THESE

présentée pour obtenir le grade de
Docteur de l'Université de Strasbourg
Discipline: Science du Vivant
Spécialité: Aspects Moléculaire et Cellulaire de la Biologie

Par

Olga KOSSINOVA

Insights into the selenocysteine incorporation mechanism in mammals

Soutenue publiquement le 30 septembre 2011

Membres du jury

Rapporteur externe:	Dr. Olivier NAMY
Rapporteur externe:	Dr. Satoko YOSHIZAWA
Rapporteur interne:	Dr. Marat YUSUPOV
Examinateur:	Dr. Christiane BRANLANT
Directeur de thèse:	Dr. Alain KROL
Directeur de thèse:	Pr. Galina KARPOVA

Aknowledgements

I would like to thank very much Dr. Olivier Namy, Dr. Satoko Yoshizawa, Dr. Marat Yusupov and Dr. Christiane Branlant for having accepted to examine and evaluate my PhD studies.

I would like to express my gratitude to the scientific advisers, Pr. Galina Karpova and Dr. Alain Krol, for giving me an opportunity to make a PhD in their laboratories. I'm grateful to Pr. Galina Karpova for her continuous support and excellent guidance. I would like to thank Dr. Alain Krol for fruitful discussions, numerous advises, for support and help in my PhD studies and in everyday life in France. I'm extremely grateful to Dr. Alexey Malygin from Novosibirsk lab for his invaluable help, optimism and extraordinary support during my PhD studies.

Thanks to everyone in the Strasbourg lab and UPR 9002. I would like to thank Christine and Stephanie for numerous advices on lab work and everyday life. Thanks to Anne for always understanding my French and also for the excellent technical assistance in the lab. Many thanks to Patryk and Anne-Sophie and to former lab members Michael, Artem, Akiko and Laurence for our conversations and jokes. Special thanks to José and Artem, who helped to rescue this manuscript from the broken laptop.

Thanks to everyone in the Novosibirsk lab, for your help, support, our wonderful celebrations in the lab and tea pauses I missed so much in France. Thanks to Julia, Tatyana and Konstantin for sending me Russian sweets for the New Year.

I would like to thank all my friends in Novosibirsk and in Strasbourg for the cheering me up and for making my life brighter!

Lastly, many thanks to my parents and my husband, who supported me during these years.

I would like to thank the French Embassy in Moscow for the partial scholarships in 2008-2010, and European Doctoral College and UPR 9002 for bridge scholarships in 2011.

Olga Kossinova was a member of the European Doctoral College of the University of Strasbourg during the preparation of her PhD, from 2008 to 2011, class name « Rosa Parks ». She has benefited from specific financial supports offered by the College and, along with her mainstream research, has followed a special course on topics of general European interests presented by international experts. This PhD research project has been led with the collaboration of two institutions: the Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia and the University of Strasbourg, France.

Table of contents

TABLE OF FIGURES

Résumé de la these en français

Thesis in English

List of abbreviations

40S, 60S, 80S	eukaryotic small and large ribosomal subunits and ribosome
A, C, G, T, U	adenine, cytosine, guanine, thymine, uracil
A-site	ribosomal site that holds incoming aminoacyl-tRNA
aa	amino acids
AcOH	acetic acid
ATP	adenosine 5’-triphosphate
aaUTP	5'-(3'-aminoallyl)UTP
bp	base pair
BP	bromophenol blue
BSA	bovine serum albumin
B.taurus	*Bos taurus*
CBB	Coomassie brilliant blue
C. elegans	*Caenorhabditis elegans*
$ClRCH_2NH_2$	4-(N-2-chloroethyl-N-methylamino)benzylamine
C.picta	*Chrysemus picta*
cryo EM	cryogenic electron microscopy
C-terminal	carboxy-terminal
DIO	iodothyronine deiodinase
D. melanogaster	*Drosophila melanogaster*
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
cDNA	complementary deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
eEF	eukaryotic elongation factor
EFSec	selenocysteine-specific elongation factor
EGTA	ethyleneglycoltetraacetic acid
eIF	eukaryotic initiation factor
E.melo	*Echinus melo*
E.sexfasciatus	*Epiplatys sexfasciatus*
eRF	eukaryotic release factor

E-site	ribosomal site that accomodates deacylated tRNA
G.gallus	*Gallus gallus*
GMPPNP	Guanosine 5′-[β,γ-imido]triphosphate
G.gorilla	*Gorilla gorilla*
GPx	glutathione peroxidase
H.colliei	*Hydrolagus colliei*
His tag	hexahistidine tag
HRP	horseradish peroxidase
H.sapiens	*Homo sapiens*
Hsp	Heat shock protein
kD	kilo dalton
Met-tRNA$^{Met}_{i}$	initiator tRNA aminoacylated with methionine
Msr	methionine sulfoxide reductase
NES	nuclear export signal
NLS	nuclear localization signal
NMD	nonsense-mediated decay
nt	nucleotide
N-terninal	amino-terminal
PAGE	electrophoresis in polyacrylamide gel
PCR	polymerase chain reaction
Phe	phenylalanine
P-site	ribosomal site that holds peptidyl tRNA
PSTK	phosphoseryl-tRNA kinase
P.troglodytes	*Pan troglodytes*
RNA	ribonucleic acid
mRNA	messenger ribonucleic acid
rRNA	ribosomal RNA
snRNA	small nuclear ribonucleic acid
snoRNA	small nucleolar ribonucleic acid
tRNA	transfer ribonucleic acid
RNP	ribonucleoprotein
mRNP	messenger ribonucleoprotein

snRNP	small nuclear ribonucleoprotein
snoRNP	small nucleolar ribonucleoprotein
RRL	rabbit reticulocyte lysate
s4-UTP	4-thiouridine 5'-triphosphate
SBP2	SECIS binding protein 2
SDS	sodium dodecyl sulfate
Sec	selenocysteine
SECIS	selenocysteine insertion sequence
SecS	selenocysteine synthase
Sel	selenoprotein
SELEX	systematic evolution of ligands by exponential enrichment
Ser	serine
SerRS	seryl-tRNA synthetase
SPS	selenophosphate synthetase
SRE	selenocysteine codon redefinition element
TR	thioredoxin reductase
UTR	untranslated region
UV	ultraviolet
X.laevis	*Xenopus laevis*
XC	xylene cyanol
λ	wave length
ψ	pseudouridine
Å	angstrem

Résumé de la thèse en français

Biosynthèse des sélénoprotéines. Interactions ARN messager-protéines-ribosome

La forme biologique de l'oligo-élément sélénium est l'acide aminé sélénocystéine. Il est codé par un triplet UGA (codon Sec) qui agit généralement comme un codon stop. Ainsi, une machinerie spécialisée est utilisée pour incorporer cet acide aminé dans les sélénoprotéines pendant la traduction (pour revue Allmang and Krol, 2006 ; Allmang et al., 2009). Ce processus implique en particulier une structure en tige-boucle située dans la région 3'-UTR des ARN messagers, appelée SelenoCysteine Insertion Sequence (SECIS), le facteur d'élongation spécialisé EFSec et d'autres facteurs protéiques. L'un de ceux-ci est la protéine SBP2 (SECIS Binding Protein 2), qui est nécessaire pour la reconnaissance du triplet UGA comme codon Sec par le ribosome. Aucun mécanisme détaillé pour décrire l'incorporation de sélénocystéine n'a encore été avancé. Toutefois, deux modèles ont déjà été proposés dans la littérature qui diffèrent par la localisation de SBP2. L'un stipule que SBP2 se lie en premier aux ribosomes puis à l'élément SECIS (Figure 1A), tandis que l'autre suggère que SBP2 est liée à l'élément SECIS en premier lieu (Figure 1B).

