA Se1 and Se3 lacked two residues compared to the initial RNAs. In addition, all the Se RNAs had a G instead of a U residue at position 38 in the conserved sequence. The same U38G base substitution was also found in several of the RNAs selected by the GST/Snu13p protein (47). By using gel-shift experiments, we verified that the seven distinct Se RNAs, that were selected, showed an affinity for the untagged C-SBP2 protein (Figure 3C). A wide range of apparent  $K_d$  values was observed (from 500 nM to >2000 nM) (Figure 3C).

### The selected RNAs all form canonical K-turn motifs

In order to understand the structural reasons for these different affinities, the possible folding of the seven selected RNAs was investigated. Remarkably, each of them could form a canonical K-turn structure with tandem G.A and A.G base pairs and a 3-nt bulge including a U residue at position 3. Most of the proposed structures were verified by enzymatic probing (Figure 4A).

They are represented in Figure 4B, where they are classified according to the values of the established apparent  $K_d$ s. The free energies of the proposed 2D structures at 37°C in 1 M NaCl were also calculated by using the M-Fold software. Based on these structures, nts 17–21 and 38–39 correspond to residues 1–5 and 6–7 of the K-turn motif. Hence, residues 1 and 2 in the bulge, the A residue of the first G.A pair in stem II and one U residue of the third pair in this stem corresponded to invariant residues in the starting RNA mixture. The G residue of the A.G pair corresponded to the above-mentioned U to G mutation at position 38 in the constant region. This G residue might have been generated by misincorporation in the course of the amplification cycles. Its selection in all the RNAs is in agreement with the high functional importance of the A.G pair in K-turn formation. All the selected RNAs had an identical UGAU sequence from position 19 to 22 in the randomized segment, which demonstrated a strong pressure for the selection of a perfectly canonical K-turn structure with two A.G and G.A pairs, and a U residue at position 3 in the bulge. In all the selected RNAs, except RNA Se3 which contains a G.U pair, the constant U<sub>37</sub> residue was always facing a U or a C residue in helix II. Interestingly, a U.U pair was almost always selected at this position of helix II in the SELEX experiment performed with Snu13p (47). Requirement of a non-Watson–Crick pair on top of the G.A and A.G pairs for binding of C-SBP2 may explain the absence of binding of C-SBP2 to yU4 RNA which has a G–C pair at this position in helix II (Figure 2B). Finally, in all the selected Se RNAs, helix II contained at least two Watson–Crick



**Figure 3.** Sequences of the RNAs recovered from the SELEX experiment and test of their affinities for C-SBP2. (A) Alignment of the WT yU3B/C RNA sequence with the degenerated N18 RNA and the selected Se1-Se7 RNAs sequences. Nucleotides in Se1-Se7 RNA, are numbered according to the positions of the homolog nucleotides in the WT yU3B/C RNA. The number of sequenced plasmids encoding each selected RNA is indicated in brackets on the right of the sequences. The nucleotides corresponding to the constant sequence are shown in gray, nucleotides in the degenerated sequence and nucleotides mutated during the RT-PCR cycles are shown in black. The GA dinucleotides are underlined. (B) The nucleotide sequences of a series of SECIS motifs from various genes and species (30,52) were aligned with the Se1 RNA sequence taking as references the UGA and GA conserved nucleotides of the K-turn structure (bold characters). A consensus sequence of the SECIS K-turn motifs is deduced from the alignment and indicated below. The positions of the conserved nucleotides in the two strands of helix II are indicated (C) Estimation of the affinity of C-SBP2 for the Se1, Se2, Se3, Se5 and Se7 RNAs by gel-shift assays. RNA-protein complexes formed with 5fmol of labeled RNA and increasing concentrations of C-SBP2 (as indicated below the lanes) were fractionated by gel electrophoresis as in Figure 1. The apparent  $K_d$  values are indicated above the autoradiograms.

base pairs. They are most frequently (RNAs Se3, 4, 5, 6 and 7) stacked on the three non-Watson-Crick base pairs. In agreement with the absence of binding of C-SBP2 to yU3B/C RNA (Figure 2A), none of the selected RNAs

had a bulge in the 3' strand and a short helix I. In contrast, no restriction on the size of the bulge, or on the length of helices I and II, was found in the SELEX experiment performed with Snu13p (47). Altogether, the data

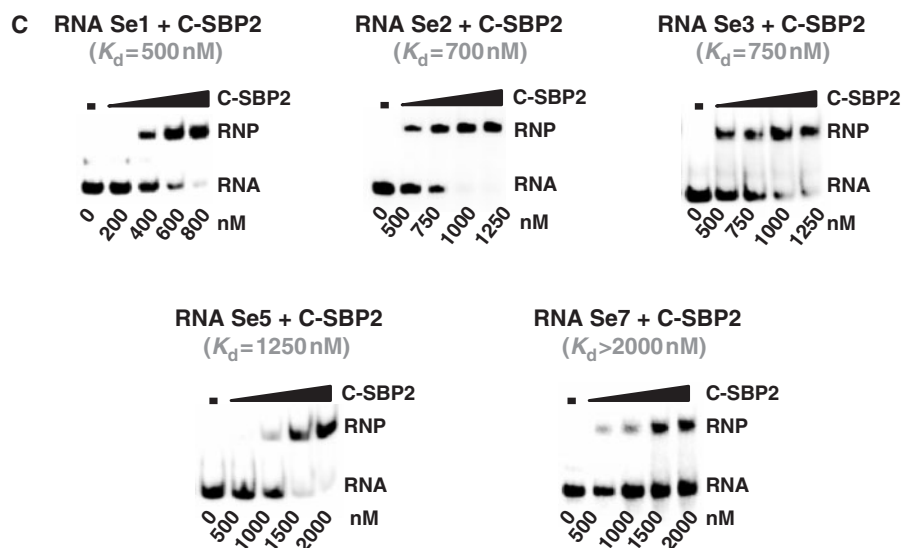


Figure 3. Continued

suggested that C-SBP2 binding requires a higher stability of the helices I and II compared to Snu13p binding. Surprisingly, the three selected RNAs, which showed the highest stabilities and also the strongest affinities for C-SBP2, were encoded by DNA sequences that were underrepresented among the cloned DNA sequences. This apparent discrepancy may be explained by the fact that RNAs Se1, 2 and 3 all have different lengths compared to the initial RNAs. They might have been generated in a late step of the selection procedure. The very low affinity found for RNA Se7, which has a stability slightly higher than those of RNAs Se4, Se5 and Se6, might be explained by sequence differences in stem II.

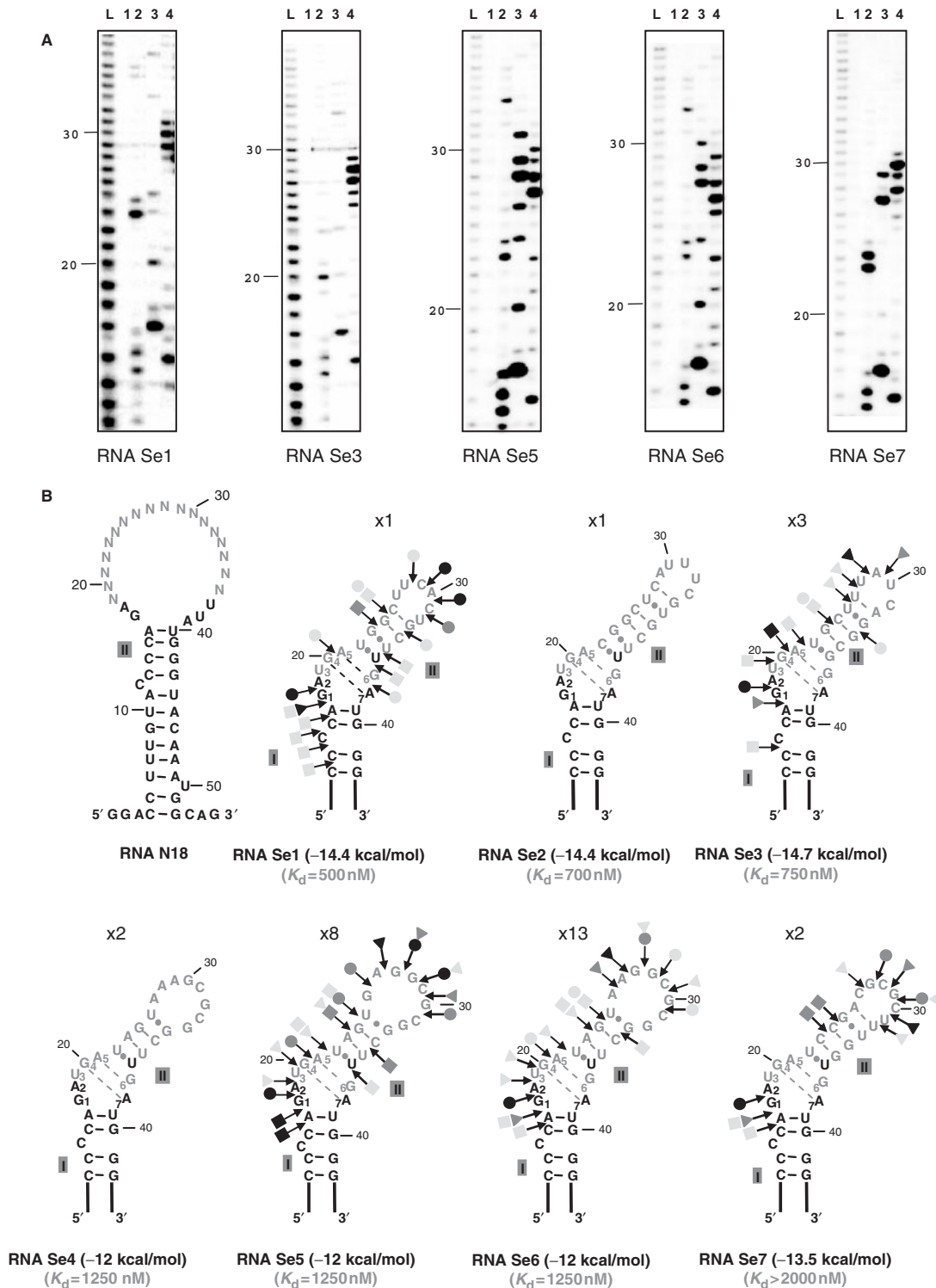
#### Specific requirements in helix II for efficient binding of protein C-SBP2

Prior to site-directed mutagenesis of Se1 RNA, we tested the influence of  $Mg^{++}$  concentration on C-SBP2 binding to this RNA. Indeed, previous data (42) established the influence of the concentration of this divalent cation on the binding of recombinant SBP2 *in vitro*. C-SBP2 binding was found to be more sensitive to the presence of  $Mg^{++}$  ions than Snu13p binding. However, the 1.5 mM  $Mg^{++}$  concentration present in the experimental buffer was found to be sufficient to ensure an efficient binding of C-SBP2 on Se1 RNA (Figure 1 in Supplementary Data). Thus subsequent experiments were performed under these conditions. To test the importance of the sequence of helix II for SBP2 binding, we mutated helix II in the winner Se1 RNA. The Se1 RNA variants produced are shown in Figure 5A. Their affinities for C-SBP2 and Snu13p were compared by gel-shift assays. Complexes were formed at different protein concentrations in order to define the apparent  $K_d$  values (Figure 5B). Interestingly, Snu13p had a very high affinity for RNA Se1. The estimated  $K_d$  (35 nM) was similar to that found for the winner RNAs in the Snu13p SELEX experiment (47). A lower affinity was found for C-SBP2 ( $K_d$  of 500 nM). Most of the base

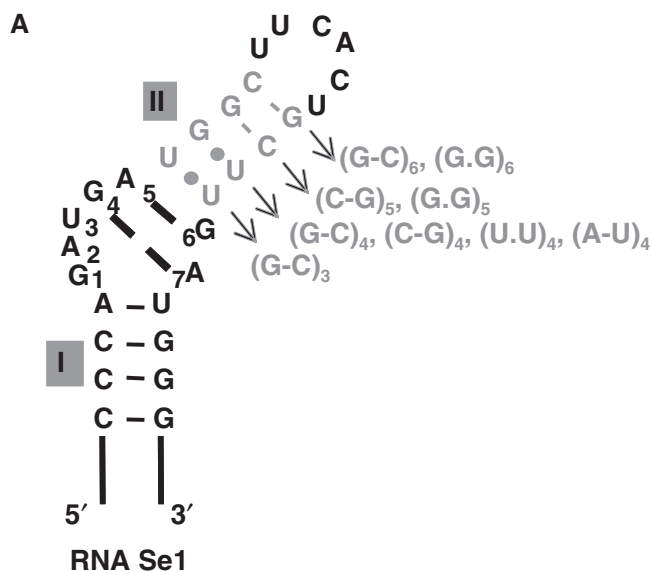
substitutions in helix II had no marked effect on Snu13p affinity. Only the strong destabilization of helix II generated by substitution of the fifth Watson–Crick base pair (G-C)<sub>5</sub> by a G.G pair had a marked deleterious effect on Snu13p affinity (factor of 20). In contrast, several base substitutions in helix II, (U.U)<sub>3</sub> to (G-C)<sub>3</sub>, (G-C)<sub>5</sub> to (G.G)<sub>5</sub> and (C-G)<sub>6</sub> to (G.G)<sub>6</sub> almost abolished C-SBP2 binding. The (G.U)<sub>4</sub> to (C-G)<sub>4</sub> and, to a lesser extent, the (G.U)<sub>4</sub> to (U.U)<sub>4</sub> substitutions, also had a marked negative effect. Hence, we concluded that C-SBP2 can interact with canonical K-turn structures, provided that helix II contains a triplet of non-Watson–Crick base pairs including the G.A and A.G sheared pairs and at least two consecutive Watson–Crick base pairs in helix II. In addition, the base pairs on top of the triplet of non-Watson–Crick base pairs should be a Pu.Py pair (G.U, G–C or A–U). This may explain why the Se7 RNA, which has a Py.Pu pair at this position, has a low affinity for protein C-SBP2.

#### The presence of a large internal loop instead of the bulge increases C-SBP2 affinity

The apparent  $K_d$  of the complex formed by C-SBP2 and the winner Se1 RNA was 3-fold lower than that found for the natural Se1N RNA (Figures 1 and 5B). Inspection of the 2D structures of these two RNAs suggested two possible explanations for the observed difference of affinity. The presence of a long stem II in Se1N RNA, and/or the presence of a large internal loop instead of a bulge in this RNA might increase C-SBP2 affinity. We tested whether the insertion of two Watson–Crick base pairs in helix II of RNA Se1 (RNA Se1:Ins) might increase the affinity of C-SBP2 (Figure 6A). Based on the observed affinities of RNA Se1:Ins for C-SBP2 and Snu13p (apparent  $K_d$ s of 300 and 25 nM, respectively), the 2 bp insertion only had a limited positive effect on C-SBP2 affinity and no marked effect on Snu13p affinity. When, in addition to the extension of stem II, the bulge of RNA Se1



**Figure 4.** All the selected RNAs that recognize C-SBP2 can form a K-turn structure. (A) Secondary structure analysis of RNAs Se1, Se3, Se5, Se6 or Se7 by enzymatic probing. The RNAs were 5'-end labeled with  $^{32}$ P, renatured and digested with V1 (0.001 U, lane 2), T1 (0.8 U, lane 3) or T2 (2.4 U, lane 4) RNases, under conditions described in the Materials and Methods section. As a control, undigested RNA was fractionated in parallel (lane 1). Lane L corresponds to the alkaline hydrolysis of the RNA used for localization of the RNase cleavage sites. Electrophoresis was performed on a 10% 8M urea-polyacrylamide gel. Nucleotide positions are indicated on the left. (B) Secondary structure models proposed for the selected RNAs. Models were proposed based on thermodynamic considerations and the results of the enzymatic digestions are shown in A. Regions corresponding to the degenerated sequences are shown by gray characters. For RNAs Se1, 3, 5, 6 and 7, V1, T1 and T2 RNase cleavages are represented by arrows surmounted of squares, dots and triangles, respectively. The color of symbols reflects the intensity of cleavages (gray, dark gray and black for low, medium and strong, respectively). Nucleotide numbering is as in Figure 3A. The apparent  $K_d$  values established for each RNA by gel retardation are indicated. The free energies of the proposed secondary structures, expressed in kcal/mol, were calculated by using the M-Fold software.



**B**

	C-SBP2	Snu13p
RNA Se1	500 nM	35 nM
(G-C) <sub>3</sub>	>4000 nM	45 nM
(G-C) <sub>4</sub>	600 nM	30 nM
(C-G) <sub>4</sub>	3000 nM	20 nM
(U.U) <sub>4</sub>	2000 nM	40 nM
(A-U) <sub>4</sub>	1050 nM	20 nM
(C-G) <sub>5</sub>	2000 nM	30 nM
(G.G) <sub>5</sub>	>4000 nM	740 nM
(G-C) <sub>6</sub>	1100 nM	40 nM
(G.G) <sub>6</sub>	>4000 nM	45 nM

**Figure 5.** Mutations in helix II of RNA Se1 are more deleterious for C-SBP2 than for Snu13p binding. (A) Positions of base substitutions in the Se1 RNA are represented in gray on the proposed secondary structure. The nature of the mutations in the variant Se1 RNAs is shown on the right of helix II. (B) The affinities of C-SBP2 and Snu13p for Se1 RNA and its variants were estimated by gel-shift assays using 5'-end labeled RNAs and protein concentrations ranging from 0 to 4000 nM. The apparent  $K_d$  values obtained for each of the RNA-protein complexes are indicated.

was converted into a large internal loop (RNA Se1:Ins+loop), the affinity for C-SBP2 was increased by a factor of 4 as compared to RNA Se1. In contrast, the affinity for protein Snu13p was decreased by a factor of 18 (Figure 6B). Hence, the presence of a large internal loop is favorable for C-SBP2 binding, but not for Snu13p interaction.

Having selected an RNA (Se1:Ins+loop RNA) with an affinity for C-SBP2 similar to that of the authentic SBP2 RNA target (SelN RNA) (Figure 1B), we then tested

the effect on C-SBP2 affinity of mutations at positions 2 and 3 in the internal loop of this RNA (Figure 6C). The results obtained revealed a preference for an A and to a lesser extent a U residue at position 2. The strongest negative effect on C-SBP2 affinity was observed for an A to C substitution at position 2 and a U to G substitution at position 3 (Figure 6C). Therefore, the identity of residues at positions 2 and 3 in the internal loop has a strong influence on C-SBP2 affinity.

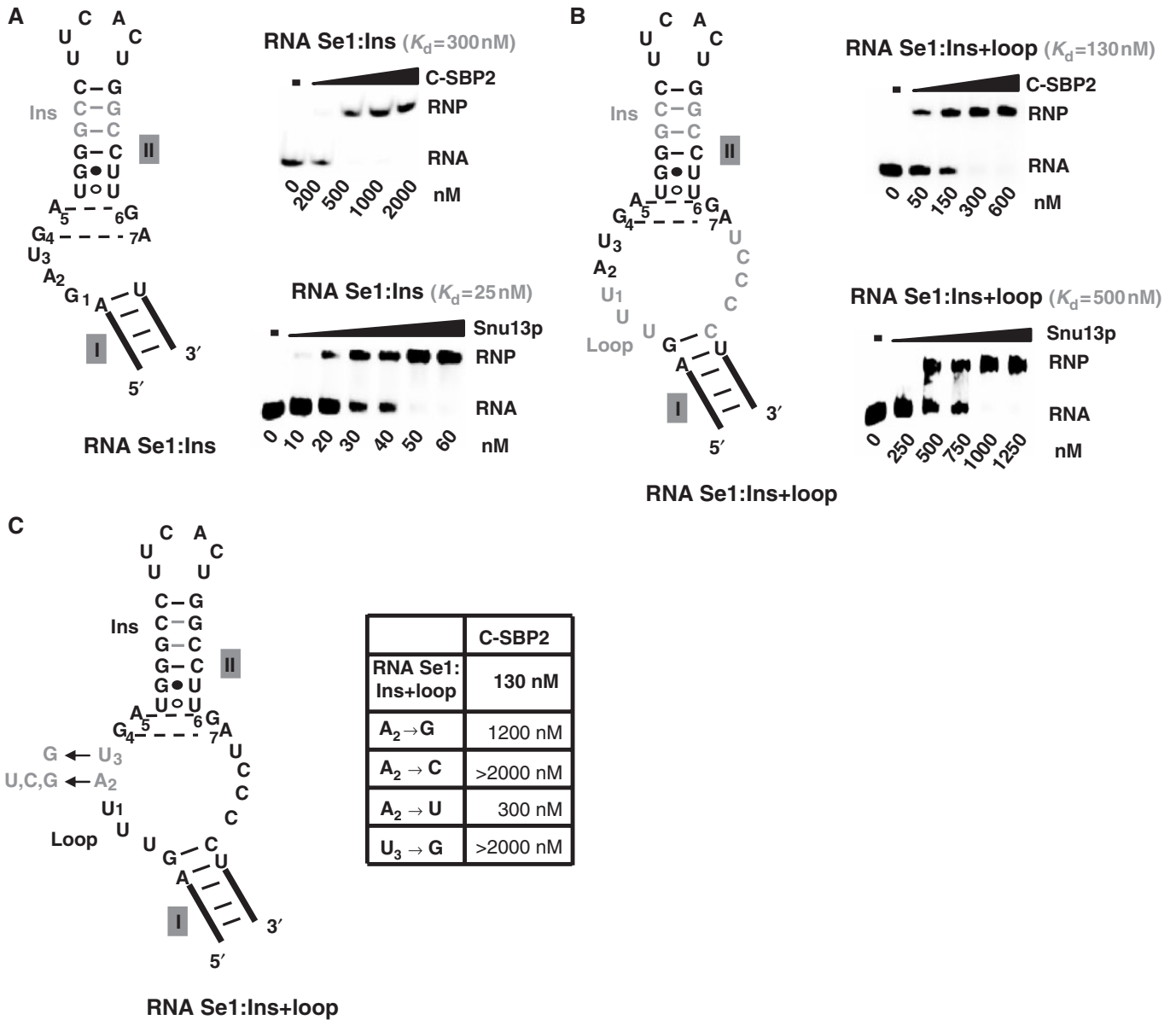
#### A large internal loop in the RNA confers a higher stability to C-SBP2-RNA complexes

Based on gel-shift experiments, Snu13p and C-SBP2 were found to have similar affinities for RNA SelN ( $K_d$ s of 180 and 160 nM, respectively) (Figure 1B). However, such apparent  $K_d$ s, established by gel-shift assays, mostly reflect the capacity of the RNA and protein partners to form a complex which is stable under electrophoresis conditions. Thus, for a better estimation of the stability of the RNP complexes, we used competition experiments. Complexes were formed, as above, with radiolabeled RNA and a protein concentration about twice that of the apparent  $K_d$ s (300 nM for C-SBP2 and 1000 nM for Snu13p, for assays on Se1:Ins+loop RNA, and two identical protein concentrations, 300 nM, for assays on SelN RNA). Cold RNA was added in excess to destabilize the complex. When complexes were formed with the Se1:Ins+loop RNA (Figure 7), a larger excess of cold Se1:Ins+loop RNA was required to dissociate C-SBP2-RNA complexes compared to Snu13p-RNA complexes and this in spite of the higher Snu13p concentration used to form the initial complex (Figure 7A). Furthermore, a much stronger difference was observed when complexes were formed with the SelN RNA: whereas a 1000-fold molar excess of SelN RNA was sufficient to destabilize the SelN-Snu13p complexes, dissociation of the SelN-C-SBP2 complexes required as much as a 40 000-fold excess of cold SelN RNA (Figure 7B). These observations revealed the high stability of complexes formed with C-SBP2.

Another approach to verify the high stability of the SelN RNA-C-SBP2 complexes was to destabilize the RNA-protein complex by addition of an excess of a competitor protein (C-SBP2 for complexes formed with Snu13p and vice versa). As seen in Figure 7C, even when added in large excess (65-fold) to the preformed SelN-C-SBP2 complex, Snu13p could not dissociate this complex. In contrast, when C-SBP2 was added at the same concentration as the Snu13p protein used to form the SelN-Snu13p complex, this complex was completely converted into a SelN RNA-C-SBP2 complex. This observation argues in favor of a strong specificity of C-SBP2 for the SECIS RNAs.

#### C-SBP2 protects a larger region of the Se1:Ins+loop and SelN RNAs than Snu13p

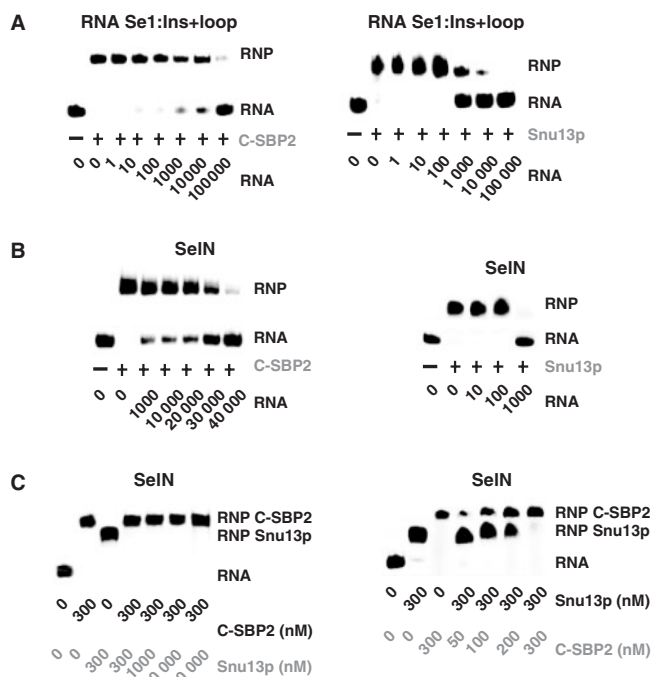
One possible explanation for the strong stability of complexes formed by protein C-SBP2 and the Se1:Ins+loop and SelN RNAs was the occurrence of more extended RNA-protein contacts in these complexes compared to those formed with Snu13p. To answer this



**Figure 6.** A K-turn motif with an extended internal loop increases C-SBP2 affinity. The variant Se1:Ins (A) and Se1:Ins+loop RNAs (B and C) are shown. The additional residues in these variant RNAs compared to Se1 RNA are shown in gray. The affinities of C-SBP2 and Snu13p for Se1:Ins (A) and Se1:Ins+loop (B) were tested by gel-shift assays. Complex formation was performed as described in Figure 1, using 5 fmol of 5'-end labeled RNA and increasing concentrations of C-SBP2 or Snu13p proteins. In Panels A and B, the apparent  $K_d$ s are indicated above the autoradiograms. (C) The base substitutions generated at positions 2 and 3 in the internal loop of the Se1:Ins+loop RNA are indicated in gray. The table gives the apparent  $K_d$  values established by gel-shift assays for complexes formed between C-SBP2 and the variant Se1:Ins+loop RNAs.

question, we probed the RNA accessibilities in the six RNP complexes formed by the Se1, Se1:Ins+loop and Se1N RNAs and each of the C-SBP2 and Snu13p proteins. We used T1 and T2 RNases under conditions such that they cleaved single-stranded regions, and V1 RNase that cleaves specifically double-stranded and stacked RNA regions. Very similar RNA protections were obtained for complexes formed by RNA Se1 and each of the proteins (Figure 8). Both proteins protected the bulge sequence, part of helix II and the 5' strand of helix I. In contrast, protections of RNAs Se1:Ins+loop and Se1N by Snu13p were very limited compared to those found for C-SBP2. Thus, with RNAs containing an

extended internal loop, the architecture of C-SBP2 allows tight RNA-protein contacts with both helices and the 5' strand of the internal loop, which is not the case for Snu13p. Interestingly also, the sensitivity to V1 RNase of the 3' strand of helix I was strongly increased by binding of C-SBP2 or Snu13p to RNA Se1. The same situation was observed upon binding of C-SBP2 to RNA Se1:Ins+loop (Figure 8). This effect was quite less marked upon Snu13p binding on this RNA. Altogether, this suggested the occurrence of a profound RNA conformational change when Snu13p or C-SBP2 bind RNA Se1 and when C-SBP2 binds RNA Se1:Ins+loop. This strong RNA conformational change is probably not



**Figure 7.** C-SBP2 forms highly stable complexes with RNAs containing an extended internal loop. The stabilities of the complexes formed between C-SBP2 and Snu13p and the Se1:Ins+loop (A) and SeIN (B) RNAs were tested by competition experiments. RNA–protein complexes were formed by using 5 fmol of 5′-end labeled Se1:Ins+loop or SeIN RNAs and C-SBP2 (300 nM) or Snu13p (1000 or 300 nM). The RNA–protein complexes were challenged with increasing concentrations of cold Se1:Ins+loop or SeIN RNAs (10–100 000- and 10–40 000-fold molar excess, respectively, as indicated below the lanes). The remaining complexes were fractionated by gel electrophoresis. (C) Comparison of the relative stabilities of the Snu13p–SeIN and C-SBP2–SeIN complexes. RNP complexes formed with C-SBP2 at 300 nM were challenged by addition of an excess of Snu13p protein and vice versa. Complexes formed with Snu13p at 300 nM were challenged by addition of an excess of C-SBP2. The remaining complexes were fractionated by gel electrophoresis. The identities and concentrations of the protein competitors used in the assays are indicated below the lanes.

induced upon binding of Snu13p to an RNA with a large internal loop. Binding of C-SBP2 to SeIN RNA also induced a hypersensitivity to V1 RNase, but the RNA segment concerned was different (extremity of the 5′ strand of helix II). No significant hypersensitivity to V1 RNase was observed upon Snu13p binding to SeIN RNA, which reinforces the idea that only C-SBP2 can establish tight contacts with RNAs containing a large internal loop and as a consequence remodel their conformation. The archaeal protein L7Ae is known to interact with both canonical K-turn and K-loop structures formed in terminal loops (9–11,15,25,48). Thus, by footprinting assays, we tested whether L7Ae can establish a tight interaction with the SeIN RNA, as does C-SBP2 (Figure 8). The apparent  $K_d$  established by gel-shift assays for the SeIN–L7Ae complex revealed a high affinity ( $K_d$  of 35 nM, not shown). According to enzymatic footprinting assays (Figure 8), this high affinity may be due to the presence of two L7Ae-binding sites in SeIN RNA: one of them corresponds to the quartet of non-Watson–Crick base

pairs, the other one to the terminal loop. Due to the presence of two G.A dinucleotides in this loop, a K-loop recognized by protein L7Ae can be formed. Interestingly, the protections found in the 5′ strand of the internal loop, helix I, and the quartet of non-Watson–Crick base pairs, are very similar in the C-SBP2–SeIN and L7Ae–SeIN complexes. Protein L7Ae protects two additional residues in the 3′ strand of the internal loop as compared to C-SBP2. Hence, concerning the recognition of RNAs with an internal loop, the behavior of protein L7Ae is closer to that of C-SBP2 than that of Snu13p.

### Mutations in helix II of SeIN RNA limit C-SBP2 affinity

Since our data suggested a functional importance of helix II for C-SBP2 binding, we tested the effects of mutations in helix II of the authentic SeIN SECIS motif on C-SBP2 binding. Substitution of the fifth G.U pair in helix II by a C–G pair as well as substitution of the sixth G–C pair by C–G pair, had less negative effects on C-SBP2 binding (factor of 2) (Figure 9) compared to those found for the corresponding substitution in RNA Se1 (factor of 4) (Figure 5). However, substitutions of the sixth G–C pair and of the seventh C–G pair by G.G pairs had strong negative effects on C-SBP2 binding ( $K_d$ s of 800 and 780 nM instead of 160 nM for the WT RNA). Therefore, mutations in an authentic SECIS RNA confirmed our observation of the importance of the stability and the sequence of helix II for C-SBP2 binding. Accordingly, Pu–Py pairs are the most frequently observed base pairs at the fifth and sixth positions in helix II of SECIS elements (Figure 3B).

## DISCUSSION

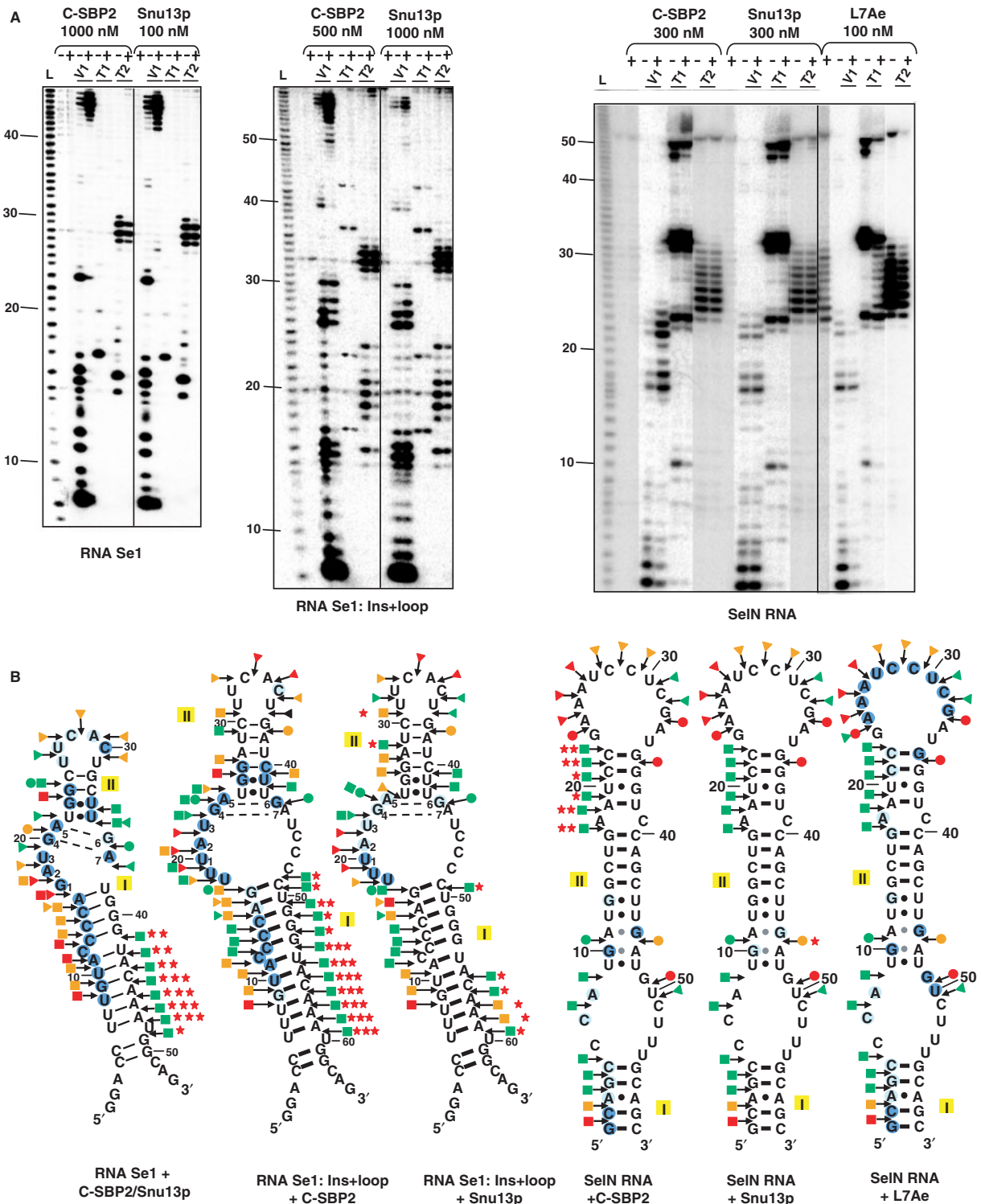
The present data based on SELEX and site-directed mutagenesis experiments improve our understanding of the sequence and structural features required for efficient interaction of SBP2 with RNA. These findings bring new insights that will facilitate the understanding of its mechanism of action in the selenocysteine incorporation machinery.

When used for studying RNA–protein interactions, the SELEX approach most generally leads to the establishment of an RNA consensus sequence. Here, despite the wide diversity of the initial RNA mixture ( $18^4$ ), only seven different sequences were selected, and several of them were very similar. All of them folded into very similar 2D structures that contained a canonical K-turn motif. This limited diversity of the selected sequences indicated narrow RNA structure requirements for efficient binding of SBP2. We confirmed this hypothesis by several experimental approaches.

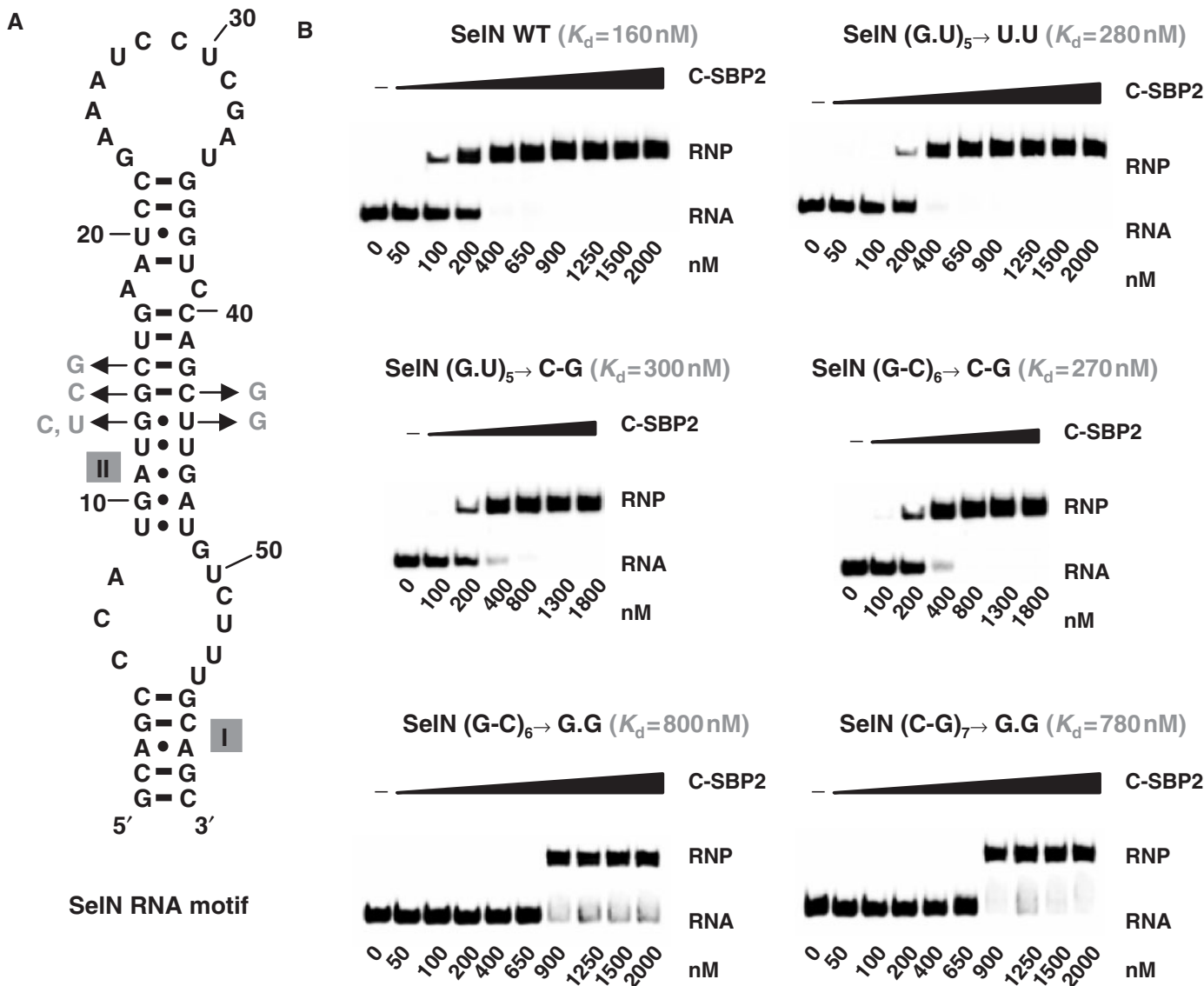
### Dual $Mg^{++}$ dependence of SBP2 binding to different RNA substrates

Earlier work (49) established that SBP2 contained in testis extracts displayed high sensitivity to  $Mg^{++}$  concentration for SECIS binding, the  $IC_{50}$  being around 4 mM. This sensitivity was however less pronounced ( $IC_{50} > 20$  mM) with a shorter, recombinant version of





**Figure 8.** C-SBP2 protects a larger region of the Se1:Ins+loop and SeIN RNAs than Snu13p. (A) The *in vitro* transcribed 5'-end labeled Se1, Se1:Ins+loop and SeIN RNAs (25 fmol) were incubated in the absence (–) or presence (+) of C-SBP2, Snu13p or L7Ae. The protein concentrations used in the assays are indicated above each panel, 2 µg of tRNA were added in each assay and the digestion was carried out for 6 min at 20°C, in buffer D, in the presence of 0.8 U RNase T1, 2.4 U RNase T2 or 0.001 U RNase V1, as described in the Materials and Methods section. The cleavage products were fractionated on a 10% polyacrylamide-8M urea gel. L: alkaline hydrolysis. Nucleotide positions are indicated on the left. (B) Schematic representation of the results shown in panel A on the secondary structures proposed for the three studied RNAs. Helices I and II are indicated. V1, T1 and T2 RNase cleavages are represented by arrows surmounted of squares, dots and triangles, respectively. The color of symbols reflects the intensity of cleavages (green, orange and red for low, medium and strong, respectively). Nucleotides with decreased sensitivity to RNase in the presence of the proteins are circled in blue (pale and dark for low and strong protection, respectively). Nucleotides with increased sensitivity to RNase in the presence of the proteins are indicated by a red star. The number of stars reflects the increased sensitivity to cleavage (one, two and three representing low, medium and strong, respectively).



**Figure 9.** The sequence and stability of helix II are important for C-SBP2 binding onto SeIN RNA. (A) The base-pair substitutions generated at positions 5, 6 and 7 in helix II of the SeIN RNA are shown. (B) Complexes were formed with 5 fmol of radiolabeled WT or mutated SeIN RNA and increasing concentrations of the C-SBP2 protein (from 50 to 2000 nM). The RNP complexes were fractionated on 6% polyacrylamide 8-M urea gel and apparent dissociation constants were determined by measuring the radioactivity in the bands of gel corresponding to free RNA and the RNP. The determined  $K_d$ s are indicated above each autoradiogram.

SBP2, and PHGPx SECIS RNA as the RNA partner (42). Interestingly, here we found that binding of C-SBP2 to the SeI RNA, which forms a canonical K-turn structure, requires a 1.5 mM  $Mg^{++}$  concentration, higher concentrations being innocuous. At first glance, the two series of results may appear contradictory. Nevertheless, these differential behaviors toward  $Mg^{++}$  are likely explained by the use of different RNA partners. SeI RNA is a genuine K-turn RNA, and it is known that divalent cations favor the closed conformation of canonical K-turn motifs (50). SECIS RNAs possess a large internal loop and thus contain a K-turn like motif (32). A high  $Mg^{++}$  concentration may induce a conformational change into SECIS RNAs, which is not favorable for SBP2 binding. For instance, based on our data, we can imagine that a

high  $Mg^{++}$  concentration promotes closing of the internal loop, and we show that a large internal loop is needed for maximum binding efficiency of SBP2. The SeI RNA is a typical Snu13p partner. As expected, no marked  $Mg^{++}$  requirement is observed for Snu13p binding to this RNA. In contrast, as SeI RNA does not contain an internal loop, a prior stabilization of the kink structure may be needed to reinforce SBP2 binding. Altogether, the previous and present data strongly suggest that each member of the L7A/L30 family is perfectly suited for binding to its authentic partner at the physiological concentration of divalent cations. When RNA partners are exchanged in *in vitro* experiments, the  $Mg^{++}$  ion concentration has to be adapted in order to form the heterologous interaction. Accordingly, a high  $Mg^{++}$  ion concentration

was found to be required for efficient *in vitro* binding of protein L30 to a SECIS element, in the presence of SBP2 (42).

### Specific sequence requirements in helix II

Site-directed mutagenesis, performed on the winner Se1 RNA obtained by SELEX experiments, demonstrated that binding of C-SBP2 requires the presence of a stable helix II containing at least two Watson–Crick base pairs. In agreement with this observation, all the SECIS elements identified so far contain a series of Watson–Crick base pairs on top of the non-Watson–Crick base-pair quartet (30,51–53; A.K., unpublished data). Accordingly, we showed that their individual disruption in SelN RNA decreases C-SBP2 affinity. Not only helix II stability but also its sequence has an influence on C-SBP2 affinity. The presence of a Pu–Py pair at the fourth position in helix II was found to be of high importance for C-SBP2 binding to Se1 RNA, a Pu–Py pair at this position being also more favorable for C-SBP2 binding to SelN RNA. This is in contrast with the absence of sequence requirement in helix II, except for the G.A and A.G base pairs and the adjacent U.U pair found for proteins Snu13p/15.5K and L7Ae (6,7,16,24,47,54–56). Up to now, little attention was given to the importance of the identity of base pairs in the upper part of helix II of SECIS elements. However, at position 4 of helix II, a Pu residue (most frequently a G residue) is almost always found in the 5' strand and a Py residue (most frequently a C residue) is observed in the 3' strand (30,51–53; A.K., unpublished data). Although less strictly conserved, the fifth base pair in helix II is predominantly a Pu–Py pair (Figure 3B). Together with our experimental data, these phylogenetic observations strongly suggest a functional importance of these conserved Pu.Py base pairs at positions 4 and 5 in helix II. In accordance with this hypothesis, the G–C pair at the fourth position in RNA Se1 was protected against V1 RNase digestion in the complex formed with C-SBP2, but not in the complex formed with Snu13p (Figure 8). Accordingly, the very limited V1 RNase cleavage, which is located between residues G13 and G14 in free SelN RNA, disappeared in the presence of C-SBP2, but not with Snu13p. Remarkably, this V1 RNase cleavage was also abolished in the presence of protein L7Ae.

Comparison of the Se1 to Se7 RNAs and site-directed mutagenesis of the Se1 RNA also show the importance for a non-Watson–Crick base pair on top of the A.G and G.A pair tandem (Figures 4 and 5). Accordingly, U.U pairs are frequently encountered pairs at this position in SECIS elements (30,51–53; A.K., unpublished data) and a U.U pair was also preferentially selected at this position of helix II, in the SELEX experiment performed with Snu13p. The presence of a U.U base pair at this position, with a C1'–C1' distance of the ribose ring close to that in G.A pairs, is very likely required to favor the smooth transition from the non-Watson–Crick to the Watson–Crick section of helix II. Noticeably also, in K-turn structures found in ribosomal RNAs, the nucleobase of one of this unpaired couple of nucleotides

was proposed to interact with one nucleobase in helix I, and thus to reinforce the inter-helical angle between helix I and helix II (57).

### Importance of a large SECIS internal loop

Increasing the size of both helix II and the internal loop of the winner Se1 RNA obtained by SELEX, yielded an RNA with an affinity similar to that of SelN RNA (Figure 6B). Such an RNA could not be obtained in the SELEX experiment, because of limitation in size of the degenerated sequence that can be used (18 nt) in these experiments. Indeed, due to the necessity to cover all the possible sequences during the screening, one cannot use largely extended degenerated sequences (58,59). In agreement with the importance of the size of helix II, all the identified SECIS elements contain a long helix II. Based on our footprinting data, the high affinity of C-SBP2 for RNAs with an internal loop, as well as the stability of the complexes formed, are due to its capacity to contact helices I and II and the 5' strand of the internal loop in these RNAs (Figure 8). Interestingly, Martin *et al.* (38) showed that closing of the internal loop of the rat D1 SECIS element almost completely abolished selenocysteine incorporation *in vivo*. In agreement with the observed requirement of at least one base pair closing the 3-nt bulge loop of K-turn motifs for efficient binding of Snu13p (44), Snu13p establishes very loose contacts with RNAs containing an internal loop. The presence of a closing base pair is not required for L7Ae and this protein is able to bind open K-loop structures (9,44,48). The presence of an arginine at position 95 in the 15.5K/Snu13p protein, that forms a salt bridge with the 5' phosphate of the residue at position 1 in the bulge, and its replacement by a valine in L7Ae, were proposed to explain this difference between proteins 15.5K/Snu13p and L7Ae (60). Interestingly, like L7Ae, SBP2 contains a valine at the corresponding position in the L7A/L30 domain (29). This may explain our observation of similar binding properties of proteins SBP2 and L7Ae on RNAs containing a large internal loop.

In free RNAs containing a canonical K-turn structure with a bulge loop, helices I and II form a 76° angle. Upon Snu13p/15.5K binding, the RNA undergoes further folding, so that the helix I–helix II angle is reduced to 48° (56,61). This folding likely explains the tight contact of Snu13p with both helices of RNA Se1 that we detected by footprinting assay. As very similar footprinting results were obtained with C-SBP2, it probably also induces a folding of this RNA. However, C-SBP2 but not Snu13p may induce a similar folding in both the Se1:Ins+loop and SelN RNAs.

### Sequence requirement in the internal loop

Our site-directed mutagenesis experiments on the Se1:Ins+loop RNA revealed the importance of the identity of residues at positions 2 and 3 in the internal loop for C-SBP2 binding. The most deleterious base substitution at position 2 was the A to C replacement. Interestingly, a 66% decrease of selenocysteine incorporation was observed when the same A to C substitution

was generated in the SECIS element of the rat GPx mRNA while A to G and A to U changes only led to a loss of 30 and 22% of the incorporation, respectively (52). In accordance with the decrease of the C-SBP2 affinity upon U to G substitution at position 3 in RNA Se1:Ins+loop, selenocysteine incorporation was decreased by 88% when this base substitution was generated in the SECIS element of the rat GPx mRNA (38). In addition, a U to C substitution at this position abolished the binding of SBP2 to SeIN RNA and is responsible for a human genetic disease, the rigid spine muscular dystrophy (62). As the residue at position 3 in canonical K-turns is located in the protein pocket, its mutation also has a deleterious effect on 15.5K/Snu13p and L7Ae binding (6,24–26). Residues E61 and K86 in 15.5K, and D54 and K79 in *Archaeoglobus fulgidus* L7Ae, are involved in the interaction with the nucleobase at position 3. Their counterparts in SBP2 (E699 and R730) probably play a similar role, since they are crucial for binding to SECIS RNAs (5,29,56). The specificity of L7Ae/L30 protein members towards the residue at position 2 in the K-turn motif is variable. Whereas an A or G residue at position 2 increases 15.5K/Snu13p affinity, substitutions at position 2 have no marked effect on L7Ae affinity (25). Concerning position 2, SBP2 exhibits a behavior closer to that of 15.5K/Snu13p than to L7Ae.

The ribosomal protein L30 was recently shown to compete with SBP2 for binding to SECIS RNA (42). L30 was found to recognize a K-turn structure of its pre-mRNA that contains a protruding A residue in a small internal loop (21–23). SELEX experiments performed with L30 revealed its preference for K-turn motifs with protruding C or A residues (27). Later, it was shown that L30 also has the ability to accommodate K-turn structures with a protruding G (28). However, its interaction with K-turn motifs containing a protruding U residue has not been demonstrated yet. Consequently, the binding of L30 to SECIS elements, that all contain a U residue at position 3, raises the question of how it can achieve this.

## CONCLUSION

Assembly of the selenocysteine incorporation machinery is proposed to be initiated by SBP2 association to SECIS elements in the nucleus, and more likely in the nucleolus (63). Protein 15.5K/Snu13p is abundant in the nucleolus and SBP2 shares several common features with Snu13/15.5K. However, our data reveal important differences in RNA specificities that may ensure the specific association of SBP2 to SECIS elements in the nuclear compartment.

## ACKNOWLEDGEMENTS

V. Senty-Ségault and S. Massenet are acknowledged for helpful discussions. S. Sonkaria is thanked for careful reading of the manuscript. A. Cléry was a fellow from the French ‘Ministère de la Recherche et des Nouvelles Technologies’. The work was supported by the Centre National de la Recherche Scientifique, the French

‘Ministère de la Recherche et des Nouvelles Technologies’, the ACI ‘Biologie Cellulaire, Moléculaire et Structurale’ n°BCMS226, the PRST ‘Bioingénierie’ of the ‘Conseil Régional Lorrain’ and the ToxNuc-E Programme ‘Toxicologie Nucléaire Environnementale’. Funding to pay the Open Access publication charge was provided by CNRS-Sciences de la vie.

*Conflict of interest statement.* None declared.

## REFERENCES

- Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
- Wimberly,B.T., Brodersen,D.E., Clemons,W.M.Jr, Morgan-Warren,R.J., Carter,A.P., Vonnrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
- Schluenzen,F., Tocilj,A., Zarivach,R., Harms,J., Gluehmann,M., Janell,D., Bashan,A., Bartels,H., Agmon,I. *et al.* (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, **102**, 615–623.
- Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
- Vidovic,I., Nottrott,S., Hartmuth,K., Luhrmann,R. and Ficner,R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell*, **6**, 1331–1342.
- Nottrott,S., Hartmuth,K., Fabrizio,P., Urlaub,H., Vidovic,I., Ficner,R. and Luhrmann,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.
- Watkins,N.J., Segault,V., Charpentier,B., Nottrott,S., Fabrizio,P., Bachi,A., Wilm,M., Rosbash,M., Branlant,C. *et al.* (2000) A common core RNP structure shared between the small nuclear box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, **103**, 457–466.
- Terns,M.P. and Terns,R.M. (2002) Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr.*, **10**, 17–39.
- Charpentier,B., Muller,S. and Branlant,C. (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res.*, **33**, 3133–3144.
- Omer,A.D., Ziesche,S., Ebhardt,H. and Dennis,P.P. (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 5289–5294.
- Rozhdetsvensky,T.S., Tang,T.H., Tchirkova,I.V., Brosius,J., Bachellerie,J.P. and Huttenhofer,A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.
- Granneman,S., Puijij,G.J., Horstman,W., van Venrooij,W.J., Luhrmann,R. and Watkins,N.J. (2002) The hU3-55K protein requires 15.5K binding to the box B/C motif as well as flanking RNA elements for its association with the U3 small nucleolar RNA in Vitro. *J. Biol. Chem.*, **277**, 48490–48500.
- Nottrott,S., Urlaub,H. and Luhrmann,R. (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.*, **21**, 5527–5538.
- Watkins,N.J., Dickmanns,A. and Luhrmann,R. (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol. Cell Biol.*, **22**, 8342–8352.
- Baker,D.L., Youssef,O.A., Chastkofsky,M.I., Dy,D.A., Terns,R.M. and Terns,M.P. (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev.*, **19**, 1238–1248.
- Tran,E.J., Zhang,X. and Maxwell,E.S. (2003) Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J.*, **22**, 3930–3940.