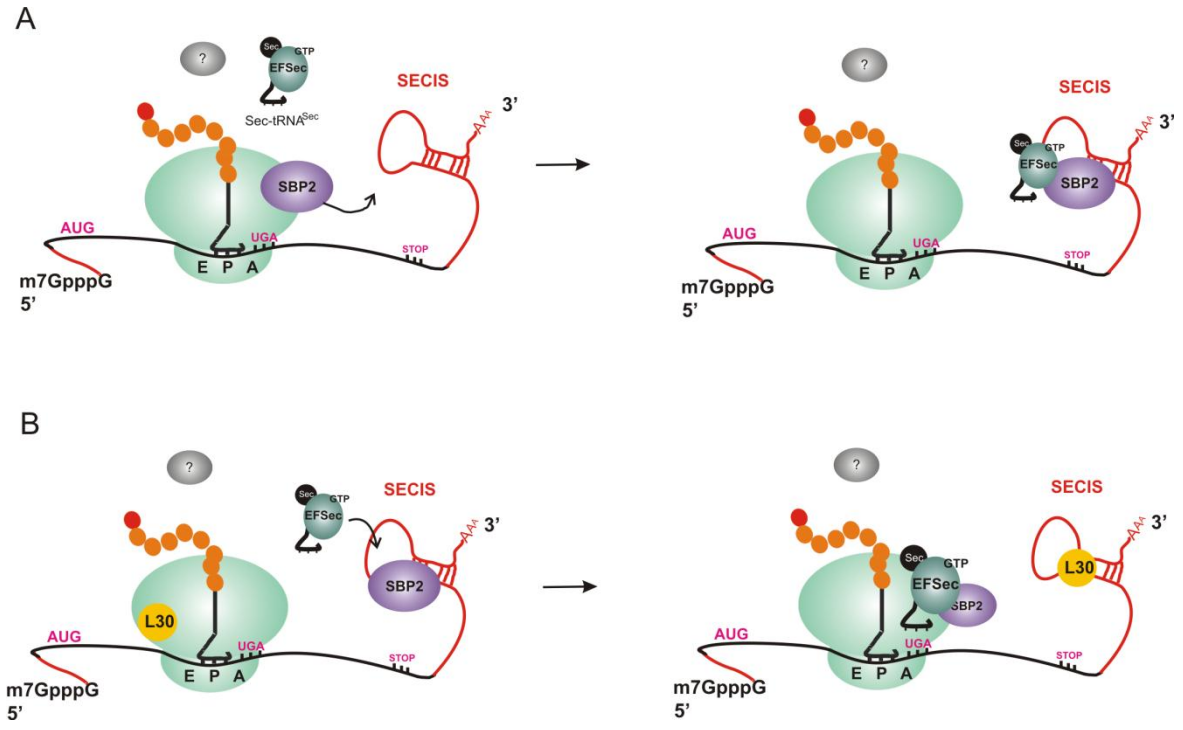

Figure 1. Mécanisme d'incorporation de la sélénocysteine dans les protéines chez les eucaryotes. (A) SBP2 est liée au ribosome, puis interagit avec l'élément SECIS et le complexe EFSec/Sec-ARNt^{Sec} afin de le présenter au site A du ribosome. (Donovan et al., 2008; Kinzy et al., 2005). (B) SBP2 liée à l'élément SECIS recrute le facteur EFSec associé au Sec-ARNt^{Sec}. L'échange entre la protéine SBP2 et la protéine ribosomique L30 permet d'apporter le Sec-ARNt^{Sec} au site ribosomique A (Chavatte et al., 2005).

Pour obtenir des avancées dans la compréhension du mécanisme et, en particulier, pour identifier de nouveaux partenaires d'interaction de l'élément SECIS, j'ai développé deux stratégies.

1/ Dans une première approche, il s'agissait d'identifier les contacts de l'élément SECIS au niveau du ribosome (environnement protéique et/ou acide nucléique) par le biais de pontages covalents. Dans ce but, j'ai construit un ensemble d'ARN messagers sélénoprotéine modèles, contenant une 5' UTR courte, à l'instar de celles des ARNm de sélénoprotéine (5' UTR naturelle ou 5' UTR riche en A), suivie par la séquence codant pour le tétrapeptide Met-Sec-Phe-Phe (ou Met-Phe-Phe-Phe pour l'ARNm contrôle), puis d'une séquence d'espacement et enfin l'élément SECIS en 3'. Pour préparer cet ARNm, une matrice d'ADN correspondant à l'ARNm sans SECIS a été synthétisée, et l'ARN correspondant a été obtenu par transcription par l'ARN polymérase du phage T7. L'ARNm avec SECIS a ensuite été obtenu par ligature de deux fragments d'ARN au moyen d'une attelle ADN, l'un contenant l'ARNm et l'autre l'élément SECIS. L'applicabilité de ces ARNm modèles pour l'étude de l'insertion de sélénocystéine a été montrée à l'aide de lysat de réticulocytes de lapin et d'inhibiteurs d'élongation de la traduction: (1) le GMPPNP, analogue non-hydrolysable du GTP (utilisé pour former le complex 48S de pré-initiation) ; (2) les antibiotiques anisomycine (bloquant l'étape de transpeptidation) et émétine (bloquant l'étape de translocation lors de la phase d'élongation) (Figure 2A). Nous avons constaté que les ARNm modèles contenant une 5' UTR riche en A possèdent une efficacité supérieure pour la liaison au ribosome par rapport à ceux contenant une 5' UTR courte naturelle (35-40% contre 10%) (Figure 2B).

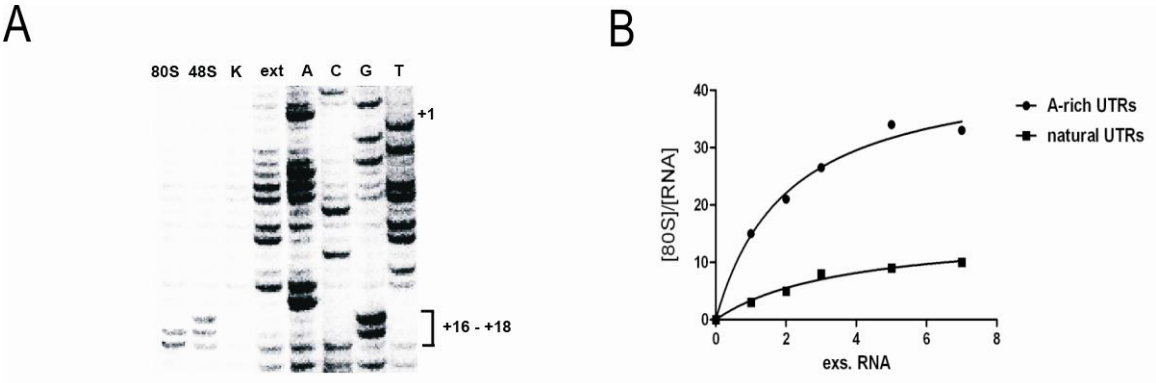

Figure 2. Caractéristiques des ARN messagers sélénoprotéine modèles. (A) Les expériences de « toe-print » montrent que les complexes 48S et 80S sont formés dans le lysat de réticulocytes de lapin (RRL) avec l'ARNm sélénoprotéine modèle (flSec) en présence de SBP2. « 80S » correspond au complexe formé en présence de 1 mM émétine ; « 48S » correspond à 2 mM GMPPNP ; aucun inhibiteur de traduction n'a été ajouté dans « K ». ext,A,C,G,T représentent les pistes de séquençage. (B) Les ARNm flSec avec une 5'UTR riche en A (cercles) possèdent une efficacité supérieure pour la liaison au ribosome par rapport à ceux contenant une 5' UTR courte naturelle (carrés).

Cette étape préalable étant validée, j'ai préparé des ARNm modèles avec éléments SECIS contenant des agents pontants : la 4-thiouridine ou bien un groupe perfluoroazoture. Ils ont été incorporés soit statistiquement (4-thiouridine) soit au niveau de positions prédéfinies dans l'élément SECIS à l'aide d'oligodésoxynucléotides complémentaires aux régions ciblées. En comparant le rendement d'obtention des produits dérivés, il a été décidé d'utiliser uniquement des ARNm avec 4-thiouridines. Après pontage aux ultra-violets dans le lysat de réticulocytes de lapin et analyse par Western blot (WB), il s'avère que SBP2 est liée à l'élément SECIS dans les complexes 48S de pré-initiation et 80S de pré-translocation. Lorsque la formation de la liaison peptidique est bloquée par l'antibiotique anisomycine, l'élément SECIS n'est plus lié à SBP2 mais à un ensemble de protéines ribosomiques. Néanmoins, SBP2 est présente dans le complexe, comme indiqué par WB. La faible quantité de protéines ribosomiques pontées n'a pas encore pu conduire à leur identification par spectrométrie de masse, en dépit de multiples tentatives d'optimisation. Ceci nous a cependant permis de proposer l'interprétation suivante (Figure 3). Pendant l'étape de transpeptidation, SBP2 est associée au ribosome mais après transpeptidation, SBP2 quitte le ribosome (peut-être suite à un changement conformationnel sur le ribosome et pas nécessairement avec l'aide d'un facteur) et, en raison de sa forte affinité pour SECIS, se lie ensuite à celui-ci à l'étape de pré-translocation.

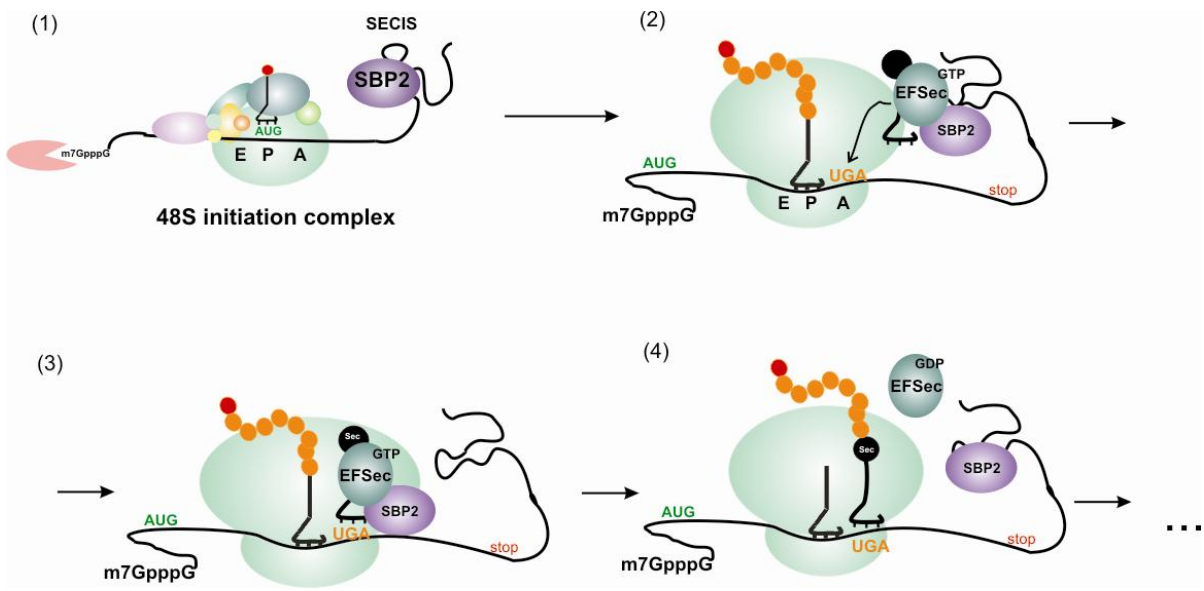

Figure 3. Modèle d'incorporation de sélénocystéine chez les eucaryotes. SBP2 est liée à l'élément SECIS (1). Lorsque le ribosome atteint le codon UGA Sec, l'élément SECIS peut contacter le ribosome de manière plutôt transitoire (2) pour délivrer le complexe SBP2/EFSec/tRNA^{Sec}/GTP au site ribosomique A (3). Après entrée de l'ARNt^{Sec} au site A, ou après formation de la liaison peptidique, SBP2 quitte le ribosome et, en raison de son affinité, revient se lier au SECIS (4).