17. Rashid,R., Aittaleb,M., Chen,Q., Spiegel,K., Demeler,B. and Li,H. (2003) Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. *J. Mol. Biol.*, **333**, 295–306.
18. Bortolin,M.L., Bachellerie,J.P. and Clouet-d'Orval,B. (2003) In vitro RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNA(Trp). *Nucleic Acids Res.*, **31**, 6524–6535.
19. Schultz,A., Nottrott,S., Hartmuth,K. and Luhrmann,R. (2006) RNA structural requirements for the association of the spliceosomal hPrp31 protein with the U4 and U4atac small nuclear ribonucleoproteins. *J. Biol. Chem.*, **281**, 28278–28286.
20. 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.
21. Vilardell,J. and Warner,J.R. (1994) Regulation of splicing at an intermediate step in the formation of the spliceosome. *Genes Dev.*, **8**, 211–220.
22. Chao,J.A. and Williamson,J.R. (2004) Joint X-ray and NMR refinement of the yeast L30e-mRNA complex. *Structure*, **12**, 1165–1176.
23. 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.
24. Kuhn,J.F., Tran,E.J. and Maxwell,E.S. (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res.*, **30**, 931–941.
25. Charron,C., Manival,X., Clery,A., Senty-Segault,V., Charpentier,B., Marmier-Gourrier,N., Branlant,C. and Aubry,A. (2004) The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J. Mol. Biol.*, **342**, 757–773.
26. Marmier-Gourrier,N., Clery,A., Senty-Segault,V., Charpentier,B., Schlotter,F., Leclerc,F., Fournier,R. and Branlant,C. (2003) A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA*, **9**, 821–838.
27. Li,H. and White,S.A. (1997) RNA aptamers for yeast ribosomal protein L32 have a conserved purine-rich internal loop. *RNA*, **3**, 245–254.
28. White,S.A., Hoeger,M., Schweppe,J.J., Shillingford,A., Shipilov,V. and Zarutskie,J. (2004) Internal loop mutations in the ribosomal protein L30 binding site of the yeast L30 RNA transcript. *RNA*, **10**, 369–377.
29. Allmang,C., Carbon,P. and Krol,A. (2002) The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *RNA*, **8**, 1308–1318.
30. Fletcher,J.E., Copeland,P.R., Driscoll,D.M. and Krol,A. (2001) The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA*, **7**, 1442–1453.
31. Walczak,R., Carbon,P. and Krol,A. (1998) An essential non-Watson–Crick base pair motif in 3'UTR to mediate selenoprotein translation. *RNA*, **4**, 74–84.
32. Allmang,C. and Krol,A. (2006) Selenoprotein synthesis: UGA does not end the story. *Biochimie*, **88**, 1561–1571.
33. Krol,A. (2002) Evolutionarily different RNA motifs and RNA-protein complexes to achieve selenoprotein synthesis. *Biochimie*, **84**, 765–774.
34. Gladyshev,V.N. (2001). Selenium in biology and human health: controversies and perspectives. In Hatfield,D.L. (ed), *Selenium: Its Molecular Biology and Role in Human Health*. Kluwer, Boston, pp. 313–317.
35. Flohe,L., Andreesen,J.R., Brigelius-Flohe,R., Maiorino,M. and Ursini,F. (2000) Selenium, the element of the moon, in life on earth. *IUBMB Life*, **49**, 411–420.
36. Rayman,M.P. (2000) The importance of selenium to human health. *Lancet*, **356**, 233–241.
37. Rederstorff,M., Krol,A. and Lescure,A. (2006) Understanding the importance of selenium and selenoproteins in muscle function. *Cell Mol. Life Sci.*, **63**, 52–59.
38. Martin,G.W.III, Harney,J.W. and Berry,M.J. (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.
39. Tujebajeva,R.M., Copeland,P.R., Xu,X.M., Carlson,B.A., Harney,J.W., Driscoll,D.M., Hatfield,D.L. and Berry,M.J. (2000) Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep.*, **1**, 158–163.
40. Fagegaltier,D., Hubert,N., Yamada,K., Mizutani,T., Carbon,P. and Krol,A. (2000) Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J.*, **19**, 4796–4805.
41. Zavacki,A.M., Mansell,J.B., Chung,M., Klimovitsky,B., Harney,J.W. and Berry,M.J. (2003) Coupled tRNA(Sec)-dependent assembly of the selenocysteine decoding apparatus. *Mol. Cell*, **11**, 773–781.
42. Chavatte,L., Brown,B.A. and Driscoll,D.M. (2005) Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat. Struct. Mol. Biol.*, **12**, 408–416.
43. Mougou,A., Gottschalk,A., Fabrizio,P., Luhrmann,R. and Branlant,C. (2002) Direct probing of RNA structure and RNA-protein interactions in purified HeLa cells and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J. Mol. Biol.*, **317**, 631–649.
44. Charron,C., Manival,X., Charpentier,B., Branlant,C. and Aubry,A. (2004) Purification, crystallization and preliminary X-ray diffraction data of L7Ae sRNP core protein from *Pyrococcus abyssi*. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 122–124.
45. Lescure,A., Allmang,C., Yamada,K., Carbon,P. and Krol,A. (2002) cDNA cloning, expression pattern and RNA binding analysis of human selenocysteine insertion sequence (SECIS) binding protein 2. *Gene*, **291**, 279–285.
46. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
47. Clery,A., Senty-Segault,V., Leclerc,F., Raue,H.A. and Branlant,C. (2007) Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. *Mol. Cell Biol.*, **27**, 1191–1206.
48. Nolivos,S., Carposis,A.J. and Clouet-d'Orval,B. (2005) The K-loop, a general feature of the *Pyrococcus* C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Res.*, **33**, 6507–6514.
49. Copeland,P.R. and Driscoll,D.M. (1999) Purification, redox sensitivity, and RNA binding properties of SECIS-binding protein 2, a protein involved in selenoprotein biosynthesis. *J. Biol. Chem.*, **274**, 25447–25454.
50. Matsumura,S., Ikawa,Y. and Inoue,T. (2003) Biochemical characterization of the kink-turn RNA motif. *Nucleic Acids Res.*, **31**, 5544–5551.
51. Kryukov,G.V., Castellano,S., Novoselov,S.V., Lobanov,A.V., Zehetab,O., Guigo,R. and Gladyshev,V.N. (2003) Characterization of mammalian selenoproteomes. *Science*, **300**, 1439–1443.
52. Fagegaltier,D., Lescure,A., Walczak,R., Carbon,P. and Krol,A. (2000) Structural analysis of new local features in SECIS RNA hairpins. *Nucleic Acids Res.*, **28**, 2679–2689.
53. Lescure,A., Gautheret,D., Carbon,P. and Krol,A. (1999) Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif. *J. Biol. Chem.*, **274**, 38147–38154.
54. Szewczak,L.B., DeGregorio,S.J., Strobel,S.A. and Steitz,J.A. (2002) Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem. Biol.*, **9**, 1095–1107.
55. Szewczak,L.B., Gabrielsen,J.S., Degregorio,S.J., Strobel,S.A. and Steitz,J.A. (2005) Molecular basis for RNA kink-turn recognition by the h15.5K small RNP protein. *RNA*, **11**, 1407–1419.
56. Moore,T., Zhang,Y., Fenley,M.O. and Li,H. (2004) Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure*, **12**, 807–818.
57. Razga,F., Spackova,N., Reblova,K., Koca,J., Leontis,N.B. and Sponer,J. (2004) Ribosomal RNA kink-turn motif – a flexible molecular hinge. *J. Biomol. Struct. Dyn.*, **22**, 183–194.

58. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
59. Klug, S.J. and Famulok, M. (1994) All you wanted to know about SELEX. *Mol. Biol. Rep.*, **20**, 97–107.
60. Hamma, T. and Ferre-D'Amare, A.R. (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure*, **12**, 893–903.
61. Wozniak, A.K., Nottrott, S., Kuhn-Holsken, E., Schroder, G.F., Grubmuller, H., Luhrmann, R., Seidel, C.A. and Oesterhelt, F. (2005) Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. *RNA*, **11**, 1545–1554.
62. Allamand, V., Richard, P., Lescure, A., Ledeuil, C., Desjardin, D., Petit, N., Gartioux, C., Ferreiro, A., Krol, A. *et al.* (2006) A single homozygous point mutation in a 3' untranslated region motif of selenoprotein N mRNA causes SEPNI-related myopathy. *EMBO Rep.*, **7**, 450–454.
63. de Jesus, L.A., Hoffmann, P.R., Michaud, T., Forry, E.P., Small-Howard, A., Stillwell, R.J., Morozova, N., Harney, J.W. and Berry, M.J. (2006) Nuclear assembly of UGA decoding complexes on selenoprotein mRNAs: a mechanism for eluding nonsense-mediated decay? *Mol. Cell. Biol.*, **26**, 1795–1805.



# The yeast exosome and human PM-Scl are related complexes of 3' → 5' exonucleases

Christine Allmang,<sup>1</sup> Elisabeth Petfalski,<sup>1</sup> Alexandre Podtelejnikov,<sup>2</sup> Matthias Mann,<sup>2</sup> David Tollervey,<sup>1,3</sup> and Philip Mitchell<sup>1</sup>

<sup>1</sup>Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR UK; <sup>2</sup>CEBI Odense University, DK-5230 Odense M, Denmark

We previously identified a complex of 3' → 5' exoribonucleases, designated the exosome, that is expected to play a major role in diverse RNA processing and degradation pathways. Further biochemical and genetic analyses have revealed six novel components of the complex. Therefore, the complex contains 11 components, 10 of which are predicted to be 3' → 5' exoribonucleases on the basis of sequence homology. Human homologs were identified for 9 of the 11 yeast exosome components, three of which complement mutations in the respective yeast genes. Two of the newly identified exosome components are homologous to known components of the PM-Scl particle, a multisubunit complex recognized by autoimmune sera of patients suffering from polymyositis-scleroderma overlap syndrome. We demonstrate that the homolog of the Rrp4p exosome subunit is also a component of the PM-Scl complex, thereby providing compelling evidence that the yeast exosome and human PM-Scl complexes are functionally equivalent. The two complexes are similar in size, and biochemical fractionation and indirect immunofluorescence experiments show that, in both yeast and humans, nuclear and cytoplasmic forms of the complex exist that differ only by the presence of the Rrp6p/PM-Scl100 subunit exclusively in the nuclear complex.

[Key Words: Exoribonucleases; exosome; polymyositis-scleroderma; *RRP4*]

Received April 21, 1999; revised version accepted July 2, 1999.

The *RRP4* gene was identified initially in the yeast *Saccharomyces cerevisiae*, via a mutation that resulted in defective pre-rRNA processing (Mitchell et al. 1996). Biochemical analyses revealed that Rrp4p is a component of a protein complex that was designated the exosome (Mitchell et al. 1997). Initial characterization identified five components of the exosome; Rrp4p, Rrp41p (Ski6p), Rrp42p, Rrp43p, and Rrp44p (Dis3p). Of these, recombinant Rrp4p, Rrp41p, and Rrp44p were each demonstrated to have 3' → 5' exonuclease activity in vitro (Mitchell et al. 1997). The in vitro activities shown by the recombinant proteins were not, however, identical. Rrp4p is a distributive, hydrolytic enzyme, Rrp44p is a processive, hydrolytic enzyme, and Rrp41p is a processive, phosphorolytic enzyme. Consistent with this activity, Rrp44p is homologous to *Escherichia coli* RNase R (vacB), a member of the RNase II family of processive, hydrolytic exonucleases (Cheng et al. 1998), whereas Rrp41p is homologous to *E. coli* RNase PH, a phosphorolytic exonuclease (Mian 1997; Mitchell et al. 1997). Rrp42p and Rrp43p are also homologous to RNase PH (Mian 1997; Mitchell et al. 1997), and, therefore, the five initial mem-

bers of the complex were all known or strongly predicted to be 3' → 5' exonucleases. It was, however, notable that the purified exosome complex exhibited only a distributive, hydrolytic activity in vitro; no processive or phosphorolytic activities were observed (Mitchell et al. 1996, 1997). This observation suggested that a reason for the assembly of multiple activities into one complex might be to allow their coordinate repression in the absence of activation by specific cofactors.

In all eukaryotes, the mature 5.8S, 18S, 25S/28S rRNAs are generated from a single large pre-rRNA by post-transcriptional processing. The five components of the exosome that were identified initially were all shown to be required for the 3' processing of the 7S pre-rRNA to the mature 5.8S rRNA; genetic depletion of each gave a very similar processing defect, which closely resembled that seen in the original *rrp4-1* mutation (Mitchell et al. 1996, 1997). Subsequent analyses revealed that the exosome functions not only as an RNA processing complex but is also required for specific RNA turnover pathways. The degradation of the excised spacer fragment extending from the 5' end of the 35S primary transcript to cleavage site A<sub>0</sub> within the 5' external transcribed spacer (5' ETS) region is defective in the *rrp4-1* strain and in strains depleted of Rrp4p,

<sup>3</sup>Corresponding author.  
E-MAIL d.tollervey@ed.ac.uk; FAX 131 650 7040.



Rrp41p, Rrp42p, Rrp43p, or Rrp44p (de la Cruz et al. 1998). A wider role for the exosome in RNA metabolism was revealed by analyses that showed that Rrp4p and Rrp41p (Ski6p) both function in the 3' → 5' pathway of mRNA degradation (Anderson and Parker 1998). From these observations, the exosome complex, or related complexes, were predicted to be present in both the nucleolus and the cytoplasm.

Expression of the human homolog of Rrp4p, hRrp4p, in yeast was shown to complement a *rrp4-1* mutation and glycerol gradient centrifugation indicated that hRrp4p was present in HeLa cell lysates in a complex of similar size to the yeast exosome (Mitchell et al. 1997). These data suggested that a complex homologous to the exosome was present in human cells.

A large number of human autoimmune diseases have been identified. Some of these, notably scleroderma, are associated with the development of antibodies directed against nucleolar epitopes (for review, see Reimer 1990). In a relatively rare autoimmune disease, polymyositis-scleroderma overlap syndrome (Reimer et al. 1986), patients frequently develop antibodies directed against a 100-kD protein, PM-Scl100 (Bluthner and Bautz 1992; Ge et al. 1992). Less frequently another protein, PM-Scl75 (Alderuccio et al. 1991), is also targeted. These two proteins are components of a large complex, designated the PM-Scl complex, that was estimated to have between 11 (Reimer et al. 1986) and 16 (Gelpi et al. 1990) components. Interestingly, PM-Scl100 is homologous to the *E. coli* 3' → 5' exoribonuclease, RNase D (Briggs et al. 1998), whereas PM-Scl75 shows homology to RNase PH (Mian 1997).

Here, we report the identification of six new components of the yeast exosome and characterize distinct nuclear and cytoplasmic forms of this complex. Two of the newly defined exosome subunits are homologous to the human PM-Scl100 and PM-Scl75 autoantigens, and these proteins are associated with the human homolog of another exosome component. Moreover, like the yeast exosome, related human complexes are localized in nucleus and cytoplasm. Together, these data provide strong evidence that the PM-Scl complex is directly homologous to the yeast exosome.

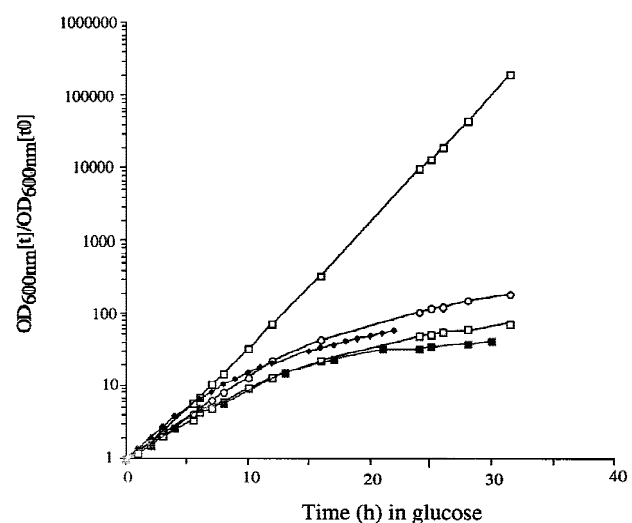
## Results

### Identification of new components of the exosome complex

The initial characterization of components that coprecipitated with protein A-tagged Rrp4p (ProtA-Rrp4p) identified four proteins (Mitchell et al. 1997). Three of these proteins, Rrp41p, Rrp42p, and Rrp43p, were homologous to *E. coli* RNase PH. However, the yeast genome contains three other putative open reading frames (ORFs) with homology to RNase PH; YDR280w (*RRP45*), YGR095c (*RRP46*; Mian 1997), and YGR158c (*MTR3*). The *RRP45* and *RRP46* ORFs were each precisely deleted in diploid strains of yeast (see Materials and Methods). On sporulation of each diploid, only two viable spores

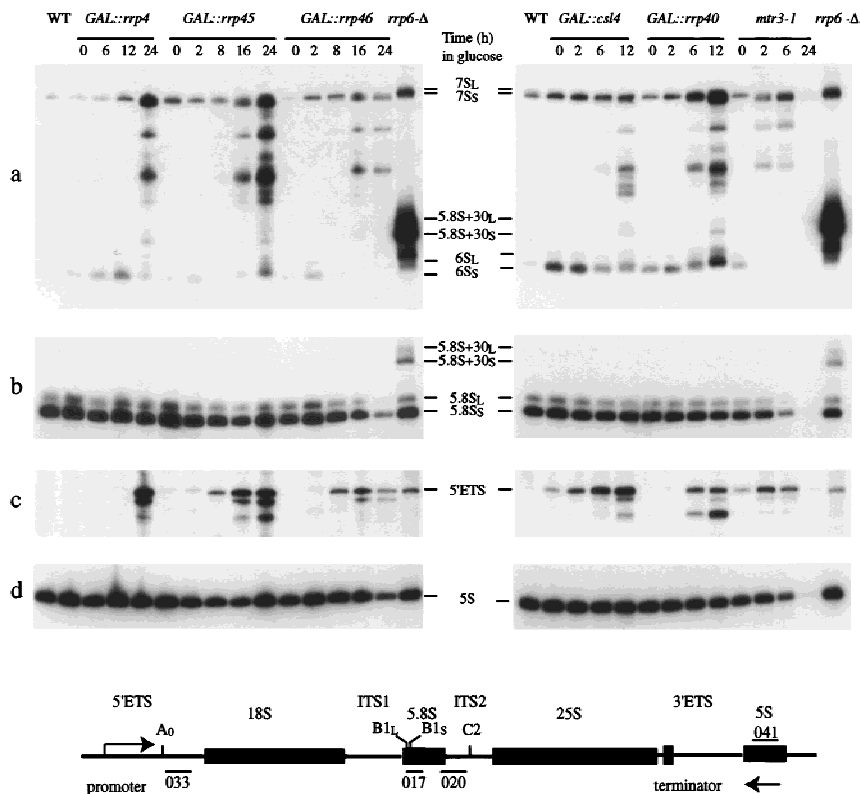
were recovered per tetrad and in each case the viable spores carried the wild-type allele. We conclude that *RRP45* and *RRP46* are both essential, at least for spore viability. Conditional alleles were constructed by placing *RRP45* and *RRP46* under the control of a repressible *GAL10* promoter (see Materials and Methods). In each case, the strains formed only microcolonies on solid medium containing 2% glucose (data not shown) and ceased growth following transfer from liquid RSG (raffinose/sucrose/galactose) medium to liquid glucose medium (Fig. 1). We conclude that Rrp45p and Rrp46p are essential for viability.

The strains depleted of Rrp45p or Rrp46p showed an accumulation of 3' extended forms of the 5.8S rRNA that extended in a ladder to the size of the 7S pre-rRNA but not beyond (Fig. 2). This phenotype is essentially identical to that seen in strains depleted for Rrp4p (Fig. 2a) or the four other components of the exosome identified previously (Mitchell et al. 1997). *Mtr3p* is essential for viability (Kadowaki et al. 1995), and a strain carrying a temperature-sensitive lethal *mtr3-1* allele (generously provided by A.M. Tartakoff, Case Western Reserve University, Cleveland, OH) was analyzed. This strain also accumulated 3' extended forms of the 5.8S rRNA after transfer to the nonpermissive temperature (37°C; Fig. 2a). The *mtr3-1* strain rapidly ceases growth following transfer to 37°C, and little pre-rRNA was recovered at the 24 hr time point, presumably because of the very low growth rate. In addition, the strains depleted of Rrp4p, Rrp45p, or Rrp46p or carrying *mtr3-1* each accumulated the excised 5' ETS region of the pre-rRNA, extending from the 5' end of the primary transcript to cleavage site



**Figure 1.** The newly identified components of the exosome complex are required for viability. Growth curves of *GAL*-regulated constructs following transfer to glucose medium. Strains were pregrown in permissive, RSG medium and transferred to repressive, glucose medium for the times indicated. Strains were maintained in exponential growth by dilution with prewarmed medium. Cell densities measured by  $OD_{600}$  are shown corrected for dilution. ( $\diamond$ ) Wild type; ( $\circ$ ) *GAL::rrp45*; ( $\square$ ) *GAL::rrp46*; ( $\blacklozenge$ ) *GAL::cs14*; ( $\blacksquare$ ) *GAL::rrp40*.

**Figure 2.** The newly identified components of the exosome complex are required for pre-rRNA processing. Northern analysis of processing of the 5.8S and degradation of the 5'ETS region of the pre-rRNA in exosome mutants. RNA was extracted from strains carrying *GAL*-regulated constructs following transfer from permissive, RSG medium to repressive, glucose medium for the times indicated, or from the *mtr3-1* strain following transfer from 25°C to 37°C for the times indicated. RNA was separated on a 6% polyacrylamide gel and hybridized with: (a) oligonucleotide 020 (complementary to the 5.8S/ITS2 boundary), (b) oligonucleotide 017 (hybridizing to the mature 5.8S rRNA), (c) oligonucleotide 033 (hybridizing to the 5'ETS around position +278). (d) oligonucleotide 041 (hybridizing to the 5S rRNA). The position of migration of the pre-rRNA species is indicated. The species labeled 5' ETS extends from the transcription start site to site A<sub>0</sub> (+610). Also shown is a cartoon of the rDNA (not to scale) with the mature rRNA regions as rectangles and the transcribed spacers as lines. The 18S, 5.8S, and 25S rRNAs are cotranscribed, separated by the internal transcribed spacers (ITS1 and ITS2) and flanked by the external transcribed spacers (5'ETS and 3'ETS). The 5S rRNA is independently transcribed in the opposite direction. The mature 5.8S rRNA is synthesized from the 7S pre-rRNA, which is 3' extended to site C<sub>2</sub> in ITS2. The 5' end of the 5.8S rRNA is generated by processing at sites at B<sub>1L</sub> and B<sub>1S</sub>, which lie about 8 nucleotides apart, generating 5.8S<sub>L</sub> and 5.8S<sub>S</sub>, respectively. Because this event precedes 3' processing, the 7S, 6S, and 5.8S + 30 pre-rRNAs all show 5' heterogeneity, generating, e.g., 6S<sub>L</sub> and 6S<sub>S</sub>.

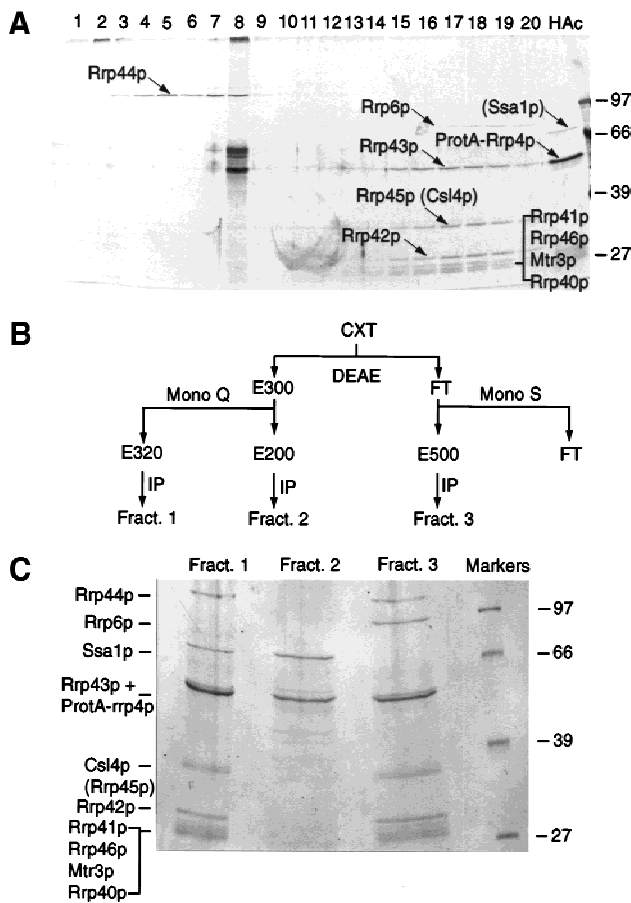


A<sub>0</sub> (Fig. 2c; 5' ETS), as well as degradation intermediates (see also de la Cruz et al. 1998). We conclude that the RNase PH homologs Rrp45p, Rrp46p, and Mtr3p are each required for the function of the exosome complex.

These observations prompted us to re-examine the biochemical purification of the exosome complex. A whole-cell extract from a strain expressing ProtA-Rrp4p under the control of the endogenous *RRP4* promoter from a low-copy-number CEN plasmid (Mitchell et al. 1997) was passed over an IgG-Sepharose column, and proteins were eluted from the bound IgG-ProtA-Rrp4p complex by use of a gradient of Mg<sup>2+</sup> (Görlich et al. 1996). Proteins were separated by SDS-PAGE, and bands were excised and subjected to sequencing analysis by mass spectrometry (see Kuster and Mann 1998; Shevchenko et al. 1996). Most bands were identified by high mass accuracy peptide mass mapping as described by Jensen et al. (1996). Several of the bands contained more than one gene product that were, however, identified without recourse to mass spectrometric peptide sequencing. In these cases, an iterative approach was used. First all tryptic peptide masses were searched against a comprehensive protein database, identifying one yeast protein. The peptide masses remaining after detailed comparison of the spectrum against the found sequence (second pass search), were again searched in the database,

yielding another yeast protein. In some cases MALDI peptide mapping did not unequivocally identify the components in a band. In these cases, nano-electrospray on a novel quadrupole Time of Flight instrument was performed (Shevchenko et al. 1997a; Wilm et al. 1996). Two broad peaks of eluted proteins were observed; Rrp44p eluted at around 500 mM MgCl<sub>2</sub> (Fig. 3A, lanes 4–6) whereas Rrp41p, Rrp42p, Rrp43p, Rrp45p, Rrp46p, and Mtr3p coeluted at around 1.6–1.8 M MgCl<sub>2</sub> (Fig. 3A, lanes 16–18). Two other proteins observed in the 1.6–1.8 M MgCl<sub>2</sub> fractions were identified as Rrp6p (YOR001w) and Rrp40p (YOL142w). The coelution of these components supports their presence in a single complex. ProtA-Rrp4p was eluted only in the acid wash of the column (Fig. 3A, lane HAC).

Following immunoprecipitation of ProtA-Rrp4p, all of the components were recovered with apparent stoichiometry, with the exception of Rrp6p, which was estimated from Coomassie staining to be approximately fivefold less abundant than the other components (data not shown). Because Rrp6p was eluted only at 1.6–1.8 M MgCl<sub>2</sub>, along with most other exosome components, it seemed unlikely that this low abundance was due to a weaker association with the exosome complex. Therefore, Rrp6p might be associated with only a subfraction of the exosome complex. To test this model, a whole-cell



**Figure 3.** Fractionation of the exosome complex and identification of new components. (A) Proteins associated with IgG-Sepharose via binding to ProtA-Rrp4p were eluted using a gradient of MgCl<sub>2</sub> and analyzed by SDS-PAGE. (Lanes 1–20) Material eluted with a 100 mM step gradient of MgCl<sub>2</sub> concentration from 100 mM (lane 1) to 2 M (lane 20). (HAc) Proteins eluted by the acid wash. Proteins are visualized by silver staining. The strong bands specifically seen in lane 8 were not observed in other experiments. (B) Purification scheme. A whole-cell extract (CXT) was batch-bound to DEAE-Sepharose FF. Bound material was eluted (E300) with TMN buffer containing 300 mM NaCl/10% glycerol (TMN-300). The eluate, in TMN-100, was passed through a Mono Q column and bound material was eluted stepwise with TMN-150, TMN-200 (E200), TMN-320 (E320), and TMN-500. Material that failed to bind to DEAE-Sepharose FF (FT) was passed through a Mono S column and bound material was eluted with TMN-500 (E500). Each sample was immunoprecipitated on IgG-Sepharose. (C) Proteins present in fractions 1, 2, and 3, obtained as outlined in B, were separated by SDS-PAGE. Approximately twofold more of the material recovered in fractions 2 and 3 was loaded onto the gel, as compared with fraction 1. Proteins positively identified by mass spectroscopy are indicated. Species in brackets were not identified in the preparations shown but are predicted to be present from other analyses. Molecular weight markers are also shown. Proteins are visualized by Coomassie staining.

extract from the ProtA-Rrp4p strain was fractionated by column chromatography (see Fig. 3B). Three fractions containing ProtA-Rrp4p were recovered (Fig. 3C). The

most abundant complex was recovered in fraction 1; this complex probably corresponds to the major complex purified previously by glycerol gradient centrifugation and immunoprecipitation (Mitchell et al. 1997). In addition to the previously characterized components of the exosome, fraction 1 contained Rrp40p, Rrp46p, and Mtr3p. Other protein bands in fraction 1 were identified as Csl4p (YNL232w) and the cytoplasmic Hsp70-like protein Ssa1p (YAL005c), but Rrp6p was not present. Fraction 3 contained the same exosome components as fraction 1, but lacked Ssa1p and contained Rrp6p. Rrp43p, which comigrates with ProtA-Rrp4p and the IgG heavy chain (Mitchell et al. 1997), was identified in fraction 3 but not in fraction 1 (Fig. 3C). Csl4p and Rrp45p also appear to comigrate in SDS-PAGE; from the band marked Csl4p + Rrp45p, only Rrp45p was identified from the preparation shown in Figure 3A, whereas only Csl4p was identified from the preparations shown in Figure 3C. It is, however, very likely that Rrp43p and Rrp45p are components of both complexes (see also below). Consistent with the recovery of Rrp6p in the total immunoprecipitate (Fig. 3A), approximately threefold less ProtA-Rrp4p was recovered in fraction 3 than in fraction 1 (twofold less of the material recovered in fraction 1 was loaded onto the gel in Fig. 3C than of the material in fractions 2 and 3). Fraction 2 comprises only ProtA-Rrp4p with Ssa1p, and was approximately fourfold less abundant than fraction 1. Consistent with glycerol gradient centrifugation (Mitchell et al. 1997), no free ProtA-Rrp4p was recovered. The ProtA-Rrp4p-Ssa1p complex was detected in variable yield on glycerol gradients (typically 5%–10% of total ProtA-Rrp4p; Mitchell et al. 1997) and may be due to dissociation of ProtA-Rrp4p from the complex during purification, possibly related to the presence of the protein A tag.

*CSL4* was identified previously in a screen for synthetic lethality with the chromatin protein Cep1p and is essential for viability (Baker et al. 1998). Conditional alleles of *CSL4* and *RRP40* were constructed by placing their expression under the control of a *GAL10* promoter (see Materials and Methods). In each case, the strains formed only microcolonies on solid medium containing 2% glucose (data not shown). Following transfer from liquid RSG medium to liquid glucose medium (Fig. 1) the strains ceased growth and 3' extended forms of the 5.8S rRNA accumulated (Fig. 2a), showing a pattern of intermediates similar to other exosome mutants. Depletion of Rrp40p or Csl4p also led to the accumulation of the 5' ETS pre-rRNA spacer fragment (Fig. 2c). Therefore, genetic depletion of any of the 10 essential components identified by copurification results in very similar defects in the processing of the 5.8S rRNA, showing that they form a single complex.

*RRP6* is not essential for viability (Briggs et al. 1998), and a strain carrying a precise deletion of *RRP6* was constructed (see Materials and Methods). This strain was impaired in growth at all temperatures and was nonviable at 37°C (temperature-sensitive lethal; data not shown). The *rrp6-Δ* strain was defective in the 3' processing of the 5.8S rRNA, but differed from the other

components of the exosome insofar as it accumulated a discrete species, 5.8S + 30, that was 3' extended by ~30 nucleotides (Fig. 2a; Briggs et al. 1998). The *rrp6-Δ* strain also accumulated the 5' ETS region of the pre-rRNA (Fig. 2c). We conclude that the exosome includes at least 11 components, all of which are required for normal 3' processing of the 5.8S rRNA and degradation of the 5' ETS region. Ten of these are essential for viability, whereas the absence of Rrp6p results in temperature-sensitive lethality (see Table 1).

It is unclear whether Ssa1p is a genuine component of the complex or associates with the exosome as a consequence of the protein A tag present on Rrp4p. In either event, because Ssa1p is predominantly cytoplasmic (Chirico et al. 1988; Deshaies et al. 1988), one obvious possibility was that fractions 1 and 3 contained cytoplasmic and nuclear forms of the exosome, respectively. To test this possibility, a ProtA-Rrp6p fusion was constructed (see Materials and Methods). The ProtA-Rrp6p construct complemented the temperature-sensitive lethal growth phenotype of the *rrp6-Δ* mutation, largely suppressed the accumulation of the 5.8S + 30 species in this strain, and cosedimented with ProtA-Rrp4p through glycerol gradients (data not shown). Therefore, we conclude that the protein A epitope does not grossly impair the ability of Rrp6p to associate with the exosome or to function in the cell.

Immunolocalization of the ProtA-Rrp6p and ProtA-Rrp4p (Mitchell et al. 1996) fusion proteins was compared to the nucleolar marker ProtA-Nop1p (Grandi et al. 1993) and staining of the nucleoplasm with DAPI. ProtA-Rrp6p gave a nuclear signal, with nucleolar en-

richment and a punctate nucleoplasmic staining. ProtA-Rrp4p was also observed in the nucleoplasm and nucleolus, but was additionally detected in the cytoplasm (Fig. 4). Notably, a GFP-Rrp43p fusion protein has recently been reported to be localized to both the nucleus and cytoplasm (Zanchin and Goldfarb 1999).

We conclude that two major forms of the exosome can be purified that contain at least 10 common components, Rrp4p, Rrp40-Rrp46p, Mtr3p, and Csl4p, all of which are essential for viability and are required for exosome function. Rrp6p is present only in a subfraction of the complex that is confined to the nucleus.

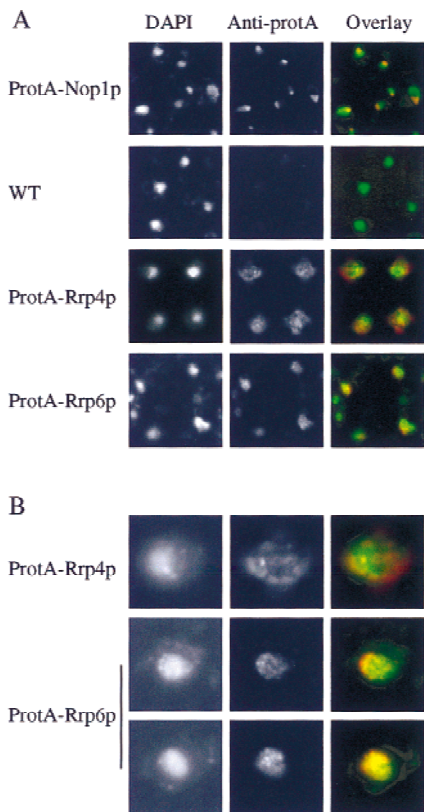
#### Characterization of the human PM-Scl complex

Rrp6p shows substantial homology to the human protein PM-Scl100 (Briggs et al. 1998), whereas Rrp45p is homologous to PM-Scl75 (Mian 1997), both of which are targets of autoimmune antibodies in patients suffering from polymyositis-scleroderma overlap syndrome (Alderuccio et al. 1991; Bluthner and Bautz 1992; Ge et al. 1992). Moreover, human orthologs have been identified for the Rrp4p, Rrp44p and Csl4p components of the exosome (Mitchell et al. 1997; Baker et al. 1998; Shiomi et al. 1998). Strikingly, expression of each of these cDNAs can suppress the phenotypes of mutations in the corresponding yeast genes, demonstrating their functional conservation (Mitchell et al. 1997; Baker et al. 1998; Shiomi et al. 1998). Translational searches of the human EST banks (see Materials and Methods) allowed virtual cDNAs to be assembled for hRrp40p, hRrp41p, hRrp42p, and hRrp46p; in each case, the putative human protein

**Table 1.** Components of the exosome

Protein	Gene	Phenotype	<i>E. coli</i> homolog	Mammalian homolog	Comments
Rrp4p	YHR069c	essential	S1 RNA BD	hRrp4p 43% (52%)	hRrp4p complements <i>rrp4-1</i>
Rrp40p	YOL142w	essential	S1 RNA BD	hRrp40p 35% (48%)	homologous to Rrp4p
Rrp41p/Ski6p	YGR195w	essential	RNase PH	hRrp41p 35% (55%)	
Rrp42p	YDL111c	essential	RNase PH	hRrp42p 25% (51%)	
Rrp43p	YCR035c	essential	RNase PH	PM-Scl75 38% (64%)	human KIAA0116 and <i>OIP2</i> also homologous
Rrp45p	YDR280w	essential	RNase PH	hRrp46p 35% (48%)	
Rrp46p	YGR095c	essential	RNase PH		
Mtr3p	YGR158c	essential	RNase PH		
Rrp44p/Dis3p	YOL021c	essential	RNase R (RNase II family)	hDis3p 45%	hDis3p complements <i>dis3-81</i>
Csl4p	YNL232w	essential	S1 RNA BD	hCs14p 48% (56%)	hCs14p complements <i>csl4-1</i>
Rrp6p	YOR001w	ts lethal	RNase D	PM-Scl100 32% (52%)	component only of nuclear complex

Rrp4p, Rrp40p, and Csl4p are not clearly homologous to known exonucleases from *E. coli* but are predicted to include regions homologous to the S1 RNA-binding domain (S1 RNA BD). For the human homologs numbers represent the percentage identity (similarity). In the case of Csl4p (Baker et al. 1998), Rrp40p, Rrp41p, Rrp42p, and Rrp46p, the numbers are based on consensus cDNAs assembled from ESTs and may not be fully accurate. (See text for references.)



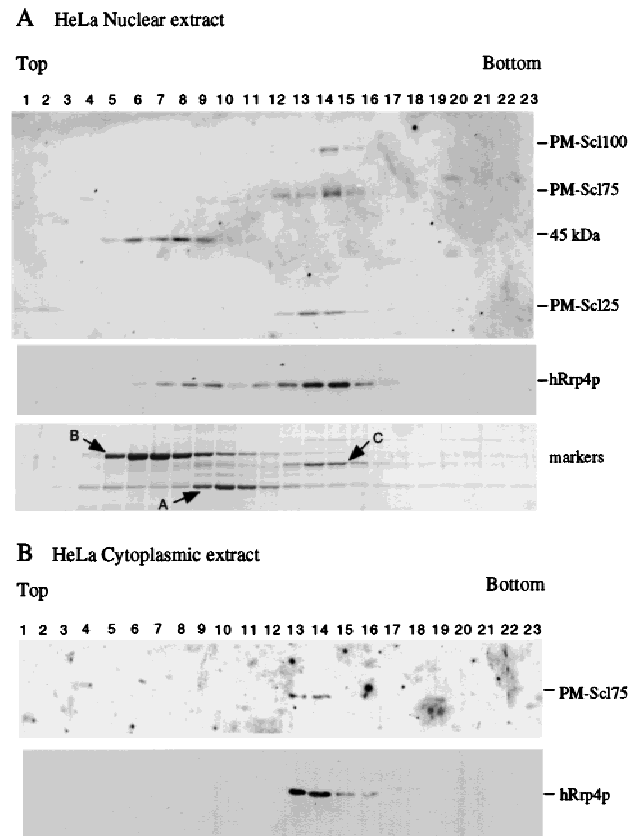
**Figure 4.** Rrp4p and Rrp6p differ in their nuclear-cytoplasmic distribution. (A) Strains expressing ProtA-Rrp4p, ProtA-Rrp6p, or ProtA-Nop1p were examined by indirect immunofluorescence using an anti-protein A antibody coupled to Texas Red. Also shown is the position of the DNA, visualized by DAPI staining. The combined image is pseudocolored with DAPI in green and Texas Red in red. For each tagged strain an otherwise isogenic wild-type control strain was also analyzed. The wild-type strain shown (P51) is isogenic with the ProtA-Rrp4p strain (see Table 2). (B) Higher resolution images are shown for the ProtA-Rrp4p and ProtA-Rrp6p to show the punctate staining pattern.

showed high homology to the yeast protein (Table 1). In addition, two other genes, KIAA0116 and *OIP2*, were found to be homologous to Rrp45p, although less so than PM-Scl75. For hRrp40p and hRrp46p, apparent products of alternative splicing were evident when the EST sequences were assembled into contigs (data not shown). In some cases, these alternative forms may have led to an overestimation of the number of discrete protein species in the PM-Scl complex.

Previous analyses showed that hRrp4p is present in a large complex (Mitchell et al. 1997). To determine whether the human homologs of other exosome components are present in the same complex, HeLa cell nuclear and cytoplasmic extracts (generously provided by Juan Valcárcel, EMBL, or prepared as described in Materials and Methods) were fractionated by glycerol gradient centrifugation. Fractions were analyzed by Western blotting with human autoimmune serum (generously provided by Walter van Venrooij, University of Nijmegen, The

Netherlands) or antibodies raised against recombinant hRrp4p (Mitchell et al. 1997; Fig. 5). In the nuclear extract, PM-Scl75 and an uncharacterized protein of ~25 kD that is also a target of the autoimmune serum (PM-Scl25) cosedimented with hRrp4p, with a peak in fractions 13 and 14. PM-Scl100 also showed substantial cosedimentation (Fig. 5A). The band at 45 kD is likely to be the species previously reported to cross-react with anti-PM-Scl75 antibodies (Alderuccio et al. 1991). PM-Scl100 was not detected in the cytoplasmic extract, but PM-Scl75 and hRrp4p cosedimented (Fig. 5B), as did PM-Scl25 (data not shown), with a peak in fractions 13 and 14.

To confirm the association between PM-Scl100 and hRrp4p in the HeLa cell nuclear extract, immunoprecipitation was performed (Fig. 6). Three different autoimmune sera (sera 1–3) were used, each of which reacted



**Figure 5.** Cosedimentation of hRrp4p and the PM-Scl complex. (A) HeLa cell nuclear extract. (B) HeLa cell cytoplasmic extract. Cell extracts were fractionated by glycerol gradient centrifugation. Samples were analyzed by Western blotting decorated with human autoimmune antisera reactive against PM-Scl100, PM-Scl75, and a previously uncharacterized human protein (PM-Scl25) or with rabbit antiserum raised against recombinant hRrp4p. The serum also cross-reacts with an unrelated 45-kD protein. Also shown is the sedimentation of molecular weight markers on a gradient run in parallel with the nuclear extract. Markers: (A) alcohol dehydrogenase from yeast (7.4S); (B) bovine serum albumin (4.3S); (C) bovine catalase (11.3S; Siegel and Monty 1966).



with PM-Scl75 and PM-Scl25 in both nuclear and cytoplasmic extracts. Homologs of at least three components of the exosome are present in the PM-Scl complex, providing strong evidence that these complexes are directly homologous.

Six human homologs of RNase PH were identified. These do not, however, have a 1:1 relationship with the six RNase PH homologs in the exosome. No clear human homologs were identified for Rrp43p or Mtr3p. Searches of the EST banks with these proteins identified ESTs related to KIAA0116 and *OIP2*; these sequences are, however, more homologous to Rrp45p than to the other yeast PH homologs (although less so than PM-Scl75). A probable interpretation is that yeast and humans have the same number of RNase PH homologs, but that some drift has occurred with duplicates of the *RRP45/PM-Scl75* gene replacing other species.

Mutations in individual components of the yeast exosome inhibited both nucleolar pre-rRNA processing and cytoplasmic mRNA turnover (Anderson and Parker 1998), indicating that related complexes are present in the nucleus and the cytoplasm. Moreover, a mutation in Mtr3p leads to nuclear accumulation of poly(A)<sup>+</sup> RNA (Kadowaki et al. 1995), as does a mutation in Dob1p/Mtr4p (de la Cruz et al. 1998; Liang et al. 1996), a putative RNA helicase required in addition to the exosome for 5.8S rRNA 3' end maturation and degradation of the 5' ETS fragment (de la Cruz et al. 1998). These observations suggest that the exosome may also play some role in nucleoplasmic RNA turnover or processing. Consistent with this hypothesis, GFP-Rrp43p (Zanchin and Goldfarb 1999) and ProtA-Rrp4p were detected in the nucleolus, nucleoplasm, and cytoplasm. In contrast, ProtA-Rrp6p was found to be exclusively nuclear, with a nucleolar enrichment. Two complexes could also be separated biochemically; these include 10 common components and differ in the presence of either Ssa1p, a cytoplasmic Hsp70-like protein (Chirico et al. 1988; Deshaies et al. 1988), or Rrp6p. The form lacking Rrp6p is presumed to be the cytoplasmic exosome complex, a proposal supported by the presence of Ssa1p. Approximately threefold more of this complex was recovered than the putative nuclear exosome that includes Rrp6p. Human PM-Scl100 was also restricted to the nucleus, while PM-Scl75, PM-Scl25, and hRrp4p partition between the nucleus and cytoplasm. The reported nucleolar enrichment of the human PM-Scl complex is probably a consequence of the immunodominance of PM-Scl100 in autoimmune sera (Ge et al. 1992; Gelpi et al. 1990). In fact, approximately equal amounts of the human nuclear and cytoplasmic complexes were recovered following subcellular fractionation.

We conclude that there are two forms of the exosome/PM-Scl complex in the nucleus and the cytoplasm that can be distinguished by the presence of Rrp6p/PM-Scl100 specifically in the nuclear form.

Rrp6p is not essential for viability, in contrast to the other 10 components of the exosome complex, although *rrp6-Δ* strains are severely impaired in growth and are temperature sensitive. Therefore, the exosome is there-

fore predicted to retain at least partial function in the absence of Rrp6p, a view supported by the observation that the major form of the complex lacks this protein. Conversely, all of the PM-Scl100 present in HeLa cell lysates appeared to be associated with the PM-Scl complex, suggesting that Rrp6p/PM-Scl100 may not function independently of the complex in vivo.

In *E. coli*, the homologs of the exosome components are not present in a related complex. However, the degradosome complex includes another 3' → 5' exonuclease, PNPase, together with the endonuclease and exonuclease RNase E and the putative RNA helicase RhlB (Carpousis et al. 1994; Py et al. 1996; Mackie 1998; Vanzo et al. 1998). It appears that throughout evolution, major activities involved in RNA processing and degradation have been assembled into large complexes, possibly to allow their coordinate regulation. The composition of these complexes are, however, very different in bacteria and eukaryotes.

## Materials and methods

### Strains and media

Except where stated, strains were grown in liquid or on solid minimal medium containing 0.67% yeast nitrogen base (DIFCO) and 2% glucose with appropriate supplements. For depletion, strains carrying GAL-regulated constructs were pregrown in RSG (2% peptone, 1% yeast extract, 2% raffinose, 2% sucrose, 2% galactose) and transferred to YPD (2% peptone, 1% yeast extract, 2% glucose).

Yeast strains used and constructed in this study are listed in Table 2. Gene disruptions of *RRP45* and *RRP46* were generated by a PCR strategy in the diploid strain BMA38 (Baudin et al. 1993) resulting in the replacement of the complete ORF by an auxotrophic marker (see Table 2). Successful disruption was confirmed by Southern hybridization. Chromosomal DNA from the *RRP45/rrp45::TRP1* and *RRP46/rrp46::HIS3* strains was digested by *EcoRI-HindIII* or *KpnI-EcoRI*, respectively, and hybridized with a probe derived from the PCR products that were used for transformation. Twelve tetrads from the *RRP45/rrp45::TRP1* strain and eight tetrads from *RRP46/rrp46::HIS3* strain were dissected on YPD plates and incubated for 6 days at 23°C. Each showed 2:2 segregation for spore viability. All viable spores were auxotrophic for tryptophan or histidine, respectively, indicating that the disrupted alleles were lethal. The nonessential *RRP6* gene was disrupted in the haploid strain YBD38 (see Table 2) by use of the *Kluyveromyces lactis* TRP marker, obtained by PCR amplification from plasmid pBS1408 (generously provided by Bertrand Séraphin, EMBL, Heidelberg, Germany).

The oligonucleotides used to construct and test the gene disruptions were 5'*RRP45::TRP1* (807); 3'*RRP45::TRP* (808); 5'*RRP46::HIS3* (809); 3'*RRP46::HIS3* (810); 5'*RRP6::KI TRP* (811); 3'*RRP6::KI TRP* (812). Test oligonucleotides were 3'*RRP45* (813); 3'*RRP6* (815); Sc TRP (816); HIS (817); KI TRP (818) (full sequences are available from the authors).

Conditional mutants under the control of the inducible *GAL10* promoter were generated for the *RRP40*, *RRP45*, *RRP46*, and *CSL4* genes by a one-step PCR strategy in the YDL401 strain (Lafontaine and Tollervey 1996). Transformants were selected for His<sup>+</sup> prototrophy and screened by PCR.

The oligonucleotides used to construct the conditional mutants were 5'*GAL-RRP45* (819); 3'*GAL-RRP45* (820); 5'*GAL-*

**Table 2.** *Strains used in this study*

Strain	Genotype	Reference/Note
BMA38	<i>MATa/α ade2-1/ade2-1 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 trp1-1/trp1-1 ura3-1/ura3-1 can1-100/can1-100</i>	Baudin et al. (1993)
YCA10	as BMA38 but <i>RRP45/RRP45::TRP1</i>	this study
YCA11	as BMA38 but <i>RRP46/RRP46::HIS3</i>	this study
YCA12	<i>MATa ade2-1 his3-Δ200 leu2-3,112 trp1-1 ura3-1 can1-100 RRP6::K1 TRP1</i>	this study
YTK100	<i>MATa mtr3-1 ura3-52</i>	Kadowaki et al. (1995)
YDL401	<i>MATa his3Δ200 leu2Δ1 trp1 ura3-52 gal2 galΔ108</i>	Lafontaine and Tollervey (1996)
YCA20	as YDL401 but <i>GAL10::RRP45</i>	this study
YCA21	as YDL401 but <i>GAL10::RRP46</i>	this study
P79	as YDL401 but <i>GAL10::protA-RRP4</i>	Mitchell et al. (1997)
P147	as YDL401 but <i>GAL10::RRP40</i>	this study
P170	as YDL401 but <i>GAL10::CSL4</i>	this study
GAL::DOB1	<i>MATα ura3-1 ade2-1 his3-11,15 leu2-3,112 trp1-1 dob1::HIS3MX6 + [pAS24-DOB1]</i>	de la Cruz et al. (1998)
P49	<i>MATα ade2-1 his3-11 leu2-3 trp1-1 ura3-52 can1-100 rrp4Δ::HIS3 + [pRS416/protA-RRP4]</i>	Mitchell et al. (1996)
P51	<i>MATα ade2-1 his3-11 leu2-3 trp1-1 ura3-52 can1-100 rrp4Δ::HIS3 + [pRS415/RRP4]</i>	Mitchell et al. (1996)
YCA40	<i>MATa ade2-1 his3-Δ200 leu2-3,112, trp1-1 ura3-1 can1-100 RRP6::K1 TRP1 + [pRS416/protA-RRP6]</i>	this study
ProtA-Nop1	<i>MATα ade leu trp lys ura3 nop1::URA3 + [pUN100-protA-NOP1]</i>	Jansen et al. (1993)

RRP46 (821); 3'GAL-RRP46 (822); 3'GAL-ProtA-RRP46 (823); 5'GAL-RRP40 (824); 3'GAL-RRP40 (825); 5'GAL-CSL4 (826); 3'GAL-CSL4 (827). The amplification of *RRP6::TRP* was done with oligonucleotides 5'RRP6 (834) and 3'RRP6 (835) (full sequences are available from the authors).

#### Construction of the *ProtA-Rrp6p* fusion

To construct the *ProtA-RRP6* fusion gene, the *RRP4* ORF was excised from plasmid pPM46 (Mitchell et al. 1997) by restriction cleavage at sites *EcoRI* and *HindIII* and replaced by the *RRP6* ORF amplified by PCR from wild-type genomic DNA and flanked by the same restriction sites. The resulting plasmid was transformed into the haploid *RRP6::TRP* strain and shown to complement fully the RNA processing and growth phenotypes of the deleted strain. The oligonucleotides used for the PCR were 5'PRS (836) and 3'PRS (837).

#### Fractionation of *ProtA-Rrp4p* complexes

Lysate from 5-liter YPD cultures of strain P49 was prepared in TMN buffer [10 mM Tris-HCl (pH 7.6), 5 mM MgCl<sub>2</sub>, 0.1% NP-40] containing 150 mM NaCl, 1 mM PMSF, and 10% glycerol, as described (Mitchell et al. 1996). *ProtA-Rrp4p* complexes were purified by immunoprecipitation with IgG-Sepharose, either from clarified lysates or after fractionation by low-pressure column chromatography. Purification procedures were carried out at 4°C in buffers containing 0.5 mM PMSF and fractions were screened for the presence of *ProtA-Rrp4p* by Western blot analyses, using peroxidase-antiperoxidase rabbit antibody (Sigma).

Cleared lysate was applied directly to a 100-μl IgG-Sepharose 6 FF column (Pharmacia) and washed with 100 ml of TMN-150, bound material was eluted with a 0.1–2 M MgCl<sub>2</sub> gradient (Görlich et al. 1996) in TMN-150 buffer (20 fractions of 150 μl at increments of 100 mM MgCl<sub>2</sub>). Aliquots of 5 μl were resolved by SDS-PAGE and visualized by silver staining. Fractions containing the proteins of interest were precipitated with 9 vol of isopropanol, pooled, and analyzed on 10% polyacrylamide gels containing SDS.

For fractionation, cleared lysate (30 ml) diluted to 100 mM NaCl was batch-bound to DEAE-Sepharose FF (Pharmacia).

Bound material was washed three times with 30 ml of TMN buffer containing 100 mM NaCl (TMN-100), eluted with 5 × 30 ml TMN-300/10% glycerol (E300) and then frozen at -80°C. The pooled eluates were diluted to 100 mM NaCl and passed through a 10-ml Q-Sepharose FF column (Mono Q; Pharmacia). Bound material was eluted stepwise with 50 ml of TMN-150, TMN-200 (E200), TMN-320 (E320), and TMN-500. Material that failed to bind to DEAE-Sepharose FF (FT) was passed through a 10-ml SP-Sepharose FF column (Mono S; Pharmacia). After washing with 50 ml of TMN-300, bound material was eluted with 50 ml of TMN-500 (E500). Eluates from the Mono Q and Mono S columns were diluted to 150 mM NaCl and passed through small (100 μl) IgG-Sepharose 6 FF columns. Bound material was washed with 100 ml of TMN-150, and retained proteins were eluted with 1 ml of 0.5 M acetic acid. The eluates were concentrated by centrifugation under vacuum and analyzed by SDS-PAGE and nanospray mass spectrometry, as above.

#### Mass spectrometric analysis

Protein bands were excised from the gel, digested in the gel, and analyzed according to the strategy described elsewhere (Shevchenko et al. 1996). High mass accuracy MALDI peptide mapping (Jensen et al. 1996) was performed on a Bruker Reflex III mass spectrometer (Bruker Daltonics, Bremen, Germany). To resolve protein mixtures an iterative approach (Jensen et al. 1997) was used. In case of uncertainty identifications were confirmed by nano-electrospray tandem mass spectrometry on a pilot QqTOF instrument (SCIEX, Toronto, Canada; Shevchenko et al. 1997b). PeptideSearch software, developed in house, was used for protein database searching.

#### Glycerol gradient analysis of a *HeLa* cell extracts

*HeLa* cell lysates were prepared according to standard procedures (Dignam et al. 1983; Lee et al. 1988). Nuclear and cytoplasmic extracts were centrifuged through 12-ml glycerol density gradients as described previously (Mitchell et al. 1997). Gradient fractions were analyzed by Western blotting analysis with rabbit anti-hRrp4p serum or sera of patients suffering from poly-



myositis-scleroderma overlap syndrome [kindly provided by Dr. W. van Venrooij and obtained from the University Hospital (St. Radboud) of Nijmegen].

#### Immunofluorescence

Cells were grown in selective medium to mid-exponential phase, fixed by incubation in 4% (vol/vol) formaldehyde for 1 hr at room temperature, and spheroplasted. Then immunofluorescence was then performed as described previously (Bergès et al. 1994; Grandi et al. 1993). Protein A fusions were detected with a rabbit anti-protein A antibody (Sigma) and a secondary goat anti-rabbit antibody coupled to Texas Red (Dianova) at a 1:100 and 1:200 dilution, respectively. To stain nuclear DNA, DAPI was included in the mounting medium (Vectashield, Vector Laboratories).

#### Immunoprecipitation of the PM-Scl complex with patient sera

Patient sera directed specifically against PM-Scl100 (kindly provided by Dr. W. van Venrooij) were used for the immunoprecipitation experiments. HeLa cell lysates were prepared as described above. A 50% solution of protein A-Sepharose beads (10 µl, Pharmacia) was washed three times in IPP 500 [500 mM NaCl, 10 mM Tris-HCl (pH 8), 0.1% NP-40, 0.5 mM PMSF] and incubated for 1 hr at room temperature with 5 µl of human autoimmune sera. Beads were washed three times with IPP500, transferred in 10 µl of IPP150 ([50 mM NaCl, 10 mM Tris-HCl (pH 8), 0.1% NP-40, 0.5 mM PMSF] and then added to 10 µl of HeLa cell nuclear extract. After incubation for 2 hr at 4°C, the supernatant was recovered and beads were washed four times with IPP150. Bound proteins were eluted from the beads by a 5 min boiling in protein gel loading buffer. Total, supernatant, and pellet proteins were analyzed by SDS-PAGE and Western blotting analysis with anti-hRrp4p serum or affinity-purified anti-hPop1 antibodies (Lygerou et al. 1996; Mitchell et al. 1997).

#### RNA analysis

RNA isolation and Northern blot hybridization were performed as described previously (Beltrame and Tollervey 1992; Tollervey 1987). Oligonucleotides used for rRNA and pre-rRNA analysis were 5'-TGAGAAGGAAATGACGCT (oligonucleotide 020), 5'-GCGTTGTTTCATCGATGC (oligonucleotide 017), 5'-CGC-TGCTCACAATGG (oligonucleotide 033), and 5'-CTACTCG-GTCAGGCTC (oligonucleotide 014).

#### Database searches

The human EST banks were searched using the EFEAME p2n program for translational frame-shifting, on the Bioaccelerator of the European Molecular Biology Laboratory (<http://www.embl-heidelberg.de>). Contigs were assembled from the retrieved ESTs by use of the Gene JockeyII program. Homology was calculated by use of using the Bestfit program [Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, WI].

The ESTs used for the alignments were *hRRP40*: HS103148, AA916866, AA715297, AA909843, AA829746, AA760696, AA748308, AA747303, HSA01383, HS479237, HS417169, HS1213865, HS1191331, HS1186630, AA937191, AA741488, HSA57832, HSA01383, HS620247, HS617138, AA736510, HS1300540, HS1273716, HS1269362, HS1229711, HS1198690, HS1191331, and HS1174014; *hRRP41*: HS0229, HSZZ84720, HS462881, HS1210855, HS060127, and HSAA29848; *hRRP42*:

AA654791, HS599371, HSZZ85135, HS20834, AA581010, HS414162, HS979316, and HSZZ84357; *hRRP46*: HS078341, HS84856, HS1255212, HS1226957, HSZZ41259, HS1256223, HS1225454, HS1249336, and HS1172072.

#### Acknowledgments

We thank Walter van Venrooij for providing the PM-Scl sera, Juan Balcácer for providing HeLa cell nuclear and cytoplasmic extract, Alan Tartakoff for the *mtr3-1* strain and Bertrand Sefaphin for pBS1408. We thank Roy Parker for pointing out the homology between Rrp41p and Mtr3p and Saira Mian for pointing out the homology between Rrp4p and Rrp40p, as well as the putative S1 DNA binding domains in Rrp4p, Rrp40p, and Csl4p. This work was supported by the Wellcome Trust.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked 'advertisement' in accordance with 18 USC section 1734 solely to indicate this fact.