2/ Lors d'une collaboration antérieure entre les deux laboratoires, SBP2 avait été montrée se liant spécifiquement à la sous-unité 60S humaine purifiée et non pas à la sous-unité 40S. Dans cette même étude, la région de SBP2 nécessaire à cette interaction a été identifiée. Il s'agit d'une séquence pentapeptide située dans le domaine riche en lysines (K-rich domain). Cependant, le site de liaison de SBP2 sur la sous-unité 60S n'avait pas encore été localisé. Nous avons entrepris ce projet avec pour but d'identifier aussi bien les protéines ribosomiques que les régions de l'ARN ribosomique 28S qui pourraient constituer ce site de liaison. Pour identifier les protéines ribosomiques, les complexes SBP2-40S, 60S-SBP2 et 80S-SBP2 ont été soumis à l'action de réactifs de pontage. J'ai effectué soit un pontage direct par les UV (qui possèdent l'avantage de générer des pontages entre deux partenaires quasiment en contact direct) soit des pontages générés par les réactifs bifonctionnels (diépoxybutane, rayon de pontage environ 4Å) ou 2-iminothiolane (7Å pour un pontage ARN-protéines et 14Å pour protéine-protéine). Ainsi, nous avons pu montrer que SBP2 se lie à la sous-unité 60S mais pas à la sous-unité 40S, que ce soit dans les complexes 40S-SBP2 ou 80S-SBP2. Ceci est en bon accord et confirme les données obtenues précédemment sur l'association SBP2-60S. Il est intéressant de noter qu'après pontage, SBP2 est liée à la sous-unité 60S, soit à l'état libre soit dans le ribosome 80S. En outre, le rendement de pontage ARNr 28S-SBP2 fut plus de 10 fois supérieur à celui obtenu avec les protéines ribosomiques, indiquant que l'ARNr 28S contribue davantage au site de liaison de SBP2 sur la 60S que les protéines ribosomiques.

Afin d'identifier cette (ces) région(s) de l'ARNr 28S, nous avons employé l'approche utilisant les radicaux hydroxyles. Ces radicaux, générés par les électrons libérés par oxydation du Fe^{2+} en Fe^{3+} dans le complexe Fe (II)-EDTA libre en présence d'un oxydant (eau oxygénée), sont largement utilisés pour sonder l'accessibilité du ribose dans le squelette phosphodiester d'un ARN, en présence ou absence d'un ligand. Les positions des scissions induites dans la liaison phosphodiester sont ensuite identifiées par une réaction de transcription inverse au moyen d'oligodésoxynucléotides marqués en 5'au ^{32}P.

Après interaction ARNr 28S-SBP2 in vitro et action des radicaux hydroxyles, les 5000 nucléotides de l'ARNr 28S humain ont été examinés. L'interprétation des données indique que la liaison de SBP2 à la sous-unité 60S humaine protège deux petites régions discrètes dans deux hélices du très grand segment d'expansion 7 (ES7). Malheureusement, ces hélices n'ont pas été entièrement modélisées dans les structures cryo-EM des ribosomes eucaryotes disponibles à l'heure actuelle. Néanmoins, on peut conclure qu'elles résident sur le côté solvant de la sous-unité 60S du ribosome, à proximité du site ribosomique A (Figure 4). Ces résultats sont en bon accord avec notre suggestion que SBP2 peut être associée au ribosome

au cours de l'étape de transpeptidation et qu'elle quitte le ribosome en raison de changements conformationnels survenant après transpeptidation.

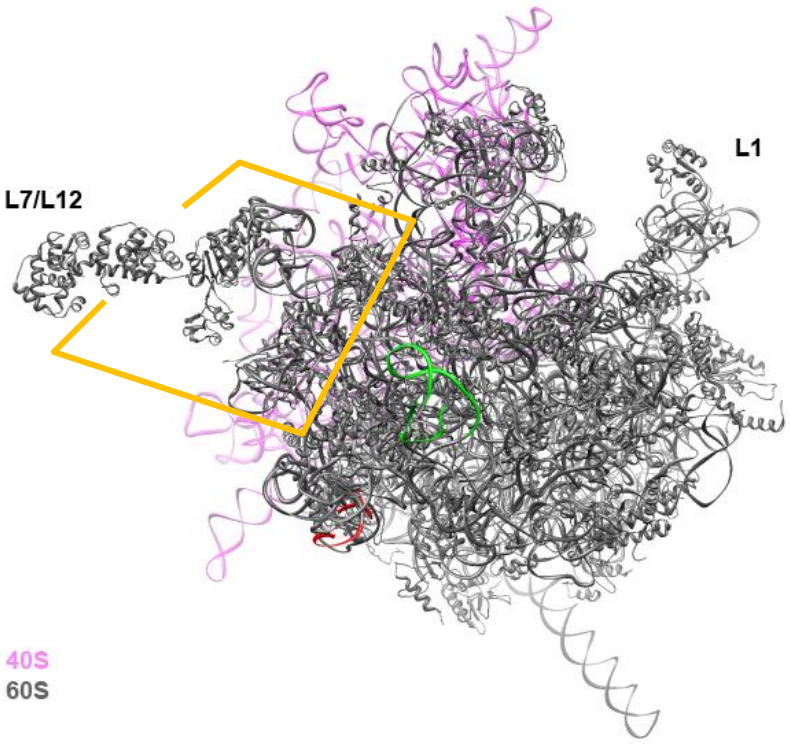

Figure 4. Modèle 3D du ribosome de blé: vue du côté solvant de la sous-unité 60S (Arnache et al., 2010, numéros d'accession PDB 3IZR, 3IZ6, 3IZ7 et 3IZ9). La sous-unité 60S est en gris, la 40S en rose (seul le squelette phosphodiester d'un ARN ribosomique est montré). La partie du segment d'expansion ES7 est en vert, l'hélice de ES39 qui devient plus accessible aux radicaux hydroxyles est en rouge. La zone possible d'interaction SBP2-ribosome est entourée en jaune.

Deux approches ont été prises pour élucider les interactions moléculaires de SBP2 sur le ribosome 60S. (1) Un travail est en cours en collaboration avec le groupe Westhof (IBMC, Strasbourg) pour dériver un modèle de structure secondaire du segment d'expansion 7, par des alignements de séquence basés sur la structure, à l'aide d'environ 100 séquences d'ARNr 28S. (2) Simultanément, une collaboration a été entreprise avec le groupe de Bruno Klaholz (IGBMC, Strasbourg) pour localiser par cryo-microscopie électronique le site de SBP2 sur la sous-unité 60S humaine.

L'ensemble de mes travaux de thèse ont fourni, et fourniront de par leurs développements ultérieurs, des données mécanistiques appréciables pour une meilleure compréhension du mécanisme complexe de recodage du codon UGA en sélénocystéine. En particulier, j'ai pu apporter une contribution à l'élucidation des principes d'interactions moléculaires entre la protéine SBP2 et le ribosome humain lors du mécanisme de synthèse des sélénoprotéines.

Part I

Introduction

Introduction

1. Selenium and its biological function

1.1. Selenium and selenocysteine

Selenium is an essential micronutrient that exerts significant health benefits. In humans, selenium deficiency is linked to the Keshan (a potentially fatal juvenile cardiomyopathy) or Kashin-Beck diseases (an osteoarticular disease) that occur in several Asian regions, where dietary selenium is very low because the soil is deprived of this element (Bellinger et al., 2009). Selenium has also been implicated in the prevention of viral infections, cancer and infertility; it has been shown as an important factor for thyroid hormone maturation, the immune system and muscle development and function (Hatfield et al., 2006; Lescure et al., 2009). The main biological form of selenium is the 21st amino acid selenocysteine that has been discovered in the 1970s (Cone et al., 1976; Hatfield and Diamond, 1993).