#### References

- Alderuccio, F., E.K. Chan, and E.M. Tan. 1991. Molecular characterization of an autoantigen of PM-Scl in the polymyositis/scleroderma overlap syndrome: A unique and complete human cDNA encoding an apparent 75-kD acidic protein of the nucleolar complex. *J. Exp. Med.* **173**: 941-952.
- Anderson, J.S.J. and R.P. Parker. 1998. The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J.* **17**: 1497-1506.
- Baker, R.E., K. Harris, and K. Zhang. 1998. Mutations synthetically lethal with *cep1* target *S. cerevisiae* kinetochore components. *Genetics* **149**: 73-85.
- Baudin, A., O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, and C. Cullin. 1993. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **21**: 3329-3330.
- Beltrame, M. and D. Tollervey. 1992. Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J.* **11**: 1531-1542.
- Bergès, T., E. Petfalski, D. Tollervey, and E.C. Hurt. 1994. Synthetic lethality with fibrillarins identifies NOP77p, a nucleolar protein required for pre-rRNA processing and modification. *EMBO J.* **13**: 3136-3148.
- Bluthner, M. and F.A. Bautz. 1992. Cloning and characterization of the cDNA coding for a polymyositis-scleroderma overlap syndrome-related nucleolar 100-kD protein. *J. Exp. Med.* **176**: 973-980.
- Briggs, M.W., K.T. Burkard, and J.S. Butler. 1998. Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J. Biol. Chem.* **273**: 13255-13263.
- Bycroft, M., T.J. Hubbard, M. Proctor, S.M. Freund, and A.G. Murzin. 1997. The solution structure of the S1 RNA binding domain: A member of an ancient nucleic acid-binding fold. *Cell* **88**: 235-342.
- Carpousis, A.J., G. Van Houwe, C. Ehretsmann, and H.M. Krisch. 1994. Copurification of *E. coli* RNAase E and PN-Pase: Evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell* **76**: 889-900.
- Cheng, Z.F., Y. Zuo, Z. Li, K.E. Rudd, and M.P. Deutscher. 1998. The *vacB* gene required for virulence in *Shigella flexneri* and

- Escherichia coli encodes the exoribonuclease RNase R. *J. Biol. Chem.* **273**: 14077–14080.
- Chirico, W.J., M.G. Waters, and G. Blobel. 1988. 70K heat shock related proteins stimulate protein translocation into microsomes. *Nature* **332**: 805–810.
- de la Cruz, J., D. Kressler, D. Tollervey, and P. Linder. 1998. Dobl1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in *Saccharomyces cerevisiae*. *EMBO J* **17**: 1128–1140.
- Deshaiya, R.J., B.D. Koch, M. Werner-Washburne, E.A. Craig, and R. Schekman. 1988. A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. *Nature* **332**: 800–805.
- Dignam, J.D., P.L. Martin, B.S. Shastry, and R.G. Roeder. 1983. Eukaryotic gene transcription with purified components. *Methods Enzymol.* **101**: 582–598.
- Ge, Q., M.B. Frank, C. O'Brien, and I.N. Targoff. 1992. Cloning of a complementary DNA coding for the 100-kD antigenic protein of the PM-Scl autoantigen. *J. Clin. Invest.* **90**: 559–570.
- Gelpi, C., A. Alguero, M. Angeles Martinez, S. Vidal, C. Juarez, and J.L. Rodriguez-Sanchez. 1990. Identification of protein components reactive with anti-PM/Scl autoantibodies. *Clin. Exp. Immunol.* **81**: 59–64.
- Görlich, D., R. Kraft, S. Kostka, F. Vogel, E. Hartmann, R.A. Laskey, I.W. Mattaj, and E. Izaurraide. 1996. Importin provides a link between nuclear protein import and U snRNA export. *Cell* **87**: 21–32.
- Grandi, P., V. Doyl, and E.C. Hurt. 1993. Purification of NSP1 reveals complex formation with 'GLFG' nucleoporins and a novel nuclear pore protein NIC96. *EMBO J.* **12**: 3061–3071.
- Jansen, R., D. Tollervey, and E.C. Hurt. 1993. A U3 snoRNP protein with homology to splicing factor PRP4 and Gb domains is required for ribosomal RNA processing. *EMBO J.* **12**: 2549–2558.
- Jensen, O.N., A. Podtelejnikov, and M. Mann. 1996. Delayed extraction improves specificity in database searches by matrix-assisted laser desorption/ionization peptide maps. *Rapid Commun. Mass Spectrom.* **10**: 1371–1378.
- . 1997. Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. *Anal. Chem.* **69**: 4741–4750.
- Kadowaki, T., R. Schneider, M. Hitomi, and A.M. Tartakoff. 1995. Mutations in nucleolar proteins lead to nucleolar accumulation of polyA+ RNA in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **6**: 1103–1110.
- Kuster, B. and M. Mann. 1998. Identifying proteins and post-translational modifications by mass spectrometry. *Curr. Opin. Struct. Biol.* **8**: 393–400.
- Lafontaine, D. and D. Tollervey. 1996. One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. *Nucleic Acids Res.* **24**: 3469–3472.
- Lee, K.A., A. Bindereif, and M.R. Green. 1988. A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. *Gene Anal. Tech.* **5**: 22–31.
- Liang, S., M. Hitomi, Y.H. Hu, Y. Liu, and A.M. Tartakoff. 1996. A DEAD-box-family protein is required for nucleocytoplasmic transport of yeast mRNA. *Mol. Cell. Biol.* **16**: 5139–5146.
- Lygerou, Z., H. Pluk, W.J. van Venrooij, and B. Séraphin. 1996. hPop1: An autoantigenic protein subunit shared by the human RNase P and RNase MRP ribonucleoproteins. *EMBO J.* **15**: 5936–5948.
- Mackie, G.A. 1998. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**: 720–723.
- Mian, I.S. 1997. Comparative sequence analysis of ribonucleases HII, III, II PH and D. *Nucleic Acids Res.* **25**: 3187–3195.
- Mitchell, P., E. Petfalski, and D. Tollervey. 1996. The 3'-end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes & Dev.* **10**: 502–513.
- Mitchell, P., E. Petfalski, A. Shevchenko, M. Mann, and D. Tollervey. 1997. The exosome: A conserved eukaryotic RNA processing complex containing multiple 3' → 5' exoribonuclease activities. *Cell* **91**: 457–466.
- Py, B., C.F. Higgins, H.M. Krisch, and A.J. Carpousis. 1996. A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature* **381**: 169–172.
- Reimer, G. 1990. Autoantibodies against nuclear, nucleolar, and mitochondrial antigens in systemic sclerosis (scleroderma). *Rheum. Dis. North Am.* **16**: 169–183.
- Reimer, G., U. Scheer, J.M. Peters, and E.M. Tan. 1986. Immunolocalization and partial characterization of a nucleolar autoantigen (PM-Scl) associated with polymyositis/scleroderma overlap syndromes. *J. Immunol.* **137**: 3802–3808.
- Shevchenko, A., O.N. Jensen, A.V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, and M. Mann. 1996. Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci.* **93**: 14440–14445.
- Shevchenko, A., I. Chernushevich, W. Ens, K.G. Standing, B. Thomson, M. Wilm, and M. Mann. 1997a. Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **11**: 1015–1024.
- Shevchenko, A., M. Wilm, and M. Mann. 1997b. Peptide sequencing by mass spectrometry for homology searches and cloning of genes. *J. Protein Chem.* **16**: 481–490.
- Shiomi, T., K. Fukushima, N. Suzuki, N. Nakashima, E. Noguchi, and T. Nishimoto. 1998. Human Dis3p, which binds to either GTP- or GDP-Ran, complements *Saccharomyces cerevisiae* dis3. *J. Biochem.* **123**: 883–890.
- Siegel, L.M. and K.J. Monty. 1966. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim. Biophys. Acta.* **112**: 346–362.
- Tollervey, D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J.* **6**: 4169–4175.
- Vanzo, N.F., Y.S. Li, B. Py, E. Blum, C.F. Higgins, L.C. Raynal, H.M. Krisch, and A.J. Carpousis. 1998. Ribonuclease E organizes the protein interactions in the *Escherichia coli* RNA degradosome. *Genes & Dev.* **12**: 2770–2781.
- Wilm, M., A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigener, T. Fotsis, and M. Mann. 1996. Femtomole sequencing of proteins from polyacrylamide gels by nanoelectrospray mass spectrometry. *Nature* **379**: 466–469.
- Zanchin, N.I. and D.S. Goldfarb. 1999. Nip7p interacts with Nop8p, an essential nucleolar protein required for 60S ribosome biogenesis, and the exosome subunit Rrp43p. *Mol. Cell. Biol.* **19**: 1518–1525.



# Functions of the exosome in rRNA, snoRNA and snRNA synthesis

Christine Allmann, Joanna Kufel, Guillaume Chanfreau<sup>1,2</sup>, Philip Mitchell, Elisabeth Petfalski and David Tollervey<sup>3</sup>

Institute of Cell and Molecular Biology, University of Edinburgh, Swann Building, King's Buildings, Edinburgh EH9 3JR, UK and <sup>1</sup>GIM-Biotechnologies, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

<sup>2</sup>Present address: Department of Chemistry and Biochemistry, UCLA, Los Angeles CA 90095-1569, USA

<sup>3</sup>Corresponding author  
e-mail: d.tollervey@ed.ac.uk

C.Allmann and J.Kufel contributed equally to this work

**The yeast nuclear exosome contains multiple 3'→5' exoribonucleases, raising the question of why so many activities are present in the complex. All components are required during the 3' processing of the 5.8S rRNA, together with the putative RNA helicase Dob1p/Mtr4p. During this processing three distinct steps can be resolved, and hand-over between different exonucleases appears to occur at least twice. 3' processing of snoRNAs (small nucleolar RNAs) that are excised from polycistronic precursors or from mRNA introns is also a multi-step process that involves the exosome, with final trimming specifically dependent on the Rrp6p component. The spliceosomal U4 snRNA (small nuclear RNA) is synthesized from a 3' extended precursor that is cleaved by Rnt1p at sites 135 and 169 nt downstream of the mature 3' end. This cleavage is followed by 3'→5' processing of the pre-snRNA involving the exosome complex and Dob1p. The exosome, together with Rnt1p, also participates in the 3' processing of the U1 and U5 snRNAs. We conclude that the exosome is involved in the processing of many RNA substrates and that different components can have distinct functions.**

**Keywords:** pre-rRNA/RNA processing/*Saccharomyces cerevisiae*/snoRNA/snRNA

## Introduction

Eukaryotic cells contain a large number of stable RNA species, nearly all of which are synthesized by post-transcriptional processing from larger precursors. This has long been known for the highly abundant cytoplasmic RNAs, tRNAs and rRNAs, but more recently it has become clear that this is also the case for the small nuclear RNAs (snRNAs), which participate in pre-mRNA splicing, and the small nucleolar RNAs (snoRNAs), which participate in rRNA processing and modification.

Analyses of 3' processing of the 5.8S rRNA in *Saccharomyces cerevisiae* led to the identification of the exosome

complex of 3'→5' exonucleases (Mitchell *et al.*, 1996, 1997; Allmann *et al.*, 1999). Originally reported to contain five different 3'→5' exonucleases, it is now likely that the exosome contains at least 10 exonucleases (Allmann *et al.*, 1999). These include six homologues of *Escherichia coli* RNase PH (Rrp41p, Rrp42p, Rrp43p, Rrp45p, Rrp46p and Mtr3p), a homologue of *E.coli* RNase R and RNase II (Rrp44p/Dis3p), and a homologue of *E.coli* RNase D (Rrp6p). Two other components, Rrp4p and Rrp40p, are homologous to each other, and Rrp4p has been shown to be a 3'→5' exonuclease (Mitchell *et al.*, 1997). The remaining component is Csl4p, which has not been reported to have nuclease activity but does contain a potential S1 RNA binding domain (S.Mian, personal communication), indicating that it is also likely to bind RNA directly. All components of the exosome are essential for viability (Mitchell *et al.*, 1996, 1997; Noguchi *et al.*, 1996; Baker *et al.*, 1998; Allmann *et al.*, 1999) with the exception of Rrp6p, the absence of which results in temperature-sensitive (ts) lethality and impaired growth at all temperatures (Briggs *et al.*, 1998). Normal processing of the 7S pre-rRNA to the mature 5.8S rRNA requires all components of the exosome, but the phenotype of the *rrp6*-Δ strain differs substantially from that of the other mutants, making it unclear whether these function in the same or parallel pathways. In addition to the components of the exosome, the yeast genome contains at least six other open reading frames that are predicted to encode 3'→5' exonucleases, based on sequence comparisons with *E.coli* enzymes (Mian, 1997; Moser *et al.*, 1997).

5' processing of the 5.8S rRNA requires the activity of two homologous 5'→3' exonucleases, Rat1p and Xrn1p, with the major role probably being played by Rat1p (Henry *et al.*, 1994). The same exonucleases are required for the 5' processing of several snoRNA species, many of which are either synthesized from polycistronic pre-snoRNA transcripts, or are excised from the introns of pre-mRNAs following intron lariat debranching (Ooi *et al.*, 1998; Petfalski *et al.*, 1998). All characterized yeast polycistronic snoRNAs are initially processed by endonuclease cleavage by Rnt1p (Chanfreau *et al.*, 1998a,b; Qu *et al.*, 1999), the yeast homologue of *E.coli* RNase III (Abou Elela *et al.*, 1996), which separates the individual pre-snoRNAs. Rnt1p also processes the pre-rRNA in the 3' external transcribed spacer (3'-ETS) (Abou Elela *et al.*, 1996; Kufel *et al.*, 1999) and cleaves 3' extended precursors to the U1, U2 and U5 snRNAs (Chanfreau *et al.*, 1997; Abou Elela and Ares, 1998; Seipelt *et al.*, 1999). In the absence of Rnt1p cleavage, polyadenylated forms of U1 and U2 are synthesized (Abou Elela and Ares, 1998; Seipelt *et al.*, 1999). In *rnt1*-Δ strains the processing of the 3'-ETS and polycistronic pre-snoRNAs is almost completely inhibited, with severe effects on rRNA and snoRNA synthesis. However, 3' processing of the snRNAs

continues, indicating the existence of alternative processing pathways or activities. The existence of alternative 3' processing pathways has also been shown for yeast tRNAs (Yoo and Wolin, 1997) and multiple activities can carry out 3' processing of many small RNAs in *E.coli* (see Li *et al.*, 1998 and references therein).

Here, we investigate the roles of the different exosome components in the 3' processing of the 5.8S rRNA and pre-rRNA spacer degradation, and present data indicating that the exosome also participates in the 3' processing of many snRNA and snoRNA species.

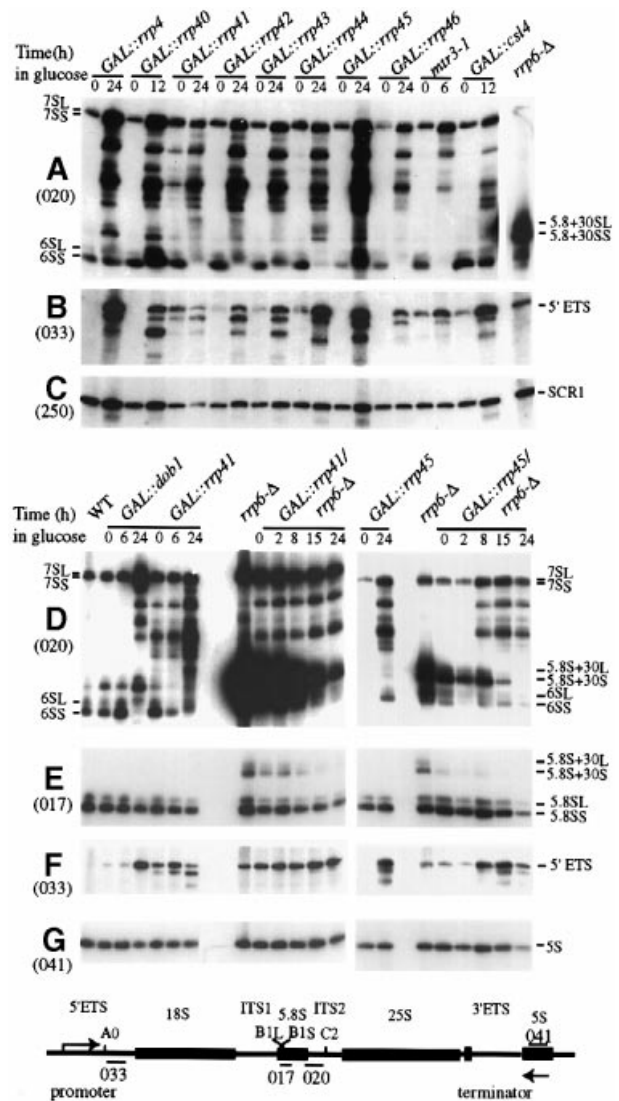
## Results

### The pathway of 5.8S 3' processing

Three different 3' extended forms of the 5.8S rRNA can be detected in wild-type strains of yeast. The 7S pre-rRNAs are 3' extended by ~134 nt to site C<sub>2</sub> in ITS2 (Veldman *et al.*, 1981), while the 6S pre-rRNAs represent 5.8S species with short, probably heterogeneous, 3' extensions of ~8 nt (Mitchell *et al.*, 1996) (see Figure 1 for a schematic showing the pre-rRNA and processing sites). In addition, the 5.8S + 30 species (Briggs *et al.*, 1998) can be detected at low levels in the wild type (Figure 1D). The 5' end of the 5.8S rRNA is heterogeneous, with two major forms that differ by 8 nt, designated 5.8S<sub>L</sub> and 5.8S<sub>S</sub> (Henry *et al.*, 1994). Since 5' processing of the 5.8S rRNA precedes 3' processing, the 7S pre-rRNA, 6S pre-rRNA and 5.8S + 30 species all show long and short forms, e.g. 5.8S + 30<sub>L</sub> and 5.8S + 30<sub>S</sub> (Figure 1A and D).

Processing of the 5.8S rRNA was compared in strains carrying conditional mutations for the 10 essential components of the exosome, using the ts-lethal *mtr3-1* allele and *GAL*-regulated constructs allowing depletion of Rrp4p, Rrp40p, Rrp41p, Rrp42p, Rrp43p, Rrp44p, Rrp45p, Rrp46p or Csl4p (Figure 1A). As previously reported (Mitchell *et al.*, 1997; Allmang *et al.*, 1999), similar 3' extended intermediates were observed in each case, forming a ladder up to the position of the 7S pre-rRNAs. The *GAL::rrp41* strain underexpresses Rrp41p in permissive, RSG medium, and therefore shows some accumulation of the extended species in the 0 h sample. Strong accumulation of the 6S pre-rRNA was seen in the strains depleted of Rrp40p or Rrp45p, while 6S was reduced in the Rrp41p-, Rrp44p- or Rrp46p-depleted and *mtr3-1* strains and little altered in strains depleted of Rrp4p or Csl4p. Moreover, in strains depleted of Rrp41p, Rrp42p or Rrp43p the position of the 6S pre-rRNA appeared to be displaced up the gel, corresponding to an increase in size of ~3 nt. We conclude that different components of the exosome do not play identical roles in processing of the 6S pre-rRNA.

In *rrp6-Δ* strains, a distinctly different pattern of processing was observed (Figure 1A and D) (Briggs *et al.*, 1998) with accumulation of high levels of the 5.8S + 30 pre-rRNAs. To determine whether Rrp6p and the other exosome components act on the same pre-rRNA processing pathway or function in independent parallel pathways, double-mutant strains were constructed carrying the *rrp6-Δ* allele and either the *GAL::rrp41* (Mitchell *et al.*, 1997) or the *GAL::rrp45* allele. Depletion of either Rrp41p or Rrp45p from a strain lacking Rrp6p led to the progressive loss of the 5.8S + 30 processing intermediate, clearly



**Fig. 1.** Northern analysis of processing of the 5.8S rRNA and degradation of the 5' ETS region of the pre-rRNA in exosome mutants. (A and D) Hybridization with probe 020, complementary to the 5.8S-ITS-2 boundary. (B and F) Hybridization with probe 033, complementary to the 5'-ETS region around + 270. (E) Hybridization with probe 017, complementary to the 5' region of the mature 5.8S rRNA. (C) Hybridization with probe 250, complementary to SCR1 RNA. (G) Hybridization with probe 041, complementary to the mature 5S rRNA. Probe names are indicated in parentheses on the left. RNA was extracted from strains carrying *GAL*-regulated constructs following transfer from permissive, RSG medium to repressive, glucose medium for the times indicated. The *mtr3-1* strain was grown in glucose medium at 25°C or transferred to 37°C for 6 h. The *rrp6-Δ* strain was grown on glucose medium at 30°C.

showing that Rrp41p and Rrp45p act epistatically to Rrp6p in the 5.8S processing pathway (Figure 1D). Metabolic labelling of an *rrp6-Δ* strain also indicated that Rrp6p participates in the major 5.8S rRNA processing pathway (Briggs *et al.*, 1998).

Mutations in the putative ATP-dependent RNA helicase Dob1p (Mtr4p) also interfere with 3' processing of the 5.8S rRNA (de la Cruz *et al.*, 1998). A *GAL::dobl* strain genetically depleted of Dob1p accumulated both the 5.8S + 30 species and larger intermediates that are seen in other exosome mutants (Figure 1D). The 6S pre-rRNAs accumulated in the *GAL::dobl* strain 2 and 6 h after

transfer to glucose medium but were reduced after 24 h. This probably occurred because processing of the pre-rRNA is strongly inhibited prior to synthesis of 6S, as shown by the high accumulation of the 7S and 5.8S + 30 pre-rRNAs.

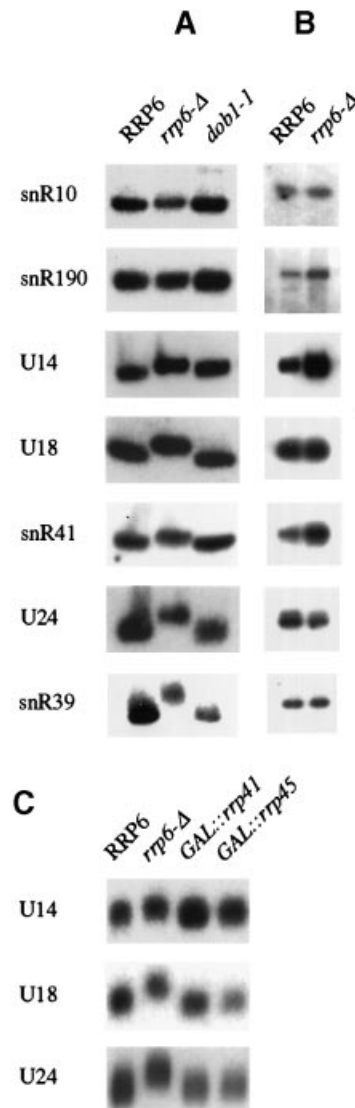
Strains depleted of exosome components or Dob1p, or carrying the *mtr3-1* or *rrp6-Δ* mutations, accumulated the excised 5' ETS region and degradation intermediates (Figure 1B and F) (de la Cruz *et al.*, 1998; Allmang *et al.*, 1999). The levels of the degradation intermediates were quite variable among the exosome mutants. This indicates that, while degradation of the 5' ETS involves the entire exosome complex, different components do not have identical functions during this activity.

### Pre-snoRNA processing

Many yeast snoRNAs are synthesized by post-transcriptional processing, either from the excised introns of pre-mRNAs or from polycistronic transcripts that include multiple snoRNAs. In higher eukaryotes, both 5' and 3' processing of pre-snoRNAs involves exonuclease activities (Caffarelli *et al.*, 1994, 1996; Cecconi *et al.*, 1995; Cavaillé and Bachellerie, 1996; Kiss *et al.*, 1996). 5' processing of several yeast pre-snoRNAs was shown to require the 5'→3' exonucleases Rat1p and Xrn1p, with the major role being performed by Rat1p (Larimer *et al.*, 1992; Kenna *et al.*, 1993; Petfalski *et al.*, 1998; Villa *et al.*, 1998). Northern analysis of RNA extracted from the *rrp6-Δ* strain showed many species with a discrete shift in gel mobility that would correspond to an increase in length of 3 nt (Figure 2A). This was observed for the intronic snoRNAs U18, U24 and snR39, as well as U14 and snR41, which are encoded in dicistronic and polycistronic transcripts, respectively (Figure 2A). In contrast, the gel mobilities of snoRNAs that are transcribed from their own promoter and terminator, snR10 (Figure 2A) and U3 (data not shown), were unaffected by deletion of *RRP6*. The dicistronic snoRNA, snR190, which is cotranscribed with U14, was also not affected (Figure 2A).

Primer extension revealed that the position of the 5' end was unaffected for each of these snoRNAs (Figure 2B) indicating that the altered gel mobility represents a failure in the 3' trimming of the snoRNA. For U24 the presence of a 3' extension was confirmed by RNase protection (data not shown).

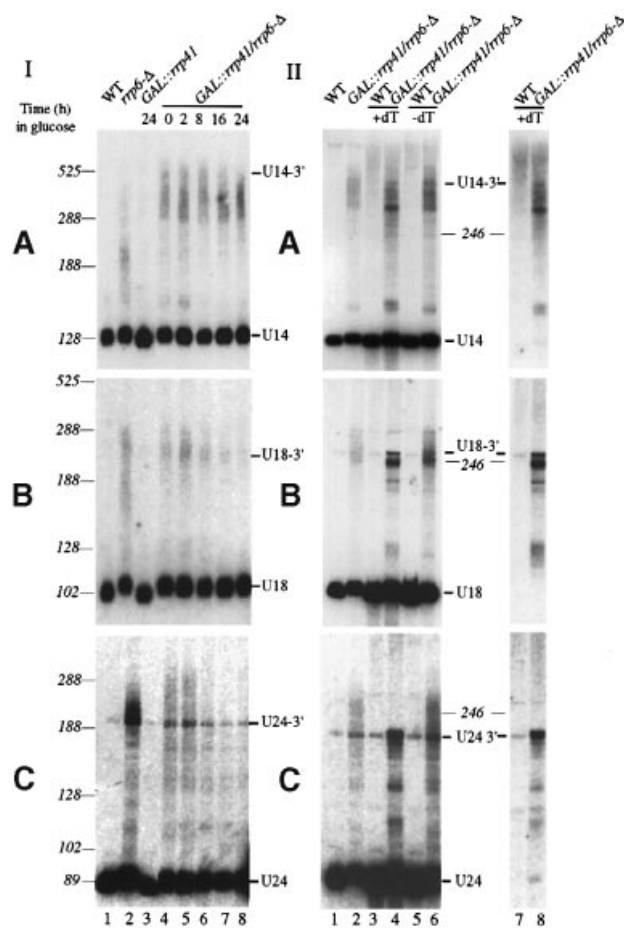
Strains individually depleted for each of the other exosome components or Dob1p, or carrying the *rrp4-1*, *mtr3-1* or *dob1-1* mutations at non-permissive temperature, were analysed for processing of U14, U18, U24 and snR190 (shown for *GAL::rrp41* and *GAL::rrp45* in Figure 2C and *dob1-1* in Figure 2A). No clear alteration in the length of the mature snoRNAs was observed, showing that depletion or mutation of other individual components of the exosome or Dob1p does not inhibit snoRNA 3' end trimming. Double-mutant strains lacking Rrp6p and depleted of either Rrp41p or Rrp45p were also analysed (shown for *GAL::rrp41/rrp6-Δ* in Figure 3I). The length of the 'almost-mature' snoRNAs in these strains was the same as in strains lacking only Rrp6p. We conclude that 3' trimming of the snoRNAs requires specifically the Rrp6p component of the exosome complex. In addition, longer extended forms were observed for U14, U18 and U24, but not for snR190 (Figure 3 and data not shown).



**Fig. 2.** Deletion of *RRP6* inhibits 3' trimming of pre-snoRNAs.

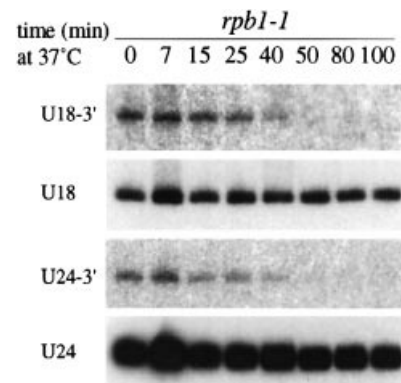
(A and C) Northern hybridization of snoRNAs. (B) Primer extension on snoRNAs. RNA was extracted from the *RRP6* and *rrp6-Δ* strains following growth at 30°C, from the *dob1-1* strain 6 h after transfer to 37°C and from the *GAL::rrp41* and *GAL::rrp45* strains following growth for 24 h on glucose medium. The gel migration shown for snR10 and snR190 in (A) is longer than that shown for the other, smaller RNA species to confirm that these were not extended in the *rrp6-Δ* strain.

Yeast U18 and U24 are intron encoded (Maxwell and Fournier, 1995; Qu *et al.*, 1995; Kiss-László *et al.*, 1996) and are synthesized predominantly from the debranched intron lariats (Ooi *et al.*, 1998; Petfalski *et al.*, 1998). As previously reported (Petfalski *et al.*, 1998), the species corresponding to the introns that are 3' unprocessed but 5' processed to the end of the mature snoRNAs [U18-3' (253 nt) and U24-3' (192 nt)] are detected in the wild-type strain (Figure 3IIB and IIC, lanes 1, 3, 5 and 7). Both U24-3' and U18-3' can be detected on Northern blots with probes that hybridize specifically with 3' extended species (Figure 3II, lanes 7 and 8) and both are lost in strains carrying mutations in the 5'→3' exonucleases Rat1p and Xrn1p (Petfalski *et al.*, 1998). In strains lacking Rrp6p, the level of U24-3' was increased and ladders of intermediates appeared both below and above



**Fig. 3.** Long 3' extended forms of snRNAs accumulate in exosome mutants. (I and II) Northern hybridization with probes directed against (A) mature U14 (202); (B) mature U18 (215); (C) mature U24 (204). RNA was extracted from the *RRP6* and *rrp6-Δ* strains following growth at 30°C, from the *GAL::rrp41* strain following growth for 24 h on glucose medium and from the *GAL::rrp41/rrp6-Δ* strain following transfer from RSG medium (0 h) to glucose medium for the times indicated. The positions of migration of SCR1 (525 nt), 7S pre-rRNA (288 nt), snR10 (246 nt) and 5.8S + 30 pre-rRNA (188 nt) determined by hybridization of the same filters are indicated as size markers. Mature U18 is 102 nt, U24 is 89 nt and U14 is 126 nt. (II) RNase H treatment of RNA samples. Lanes 1 and 2, untreated samples; lanes 3 and 4, samples treated with RNase H and oligo(dT); lanes 5 and 6, samples treated with RNase H in the absence of added oligo(dT); lanes 7 and 8, the samples shown in lanes 3 and 4 were rehybridized with probes across the ends of the mature snoRNAs that hybridize specifically with 3' extended species (210, 206, 213). RNA from the wild type (WT) and *GAL::rrp41/rrp6-Δ* 2 h samples shown in (I) was used.

this species (Figure 3IC, lane 2). To test the possibility that the species observed above U24-3' correspond to polyadenylated forms, RNA from the wild-type and the *GAL::rrp41/rrp6-Δ* strain 2 h after transfer to glucose medium was deadenylated *in vitro* with oligo(dT) and RNase H. On deadenylation the heterogeneous species observed above the U24-3' band were lost and the U24-3' signal was increased (Figure 3IIC, compare lanes 4 and 6). Strains depleted of Rrp41p (Figure 3IC, lane 3) did not clearly accumulate U24-3', whereas accumulation of U24-3' and a ladder of smaller intermediates were detected on depletion of Rrp45p and in a *rrp4-1* strain (data not shown), demonstrating that processing of U24-3' is not specific for Rrp6p. Extended forms of U24-3'



**Fig. 4.** Transcriptional inhibition leads to the loss of the 3' extended snoRNA species. RNA was extracted from the *rbp1-1* strain pre-grown at 23°C following transfer to 37°C for the times indicated. Northern hybridization was performed with oligonucleotide probes directed against mature U18 (205), mature U24 (214), U18-3' (206) and U24-3' (213).

were not observed in these strains, and the formation of polyadenylated species may be specific for *rrp6-Δ*. It is notable that *RRP6* was originally identified as a mutation that suppressed a defect in polyadenylation (Briggs *et al.*, 1998).

For U18, heterogeneous 3' extended species were observed in the *rrp6-Δ* strain, and these were longer and more discrete in the *GAL::rrp41/rrp6-Δ* double mutant (Figure 3IB). These were estimated to be in the range 250–300 nt. Following deadenylation, the U18-3' RNA is clearly seen to be strongly accumulated in the *GAL::rrp41/rrp6-Δ* strain, with a ladder of intermediates extending up to this position (Figure 3IIB, lanes 4 and 8). The same species are detected in the non-deadenylated sample (Figure 3IIB, lane 6), but are less clear.

snR190 and U14 are cotranscribed, separated by Rnt1p cleavage and then 5' processed by Rat1p (Zagorski *et al.*, 1988; Chanfreau *et al.*, 1998b; Petfalski *et al.*, 1998). Extended forms of U14 were observed in *rrp6-Δ* strains but were both more abundant and longer in *GAL::rrp41/rrp6-Δ* strains that were also depleted of Rrp41p (Figure 3A). On deadenylation these species formed a ladder up to the U14-3' RNA, which was estimated to be ~525 nt in length. It seems probable that U14-3' extends to the transcription termination site, which has not yet been located. Extended snoRNAs were not observed in strains depleted of Dob1p (data not shown).

To determine whether U18-3' and U24-3' species are processing intermediates or dead-end products, a transcription inhibition experiment was performed (Figure 4). RNA was recovered from a *rbp1-1* strain, carrying a *ts* mutation in RNA polymerase II, at time points after shift to the non-permissive temperature. Following transcription inhibition, U18-3' and U24-3' were progressively lost, indicating that they are normal processing intermediates.

We conclude that the exosome participates in the 3' processing of snoRNAs, processing the primary transcript of the dicistronic snR190-U14 snoRNAs and the debranched intron lariats containing the U18 and U24 snoRNAs. As seen for 5.8S synthesis, snoRNA processing is at least biphasic. Initial processing is partially inhibited, but not blocked, by different mutations in the exosome, whereas trimming of the final 3 nt specifically requires

Rrp6p. The additive effect of the *rrp41* and *rrp6* mutations on U18 and U14 synthesis is, however, in contrast to the epistatic interactions observed for 5.8S processing. In the absence of Rrp6p some polyadenylation of processing intermediates also occurs.

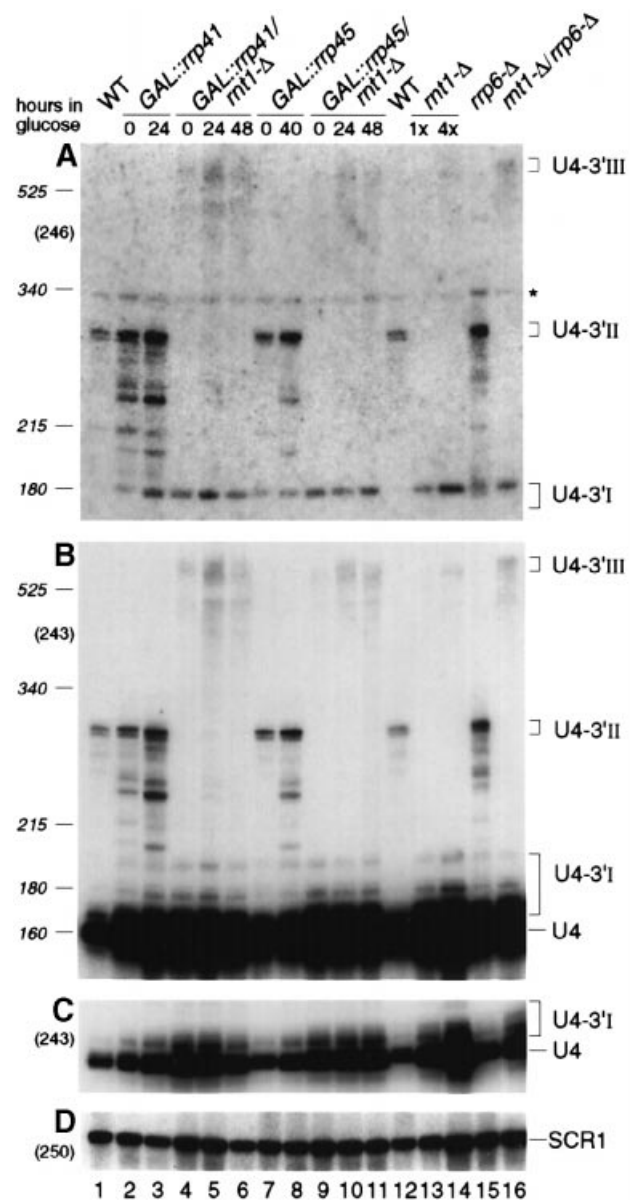
### snRNA synthesis

Since a number of snRNAs have been shown to undergo 3' processing we analysed snRNA species for alterations in exosome mutants.

Analysis of the U4 snRNA showed the existence of low levels of two longer forms (U4-3'I and U4-3'II) in the wild type (Figure 5, lanes 1 and 12). Both species were detected with either the internal U4 probe (oligo 243) (Figure 5B) or a probe complementary to the sequence across the 3' end of the mature U4 (oligo 246), which hybridizes only to 3' extended species (Figure 5A). U4-3'I is a set of heterogeneous species carrying short 3' extensions similar in length to those detected in the 6S pre-rRNA. Oligo 246 detects only the longer forms of U4-3'I. The U4-3'II species is ~140 nt larger than mature U4 and is presumably a normal processing intermediate since it is detected in wild-type cells. In strains depleted of the exosome components Rrp41p (Figure 5, lanes 2 and 3) or Rrp45p (Figure 5, lanes 7 and 8), or lacking Rrp6p (Figure 5, lane 15), the levels of U4-3'I and U4-3'II were increased and a ladder of intermediates was observed that extended from the size of U4-3'I to that of U4-3'II. The *GAL::rrp41* strain shows some accumulation of the extended species in the 0 h sample due to under-expression of Rrp41p (Figure 5, lane 2). The same species were accumulated in a strain depleted of Dob1p (data not shown). The accumulation of these intermediates was not stronger in the *rrp6*- $\Delta$  strain that was also depleted of Rrp41p or Rrp45p than in the *rrp6*- $\Delta$  single-mutant strain (data not shown). In strains lacking Rnt1p, U4-3'II RNA was absent whilst the levels of the U4-3'I species were increased. These observations suggested that the U4-3'II is generated by Rnt1p cleavage and acts as an entry site for the exosome complex.

Strains lacking Rnt1p also accumulated longer 3' extended forms of U4 (Figure 5, lanes 13 and 14; 4-fold more RNA was loaded in lane 14 to allow visualization of the longer 3' extended species). These were more abundant in *mnt1*- $\Delta$  strains that were also depleted of Rrp41p (Figure 5, lanes 4–6), Rrp45p (Figure 5, lanes 9–11) or Rrp6p (Figure 5, lane 16). Particularly in the *mnt1*- $\Delta$ /*rrp6*- $\Delta$  strain (Figure 5B, lane 16) these were seen to extend to a distinct species (U4-3'III) of ~590 nt. It seems likely that these represent intermediates in the 3' processing of U4 from U4-3'III, which most probably extends to the transcription termination site. Little effect on the 3' extended forms of U4 was seen on treatment with RNase H and oligo(dT) (data not shown).

Rnt1p cleaves on both sides of extended, imperfect stems with closing AGNN tetraloops (Chanfreau *et al.*, 1998a). Inspection of the U4 3' flanking region identified good consensus Rnt1p cleavage sites (Figure 6D). To determine whether this represented a genuine substrate for Rnt1p, cleavage was assayed *in vitro*. The region of the U4 3' flanking sequence from positions +49 to +235, which includes the putative cleavage sites, was transcribed *in vitro* (see Materials and methods). The sites cleaved



**Fig. 5.** Northern analysis of processing of U4 snRNA in *mnt1*- $\Delta$  and exosome mutants. RNA was extracted from strains carrying *GAL*-regulated constructs following transfer from permissive, RSG medium to repressive, glucose medium at 30°C for the times indicated, or from the wild-type (WT), *mnt1*- $\Delta$  and *rrp6*- $\Delta$  strains grown on glucose medium at 30°C. RNA was separated on a 6% polyacrylamide gel and hybridized with oligonucleotide probes. (A) Oligo 246 complementary to the region across the 3' end of the mature U4 snRNA. (B and C) Oligo 243 complementary to the mature U4 snRNA. (D) Oligo 250 complementary to the mature SCR1 RNA; the panels show successive hybridizations of the same filter. Probe names are indicated in parentheses on the left and the positions of detected RNA species are indicated on the right. (C) presents a weaker exposure of the same gel as (B). (C and D) present only relevant regions of the Northern blots. The species marked with \* in (A) probably results from a cross-hybridization with another RNA since it was not detected in (B). The amount of total RNA loaded in lane 14 is 4-fold higher than in lane 13 and other lanes. The positions of migration of SCR1 (525 nt), MRP RNA (340 nt), U5<sub>L</sub> (215 nt) and U5<sub>S</sub> (180 nt) determined by hybridization of the same filter are indicated as size markers. Mature U4 is 160 nt.

*in vitro* by recombinant His<sub>6</sub>-Rnt1p were identified by primer extension with oligo 249 (Figure 6A). Two major stops, corresponding to sites of cleavage between

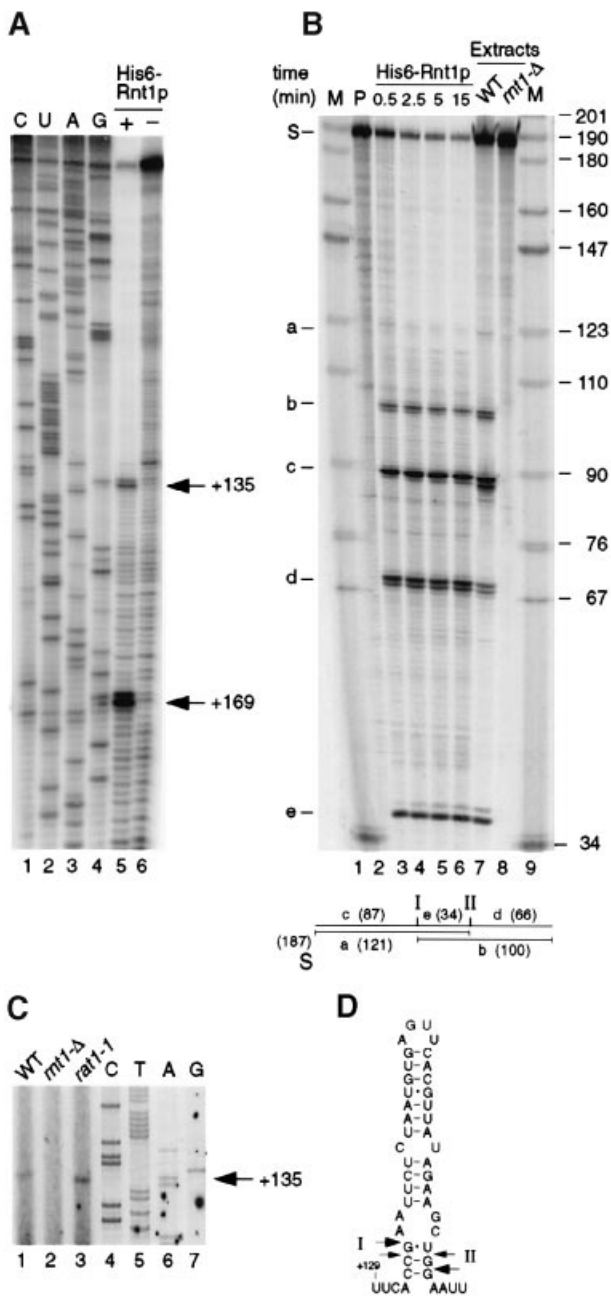


nucleotides +135/+136 (site I in Figure 6D) and +169/+170 (site II in Figure 6D) with respect to the 3' end of the mature U4, were detected following incubation with His<sub>6</sub>-Rnt1p (Figure 6A, lane 5); these sites were not detected in the no-enzyme controls (Figure 6A, lane 6). The sites lie on both sides of the predicted stem-loop structure and are in good agreement with the consensus for other Rnt1p cleavage sites (Figure 6D). Additional minor stops were detected one nucleotide 5', corresponding to positions +134/+135 and +168/+169 (smaller arrows in Figure 6D). To confirm that these stops represented sites of endonuclease cleavage, internally labelled RNA transcripts (Figure 6B) were also assayed using either recombinant His<sub>6</sub>-Rnt1p or extracts prepared from *RNT1*<sup>+</sup> and *rnt1*-Δ strains of yeast (see Materials and methods). Incubation with either recombinant Rnt1p (Figure 6B, lanes 3–6) or the yeast extract containing Rnt1p (Figure 6B, lane 7), resulted in the appearance of discrete cleavage

products that were not observed in the input RNA (Figure 6B, lane 2) or with the yeast extract lacking Rnt1p (Figure 6B, lane 8). The estimated sizes of these products are in good agreement with the predicted sizes of the products of cleavage at both sites I and II (Figure 6B; the predicted locations and sizes of the cleavage products are indicated below the gel).

To identify the position of *in vivo* processing, we performed primer extension using an oligo that hybridizes to the U4 flanking sequence 3' to the Rnt1p cleavage site (oligo 249). In the wild-type strain, a primer extension stop was detected between nucleotides +135 and +136, precisely matching *in vitro* cleavage site I (Figure 6C, lane 1). This stop was absent from the *rnt1*-Δ strain (Figure 6C, lane 2). *In vivo* cleavage could not be detected at site II, but the primer gave a high background in this region (data not shown). Following cleavage of the pre-rRNA in the 3'-ETS by Rnt1p, the excised 3' fragment is degraded 5'→3' by Rat1p (Kufel *et al.*, 1999). The primer extension stop observed in the *rat1-1* strain at position +135/+136 was stronger than in the wild type, supporting the conclusion that this is a site of *in vivo* endonuclease cleavage (Figure 6C, lane 3).

We conclude that Rnt1p cleaves the 3' flanking sequence of the U4 snRNA. The predominant *in vivo* cleavage is at position +135/+136, generating the U4-3'II RNA. Cleavage may also occur at +169/+170, although no species corresponding to this cleavage was detected by Northern hybridization. The +135/+136 cleavage acts as a site of entry for the exosome complex. Since U4-3'II is detected in wild-type cells it is presumably a normal processing intermediate. In the absence of Rnt1p, longer transcripts are detected; these are also substrates for the exosome since they accumulate at higher levels in double-mutant strains, but can be efficiently processed to mature U4 by another pathway(s). It seems likely that the accumu-



**Fig. 6.** Rnt1p cleaves the 3' end of the U4 precursor. **(A)** Mapping of the *in vitro* Rnt1p cleavage sites. Primer extension was performed using oligo 249 on the model U4-3' RNA incubated with buffer (lane 6) or recombinant His<sub>6</sub>-Rnt1p (lane 5) as described in Materials and methods. DNA sequencing reaction on a PCR product encompassing the region of the 3' end of U4 from position +49 to +235, using the same primer, was run in parallel (lanes 1–4). The primer extension stops at positions +135 and +169 are indicated. **(B)** *In vitro* cleavage of an internally labelled model U4-3' RNA substrate by Rnt1p. <sup>32</sup>P-labelled U4-3' RNA was incubated at 23°C in the following conditions: lane 2, Rnt1p buffer; lanes 3–6, Rnt1p buffer with 20 ng of recombinant His<sub>6</sub>-Rnt1p for the times indicated; lane 7, with whole-cell extract from a wild-type (WT) strain of yeast; lane 8, with whole-cell extract from a *rnt1*-Δ strain; lanes 1 and 9, RNA size markers. The positions of DNA size markers are indicated on the right of the gel. The cleavage products obtained are labelled a–e on the left and the predicted origin of each species is indicated below the gel. S: substrate, 187 nt; a: 5' end of transcript to site II, 121 nt; b: 3' end of transcript to site I, 100 nt; c: 5' end of transcript to site I, 87 nt; d: 3' end of transcript to site II, 66 nt; e: site I to site II, 34 nt. **(C)** Mapping of the Rnt1p 5' cleavage site *in vivo*. Primer extension analysis using the 3' end of the pre-U4 was performed using primer 249 hybridizing downstream of position +212. RNA was extracted from wild-type (WT, lane 1) and *rnt1*-Δ (lane 2) strains grown at 30°C and from a *rat1-1* strain following transfer to 37°C for 2 h (lane 3). DNA sequencing reactions were run in parallel (lanes 4–7). The primer extension stop at position +135 is indicated. **(D)** Computer-predicted RNA structure in the U4 3' flanking region that contains the Rnt1p cleavage sites. The major cleavage sites I (between nucleotides 135 and 136) and II (between nucleotides 169 and 170) are indicated by arrows.

**Table I.** Phosphoimager quantitation of Northern hybridization data from Figures 5 and 7

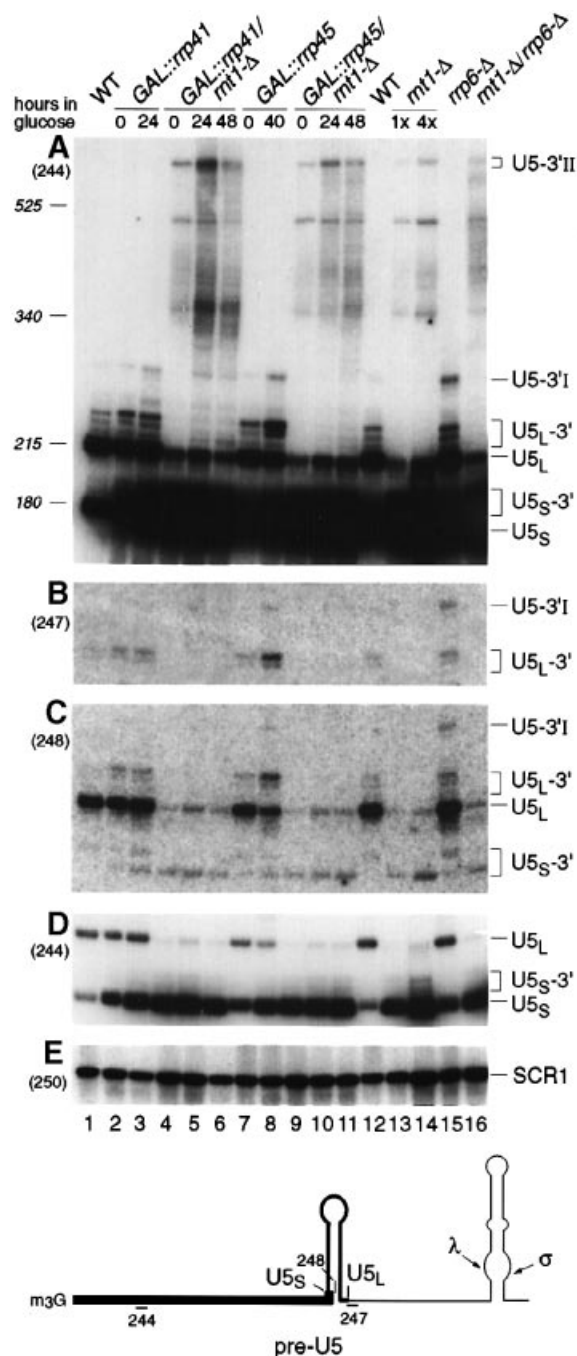
	WT	<i>GAL::rrp41</i> (24 h)	<i>GAL::rrp45</i> (40 h)	<i>rrp6-Δ</i>	<i>rnt1-Δ</i>
U5 <sub>L</sub>	1.27	1.63	0.63	1.9	0.025
U5 <sub>S</sub>	1	13.5	6.1	4.5	9.4
U5 <sub>S</sub> /U5 <sub>L</sub>	0.79	8.28	9.68	2.37	376
U4	1	3.8	2.9	2.5	6.5
SCR1	1	1.05	1.27	1.59	1.46

U5<sub>L</sub> and U5<sub>S</sub> levels are expressed relative to the signal for U5<sub>S</sub> in the wild-type strains.

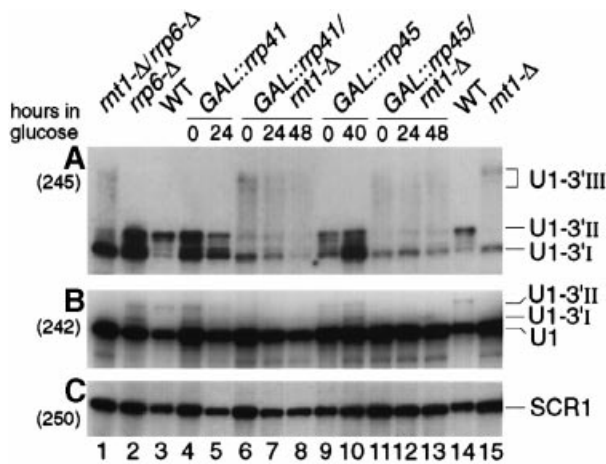
lation of the short 3'-extended U4-3'1 species is not a direct product of inhibition of the exosome complex, but is associated with the activation of an alternative processing pathway, since these species are increased relative to the wild type in strains lacking either exosome components or Rnt1p.

Quantitation of the Northern data (Table I) revealed that the mature U4 accumulates above the wild-type level in both the exosome and Rnt1p mutant strains relative to the cytoplasmic 7SL RNA homologue SCR1 (Table I and Figure 5D) or the nucleolar MRP RNA (data not shown). We conclude that a significant fraction of the U4 or pre-U4 population is normally degraded by an exosome-dependent pathway in wild-type strains.

Two forms of the U5 snRNA, U5<sub>L</sub> and U5<sub>S</sub>, which differ at their 3' ends, are observed in wild-type yeast strains (Patterson and Guthrie, 1987). Species with short, heterogeneous 3' extensions were observed for both U5<sub>L</sub> (U5<sub>L</sub>-3') and U5<sub>S</sub> (U5<sub>S</sub>-3'). These species are detected with an internal U5 probe (oligo 244; Figure 7A and D), and also with probes across the 3' end of U5<sub>L</sub> (oligo 247; Figure 7B) or across the 3' end of U5<sub>S</sub> (oligo 248; Figure 7C), although only the longer forms are detected by oligos 247 and 248. Both U5<sub>L</sub>-3' and U5<sub>S</sub>-3' were detected at low levels in wild-type strains (Figure 7, lanes 1 and 12). These, and a longer species, U5-3'1, were accumulated in strains depleted of Rrp41p (Figure 7, lanes 2 and 3) or Rrp45p (Figure 7, lanes 7 and 8), or lacking Rrp6p (Figure 7, lane 15) and were mildly accumulated in strains depleted of Dob1p (data not shown). In *rrp6-Δ* strains that were also depleted of Rrp41p, the accumulation of these intermediates was not clearly different from the *rrp6-Δ* single-mutant strains (data not shown). Cleavage sites for Rnt1p are present in the U5 3' flanking region (Chanfreau *et al.*, 1997). In the absence of Rnt1p, U5<sub>L</sub>-3' and U5<sub>S</sub>-3' were absent and the level of U5<sub>L</sub> was strongly reduced (Figure 7D, lanes 13 and 14 and Table I). Based on their gel mobilities, the U5<sub>L</sub>-3' species extend up to a position close to the 5' Rnt1p cleavage site [site λ in Figure 7 and Chanfreau *et al.* (1997)], while the larger U5-3'1 species extends to a position close to the 3' Rnt1p cleavage site [site σ in Figure 7 and Chanfreau *et al.* (1997)]. The accumulation of these species in exosome mutants suggests that the Rnt1p cleavages normally act as entry sites for the exosome. In the absence of Rnt1p (Figure 7A, lanes 13 and 14), longer 3' extended forms of U5 were detected in a ladder to a species designated U5-3'II of ~690 nt (Figure 7A). These were strongly increased in *rnt1-Δ* strains also depleted of Rrp41p, Rrp45p or Rrp6p (Figure



**Fig. 7.** Northern analysis of processing of U5 snRNA in *rnt1-Δ* and exosome mutants. Strains were grown and RNA was prepared as described for Figure 5. (A and D) Hybridization with oligo 244 complementary to the mature U5 snRNA. (B) Hybridization with oligo 247 complementary to the region across the 3' end of mature U5<sub>L</sub> snRNA. (C) Hybridization with oligo 248 complementary to the region across the 3' end of mature U5<sub>S</sub> snRNA. Probe names are indicated in parentheses on the left and the positions of detected RNA species are indicated on the right. (B, C and D) present only relevant regions of the Northern blot. The amount of total RNA loaded in lane 14 is 4-fold greater than in lane 13 and other lanes. (E) Oligo 250 complementary to the mature SCR1 RNA; the panels show successive hybridizations of the same filter. The positions of migration of SCR1 (525 nt), MRP RNA (340 nt) determined by hybridization of the same filter are indicated as size markers. Mature U5<sub>L</sub> is 215 nt and U5<sub>S</sub> is 180 nt.



**Fig. 8.** Northern analysis of processing of U1 snRNA in *mt1-Δ* and exosome mutants. Strains were grown and RNA was prepared as described for Figure 5. (A) Hybridization with oligo 245 complementary to the region across the 3' end of mature U1 snRNA. (B) Hybridization with oligo 242 complementary to the mature U1 snRNA. Probe names are indicated on the left and the positions of RNA species detected are indicated on the right. (C) Oligo 250 complementary to the mature SCR1 RNA; the panels show successive hybridizations of the same filter.

7A). It seems likely that U5-3'II extends to the transcription termination site and that the ladder represents intermediates in its 3' processing to U5<sub>S</sub>. Little effect on the 3' extended forms of U5 was seen on treatment with RNase H and oligo(dT) (data not shown).

The signal for U5<sub>S</sub> was strikingly increased relative to U5<sub>L</sub>, SCR1 RNA or MRP RNA (data not shown) in strains depleted of Rrp41p (Figure 7, lanes 2 and 3), Rrp45p (Figure 7, lanes 7 and 8) or Dob1p (data not shown), or lacking Rrp6p (Figure 7, lane 15) or Rnt1p (Figure 7D, lane 13). These results are quantified in Table I; the U5<sub>S</sub>:U5<sub>L</sub> ratio is changed >10-fold on depletion of Rrp41p or Rrp45p and the total amount of U5<sub>L</sub> + U5<sub>S</sub> synthesized is substantially more than wild type. We conclude that when the normal processing pathway is active, a large fraction of the pre-U5<sub>S</sub> or U5<sub>S</sub> population is degraded. U5<sub>S</sub> is unusual among snRNAs in not having a terminal stem-loop structure beyond the Sm binding site to stabilize the 3' end, which may make it particularly liable to degradation.

A 3' extended form of the U1 snRNA that is likely to extend to the Rnt1p cleavage site has been reported for wild-type strains (Seipelt *et al.*, 1999). This presumably corresponds to the species designated U1-3'II that we detect on Northern hybridization (Figure 8, lanes 3 and 14), which is absent in *mt1-Δ* strains (Figure 8, lane 15). The yeast U1 RNA is 568 nt in length (Kretzner *et al.*, 1987; Siliciano *et al.*, 1987), substantially larger than U4 and U5, and the precursors are therefore less well separated in Figure 8 than in Figures 5 and 7. In addition, we see shorter 3' extended forms, U1-3'I, which are strongly accumulated in strains depleted of Rrp41p (Figure 8, lanes 4 and 5) or Rrp45p (Figure 8, lanes 9 and 10), or lacking Rrp6p (Figure 8, lane 2). Unlike the longer U1-3'II species, the shorter U1-3'I persists in *mt1-Δ* strains (Figure 8, lane 15). The *mt1-Δ* strain accumulates longer, heterogeneous 3' extended species, U1-3'III, which are reported to be polyadenylated (Seipelt *et al.*, 1999). As for U4 and U5,

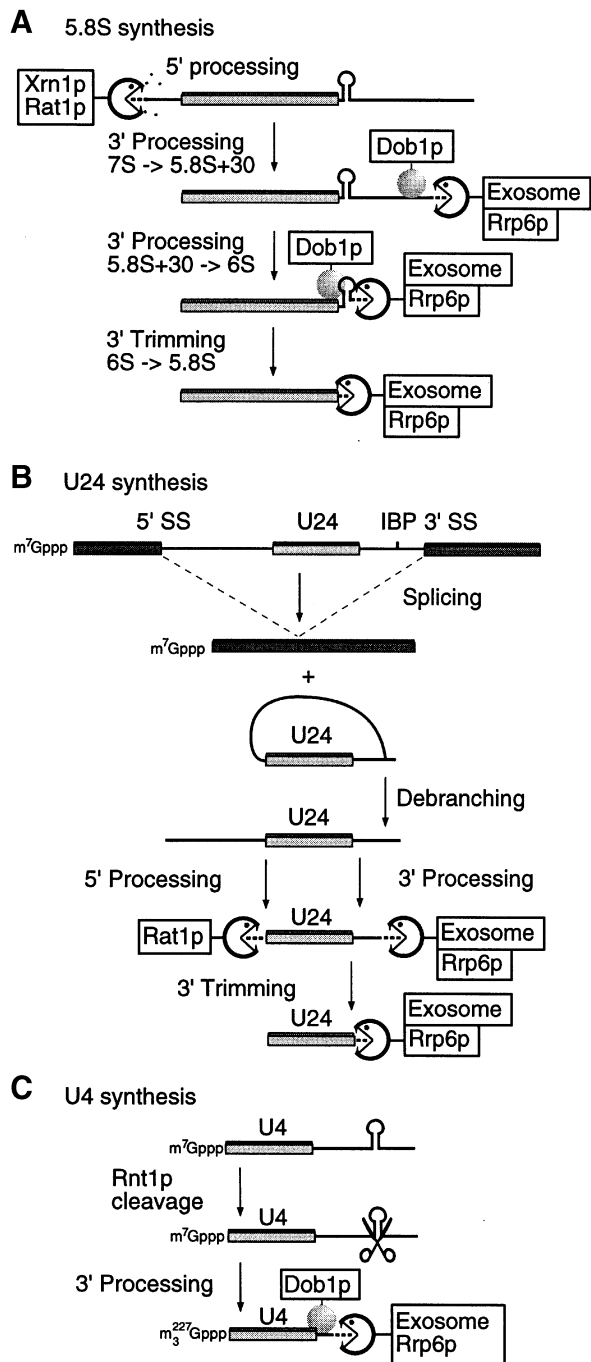
these long species are more strongly accumulated in *mt1-Δ* strains that also lack exosome components. In contrast to the results for U4 and U5, the mature U1 is not clearly accumulated above wild-type levels in the exosome mutants.

We conclude that 3' processing of U1, like U4 and U5, involves 3' cleavage by Rnt1p and processing by the exosome, although mutations in individual components of the complex do not block processing.

## Discussion

The characterization of the exosome complex raised an obvious question: why are so many different exonucleases present in the complex? Possible explanations are that multiple enzymes might function in the processing of single RNA substrates or that different enzymes might be preferentially active on different substrates. We have presented initial evidence for both of these phenomena, as well as identifying a large number of additional substrates for the complex.

During the processing of the 7S pre-rRNA to the 5.8S rRNA, the specificity of the exonuclease appears to change at least twice (see Figure 9A), a phenomenon that we refer to as exonuclease hand-over. Moreover, the effects on the 6S pre-rRNA vary between different mutants, showing that they do not have identical functions. The putative RNA helicase Dob1p/Mtr4p appears to be required for each of these processing steps. The 5.8S + 30 pre-rRNAs extend to the 3' side of a predicted stable stem-loop structure, which includes the two terminal nucleotides of the mature 5.8S rRNA (Yeh and Lee, 1990). It is, however, unclear whether processing is inhibited by the stem-loop structure itself, or by the consequent very close proximity to the 3' region of the mature 5.8S rRNA, which is likely to be associated with ribosomal proteins. The simplest interpretation of the data would be that all of the essential exosome components, but not Rrp6p, are required for normal processing of the 7S pre-rRNA to 5.8S + 30. At this point, Rrp6p may take over the major role and process the 5.8S + 30 species to 6S pre-rRNA. In the absence of the exosome components some other activity is able to digest the 7S pre-rRNAs partially, producing the observed ladder of intermediates. This alternative activity is not provided by Rrp6p, since these species are detected in the *GAL::rrp41/rrp6-Δ* and *GAL::rrp45/rrp6-Δ* strains. Moreover, both the remaining exosome components and any alternative activity can poorly process the 5.8S + 30 pre-rRNAs since these accumulate to high levels in the *rrp6-Δ* strain. It is notable that the 3' end of this species is predicted to lie in a stem structure, which may inhibit its processing if released as a free RNA. Unlike the larger intermediates that are specifically detected in the exosome mutants, the 6S pre-rRNAs are detected in the wild type. This may represent a site at which rapid processive processing is normally replaced by slower distributive trimming of the RNAs. The heterogeneity shown between different exosome mutant strains indicates that 6S pre-rRNA is processed by the exosome, rather than being the product of an alternative pathway that is activated in the absence of exosome activity. The pattern of intermediates observed during degradation of the 5' ETS region of the pre-rRNA also



**Fig. 9.** Models for RNA processing pathways. (A) Processing of the 7S pre-rRNA to 5.8S rRNA. The mature rRNA is shown as a box and the transcribed spacers as lines. Processing of the 7S pre-rRNA to 5.8S + 30 requires all essential components of the exosome complex. Processing from 5.8S + 30 to 5.8S + 8 (6S pre-rRNA) specifically requires Rrp6p. The final trimming to the mature 5.8S again requires multiple exosome components. Each step requires the putative RNA helicase Dob1p/Mtr4p. 5.8S + 30 lies at the 3' side of a predicted stem-loop structure. (B) Processing of the U24 snoRNA from the debranched intron lariat following mRNA splicing. The 5' and 3' exons are shown as dark boxes, the mature U24 is shown as a lighter box and the remainder of the intron as lines. (C) Processing of the U4 snRNA. An Rnt1p cleavage site lies in the 3' flanking sequence and may act as an entry site for the exosome, acting together with Dob1p/Mtr4p. The timing of cap trimethylation of U4 is not clear. 3' processing of snoRNAs and snRNAs is not blocked by mutation of individual components of the exosome indicating that other extrinsic or intrinsic activities can functionally replace these.

varies amongst the different exosome mutants, suggesting that exonuclease hand-over may be occurring during this processing.

Analysis of pre-snRNA processing reveals a complex picture. Each of the RNA polymerase II-transcribed snRNAs in yeast, U1, U2, U4 and U5, has a cleavage site for Rnt1p in the 3' flanking region (Chanfreau *et al.*, 1997; Abou Elela and Ares, 1998; Seipelt *et al.*, 1999 and this work). For U1, U4 and U5 this may act as an entry site for the exosome complex, acting together with the Dob1p RNA helicase (see Figure 9C). However, in no case is synthesis of the snRNA blocked by the inhibition of cleavage by Rnt1p or by mutations in exosome components, indicating that alternative processing pathways exist.

In strains lacking Rnt1p the synthesis of long 3' extended forms of U1 and U2 has been reported (Abou Elela and Ares, 1998; Seipelt *et al.*, 1999) and we show here that this is also the case for U4 and U5. During pre-rRNA processing, the 3' end of the 35S pre-rRNA is normally cleaved cotranscriptionally by Rnt1p (Abou Elela *et al.*, 1996; Kufel *et al.*, 1999). In the absence of Rnt1p the pre-rRNA transcripts extend to a position close to the site of transcription termination. The extended U4 and U5 species form a ladder to a discrete size that we speculate represents the transcription termination site. These are not detected in the wild type, suggesting that pre-U4 and pre-U5 may also be cleaved cotranscriptionally. The extended species are substrates for the exosome since they accumulate at higher levels in *rnt1-Δ* strains that are depleted for exosome components. For both the pre-rRNA and U4, the excised 3' fragments generated by Rnt1p cleavage are degraded by the 5'→3' exonuclease Rat1p, which also processes the 5' end of the 5.8S<sub>s</sub> rRNA and many snoRNAs (see Figure 9B) (Amberg *et al.*, 1992; Henry *et al.*, 1994; Petfalski *et al.*, 1998; Villa *et al.*, 1998).

3' maturation of snoRNAs that are excised from mRNA introns (U18 and U24) or from a dicistronic transcript (U14) also involves the exosome. U18 and U24 are predominately processed from the debranched intron lariat (Ooi *et al.*, 1998; Petfalski *et al.*, 1998; Villa *et al.*, 1998). In the exosome mutants we see accumulation of the species in which the 5' end of the snoRNA has been matured but the intron is 3' unprocessed (U18-3' and U24-3'), together with a ladder that probably represents intermediates in the 3' processing of these to the mature snoRNAs. The U18-3' and U24-3' RNAs also undergo some polyadenylation in strains lacking Rrp6p. In Rnt1p mutants, the 3' extended forms of U1 and U2 that are generated also become polyadenylated (Abou Elela and Ares, 1998; Seipelt *et al.*, 1999), so this seems to be a general phenomenon in yeast. We conclude that 3' processing of the debranched intron lariats containing U18 and U24 normally involves the exosome (see Figure 9B). Apparent intermediates in the 3' processing of U14 are also observed, particularly in strains lacking both Rrp6p and Rrp41p. These may extend to the transcription termination site, but this has not yet been localized. Final trimming of each of these snoRNAs specifically requires Rrp6p. This activity apparently cannot be substituted by other exonucleases, since the entire snoRNA population is shifted in size by ~3 nt. This trimming activity is not clearly inhibited by mutations in other single components

**Table II.** Yeast strains used in this work

Strain	Genotype	Reference
BMA 38	<i>MATα ade2-1/ade2-1 his3-Δ200/his3-Δ200 leu2-3, 112/leu2-3, 112 trp1-1/trp1-1 ura3-1/ura3-1 can1-100/can1-100</i>	Baudin et al. (1993)
YCA12	<i>MATα ade2-1 his3-Δ200 leu2-3, 112 trp1-1 ura3-1 can1-100 RRP6::Kl TRP1</i>	Allmang et al. (1999)
YDL401	<i>MATα his3-Δ200 leu2-Δ1 trp1 ura3-52 gal2 galΔ108</i>	Lafontaine and Tollervey (1996)
P118	as YDL401 but <i>GAL10::prot.A-RRP41</i>	Lafontaine and Tollervey (1996)
YCA20	as YDL401 but <i>GAL10::RRP45</i>	Allmang et al. (1999)
YCA30	as YCA20 but <i>RRP6::Kl TRP1</i>	this study
YCA31	as P118 but <i>RRP6::Kl TRP1</i>	this study
YJK10	as YDL401 but <i>RNT1::TRP1</i>	this study
YJK11	as P118 but <i>RNT1::TRP1</i>	this study
YJK12	as YCA20 but <i>RNT1::TRP1</i>	this study
YJK13	as YCA12 but <i>RNT1::TRP1</i>	this study
GAL::DOB1	<i>MATα ura3-1 ade2-1 his3-11,15 leu2-3, 112 trp1-1 dob1::HIS3MX6 + [pAS24-DOB1]</i>	de la Cruz et al. (1998)
<i>rat1-1</i>	<i>MATα, his3-Δ200, leu2-Δ1, ura3-52, rat1-1</i>	Amberg et al. (1992)
RP582	<i>MATα leu2-3, 112 ura3-52 rpb1-1</i>	Decker and Parker (1993)

of the exosome. It is notable that each of the enzymatic activities shown to be involved in pre-snoRNA processing, the exosome, the 5'→3' exonuclease Rat1p and the endonuclease Rnt1p, also participate in pre-rRNA processing.

There are apparent similarities between 3' processing of the 5.8S rRNA, snoRNAs and snRNAs. In each case there is a downstream cleavage, by endonuclease or splicing, which acts as a site of entry for exonucleases. Processing is at least biphasic with short 3' extended forms accumulating, and each appears to involve the activities of more than one component of the exosome. In each case, other activities can partially substitute for the mutant exosome components. For the 7S pre-rRNA, and particularly the 5.8S + 30 pre-rRNA, this is inefficient, and synthesis of mature 5.8S is strongly inhibited. The final trimming of snoRNAs is apparently completely dependent on the activity of Rrp6p, but the processing of further 3' extended pre-snoRNAs and pre-snRNAs can be carried out by other activities with good efficiency, as shown by the wild-type levels of the mature RNAs. Indeed, U4, and particularly U5s, are synthesized at substantially elevated levels in exosome mutants, indicating competition between synthesis of mature snRNA and degradation of the precursors in the wild type.

It is unclear whether residual processing in the exosome mutant strains is carried out by other components of the complex. It is notable that double-mutant strains lacking both Rrp6p and Rrp41p show stronger phenotypes for some activities (e.g. 3' processing of U14 and U18) than does either single mutant, indicating that the absence of one component does not necessarily inactivate the entire complex. Alternatively, the yeast genome contains several other predicted 3'→5' exonucleases (Mian, 1997; Moser et al., 1997) that may be able to partially substitute for the exosome. The combination of mutations in these genes with mutations in exosome components will now be needed to analyse their interactions and substrates.

## Materials and methods

### Strains

Growth and handling of *S.cerevisiae* were by standard techniques. The transformation procedure was according to Gietz et al. (1992). Except where stated, strains were grown in liquid and solid minimal medium containing 0.67% yeast nitrogen base (Difco) and 2% glucose. *GAL*-

regulated strains were pre-grown in RSG medium (2% raffinose, 2% sucrose, 2% galactose, 0.67% yeast nitrogen base) and harvested at intervals following a shift to 2% glucose.

Yeast strains used and constructed in this study are listed in Table II. To construct the double mutants *GAL10::rrp45/rrp6-Δ* (YCA30) and *GAL10::rrp41/rrp6-Δ* (YCA31), the *RRP6::TRP* disrupted allele was amplified by PCR from genomic DNA extracted from *rrp6-Δ* (YCA12). The PCR product was transformed into the corresponding conditional mutant strains P118 and YCA20, respectively. The amplification of *RRP6::TRP* was done with oligos: 5'RRP6, CAGTAATGAATATTAAT-GTTCATCTGAAGATAGACG; 3'RRP6, ATGGTGTGCATGGGGG-AGCCATAACTCCATGACACA. Strains YDL401, P118 and YCA20 were used to construct the mutants *rnt1-Δ* (YJK10), *GAL10::rrp41/rnt1-Δ* (YJK11) and *GAL10::rrp45/rnt1-Δ* (YJK12), using the same PCR strategy. Oligonucleotides 5'RNT1, 5'-GAAGACATATCCGAAGTG-ACA and 3'RNT1, 5'-GGATTTCTATACCCTCGAGGAG, complementary to sequences beyond the *RNT1* gene, were used for the amplification with genomic DNA extracted from *rnt1-Δ* strain generously provided by G.Chanfreau (Chanfreau et al., 1998b). Transformants were selected for Trp<sup>+</sup> prototrophy and were screened by PCR. The phenotypes of respective constructs were confirmed by Northern hybridization. Strain *rrp6-Δ/rnt1-Δ* (YJK13) was constructed by crossing YJK10 and YCA12 strains. The double-mutant strain was selected from dissected full tetrads by testing for the pre-rRNA and snoRNA processing phenotypes by Northern hybridization. Wild-type *RNT1* and *rnt1-Δ* sister strains (Chanfreau et al., 1998b) were used to prepare whole-cell extract. Strain *rat1-1* was kindly provided by C.Cole (Amberg et al., 1992).

### RNA extraction, Northern hybridization and primer extension

RNA was extracted as described previously (Tollervey and Mattaj, 1987). Northern hybridization (Tollervey, 1987) and primer extension (Beltrame and Tollervey, 1992) were as described previously. Standard 6 or 8% acrylamide gels were used to analyse low molecular weight rRNA species and primer extension reactions.

For pre-rRNA and rRNA analysis the following oligonucleotides were used: 017 5'-GCGTTGTTTCATCGATGC; 020 5'-TGAGAAGGAAATG-ACGCT; 033 5'-CGCTGCTACCAATGG; 041 5'-CTACTCGGTCA-GGCTC.

The oligonucleotides used for Northern blot hybridization and primer extensions on other small RNAs were as follows: 031 (MRP) 5'-AATAGAGGTACCAGGTCAAGAAGC; 201 (snR190) 5'-CGTCAT-GGTCGAATCGG; 202 (U14) 5'-TCACTCAGACATCCTAGG; 205 (U18) 5'-GTCAGATACTGTGATAGTC; 206 (U18-3') 5'-GCTCTG-TGTGCTATCGTC; 210 (U14-3') 5'-GTATACGATCACTCAGAC; 213 (U24-3') 5'-AAACCATTTCATCAGAG; 214 (U24) 5'-TCAGAGAT-CTTGGTGATAAT; 218 (snR10, 2'-O-methyl RNA) 5'-CUIUUAUU-UUICIUU; 236 (snR39) 5'-GGTGATAAGTTACGACAGC; 238 (snR41) 5'-GGGTTGTGACATGTAGTTA; 242 (U1) 5'-CACGCCTTCCGCG-CCGT; 243 (U4) 5'-CCGTGCATAAGGAT; 244 (U5) 5'-AATATG-GCAAGCCC; 245 (3'Ex-U1) 5'-TGTTCCATTTATTTCTGAAA; 246 (3'Ex-U4) 5'-AAAGATGAATATCGGTAATG; 247 (3'Ex-U5s) 5'-GAGAAAAAGGGCAGAAAAG; 248 (3'Ex-U5L) 5'-TAGAAAAGAT-AAACGCCCT; 249 (U4DS) 5'-GACACACAAGAGGAGAACAACCTC; 250 (SCR1) 5'-AAGACCCAGAACTACCTTG.

### RNase H treatment

Deadenylation was performed essentially as described (Decker and Parker, 1993). Samples of 30 µg of RNA were annealed with 750 ng oligo(dT) at 65°C for 1 h and digested with 6 U RNase H at 30°C for 1 h. The control samples were treated identically, except that the oligo(dT) was omitted.

### In vitro processing reactions

Synthetic U4-3' RNA was obtained by *in vitro* transcription as described (Chanfreau *et al.*, 1998b) using PCR product as template. PCR product was generated from genomic DNA using a forward primer carrying a T7 promoter (T7U4DS, 5'-GCGAATTCTAATACGACTCACTATAGG-AAGTAATATCAAAAAATAGG) and a reverse primer U4DS.

Whole-cell extracts were prepared from wild-type and *mtl1Δ* sister strains as described (Chanfreau *et al.*, 1998b).

Recombinant His<sub>6</sub>-Rnt1p was prepared by cloning a PCR-amplified *RNT1* gene into pET16B (Novagen), using *Nde*I and *Bam*HI restriction sites added into the primers (*Nde*I-Rnt1, 5'-GGGAATTCATATGGGCT-CAAAAGTAGCAGG; *Bam*-Rnt1, 3'-CGGGATCCTCAGCTGTAT-CTGAGAATTTCTTTCTTATTTC). Expression of His<sub>6</sub>-Rnt1p in *E. coli* strain BL21 was induced by addition of isopropyl-β-D-galactopyranoside at OD<sub>0.5</sub> (final concentration 0.5 mM). After 3 h of expression at 30°C, cells were harvested and pellets were kept at -80°C. Pellets were resuspended in 40 ml of Start buffer (20 mM sodium phosphate pH 7.0, 10 mM imidazole) and cells were further lysed by passing through a French press. Cell debris were pelleted and the supernatant was loaded into a Hi-Trap Chelating column (Pharmacia) pre-equilibrated with the Start buffer. The column was washed with 10 ml Start buffer and proteins were eluted with sodium phosphate buffer with increasing imidazole concentration (20, 40, 60, 100, 300 and 500 mM) or a linear gradient of imidazole (10–500 mM). Peak fractions were pooled and the protein was dialysed twice against the storage buffer (50% glycerol, 50 mM Tris-HCl pH 7.6, 200 mM KCl, 0.5 mM dithiothreitol, 0.5 mM EDTA pH 8.5). The protein was stored at a concentration of 2 mg/ml at -20°C and remained active for several months after storage.

*In vitro* processing of U4-3' RNA in cell extracts or with recombinant His<sub>6</sub>-Rnt1p, and mapping of the cleavage sites using primer extension, was performed as described (Chanfreau *et al.*, 1998a,b). Prior to the reaction, gel-purified RNA substrates (2 nM) were denatured for 2 min at 85°C in the Rnt1p buffer (50 mM Tris-HCl pH 7.6, 200 mM KCl, 0.1 mg/ml wheat-germ tRNA, 5 mM MgCl<sub>2</sub>) and cooled to 23°C. The cleavage reaction was performed either at 23°C using from 50 to 200 fmol of recombinant His<sub>6</sub>-Rnt1 or by incubation in the whole-cell extracts.

### Acknowledgements

J.K. was the recipient of a long-term EMBO fellowship. This work was supported by the Wellcome Trust.

### References

Abou Elela, S. and Ares, M., Jr (1998) Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *EMBO J.*, **17**, 3738–3746.

Abou Elela, S., Igel, H. and Ares, M., Jr (1996) RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell*, **85**, 115–124.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D. and Mitchell, P. (1999) The yeast exosome and human PM-Scl are related complexes of 3'→5' exonucleases. *Genes Dev.*, **13**, 2148–2158.

Amberg, D.C., Goldstein, A.L. and Cole, C.N. (1992) Isolation and characterization of *RAT1*: an essential gene of *Saccharomyces cerevisiae* required for the efficient nucleocytoplasmic trafficking of mRNA. *Genes Dev.*, **6**, 1173–1189.

Baker, R.E., Harris, K. and Zhang, K. (1998) Mutations synthetically lethal with *cep1* target *S. cerevisiae* kinetochore components. *Genetics*, **149**, 73–85.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. and Cullin, C. (1993) A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **21**, 3329–3330.

Beltrame, M. and Tollervey, D. (1992) Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J.*, **11**, 1531–1542.

Briggs, M.W., Burkard, K.T. and Butler, J.S. (1998) Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential

for efficient 5.8 S rRNA 3' end formation. *J. Biol. Chem.*, **273**, 13255–13263.

Caffarelli, E., Ares, M., Santoro, B., Fragapane, P. and Bozzoni, I. (1994) *In vitro* study of processing of the intron-encoded U16 small nucleolar RNA in *Xenopus laevis*. *Mol. Cell. Biol.*, **14**, 2966–2974.

Caffarelli, E., De Gregorio, E., Fatica, A., Prislei, S., Fragapane, P. and Bozzoni, I. (1996) Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reactions and the stability of the mature snoRNAs. *EMBO J.*, **15**, 1121–1131.

Cavaillé, J. and Bachellerie, J.-P. (1996) Processing of fibrillar-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by the common stem-box terminal structure. *Biochimie*, **78**, 443–456.

Cecconi, F., Mariottini, P. and Amaldi, F. (1995) The *Xenopus* intron-encoded U17 snoRNA is produced by exonucleolytic processing of its precursor in oocytes. *Nucleic Acids Res.*, **23**, 4670–4676.

Chanfreau, G., Elela, S.A., Ares, M., Jr and Guthrie, C. (1997) Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev.*, **11**, 2741–2751.

Chanfreau, G., Legrain, P. and Jacquier, A. (1998a) Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J. Mol. Biol.*, **284**, 975–988.

Chanfreau, G., Rotondo, G., Legrain, P. and Jacquier, A. (1998b) Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J.*, **17**, 3726–3737.

de la Cruz, J., Kressler, D., Tollervey, D. and Linder, P. (1998) Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in *Saccharomyces cerevisiae*. *EMBO J.*, **17**, 1128–1140.

Decker, C.J. and Parker, R. (1993) A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev.*, **7**, 1632–1643.

Gietz, D., St Jean, A., Woods, R.A. and Schiestl, R.H. (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.*, **20**, 1425.

Henry, Y., Wood, H., Morrissey, J.P., Petfalski, E., Kearsey, S. and Tollervey, D. (1994) The 5' end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *EMBO J.*, **13**, 2452–2463.

Kenna, M., Stevens, A., McCammon, M. and Douglas, M.G. (1993) An essential yeast gene with homology to the exonuclease-encoding *XRN1/KEM1* gene also encodes a protein with exoribonuclease activity. *Mol. Cell. Biol.*, **13**, 341–350.

Kiss, T., Bortolini, M.-L. and Filipowicz, W. (1996) Characterization of the intron-encoded U19 RNA, a new mammalian small nucleolar RNA that is not associated with fibrillarin. *Mol. Cell. Biol.*, **16**, 1391–1400.

Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M. and Kiss, T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.

Kretzner, L., Rymond, B.C. and Rosbash, M. (1987) *S. cerevisiae* U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. *Cell*, **50**, 593–602.

Kufel, J., Dichtl, B. and Tollervey, D. (1999) Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA*, **5**, 909–917.

Lafontaine, D. and Tollervey, D. (1996) One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. *Nucleic Acids Res.*, **24**, 3469–3472.

Larimer, F.W., Hsu, C.L., Maupin, M.K. and Stevens, A. (1992) Characterization of the *XRN1* gene encoding a 5'→3' exoribonuclease: sequence data and analysis of disparate protein and mRNA levels of gene-disrupted yeast cells. *Gene*, **120**, 51–57.

Li, Z., Pandit, S. and Deutscher, M.P. (1998) 3' exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **95**, 2856–2861.

Maxwell, E.S. and Fournier, M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.*, **35**, 897–934.

Mian, I.S. (1997) Comparative sequence analysis of ribonucleases HII, III, II PH and D. *Nucleic Acids Res.*, **25**, 3187–3195.

Mitchell, P., Petfalski, E. and Tollervey, D. (1996) The 3'-end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev.*, **10**, 502–513.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. and Tollervey, D. (1997) The exosome; a conserved eukaryotic RNA processing complex

- containing multiple 3'→5' exonuclease activities. *Cell*, **91**, 457–466.
- Moser,M.J., Holley,W.R., Chatterjee,A. and Mian,I.S. (1997) The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucleic Acids Res.*, **25**, 5110–5118.
- Noguchi,E. *et al.* (1996) Dis3, implicated in mitotic control, binds directly to Ran and enhances the GEF activity of RCC1. *EMBO J.*, **15**, 5595–5605.
- Ooi,S.L., Samarsky,D.A., Fournier,M.J. and Boeke,J.D. (1998) Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA*, **4**, 1096–1110.
- Patterson,B. and Guthrie,C. (1987) An essential yeast snRNA with a U5-like domain is required for splicing *in vivo*. *Cell*, **49**, 613–624.
- Petfalski,E., Dandekar,T., Henry,Y. and Tollervey,D. (1998) Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol. Cell. Biol.*, **18**, 1181–1189.
- Qu,L.H., Henry,Y., Nicoloso,M., Michot,B., Azum,M.C., Renalier,M.H., Caizergues-Ferrer,M. and Bachelierie,J.P. (1995) U24, a novel intron-encoded small nucleolar RNA with two 12 nt long, phylogenetically conserved complementarities to 28S rRNA. *Nucleic Acids Res.*, **23**, 2669–2676.
- Qu,L.H. *et al.* (1999) Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol. Cell. Biol.*, **19**, 1144–1158.
- Seipelt,R.L., Zheng,B., Asuru,A. and Rymond,B.C. (1999) U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway. *Nucleic Acids Res.*, **27**, 587–595.
- Siliciano,P.G., Jones,M.H. and Guthrie,C. (1987) *Saccharomyces cerevisiae* has a U1-like small nuclear RNA with unexpected properties. *Science*, **237**, 1484–1487.
- Tollervey,D. (1987) A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J.*, **6**, 4169–4175.
- Tollervey,D. and Mattaj,I.W. (1987) Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J.*, **6**, 469–476.
- Veldman,G.M., Klootwijk,J., van Heerhuizen,H. and Planta,R.J. (1981) The nucleotide sequence of the intergenic region between the 5.8S and 26S rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8S and 26S rRNA and the processing of the primary transcript. *Nucleic Acids Res.*, **9**, 4847–4862.
- Villa,T., Ceradini,F., Presutti,C. and Bozzoni,I. (1998) Processing of the intron-encoded U18 small nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exo- and endonucleolytic activities. *Mol. Cell. Biol.*, **18**, 3376–3383.
- Yeh,L.C. and Lee,J.C. (1990) Structural analysis of the internal transcribed spacer 2 of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. *J. Mol. Biol.*, **211**, 699–712.
- Yoo,C.J. and Wolin,S.L. (1997) The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell*, **89**, 393–402.
- Zagorski,J., Tollervey,D. and Fournier,M.J. (1988) Characterization of an SNR gene locus in *Saccharomyces cerevisiae* that specifies both dispensable and essential small nuclear RNAs. *Mol. Cell. Biol.*, **8**, 3282–3290.

Received June 24, 1999; revised and accepted August 10, 1999

## Precursors to the U3 Small Nucleolar RNA Lack Small Nucleolar RNP Proteins but Are Stabilized by La Binding

JOANNA KUFEL,<sup>1</sup> CHRISTINE ALLMANG,<sup>1</sup> GUILLAUME CHANFREAU,<sup>2†</sup> ELISABETH PETFALSKI,<sup>1</sup>  
DENIS L. J. LAFONTAINE,<sup>1</sup> AND DAVID TOLLERVEY<sup>1\*</sup>

Wellcome Trust Centre for Cell Biology, ICMB, The University of Edinburgh, Edinburgh EH9 3JR, Scotland,<sup>1</sup> and  
GIM-Biotechnologies, Institute Pasteur, 75724 Paris Cedex 15, France<sup>2</sup>

Received 10 March 2000/Returned for modification 10 April 2000/Accepted 2 May 2000

**Almost all small eukaryotic RNAs are processed from transiently stabilized 3'-extended forms. A key question is how and why such intermediates are stabilized and how they can then be processed to the mature RNA. Here we report that yeast U3 is also processed from a 3'-extended precursor. The major 3'-extended forms of U3 (U3-3'I and -II) lack the cap trimethylation present in mature U3 and are not associated with small nucleolar RNP (snoRNP) proteins that bind mature U3, i.e., Nop1p, Nop56p, and Nop58p. Depletion of Nop58p leads to the loss of mature U3 but increases the level of U3-3'I and -II, indicating a requirement for the snoRNP proteins for final maturation. Pre-U3 is cleaved by the endonuclease Rnt1p, but U3-3'I and -II do not extend to the Rnt1p cleavage sites. Rather, they terminate at poly(U) tracts, suggesting that they might be bound by Lhp1p (the yeast homologue of La). Immunoprecipitation of Lhp1p fused to *Staphylococcus aureus* protein A resulted in coprecipitation of both U3-3'I and -II. Deletion of *LHP1*, which is nonessential, led to the loss of U3-3'I and -II. We conclude that pre-U3 is cleaved by Rnt1p, followed by exonuclease digestion to U3-3'I and -II. These species are stabilized against continued degradation by binding of Lhp1p. Displacement of Lhp1p by binding of the snoRNP proteins allows final maturation, which involves the exosome complex of 3'→5' exonucleases.**

Eukaryotic cells contain a large number of stable RNA species, nearly all of which are synthesized by posttranscriptional processing from larger precursors. This has long been known for the highly abundant cytoplasmic RNAs, tRNAs, and rRNAs, but more recently it has become clear that is also the case for the small nuclear RNAs (snRNAs), which participate in pre-mRNA splicing, and the small nucleolar RNAs (snoRNAs), which participate in rRNA processing and modification. It is a long-standing mystery why cells use such a strategy, rather than simply terminating transcription at the end of the mature RNA sequence. We will offer a potential explanation for this, at least in the case of the U3 snoRNA.

Analyses of the 3' end processing of the 5.8S rRNA in *Saccharomyces cerevisiae* led to the identification of the exosome complex, composed of 11 different 3'→5' exonucleases (6, 36, 37; E. Petfalski and D. Tollervy, unpublished data). Subsequent work showed that the exosome participates in the 3' processing of other RNA substrates, including many snRNAs and snoRNAs (5, 55), and also participates in mRNA turnover (9). A homologous complex, designated the PM-Scl complex, is present in human cells and is a target for autoimmune antibodies (6).

In addition to the exosome, normal 3' processing of the U1, U2, U4, and U5 snRNAs involves cleavage by the endonuclease Rnt1p (1, 5, 14, 45), the yeast homologue of *Escherichia coli* RNase III (2). Rnt1p cleaves on both sides of extended stem-loop structures with closing AGNN tetraloops (15), and these cleavages are likely to act as entry sites for the exosome

complex, with the final trimming performed by the Rex exonucleases and/or the exosome component Rrp6p (5, 54). Rnt1p also acts to separate the individual pre-snoRNAs from polycistronic precursors (15, 16) and processes the 3' external transcribed spacer of the yeast pre-rRNA (2, 28).

Another 3' processing factor, the La phosphoprotein, was identified as the target of human autoimmune antibodies and was shown to bind to the poly(U) tracts located at the 3' ends of all RNA polymerase III transcripts (42, 48). La also binds to 3' extended precursors to human U1 and the yeast snRNAs (34, 58) and to internal sequences of several viral RNAs, in some cases at sequences that lack poly(U) tracts (4, 23). The yeast homologue of La, Lhp1p (La-homologous protein), is nonessential for viability but is required for normal 3' processing of tRNAs (56, 59). In the presence of Lhp1p, processing is endonucleolytic, whereas in the absence of Lhp1p this cleavage is inhibited and an alternative, exonucleolytic pathway takes over tRNA 3' maturation (59). Lhp1p also associates with the newly transcribed U6 snRNA, which is transcribed by RNA polymerase III (39).

Here we show how these factors collaborate in the 3' processing of the U3 snoRNA.

### MATERIALS AND METHODS

**Strains.** Growth and handling of *S. cerevisiae* were by standard techniques. The transformation procedure was as described elsewhere (21). Yeast strains used and constructed in this study are listed in Table 1. Wild-type *RNT1* and *mt1-Δ* sister strains (15) were used to prepare whole-cell extract. Strain *rat1-1* was kindly provided by C. Cole (7). The nonessential gene *LHP1* was disrupted and tagged with *Staphylococcus aureus* protein A ("ProtA" in construct designations) at the carboxy-terminal end of Lhp1p by a PCR strategy (29) in the haploid strain YDL401, using the *Kluyveromyces lactis* *URA3* marker.

The oligonucleotides used to construct and test the gene disruption and protein A tagging were 838 (5' LHP1::URA), 5'-TCTATTTGGTTCTACTGGAACTAAAGTAGCATCTGCAAAGAAGTAGAGAAGTTTGAGAGGGC; 839 (3' LHP1::URA), 5'-ATATGCTATGATAATGAGATACGAGAACCAGAA GAAACACAAGAAGCTGGGTAGAAAGATCGGTC; 840 (5' LHP1 test), 5'-A CAGAGTCGCATCTCATCGC; 841 (3' KI URA), 5'-GGTAGAAGATCGG

\* Corresponding author. Mailing address: Wellcome Trust Centre for Cell Biology, ICMB, Swann Building, King's Buildings, The University of Edinburgh, Edinburgh EH9 3JR, Scotland. Phone: 44 131 650 7092. Fax: 44 131 650 7040. E-mail: d.tollervy@ed.ac.uk.

† Present address: Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095-1569.



TABLE 1. Yeast strains used in this work

Strain	Genotype	Reference or source
YCA12	<i>MATa ade2-1 his3-Δ200 leu2-3,112 trp1-1 ura3-1 can1-100 RRP6::KI TRP1</i>	6
YDL401	<i>MATa his3Δ200 leu2Δ1 trp1 ura3-52 gal2 galΔ108</i>	29
D150	<i>MATa ura3-52 leu2-3,112 ade1-100 his4-519</i>	L. Guarente, personal communication
P118	As YDL401 but <i>GAL10::ProtA-RRP41</i>	29
YCA20	As YDL401 but <i>GAL10::RRP45</i>	6
YJK10	As YDL401 but <i>RNT1::TRP1</i>	5
YJK11	As P118 but <i>RNT1::TRP1</i>	5
YJK12	As YCA20 but <i>RNT1::TRP1</i>	5
YJK13	As YCA12 but <i>RNT1::TRP1</i>	5
<i>rat1-1</i>	<i>MATα his3-Δ200 leu2-Δ1 ura3-52 rat1-1</i>	6
RP582	<i>MATa leu2-3,112 ura3-52 rpb1-1</i>	18
ProtA-Nop1	<i>MATα ade leu trp lys ura3 nop1::URA3 pUN100-ProtA-NOP1</i>	26
ProtA-Nop58	<i>MATa ade2 ade3 leu2 ura3 can1 nop58::HIS3 pRS315-ProtA-NOP58</i>	20
ProtA-Nop56	<i>MATa ade2 ade3 leu2 ura3 nop56::HIS3 pRS315-ProtA-NOP56</i>	20
<i>GAL::nop58</i>	As YDL401 but <i>GAL10::NOP58</i>	30
<i>GAL::nop56</i>	As YDL401 but <i>GAL10::NOP56</i>	31
Lhp1p-ProtA	As YDL401 but <i>LHP1::ProtA</i>	This work
YCA 35	<i>MATa his3Δ200 leu2Δ1 trp1 ura3-52 gal2 galΔ108 LHP1::KI URA</i>	This work

TC; 842 (5' LHP1::ProtA); 5'-GAGGACTCTTCTGCCATTGCCGATGACGA TGAGGAGCACAAGGAGGCGTGGACAAACAATT; and 843 (3' LHP1::ProtA), 5'-TCCATTTTAACCAAGTAACGGTAATTTTAATACTAA TAAAAAAGCTGGGTAGAAGATCGGTC.

**RNA extraction, Northern hybridization, and primer extension.** For depletion of Rrp41p and Rrp45p, cells were harvested at intervals following the shift from RSG medium (2% galactose, 2% sucrose, 2% raffinose) to medium containing 2% glucose. Otherwise strains were grown in YPD medium. RNA was extracted as described previously (52). Northern hybridization and primer extension were as described previously (12, 51). Standard 6 or 8% acrylamide gels were used to analyze low-molecular-weight RNA species and primer extension reactions. For RNA hybridization and primer extension, the following oligonucleotides were used: 200 (U3), 5'-UUUAGGGGACUUGUU; 203 (5'U3), 5'-CUAUAGAAU GAUCCU; 218 (snR10), 5'-CUUUUUUUUUUICU; 230 (anti-U3sub6), 5'-GATTCCTATAGAAACACAG; 250 (scR1), 5'-ATCCCGCCGCCCTCCATC AC; 251 (3'Ex-U3), 5'-GTGGTAACTTGTC; 252 (U3ADS), 5'-TTTGTTT TCGCATCCGTCGCTC; 253 (U3DS), 5'-GGAGTCATACTATCAAGAAC; 254 (3'U3), 5'-CCAACCTGTGACACTGCCATT; 260 (U3 intron), 5'-CAAA AGCTGCTGCAATGG; 261 (U6), 5'-AAAACGAAATAAATCTTTGTAA AC; and 310 (tRNA<sup>Tyr</sup><sub>GVA</sub>-intron), 5'-AAGATTTCTGTAGTGATAA.

Oligonucleotides 200, 203, and 218 are largely composed of 2'-O-methyl RNA. **Expression of the U3 cDNA.** The synthesis of U3A from cDNA constructs was analyzed by expression of the ARS-CEN pU3-wt plasmid carrying an *ADE2* marker (11). This U3 intronless construct is under the control of the natural promoter and terminator regions. Expression was analyzed in the *GAL::snr17A* strain JH84 (24; J. Hughes, personal communication), from which the endogenous U3A was depleted by growth on glucose medium. Alternatively, the pU3 sub6-CBS1 plasmid, which carries the viable mutations U3sub6 and CBS1 (11, 47), was expressed in wild-type yeast strains. U3 synthesized from the cDNA construct was detected by hybridization with a probe specific for the sub6 mutation (47).

**In vitro processing reactions.** Synthetic U3-3' RNAs were obtained by in vitro transcription as described elsewhere (16), using a PCR product as template. The PCR product was generated from genomic DNA using a forward primer carrying a T7 promoter (T7U3DS; 5'-GCGAATCTAATACGACTCACTATAGGTACT TTCTTTTGAAGGAT) and reverse primers 252 (U3ADS) for a longer U3(-60/+177) transcript or 253 (U3DS) for a shorter U3(-60/+139) transcript. Whole-cell extracts were prepared from wild-type and *mt1-Δ* sister strains as described previously (16), and recombinant His<sub>6</sub>-Rnt1p was purified as described previously (5, 16).

In vitro processing of U3-3' RNAs in cell extracts or with recombinant His<sub>6</sub>-Rnt1p and mapping of the cleavage sites using primer extension were performed as described elsewhere (16). Prior to the reaction, gel-purified RNA substrates (2 nM) were denatured for 2 min at 85°C in Rnt1p buffer (50 mM Tris-HCl [pH 7.6], 200 mM KCl, 0.1 mg of wheat-germ tRNA/ml, 5 mM MgCl<sub>2</sub>) and cooled to 23°C. The cleavage reaction was performed at 23°C using 100 fmol of recombinant His<sub>6</sub>-Rnt1p or by incubation in the whole-cell extracts.

**RNAse A/T<sub>1</sub> mapping.** RNAse A/T<sub>1</sub> mapping was performed as described elsewhere (22). The <sup>32</sup>P-labeled antisense probe was transcribed in vitro with T7 polymerase using a PCR template as described above. The PCR product was generated from genomic DNA using forward primer T7antiU3, carrying a T7 promoter (5'-GCGAATCTAATACGACTCACTATAGGTTTTAAACAATT TAGAAAAGG), and reverse primer 3'antiU3 (5'-GGGCTCTATGGGTGGG TAC). The RNA transcript was gel purified and hybridized to 8 μg of total RNA

in 30 μl of piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES) buffer (40 mM PIPES) [pH 6.7], 400 mM NaCl, 1 mM EDTA and 50% formamide. Annealing was performed by heating at 95°C for 2 min followed by incubation at 48°C for several hours. Digestion in RNase buffer (10 mM Tris-HCl [pH 7.5], 300 mM NaCl, 1 mM EDTA) was performed with 5 to 15 U of RNase T<sub>2</sub>, 0.4 to 2.5 units of RNase T<sub>1</sub>, and 0.1 to 0.5 μg of RNase A (RNase T<sub>2</sub> from GibcoBRL; RNases T<sub>1</sub> and A from Boehringer) for 30 min at 25°C. Protected products were recovered by guanidium thiocyanate-phenol-chloroform extraction and separated on an 10% polyacrylamide gel.

**RNAse H treatment.** Deadenylation was performed essentially as described elsewhere (18). Samples of 30 μg of RNA were annealed with 750 ng of oligo(dT) at 65°C for 1 h and digested with 6 U of RNase H at 30°C for 1 h. The control samples were treated identically except that the oligo(dT) was omitted.

**Immunoprecipitation.** For immunoprecipitation of ProtA-Nop1p, ProtA-Nop58p, ProtA-Nop56p, Lhp1p-ProtA, and m<sup>3</sup>:<sup>2,7</sup>G-capped RNAs, yeast whole-cell extracts were prepared as described elsewhere (46) except that for immunoprecipitation of m<sup>3</sup>:<sup>2,7</sup>G-capped RNAs, cells were resuspended in buffer A (150 mM potassium acetate [KAc; pH 7.5], 20 mM Tris-Ac, 5 mM MgAc) with 1 mM dithiothreitol, 0.5% Triton X-100, and 5 mM phenylmethylsulfonyl fluoride. Immunoprecipitation of ProtA-Nop1p, ProtA-Nop58p, ProtA-Nop56p, and Lhp1p-ProtA with rabbit immunoglobulin G (IgG) agarose beads (Sigma) was performed as previously described (33) at 150 mM salt (KAc) concentration. For immunoprecipitation with m<sup>3</sup>:<sup>2,7</sup>G-cap-specific serum (R1131; kindly provided by R. Lührmann), 30 μl of suspension of protein G-Sepharose was washed with phosphate-buffered saline buffer and incubated on a rotating wheel with extract equivalent to 4 units of optical density at 600 nm of cells in 120 μl of buffer A for 2 h at 4°C. After the pellet was washed in buffer A, bound m<sup>3</sup>:<sup>2,7</sup>G-capped RNAs were eluted with 10 mM m<sup>7</sup>G(5')ppp(5')G (Pharmacia) in 30 μl of buffer A. The RNAs were extracted with GTC/phenol-chloroform and ethanol precipitated.

## RESULTS

**Yeast cells contain 3'-extended forms of U3.** Yeast U3 is encoded by two genes, *SNR17A*, encoding U3A, and *SNR17B*, encoding U3B (25). U3A is approximately 10-fold more abundant than U3B (25), and all analyses have been performed for U3A. On Northern hybridization, probe 200, to mature U3A, was observed to hybridize to two RNA species of slower gel mobility (U3-3'I and U3-3'II) in total yeast RNA preparations (Fig. 1A, lane 1) that were estimated to be approximately 10 and 20 nucleotides (nt), respectively, longer than the mature U3 (333 nt). A probe complementary to the sequence across the 3' end of the mature U3A (probe 251), which hybridizes specifically to 3'-extended species, also detected these RNAs as well as a longer species (U3-int 3') of approximately 470 nt. Both *SNR17A* and *SNR17B* contain introns that are excised by the pre-mRNA splicing machinery (38). The size and hybridization pattern of U3-int 3' indicates that it corresponds to a 3'-extended precursor that retains the intron (Fig. 1D and 6B).

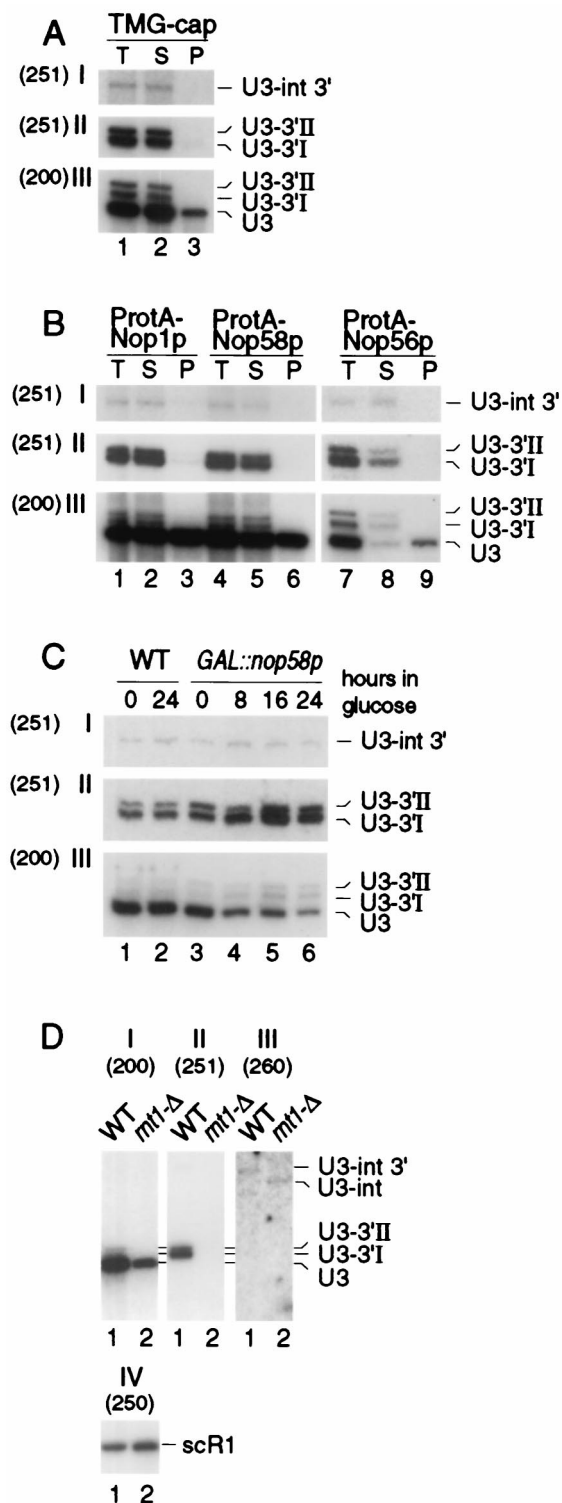


FIG. 1. Northern analysis of 3'-extended forms of U3 snoRNA. Probes (indicated in parentheses): 251, complementary to the region across the 3' end of the mature U3A; 200, complementary to mature U3; 260, complementary to the U3A intron; 250, complementary to the scR1 RNA. For panels A and B, input lysates were estimated to contain comparable amounts of U3 snoRNA, and equal fractions of the preparation were loaded for each lane; panels C and D, constant amounts of total RNA were loaded in each lane. (A) Immunoprecipitation with m<sup>3</sup>:<sup>2,7</sup>G cap-specific antibody (R1131) on lysates from the wild-type D150 strain. (B) Immunoprecipitation of lysates from strains expressing epitope-tagged fusion proteins ProtA-Nop1p, ProtA-Nop58p, and ProtA-Nop56p. (C) Stability of mature and 3'-extended U3 upon depletion of Nop58p. RNA was

It is not clear whether U3-int 3' has 3' ends identical to those of U3-3'I and U3-3'II. Synthesis of the U3-3'I and U3-3'II RNAs was not affected by the presence or absence of the intron in the pre-snoRNA, since identical species were observed in strains expressing U3 cDNA constructs (see Materials and Methods) (data not shown).

The mature U3 carries a 5' trimethyl guanosine (TMG) cap structure (25) and was precipitated with anti-TMG antibodies (Fig. 1A, lane 3) (generously provided by R. Lüthmann, University of Marburg). In contrast, the U3-3'I, U3-3'II, and U3-int 3' RNAs were not precipitated with anti-TMG and were recovered exclusively in the immune supernatant (Fig. 1A, lane 2). Mature yeast U3, like all box C+D snoRNAs, is associated with Nop1p, Nop56p, and Nop58p (30, 31, 44) and was coprecipitated with protein A-tagged fusion proteins (Fig. 1B, lanes 3, 6, and 9). No association of U3-3'I, U3-3'II, or U3-int 3' with these proteins was observed, and the RNAs were again recovered exclusively in the immune supernatants (Fig. 1B, lanes 2, 5, and 8).

Genetic depletion of Nop58p leads to the loss of all tested box C+D snoRNAs including U3 (30). The *GAL::nop58* strain was pregrown on permissive, galactose medium (0-h sample) and then transferred to glucose to repress synthesis of Nop58p (Fig. 1C). Mature U3 was codepleted with Nop58p, whereas the levels of the U3-3'I and U3-3'II RNAs were increased. The U3-int 3' species was unaffected.

We conclude that the U3 snoRNA is synthesized from 3' extended precursors that lack the TMG cap structure. The pre-U3 species are not associated with snoRNP proteins and, unlike the mature snoRNA, do not require Nop58p for stability. Indeed, the accumulation of U3-3'I and U3-3'II in strains depleted of Nop58p indicates that their normal maturation to U3 requires Nop58p binding.

**3' processing of U3 involves cleavage by Rnt1p.** Rnt1p cleaves 3'-extended precursors to the U1, U2, U4, and U5 snRNAs and processes polycistronic pre-snoRNAs. We therefore determined whether it is also involved in the 3' processing of pre-U3 species. In strains carrying a complete deletion of the *RNT1* gene, the level of mature U3 was reduced approximately threefold (Fig. 1D, I; see also Table 2). Strains carrying *mt1-Δ* lacked the U3-3'I and U3-3'II RNAs (Fig. 1D, II) and we observed a heterogeneous group of RNAs extending to approximately 600 nt (see Fig. 6A, lane 16, where more RNA is loaded). In addition, the intron-containing precursor was found to be 3' processed in the *mt1-Δ* strain, in contrast to the 3'-extended form seen in the wild type (Fig. 1D, III, lane 2; see also Fig. 6C, lanes 12 to 14). The reduced levels of U3 in the *mt1-Δ* strain were initially postulated to be due to impaired splicing (15). However, subsequent work indicated that splicing was not defective in the *mt1-Δ* strain (45) and, as shown in Fig. 1D, there is no overall accumulation of intron-containing forms of U3.

We conclude that 3' processing of U3 normally involves cleavage by Rnt1p. In the absence of cleavage, long 3'-extended forms are synthesized. The time required for these to be synthesized and then processed may allow assembly of the mature snoRNP proteins, and processing proceeds directly to the 3' end of the mature snoRNA. This processing is, however,

extracted from the *GAL::nop58* and wild-type (WT) strains following transfer from permissive, galactose medium to repressive, glucose medium for the times indicated. (D) Effects of *mt1-Δ* on 3'-extended U3. The level of scR1 RNA is shown as a control for loading. T, total cell lysate; S, immune supernatant; P, immunoprecipitate.



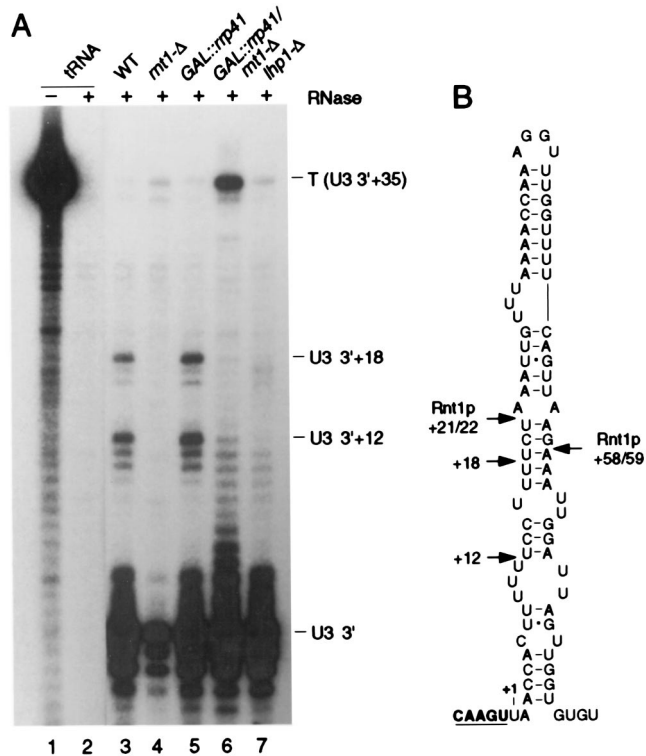


FIG. 3. Mapping of the 3'-extended forms of U3 by RNase protection. (A) RNA was extracted from wild-type (WT), *mt1-Δ*, and *lhp1-Δ* strains grown at 30°C and from *GAL::rnp41* and *GAL::rnp41/mt1-Δ* strains following transfer from permissive, RSG medium to repressive, glucose medium at 30°C for 24 and 48 h, respectively. Total *E. coli* tRNA was used as a control RNA. Positions of the Rnt1p-dependent protected species at +12 and +18 are indicated. (B) Schematic of the U3 3' flanking region showing the ends of the protected regions and the Rnt1p cleavage sites.

degraded by Rnt1p. The level of the mature U3 is reduced in strains lacking Rnt1p, indicating that this is normally the major synthesis pathway.

**The major 3'-extended forms of U3 do not extend to the Rnt1p cleavage sites.** High-resolution Northern hybridization showed that the U3-3'I band was too small to extend to the Rnt1p cleavage sites, and even the larger U3-3' II species appeared to be slightly smaller than expected. The 3' ends of these species were therefore determined by RNase protection. For this, the region of *SNR17A* from 295 to +36 was amplified by PCR using a primer that incorporated a T7 promoter (see Materials and Methods). In addition to the band corresponding to the mature 3' end of U3, two major protected fragments were detected in RNA from the wild-type strain (Fig. 3A, lane 3) but were absent from the *mt1-Δ* strain (Fig. 3A, lane 4). The sizes to these bands correspond to species that extend to U3+12 and U3+18, in good agreement with the gel mobilities of the U3-3'I and U3-3'II RNAs, respectively.

We conclude that following Rnt1p cleavage, the pre-U3 undergoes rapid trimming back to +12 and +18.