The amino acid selenocysteine is an analog of cysteine where the sulfur atom is replaced by selenium (Figure 1). It is encoded by a UGA codon (Chambers et al., 1986) and co-translationally incorporated into the nascent polypeptide chains by a specialized mechanism that will be discussed in section 4. Over the last decade, a lot of selenium-containing proteins, selenoproteins, have been identified and the cellular functions of some of them characterized, leading to a better understanding of the biological role of selenium.

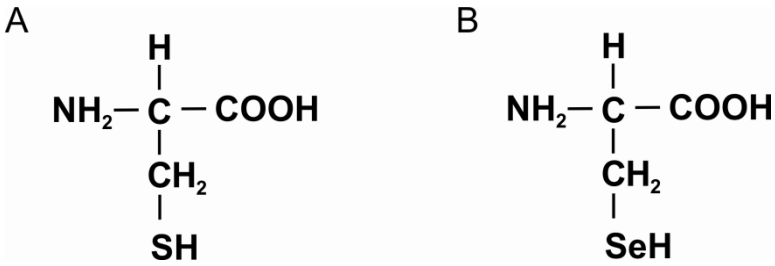

Figure 1. Chemical structures of cysteine (A) and selenocysteine (B).

1.2. The eukaryotic selenoproteome

Selenoproteins are found in all three domains of life, but not in all species. For example, the bacterial proteome contains up to 39 selenoproteins; 7 to 10 selenoproteins are found in archaea, and 3 to 37 selenoprotein genes are contained in eukaryotes with about 25 in mammals (Lobanov et al., 2009). Extreme cases can be found: the nematode *C.elegans*

expresses one single selenoprotein (thioredoxin reductase) (Gladyshev et al., 1999) while up to 37 selenoproteins are present in aquatic organisms like fishes and algae (Zhang and Gladyshev, 2010). In general, sea-water organisms possess more selenium-containing proteins than terrestrial ones. All land plants, fungi, some protists and insects do not have selenoproteins at all (reviewed in Kryukov et al., 2003; Lobanov et al., 2009).

Eukaryotic selenoproteins (Table 1) can be divided into two groups. In the first one, the Sec residue is located in the N-terminal part of the protein, often in the CXXU motif (C=cysteine, X=any amino acid, U=selenocysteine), that is a redox-active thioredoxin-like motif (Johannsson et al., 2005). This group comprises glutathione peroxidases (GPx 1-4 and 6), iodothyronine deiodinases (DI 1-3), methionine sulfoxide reductase A, selenophosphate synthetase SPS2, Sep15 and selenoproteins H, M, N, T, U, V and W. Interestingly, SPS2 being a selenoprotein itself, is involved in selenocysteine biosynthesis (discussed in section 3.2). Another interesting example is SelN that was discovered in our laboratory (Lescure et al., 1999). SelN was the first selenoprotein shown to be involved in a genetic disorder (Moghadaszadeh et al., 2001). Mutations within the coding region of the SelN gene (SEPN1) are associated with early-developing muscular disorders now referred to as SEPN1-related myopathies. However, the SelN catalytic function still remains unknown. Apart from a possible antioxidant function, it was suggested that SelN is involved in muscle biogenesis and regulation of Ca^{2+} homeostasis in the muscle (reviewed in Castets et al., 2009; Lescure et al., 2009).

Protein	Number of Sec; location	Function	Linked pathology	Other remarks
GPx1	single, C-ternimal	free-radical reduction	colon and breast cancer	
GPx2		free-radical reduction		
GPx3		free-radical reduction		
GPx4		oxidized phospholipid reduction and chromatin condensation during spermatogenesis	male infertility	
GPx6		free-radical reduction		Selenoprotein in humans and pig only
DI1		thyroid hormone maturation		
DI2		thyroid hormone maturation		
DI3		thyroid hormone maturation		

Protein	Number of Sec; location	Function	Linked pathology	Other remarks
SPS2	single, C-ternimal	selenocysteine synthesis		
Sep15		involved in cell apoptosis		
SelH		unknown		
SelM		unknown	Alzheimer disease	
SelN		antioxidant function, muscle biogenesis	myopathies, muscular dystrophy	
SelT		regulation of Ca^{2+} homeostasis		
SelU		unknown		
SelV		unknown, highly expressed in testis		
SelW		antioxidant function		
Msr A		methionine-S-sulfoxide reductase		Selenoprotein only in *C.reinhardtii*
TrxR1	single, N-terminal	redox enzymes, involved in multiple pathways and regulation mechanisms	Alzheimer disease	
TrxR2				
TrxR3				
SelI		unknown		
SelJ		possibly structural function		Actinopterygian fishes and sea urchin
SelK		unknown		
SelO		unknown		
SelS		possible antioxidant, inflammation regulation		
MsrB		methionine-R-sulfoxide reductase	Alzheimer disease	
SelL	2 Sec	unknown		Present in aquatic organisms only
MsrB	4 Sec	oxidoreductase		Present in *M. senile* only
SelP	Multiple Sec, N- and C-terminal	Nitrite reductase, plasmatic selenium transport		

Table 1. The eukaryotic selenoproteome. Abbreviations: GPx, glutathione peroxidase; DI, iodothyronine deiodinase; TrxR, thioredoxin reductase, SPS2, selenophosphate synthetase, Sel, selenoprotein. Data taken from Kryukov et al., (2003); Rederstorff et al., (2006); Lobanov et al., (2009); Lee et al., (2011).

In the second group, the Sec residue resides in the C-terminal part of the protein. The members of this group are thioredoxin reductases TrxR1-3, selenoproteins K, S, O, I, J, MsrB (methionine sulfoxide reductase B). The functions of these proteins, when known, are summarized in Table 1.

There are several examples of multiple Sec-containing proteins. SelL, found in diverse aquatic organisms such as fishes and invertebrates, contains two Sec residues organized in the UXXU motif. This is the first selenoprotein discovered in which a diselenide bond can be formed (Shchedrina et al., 2007). Another selenoprotein with multiple selenocysteine residues is SelP. The amount of Sec residues can vary from 7 to 28 in different species. In this protein, one Sec is always located in the N-terminal part of SelP, the others being found in the C-terminal part. It was proposed that this protein plays a key role in the Se transport from the liver to other organs (Hoffmann et al., 2007). Very recently, the Gladyshev group computationally identified a new multiple Sec-containing protein, the MsrB homolog from *Metridium senile*, which has four in-frame UGA codons and two SECIS elements (Lee et al., 2011). One of the UGA codons corresponds to the conserved catalytic Sec or Cys in MsrBs, whereas the other three have no protein homolog either with Sec or Cys at these positions. The authors proposed that some if not all the Sec residues have a non-catalytic function, for example metal cation binding.