**The major 3'-extended forms of U3 are stabilized by Lhp1p.** It seemed very likely that some RNA binding factor was responsible for stabilizing the 3' ends of the U3-3'I and -II species. Inspection of the sequence showed that these RNAs possessed 3' poly(U) tracts (Fig. 3B). The 3' poly(U) tracts of RNAs transcribed by RNA polymerase III are bound by the La protein (42, 48), as are the 3' extended precursors to human U1 (34) and yeast (58) snRNAs. We therefore tested whether

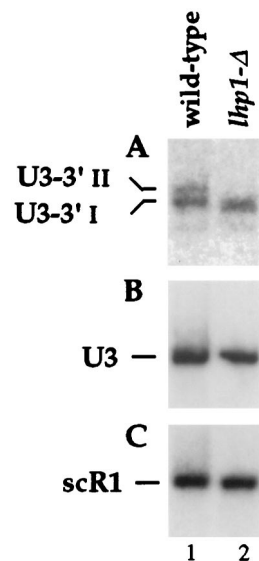


FIG. 4. 3'-extended forms of U3 are stabilized by Lhp1p. Lane 1, *LHP1* strain; lane 2, *lhp1-Δ* strain. Total RNA was analyzed by Northern hybridization with probe 251, specific for the 3'-extended U3 (A), probe 200, which hybridizes to the mature U3 (B), and probe 250, which hybridizes to the scR1 RNA (C).

the U3-3'I and -II RNAs were being stabilized by binding to Lhp1p, the yeast homologue of La (39, 59).

The *LHP1* gene is nonessential (59), and a gene disruption was performed by a one-step PCR approach (10) using the *K. lactis* *URA3* marker (see Materials and Methods). RNase protection analysis of RNA from the *lhp1-Δ* strain showed the loss of the major 3'-extended ends at +18 and +12 and the appearance of shorter, heterogeneous protected fragments corresponding to RNAs from U3+8 to U3+11 (Fig. 3A, lane 7). This result was confirmed by Northern hybridization (Fig. 4). The U3-3'II and U3-3'I species were absent from the *lhp1-Δ* strain (Fig. 4A), and a species slightly shorter than U3-3'I was detected. The level of mature U3 was unaffected in the *lhp1-Δ* strain (Figs. 3A and 4B), as were the levels of the truncated U3 degradation intermediates seen in wild-type cells (see Fig. 6; data not shown). These data suggested that both U3-3'I and U3-3'II were stabilized by binding Lhp1p.

To confirm this, a C-terminal fusion between Lhp1p and two copies of the Z domain of *S. aureus* protein A was constructed and integrated at the chromosomal *LHP1* locus by a one-step PCR approach (29) (see Materials and Methods). Western blotting confirmed that the fusion protein was expressed and could be efficiently immunoprecipitated with IgG agarose (data not shown). Immunoprecipitation was performed on two independently isolated Lhp1p-ProtA strains; data are presented for only one strain in Fig. 5. Processing of pre-tRNA<sup>Tyr</sup> appeared to be the same in the strain expressing only Lhp1p-ProtA and the wild type (Fig. 5D); however, some accumulation of the shorter 3'-extended pre-U3 species was visible (Fig. 5A), suggesting that the Lhp1p-ProtA fusion protein is under-expressed or otherwise not fully functional.

As expected, the tRNA<sup>Tyr</sup> primary transcript (Fig. 5D) and the U6 snRNA (Fig. 5E) were immunoprecipitated on IgG agarose from the strain expressing Lhp1p-ProtA (lane 6) but not from the wild type (lane 3). Both U3-3'I and U3-3'II were coprecipitated with Lhp1p-ProtA (Fig. 5A), as were U3-int 3' and a species of approximately 800 nt designated U3-3'III (Fig. 5B). The species shorter than U3-3'I seen in the Lhp1p-ProtA strain was not coprecipitated and remained in the immune

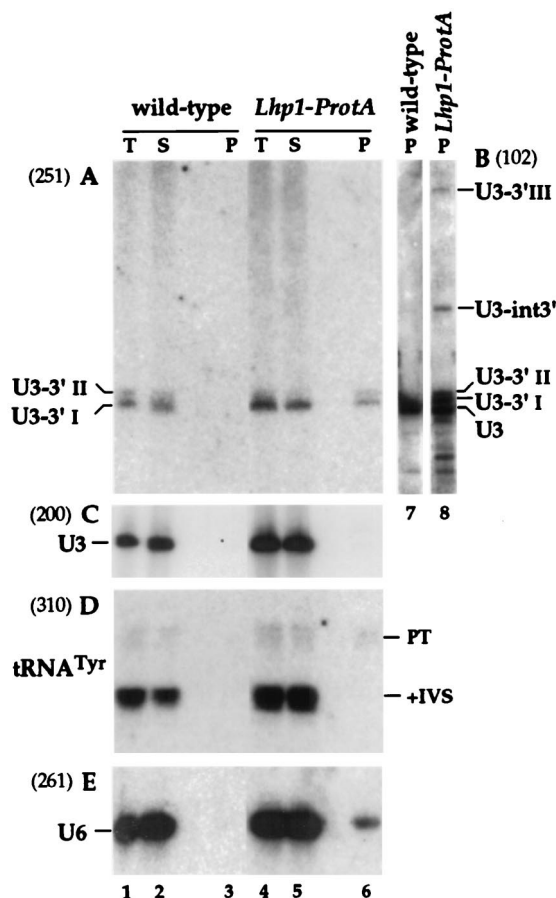


FIG. 5. 3'-extended forms of U3 are coprecipitated with Lhp1p-ProtA. Lysates from the *LHP1*<sup>+</sup> and *LHP1::ProtA* strains were immunoprecipitated using IgG agarose. RNA was recovered from the total cell lysate (T), immune supernatant (S), and immunoprecipitate (P) and analyzed by Northern hybridization. Probes are indicated in parentheses and described in Materials and Methods. On prolonged exposure, background precipitation of mature U3 is seen for both the wild-type and Lhp1-ProtA strains (lanes 7 and 8). In panel B, the total and supernatant lanes were heavily overexposed at the exposure needed to visualize the U3-int 3' and U3-3'III RNAs and were omitted. Approximately fourfold more cell equivalents are loaded for the bound material.

supernatant (Fig. 5A, lane 5). Mature U3 (Fig. 5B and C) and the 3' processed, intron-containing pre-tRNA<sup>Tyr</sup> (Fig. 5D) were recovered at the same low levels in the wild-type and Lhp1-ProtA precipitates. The pre-U3 and pre-tRNA species were more efficiently precipitated than U6, presumably because only the newly synthesized U6 is associated with Lhp1p (35, 39).

We conclude that Lhp1p binds and stabilizes the major 3'-extended forms of U3.

**The exosome participates in 3' processing of U3.** The levels of 3'-extended precursors to other snoRNAs and snRNAs are elevated in strains carrying mutations in the exosome complex (5, 55). To assess the effects of genetic depletion of exosome components on the 3'-extended forms of U3, Rrp41p and Rrp45p were depleted by transfer of *GAL::rrp41* and *GAL::rrp45* strains (6, 36) from permissive RSG medium (0-h samples) to repressive, glucose medium for the times indicated. A strain deleted for the gene encoding the Rrp6p component of the exosome (6) was also analyzed. In the strains lacking Rrp41p (Fig. 6A and C, lanes 5), Rrp45p (lanes 10), or Rrp6p (lanes 2), the levels of U3-3'I and U3-3'II were higher than in

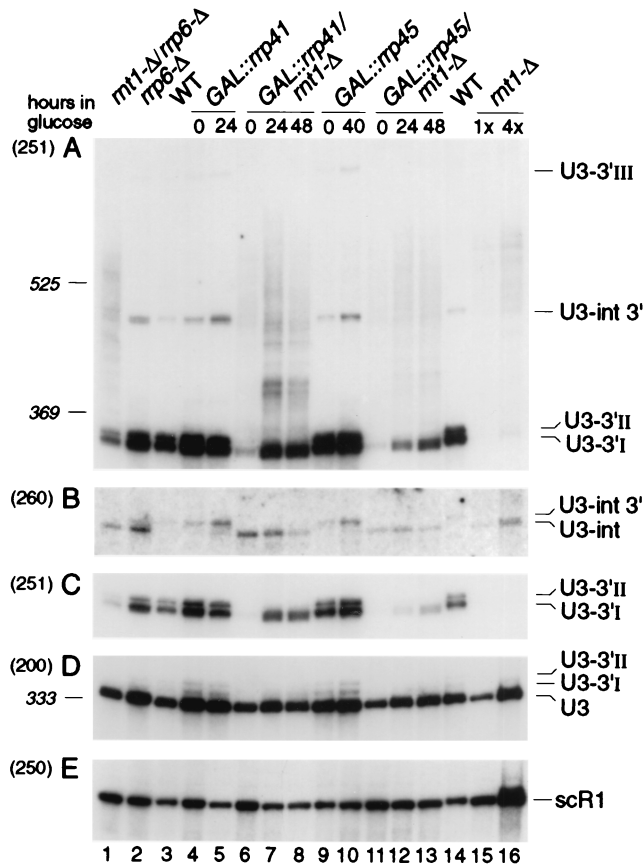


FIG. 6. Northern analysis of processing of U3 snoRNA in exosome mutants. RNA was extracted from strains carrying *GAL*-regulated constructs following transfer from permissive, RSG medium to repressive, glucose medium at 30°C for the times indicated, or from the wild-type (WT), *mt1-Δ*, *rrp6-Δ*, and *mt1-Δ/rrp6-Δ* strains grown on glucose medium at 30°C. RNA was separated on a 6% polyacrylamide gel and hybridized with oligonucleotide probes. The panels show successive hybridization of the same filter. Probes are indicated in parentheses on the left and described in Materials and Methods; the positions of RNA species detected are indicated on the right. Panel C presents a weaker exposure of the same gel as panel A. Panels B to E present only relevant regions of the Northern blots. The amount of total RNA loaded in lane 16 is fourfold greater than in lane 15 and other lanes. The positions of migration of scR1 (525 nt) and P (369 nt) RNAs determined by hybridization of the same filter are indicated as size markers. Mature U3 is 333 nt.

the isogenic wild-type control strains (lanes 3 and 14); these results are quantitated in Table 2. For the *GAL::rrp41* strain, this increase was confirmed by RNase protection (Fig. 3A, lane 4), which showed that the accumulated precursors were identical to U3-3'I and -II. Rrp41p is underexpressed in the *GAL::rrp41* strain in RSG medium and therefore shows some accumulation of the extended species in the 0-h sample (6, 36). In strains genetically depleted of other exosome components, Rrp4p, Rrp40p, Rrp46p, or Csl4p, increased levels of U3-3'I and -II were also observed (data not shown). In addition, an RNA species that comigrated with the U3-3'III RNA, seen on Lhp1p-ProtA precipitation (Fig. 5), was accumulated in the exosome mutants. On prolonged exposure, this species could also be detected at low levels in wild-type cells. Depletion of the exosome components did not lead to depletion of the mature U3. Indeed, as was previously observed for the U4 and U5 snRNAs, depletion of exosome components led to an increase in the mature U3 snoRNA of approximately twofold (Table 2).

In strains lacking exosome components, the 3' processed, intron-containing precursor is clearly detected. This is most visible for Rrp6p (Fig. 6B, lane 2) but was also seen for several other exosome mutants (Fig. 6B and data not shown). This species is not detected in the wild type, and we speculate that this processing intermediate is normally a dead-end product that is degraded by the exosome. 3' processing appears to be dependent on snoRNP protein binding, but assembly with the mature snoRNP proteins may be incompatible with assembly of a functional spliceosome. The exosome also degrades other stalled, intron-containing pre-mRNAs (C. Bousquet-Antonelli, C. Presutti, and D. Tollervey, submitted for publication).

The combination of the deletion of both *RNT1* and *RRP6* (Fig. 6, lane 1) partially restored synthesis of species with the same gel mobility as the U3-3'I and U3-3'II RNAs. Depletion of Rrp41p or Rrp45p from the strain lacking Rnt1p (Fig. 6A and C, lanes 7, 8, 12, and 13) led to the appearance of heterogeneous RNA species slightly smaller than U3-3'I, similar in size to the species seen in the *lhp1*- $\Delta$  strain (Fig. 4). Consistent with this, RNase protection analysis in the *GAL::rrp41/mt1*- $\Delta$  strain reveals a ladder of protected RNA fragments extending from mature U3 to position U3+12 (Fig. 3A, lane 6); due to the location of the hybridization probe, only the longer RNAs were detected by Northern hybridization (Fig. 6). A stronger ladder of RNA species extending up to the position of U3-3'III was observed by Northern hybridization (Fig. 6, lanes 7, 8, 12, and 13), which was reflected by the strong protection of the full-length antisense probe (Fig. 3A, lane 6). The combination of each of exosome mutations with *mt1*- $\Delta$  partially restored the mature U3 levels compared to the *mt1*- $\Delta$  single mutant strain (Table 2).

We conclude that the exosome complex of 3'→5' exonucleases participates in the 3' processing of U3. This processing pathway closely resembles that of the U1, U4, and U5 snRNAs (5, 14, 45, 55). In each case, synthesis of the mature RNA continues in strains depleted of single components of the exosome, indicating either that different components of the complex are partially functionally redundant or that other exonucleases can largely substitute for the exosome.

The level of the mature U3 is elevated in the exosome mutants, indicating competition between the synthetic pathway and degradation of the pre-U3. This was also seen for the U4 and U5 snRNAs (5). Consistent with this model, a truncated U3 species (U3\*\*) was observed in wild-type strains (Fig. 7, lanes 1 and 12) (24, 35). The U3\*\* species was 5' and 3' truncated, as shown by its failure to hybridize to probes directed against either the 3' end of U3 (Fig. 7B) or the 5' end of U3 (Fig. 7C). In contrast, the U3\* species that is accumulated in *rrp6*- $\Delta$ , *GAL::rrp41*, and *GAL::rrp45* strains was truncated only at the 5' end, indicating that U3 is normally 3' degraded by the exosome. The level of U3\* is further elevated

in exosome mutants that also lack Rnt1p, consistent with the model that degradation of pre-U3 is increased in *mt1*- $\Delta$  strains. The 5' degradation activity has not been further characterized but is likely due to the 5'→3' exonuclease Rat1p, which 5' processes other snoRNAs and degrades pre-rRNA spacer fragments (41).

In strains lacking Rnt1p, 3' extended forms of U1 and U2 snRNAs undergo a low level of polyadenylation (1, 45), and the precursors to several snRNAs and snoRNAs are polyadenylated in exosome mutants (5, 55). To determine whether this was also the case for the 3'-extended U3, RNA was treated in vitro with oligo(dT) and RNase H. Following this deadenylation treatment, the longer 3'-extended species detected in the *mt1*- $\Delta$ /*GAL::rrp41* strain became shorter and more discrete (data not shown), indicating that a low level of polyadenylation had indeed occurred.

## DISCUSSION

**How is U3 processed?** A model for 3' processing of the U3A snoRNA is presented in Fig. 8. We postulate that processing is normally initiated by cotranscriptional cleavage by Rnt1p across a stem structure at positions +21 and +58 with respect to the 3' end of U3. The released 3' fragment is degraded by the 5'→3' exonuclease Rat1p, as shown by its accumulation in the *rat1*-1 strain. The 3' extended pre-snoRNA is rapidly processed to +12 and +18, since the species extended to +21 is not readily detected in total RNA. The products of Rnt1p cleavage of pre-U4 and pre-U5 are elevated in strains deleted for components of the exosome (5), and we think it probable that the exosome complex also carries out the initial shortening of the pre-U3. We cannot, however, exclude the participation of other exonucleases, such as the Rex1-3p family that carry out the final trimming of several small RNA species (54). The pre-U3 is stabilized against further 3' degradation by binding of Lhp1p to the 3' poly(U) tracts at +19 and +13; whether Lhp1p binds to internal poly(U) tracts prior to the start of digestion, or binds to free 3' poly(U) tracts generated during digestion, cannot be determined at present. The larger U3-3'III species is bound by Lhp1p, suggesting that Lhp1p does bind to internal poly(U) sequences prior to processing, but the endpoints of this have not been mapped and we cannot exclude the possibility that it has a terminal poly(U) tract. It is likely that the poly(U) tracts at +19 and +13 can each bind Lhp1p, although binding may be mutually exclusive.

The box C+D snoRNAs, including U3, bind a set of common proteins, Nop1p, Nop56p, and Nop58p (13, 30–32, 40, 44, 53) that probably bind to the box D sequence close to the 3' end of the snoRNA and the 3'-terminal stem (13, 57). These proteins are not associated with the 3'-extended U3 species, and we propose that their binding displaces Lhp1p from the 3'

TABLE 2. PhosphorImager quantification of Northern hybridization data from Fig. 6<sup>a</sup>

Construct	Quantification						
	<i>GAL::rrp41</i>	<i>GAL::rrp41/mt1</i> - $\Delta$	<i>GAL::rrp45</i>	<i>GAL::rrp45/mt1</i> - $\Delta$	<i>mt1</i> - $\Delta$	<i>rrp6</i> - $\Delta$	<i>rrp6</i> - $\Delta$ / <i>mt1</i> - $\Delta$
U3	2.1	1.16	2.14	1	0.31	2.82	0.97
scR1	0.83	0.87	1.66	1.85	2.27	2.16	2.34
U3/scR1	2.53	1.33	1.29	0.54	0.11	1.3	0.41
U3-3'I + -II	2.3	0.57 <sup>b</sup>	2.89	0.27 <sup>b</sup>	0.044	2.78	0.31
U3-3'I + -II/scR1	2.77	0.65	1.74	0.15	0.019	1.29	0.13

<sup>a</sup> The U3-3'I and -II doublet was quantified as one species. The *GAL::rrp41* and *GAL::rrp41/mt1*- $\Delta$  data are from the 24-h time points; the *GAL::rrp45* and *GAL::rrp45/mt1*- $\Delta$  data are from the 40- and 48-h time points, respectively. Values are relative to the wild-type level, assigned a value of 1.

<sup>b</sup> Species shorter than U3-3'I that appears in the *GAL::rrp41/mt1*- $\Delta$  and *GAL::rrp45/mt1*- $\Delta$  strains.

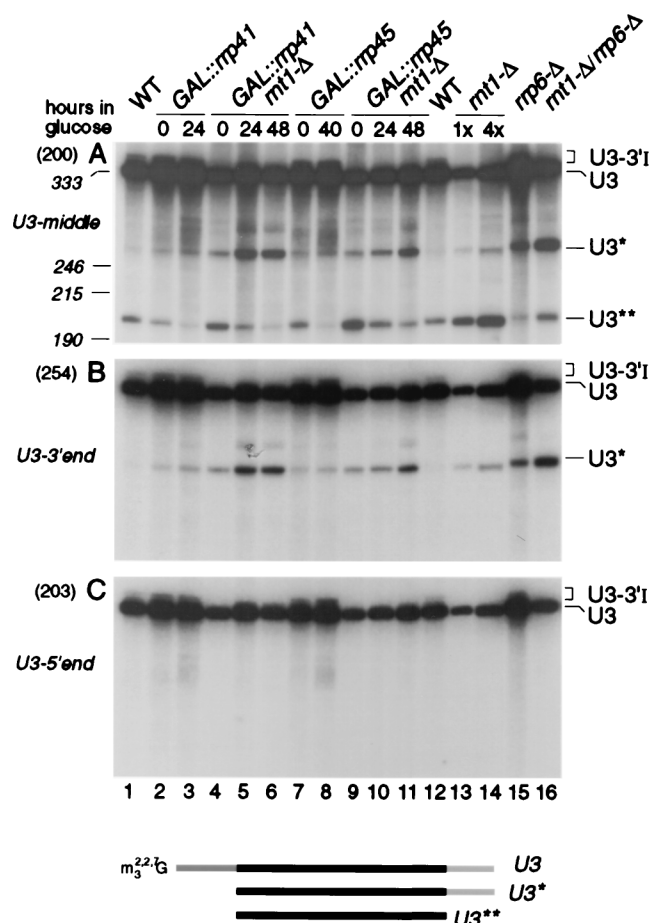


FIG. 7. Exosome components participate in the degradation of U3 snoRNA. For Northern analysis of U3 snoRNA in wild-type (WT) and *mt1-Δ* and exosome mutant strains. RNA was extracted as described for Fig. 2, separated on an 6% polyacrylamide gel, and hybridized with oligonucleotide probes. The panels show successive hybridization of the same filter. Probes are indicated in parentheses on the left and described in Materials and Methods; the positions of RNA species detected are indicated on the right. The amount of total RNA loaded in lane 14 is fourfold greater than in lane 13 and other lanes. The positions of migration of snRNA190 (190 nt), U5<sub>L</sub> (215 nt), and snR10 (246 nt) determined by hybridization of the same filter are indicated as size markers. Mature U3 is 333 nt. The locations of the oligonucleotide probes and the predicted structures of the degradation intermediates are shown schematically.

flanking sequence. Since the snoRNP proteins bind at the very 3' end of the snoRNA, this displacement may be steric. Removal of Lhp1p is envisaged to allow the exosome to resume processing, generating the mature snoRNA 3' end. This is followed by cap trimethylation; in vertebrates this snoRNA modification requires the conserved box C+D snoRNAs (50), probably acting via binding the mature snoRNP proteins. The yeast U3 genes are unusual in that they contain an intron that is excised by the normal pre-mRNA splicing machinery. In wild-type cells this is spliced from the 3'-extended pre-U3, since only the 3'-extended, intron-containing species is detected. The endpoints of the U3-int 3' species have not been determined, but these species are associated with Lhp1p, suggesting that they may have been largely processed to +18 and +12.

Deletion of Rnt1p strongly reduces synthesis of mature U3 (Table 2). Processing of the long 3'-extended pre-U3 species generated in the absence of Rnt1p cleavage involves the exosome, as shown by their increased levels in *mt1-Δ* strains lack-

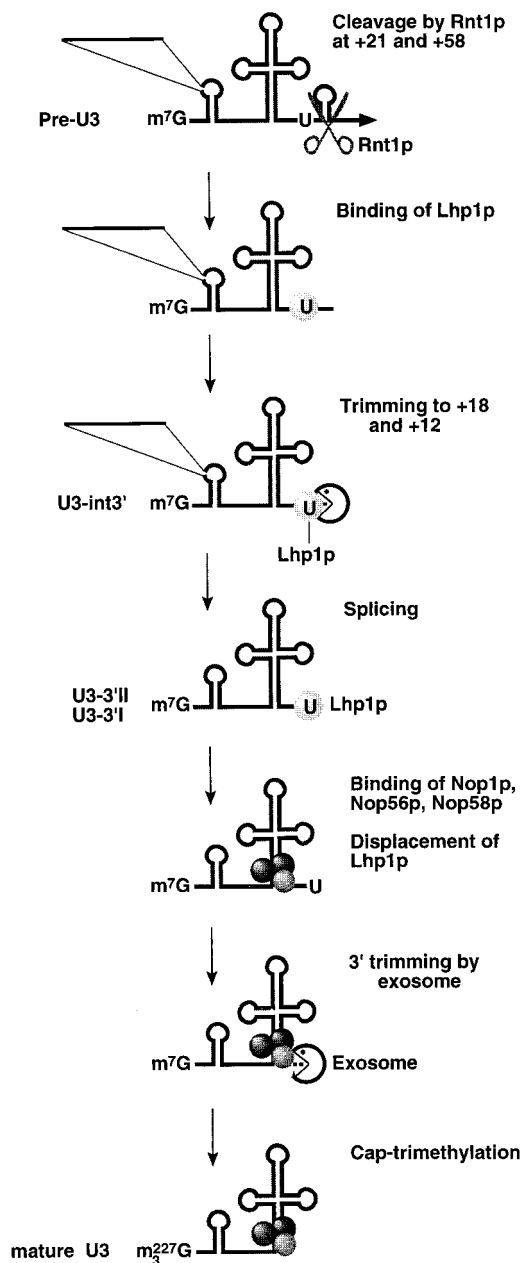


FIG. 8. Model for the 3' processing of the U3A snoRNA. The presence of the poly(U) tracts and stem-loop structure in the 3' flanking sequence and the intron are indicated. For simplicity, only one poly(U) tract is indicated. In reality, two tracts are present, at +19 and +13, each of which is likely to act as a binding site for Lhp1p. The activity that carries out the initial trimming to +18 and +12 has not been determined but is likely to be the exosome. The endpoints of the U3-int 3' species have not been determined, but the finding that these species are associated with Lhp1p suggests that they are processed to +18 and +12.

ing exosome components. We speculate that a processive exosome complex assembles on the long 3'-extended pre-U3, which is able to substantially displace bound Lhp1p and/or the snoRNP proteins and therefore degrades most of the pre-U3 population. Consistent with this model, depletion of exosome components from *mt1-Δ* strains restored mature U3 to the wild-type level (Table 2).

In the absence of Lhp1p, the U3 snoRNA was still 3' pro-





- of 3'→5' exonucleases. *Genes Dev.* **13**:2148–2158.
7. Amberg, D. C., A. L. Goldstein, and C. N. Cole. 1992. Isolation and characterization of RAT1: an essential gene of *Saccharomyces cerevisiae* required for the efficient nucleocytoplasmic trafficking of mRNA. *Genes Dev.* **6**:1173–1189.
  8. Anderson, J., L. Phan, R. Cuesta, B. A. Carlson, M. Pak, K. Asano, G. R. Bjork, M. Tamame, and A. G. Hinnebusch. 1998. The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes Dev.* **12**:3650–3662.
  9. Anderson, J. S. J., and R. P. Parker. 1998. The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SK12 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J.* **17**:1497–1506.
  10. Baudin, A., O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, and C. Culin. 1993. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **21**:3329–3330.
  11. Beltrame, M., and D. Tollervey. 1995. Base-pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *EMBO J.* **14**:4350–4356.
  12. Beltrame, M., and D. Tollervey. 1992. Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J.* **11**:1531–1542.
  13. Caffarelli, E., M. Losito, C. Giorgi, A. Fatica, and I. Bozzoni. 1998. In vivo identification of nuclear factors interacting with the conserved elements of box C/D small nucleolar RNAs. *Mol. Cell. Biol.* **18**:1023–1028.
  14. Chanfreau, G., S. A. Elela, M. Ares, Jr., and C. Guthrie. 1997. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev.* **11**:2741–2751.
  15. Chanfreau, G., P. Legrain, and A. Jacquier. 1998. Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J. Mol. Biol.* **284**:975–988.
  16. Chanfreau, G., G. Rotondo, P. Legrain, and A. Jacquier. 1998. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J.* **17**:3726–3737.
  17. Cooper, M., L. H. Johnston, and J. D. Beggs. 1995. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J.* **14**:2066–2075.
  18. Decker, C. J., and R. Parker. 1993. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev.* **7**:1632–1643.
  19. Fan, H., J. L. Goodier, J. R. Chamberlain, D. R. Engelke, and R. J. Maraia. 1998. 5' processing of tRNA precursors can be modulated by the human La antigen phosphoprotein. *Mol. Cell. Biol.* **18**:3201–3211.
  20. Gautier, T., T. Berges, D. Tollervey, and E. Hurt. 1997. The nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol. Cell. Biol.* **17**:7088–7098.
  21. Gietz, D., A. St. Jean, R. A. Woods, and R. H. Schiestl. 1992. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.* **20**:1425.
  22. Goodall, G. J., K. Wiebauer, and W. Filipowicz. 1990. Analysis of pre-mRNA processing in transfected plant protoplasts. *Methods Enzymol.* **181**:148–161.
  23. Heise, T., L. G. Guidotti, and F. V. Chisari. 1999. La autoantigen specifically recognizes a predicted stem-loop in hepatitis B virus RNA. *J. Virol.* **73**:5767–5776.
  24. Hughes, J. M. X., and M. J. Ares. 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J.* **10**:4231–4239.
  25. Hughes, J. M. X., D. A. M. Konings, and G. Cesareni. 1987. The yeast homologue of U3 snRNA. *EMBO J.* **6**:2145–2155.
  26. Jansen, R., D. Tollervey, and E. C. Hurt. 1993. A U3 snoRNP protein with homology to splicing factor PRP4 and Gβ domains is required for ribosomal RNA processing. *EMBO J.* **12**:2549–2558.
  27. Kiss, T., C. Marshallsay, and W. Filipowicz. 1991. Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and in vitro. *Cell* **65**:517–526.
  28. Kufel, J., B. Dichtl, and D. Tollervey. 1999. Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA* **5**:909–917.
  29. Lafontaine, D., and D. Tollervey. 1996. One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. *Nucleic Acids Res.* **24**:3469–3472.
  30. Lafontaine, D. L. J., and D. Tollervey. 1999. Nop58p is a common component of the box C+D snoRNPs that is required for stability of the snoRNAs. *RNA* **5**:455–467.
  31. Lafontaine, D. L. J., and D. Tollervey. 2000. Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol. Cell. Biol.* **20**:2650–2659.
  32. Lischwe, M. A., R. L. Ochs, R. Reddy, R. G. Cook, L. C. Yeoman, E. M. Tan, M. Reichlin, and H. Busch. 1985. Purification and partial characterization of a nucleolar scleroderma antigen (Mr=34,000; pI, 8.5) rich in NG, NG-dimethylarginine. *J. Biol. Chem.* **260**:14304–14310.
  33. Lygerou, Z., P. Mitchell, E. Petfalski, B. Séraphin, and D. Tollervey. 1994. The *POP1* gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. *Genes Dev.* **8**:1423–1433.
  34. Madore, S. J., E. D. Wieben, and T. Pederson. 1984. Eukaryotic small ribonucleoproteins. Anti-La human autoantibodies react with U1 RNA-protein complexes. *J. Biol. Chem.* **259**:1929–1933.
  35. Mayes, A. E., L. Verdone, P. Legrain, and J. D. Beggs. 1999. Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J.* **18**:4321–4331.
  36. Mitchell, P., E. Petfalski, A. Shevchenko, M. Mann, and D. Tollervey. 1997. The exosome; a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonuclease activities. *Cell* **91**:457–466.
  37. Mitchell, P., E. Petfalski, and D. Tollervey. 1996. The 3'-end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev.* **10**:502–513.
  38. Myslinski, E., V. Ségault, and C. Branlant. 1990. An intron in the genes for U3 small nucleolar RNAs of the yeast *Saccharomyces cerevisiae*. *Science* **247**:1213–1216.
  39. Pannone, B. K., D. Xue, and S. L. Wolin. 1998. A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J.* **17**:7442–7453.
  40. Parker, K. A., and J. A. Steitz. 1987. Structural analysis of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. *Mol. Cell. Biol.* **7**:2899–2913.
  41. Petfalski, E., T. Dandekar, Y. Henry, and D. Tollervey. 1998. Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol. Cell. Biol.* **18**:1181–1189.
  42. Rinke, J., and J. A. Steitz. 1982. Precursor molecules of both human 5S ribosomal RNA and transfer RNAs are bound by a cellular protein reactive with anti-La lupus antibodies. *Cell* **29**:149–159.
  43. Salgado-Garrido, J., E. Bragado-Nilsson, S. Kandels-Lewis, and B. Séraphin. 1999. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J.* **18**:3451–3462.
  44. Schimmang, T., D. Tollervey, H. Kern, R. Frank, and E. C. Hurt. 1989. A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J.* **8**:4015–4024.
  45. Seipelt, R. L., B. Zheng, A. Asuru, and B. C. Rymond. 1999. U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway. *Nucleic Acids Res.* **27**:587–595.
  46. Séraphin, B., and M. Roshash. 1989. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**:349–358.
  47. Sharma, K., and D. Tollervey. 1999. Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. *Mol. Cell. Biol.* **19**:6012–6019.
  48. Stefano, J. E. 1984. Purified lupus antigen La recognizes an oligouridylylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell* **36**:145–154.
  49. Suh, D., H. Busch, and R. Reddy. 1986. Isolation and characterization of a human U3 small nucleolar RNA gene. *Biochem. Biophys. Res. Commun.* **137**:1133–1140.
  50. Terns, M. P., C. Grimm, E. Lund, and J. E. Dahlberg. 1995. A common maturation pathway for small nucleolar RNAs. *EMBO J.* **14**:4860–4871.
  51. Tollervey, D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J.* **6**:4169–4175.
  52. Tollervey, D., and I. W. Mattaj. 1987. Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J.* **6**:469–476.
  53. Tyc, K., and J. A. Steitz. 1989. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J.* **8**:3113–3119.
  54. van Hoof, A., P. Lennertz, and R. Parker. 2000. Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. *EMBO J.* **19**:1357–1365.
  55. van Hoof, A., P. Lennertz, and R. Parker. 2000. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol. Cell. Biol.* **20**:441–452.
  56. Van Horn, D. J., C. J. Yoo, D. Xue, H. Shi, and S. L. Wolin. 1997. The La protein in *Schizosaccharomyces pombe*: a conserved yet dispensable phosphoprotein that functions in tRNA maturation. *RNA* **3**:1434–1443.
  57. Watkins, N. J., D. R. Newman, J. F. Kuhn, and E. S. Maxwell. 1998. In vitro assembly of the mouse U14 snoRNP core complex and identification of a 65-kDa box C/D-binding protein. *RNA* **4**:582–593.
  58. Xue, D., D. Rubinson, B. K. Pannone, C. J. Yoo, and S. L. Wolin. 2000. U snRNP assembly in yeast involves the La protein. *EMBO J.* **19**:1650–1660.
  59. Yoo, C. J., and S. L. Wolin. 1997. The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell* **89**:393–402.

## A Nuclear Surveillance Pathway for mRNAs with Defective Polyadenylation<sup>¶</sup>

Laura Milligan,<sup>†</sup> Claire Torchet,<sup>†‡</sup> Christine Allmang,<sup>§</sup> Tracey Shipman,  
and David Tollervey<sup>\*</sup>

Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom

Received 11 March 2005/Returned for modification 27 April 2005/Accepted 10 August 2005

**The *pap1-5* mutation in poly(A) polymerase causes rapid depletion of mRNAs at restrictive temperatures. Residual mRNAs are polyadenylated, indicating that Pap1-5p retains at least partial activity. In *pap1-5* strains lacking Rrp6p, a nucleus-specific component of the exosome complex of 3'-5' exonucleases, accumulation of poly(A)<sup>+</sup> mRNA was largely restored and growth was improved. The catalytically inactive mutant Rrp6-1p did not increase growth of the *pap1-5* strain and conferred much less mRNA stabilization than *rrp6Δ*. This may indicate that the major function of Rrp6p is in RNA surveillance. Inactivation of core exosome components, Rrp41p and Mtr3p, or the nuclear RNA helicase Mtr4p gave different phenotypes, with accumulation of deadenylated and 3'-truncated mRNAs. We speculate that slowed mRNA polyadenylation in the *pap1-5* strain is detected by a surveillance activity of Rrp6p, triggering rapid deadenylation and exosome-mediated degradation. In wild-type strains, assembly of the cleavage and polyadenylation complex might be suboptimal at cryptic polyadenylation sites, causing slowed polyadenylation.**

The exosome is a complex of 3'-5' exonucleases that is conserved in eukaryotes (31) and archaea (25). In yeast, nuclear and cytoplasmic forms of the exosome share 10 components. All of these proteins are essential for viability and have sequence homology to known 3'-5' exoribonucleases, and several have been shown to function as ribonucleases in vitro. Genetic depletion or mutation of any of these proteins results in very similar defects in RNA maturation and degradation (2), and for convenience they are often referred to as the "core" exosome components (reviewed in references 12, 32, and 48). In addition, the cytoplasmic complex is associated with the GTPase Ski7p (3, 49), while the nuclear complex is associated with an additional exonuclease, Rrp6p (2, 11), and a nucleic acid binding protein, Lrp1p/Rrp47p (30, 35).

Ski7p functions together with the core exosome in cytoplasmic mRNA turnover and RNA surveillance pathways (3, 49). In contrast, the functions of Rrp6p and Lrp1p/Rrp47p are distinct from those of the core components of the exosome during nuclear 3' processing of several small stable RNAs, including the 5.8S rRNA (1, 2, 9, 30, 35, 47). In these cases, Rrp6p specifically processes RNA intermediates that are generated by the activity of the core exosome.

In eukaryotic mRNAs, the 3' poly(A) tail plays key roles in translation, mRNA stability, and, at least in some cases, nu-

clear export. The poly(A) tail is added to the 3' ends of mRNAs by poly(A) polymerase, Pap1p in yeast (34), within a large processing complex in a reaction that is normally coupled to cotranscriptional mRNA cleavage and transcription termination (6, 53; reviewed in references 27 and 37). In some strains with defects in pre-mRNA cleavage, long 3'-extended transcripts that are rapidly degraded by the nuclear exosome are generated (42). In certain cases, subsequent polyadenylation that is uncoupled to pre-mRNA cleavage can apparently generate functional mRNAs from pre-mRNAs that have been 3' processed by the exosome. Rrp6p is not required for the initial processing of the 3'-extended transcripts (42). However, in strains defective in mRNA cleavage due to the *mal4-1* mutation (28, 29), Rrp6p plays a distinct role in pre-mRNA degradation following initial processing by the exosome, apparently antagonizing polyadenylation. However, recombinant Rrp6p was reported to show no preference for poly(A)<sup>+</sup> RNAs in vitro (11), so any direct role in deadenylation is unlikely to result from the intrinsic specificity of the Rrp6p exonuclease activity. A different role for RNA polyadenylation in stimulating nuclear RNA degradation by the exosome has been described recently (21, 22, 50, 52). This involves a distinct nuclear poly(A) polymerase, Trf4p (22, 39, 50, 52).

A previous analysis identified the *rrp6-1* point mutation, which alters a key residue in the catalytic region of Rrp6p, as a suppressor of the temperature-sensitive (TS) lethal mutation *pap1-1* (9, 11, 34, 36). This suggested that Rrp6p, and perhaps the nuclear exosome, plays a role in degrading mRNAs that have failed to undergo polyadenylation. Consistent with this model, Rrp6p was required to restrict mRNAs synthesized in *pap1-1* strains to a nuclear region that was proposed to lie close to the site of transcription (18).

To better define the role of the exosome in the degradation of mRNAs with defects in polyadenylation, we examined poly(A) tail length and mRNA degradation in strains carrying a collection of reported TS lethal alleles of *PAP1* (28). In each

\* Corresponding author. Mailing address: Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom. Phone: (44) 131 650 7092. Fax: (44) 131 650 7040. E-mail: d.tollervey@ed.ac.uk.

<sup>†</sup> These authors contributed equally to the work.

<sup>‡</sup> Present address: Unité de Génétique des Interactions Macromoléculaires, URA 2171-CNRS, Institut Pasteur, 25-28 rue du Docteur ROUX, F-75724 Paris Cedex 15, France.

<sup>§</sup> Present address: UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue Rene Descartes, F-67084 Strasbourg Cedex, France.

<sup>¶</sup> Supplemental material for this article may be found at <http://mcb.asm.org/>.

mutant, mRNA levels were rapidly reduced after transfer to nonpermissive temperature. However, in *pap1-5*, but not *pap1-2*, strains, the reduced level of residual mRNAs appeared to be substantially polyadenylated at the nonpermissive temperature. Further analyses led to the conclusion that the deficiency in mRNAs in the *pap1-5* strain is not due to the inability to synthesize poly(A) tails per se but to an RNA surveillance pathway that triggers nuclear deadenylation and exosome-mediated degradation of the newly synthesized pre-mRNAs.

#### MATERIALS AND METHODS

**Strains, media, and yeast genetics.** Strains were grown in YPD medium, containing 2% peptone, 1% yeast extract, and 2% glucose, or YPGal, containing 2% peptone, 1% yeast extract, and 2% galactose. Transformation was performed as described previously (17), except that 6% dimethyl sulfoxide was added prior to heat shock and the final pellet was resuspended in 0.15 M NaCl. For the strains of *Saccharomyces cerevisiae* used in this study, see Table S1 in the supplemental material. To make strain YCA42 (*pap1-5/rrp6Δ*), the *RRP6* open reading frame was replaced by *Kluyveromyces lactis URA3* in strain *pap1-5* by using primers 5'RRP6::URA (849) (see Table S2 in the supplemental material for the sequence) and 3'RRP6::URA (850). Transformants were selected for Ura<sup>+</sup> prototrophy and analyzed by Northern blotting for 5.8S rRNA processing defects. To make strains YCT56 (*pap1-5/GAL::rrp41*) and YCT59 (*pap1-5/rrp6Δ/GAL::rrp41*), the HIS3-GAL10-ProtA-RRP41 cassette was amplified by PCR from strain P118 with primers RRP41-1 (842) and RRP41-2 (843) and transformed into strains *pap1-5* and YCA42. Correct gene deletion was confirmed by analysis of the 5.8S rRNA processing defect. Strain YCT83 (*pap1-5/mur3-1*) was obtained by sporulation of diploids resulting from crossing *pap1-5* with YCT73. The KAN-GAL-3HA-MTR4 construct was generated from strain YCBA81 by one-step PCR (22a) in strain *pap1-5* with primers MTR4-F4 (991) and MTR4-R3 (992). Transformants were selected for kanamycin resistance and analyzed by Northern blotting for 5.8S rRNA processing defects. One transformant, YCT109 (*pap1-5/GAL::MTR4*), was selected. To make strains YCT68 (*ski7Δ*) and YCT71 (*pap1-5/ski7Δ*), the *KAN::ski7* cassette was amplified by PCR from strain Y01852 (EUROSCARF) with primers SKI7-1 (993) and SKI7-2 (994) and transformed into D270 and *pap1-5*, respectively. Correct integration was confirmed by PCR. To make strains YLM122 (*ccr4Δ*) and YLM124 (*pap1-5/ccr4Δ*), the *KAN::ccr4* cassette was amplified by PCR from strain Y00387 (EUROSCARF) with primers CCR4-1 (1101) and CCR4-2 (1102) and transformed into D270 and *pap1-5*, respectively; correct integration was confirmed by PCR. To make strains YLM127 (*pan2Δ*) and YLM129 (*pap1-5/pan2Δ*), the *KAN::pan2* cassette was amplified by PCR from strain Y04461 (EUROSCARF) with primers PAN2-1 (1104) and PAN2-2 (1105) and transformed into D270 and *pap1-5*, respectively; correct integration was confirmed by PCR. Strain YLM121 (*pap1-5/rrp6-1*) was obtained by sporulation of diploids resulting from crossing *pap1-5* with a strain carrying the *rrp6-1* allele (11).

**RNA extraction and analysis.** RNA extractions were performed as described previously (41). Seven micrograms of total RNA was analyzed for each sample. Small RNAs were separated on a 6% acrylamide gel containing 8.3 M urea and transferred to a Hybond N<sup>+</sup> membrane by electrotransfer. High-molecular-weight RNAs were analyzed on 1.2% agarose gels and transferred by capillary elution. For the oligonucleotides used, see Table S2 in the supplemental material.

For poly(A) tail length analysis of mRNA, 7 μg of total RNA was digested with 10 μg of RNase A and 250 units of RNase T1 in 10 mM Tris, pH 8, 300 mM NaCl. The digestion was stopped by adding 10 mM EDTA, 0.25% sodium dodecyl sulfate, 25 μg/ml proteinase K, and 0.5 mg/ml glycogen. Samples were precipitated, and then 3' end labeling of the poly(A) tails was carried out overnight at 4°C with 10 μCi [<sup>32</sup>P]pCp (cytidine-3',5'-bisphosphate) and 40 units of T4 RNA ligase in 50 mM Tris-HCl, pH 7.9, 10 mM MgCl<sub>2</sub>, 3.3 mM dithiothreitol, 10 μg/ml bovine serum albumin, and 10% dimethyl sulfoxide. Samples were then phenol-chloroform extracted and precipitated, and electrophoretic separation was analyzed on a 12% acrylamide-8 M urea gel. For total poly(A) tail analysis, 7 μg of total RNA was extensively hydrolyzed with RNase A and RNase T1. Following ethanol precipitation to remove free nucleotides, residual poly(A) tracts were 3' end labeled with [<sup>32</sup>P]pCp and RNA ligase and resolved on a 12% polyacrylamide gel containing 8 M urea. Similar results were obtained with two independent experiments.

**RNase H treatment.** Deadenylation was performed essentially as described previously (33). Samples (20 μg) of RNA were annealed with 400 ng oligo(dT) at 68°C for 10 min and digested with 1.5 U RNase H at 30°C for 1 h.

#### RESULTS

**In *pap1-5* mutant strains, poly(A)<sup>+</sup> mRNAs are degraded by the nuclear exosome.** The *pap1-2* and *pap1-5* mutations each result in tight TS lethality at 37°C, but they have little effect on growth at 23°C (28). Both alleles have multiple mutations, and it is not clear which of them give rise to the TS phenotype. The lengths of the poly(A) tracts present in total RNA were assessed by 3' labeling with [<sup>32</sup>P]pCp following digestion with RNase A and RNase T1 (see Materials and Methods). During growth at 23°C, little difference was seen in the average poly(A) tail length between the *pap1-2* and *pap1-5* strains, both of which showed maximal poly(A) tail lengths around 20 nucleotides shorter than that of the wild type (Fig. 1A). Following transfer to 37°C for 30 min, the *pap1-2* strain retained only low levels of poly(A) (Fig. 1A, lane 5). In contrast, *pap1-5* strains retained substantial poly(A) tracts (Fig. 1A, lane 6). The signal strength was reduced at 37°C relative to 23°C, consistent with overall loss of mRNAs, but the maximal tail length distribution was not greatly shortened (see Fig. S1 in the supplemental material for PhosphorImager quantification of these data). Similar poly(A) length distribution was seen even after 90 min at 37°C (data not shown). This result would be consistent with synthesis of a reduced level of poly(A)<sup>+</sup> RNAs, which continue to be normally deadenylated in the cytoplasm. We conclude that the *pap1-5* strain, but not the *pap1-2* strain, retains significant polyadenylation activity at 37°C. This suggested that lethality in the *pap1-5* strain did not result simply from an inability to generate poly(A) tails.

The *pap1-2* and *pap1-5* alleles were combined with deletion of the *RRP6* gene. RNAs from the single- and double-mutant strains were analyzed by Northern blotting 2 h after transfer to 37°C (Fig. 1B). Increased *CYH2* and *ACT1* mRNA levels were seen for the *pap1-5 rrp6Δ* strain relative to the *pap1-5* single mutant. In contrast, no increases were seen in the *pap1-2 rrp6Δ* strain relative to *pap1-2* alone. Quantification is shown for *CYH2* transcript and is standardized relative to scR1 RNA, a component of the cytoplasmic signal recognition particle.

These observations suggested a role for the nuclear exosome in the degradation of newly synthesized poly(A)<sup>+</sup> pre-mRNAs in the *pap1-5* mutant strain. To confirm the nuclear localization of this degradation, the *pap1-5* allele was also combined with a deletion of the gene encoding the cytoplasmic exosome component Ski7p, which is required for 3' degradation of cytoplasmic mRNAs (3, 5, 45, 49). No clear mRNA stabilization was conferred by the absence of Ski7p, and no truncated RNA species were observed (Fig. 1C).

The loss of mRNA from the *pap1-5* strain was assessed during a time course following transfer to 37°C. Several mRNAs tested (*ACT1*, *CYH2*, *RPL25*, and *MFA2* and *RPL30*, *RPS26a*, and *CYCI*) (Fig. 2A and B, lanes 8 to 14, and data not shown) were all progressively depleted at 37°C in the *pap1-5* strain, indicating that accumulation of new mRNA was inhibited. However, even at late time points, residual mRNAs were present in the *pap1-5* strain, indicating a reduced level of

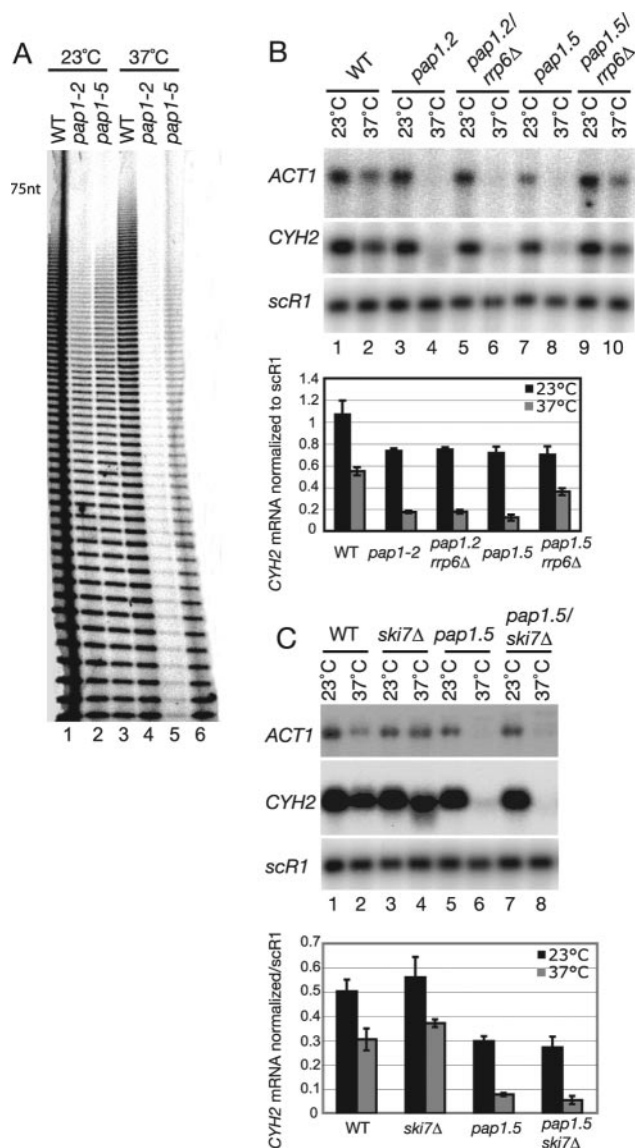


FIG. 1. The *pap1-5*, but not *pap1-2*, mutation allows poly(A) synthesis at the nonpermissive temperature and is suppressed by loss of Rrp6p. (A) Poly(A) tail length analysis of mRNAs from strains carrying the *pap1-2* and *pap1-5* mutations. Poly(A) tracts present in 7  $\mu$ g of total RNA were labeled and analyzed on a 12% acrylamide-8 M urea gel. The size marker was tRNA from end-labeled total RNA. WT, wild type. (B and C) Northern blot analyses. The strains indicated were grown on glucose medium at 23°C (23°C lanes) and then shifted to 37°C for 2 h (37°C lanes). For each lane, 7  $\mu$ g of total RNA was separated on a 1.2% agarose-formaldehyde gel and analyzed by Northern hybridization using the probes indicated on the left. The graphs show mean values  $\pm$  standard deviations for the *CYH2* transcript, obtained from PhosphorImager quantification of three independent experiments and normalized to the *scR1* loading control.

ongoing mRNA synthesis. Quantification is shown for the *CYH2* transcript, standardized to *scR1* RNA.

This conclusion was greatly strengthened by analysis of the *pap1-5 rrp6Δ* strain (Fig. 2A and B, lanes 15 to 21). For all mRNAs tested, synthesis in the *pap1-5* strain at 37°C was substantially increased by the absence of Rrp6p. The *rrp6Δ*

mutation alone did not strongly affect these mRNA species (data not shown).

The *pap1-5* mutation was also combined with *GAL::RRP41* to allow depletion of the core exosome component Rrp41p. In *pap1-5* strains depleted of Rrp41p, a substantially different phenotype was observed (Fig. 2A and B, lanes 22 to 29). For most mRNAs tested, truncated RNA species were observed that migrated at positions below the size range of mRNAs in the wild type. Such truncated RNAs are not seen in strains lacking only Rrp41p (data not shown; see references 8 and 42), showing that they are a consequence of some defect in mRNA synthesis in the *pap1-5* strain. An exception was the *ACT1* (actin) mRNA, for which truncated RNAs were not detected by Northern hybridization of the full-length mRNA but were observed following truncation by RNase H cleavage (data not shown). A 3'-extended RNA species was visible in strains lacking Rrp41p. This RNA was also detected with a probe to the *ACT1* 3' flanking sequence (data not shown). An increased level of this RNA species was previously observed in the *pap1-1* strain (26).

In the *pap1-5 GAL::RRP41* strain, very rapid loss of many mRNAs was seen after transfer to 37°C. This effect was not dependent on the *pap1-5* mutation and was also seen in *PAP1*<sup>+</sup> strains depleted of core exosome components (data not shown) or the exosome cofactor Mtr4p (see Fig. 4). In some preparations, a substantial decline in mRNA levels was also seen in wild-type strains. The mechanism underlying these precipitous reductions in cytoplasmic mRNA levels is still unclear. The mRNAs detected at later time points are presumably synthesized de novo at the nonpermissive temperature, confirming the continued synthesis of poly(A)<sup>+</sup> RNAs.

The *pap1-5* strains were also tested for decay of heat shock mRNAs, for which a pseudo-pulse-chase analysis can be performed by induction at 42°C, followed by incubation at 37°C (Fig. 2C). *SSA3* showed robust induction in the *pap1-5* strain (Fig. 2C, lanes 7 to 12). The absence of Rrp6p from the *pap1-5* strain increased the expression of *SSA3* (Fig. 2C, lanes 13 to 18). Truncated forms of *SSA3* were also evident in the *pap1-5* strain depleted of Rrp41p (Fig. 2C, lanes 19 to 24).

These results indicated that the exosome components Rrp6p and Rrp41p play distinct roles during mRNA degradation in the *pap1-5* strain. To determine whether Rrp6p acts prior to Rrp41p and the core exosome, we constructed a *pap1-5* strain that lacked Rrp6p and could be depleted of Rrp41p (Fig. 2A and B, lanes 30 to 37, and Fig. 2C, lanes 25 to 30). The absence of Rrp6p suppressed accumulation of the truncated RNA species, which were seen in the *pap1-5* strain depleted of Rrp41p alone, for most mRNAs tested, *CYH2* and *SSA3* (Fig. 2A to C) and *CYC1* and *RPL30* (data not shown). In the experiment shown in Fig. 2, the absence of Rrp6p plus Rrp41p from *pap1-5* had little effect on the *RPL25* mRNAs relative to depletion of Rrp41p alone (Fig. 2B). However, loss of the truncated *RPL25* species was evident in other experiments; the reason for this variability is still unclear.

It is notable that the reduction in mRNA levels in *pap1-5* mutant strains, and the degree of restoration in *rrp6Δ* mutants, showed substantial variation for different mRNA species. Heterogeneity has been seen in the nuclear degradation of unspliced pre-mRNAs (8). As in the cytoplasm, mRNA degradation in the nucleus apparently proceeds at species-specific rates

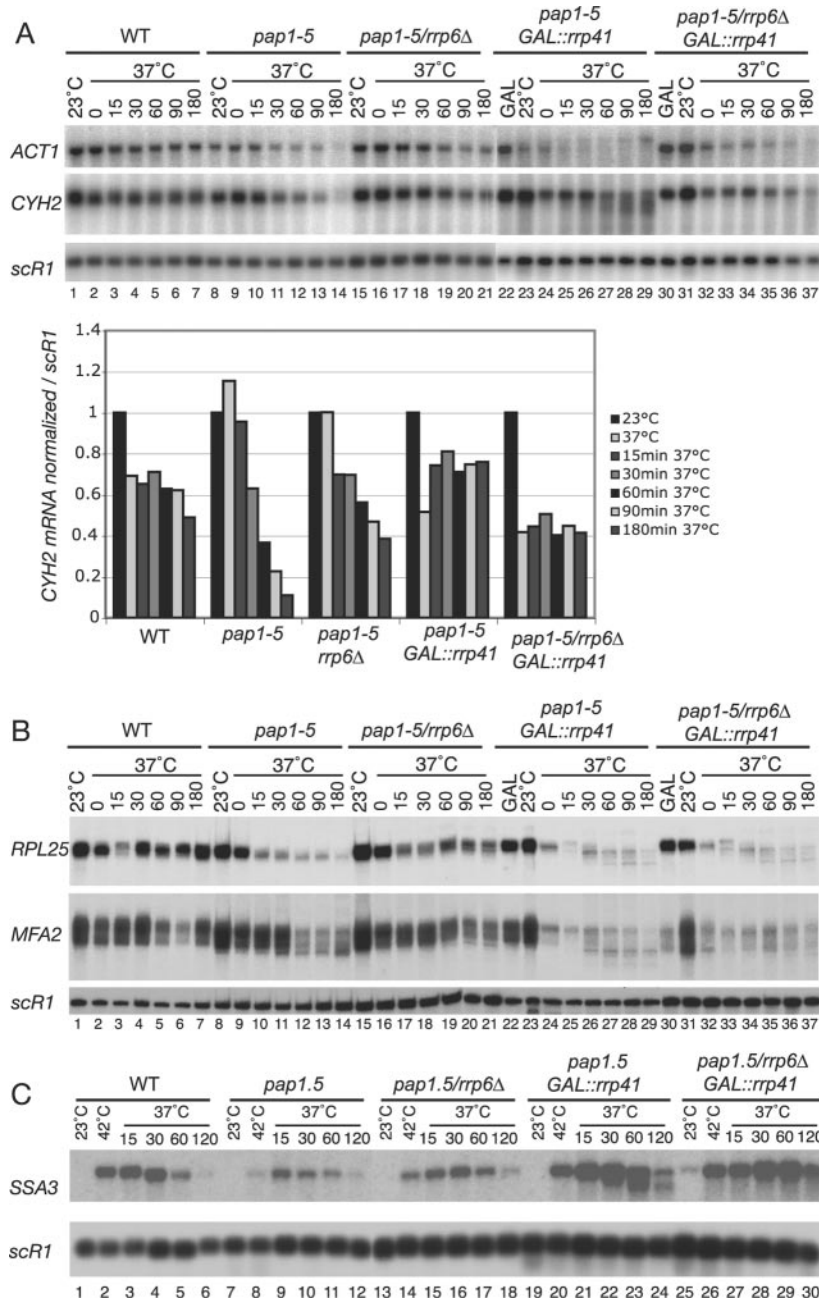


FIG. 2. Rrp6p acts prior to Rrp41p in the same degradation pathway. (A and B) Total RNA was extracted from the wild-type (WT), *pap1-5*, and *pap1-5/rrp6Δ* strains grown on glucose medium at 23°C and after shift to 37°C for the times indicated. Strains *pap1-5/GAL::rrp41* and *pap1-5/rrp6Δ/GAL::rrp41* were pregrown in galactose medium at 23°C (GAL lanes), transferred to glucose medium at 23°C for 20 h (23°C lanes), and then shifted to 37°C for the times indicated. Northern blot analysis was performed on 7 μg of total RNA separated on a 6% acrylamide-8.3 M urea gel (A) or a 1.2% agarose-formaldehyde gel (B). The graph shows levels of the *CYH2* transcripts obtained by PhosphorImager quantification of the data presented in panel A normalized to the *scR1* loading control. Values obtained at 23°C were arbitrarily set as 1. (C) Northern blot of heat shock-inducible mRNA. The wild-type, *pap1-5*, and *pap1-5/rrp6Δ* strains were pregrown on glucose medium at 23°C. The *pap1-5/GAL::rrp41* and *pap1-5/rrp6Δ/GAL::rrp41* strains were pregrown in galactose medium at 23°C and transferred to glucose medium at 23°C for 20 h (23°C lanes). All strains were then shifted to 42°C for 15 min (42°C lanes), followed by transfer to 37°C for the times indicated.

and shows variations in response to mutations in the degradation machinery. This presumably reflects differences in RNP structure.

**mRNAs present in the *pap1-5* strain at nonpermissive temperature are polyadenylated.** To assess the polyadenylation states of mRNAs present in the *pap1-5* strains at 37°C, dead-

enylation was performed in vitro using RNase H and oligo(dT) (Fig. 3A; quantification is shown for the *RPL25* and *RPL30* transcripts in Fig. 3B). Deadenylation of the wild-type samples resulted in increased in-gel mobility and more coherent RNA distribution for the *RPL25*, *RPL30*, and *MFA2* mRNAs, as expected. This was also the case for RNA in the *pap1-5/rrp6Δ*

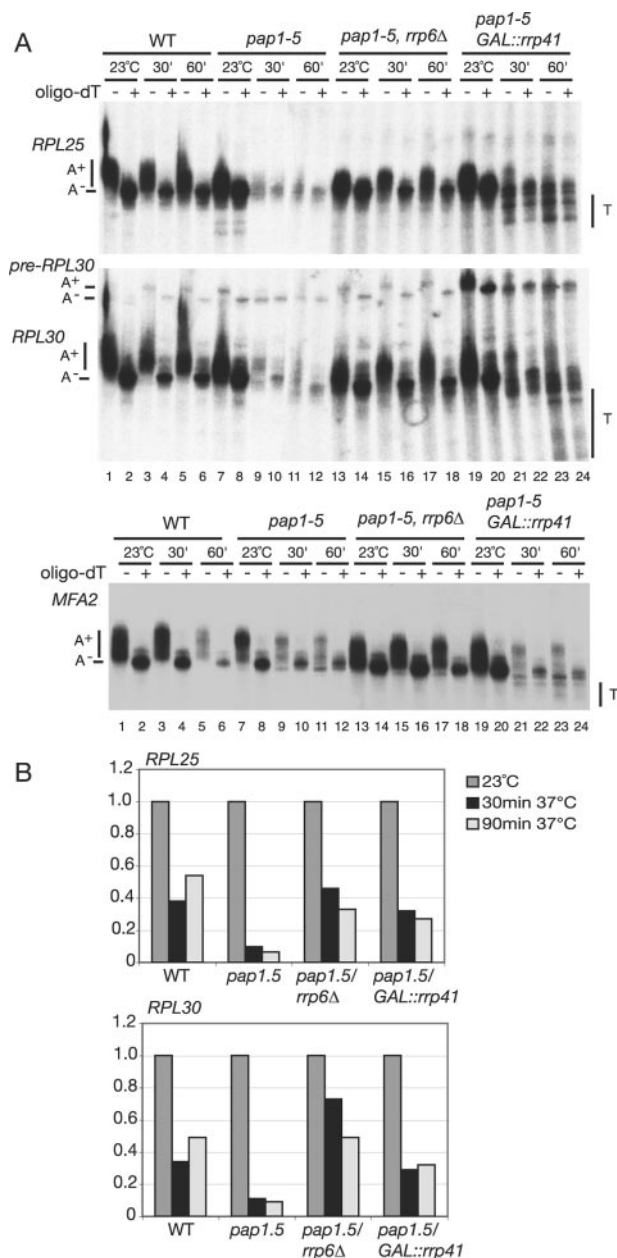


FIG. 3. Specific mRNAs in the *pap1-5* strain are polyadenylated. (A) Total RNA was extracted from the wild-type (WT), *pap1-5*, and *pap1-5/rrp6Δ* strains and grown in glucose medium for 30 min after transfer to 37°C. The *pap1-5/GAL::rrp41* strain was pregrown in galactose medium at 23°C, transferred to glucose medium at 23°C for 20 h, and then shifted to 37°C for 30 min. Samples were treated with RNase H plus oligo(dT) (+ lanes) and compared with untreated samples (– lanes). Samples were separated on a 6% acrylamide-8.3 M urea gel, transferred to nylon, and hybridized with *RPL25*, *RPL30*, and *MFA2* probes. T, deadenylated and truncated species. (B) PhosphorImager quantification of data from panel A. Deadenylated *RPL25* and *RPL30* mRNA (oligo-dT + lanes) was quantified using a PhosphorImager and standardized to *scR1* RNA.

strain at nonpermissive temperature, confirming that polyadenylated mRNAs continue to be synthesized. In the *pap1-5* single mutant, residual *RPL30* and *MFA2* were still polyadenylated at the nonpermissive temperature, but this was less

clear for *RPL25*. The *pap1-5* strain depleted of Rrp41p showed little accumulation of poly(A)<sup>+</sup> RNA relative to the *pap1-5* single mutant but accumulated deadenylated and truncated species (Fig. 3A).

The size heterogeneity of the mRNA populations complicates quantification of their abundance. We therefore compared the signals obtained for the deadenylated RNAs in Fig. 3A. *RPL25* and *RPL30* mRNA levels were quantified after deadenylation and standardized relative to *scR1* RNA. Quantification (Fig. 3B) showed that even in the wild type, total levels of *RPL25* and *RPL30* mRNAs were reduced following transfer to 37°C. This reduction was substantially greater in the *pap1-5* single mutants but was largely suppressed in the absence of Rrp6p or following depletion of Rrp41p.

We conclude that in the *pap1-5* strain these mRNAs largely undergo Rrp6p-dependent deadenylation followed by Rrp41p-dependent degradation. This suggests that the drastic reductions in mRNA levels seen in the *pap1-5* strain following transfer to 37°C are not primarily due to an inability to synthesize poly(A)<sup>+</sup> mRNAs. Rather, the newly synthesized mRNAs are rapidly identified by an RNA surveillance mechanism that requires Rrp6p.

**Degradation following deadenylation requires Mtr3p and Mtr4p.** To confirm that the phenotypes seen on depletion of Rrp41p were due to defects in the function of the nuclear exosome, the *pap1-5* allele was combined with a TS lethal mutation in the core exosome component Mtr3p and with the *GAL::MTR4* allele, which allows genetic depletion of Mtr4p/Dob1p (15), a putative RNA helicase and cofactor for the nuclear exosome (Fig. 4). In the *pap1-5 mtr3-1* double-mutant strain, loss of the polyadenylated mRNA after transfer to 37°C was accompanied by accumulation of deadenylated and truncated mRNAs (Fig. 4, lanes 22 to 28). Quantification is shown for the *CYH2* mRNA and is standardized relative to *scR1* RNA.

In the Mtr4p-depleted strains, the *RPL30* and *RPL25* mRNAs were very rapidly lost following transfer to 37°C. In the *pap1-5* strain depleted of Mtr4p, the appearance of de novo-synthesized mRNAs that were deadenylated and truncated was seen at later time points (Fig. 4, lanes 37 to 44). This phenotype closely resembles that seen in the *pap1-5* strain depleted of Rrp41p (Fig. 2).

We conclude that in *pap1-5* strains newly synthesized poly(A)<sup>+</sup> mRNAs are rapidly deadenylated, followed by 3'→5' degradation by the nuclear exosome, acting together with its cofactor Mtr4p.

**Deadenylation does not require the catalytic activity of Rrp6p.** In vitro analyses have shown that the residue altered in the *rrp6-1* allele is critically required for catalysis, and the mutant protein is therefore unlikely to show exonuclease activity in vivo (11, 36). To determine whether Rrp6p is directly responsible for mRNA deadenylation in the *pap1-5* strain, a *pap1-5 rrp6-1* strain was constructed. The levels of the *RPL30* and *RPS26a* mRNAs were mildly elevated in the *pap1-5 rrp6-1* strain relative to *pap1-5* alone, but *rrp6-1* had much less effect than *rrp6Δ* (Fig. 5A and B) and other mRNAs. The exonuclease activity of Rrp6p may participate in deadenylation but is apparently not required for degradation to occur.

In growth tests in liquid culture, the absence of Rrp6p partially suppressed the growth defect of strains carrying *pap1-5* at

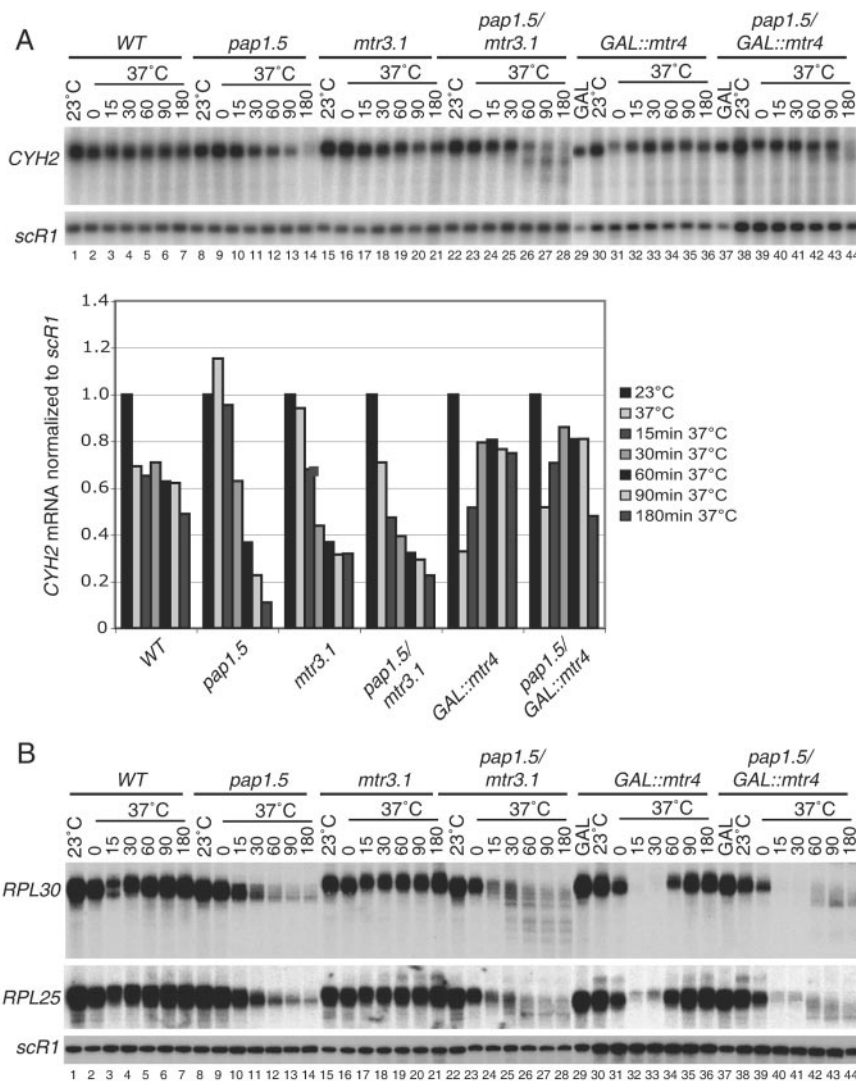


FIG. 4. 3' degradation requires Mtr3p and Mtr4p. (A and B) Total RNA was extracted from the wild-type (WT), *pap1-5*, *mtr3-1*, and *pap1-5/mtr3-1* strains grown on glucose medium at 23°C and after shift to 37°C for the times indicated. Strains *GAL::mtr4* and *pap1-5/GAL::mtr4* were pregrown in galactose medium at 23°C (GAL lanes), transferred to glucose medium at 23°C for 20 h (23°C lanes), and then shifted to 37°C for the times indicated. Northern blot analysis was performed on 7  $\mu$ g of total RNA separated on a 6% acrylamide-8.3 M urea gel (A) or a 1.2% agarose gel (B). The graph shows levels of the *CYH2* transcripts obtained by PhosphorImager quantification of the data presented in panel A, normalized to the *scR1* loading control. Values obtained at 23°C were arbitrarily set as 1.

either 34°C or 37°C (shown for 37°C in Fig. 6). The *rrp6 $\Delta$*  single-mutant strain is impaired in growth, but despite this, the *rrp6 $\Delta$  pap1-5* strain clearly grew better than the *pap1-5* single mutant. In contrast, growth of the *rrp6-1 pap1-5* strain was indistinguishable from that of the strain with *pap1-5* alone, consistent with the low level of suppression seen in Northern analyses. These results suggest that Rrp6p has an RNA surveillance function that is distinct from its exonuclease activities and is required to identify mRNAs synthesized in the *pap1-5* strain as being defective and to target them for degradation.

Two deadenylase complexes characterized in yeast are the Ccr4p-Caf1p-Not complex, which is probably the major cytoplasmic deadenylase, and Pan2p-Pan3p, which has been implicated in nuclear poly(A) length control (7, 10, 14, 43, 44). To assess their participation in deadenylation, *ccr4 $\Delta$*  and *pan2 $\Delta$*

mutations were each combined with *pap1-5*. Neither mutation conferred significant mRNA stabilization in the *pap1-5* strain (see Fig. S2 and S3 in the supplemental material). Consistent with this, neither double mutant showed increased growth in liquid culture relative to *pap1-5* alone (data not shown). The combination of *pan2 $\Delta$*  with *rrp6 $\Delta$*  and *pap1-5* in a triple-mutant strain failed to increase mRNA synthesis or growth relative to the *rrp6 $\Delta$  pap1-5* double mutant (data not shown; see Fig. S2 in the supplemental material).

## DISCUSSION

**Nuclear mRNAs are initially deadenylated and then degraded by the exosome.** We had anticipated that mRNAs synthesized at the nonpermissive temperature in strains with de-

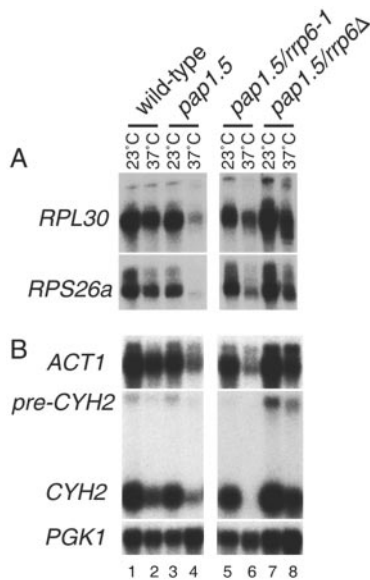


FIG. 5. mRNA levels in *pap1-5* strains lacking the exonuclease activity of Rrp6p. RNA was extracted from the strains indicated growing at 23°C and 1 h after transfer to 37°C. (A) RNA separated on 8% polyacrylamide-urea gel. (B) RNA separated on 1.2% agarose gels.

fects in poly(A) polymerase would predominantly lack poly(A) tails. However, the analysis of total poly(A) RNA and individual mRNAs indicates that the TS lethal *pap1-5* strain retains substantial polyadenylation activity in vivo at the restrictive temperature. The loss of mRNAs from *pap1-5* strains is apparently due to the targeting of newly synthesized, polyadenylated mRNAs for nuclear degradation. The mRNAs are initially deadenylated by an activity that requires the nucleus-specific exosome component Rrp6p and then 3' degraded by the core exosome, acting together with the nuclear RNA helicase Mtr4p (Fig. 7). In contrast, the *pap1-2* allele retained little poly(A)<sup>+</sup> RNA at the nonpermissive temperature, and mRNA synthesis was not clearly restored by loss of Rrp6p. While we cannot exclude the possibility that poly(A) tails detected in the *pap1* strains at nonpermissive temperature are synthesized by the poly(A) polymerase activity of Trf4p (22, 39, 50, 52), the allele specificity makes this less likely.

Deadenylation of poly(A)<sup>+</sup> mRNAs synthesized in the *pap1-5* strain at 37°C required the nuclear exosome component Rrp6p. This contains an exonuclease domain that is related to *Escherichia coli* RNase D and shows 3'-exonuclease activity in vitro (11, 36), suggesting that it might act directly as the deadenylase. However, previous analyses reported that recombinant Rrp6p shows no preference for poly(A)<sup>+</sup> RNAs in vitro (11), so any specific role in deadenylation was unlikely to result from its intrinsic activity. The *rrp6-1* point mutation is strongly predicted to inactivate the in vivo exonuclease activity of Rrp6p and phenocopies *rrp6Δ* for defects in stable RNA synthesis (9, 11, 36). However, the presence of the *rrp6-1* mutation conferred little suppression of mRNA synthesis in *pap1-5* strains and did not lead to detectable accumulation of degradation intermediates. Moreover, unlike *rrp6Δ*, the *rrp6-1* mutation conferred no growth suppression in *pap1-5* strains. This indicates that while Rrp6p is required for surveillance of the

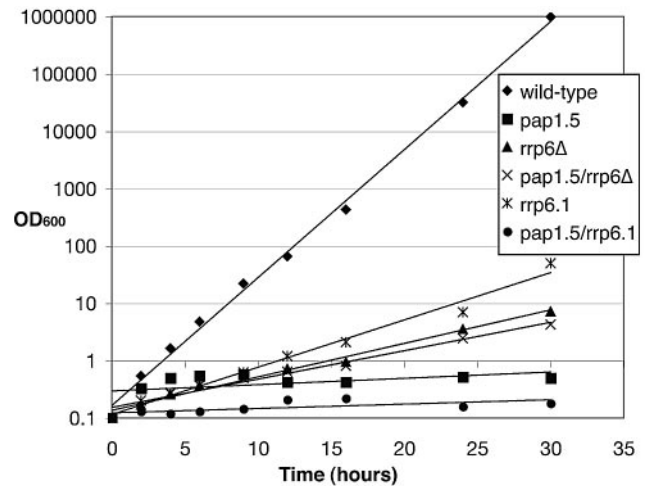


FIG. 6. Growth curves obtained following transfer to 37°C. Cells were pregrown in rich YPD medium at 23°C and transferred to 37°C at time zero. The cells were maintained in exponential growth by addition of prewarmed medium. OD<sub>600</sub>, optical density at 600 nm.

mRNAs synthesized in the *pap1-5* strain, its exonuclease activity is not required for their degradation. Rrp6p is comprised of two distinct domains, with an amino-terminal exonuclease domain and a C-terminal HRDC (helicase and RNase D C-terminal) domain. The HRDC domain has been proposed to play a regulatory role in Rrp6p function (36) and is likely to have nucleic acid binding activity (20). It is therefore possible that the HRDC domain specifically functions in surveillance in the *pap1-5* strain. Since Rrp6p is apparently not required for mRNA deadenylation, we tested two other characterized yeast deadenylases, Ccr4p and Pan2p (7, 10, 44). However, the decline in the levels of most mRNAs tested following transfer to 37°C was indistinguishable in *pap1-5* strains and *pap1-5 ccr4Δ* or *pap1-5 pan2Δ* double mutants. Moreover, mRNA levels in the *pap1-5 rrp6Δ pan2Δ* triple mutant were not different from those in the *pap1-5 rrp6Δ* double-mutant strain. Consistent with this observation, *pan2Δ* also conferred no growth increase in the *pap1-5* strain.

This indicates that Ccr4p and Pan2p are not individually responsible for nuclear deadenylation in the *pap1-5* background. It may be that once an mRNA has been targeted for degradation in an Rrp6p-dependent process, multiple proteins can participate in the deadenylation. During 3' maturation of the yeast 5.8S rRNA, several 3' exonucleases participate in the final trimming (9, 31, 46), and this is also the case for many RNA-processing and degradation steps in bacteria (24).

The features that make nuclear pre-mRNAs targets for degradation in *pap1-5* strains have not yet been established. Candidates for features that might be recognized include defects in the structure of the cleavage and polyadenylation machinery, the presence of shortened poly(A) tails, and a reduced rate of polyadenylation. We favor the last possibility and speculate that assembly of the cleavage and polyadenylation machinery at suboptimal, and therefore potentially inappropriate, sites may be correlated with a lower rate of poly(A) addition and/or reduced processivity of the reaction. The low sequence complexity of polyadenylation sites suggests that many potential



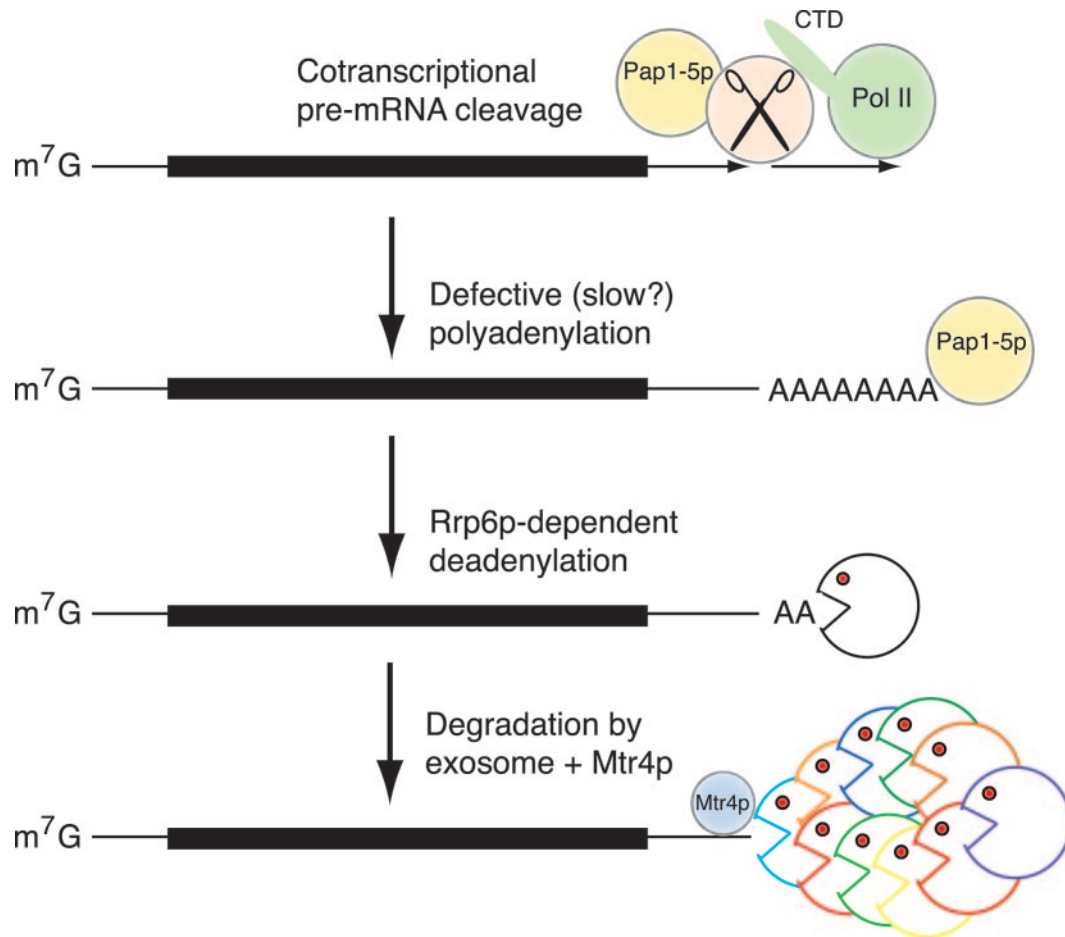


FIG. 7. Model for the degradation of pre-mRNAs in *pap1-5* strains. Several pre-mRNA 3' cleavage and polyadenylation factors bind to the C-terminal domain (CTD) of RNA polymerase II prior to recognition of the target site on the nascent RNA transcript (reviewed in references 19 and 38). In the strains expressing the partially defective Pap1-5p, cleavage and polyadenylation still occur, but pre-mRNA surveillance is triggered. We speculate that this is a consequence of slowed polyadenylation. Deadenylation of the pre-mRNA requires the nucleus-specific exonuclease Rrp6p. Subsequent degradation requires the nuclear exosome complex indicated by the symbols on the right, and the putative RNA helicase Mtr4p.

cryptic sites exist. Glutathione *S*-transferase-tagged Rrp6p has been reported to coprecipitate with Pap1p from cell lysates (11), indicating that they can physically interact. It is conceivable that prolonged association of Pap1p with the pre-mRNA, due to slowed polyadenylation, might be sufficient to recruit Rrp6p and the exosome.

In multicellular organisms, regulated and alternative poly(A) site choice has been reported and can have important developmental consequences (4, 13, 16, 23, 40). In such cases, the nuclear RNA surveillance pathway we report here may be important in determining the relative levels of the mRNAs synthesized.

#### ACKNOWLEDGMENTS

We thank P. J. Preker, T. Wiederkehr, and W. Keller for generously providing the *pap1-2* and *pap1-5* strains and communicating unpublished results; J. S. Butler for kindly providing the *rrp6-1* strain; and the laboratory of A. Jacquier, in which some of the experiments were carried out.

C.T. was the recipient of a fellowship from FEBS. This work was supported by the Wellcome Trust and EU grant QL2-CT-2001-01554.

#### REFERENCES

- Allmann, C., J. Kufel, G. Chanfreau, P. Mitchell, E. Petfalski, and D. Tollervey. 1999. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**:5399–5410.
- Allmann, C., E. Petfalski, A. Podtelejnikov, M. Mann, D. Tollervey, and P. Mitchell. 1999. The yeast exosome and human PM-Scl are related complexes of 3'→5' exonucleases. *Genes Dev.* **13**:2148–2158.
- Araki, Y., S. Takahashi, T. Kobayashi, H. Kajihio, S. Hoshino, and T. Katada. 2001. Ski7p G protein interacts with the exosome and the Ski complex for 3'→5' mRNA decay in yeast. *EMBO J.* **20**:4684–4693.
- Audibert, A., and M. Simonelig. 1998. Autoregulation at the level of mRNA 3' end formation of the suppressor of forked gene of *Drosophila melanogaster* is conserved in *Drosophila virilis*. *Proc. Natl. Acad. Sci. USA* **95**:14302–14307.
- Benard, L., K. Carroll, R. C. P. Valle, D. C. Masison, and R. B. Wickner. 1999. The ski7 antiviral protein is an EF1- $\alpha$  homolog that blocks expression of non-poly(A) mRNA in *Saccharomyces cerevisiae*. *J. Virol.* **73**:2893–2900.
- Birse, C. E., L. Minville-Sebastia, B. A. Lee, W. Keller, and N. J. Proudfoot. 1998. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280**:298–301.
- Boeck, R., S. J. Tarun, M. Rieger, J. A. Deardorff, S. Müller-Auer, and A. B. Sachs. 1996. The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *J. Biol. Chem.* **271**:432–438.
- Bousquet-Antonelli, C., C. Presutti, and D. Tollervey. 2000. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**:765–775.
- Briggs, M. W., K. T. Burkard, and J. S. Butler. 1998. Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J. Biol. Chem.* **273**:13255–13263.

10. Brown, C. E., S. Z. Tarun, Jr., R. Boeck, and A. B. Sachs. 1996. PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**:5744–5753.
11. Burkard, K. T., and J. S. Butler. 2000. A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with poly(A) polymerase and the hnRNA protein Npl3p. *Mol. Cell. Biol.* **20**:604–616.
12. Butler, S. 2002. The ying and yang of the exosome. *Trends Biochem. Sci.* **12**:90–96.
13. Castelo-Branco, P., A. Furger, M. Wollerton, C. Smith, A. Moreira, and N. Proudfoot. 2004. Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol. Cell. Biol.* **24**:4174–4183.
14. Dageron, M. C., F. Mauxion, and B. Seraphin. 2001. The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. *Nucleic Acids Res.* **29**:2448–2455.
15. de la Cruz, J., D. Kressler, D. Tollervey, and P. Linder. 1998. Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in *Saccharomyces cerevisiae*. *EMBO J.* **17**:1128–1140.
16. Edwalds-Gilbert, G., K. L. Veraldi, and C. Milcarek. 1997. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* **25**:2547–2561.
17. Gietz, R. D., R. H. Schiestl, A. R. Willems, and R. A. Woods. 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**:355–360.
18. Hilleren, P., T. McCarthy, M. Rosbash, R. Parker, and T. H. Jensen. 2001. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* **413**:538–542.
19. Hirose, Y., and J. L. Manley. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev.* **14**:1415–1429.
20. Jancsak, P., P. L. Garcia, F. Hamburger, Y. Makuta, K. Shiraiishi, Y. Imai, H. Ikeda, and T. A. Bickle. 2003. Characterization and mutational analysis of the RecQ core of the bloom syndrome protein. *J. Mol. Biol.* **330**:29–42.
21. Kadaba, S., A. Krueger, T. Trice, A. M. Krecic, A. G. Hinnebusch, and J. Anderson. 2004. Nuclear surveillance and degradation of hypomodified initiator tRNA<sup>Met</sup> in *S. cerevisiae*. *Genes Dev.* **18**:1227–1240.
22. LaCava, J., J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier, and D. Tollervey. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**:713–724.
- 22a. LaFontaine, D., and D. Tollervey. 1996. One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. *Nucleic Acids Res.* **24**:3469–3472.
23. Legendre, M., and D. Gautheret. 2003. Sequence determinants in human polyadenylation site selection. *BMC Genomics* **4**:7.
24. Li, Z., S. Pandit, and M. P. Deutscher. 1998. 3' Exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **95**:2856–2861.
25. Lorentzen, E., P. Walter, S. Fribourg, E. Evgueniev-Hackenberg, G. Klug, and E. Conti. 2005. The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat. Struct. Mol. Biol.* **12**:575–581.
26. Mandart, E., and R. Parker. 1995. Effects of mutations in the *Saccharomyces cerevisiae* RNA14, RNA15, and PAP1 genes on polyadenylation in vivo. *Mol. Cell. Biol.* **15**:6979–6986.
27. Minvielle-Sebastia, L., and W. Keller. 1999. mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.* **11**:352–357.
28. Minvielle-Sebastia, L., P. J. Preker, and W. Keller. 1994. RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. *Science* **266**:1702–1705.
29. Minvielle-Sebastia, L., B. Winsor, N. Bonneaud, and F. Lacroute. 1991. Mutations in the yeast RNA14 and RNA15 genes result in an abnormal mRNA decay rate; sequence analysis reveals an RNA-binding domain in the RNA15 protein. *Mol. Cell. Biol.* **11**:3075–3087.
30. Mitchell, P., E. Petfalski, R. Houalla, A. Podtelejnikov, M. Mann, and D. Tollervey. 2003. Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. *Mol. Cell. Biol.* **23**:6982–6992.
31. Mitchell, P., E. Petfalski, A. Shevchenko, M. Mann, and D. Tollervey. 1997. The exosome; a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonuclease activities. *Cell* **91**:457–466.
32. Mitchell, P., and D. Tollervey. 2000. Musing on the structural organization of the exosome complex. *Nat. Struct. Biol.* **7**:843–846.
33. Muhlrads, D., and R. Parker. 1992. Mutations affecting stability and deadenylation of the yeast MFA2 transcript. *Genes Dev.* **6**:2100–2111.
34. Patel, D., and J. S. Butler. 1992. Conditional defect in mRNA 3' end processing caused by a mutation in the gene for poly(A) polymerase. *Mol. Cell. Biol.* **12**:3297–3304.
35. Peng, W. T., M. D. Robinson, S. Mnaimneh, N. J. Krogan, G. Cagney, Q. Morris, A. P. Davierwala, J. Grigull, X. Yang, W. Zhang, N. Mitsakakis, O. W. Ryan, N. Datta, V. Jovic, C. Pal, V. Canadian, D. Richards, B. Beattie, L. F. Wu, S. J. Altschuler, S. Roweis, B. J. Frey, A. Emili, J. F. Greenblatt, and T. R. Hughes. 2003. A panoramic view of yeast noncoding RNA processing. *Cell* **113**:919–933.
36. Phillips, S., and J. S. Butler. 2003. Contribution of domain structure to the RNA 3' end processing and degradation functions of the nuclear exosome subunit Rrp6p. *RNA* **9**:1098–1107.
37. Proudfoot, N. 2000. Connecting transcription to messenger RNA processing. *Trends Biochem. Sci.* **25**:290–293.
38. Proudfoot, N. J., A. Furger, and M. J. Dye. 2002. Integrating mRNA processing with transcription. *Cell* **108**:501–512.
39. Saitoh, S., A. Chabes, W. H. McDonald, L. Thelander, J. R. Yates, and P. Russell. 2002. Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. *Cell* **109**:563–573.
40. Simpson, G. G., P. P. Dijkwel, V. Quesada, I. Henderson, and C. Dean. 2003. FY is an RNA 3' end-processing factor that interacts with FCA to control the Arabidopsis floral transition. *Cell* **113**:777–787.
41. Tollervey, D., and I. W. Mattaj. 1987. Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J.* **6**:469–476.
42. Torchet, C., C. Bousquet-Antonelli, L. Milligan, E. Thompson, J. Kufel, and D. Tollervey. 2002. Processing of 3' extended read-through transcripts by the exosome can generate functional mRNAs. *Mol. Cell* **9**:1285–1296.
43. Tucker, M., R. R. Staples, M. A. Valencia-Sanchez, D. Muhlrads, and R. Parker. 2002. Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. *EMBO J.* **21**:1427–1436.
44. Tucker, M., M. A. Valencia-Sanchez, R. R. Staples, J. Chen, C. L. Denis, and R. Parker. 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* **104**:377–386.
45. van Hoof, A., P. A. Frischmeyer, H. C. Dietz, and R. Parker. 2002. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**:2262–2264.
46. van Hoof, A., P. Lennertz, and R. Parker. 2000. Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. *EMBO J.* **19**:1357–1365.
47. van Hoof, A., P. Lennertz, and R. Parker. 2000. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol. Cell. Biol.* **20**:441–452.
48. van Hoof, A., and R. Parker. 1999. The exosome: a proteasome for RNA? *Cell* **99**:347–350.
49. van Hoof, A., R. R. Staples, R. E. Baker, and R. Parker. 2000. Function of the ski4p (Csl4p) and Ski7p proteins in 3'-to-5' degradation of mRNA. *Mol. Cell. Biol.* **20**:8230–8243.
50. Vanacova, S., J. Wolf, G. Martin, D. Blank, S. Dettwiler, A. Friedlein, H. Langen, G. Keith, and W. Keller. 2005. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol.* **3**:e189.
51. Reference deleted.
52. Wyers, F., M. Rougemaille, G. Badis, J.-C. Rouselle, M.-E. Dufour, J. Boulay, B. Régnault, F. Devaux, A. Namane, B. Séraphin, D. Libri, and A. Jacquier. 2005. Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**:725–737.
53. Yonaha, M., and N. J. Proudfoot. 2000. Transcriptional termination and coupled polyadenylation in vitro. *EMBO J.* **19**:3770–3777.



# Accurate Processing of a Eukaryotic Precursor Ribosomal RNA by Ribonuclease MRP in Vitro

Zoi Lygerou, Christine Allmang, David Tollervey, Bertrand Séraphin\*

Very few of the enzymes required for eukaryotic precursor ribosomal RNA (pre-rRNA) processing have been identified. Ribonuclease (RNase) MRP was characterized as a nuclease that cleaves mitochondrial replication primers, but it is predominantly nucleolar. Previous genetic evidence revealed that this ribonucleoprotein is required, directly or indirectly, for cleavage of the yeast pre-rRNA in vivo at site A<sub>3</sub>. Here, an in vitro processing system that accurately reproduces this cleavage is described. Biochemical purification and the use of extracts depleted of the MRP RNA demonstrate that endonucleolytic cleavage of the pre-rRNA is directly mediated by RNase MRP. This establishes a role for RNase MRP in the nucleolus.

Three of the four eukaryotic ribosomal RNAs are produced by processing a long precursor RNA (Fig. 1A). Genetic analysis in the yeast *Saccharomyces cerevisiae* provides a means to dissect this processing pathway and identify the factors and steps involved (1). Nevertheless, study of the biochemical mechanisms underlying pre-rRNA processing would be facilitated by the development of tractable in vitro systems. The ribonucleoprotein RNase MRP was identified as an endonuclease that cleaves mitochondrial replication primers in vitro (2). However, its predominantly nucleolar localization (3) and the reported existence of another enzyme able to cleave

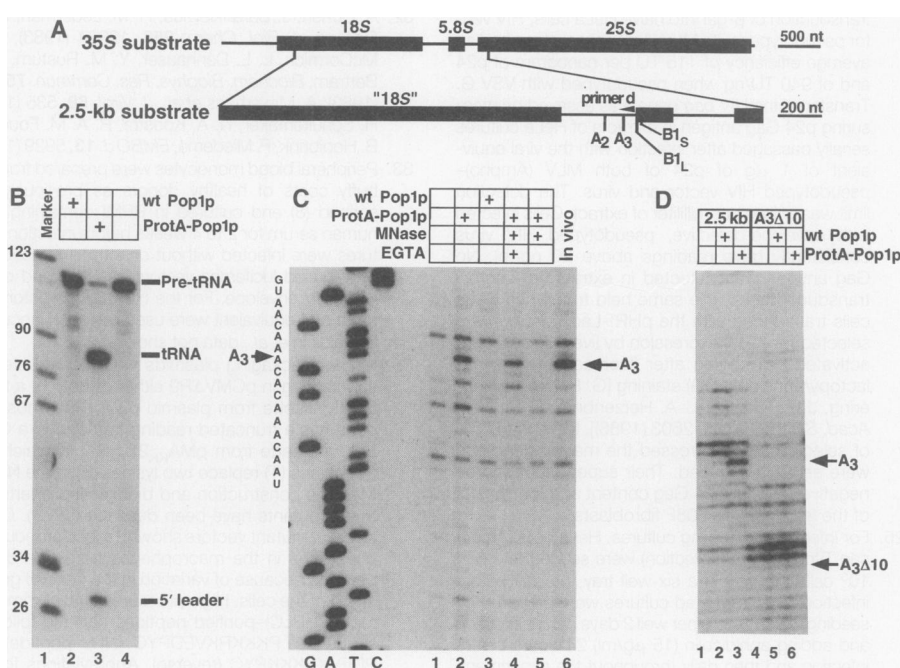
mitochondrial primers (4) have led to some controversy about RNase MRP's cellular function. Mutations in two components of yeast RNase MRP, the MRP RNA (5, 6) or Pop1p protein (7), inhibit in vivo cleavage of the pre-rRNA at a site, designated A<sub>3</sub>, located upstream of the 5.8S rRNA (7, 8). It was, however, unclear whether RNase MRP participated directly in this cleavage event.

Pop1p is a component of both RNase P and RNase MRP (7). A tagged version of Pop1p fused to two immunoglobulin G (IgG)-binding regions of *Staphylococcus aureus* protein A (ProtA-Pop1p) is functional in vivo and efficiently coprecipitates the RNase P and MRP RNAs (7).

We reasoned that the activities of both RNase P (9) and RNase MRP might be enriched from extracts containing ProtA-Pop1p by affinity selection with IgG agarose beads (10). We first tested whether yeast RNase P activity (11) could be detected by this strategy. A ProtA-Pop1p precipitate cleaved a radiolabeled pre-tRNA (Fig. 1B). This reaction was most likely mediated by RNase P for the following reasons. (i) Cleavage was dependent on the presence of ProtA-Pop1p (Fig. 1B). (ii) Cleavage was accurate (12, 13). (iii) Micrococcal nuclease treatment of the precipitate inhibited cleavage (13). (iv) The sup3e-A1 mutant pre-tRNA, which is defective for cleavage by RNase P (14), was not processed in our assay (13). Thus, affinity selection of ProtA-Pop1p can be used to detect associated enzymatic activities.

We tested next whether the same precipitates could process the 35S pre-rRNA. Because of the large size of the pre-rRNA substrate (7 kb), the products of the reaction were analyzed by primer extension (10). A primer extension stop appeared after incubation of the 35S pre-rRNA substrate with a ProtA-Pop1p precipitate (Fig. 1C). This stop mapped to site A<sub>3</sub> (Fig. 1C) and was not detected when extracts from a strain expressing nontagged Pop1p were used (Fig. 1C), showing that ProtA-Pop1p or associated factors (or both) mediate this reaction (15). The processing activity contains an essential RNA

**Fig. 1.** In vitro processing of pre-tRNA and pre-rRNA by affinity-selected ProtA-Pop1p pellets. **(A)** Structure of the 35S pre-rRNA transcript and of the 2.5-kb substrate. Mature rRNAs are shown as boxes and spacers as lines. The A<sub>2</sub> and A<sub>3</sub> cleavage sites, the 5' ends of the 5.8S rRNA (B1<sub>S</sub> and B1<sub>L</sub>), and primer d are indicated on an enlarged drawing of the 2.5-kb transcript region. Scale bars are on the right. **(B)** Endonucleolytic cleavage of a <sup>32</sup>P-labeled SupS1 pre-tRNA transcript (24). Lane 1, molecular size marker with sizes indicated on the left in nucleotides; lane 2, control reaction with a precipitate from a wild-type extract (wt Pop1p); lane 3, transcript processed with a ProtA-Pop1p precipitate; lane 4, mock-treated substrate. The pre-tRNA, mature tRNA, and 5' leader are indicated. **(C)** Processing of a nonlabeled 35S pre-rRNA at site A<sub>3</sub> analyzed by primer extension. Lane 1, mock-treated substrate depicting nonspecific primer extension stops (for example, because of secondary structure); lane 2, transcript processed with a ProtA-Pop1p precipitate; lane 3, control reaction with a precipitate from a wild-type extract; lane 4, EGTA addition prevents the micrococcal nuclease inactivation of the A<sub>3</sub> processing activity; lane 5, the A<sub>3</sub> processing activity of the ProtA-Pop1p precipitate is micrococcal nuclease (MNase)-sensitive; lane 6, primer extension on cellular RNA depicting the primer extension stop corresponding to the in vivo A<sub>3</sub> cleavage. Lanes G, A, T, and C are the cognate sequence ladder. **(D)** Mutant A<sub>3</sub>Δ10 is not processed in vitro. A 2.5-kb wild-type transcript (lanes 1 to 3) and a mutant derivative bearing a 10-nt deletion immediately 3' to site A<sub>3</sub> (lanes 4 to 6) were assayed by primer



extension for in vitro cleavage at site A<sub>3</sub>. Lanes 1 and 4, mock-treated substrates; lanes 2 and 5, substrate incubated with the IgG precipitate from a wild-type extract; lanes 3 and 6, substrate incubated with a ProtA-Pop1p precipitate. Positions corresponding to cleavage at A<sub>3</sub> of the wild-type and A<sub>3</sub>Δ10 substrates are indicated.

component because it was inactivated by micrococcal nuclease (Fig. 1C). A 10-nucleotide (nt) deletion immediately 3' to site A<sub>3</sub> prevents processing at this site in vivo (8). A 2.5-kb substrate (Fig. 1A) carrying this mutation was not processed in vitro, whereas a control wild-type 2.5-kb transcript was processed (Fig. 1D), indicating that the in vitro reaction mimics processing in vivo. Our results show that a micrococcal nuclease-sensitive activity associated with Pop1p accurately cleaves the pre-rRNA at site A<sub>3</sub> in vitro.

The RNase MRP and RNase P RNAs are the major RNA species found in a ProtA-Pop1p precipitate when assayed by end-labeling (13); therefore, one of the corresponding endonucleases is most likely to mediate the in vitro pre-rRNA processing reaction. RNase P and MRP were separated by biochemical purification (16) (Fig. 2A). In the most purified MRP fraction (M2), no RNase P RNA could be detected, whereas the purest RNase P fraction (P2) contained only trace amounts of MRP RNA (Fig. 2B). The peak RNase MRP and P fractions from each column were affinity-selected on IgG agarose beads and assayed for cleavage of a pre-tRNA substrate and the 35S pre-rRNA substrate (Fig. 2, C and D). Accurate endonucleolytic cleavage of the pre-tRNA was detected in the RNase P-containing fractions but not in the fractions highly enriched for RNase MRP (Fig. 2C). Conversely, processing of the 35S pre-rRNA was specifically detected in the RNase MRP fractions (Fig. 2D).

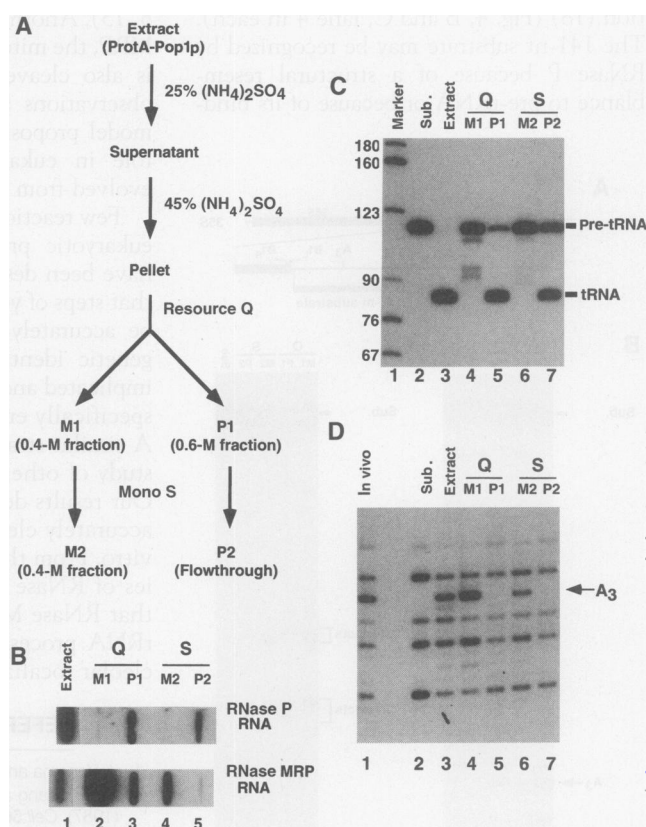
We used an in vivo depletion strategy (17) to demonstrate that RNase MRP, rather than a copurifying ribozyme, is responsible for the processing reaction. Extracts were prepared from cells that conditionally transcribe the MRP RNA (6) and also express ProtA-Pop1p. In vitro cleavage of the 35S pre-rRNA at site A<sub>3</sub> was lost in extracts prepared after in vivo depletion of MRP RNA (Fig. 3A). MRP RNA depletion did not affect RNase P activity (Fig. 3B). Extracts prepared from wild-type cells grown under the same conditions were active for processing at site A<sub>3</sub> (Fig. 3A) and RNase P activity (Fig. 3B). Therefore, depletion of the MRP RNA specifically affected processing at site A<sub>3</sub>. We conclude that RNase MRP directly and accurately processes the pre-rRNA at site A<sub>3</sub> in vitro.

To better define the substrate requirements for pre-rRNA processing, we tested a 141-nt transcript overlapping site A<sub>3</sub> (Fig. 4A). This short substrate was processed accurately at site A<sub>3</sub> by a precipitate of fraction M2 (and M1) containing

highly purified RNase MRP (Fig. 4B). Incubation of internally labeled 141-nt substrate with a purified RNase MRP precipitate produced fragments of 81 and 60 nt

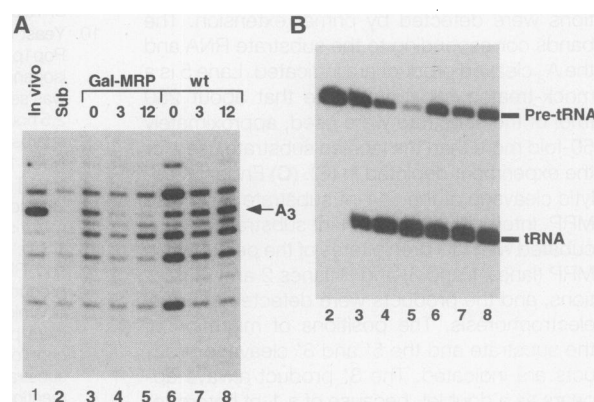
(Fig. 4C), corresponding to the 5' and 3' cleavage products, respectively. Cleavage at site A<sub>3</sub> by RNase MRP is therefore endonucleolytic, and the information re-

**Fig. 2.** Purified RNase MRP processes pre-rRNA at the A<sub>3</sub> site. **(A)** Fractionation scheme. The fractionation conditions and fraction names are indicated. Activities present in the M1, P1, M2, and P2 fractions were tested after further purification by affinity selection on IgG agarose. **(B)** Distribution of RNase P and MRP RNAs in the fractions. Lane 1, RNAs extracted from total extract; lanes 2 and 4, RNAs from the peak RNase MRP fractions after Resource Q and Mono S chromatography, respectively; lanes 3 and 5, RNAs from the peak RNase P fractions after Resource Q and Mono S chromatography, respectively. Slot-blot of the duplicate filters were hybridized either with an RNase P (top panel) or MRP RNA probe (lower panel). Lane 1 was exposed for a shorter period than were the other lanes, but identical exposure lengths are shown for the P and MRP hybridizations, allowing



direct comparison of the ratio of RNase P and MRP RNAs. Traces of RNase MRP RNA could be detected in fraction P2 after longer exposure, whereas only a background signal could be detected in the slot containing the M2 fraction probed for RNase P RNA. **(C)** Assay of the pre-tRNA processing activity by precipitates of the fractions. Precipitates from total extract (lane 3), the Resource Q M1 and P1 fractions (lanes 4 and 5), and the Mono S M2 and P2 fractions (lanes 6 and 7) were assayed for pre-tRNA processing. A mock-treated substrate is presented in lane 2 and a molecular size marker in lane 1 with the size of the corresponding bands (in nucleotides) indicated on the left. Some nonspecific degradation is apparent in lanes 4 and 6, but no specific cleavage is detected. **(D)** Assay of the pre-rRNA processing activity in precipitates of the fractions. In lanes 3 to 7, pre-rRNA processing was assayed with the same fractions as for the pre-tRNA processing in (C). Lane 2 shows a mock-treated RNA, and lane 1, RNA extracted from wild-type cells. Sub., substrate.

**Fig. 3.** Precipitates of extracts made from cells depleted in vivo of RNase MRP RNA do not process pre-rRNA at site A<sub>3</sub>. **(A)** In vitro cleavage of the pre-rRNA substrate (Sub.) by ProtA-Pop1 precipitates. A plasmid carrying the ProtA-Pop1 construct was introduced into strain MES124, which contains a galactose-regulated *NME1* gene (Gal-MRP, lanes 3 to 5), and into the otherwise isogenic wild-type strain MES123 (lanes 6 to 8) (6). Extracts were prepared from cells grown on galactose (lanes 3 and 6) and from cells grown on glucose for 3 hours (lanes 4 and 7) and 12 hours (lanes 5 and 8). Under these conditions the MRP RNA amount reaches a minimum after 8 hours (6). Lane 1, RNA extracted from wild-type cells; lane 2, mock-treated RNA. **(B)** Pre-tRNA processing by precipitates of the same extracts. Lanes are as for (A), except that lane 1 is omitted.

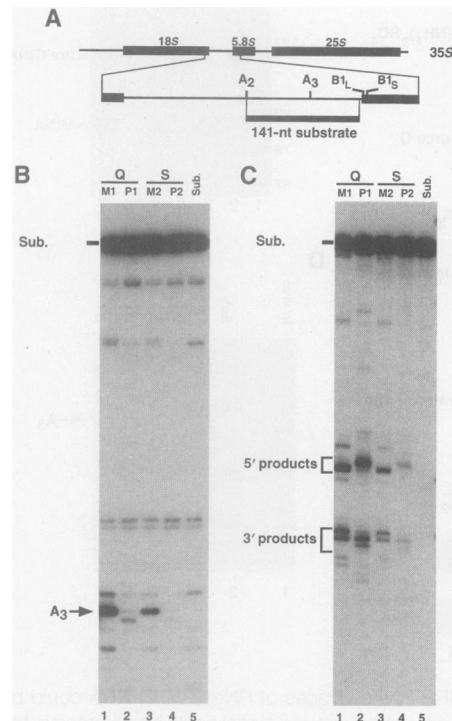


European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.

\*To whom correspondence should be addressed.

quired for substrate recognition is contained in a 141-nt fragment of the pre-rRNA.

A low level of aberrant processing of the 141-nt substrate 1 nt 3' to site A<sub>3</sub> was observed with our purest RNase P preparation (18) (Fig. 4, B and C, lane 4 in each). The 141-nt substrate may be recognized by RNase P because of a structural resemblance to pre-tRNA or because of its bind-



**Fig. 4.** A 141-nt pre-rRNA substrate is recognized and cleaved endonucleolytically by RNase MRP. (A) Location of the 141-nt substrate. The 35S pre-rRNA is shown on top, and the region surrounding site A<sub>3</sub> is shown enlarged below. The 141-nt substrate extends from 3 nt downstream of the A<sub>2</sub> site to 9 nt upstream of the B<sub>1L</sub> site (8). (B) In vitro processing of an unlabeled 141-nt substrate (Sub.). The products of the reaction with IgG precipitates of the peak RNase MRP (lanes 1 and 3) and P (lanes 2 and 4) fractions were detected by primer extension. The bands corresponding to the substrate RNA and the A<sub>3</sub> cleaved product are indicated. Lane 5 is a mock-treated substrate. Note that about 250 fmol of the substrate were used, approximately 50-fold more than the labeled substrate used for the experiment depicted in (C). (C) Endonucleolytic cleavage of the 141-nt substrate by RNase MRP. Internally labeled 141-nt substrate was incubated with IgG precipitates of the peak RNase MRP (lanes 1 and 3) and P (lanes 2 and 4) fractions, and the products were detected after gel electrophoresis. The positions of migration of the substrate and the 5' and 3' cleavage products are indicated. The 3' product always appears as a doublet, because of a 1-nt heterogeneity at the 3' end of the substrate, generated during in vitro transcription. The 3' and 5' fragments were identified by processing end-labeled substrates (13).

ing to Pop1p. Only low levels of aberrant cleavage could be detected with the longer substrates (Figs. 1 to 3), possibly because folding of these longer RNAs interferes with RNase P binding or catalysis or both. This cleavage was not detectable in vivo (7, 8, 13). Another in vitro substrate for RNase MRP, the mitochondrial replication primer, is also cleaved by RNase P (19). These observations are consistent with a recent model proposing that RNase MRP and its role in eukaryotic pre-rRNA processing evolved from RNase P (20).

Few reactions that reproduce steps of the eukaryotic pre-rRNA processing in vitro have been described (21). We have shown that steps of yeast pre-rRNA processing can be accurately reproduced in vitro by the genetic identification of the components implicated and the use of tagged proteins to specifically enrich for the desired activity. A similar strategy could be applied to the study of other complex cellular processes. Our results demonstrate that RNase MRP accurately cleaves pre-rRNA at site A<sub>3</sub> in vitro. From this and previous in vivo studies of RNase MRP mutants, we conclude that RNase MRP is directly implicated in rRNA processing, consistent with its nucleolar localization.