To summarize, selenoproteins form a small but important class of proteins involved in vital processes. At least some of them are of major importance for life as demonstrated with knockout mice for the $tRNA^{Sec}$ gene which showed early embryonic lethality (Bösl et al., 1997; Downey et al., 2009). To provide the synthesis of selenoproteins and, thereby, normal cell life, an energy-demanding synthesis machinery is needed (discussed in detail in section 5) as well as constant selenium supply. As it has been already mentioned above, some organisms manage to live without selenoproteins. In these organisms, Cys-containing homologs of almost all selenoproteins can be found (summarized in Table 2) (Castellano et al., 2005). Interestingly also, individual "selenoproteins" are not always Sec-containing between domains of life and sometimes between certain branches of organisms within one of the kingdoms. For example, the vast majority of TrxRs are selenoproteins, whereas the *D.melanogaster* TrxR contains Cys instead. Another example is selenoprotein J that is present only in actinopterygian fishes and sea urchin and has a Cys-containing homolog in Cnidarians (aquatic organisms either medusa-like or sessile polyps) (Castellano et al., 2005). It is also puzzling that C. elegans maintained the whole selenocysteine insertion machinery to synthesize the only selenoprotein thioredoxin reductase (Gladyshev et al., 1999). The question thus arises: why did certain organisms keep (or develop) these few selenoproteins?

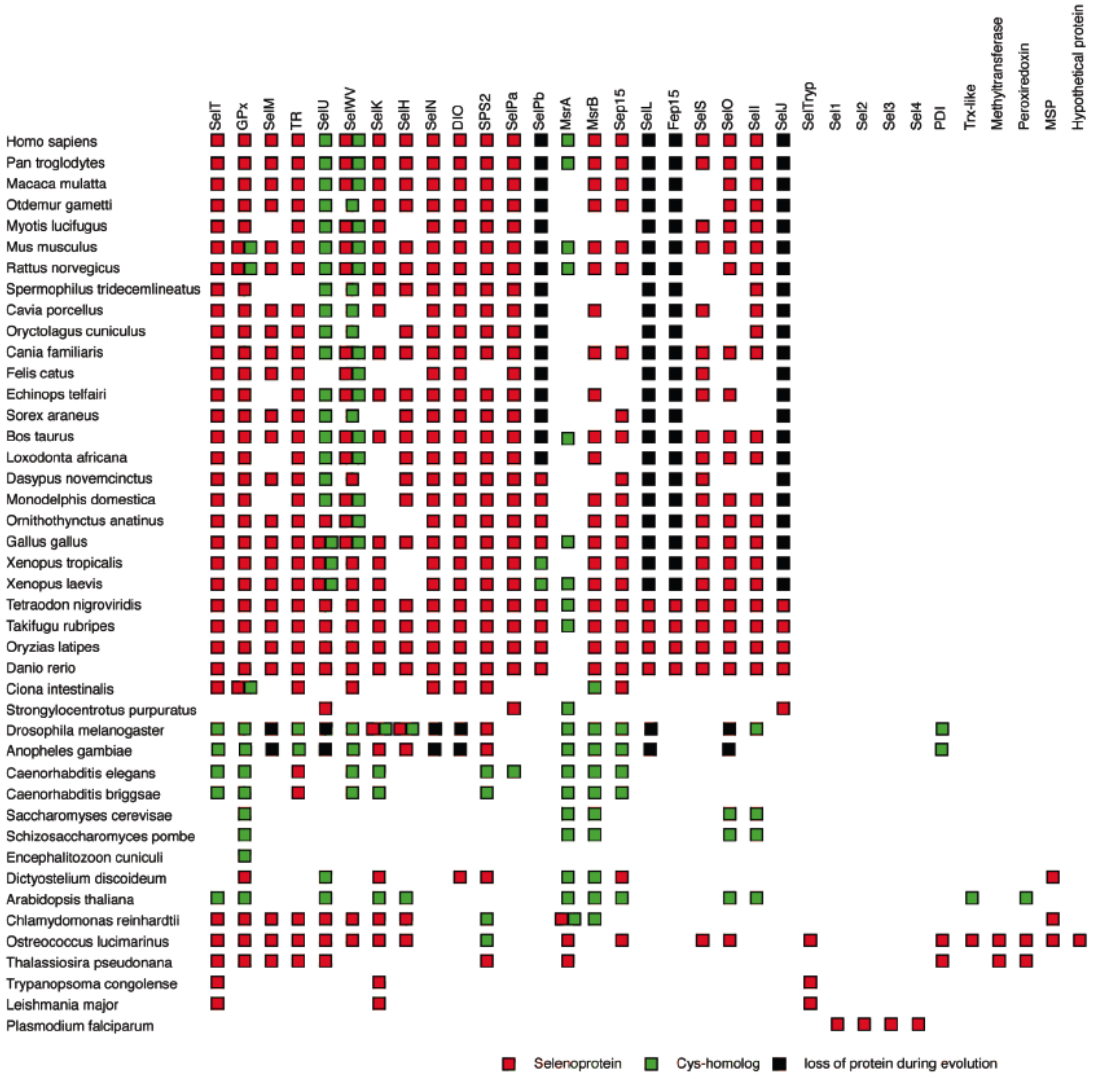

Table 2. Distribution of eukaryotic selenoproteins and Cys homologues in eukaryotes. Red, green and black boxes indicate selenoproteins, cysteine homologs and proteins lost during evolution, respectively. Adapted from the thesis of A. Takeuchi. Data taken from (Castellano et al., 2001; Kryukov et al., 2003; Castellano et al., 2004; Castellano et al., 2005; Taskov et al., 2005; Lobanov et al., 2006a; Lobanov et al., 2006b; Lobanov et al., 2007; Castellano et al., 2008, Chapple and Guigo, 2008, Dayer et al., 2008, Lobanov et al., 2008a, Lobanov et al., 2008b). Abbreviations: GPx: glutathione peroxidase, TR: Thioredoxin reductase, DIO: Iodothyronine deiodinase, SPS2: Selenophosphate synthetase, PDI: Protein disulfide isomerase, MSP: Predicted membrane sleenoprotein, Msr: methionine sulfoxide reductase.

2. Canonical mechanism of translation in eukaryotes

Prior to describing the mechanism of selenoprotein synthesis, which is the main topic of the Introduction, I would like to briefly introduce the canonical mechanism of protein synthesis in eukaryotes.

In eukaryotes, translation occurs in the cytoplasm on the 80S ribosome which is composed of the 60S and 40S subunits. The process of translation can be divided into four steps: initiation of protein synthesis, elongation, termination and ribosome recycling (reviewed in Rodnina and Wintermeyer, 2009; Sonenberg and Hinnebusch, 2009; Jackson et al., 2010).