## REFERENCES AND NOTES

1. J. Venema and D. Tollervey, *Yeast* **11**, 1629 (1995).
2. D. D. Chang and D. A. Clayton, *Science* **235**, 1178 (1987); *Cell* **56**, 131 (1989).
3. T. Kiss and W. Filipowicz, *Cell* **70**, 11 (1992); J. N. Topper, J. L. Bennett, D. A. Clayton, *ibid.*, p. 16.
4. J. Côté and A. Ruiz-Carrillo, *Science* **261**, 765 (1993).
5. K. Shuai and J. W. Warner, *Nucleic Acids Res.* **19**, 5059 (1991); L. Lindahl, R. H. Archer, J. M. Zengel, *ibid.* **20**, 295 (1992); S. Chu, R. H. Archer, J. M. Zengel, L. Lindahl, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 659 (1994).
6. M. E. Schmitt and D. A. Clayton, *Mol. Cell. Biol.* **13**, 7935 (1993).
7. Z. Lygerou, P. Mitchell, E. Petfalski, B. Séraphin, D. Tollervey, *Genes Dev.* **8**, 1423 (1994).
8. Y. Henry *et al.*, *EMBO J.* **13**, 2452 (1994).
9. S. C. Darr, J. W. Brown, N. R. Pace, *Trends Biochem. Sci.* **17**, 178 (1992); S. Altman, L. Kirsebom, S. Talbot, *FASEB J.* **7**, 7 (1993).
10. Yeast extracts (22) were prepared from the ProtA-Pop1p-expressing strain BSY414 and the otherwise isogenic wild-type strain BSY17 (7, 23). ProtA-Pop1p was selected from 0.6 µl of extract by incubation for 2.5 hours at 4°C with 6 µl of a 50% suspension of IgG agarose beads in 120 µl of IPP150 buffer [150 mM NaCl, 10 mM tris-Cl (pH 8.0), 0.1% NP-40, and 0.1% Na<sub>2</sub>S<sub>2</sub>O<sub>5</sub>]. IgG pellets were incubated with 5 fmol of labeled pre-tRNA<sup>SupS1</sup> (24) for 30 min at 37°C in 20 mM tris-Cl (pH 8), 10 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT), 50 mM KCl, bovine serum albumin (50 µg/ml), and RNasin (60 U/ml). Pre-rRNA substrates were produced by in vitro transcription of the appropriate plasmids. Five femtomoles of pre-rRNA substrate were processed as described above for tRNA, except for Fig. 4B where 250 fmol of unlabeled 141-nt substrate were used. The products of the processing reactions with unlabeled pre-rRNA substrates were extracted and analyzed by primer extension (23) with radiolabeled oligonucleotide d (8) (Fig. 1A). Micrococcal nuclease was incubated for 20 min at 37°C with IgG precipitates after the addition of CaCl<sub>2</sub> to a final concentration of 4 mM. Before addition of the substrate, EGTA was added to a final concentration of 32 mM. For the control, EGTA was added to the IgG precipitates before addition of the micrococcal nuclease. Reaction products were fractionated by gel electrophoresis (25).
11. J. Y. Lee, C. E. Rohlman, L. A. Molony, D. R. Engelke, *Mol. Cell. Biol.* **11**, 721 (1991); A. J. Tranguch, D. W. Kindelberger, C. E. Rohlman, J. Y. Lee, D. R. Engelke, *Biochemistry* **33**, 1778 (1994).
12. G. Krupp, D. Kahle, T. Vogt, S. Char, *J. Mol. Biol.* **217**, 637 (1991).
13. Z. Lygerou, C. Allmang, D. Tollervey, B. Séraphin, data not shown.
14. D. Pearson *et al.*, *Mol. Cell. Biol.* **5**, 808 (1985).
15. Note that the use of ProtA-Pop1p precipitates was essential for the detection of in vitro processing. No activity was detected in complete extracts (13), probably because of the presence of inhibitors or the occurrence of competing reactions.
16. Extract from strain BSY414 was mixed with one-eighth volume 2M potassium phosphate (pH 8), and solid (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was added to 25% saturation. After centrifugation, solid (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was added to the supernatant to 45% saturation. The precipitate, in buffer Q6-50 [20 mM histidine (pH 6), 50 mM NaCl, 0.2 mM EDTA, 10% glycerol, 0.5 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), and 2 mM benzimidazole] was loaded on a Resource Q column (Pharmacia). Elution was performed with a linear gradient of 300 to 800 mM NaCl. The RNase MRP (M1) and the RNase P (P1) fractions were loaded separately on Mono S in buffer S7-50 [50 mM NaPO<sub>4</sub> (pH 7), 50 mM NaCl, 0.2 mM EDTA, 10% glycerol, 0.5 mM DTT, 0.5 mM PMSF, and 2 mM benzimidazole]. RNase P remained in the flowthrough (P2), whereas RNase MRP eluted in a 0.4-M NaCl step (M2). During the fractionation, we followed the RNase P and MRP RNAs by slot blot hybridizations and ProtA-Pop1p by protein immunoblotting (25). The RNase P and MRP peak fractions were affinity-selected with IgG agarose. Less than 0.01% of the starting protein was present in fractions M2 and P2. Because of the high amount of IgGs present on the beads, it is not possible to directly evaluate the final level of purification after the affinity selection. The RNase MRP present in fraction P1 is a minor amount of the total RNase MRP present in extracts, which is poorly functional possibly because it is missing some protein component. This is in agreement with its aberrant chromatographic behavior compared with the bulk of RNase MRP.
17. B. Séraphin and M. Rosbash, *Cell* **59**, 349 (1989).
18. The 141-nt transcript is processed by precipitates from whole-cell extracts containing ProtA-Pop1p both at site A<sub>3</sub> and 1 nt 3' to site A<sub>3</sub>. Processing at both sites is micrococcal nuclease-sensitive (13).
19. T. Potuschak, W. Rossmann, R. Karwan, *Nucleic Acids Res.* **21**, 3239 (1993).
20. J. P. Morrissey and D. Tollervey, *Trends Biochem. Sci.* **20**, 78 (1995).
21. G. J. Hannon *et al.*, *Mol. Cell. Biol.* **9**, 4422 (1989); M. T. Yip and M. J. Holland, *J. Biol. Chem.* **264**, 4045 (1989); S. Kass, K. Tyc, J. A. Steitz, B. Sollner-Webb, *Cell* **60**, 897 (1990).
22. R.-J. Lin, A. J. Newman, S.-C. Cheng, J. Abelson, *J. Biol. Chem.* **260**, 14780 (1985).
23. B. Séraphin, L. Kretzner, M. Rosbash, *EMBO J.* **7**, 2533 (1988).
24. D. Drainas, S. Zimmerly, I. Willis, D. Soll, *FEBS Lett.* **251**, 84 (1989).
25. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).
26. We thank D. A. Clayton, M. E. Schmitt, J. Morrissey, J. Mermoud, I. Willis, and D. Söll for the gift of yeast strains, plasmids, and enzymes, D. Engelke for the exchange of unpublished information, and J. Lewis for helpful suggestions. We thank our colleagues at the European Molecular Biology Laboratory for careful reading of the manuscript and their help. C.A. is supported by a fellowship from the European Union. B.S. is on leave from Centre National de la Recherche Scientifique.

13 November 1995; accepted 8 February 1996



# The Role of the 3' External Transcribed Spacer in Yeast Pre-rRNA Processing

Christine Allmang and David Tollervey\*

*Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom*

We have undertaken a deletion analysis of the 3' external transcribed spacer (3' ETS) in the pre-rRNA of *Saccharomyces cerevisiae*. A stem loop structure immediately 3' to the 25 S rRNA region is necessary and sufficient for processing of the 3' ETS. This is believed to be by cotranscriptional cleavage by Rnt1p, the yeast homologue of RNase III. In addition, this stem-loop is required for cleavage of site A<sub>3</sub> by RNase MRP and for processing at site B<sub>1L</sub> in the 3' region of ITS1. Processing at an upstream site in ITS1, site A<sub>2</sub>, and at sites in the 5' external transcribed spacer are not affected, even by complete deletion of the 3' ETS. We conclude that processing in the 3' ETS and in ITS1 is coupled. This would constitute a quality control that prevents synthesis of the 5.8 S rRNA and 5' end maturation of the 25 S rRNA in transcripts which are incomplete due to premature transcription termination.

© 1998 Academic Press Limited

**Keywords:** rRNA; ribosomes; nucleolar; *Saccharomyces cerevisiae*; RNase MRP

\*Corresponding author

## Introduction

In eukaryotes, the ribosomal RNAs are produced in the nucleolus and cotranscribed as a single large precursor RNA (pre-rRNA) that is processed into the mature 18 S, 5.8 S and 25 S rRNAs by removal of the external transcribed spacers (5' ETS and 3' ETS) and internal transcribed spacers (ITS1 and ITS2). Removal of the transcribed spacers involves a series of processing steps carried out by endonucleases and exonucleases (Venema & Tollervey, 1995) (Figure 1).

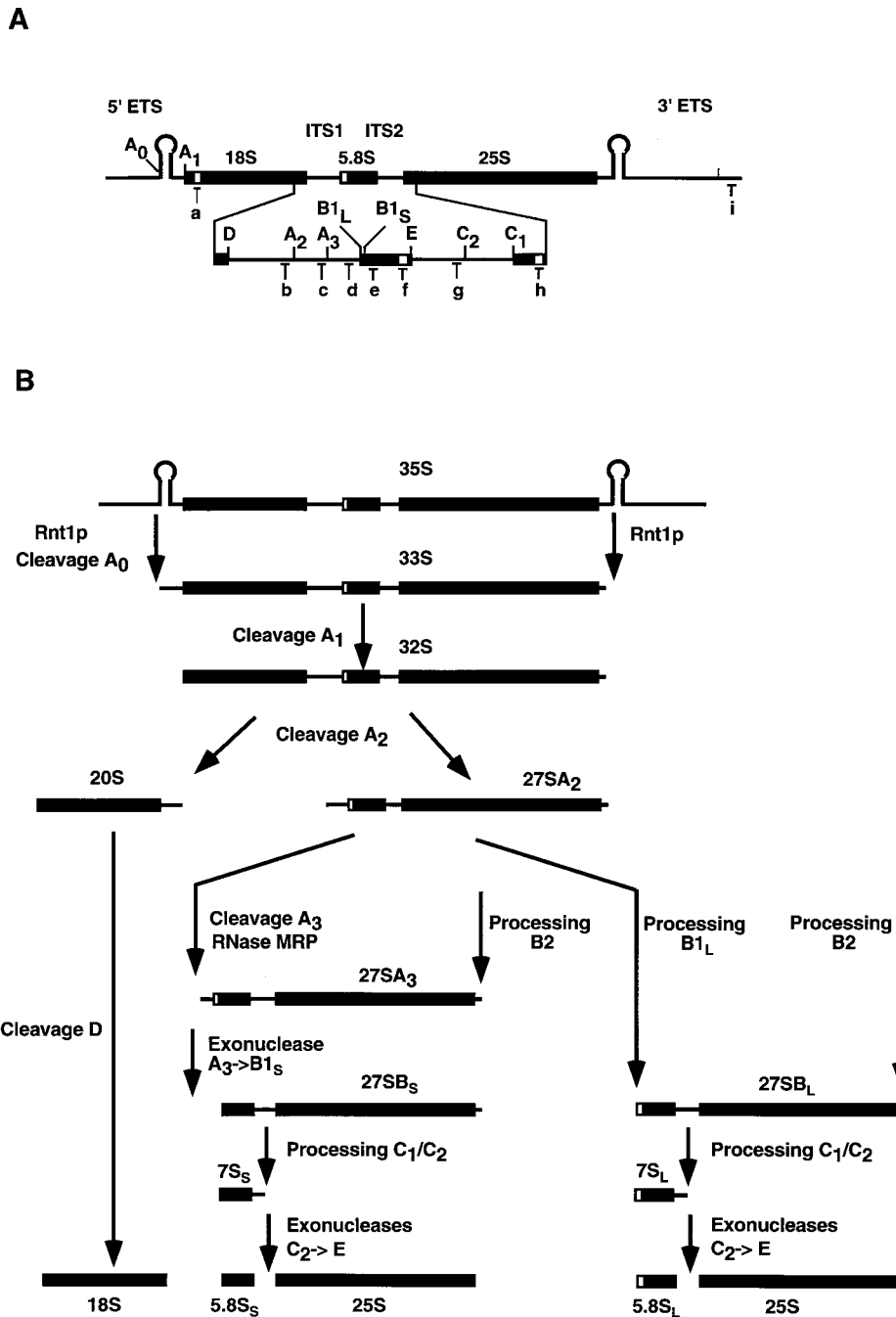
In yeast, the longest detectable pre-rRNA transcript, the 35 S pre-rRNA, is generated by cleavage in the 3' ETS and is reported to extend from the initiation site to a position seven nucleotides beyond the 3' end of 25 S rRNA. Early studies using rRNA mini-gene transcripts suggested that formation of the 3' end of 25 S rRNA is a multi-step process (Kempers-Veenstra, 1986; Veldman *et al.*, 1980). Transcription termination, mapped at position +210 relative to the 3' end of 25 S rRNA (Kempers-Veenstra, 1986), produces the 3' ETS. This is removed by endonucleolytic cleavage at sites between nucleotides +15 and +50 followed by exonucleolytic digestion to position +7

(Kempers-Veenstra, 1986). The conserved region +15 to +50 of dyad symmetry was shown to be required and sufficient for 3' end maturation of 25 S rRNA *in vitro*. This step of the yeast processing pathway was the first to be reproduced *in vitro* using partially purified yeast whole cell extracts (Yip & Holland, 1989). However, few *trans*-acting factors involved in 3' ETS processing have been identified. The product of the *RNA82* gene (Piper *et al.*, 1983) is likely to be required since *rna82-1* mutants affect 3' end formation of transcripts derived from a mini-gene reporter (Kempers-Veenstra, 1986). More recently, Abou Elela *et al.* (1996) identified the endonuclease Rnt1p, the yeast homologue of *Escherichia coli* RNase III, which cleaves the conserved stem-loop structure in the 3' ETS. In an *rnt1-1* strain, processing of the 3' ETS was inhibited leading to the accumulation of 3' extended forms of the 27 S pre-rRNAs and 25 S rRNA. A synthetic 3' ETS substrate was specifically cleaved *in vitro* by recombinant Rnt1p at a site within the stem-loop structure located 21 nt downstream of the 3' end of 25 S rRNA; a position close to, but not identical with, the reported sites of *in vivo* processing.

Rnt1p is also likely to carry out the endonucleolytic cleavage in the 5' ETS at site A<sub>0</sub> (Abou Elela *et al.*, 1996). The recombinant protein is able to cleave a synthetic 5' ETS RNA at site A<sub>0</sub> *in vitro* in the absence of cofactors. In contrast, a large number of *trans*-acting factors are required for the early

Abbreviations used: ETS, external transcribed spacer; ITS, internal transcribed spacer; pre-rRNA, precursor rRNA; snoRNP, small nucleolar ribonucleoprotein particle.





**Figure 1.** Structure and processing pathway of the pre-rRNA in *Saccharomyces cerevisiae*. **A**, Structure of the 35 S pre-rRNA and position of oligonucleotides. Thick lines represent the mature rRNA sequences and thin lines the transcribed spacers. The hybridization positions of oligonucleotides a to i are indicated. Open boxes within the 18 S, 5.8 S and 25 S rRNA regions indicate the location of the tags. **B**, Pre-rRNA processing pathway. The rDNA is transcribed into a single large pre-rRNA that undergoes sequential cleavage to generate the mature rRNAs. Cleavages by Rnt1p in the 3' ETS and at site A<sub>0</sub> in the 5' ETS generate the 35 S pre-rRNA and the 33 S pre-rRNA respectively. The 33 S pre-rRNA is subsequently processed at sites A<sub>1</sub> at the 5' end of 18 S rRNA, generating 32 S pre-rRNA, and site A<sub>2</sub> in ITS1 giving rise to the 20 S and 27 SA<sub>2</sub> precursors. A<sub>2</sub> cleavage separates the pre-rRNAs destined to form the small and large ribosomal subunit rRNAs. The 20 S precursor is then endonucleolytically cleaved at site D to yield the mature 18 S rRNA. The 27 SA<sub>2</sub> pre-rRNA is processed by two alternative pathways giving rise to the two forms of 5.8 S rRNA, the major form, 5.8 S<sub>S</sub>, and the minor form, 5.8 S<sub>L</sub>. Formation of 5.8 S<sub>S</sub> requires cleavage of 27 SA<sub>2</sub> by RNase MRP at site A<sub>3</sub> to generate the 27 SA<sub>3</sub> pre-rRNA. This site acts as an entry site for two exonucleases, Rat1p and Xrn1p, that degrade the pre-rRNA 5' → 3' to site B1<sub>S</sub>, generating the 5' end of the short form of the 27 SB pre-rRNA, 27 SB<sub>S</sub>. This pre-rRNA is subsequently processed to the 5.8 S<sub>S</sub> and 25 S rRNAs. Processing at sites C<sub>1</sub> and C<sub>2</sub> generates the 7 S<sub>S</sub>, which is converted to 5.8 S<sub>S</sub> by a complex of 3' → 5' exonucleases, the "exosome". An alternative pathway leads to the cleavage at site B1<sub>L</sub>, the 5' end of the 27 SB<sub>L</sub> pre-rRNA, which is processed to yield the 5.8 S<sub>L</sub> and 25 S rRNAs.

cleavages of the pre-rRNA *in vivo*. The major class comprises a number of small nucleolar ribonucleoprotein particles (snoRNPs; for reviews see Maxwell & Fournier, 1995; Tollervey & Kiss, 1997). Four snoRNPs, U3, U14, snR10 and snR30 have been shown to be required for the early cleavages in the 5' region of the pre-rRNA at sites A<sub>0</sub>, A<sub>1</sub> and A<sub>2</sub> (see Figure 1). Genetic depletion of any of the RNA or protein components of these snoRNP has similar effects: the inhibition of cleavage at site A<sub>0</sub>, A<sub>1</sub> and A<sub>2</sub>, resulting in the inhibition of the synthesis of mature 18 S rRNA. These snoRNP components may participate in a multi-snoRNP processing complex, probably assembling on the 5' ETS, that brings together the sequences surrounding sites A<sub>0</sub>, A<sub>1</sub> and A<sub>2</sub> in order to coordinate their cleavage (reviewed by Morrissey & Tollervey, 1995).

Subsequent cleavages further 3' in the pre-rRNA generate the 5.8 S and 25 S rRNAs. Another RNP particle, RNase MRP plays a role in these processing reactions (Chu *et al.*, 1994; Lygerou *et al.*, 1994, 1996; Schmitt & Clayton, 1993). RNase MRP directly cleaves site A<sub>3</sub> located in ITS1 (Lygerou *et al.*, 1996), providing an entry site for 5' → 3' exonuclease degradation to site B1<sub>S</sub>, the 5' end of the major form of 5.8 S<sub>S</sub> rRNA (Henry *et al.*, 1994). This trimming requires two proteins, Xrn1p and Rat1p, that exhibit a 5' → 3' exonuclease activity *in vitro* (Amberg *et al.*, 1992; Kenna *et al.*, 1993; Larimer *et al.*, 1992; Stevens & Poole, 1995). An alternative, less understood, pathway processes site B1<sub>L</sub> the 5' end of the minor 5.8 S<sub>L</sub> rRNA. Although capable of functioning independently, the RNase MRP complex assembled at site A<sub>3</sub> and the snoRNP complex assembled in the 5' ETS and site A<sub>2</sub> are believed to interact to bring about efficient A<sub>2</sub> and A<sub>3</sub> cleavage (Allmang *et al.*, 1996). This interaction may occur *via* a bridging factor, Rrp5p (Venema & Tollervey, 1996).

Strikingly, it appears that none of the pre-rRNA processing activities function on the nascent pre-rRNA; yeast pre-rRNA processing is initiated only on the fully transcribed pre-rRNA. Following completion of transcription, however, the 35 S pre-rRNA undergoes very rapid processing. We therefore speculated that a signal present in the 3' ETS might be required to initiate the processing pathway.

Various cis-acting signals are predicted to be present in the 3' ETS and in search of these signals we have undertaken an analysis of the effects of deletions in the 3' ETS on pre-rRNA processing *in vivo*.

## Results

### Nested deletions in the 3' ETS

To test the effects of deletions in the 3' ETS on ribosome synthesis we have used a system allowing the conditional expression of mutant and wild-type pre-rRNA (Henry *et al.*, 1994). Previous ana-

lyses have shown that yeast RNA polymerase I transcription terminates 210 nt beyond the mature 3' end of 25 S rRNA (Veldman *et al.*, 1980). In the GAL::rDNA construct, the GAL7 terminator region is inserted 284 nt 3' to the 25 S coding sequence and introduces a *SalI* site directly adjacent to the rDNA coding sequence (Nogi *et al.*, 1991). Nested deletions were constructed that encompassed the whole 3' ETS region, including the stem-loop structure cleaved by Rnt1p, and analyzed *in vivo*. Unidirectional deletions were generated by Exonuclease III from a synthetic linker inserted into the *SalI* site (see Materials and Methods) and enter the 3' ETS progressively from 3' to 5'. The extents of these deletions are shown Figure 2. The clearest predicted structure in this region is a strong stem-loop located immediately 3' to the 25 S rRNA coding sequence (nt +8 to +56; see Figure 2B). Mutations 3' ETS Δ1 to Δ3 leave the stem-loop structure intact, whereas 3' ETS Δ4 and Δ5 enter the stem-loop. 3' ETS Δ6 deletes the entire stem-loop structure and in addition deletes six nucleotides from the 3' end of the 25 S rRNA. In the 3' ETS Δ6 + H mutant, the stem-loop structure was recreated, with the last six nucleotides of 25 S rRNA substituted by the *SalI* restriction site sequence (see Material and Methods). Three of the last six nucleotides of the 25 S rRNA are predicted to be base-paired at the base of a helical stem within the 25 S rRNA (Gutell & Fox, 1988). The 3' ETS Δ6 + H mutation may therefore alter the structure of the 3' helix in the mature 25 S rRNA.

The mutations were functionally analysed by expression in the pGAL::rDNA construct. Expression of the pre-rRNA carrying the deletions that leave the hairpin structure intact (3' ETS Δ1, Δ2 and Δ3) supported the growth of an *rpa12* strain at the non-permissive temperature (37°C) on galactose medium (Figure 3) at rates similar to the wild-type pre-rRNA. In contrast, expression of the pre-rRNA carrying the mutations that enter the stem-loop structure (3' ETS Δ4, Δ5 and Δ6) did not support the growth of the *rpa12* strain. Restoration of the hairpin structure in 3' ETS Δ6 + H pre-rRNA did not restore the ability to support growth in an *rpa12* strain (Figure 3).

### Deletions in the 3' ETS affect 3' end processing of the pre-rRNA

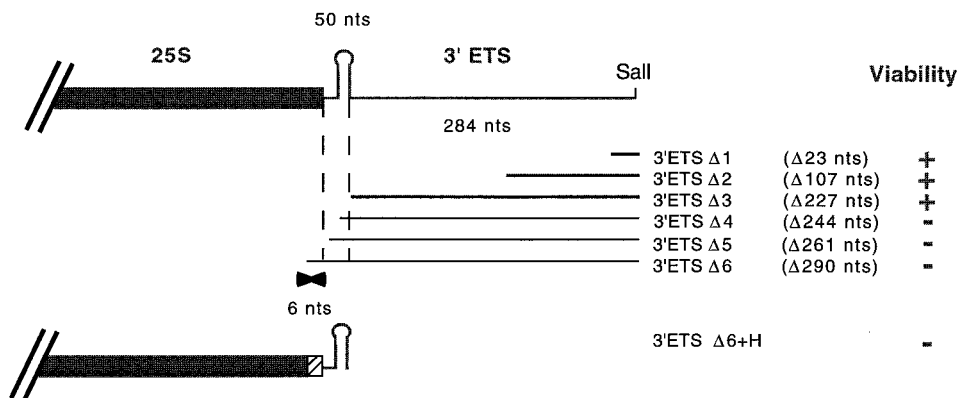
The effect of the 3' ETS deletions on pre-rRNA processing were examined by Northern hybridization six hours after transfer to 37°C (Figure 4). The steady state levels of the mature rRNAs were determined using oligonucleotides complementary to neutral tags that are present in the 25 S and 18 S rRNA (Beltrame & Tollervey, 1992), while the levels of the pre-rRNAs were determined using the oligonucleotides indicated in figure 1A. None of the 3' ETS deletions affected the accumulation of the 18 S rRNA (Figure 4VII) or altered the levels of the 20 S pre-rRNA (Figure 4VI). This showed that global processing of the pre-rRNA was not

affected by the deletions. We conclude that, contrary to our expectations, there is no signal in the 3' ETS that is required to initiate the pre-rRNA processing pathway.

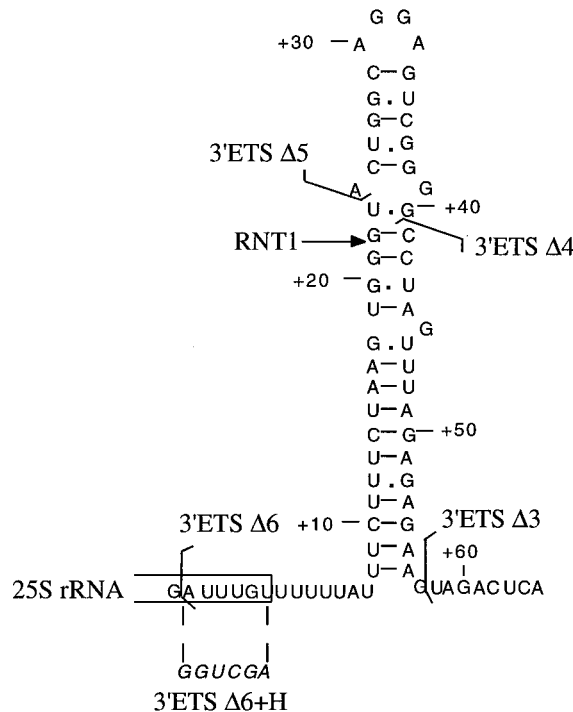
The 3' ETS  $\Delta 1$ ,  $\Delta 2$  and  $\Delta 3$  mutant pre-rRNAs generated normal levels of 25 S rRNA (Figure 4V,

lanes 3 to 5) and pre-rRNA processing was unaffected with normal levels of all the major intermediates. Therefore, no sequence 3' to the stem-loop structure is required for normal function of the 3' ETS. In contrast, synthesis of the 25 S rRNA from pre-rRNAs carrying the 3' ETS  $\Delta 4$  and  $\Delta 5$

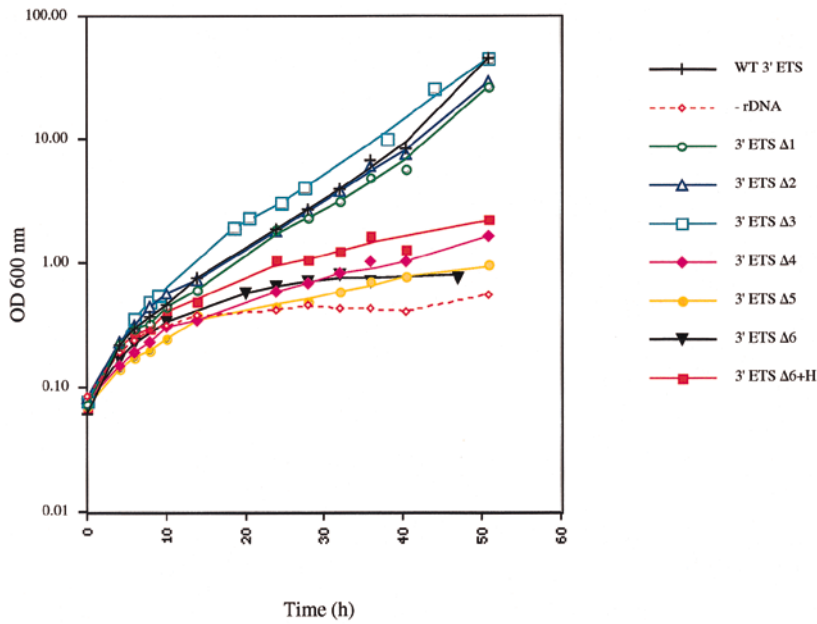
A)



B)



**Figure 2.** Deletions in the 3' ETS. A, Schematic representation of the 3' ETS region (thin line) showing the stem-loop structure and the *SalI* site located into the GAL7 terminator (Nogi *et al.*, 1991). The extent and size of the 3' → 5' deletions in mutants 3' ETS  $\Delta 1$  to 3' ETS  $\Delta 6$  is represented below. In the schematic representation of 3' ETS  $\Delta 6 + H$  the hatched box indicates the altered region in the re-introduced sequence. The viability of the *rpa12* strain at non-permissive temperature is indicated in each case; (+) representing a fully viable mutation, (–) representing a lethal mutation. B, Predicted structure of the 5' region of the 3' ETS. The major site of *in vitro* cleavage by Rnt1p (Abou Elela *et al.*, 1996) and the end points of the 3' ETS  $\Delta 3$ ,  $\Delta 4$ ,  $\Delta 5$  and  $\Delta 6$  mutations are indicated. The sequence in italics is that re-introduced in the  $\Delta 6 + H$  mutation.



**Figure 3.** Growth analysis of 3' ETS deletion mutants. Growth curves of NOY504 strains transformed by pTA1 plasmids containing the 3' ETS deletions. Cells were grown in minimal galactose medium at 23°C until reaching an  $A_{600\text{ nm}}$  of 0.07 and then shifted to 37°C, the non-permissive temperature, to repress chromosomal rDNA transcription. Growth curves were measured after the temperature shift. The symbols used for each mutant are represented beside the curves. Strains were transformed with the following vectors: (WT 3' ETS) vector containing the wild-type 3' ETS sequence; (-rDNA) vector lacking the rDNA sequence; (3' ETS  $\Delta 1$ - $\Delta 6$  and 3' ETS  $\Delta 6 + H$ ) vectors containing the 3' ETS deletions reported in Figure 2.

mutations was strongly reduced (Figure 4V, lanes 6 and 7), and no detectable 25 S rRNA was synthesized from the 3' ETS  $\Delta 6$  pre-rRNA (Figure 4V, lane 8). Restoration of the stem-loop structure in the 3' ETS  $\Delta 6 + H$  pre-rRNA only partially restored 25 S rRNA synthesis (Figure 4V, lane 9), to levels similar to those observed for the 3' ETS  $\Delta 4$  and  $\Delta 5$  constructs. Transformants carrying a plasmid without the rDNA unit did not show any background signal (Figure 4, lane 2). The 3' ETS  $\Delta 4$ ,  $\Delta 5$  and  $\Delta 6$  mutations resulted in the production of aberrant extended forms of 27 SA<sub>2</sub> pre-rRNA (Figure 4II, lanes 6 to 8) and 27 SB (Figure 4III, lanes 6 to 8). Furthermore, an extended form of 25 S rRNA was synthesized from the 3' ETS  $\Delta 4$  and  $\Delta 5$  pre-rRNAs. These aberrant intermediates, 27 SA<sub>2</sub><sup>\*</sup>, 27 SB<sup>\*</sup> and 25 S<sup>\*</sup>, represent forms of the corresponding RNAs that are 3' extended up to the GAL7 terminator since they can be detected by oligonucleotide i (complementary to the GAL7 terminator sequence; Figure 4IV, lanes 6 and 7). The levels of 27 SA<sub>2</sub><sup>\*</sup> synthesized from the 3' ETS  $\Delta 4$  and  $\Delta 5$  are comparable to those of the 27SA<sub>2</sub> synthesized from the wild-type pre-rRNA (Figure 4II), again indicating that processing at site A<sub>2</sub> is unaffected by even the complete deletion of the 3' ETS. The 27 SA<sub>2</sub><sup>\*</sup> pre-rRNA, however, slightly accumulates in  $\Delta 6$  pre-rRNAs (Figure 4II). In contrast, levels of 27 SB<sup>\*</sup> synthesized from the 3' ETS  $\Delta 4$  and  $\Delta 5$  pre-rRNAs are strongly reduced as compared to the wild-type 27 SB (Figure 4III). Moreover, no 27 SB<sup>\*</sup> and 25 S<sup>\*</sup> rRNA are synthesized from the 3' ETS  $\Delta 6$  mutant, in marked contrast to the 3' ETS  $\Delta 4$  and  $\Delta 5$  pre-rRNAs from which 25 S<sup>\*</sup> rRNA is synthesized. The ratio of 25 S<sup>\*</sup>:27 SA<sub>2</sub><sup>\*</sup> in the  $\Delta 4$  and  $\Delta 5$  pre-rRNAs is substantially lower than the ratio of 25 S:27 SA<sub>2</sub> synthesized from the wild-type pre-rRNA (data not

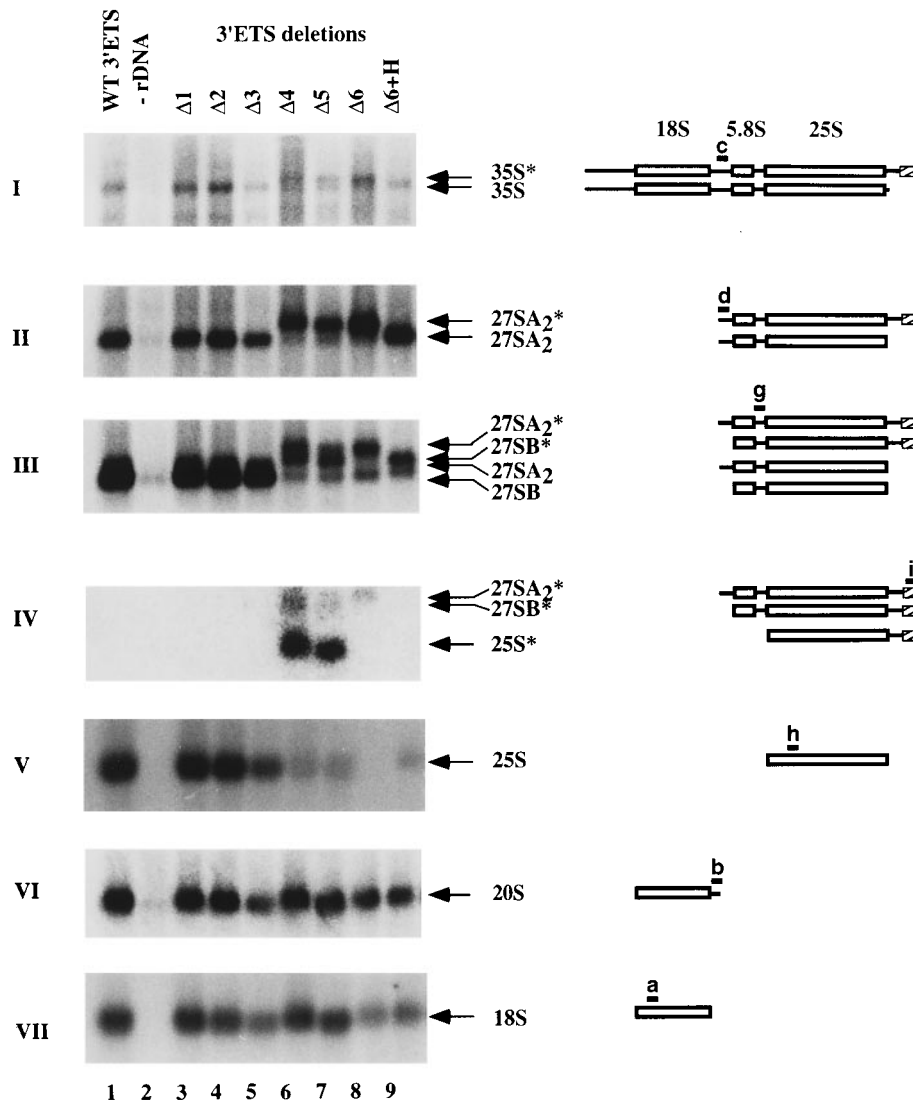
shown). This indicates that the processing of 27 SA<sub>2</sub><sup>\*</sup> to 25 S<sup>\*</sup> is relatively inefficient, consistent with the reduced level of 27 SB<sup>\*</sup>.

When the 3' ETS hairpin structure is restored in mutant 3' ETS  $\Delta 6 + H$ , the aberrant 3' extended intermediates are lost, and only normal pre-rRNA species are detected (Figure 4II, III, IV, lane 9). However, the 27 SA<sub>2</sub> pre-rRNA slightly accumulates (Figure 4II) and the level of 27 SB is strongly decreased (Figure 4III).

We conclude that an intact stem-loop structure is required for processing of the 3' ETS. Mutations which enter (3' ETS  $\Delta 4$  and  $\Delta 5$ ) or remove (3' ETS  $\Delta 6$ ) this structure prevent processing, leading to the accumulation of pre-rRNA species that are 3' extended to the transcription termination site. Processing of the 3' ETS is restored by re-insertion of the hairpin structure in the 3' ETS  $\Delta 6 + H$  pre-rRNA. In addition, both the six nucleotides at the 3' end of the 25 S rRNA and the hairpin structure are important for the processing in ITS1 that converts the 27 SA<sub>2</sub> pre-rRNA to 27 SB. Since the major processing pathway for 27 SA<sub>2</sub> is cleavage at site A<sub>3</sub> by RNase MRP, these data indicate that A<sub>3</sub> cleavage was inhibited by the mutations in the 3' ETS.

### Deletions in the 3' ETS affect processing in ITS1

The effects of the deletions on the efficiency and accuracy of the pre-rRNA cleavages in ITS1 were analysed by primer extension using oligonucleotide g that hybridizes to the pre-rRNA in ITS2 (see Figure 1). The levels of primer extension products terminating at site A<sub>2</sub> reflect the abundance of the 27 SA<sub>2</sub> and 27 SA<sub>2</sub><sup>\*</sup> pre-rRNAs; these are unaffected by any of the deletions in the 3' ETS

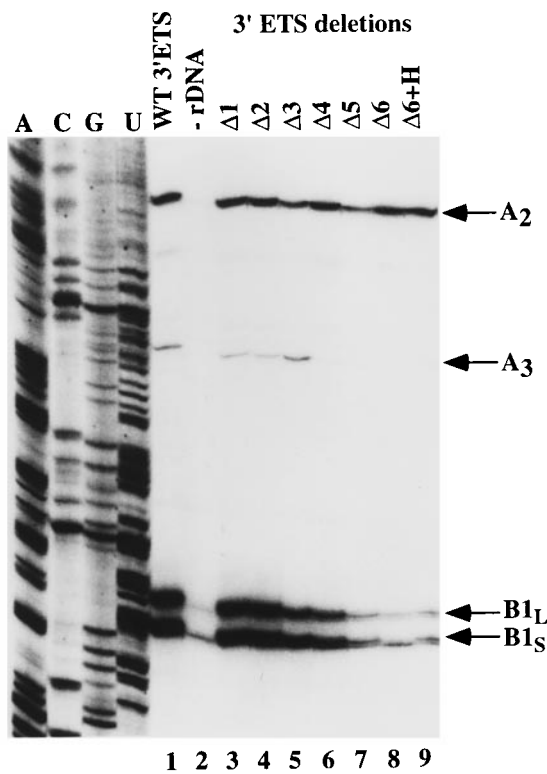


**Figure 4.** Northern analysis of high molecular weight pre-rRNA and rRNAs from strains expressing 3' ETS deletions. RNA was extracted from NOY504 transformants expressing various plasmid-borne mutant pre-rRNA after six hours growth at 37°C. RNA was separated on a 1.2% agarose gel containing formaldehyde and transferred for Northern hybridization. Lane 1, pre-rRNA with the wild-type 3' ETS sequence. Lane 2, vector lacking the rDNA sequence. Lane 3, 3' ETS  $\Delta 1$  pre-rRNA. Lane 4, 3' ETS  $\Delta 2$  pre-rRNA. Lane 5, 3' ETS  $\Delta 3$  pre-rRNA. Lane 6, 3' ETS  $\Delta 4$  pre-rRNA. Lane 7, 3' ETS  $\Delta 5$  pre-rRNA. Lane 8, 3' ETS  $\Delta 6$  pre-rRNA. Lane 9, 3' ETS  $\Delta 6 + H$  pre-rRNA. I, oligo c. II, oligo d. III, oligo g. IV, oligo i. V, oligo h. VI, oligo b. VII, oligo a. For simplicity we only show regions of each Northern; together all the high molecular weight RNAs detected are shown. The position of the various pre-rRNAs and rRNAs is indicated on the right and the RNA is schematically represented; boxes represent the mature rRNA sequences, hatched boxes the *GAL7* terminator sequences and thin lines the transcribed spacers. The hybridization site and the name of the oligonucleotide used for the detection of each species is indicated.

(Figure 5). Similarly, the early cleavages  $A_0$  and  $A_1$  were not affected by the deletions (data not shown), consistent with the Northern hybridization data. In contrast, the primer extension stop at site  $A_3$  is lost in pre-rRNAs with deletions that enter or delete the stem-loop structure (3' ETS  $\Delta 4$ ,  $\Delta 5$  or  $\Delta 6$ ; Figure 5, lanes 6 to 8) and is not restored by re-introduction of the hairpin in 3' ETS  $\Delta 6 + H$  (Figure 5, lane 9). The levels of the primer extension stops at both  $B1_S$  and  $B1_L$  are reduced to levels close to that of the  $-rDNA$  control in mutants 3' ETS  $\Delta 5$ ,  $\Delta 6$  and  $\Delta 6 + H$  (Figure 5, lanes 7 to 9), and reduced to a lesser extent in 3'

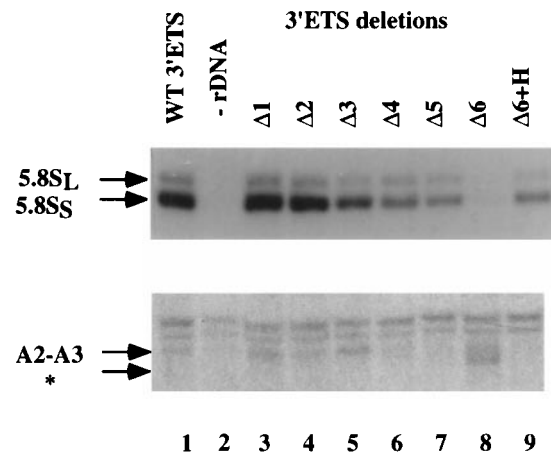
ETS  $\Delta 4$  (Figure 5, lane 6). These results are consistent with the low levels of 27 SB and 27 SB\* detected for these mutants by Northern hybridization (Figure 4).

The primer extension data for site  $A_3$  showed the loss of the 27 SA<sub>3</sub> pre-rRNA from the 3' ETS  $\Delta 4$ - $\Delta 6$  mutants. To determine whether this is due to the destabilization of the pre-rRNA or to the inhibition of pre-rRNA cleavage at site  $A_3$  we examined the level of the excised  $A_2$ - $A_3$  fragment. Northern hybridization of low molecular weight RNA using a probe hybridizing between sites  $A_2$  and  $A_3$  (oligonucleotide c in Figure 1) showed that



**Figure 5.** Primer extension analysis through ITS1 of pre-rRNA containing deletions in 3' ETS. RNA was extracted from NOY504 transformants expressing various plasmid-borne mutant pre-rRNA after six hours growth at 37°C. Primer extension was performed using oligonucleotide g hybridizing within ITS2. A dideoxynucleotide sequence generated with the same oligonucleotide was run in parallel. Lane 1, pre-rRNA with the wild-type 3' ETS sequence. Lane 2, vector lacking the rDNA sequence. Lane 3, 3' ETS Δ1 pre-rRNA. Lane 4, 3' ETS Δ2 pre-rRNA. Lane 5, 3' ETS Δ3 pre-rRNA. Lane 6, 3' ETS Δ4 pre-rRNA. Lane 7, 3' ETS Δ5 pre-rRNA. Lane 8, 3' ETS Δ6 pre-rRNA. Lane 9, 3' ETS Δ6 + H pre-rRNA. The positions of primer extension stops at sites A<sub>2</sub>, A<sub>3</sub>, B1<sub>S</sub> and B1<sub>L</sub> are indicated.

this fragment can be detected in strains expressing the wild-type or the Δ1, Δ2 and Δ3 mutant pre-rRNAs (Figure 6, lanes 3 to 5), but is not detected in the strains expressing the Δ4 or Δ5 pre-rRNAs (Figure 6, lanes 6 and 7). We conclude that the 3' ETS mutations that enter the stem-loop structure prevent excision of this fragment. Since A<sub>2</sub> cleavage was unaffected in the mutant pre-rRNAs, we conclude that cleavage at site A<sub>3</sub> was indeed specifically inhibited. In the case of the 3' ETS Δ6 mutation (Figure 6, lane 8) the A<sub>2</sub>-A<sub>3</sub> fragment was not detected but a smaller RNA species was consistently observed using either oligonucleotide c or d (marked with \* in Figure 6). This species was not detected when the stem-loop structure was re-inserted in the Δ6 + H pre-rRNA (Figure 6, lane 9). The identity of this species is currently uncertain.



<b>5.8S<sub>S</sub>:5.8S<sub>L</sub></b>	4.2	-	4.4	4.5	4.5	2.7	2.8	-	4.4
--	-----	---	-----	-----	-----	-----	-----	---	-----

**Figure 6.** Northern analysis of 5.8 S rRNA and the A<sub>2</sub>-A<sub>3</sub> RNA fragment. RNA was extracted from NOY504 transformants expressing various plasmid-borne mutant pre-rRNA after six hours growth at 37°C. Low molecular weight RNA was separated on an 8% polyacrylamide gel containing 8 M urea and transferred for Northern hybridization. Upper panel: mature 5.8 S detected with oligo f, complementary to the tag sequence. Lower panel: excised A<sub>2</sub>-A<sub>3</sub> fragment detected with oligo c. Lane 1, pre-rRNA with the wild-type 3' ETS sequence. Lane 2, vector lacking the rDNA sequence. Lane 3, 3' ETS Δ1 pre-rRNA. Lane 4, 3' ETS Δ2 pre-rRNA. Lane 5, 3' ETS Δ3 pre-rRNA. Lane 6, 3' ETS Δ4 pre-rRNA. Lane 7, 3' ETS Δ5 pre-rRNA. Lane 8, 3' ETS Δ6 pre-rRNA. Lane 9, 3' ETS Δ6 + H pre-rRNA. The ratios between 5.8 S<sub>S</sub> and 5.8 S<sub>L</sub> measured by PhosphorImager scanning (Molecular Dynamics) are indicated in the table below the Figure.

Sites B1<sub>S</sub> and B1<sub>L</sub> are the 5' end of the 5.8 S<sub>S</sub> and 5.8 S<sub>L</sub> rRNAs, respectively. Northern hybridization using a probe specific for the tag in 5.8 S rRNA (oligonucleotide f in Figure 1) shows a strong reduction in the accumulation of both 5.8 S<sub>S</sub> and 5.8 S<sub>L</sub> rRNAs synthesized from the 3' ETS Δ4 and Δ5 pre-rRNAs (Figure 6, lanes 6 and 7). In addition, these deletions under-accumulate 5.8 S<sub>S</sub> compared to 5.8 S<sub>L</sub> rRNA; the ratio is reduced by a factor of 1.5, as determined by PhosphorImager scanning (Figure 6). The ratios of 5.8 S<sub>S</sub> to 5.8 S<sub>L</sub> rRNA synthesized from the 3' ETS Δ1, Δ2 and Δ3 pre-rRNAs are identical to the wild-type. Synthesis of 5.8 S<sub>S</sub> rRNA requires upstream cleavage at site A<sub>3</sub> (Henry *et al.*, 1994) and its under accumulation in the Δ4 and Δ5 mutants confirms the inhibition of cleavage at this site. In the case of the 3' ETS Δ6 pre-rRNA, accumulation of 5.8 S rRNA was undetectable (Figure 6, lane 8). Re-insertion of the stem-loop structure in the 3' ETS Δ6 + H pre-rRNA allows formation of low levels of 5.8 S<sub>S</sub> and

5.8 S<sub>L</sub> rRNA, with a ratio similar to the wild-type control.

From these data we conclude that the integrity of the stem-loop structure in the 3' ETS is important both for processing of the 3' ETS and for cleavage of site A<sub>3</sub> in ITS1. The alternative ITS1 pathway that processes site B1<sub>L</sub> is less sensitive to disruption of the stem-loop structure, as shown by the increase in the ratio of 5.8 S<sub>L</sub> to 5.8 S<sub>S</sub> rRNA synthesized from the 3' ETS Δ4 and Δ5 pre-rRNAs, and its restoration in the Δ6 + H pre-rRNA.

## Discussion

Recent years have seen the development of several systems that allow mutations in the yeast pre-rRNA to be studied *in vivo*. The effects of mutations in several of the transcribed spacer regions on pre-rRNA processing have been studied in some detail, although interpretation of the results has frequently been complicated. The spacer which to date has been least well studied in *Saccharomyces cerevisiae* is the 3' ETS. This region of the pre-rRNA is predicted to have a number of features of importance and we have undertaken an initial deletion analysis.

Perhaps the most surprising finding to emerge from this analysis is a negative result; there does not appear to be a signal in the 3' ETS that is important for initiating the processing pathway. The basis of the apparent need for such a signal can be simply stated: Transcription of the 35 S pre-rRNA requires approximately five minutes, so the early processing sites are transcribed several minutes before their eventual processing. Why are these sites not cleaved cotranscriptionally, as is the case in *E. coli*? Moreover, the life time of the fully transcribed 35 S pre-rRNA is very short: from Northern hybridization and pulse-chase labeling we estimate this at approximately ten seconds; and during this time many or all of the ca 60 2'-O-methyl modifications in the pre-rRNA are also made (Brand *et al.*, 1977; Klootwijk *et al.*, 1972). Why are these not made cotranscriptionally? Our anticipation was that the processing pathway would be initiated by the recognition of some signal within the 3' ETS region that would indicate that transcription has been successfully completed. This does not appear to be the case. Even the pre-rRNA from which the entire 3' ETS region has been deleted (3' ETS Δ6) undergoes the early processing reaction with normal kinetics, as shown by the lack of accumulation of the full length 35 S pre-rRNA or the 33 S and 32 S pre-rRNAs, generated by processing at sites A<sub>0</sub> and A<sub>1</sub>. Moreover, an alternative hypothesis, that the termination of transcription by RNA polymerase I itself triggers the processing pathway is also unlikely in this case, as transcription is by RNA polymerase II. How the correct timing of the processing and

modification of the pre-rRNA is achieved therefore remains an enigma.

### The stem-loop structure is required for processing of the 3' ETS

The feature of the 3' ETS that the deletion mutations clearly identify as functionally important is the strong stem-loop structure predicted to form close to the end of the 25 S rRNA (see Figure 2B). Mutations that delete the entire 3' ETS up to the 3' boundary of this stem (3' ETS Δ1-3) have no detectable effect on processing of the pre-rRNA in the 3' ETS or elsewhere. In contrast, deletions which enter this stem-loop (3' ETS Δ4 and Δ5), or remove it entirely (3' ETS Δ6) are severely impaired in processing of the 3' ETS region. In these pre-rRNAs, the 35 S\* transcripts extend to the GAL7 terminator that is inserted 3' to the normal site of termination by RNA polymerase I. These species are not detected in the wild-type pre-rRNA and we conclude that the 3' ETS is normally cleaved cotranscriptionally, in marked contrast to processing at other sites. The stem-loop structure in the 3' ETS can be cleaved *in vitro* by Rnt1p (Abou Elela *et al.*, 1996) which, by analogy to *E. coli* RNase III, is expected to cleave in extended imperfect stem structures. Similar structures are cleaved by Rnt1p in the 5' ETS (Abou Elela *et al.*, 1996) and in precursors to U5 snRNA (Chanfreau *et al.*, 1997). The 3' ETS Δ4 and Δ5 mutations do not remove the site of *in vitro* cleavage but are predicted to disrupt the stem structure, thus abolishing cleavage by Rnt1p. Interestingly, the presence of these long 3' extensions does not prevent processing of the aberrant pre-rRNAs, since we detect the 27 SA\* and 27 SB\* pre-rRNAs, and even the 25 S\* rRNA, that are 3' extended to the GAL7 terminator. The stem-loop structure is both necessary and sufficient for cleavage in the 3' ETS, since its reintroduction into the 3' ETS Δ6 + H pre-rRNA restores normal processing in the 3' ETS.

### Processing in the 3' ETS and ITS1 are coordinated

These analyses also revealed interactions between the 3' ETS and pre-rRNA processing reactions that occur almost 4 kb away in ITS1. Deletions Δ4 and Δ5 in the hairpin specifically inhibited cleavage at site A<sub>3</sub> in ITS1. This was shown by the loss of the primer extension stop at site A<sub>3</sub>, showing the loss of the 27 SA<sub>3</sub> pre-rRNA, and the loss of the excised A<sub>2</sub>-A<sub>3</sub> cleavage fragment. Two forms of the 27 SB pre-rRNA and 5.8 S rRNA are generated. In the major pathway, cleavage at site A<sub>3</sub> is followed by exonuclease trimming to site B1<sub>S</sub>, the 5' end of the 27 SB<sub>S</sub> pre-rRNA and 5.8 S<sub>S</sub> rRNA. A minor pathway generates the 27 SB<sub>L</sub> pre-rRNA and 5.8 S<sub>L</sub> rRNA. Synthesis of both the 27 SB<sub>S</sub> and 27 SB<sub>L</sub> pre-rRNAs from the Δ4 and Δ5 pre-rRNAs was inhibited, resulting in reduced synthesis of the 5.8 S<sub>S</sub> and 5.8 S<sub>L</sub> rRNAs.

However, synthesis of 5.8 S<sub>5</sub> was more severely inhibited than was synthesis of 5.8 S<sub>L</sub>, indicating that A<sub>3</sub> cleavage by RNase MRP was more sensitive to mutations in the 3' ETS stem-loop than was processing at B<sub>1L</sub>.

The effects of mutations in the 3' ETS are distinct from mutations in Rnt1p. In particular, the *rnt1-1* strain accumulates the 23 S RNA, a product of cleavage at site A<sub>3</sub>, and is not clearly defective in the processing of 27 SA<sub>2</sub>\* to 27 SB\* (Abou Elela *et al.*, 1996). These observations indicate that the defects in processing of site A<sub>3</sub> in the 3' ETS Δ4 and Δ5 pre-rRNAs are not a direct consequence of the inhibition of 3' ETS processing. We conclude that the stem-loop structure is recognized by a factor other than Rnt1p, and this recognition is important for processing in ITS1.

The obvious rationale for the existence of such a system is to prevent synthesis of the 5.8 S rRNA and maturation of the 5' end of 25 S rRNA from transcripts which are incomplete due to premature transcription termination. A similar situation has been reported in *Schizosaccharomyces pombe*. Mutations that delete, or interfere with the structure of, a hairpin stem in the 3' ETS inhibited both the removal of the 3' ETS region and production of the 5.8 S and 25 S rRNAs (Hitchen *et al.*, 1997; Melekhovets *et al.*, 1994), suggesting a similar inhibition of steps in ITS1 processing. Indeed, Good *et al.* (1997a, b) have proposed that the potential for such quality control systems is a reason for the existence of pre-rRNA processing in general.

### A model for the coordination of pre-rRNA processing

In both Bacteria and Archaea the 16 S and 23 S rRNA sequences are flanked by spacer sequences which form extensive helices that generate the recognition sites for endonucleases; RNase III in bacteria (King *et al.*, 1984; Robertson & Dunn, 1975; Young & Steitz, 1978) and the bulge-helix-bulge endonuclease in Archaea (reviewed by Dennis, 1997; Garrett *et al.*, 1991). These interactions within the pre-rRNAs ensure the coordination of processing at the 5' and 3' ends of the rRNAs. In *S. cerevisiae*, the hairpin structures in the 5' ETS and 3' ETS each provide cleavage sites for Rnt1p (RNase III) but different mechanisms exist to ensure the coordination of the processing reactions at the opposite ends of the 18 S and 5.8 S/25 S rRNAs. The eukaryotic 5.8 S rRNA is clearly homologous to the 5' end of the bacterial and archaeal 23 S rRNA, indicating that the ITS2 region arose by an insertion event in an early eukaryote. The coupling reported here between processing in ITS1 and processing of the 3' ETS would therefore be functionally analogous to the coupling between processing at the 5' and 3' ends of the bacterial and archaeal 23 S provided by the requirement for base-pairing of the flanking sequences.

In *Xenopus laevis* the snoRNA U8 is required for the synthesis of 5.8 S and 25 S rRNA; depletion of

U8 leads to a phenotype similar to deletions of the 3' ETS stem-loop structure with the inhibition of processing both of the 3' ETS and in ITS1 (Peculis & Steitz, 1993). No yeast homologue of U8 has yet been identified, but we predict that a functional homologue will play a role in the processing analysed here. It is possible that this RNP binds to the hairpin structure in the 3' ETS, promoting cleavage by Rnt1p and also interacts with the RNase MRP complex bound to the pre-rRNA around site A<sub>3</sub> in ITS1.

This model resembles that proposed for the coupling of pre-rRNA processing in the 5' ETS and in ITS1. Cleavage of site A<sub>0</sub> in the 5' ETS by Rnt1p *in vivo* absolutely requires the binding of the U3 snoRNA to the pre-rRNA at a site some 140 nucleotides 5' to the cleavage site. Moreover, mutations in U3 or other snoRNAs (U14, snR10 and snR30) inhibit processing both at the 5' end of the 18 S rRNA and in ITS1 at site A<sub>2</sub>. This has led to the suggestion that a multi-snoRNP complex forms and is required for these coordinated processing reactions. Furthermore, processing at sites A<sub>1</sub> and A<sub>2</sub> is coupled to processing by RNase MRP at site A<sub>3</sub>. This coupling may involve the large Rrp5p protein functioning as a bridging factor (Venema & Tollervey, 1996).

This suggests a model (Figure 7) in which RNase MRP interacts with multiple components of the processing machinery: an snoRNP complex and Rrp5p bound to the 5' ETS and at site A<sub>2</sub> and another complex bound to the pre-rRNA in the 3' ETS, making 35 S pre-rRNA processing a highly coordinated process.

## Materials and Methods

### Strains and media

Growth and handling of *S. cerevisiae* used standard techniques. The strain used was NOY504:  $\alpha$ , *rpa12::LEU2*, *leu2-3*, 112, *ura3-1*, *trp1-1*, *his3-11*, *can1-100* (Nogi *et al.*, 1993; generously provided by M. Nomura).