2.1. Initiation

A eukaryotic mRNA carries three characteristic structural features. The first one is the presence of a modification at the 5'end, called the m^7G cap. It is constituted by a post-transcriptionally added guanine nucleotide methylated at the N7 position. It is connected to the DNA-encoded sequence via an unusual 5'-5' linkage (Figure 2). Secondly, a sequence composed of three purine bases upstream of the start-codon and a G immediately downstream the start codon (called the Kozak sequence) is present in many but not all mRNAs (Kozak, 1989). The third modification is the post-transcriptionally added polyA sequence (reviewed in Sonenberg and Hinnebusch, 2009).

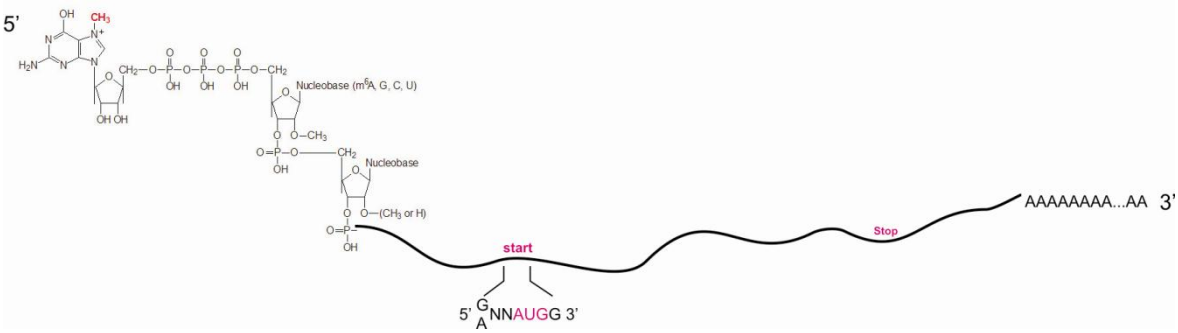

Figure 2. Eukaryotic messenger RNA. 5' cap, Kozak sequence and 3' polyA are shown.

Translation initiation eventually leads to the formation of the 80S ribosome. This step requires a number of initiation factors (Table 3) and comprises several events, summarized in Figure 3. These are (i) formation of the 43S pre-initiation complex; (ii) mRNA activation, during which the mRNA cap-proximal region is unwound in an ATP-dependent manner; (iii)

attachment of the 43S complex to the mRNA which scans the 5' UTR in the 5' to 3' direction; (iv) recognition of the AUG start codon and formation of the 48S initiation complex where the Met-tRNA$^{Met}_{i}$ base-pairs with the AUG codon of the mRNA in the P-site of the 40S subunit (at this time, the scanning complex is switched to the "closed" conformation). The final event is the GTP-dependent joining of the 48S pre-initiation complex with the 60S ribosomal subunit, resulting in the formation of the 80S initiation complex (Figure 3) (reviewed in Pestova and Kolupaeva, 2002; Sonenberg and Hinnebusch, 2009; Jackson et al., 2010).

name		Description
Factors involved in start codon selection		
eIF1		Stimulates eIF2-GTP-Met-tRNA$_i$ binding to 40S; promotes ribosomal scanning; blocks premature hydrolysis of eIF2-bound GTP
eIF1A		Together with eIF1 promotes ribosomal scanning and initiation codon selection
Factors involved in mRNA recruitment		
eIF3		Interacts with 43S pre-initiation complex components, recruits 43S complex to mRNA, promotes subsequent scanning
eIF4F cap-binding complex	eIF4A	DEAD-box ATPase and ATP-dependent RNA helicase
	eIF4E	Binds cap structure at the 5' mRNA ternimus
	eIF4G	Binds eIF4A, eIF4E, eIF3, PABP and mRNA; enhances eIF4A helicase activity
eIF4B		Enhances helicase activity of eIF4A
PABP		Interacts with both 3' polyA sequence of mRNA and eIF4G; enhances eIF4F binding to 5' cap
Factors delivering Met-tRNAi to 40S ribosomal subunit		
eIF2		Binds Met-tRNA$_i$ and GTP, mediates ribosomal recruitment of Met-tRNA$_i$.
eIF5		Induces hydrolysis of eIF2-bound GTP
Factors that mediate ribosomal subunit joining		
eIF5B		eIF5B interacts with eIF1A, stimulating subunit joining and GTP hydrolisys

Table 3. Eukaryotic initiation factors (adopted from Sonenberg and Hinnebusch, 2009; Jackson et al., 2010).