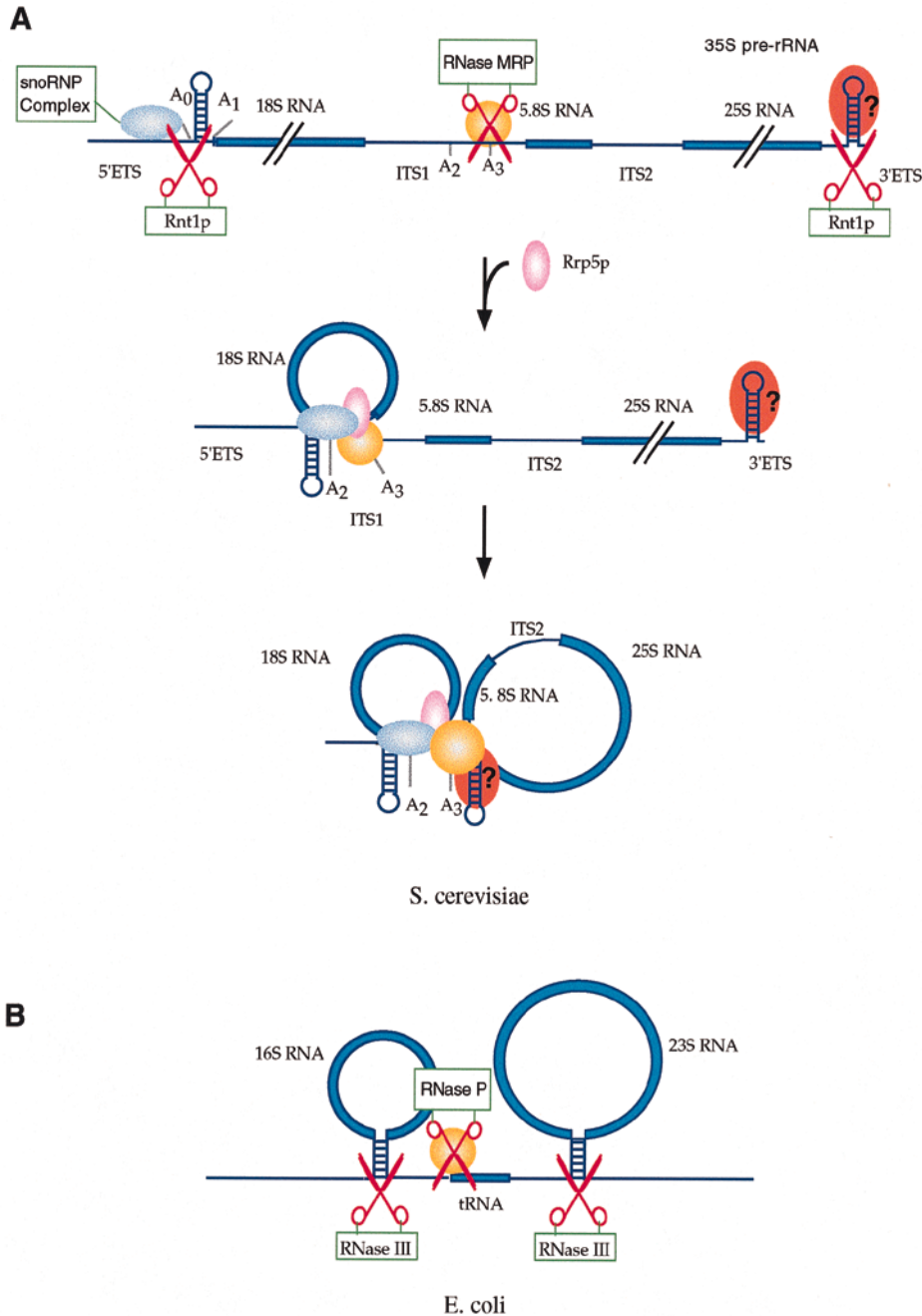
### Construction of nested deletions in the 3' ETS

The wild-type plasmid pTA1 used to generate nested deletions in the 3' ETS was derived from pGAL::rDNA (Henry *et al.*, 1994). This plasmid, contains the entire yeast rDNA unit fused to the GAL7 promoter (Nogi *et al.*, 1993) in YEp24 (2 $\mu$ m-URA3). In addition, small oligonucleotide tags have been inserted in the 18 S, 5.8 S and 25 S rRNA genes. pTA1 was obtained by insertion of the polylinker *NotI*, *BstEII*, *BstXI* (*KpnI*), *SalI* at the *SalI* site located nine nucleotides downstream of the rDNA unit in the GAL7 terminator sequence. 3' ETS nested deletions of the 3' ETS were produced by exonuclease III digestion (Erase-a-Base, Promega). The pTA1 plasmid was digested with *KpnI* and *NotI* and progressive unidirectional deletions were generated from the *NotI* site. The plasmid sequences were protected from digestion by the 3' overhang at the *KpnI* restriction site. The 3' *SalI* site was left intact and now directly flanked the rDNA sequence. The positions of the mutations are indicated in Figure 2, numbering is relative to the 3' end of 25 S



rRNA. In the 3' ETS  $\Delta 1$ ,  $\Delta 2$ ,  $\Delta 3$ ,  $\Delta 4$ ,  $\Delta 5$  and  $\Delta 6$  mutants nucleotides from +284 to +262, +178, +58, +41, +24 and -5, respectively, were deleted. Mutant 3' ETS

$\Delta 6 + H$  was obtained by re-insertion of the hairpin structure into the 3' ETS  $\Delta 6$  deletion mutant at the same nucleotide position as in the wild-type pre-rRNA. For



**Figure 7.** Model for the coordination of pre-rRNA processing in yeast and comparison with *E. coli*. The pre-rRNA processing pathways of *S. cerevisiae* (A) and *E. coli* (B) are represented. Boxes represent mature rRNAs and thin lines the transcribed spacers. Endonucleases are represented by scissors and snoRNP complexes by balls. The basic arrangement of the rRNA is conserved between bacteria and eukaryotes. However the yeast large-subunit rRNA is split by the insertion of the ITS2 spacer and the *E. coli* pre-rRNAs encode tRNAs in spacer between the 16 S and 23 S rRNAs. The spacers flanking the bacterial rRNA molecule form extensive helices, which contain the recognition sites for RNase III and ensure the coordination of processing at the 5' and 3' end of rRNAs. The tRNAs present in the spacer regions provide a site for cleavage by RNase P. In yeast, hairpin structures in the 5' ETS and 3' ETS provide cleavages sites for Rnt1p (RNase III) and small nucleolar ribonucleoprotein particle (snoRNP) complexes replace intramolecular base-pairing to ensure coordinated 5' and 3' end processing. We previously proposed that a snoRNP complex bound in the 5' ETS interacts with the RNase MRP complex that cleaves site A<sub>3</sub> in ITS1, and this coupling may involve the Rrp5p protein. Here we propose a coupling between processing in ITS1 and in the 3' ETS. A functional homologue of the U8 snoRNA is predicted to be involved in the coupling cleavage of the 3' ETS hairpin structure by Rnt1p and ITS1 cleavage by the RNase MRP complex.

this, the 3' ETS  $\Delta 6 + H$  plasmid was linearized at the *Sall* site and nucleotides 5'-TCGATTTTTATTTCTTTCT-AAGTGGGTACTGGCAGGAGTCGGGGCCTAGTTA-GAGAGAG-3' were inserted, resulting in the mutation of the last six nucleotides <sup>3387</sup>AUUUGU<sub>3392</sub> of 25 S rRNA to <sup>3387</sup>GGUCGA<sub>3392</sub> and restoration of the wild-type 3' ETS sequence and hairpin structure up to nucleotides +58 (Figure 2B).

### Growth rate measurement

For growth rate measurements cells were first grown in minimal galactose medium at 23°C (the permissive temperature for chromosomal rDNA transcription in *rpa12* strains) to an  $A_{600\text{ nm}}$  of 0.07. Cells were then shifted to the non-permissive temperature (37°C) to repress chromosomal rDNA transcription and selectively express the plasmid-encoded, mutant rDNA. Regular dilution of cells with pre-warmed medium was performed in order to maintain exponential growth and the  $A_{600\text{ nm}}$  was followed for 50 hours.

### RNA extraction

Prior to RNA extraction for Northern analysis or primer extension, NOY504 strains were transformed with pGAL::rDNA, the 3' ETS mutant plasmids or a negative control plasmid (-rDNA) YEplac 195 (2 $\mu$ m-URA3; Gietz & Sugino, 1988); and grown at 23°C in minimal galactose medium until they reached mid log phase. Cells were diluted to  $A_{600\text{ nm}}$  0.09 and shifted for six hours to 37°C (Henry *et al.*, 1994). Total RNA was extracted as previously described (Tollervey & Mattaj, 1987).

### Northern Hybridization

For each sample 8  $\mu$ g of total RNA was separated on 1.2% (w/v) agarose-formaldehyde gels and transferred to Hybond N<sup>+</sup> membranes (Amersham) for Northern hybridization as described by Tollervey & Mattaj (1987). Northern hybridization was performed as previously described using the following oligonucleotides. (a) 5'-CGAGGATCCAGGCTTT-3'; (b) 5'-GCTCTTTGCTCTT GCC-3'; (c) 5'-TGTTACCTCTGGGCC-3'; (d) 5'-CCAGTTACGAAAATTCTTG-3'; (g) 5'-GGCCAGCAATTTCAAGT-3'; (h) 5'-ACTCGAGAGCTTCAGTAC-3'; (i) 5'-AAGAATCAGATTTACAGATAAATGATGT-CATT-3'. Oligonucleotides a, and h hybridize to the tags in 18 S and 25 S rRNAs, respectively. Oligo i hybridizes to the *GAL7* terminator sequence, 47 nucleotides downstream of the *Sall* site (see Figure 1).

Low molecular weight RNAs were separated on 8% (w/v) polyacrylamide gels containing 8 M urea in 1  $\times$  TBE and electroblotted onto Hybond N<sup>+</sup> membranes (Amersham). Hybridization to the tag in 5.8 S rRNA was performed using an 2'-O-allyl oligonucleotide (f) (Lamond & Sproat, 1993). The sequence is 5'-DGDDUD-CUGGCGDdGdC-3' (Henry *et al.*, 1994), where D is 2,6-diamino-purine, which can form three hydrogen bonds to U residues (Lamm *et al.*, 1991).

### Primer extension

Primer extension was performed as described previously (Beltrame & Tollervey, 1992) on 4  $\mu$ g of total RNA using primer (d) 5'-CCAGTTACGAAAATTCTTG-3'. To identify the position of primer extension stops,

DNA sequencing reactions performed with the same oligonucleotide were run in parallel.

## Acknowledgements

We thank Dr M. Nomura (university of California, Irvine) for the *rpa12* strain and Phil Mitchell and Emanuelle Pascolo for critical reading of the manuscript. C.A. was the recipient of a grant from the E.U. This work was partly supported by the Wellcome Trust.

## References

- Abou, Elela S., Igel, H. & Ares, M. J. (1996). RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell*, **85**, 115–124.
- Allmang, C., Henry, Y., Morrissey, J. P., Wood, H., Petfalski, E. & Tollervey, D. (1996). Processing of yeast pre-rRNA at sites A<sub>2</sub> and A<sub>3</sub> is linked. *RNA*, **2**, 63–73.
- Amberg, D. C., Goldstein, A. L. & Cole, C. N. (1992). Isolation and characterization of RAT1: an essential gene of *Saccharomyces cerevisiae* required for the efficient nucleocytoplasmic trafficking of mRNA. *Genes Dev.* **6**, 1173–1189.
- Beltrame, M. & Tollervey, D. (1992). Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J.* **11**, 1531–1542.
- Brand, R. C., Klootwijk, J., van Steenberg, T. J. M., de Kok, A. J. & Planta, R. J. (1977). Secondary methylation of yeast ribosomal precursor RNA. *Eur. J. Biochem.* **75**, 311–318.
- Chanfreau, G., Abou, Elela S., Ares, M., Jr & Guthrie, C. (1997). Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev.* **11** (27), 2741–2751.
- Chu, S., Archer, R. H., Zengel, J. M. & Lindahl, L. (1994). The RNA of RNase MRP is required for normal processing of ribosomal RNA. *Proc. Natl Acad. Sci. USA*, **91**, 659–663.
- Dennis, P. P. (1997). Ancient ciphers: translation in Archaea. *Cell*, **89** (7), 1007–10.
- Garrett, R. A., Dalgaard, J., Larsen, N., Kjems, J. & Mankin, A. S. (1991). Archaeal rRNA operons. *Trends Biochem. Sci.* **16** (1), 22–6.
- Gietz, R. D. & Sugino, A. (1988). New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene*, **74**, 527–534.
- Good, L., Intine, R. V. & Nazar, R. N. (1997a). Interdependence in the processing of ribosomal RNAs in *Schizosaccharomyces pombe*. *J. Mol. Biol.* **273** (4), 782–8.
- Good, L., Intine, R. V. & Nazar, R. N. (1997b). The ribosomal-RNA-processing pathway in *Schizosaccharomyces pombe*. *Eur. J. Biochem.* **247**, 314–321.
- Gutell, R. R. & Fox, G. E. (1988). A compilation of large subunit RNA sequences presented in a structural format. *Nucl. Acids Res.* **16**(suppl), r175–r269.
- Henry, Y., Wood, H., Morrissey, J. P., Petfalski, E., Kearsley, S. & Tollervey, D. (1994). The 5' end of yeast 5.8 S rRNA is generated by exonucleases from an upstream cleavage site. *EMBO J.* **13**, 2452–2463.
- Hitchen, J., Ivakine, E., Melekhovets, Y. F., Lalev, A. & Nazar, R. N. (1997). Structural features in the 3' external transcribed spacer affecting intragenic processing of yeast rRNA. *J. Mol. Biol.* **274** (4), 481–490.
- Kempers-Veenstra, A. E., Oliemans, J., Offenber, H., Dekker, A. F., Piper, P. W., Planta, R. J. &

- Klootwijk, J. (1986). 3'-End formation of transcripts from the yeast rRNA operon. *EMBO J.* **5**, 2703–2710.
- Kenna, M., Stevens, A., McCammon, M. & Douglas, M. G. (1993). An essential yeast gene with homology to the exonuclease-encoding *XRN1/KEM1* gene also encodes a protein with exoribonuclease activity. *Mol. Cell. Biol.* **13**, 341–350.
- King, T. C., Sirdeshmukh, R. & Schlessinger, D. (1984). RNase III cleavage is obligate for maturation but not for function of *Escherichia coli* pre-23 S rRNA. *Proc. Natl Acad. Sci. U S A*, **81** (1), 185–188.
- Klootwijk, J., van der Bos, R. C. & Planta, R. J. (1972). Secondary methylation of yeast ribosomal RNA. *FEBS letters*, **27**, 102–106.
- Lamm, G. M., Blencowe, B. J., Sproat, B. S., Iribarren, A. M., Ryder, U. & Lamond, A. I. (1991). Antisense probes containing 2-aminoadenosine allow efficient depletion of U5 snRNP from HeLa splicing extracts. *Nucl. Acids Res.* **19**, 3193–3198.
- Lamond, A. I. & Sproat, B. S. (1993). Antisense oligonucleotides made of 2'-O-alkyl RNA: their properties and applications in RNA biochemistry. *FEBS Letters*, **325**, 123–127.
- Larimer, F. W., Hsu, C. L., Maupin, M. K. & Stevens, A. (1992). Characterization of the *XRN1* gene encoding a 5' → 3' exoribonuclease: sequence data and analysis of disparate protein and mRNA levels of gene-disrupted yeast cells. *Gene*, **120**, 51–57.
- Lygerou, Z., Mitchell, P., Petfalski, E., Séraphin, B. & Tollervey, D. (1994). The *POP1* gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. *Genes Dev.* **8**, 1423–1433.
- Lygerou, Z., Allmang, C., Tollervey, D. & Séraphin, B. (1996). Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science*, **222**, 268–270.
- Maxwell, E. S. & Fournier, M. J. (1995). The small nucleolar RNAs. *Annu. Rev. Biochem.* **35**, 897–934.
- Melekhovets, Y. F., Good, L., Abou, Elela S. & Nazar, R. N. (1994). Intragenic processing in yeast rRNA is dependent on the 3' external transcribed spacer. *J. Mol. Biol.* **239**, 170–180.
- Morrissey, J. P. & Tollervey, D. (1995). Birth of the snoRNPs - the evolution of RNase MRP and the eukaryotic pre-rRNA processing system. *Trends Biochem. Sci.* **20**, 78–82.
- Nogi, Y., Yano, R. & Nomura, M. (1991). Synthesis of large rRNAs by RNA polymerase II in mutants defective in RNA polymerase I. *Proc. Natl Acad. Sci. (USA)*, **88**, 3962–3966.
- Nogi, Y., Yano, R., Dodd, J., Carles, C. & Nomura, M. (1993). Gene *RRN4* in *Saccharomyces cerevisiae* encodes the A12. 2 subunit of RNA polymerase I and is essential only at high temperatures. *Mol. Cell. Biol.* **13**, 114–122.
- Peculis, B. A. & Steitz, J. A. (1993). Disruption of U8 nucleolar snRNA inhibits 5. 8 S and 28 S rRNA processing in the *Xenopus* oocyte. *Cell*, **73**, 1233–1245.
- Piper, P. W., Bellatin, J. A. & Lockheart, A. (1983). Altered maturation of sequences at the 3' terminus of 5 S gene transcripts in *Saccharomyces cerevisiae*. *EMBO J.* **2**, 353–359.
- Robertson, H. D. & Dunn, J. J. (1975). Ribonucleic acid processing activity of *Escherichia coli* ribonuclease III. *J Biol Chem*, **250** (8), 3050–3056.
- Schmitt, M. E. & Clayton, D. A. (1993). Nuclear RNase MRP is required for correct processing of pre-5. 8 S rRNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**, 7935–7941.
- Stevens, A. & Poole, T. L. (1995). 5'-exonuclease-2 of *Saccharomyces cerevisiae*. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1. *J. Biol. Chem.* **270** (27), 16063–16069.
- Tollervey, D. & Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.* **9**, 337–342.
- Tollervey, D. & Mattaj, I. W. (1987). Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J.* **6**, 469–476.
- Veldman, G. M., Klootwijk, J., De Jong, P., Leer, R. J. & Planta, R. J. (1980). The transcription termination site of the ribosomal RNA operon in yeast. *Nucl. Acids Res.* **8**, 5179–5192.
- Venema, J. & Tollervey, D. (1995). Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*. *Yeast*, **11**, 1629–1650.
- Venema, J. & Tollervey, D. (1996). RRP5 is required for the formation of both 18 S and 5. 8 S rRNA in yeast. *EMBO J.* **15**, 5701–5714.
- Yip, M. T. & Holland, M. J. (1989). *In vitro* RNA processing generates mature 3' termini of yeast 35 S and 25 S ribosomal RNAs. *J. Biol. Chem.* **264**, 4045–4051.
- Young, R. A. & Steitz, J. A. (1978). Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. *Proc. Natl. Acad. Sci. U S A*, **75** (8), 3593–3597.

*Edited by M. Yaniv*

(Received 7 November 1997; received in revised form 2 February 1998; accepted 4 February 1998)

---

# Role of conserved nucleotides in building the 16S rRNA binding site of *E.coli* ribosomal protein S8

---

Christine Allmang, Marylène Mougel, Eric Westhof, Bernard Ehresmann and Chantal Ehresmann\*

UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg cedex, France

---

Received June 6, 1994; Revised and Accepted July 29, 1994

---

## ABSTRACT

**Ribosomal protein S8 specifically recognizes a helical and irregular region of 16S rRNA that is highly evolutionary constrained. Despite its restricted size, the precise conformation of this region remains a question of debate. Here, we used chemical probing to analyze the structural consequences of mutations in this RNA region. These data, combined with computer modelling and previously published data on protein binding were used to investigate the conformation of the RNA binding site. The experimental data confirm the model in which adenines A595, A640 and A642 bulge out in the deep groove. In addition to the already proposed non canonical U598 – U641 interaction, the structure is stabilized by stacking interactions (between A595 and A640) and an array of hydrogen bonds involving bases and the sugar phosphate backbone. Mutations that alter the ability to form these interdependent interactions result in a local destabilization or reorganization. The specificity of recognition by protein S8 is provided by the irregular and distorted backbone and the two bulged adenines 640 and 642 in the deep groove. The third adenine (A595) is not a direct recognition site but must adopt a bulged position. The U598 – U641 pair should not be directly in contact with the protein.**

## INTRODUCTION

The interaction of *E.coli* ribosomal protein S8 with its 16S rRNA binding site represents an interesting model for studying the molecular mechanism of specific RNA–protein recognition. Protein S8 is capable of binding individually to the central domain of 16S rRNA and plays an important role in the early stage of ribosomal 30S subunit assembly (1–2). It participates to the formation of one early nucleation site (3), and interacts co-operatively with other ribosomal proteins (4–5). It is therefore a crucial element for the sequential assembly of RNA and proteins constituting the small ribosomal subunit. It is also able to regulate the translation of its own operon (6–8) by a feed-back mechanism.

A considerable amount of work was already devoted to the interactions between S8 and its 16S rRNA target site and to the fine structure of this site (4–5, 9–14). It was recently shown that the rRNA can be restricted to a short helical stem (nucleotides 588–605/633–651), without significantly altering the apparent affinity constant (15). The central part of this helical region (called ‘region C’) is highly evolutionary constrained and the conserved elements are also found in the target regulatory site of S8 on its mRNA (8,16). We previously proposed a three-dimensional model of region C, derived from structure probing and computer modeling (14). This model displays characteristic features: A595, A640 and A642 bulge out in the deep groove of the helix, and U598 and U641 form a non-canonical base pair. However, the conformation of this region is disputed and three other folding models have been proposed in the literature. These models essentially differ in the pairing of U598 which is either with A640 (17–18), U641 (14) or A642 (5). We favoured a U595–U641 base pair (14), since it accounts for the non reactivity of U598 and U641 and for the reactivity of A640, A642 and A595. The pair U598–A640 was recently proposed on the basis of sequence comparison (17–18). In order to agree with the reactivity data, such a U598–A640 pair should involve Hoogsteen hydrogen bonding and not Watson–Crick interactions. In addition, the non reactivity of the unpaired U641 could only be explained by additional tertiary interaction or stacking.

Recently, we investigated the role of conserved nucleotides in region C as potential determinants for S8 recognition by studying the effect of 14 single and double mutations on S8 recognition (15). Of the 14 mutants tested, only three are still efficiently recognized by S8. In order to discriminate whether the loss of recognition is due to the loss of a specific contact or to conformational rearrangement, we now report the structural consequences of the mutations, using chemical probing on the 14 RNA variants mentioned above and of two new RNA mutants (A598/U640 and A598/U640/G641). In addition, footprinting experiments were conducted on those mutants that still retain S8 binding capacity. Our results emphasize the subtleties of RNA conformation and an unexpected versatility in the structural consequences of single base mutations. An improved three-

---

\*To whom correspondence should be addressed

dimensional model is derived from the present experimental data and the results are discussed in terms of RNA folding and S8 recognition.

## MATERIALS AND METHODS

### Preparation of the biological material

Plasmids construction, RNA synthesis and purification of wild-type and mutant 16S rRNA fragments (nucleotides 584–756) are described in (15). Two additional mutants were constructed (A598/U640 and A598/U640/G641) following the same protocol. Their relative binding affinity was determined as in (15). Ribosomal protein S8 was prepared under non-denaturing conditions according to Cachia *et al.* (19).

### Chemical probing and footprinting

A standard assay contains 16 pmol RNA and 2  $\mu$ g carrier tRNA in 20  $\mu$ l of appropriate buffer. RNA was first pre-incubated for 15 min at 40°C in buffer N1 [50 mM sodium cacodylate (pH 7.5), 20 mM magnesium acetate, 250 mM potassium acetate] or N2 [50 mM sodium borate (pH 8.0); 20 mM magnesium acetate, 250 mM potassium acetate]. For each reaction, a control was treated in parallel, omitting the reagent. Modification with DMS: incubation was for 5 and 10 min in buffer N1 or for 2 and 5 min in buffer D1 [50 mM sodium cacodylate (pH 7.5), 1 mM EDTA] for semi-denaturing conditions. Modifications with CMCT: incubation was for 15 and 30 min in buffer N2 or for 2 and 5 min in buffer D2 [50 mM sodium borate (pH 8.0), 1 mM EDTA] for semi-denaturing conditions. Modifications with DEPC: incubation was for 15, 30 and 60 min in buffer N1 or for 15 and 30 min in buffer D1 (semi-denaturing conditions). All modifications were at 37°C. Footprinting experiments using CMCT and DMS were conducted on wild-type RNA and mutants allowing S8 binding. Complexes were formed in the presence of 0.4  $\mu$ M S8 for wild-type RNA, mutants U595 and A641, or 2  $\mu$ M for mutant A598–U640. Footprinting gels were scanned using the Bio-Imager Analyzer BAS 2000 (Fuji). Synthesis of primer, labeling, hybridization, reverse transcription and analysis of generated cDNA fragments were described by Mougél *et al.* (14).

### Computer modeling

The modeled molecule integrating stereochemical constraints and experimental data was constructed with the help of several computer programs and tested by comparing the theoretical accessibility of atoms with the observed experimental reactivity, as described earlier (20).

## RESULTS

### Binding strength of the new mutants

Previous results showed that both mutants A598 and U642 fail to recognize S8 (15). Here, we tested the possibility to restore S8 binding by the double mutation A598/U642. The results (not shown) show that this double mutation restores only partially S8 recognition (with a 5-fold reduced binding strength). Sequence comparison indicates that U598 is highly conserved. However, in *Rcy purpur*, nucleotide 598 is an adenine, and nucleotides 640 and 641 are simultaneously replaced by U and G, respectively. Therefore, we constructed a new mutant containing these three mutations (A598/U640/G641). This triple mutant is not recognized by S8 (results not shown).

### Conformational studies of the RNA variants

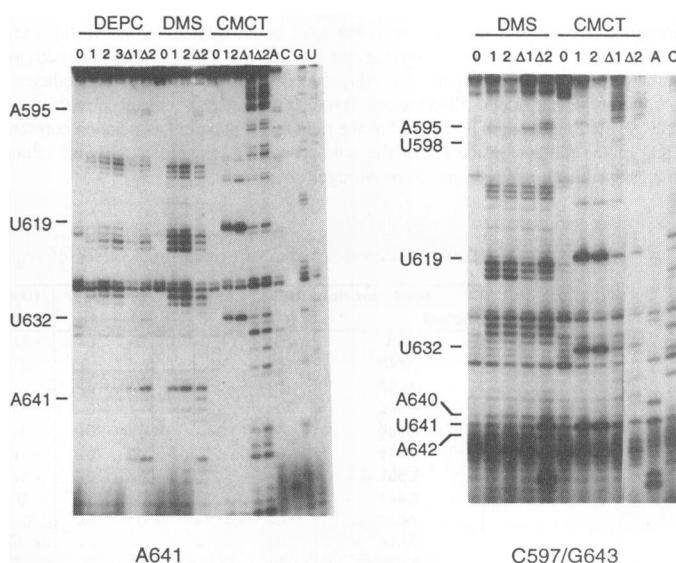
The four bases were tested for their chemical reactivity at one of their Watson–Crick positions with DMS, at A(N1) and C(N3), and with CMCT, at G(N1) and U(N3). For some mutants, position N7 of adenines was also probed with DEPC. In addition, footprinting experiments were conducted using DMS and CMCT with those RNAs that still retain S8 binding ability. A typical experiment is shown in Fig. 1. Experiments were repeated several times (from 2 to 4 times) and the degree of reactivity was evaluated from 1 to 4 by visual inspection. In the case of footprinting experiments, reactivity changes induced by S8 binding were quantified.

The reactivity changes induced by the mutations are exclusively localized in region C (nucleotides 594–599/639–645). Results are summarized in Table 1 and in Figs 2–4 which show the deduced secondary foldings of region C. One striking consequence of all the mutations tested is that U641, which is not reactive in the wild-type RNA, becomes reactive at various degrees in all mutated RNAs, with the single exception of mutant G643 (Table 1). By contrast, U598 remains unreactive in all mutants, suggesting that its N3 position is involved in H-bonding or that the residue is stacked inside the helix, preventing modification.

## DISCUSSION

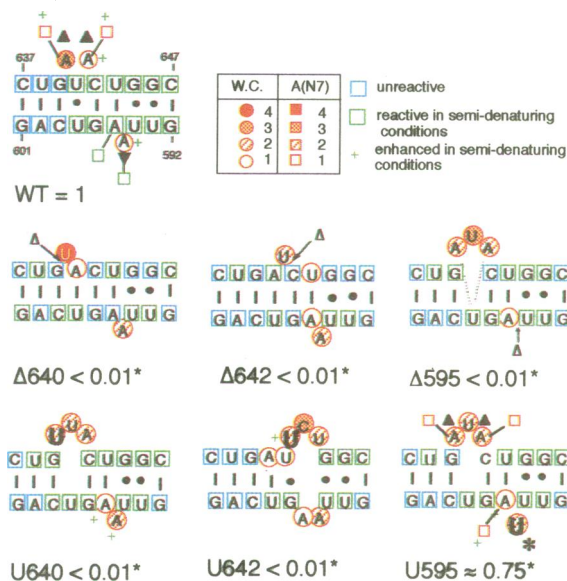
### Mutations affecting adenines 595, 640 and 642

The deletion of any of these three adenines results in a complete loss of binding (15). The deletion of either A640 or A642 induces reactivity at U641 and decreases the reactivity of A642(N1) or A640(N1), respectively (Table 1). These results suggest that nucleotide U641 is bulging out in these two mutants and that U598 pairs with either A642 or A640, respectively (Fig. 2). Moreover,



**Figure 1.** Probing on RNAs A641 and C597/G643. Modification with DMS and CMCT: Incubation control (lane 0). Native conditions: incubation with DMS or CMCT was for 5 min (lane 1) and 10 min (lane 2); semi-denaturing conditions: incubation with CMCT was for 5 min (lane Δ1) and 10 min (lane Δ2). Modification with DEPC: incubation control (lane 0); native conditions: incubation for 15 min (lane 1) and 30 min (lane 2); semi-denaturing conditions: incubation for 5 min (lane Δ1) and 10 min (lane Δ2).

the reactivity of A595(N1) increases by 2-fold, indicating that A595 is not simply bulged out in the wild-type RNA (as already hinted at by its low reactivity) but is probably involved in hydrogen-bonding or stacking, interactions which are disrupted in each deletion mutant. Thus, the observed lack of binding of protein S8 may be due to the loss of a possible contact and/or to a local structural rearrangement of region C. Unexpectedly, the deletion of A595 induces a high level of reactivity at U641 (level 3), and a 2-fold increase in the reactivity of A642 (Fig. 2), showing that the removal of the bulged A595 destabilizes the



**Figure 2.** Proposed secondary structure of mutants affecting adenines 595, 640 and 642. The wild-type RNA is shown as reference. Only nucleotides 592–601/637–647 are shown. The reactivities (estimated between 1 and 4 from marginal to high) are indicated with the color code. Substituted nucleotides are indicated in bold characters and deletions ( $\Delta$ ). Nucleotides without reactivity indicated are not determined. The S8-induced reactivity changes are indicated for wild-type and U595: filled triangle (protection); asterisk (enhanced reactivity). The S8 binding strength [expressed as the ratio of the apparent association constant ( $K_a$ ) of the mutant on the  $K_a$  of the wild-type RNA] are indicated. The values marked with an asterisk are from Mougel *et al.* (15).

interactions which involve U641. The non-reactivity of U598 suggests that it remains stacked inside the helix, either unpaired or alternatively paired with A640 or A642.

The A to U substitution at position 642 causes the disruption of the G597–C643 pair since C643 becomes highly reactive at N3 (level 3). The reactivity pattern favors the existence of two base pairs, U598–A640 and G597–U641, while nucleotides A595, A596, U642, C643 and U644 form an interior asymmetric loop (Fig. 2). Thus, the loss of binding induced by the U642 mutation results from a refolding of region C. In mutant U640, U641 becomes reactive (level 2) but less than in mutant  $\Delta$ A595 (level 3). Therefore, the interaction involving U641 might be weakened but not completely abolished. Another consequence of the A640 substitution is the 2-fold increase in reactivity of A595(N1), as already observed in mutants  $\Delta$ 640 and  $\Delta$ 642. Since the deletion of A595 has also a distal effect on U641 and A642, a structural interdependence between A595, U598, A640 and A642 can be inferred.

Mutant U595 requires a particular attention since it is still recognized by S8 with the same affinity as the wild-type RNA (15). Its reactivity pattern is rather similar to that of mutant U640 (Table 1). However, U641 becomes reactive (level 2), revealing an unexpected distal effect induced by the mutation. The fact that mutant U595, but not mutant U640, is recognized by S8 suggests that A640 is a specific determinant for S8, and that a bulged nucleotide, but not necessarily an adenine, is required at position 595. Most likely, this bulged nucleotide or the particular distortion of the backbone induced by this bulge, is necessary for a correct RNA fold. Since both U595 and U641 are reactive in this mutant (level 2), it was interesting to test their reactivity in the S8–RNA complex. The footprinting experiments show that A640 and A642 become unreactive as in the case of the wild-type RNA. However, U641 displays the same level of reactivity as in the naked RNA and the reactivity of U595 is even increased by a factor of 2 (not shown). This observation confirms that nucleotide 595 is not a specific contact but is required as a bulge. Note that U641 remains unreactive in the wild-type RNA–S8 complex.

#### Possible interactions involving U598

In the different models proposed so far, U598 is paired with either A640, U641 or A642. The U598–U641 pair was tested by

**Table 1.** Reactivity data of critical nucleotides of region C in wild-type and mutant RNAs

Nucl. position	U594	A595	A596	G597	U598	C599	G639	A640	U641	A642	C643	U644	G645
Mutant	(N3)	(N1,N7)	(N1,N7)	(N1)	(N3)	(N3)	(N1)	(N1,N7)	(N3)	(N1,N7)	(N3)	(N3)	(N1)
WT	0+	1+, 0+	0+, 0+	0+	0+	0	0	2, 1+	0+	1+, 1+	0+	0+	0+
$\Delta$ 640	0+	2	0	0	0+	0+	0	$\Delta$	4	1	0	0+	0
$\Delta$ 642	0+	2	1	0	0+	0	0	0+	2	$\Delta$	0	1	0
$\Delta$ 595	0+	$\Delta$	1	0+	0+	0	0+	2	3	2	0+	0+	0+
U640	0+	2+	1+	0+	0+	0+	0+	2	2	2	0+	0+	0+
U642	0+	2	1	0+	0+	0	0	1	1+	2	3	2	0+
U595	0+	2	1+, 1+	0+	0+	0	0	2, 1	2	2, 1	nd	0+	0+
C641	0+	2+, 2+	1+, 0+	0	0	0	0	5, 0	nd	1+, 1+	0+	1	0+
A641	0+	0+, 1+	0, 0	0+	0+	0	0+	3, 1+	0+, 0	2, 1	0+	0+	0+
A598	0+	0+, 2+	0+, 0+	0+	2, 0+	0	0	2, 2	2	2, 2	0	0+	0+
A598/U642	0+	1+, 3	0+, 0+	0	2, 0+	0+	0+	2, 2	4	3	0	0+	0+
G643	1	4	2	0+	0+	0	0	0	0+	2	nd	3	0+
C597/G643	1+	2+	1	0+	0+	0	0	2	3	2	0	0+	0+
G599	0+	1+, 4	0	0+	0+	0+	1+	3, 3	2	2, 2+	0	0+	0+
G599/C639	0+	1+	0+	0+	0+	0	0	2	1+	2+	0	0+	0+
A598/U640	0+	1+	1+	0	2	nd	0	0	1+	2	1	0+	0+
A598/U640/G641	0	0, 1+	0, 1+	0	0+, 1+	0	0	1, 3	0	3, 3	0+	0+	0+

The degree of reactivity of U(N3), G(N1), C(N3), A(N1) (first number) and A(N7) (second number) is estimated from 1 to 4, as in Fig. 3 and 7–9. Degree 5 corresponds to an hyperreactivity (boxed). Reactivity of mutated nucleotides is shadowed. (+) denotes reactivity or increase of reactivity in semi-denaturing conditions; (nd) is not determined.

substituting U641 either by C (preventing the formation of the U–U pair but not the pairing between U598 with either A640 or A642), or by A (allowing the formation of a potential canonical U–A pair). Unexpectedly, the C641 mutation results in hyperactivity of A640(N1) (with N7 unreactive), while U598 remains unreactive under native conditions (Table 1). Unfortunately, a pause of reverse transcriptase masks the mutated C641. Otherwise, the reactivity of A595 is enhanced 2-fold at both N1 and N7. Although we have no obvious explanation for the hyperactivity of A640, this results precludes the formation of a canonical U598–A640 pair in this mutant (but not a Hoogsteen pair). On the other hand, in mutant A641 the mutated adenine is unreactive at both N1 and N7, while A640 and A642 are reactive (with A640 > A642) (Fig. 1). This result indicates that U598 does form a canonical pair with the mutated A641 but not with A640 or A642, although the three adenines are potential candidates for pairing (Fig. 3). Furthermore, this mutant is perfectly recognized by S8. Moreover, A640 and A642 are clearly protected from DMS modification in the presence of S8 (not shown). On the other hand, the substitution of U598 by A leads to a loss of S8 binding (15). Interestingly, probing indicates that A598 and U641 do not form a stable inverted pair, as shown by the reactivity of these two nucleotides (level 2) (Fig. 3). It is puzzling that the U598–A641 pair can be formed, while the inverted A598–U641 cannot. This results supports the existence of an unusual U598–U641 pair, however.

The other two alternatives imply the formation of a base pair between U598 with either A640 or A642. We first showed that substituting A640 or A642 by U leads to a loss of binding and induces local rearrangements (see below). However, the double A598/U642 substitution does not restore binding (15). In fact, in this double mutant, U641 and U642 are both highly reactive (with U641 > U642), indicating that the mutated A598 which is also reactive (level 2) does not pair with any of the two potential candidates (U641 or U642). Most likely, residues 598 and 640–642 are unpaired and form an interior open loop (Fig. 3). Furthermore, the reactivity of A595 remains unchanged at N1 but becomes reactive at N7 (level 3). In this study, we showed that the double A598/U640 substitution restores only partially

S8 recognition. Probing experiments indicate that A598 is reactive (level 2) and U640 unreactive. Strikingly, U641 is only marginally reactive (2-fold less than in RNA A598), but becomes more reactive in semi-denaturing conditions. Thus, there is no clear evidence for a pairing of A598 with either U640 or U641 (Fig. 3). We also showed that the triple mutant A598/U640/G641 is not recognized by S8. Probing experiments indicate that only A642 is highly reactive at both N1 and N7 (level 3). Again, it is not clear from probing data whether A598 interacts with U640 or G641.

Our results also points out the limits of nucleotide sequence comparison while ignoring amino acid sequence co-variations in the corresponding protein. Indeed, a phylogenetic analysis of a subset of protein L23/25 and their putative respective rRNA binding sites clearly evidenced the existence of co-variations in both RNA and protein (21). Therefore, none of the postulated pairs involving U598 can be strictly proven by the classical disruption/inversion method. The only positive mutant (A641)

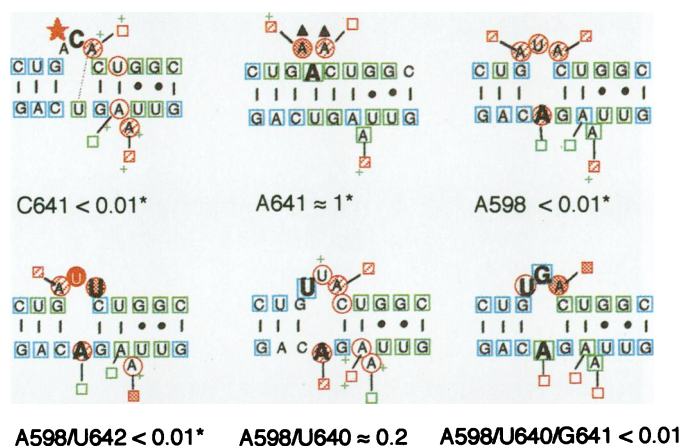


Figure 3. Proposed secondary structure of mutants affecting the possible interactions involving U598. Same legend as in Fig. 2. The red star denotes an hyperactivity.

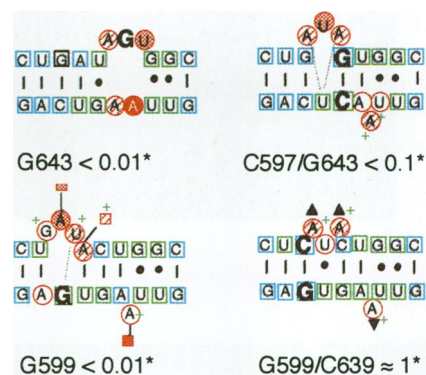


Figure 4. Proposed secondary structure of mutants affecting the G–C pairs. Same legend as in Fig. 2. The S8-induced protections from DMS are indicated for mutant G599/C639 (in the presence of 2 μM of S8).

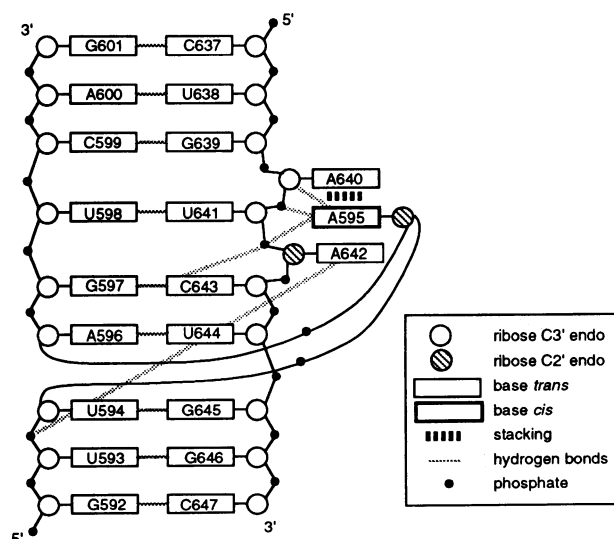
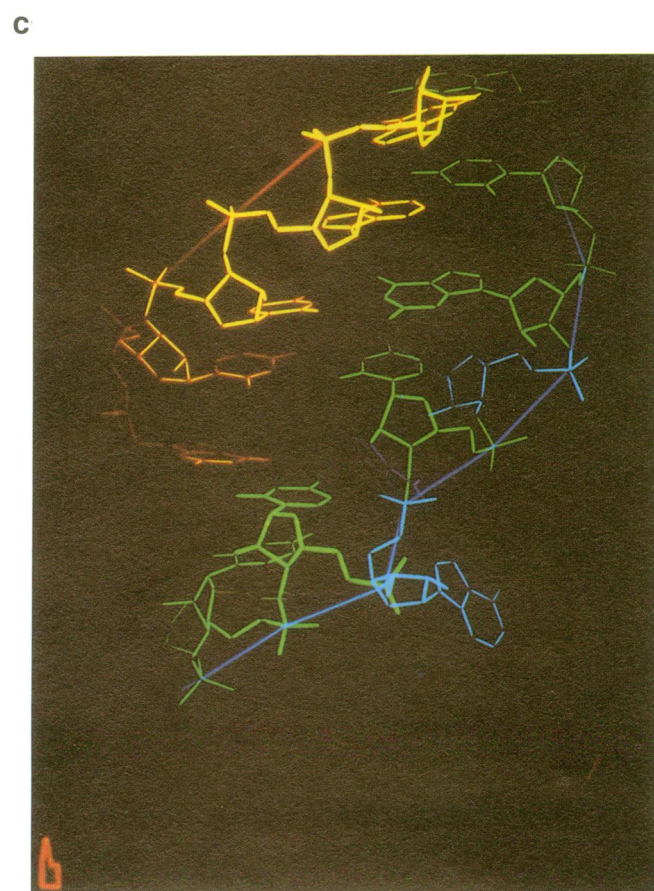
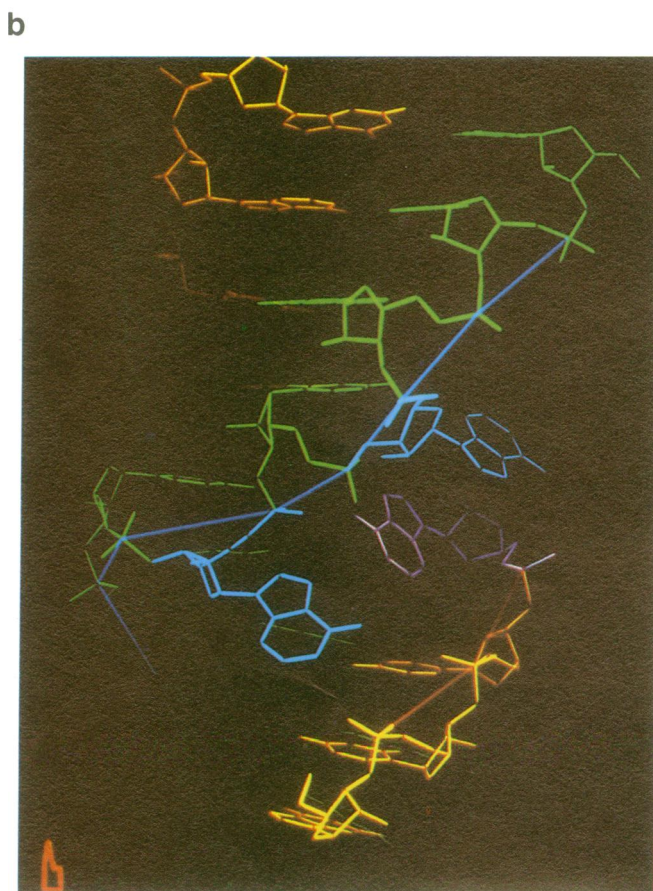
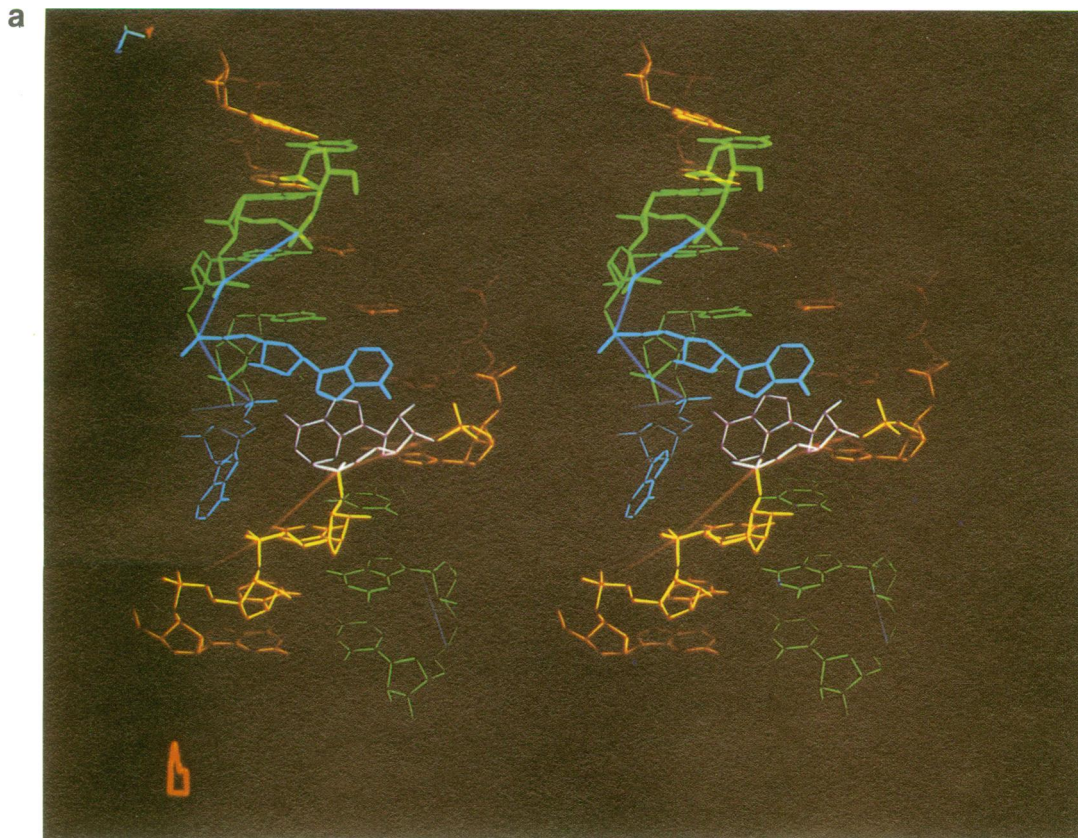


Figure 5. Schematic diagram of the proposed conformation of region C.





favors the existence of the U598–U641 pair, since the formation of a U598–A641 pair was clearly evidenced. One should note that U641 is frequently substituted by A in 16S-like rRNAs.

### Mutations affecting the G–C pairs

From our model, the G599 and G643 substitutions, which both lead to a complete loss of S8 binding (15), are expected to disrupt the two G–C pairs surrounding the U–U pair and most likely to have a destabilizing effect. Indeed, reactivity data indicate that these mutations induce conformational rearrangements. As a result of the G643 mutation, A640(N1) becomes unreactive while the reactivity of A595(N1) is enhanced by a factor of 4 (Table 1). In addition, U644 and A596(N1) become reactive (level 3 and 2, respectively). These data suggest the formation of pairs U598–A640 and G597–U641 (as in mutant U642). Unexpectedly, the potential A595–U644 pair is not formed and nucleotides 595, 596 and 642–644 form a five-base internal loop (Fig. 4). The G599 mutation induces a 2-fold increase of reactivity at A640 (both N1 and N7 positions). It also induces new reactivity at U641 (level 2) and at A595(N7) (level 4). Most likely, helix III is extended by the two meta-stable U598–A642 and G599–U641 pairs (Fig. 4).

The double mutation G599–C639 restores S8 binding (15) and gives a reactivity pattern similar to that of the wild-type RNA (Table 1). The only difference with the wild-type RNA is a marginal reactivity of U641 (level 1) in native conditions, and a 2-fold increase of the reactivity of A640(N1). Therefore, the C599–G639 base pair can be inverted without significant functional and structural effect (Fig. 4). On the contrary, the double mutation C597/G643 does not restore binding (15). Probing experiments show that the mutations cause a strong reactivity of U641 (level 3) and a 2-fold increased reactivity of A595 and A640 at N1 (Fig. 1). The high reactivity at U641 could be explained by the possible loss of interactions with U598 as a consequence of the mutations or to the alteration of a network of interactions involving other nucleotides like A595. Thus, an inverted C597/G643 pair is formed but it is not structurally equivalent to the wild-type one. However, specific contacts between S8 and this G–C pair cannot be excluded. Interestingly, the G597–C643 pair is strictly conserved. Note that a C to U transition and a single deletion at position 643, both produce over 50-fold reduction in S8 affinity and confer slow growth in *E. coli* cells *in vivo* (16).

### A possible three-dimensional model

The present results show that the fold of region C is functionally and structurally highly constrained. The effect of mutations could not be predicted by a simple secondary structure model. The mutations can be classified in 3 classes: (i) mutations that display a wild-type like folding and affinity for protein S8 (A641 and G599/C639); (ii) mutations that induce a substantial refolding ( $\Delta$ 640,  $\Delta$ 642, U642, C641, G643 and G599) and are not recognized by protein S8; (iii) mutations that induce a local opening of region C ( $\Delta$ 595, U595, U640, A598, A598/U642, A598/U640, A598/U640/G641 and C597/G643) with variable effect on S8 binding. These latter mutations seem to be responsible for the disruption of a network of interactions in

region C resulting in a destabilization of the postulated U598–U641 pair. Furthermore, there is a clear structural interdependence between nucleotides A595, U598, A640, A642 and G597 and/or C643. The new model we propose does not basically differ from the previous one, as far as the base-pairing scheme is concerned, however the conformation of the sugar-phosphate backbone is more irregular and tertiary interactions account for the present observations (Figs 5–6).

In this model, the U598(N3, O4)–U641(O2, N3) already proposed in the previous model (14) has been maintained. The three bulged adenines are still bulging out on the same side of the helix, facing the major groove, but their orientation has been modified. Both A595 and A640 adopt a C2' *endo* sugar pucker. Adenine 595, which is in a *syn* conformation, is stacked on A640 and both residues can be involved in an array of hydrogen bonds (Figs 5–6). Thus, hydrogen bonding between A595(N6) and the phosphate groups of both U641 and A642, between A595(N7) and the 2'OH of A640, as well as between A642(N6) and the phosphate group of U594 can occur. There is a very good correlation between the reactivity of A640 and A595 at both N1 and N7 and their accessibility in the model. Moreover, the postulated hydrogen bonds involving A595 and the ribose-phosphate backbone most likely stabilize its particular conformation. Thus, according to the model, deleting or substituting A595 results in the loss of these interactions and to the destabilization of the U598–U641 pair. The free hydrogen of C643(N4) can also form a bond with the phosphate group of A642. This should account for the observed increased reactivity of both A642 and U641 when inverting the G597–C643 pair. This model offers a rather satisfying solution for the observed interdependence between the three bulged adenines, the U–U pair and C643. Other hydrogen bonding possibilities cannot be excluded. Overall, the postulated structure is characterized by: (i) the known tendency of R–Y–R sequences for conformations in which the two purine residues stack on a side opposite to that of the pyrimidine (22); (ii) the added stabilization brought about by the third adenine 'intercalating' between the two bulged adenines.

### What is recognized by protein S8?

One characteristic feature of the model is the irregularity of the sugar-phosphate backbone (with one kink on the 5' strand and two kinks on 3' strand). The reason why the U598–U641 pair can be replaced by U598–A641 but not by A598–U641 is probably correlated with this particular geometry. Another consequence of the proposed conformation is the widening of the deep groove, allowing to position the three bulged adenines. Protein S8 may sit in the distorted deep groove of the RNA and probably recognizes the irregular backbone conformation. The model also fits with the idea that A640(N1), which is accessible in the naked RNA and protected in the bound form, is a specific contact. The invariant A642 is also a good candidate for specific interaction, in particular positions N6 and N1 which are both accessible in the model. It should be reminded that S8 binding is strongly affected by protonation of (a) residue(s) with a pK around 5–6 (13) and that an adenine was considered to be the best candidate. On the opposite, A595 which is buried and poorly

**Figure 6.** Proposed three-dimensional model of region C. (a) Stereoscopic view down the deep groove, with strand 637–647 in green, strand 592–601 in yellow, A595 in pink, A640 and A642 in blue. (b) Detailed view showing the coaxial stacking between A640 and A595 and possible hydrogen bonds (A595(N6)–OP641 and –OP642; A595(N7)–A640(2'OH); A642(N6)–OP594). (c) Detailed view after a rotation of 180° about the vertical axis, showing the U598–U641 and the possible hydrogen bond between C643(N4) and OP642.

accessible does not appear to be involved in direct interaction. The evidence that A595 is not a recognition site is provided by mutant U595. In this case, the reactivity of U595(N3) is even increased in the presence of S8, suggesting that it is tilted outside the helix. The fact that U641(N3) remains reactive in this mutant in the presence of S8 also indicates that U641 is probably not directly recognized. This can be explained by the particular location of the U598–U641 pair: in the proposed model, its access from the distorted deep groove is partially shielded by the bulged adenines.

The S8 binding site constitutes a typical example of RNA structural complexity used as a source of protein specific recognition. Our results highlight subtleties in the RNA conformation which cannot be explained by a simple secondary structure. In addition, they clearly show that the classical disruption/replacement method used to prove standard Watson–Crick base-pairing is inadequate for identifying non canonical interactions.

## ACKNOWLEDGEMENTS

We are indebted to F.Eyermann for skilful technical assistance, to C.Cachia for providing purified ribosomal protein S8, to J.Reinbolt for analysis of S8 and to B.Masquida for sequence alignments. P.Romby is thanked for critical reading of the manuscript.

## REFERENCES

- Nomura, M. and Held, W.A. (1990) in 'Ribosomes' (Nomura, M., Tissieres, A. and Lengyel, P. eds) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., pp 1933–223.
- Held, W.A., Ballou, B., Mizushima, S. and Nomura, M. (1974) *J. Biol. Chem.* **249**, 3103–3111.
- Moore, P.B. (1987) *Cold spring Harbor Symp Quant Biol.* **52**, 721–728.
- Gregory, R.J., Zeller, M.L., Thurlow, D.L., Gourse, R.L., Stark, M.J.R., Dahlberg, A.E. and Zimmermann, R.A. (1984) *J. Mol. Biol.* **178**, 287–302.
- Svensson, P., Changchien, L.M., Craven, G.R. and Noller, H.F. (1988) *J. Mol. Biol.* **200**, 301–308.
- Yates, J.L. Arfsten, A.E. and Nomura, M. (1980) *Proc. Natl. Acad. Sci. USA.* **77**, 1837–1841.
- Dean, D., Yates, J.L. and Nomura, M. (1981) *Nature* **289**, 89–91.
- Cerretti, D.P., Mattheakis, L.C., Kearney, K.R., Vu, L.V. and Nomura, M. (1988) *J. Mol. Biol.* **204**, 309–329.
- Ungewickell, E., Garrett, R.A., Ehresmann, C., Stiegler, P. and Fellner, P. (1975) *Eur. J. Biochem.* **51**, 165–180.
- Zimmermann, R.A., Mackie, G.A., Muto, A., Garrett, R.A., Ungewickell, E., Ehresmann, C., Stiegler, P., Ebel, J.P. and Fellner, P. (1975) *Nucleic Acids Res.* **2**, 279–302.
- Zimmermann, R.A. and Singh-Bergmann, K. (1979) *Biochim. Biophys. Acta* **563**, 422–431.
- Wower, I. and Brimacombe, R. (1983) *Nucleic Acids Res.* **11**, 1419–1437.
- Mougel, M., Ehresmann, B. and Ehresmann, C. (1986) *Biochemistry* **25**, 2756–2765.
- Mougel, M., Eyermann, F., Westhof, E., Romby, P., Expert-Bezançon, A., Ebel, J.P., Ehresmann, B. and Ehresmann, C. (1987) *J. Mol. Biol.* **198**, 91–107.
- Mougel, M., Allmang, C., Eyerman, F., Cachia, C., Ehresmann, B. and Ehresmann, C. (1993) *Eur. J. Biochem.* **215**, 787–792.
- Gregory, R.J., Cahill, P.B.F., Thurlow, D.L. and Zimmermann, R.A. (1988) *J. Mol. Biol.* **204**, 295–307.
- Gutell, R. R. (1993) *Nucleic Acids Res.* **21**, 3051–3054.
- Neefs, J.M., Van der Peer, Y., de Rijk, P., Chapelle, S. and de Wachter, R. (1993) *Nucleic Acids Res.* **21**, 3025–3049.
- Cachia, C., Flamion, P.J. and Schreiber, J.P. (1991) *Biochimie* **73**, 607–610.
- Westhof, E., Romby, P., Romaniuk, P., Ebel, J.P., Ehresmann, C. & Ehresmann, B. (1989) *J. Mol. Biol.* **207**, 417–431.
- Metzenberg, S., Joblet, C., Verspieren, P. and Agabian, N. (1993) *Nucleic Acids Res.* **21**, 4936–4940.
- van de Hoogen, Y.T., Treurniet, S.J., Roelen, H.C.P.F., de Vroom, E., van der Marel, G.A., van Boom, J.H. and Altona, C. (1988) *Eur J. Biochem.* **171**, 155–162.