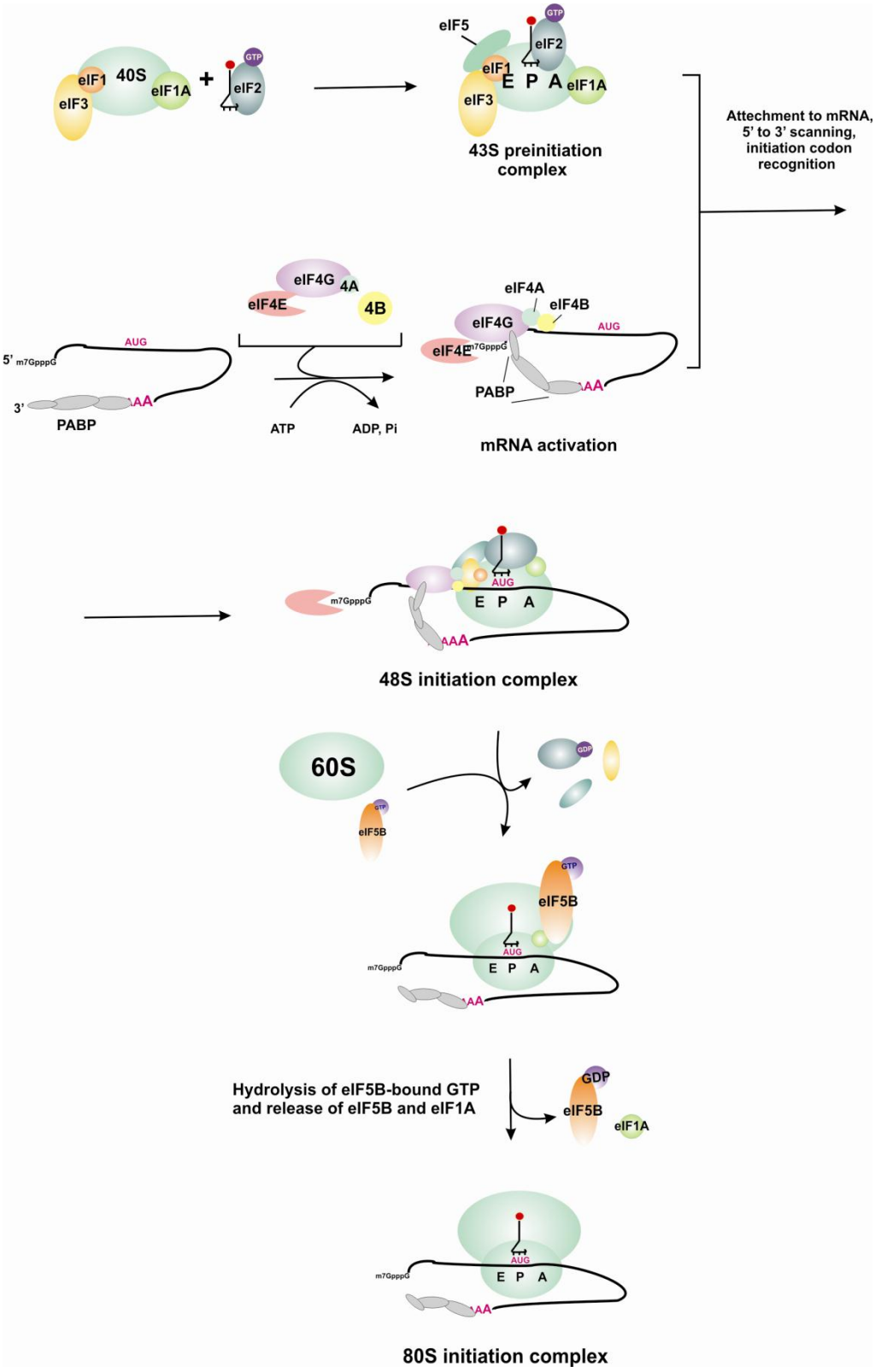

Figure 3. Simplified model of eukaryotic translation initiation, according to (Jackson et al., 2010). Scheme represents factors involved and the order of events. Abbreviations: 60S and 40S stand for large and small ribosomal subunits, respectively; A, P and E letters stand for ribosomal A-site, P-site and E-site; PABP, poly A binding protein; eIF, eukaryotic initiation factor.

2.2. Elongation

The elongation phase consists of three steps: (i) decoding of mRNA codons by the specific aminoacyl-tRNA; (ii) peptide bond formation; (iii) translocation of the peptidyl-tRNA from the A-site to the P-site and, at the same time, introduction of the next mRNA codon to the A-site. The decoding and translocation steps are facilitated by elongation factors (Figure 4). Factors involved in tRNA recruitment are eEF1A and eEF1B. eEF1A in the GTP-bound form interacts with the aminoacyl-tRNA and brings it to the A-site of the ribosome. Following GTP hydrolysis, the eEF1A/GDP is released from the ribosome. The GDP-bound eEF1A cannot bind aminoacyl-tRNAs and is recycled to the active GTP-bound form by eEF1B (for review, see Merrick and Nyborg, 2000). After the peptide-bond formation catalyzed by the ribosome, elongation factor eEF2 is needed to translocate the ribosome along the mRNA for a distance of one codon. As a result, the $tRNA^{Met}_{i}$ is moved to the E-site of the ribosome; another aminoacyl-tRNA, bearing a covalently linked dipeptide, is moved to the P-site, and the A-site is now open and ready to accept another aminoacyl-tRNA•eEF1A•GTP complex, beginning the next cycle of elongation (reviewed in Merrick and Nyborg, 2000; Browne and Proud, 2002; Rodnina and Wintermeyer, 2009).

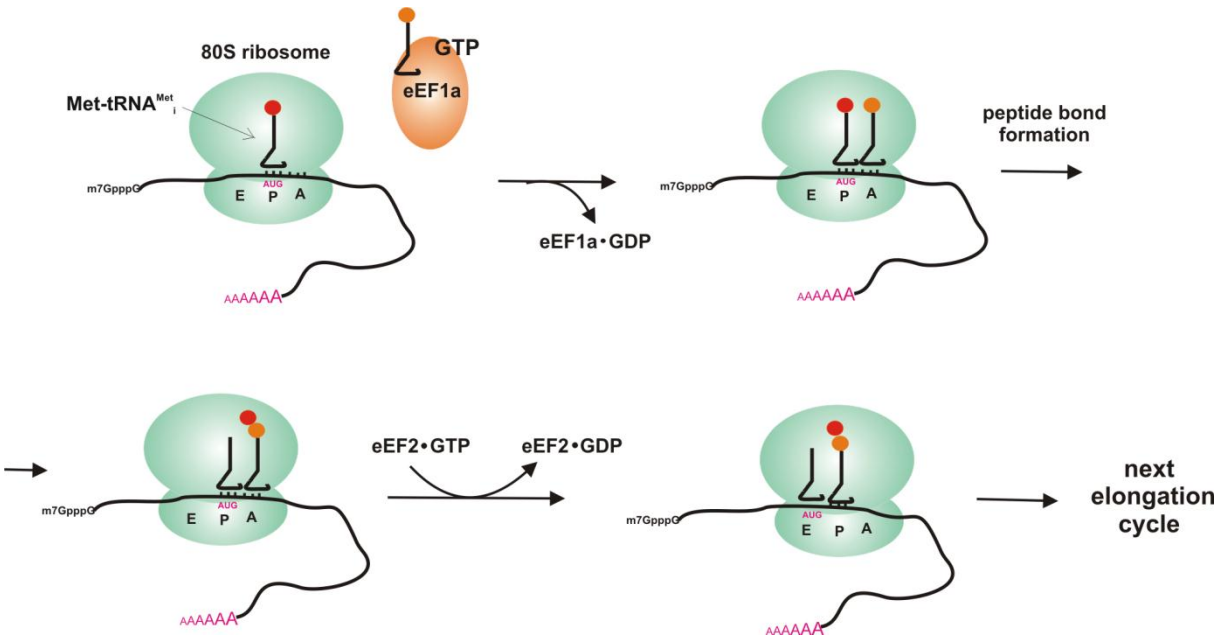

Figure 4. Cycle of peptidyl chain elongation during translation in eukaryotes. The scheme represents the factors involved and the order of events. A, P and E stand for the ribosomal A-, P-and E-sites, respectively; eEF: eukaryotic elongation factor.

2.3. Termination and recycling

The final step of translation, the termination, occurs when the ribosome reaches one of the three stop-codons: UAG, UAA or UGA. Two polypeptide chain release factors exist in eukaryotes, eRF1 and the GTPase eRF3, that are needed for this step (Figure 5). The eRF3•GTP acts together with eRF1. When the eRF1•eRF3•GTP complex reaches the ribosome, eRF1 binds the ribosomal A-site and recognizes the stop-codon. This event is made possible because the 3D structure of eRF1 mimics that of a tRNA (Song et al., 2000). GTP hydrolysis by eRF3 is a prerequisite for peptide release (Salas-Marco et al., 2004; Alkalaeva et al., 2006).

After the completed protein chain is released from the ribosome, the recycling stage of translation occurs, freeing the ribosomal subunits, the mRNA and tRNAs for another round of translation. In the current model of recycling, a protein called ABCE1 binds the eRF1-bound post-termination ribosomes and promotes their dissociation into free 60S ribosomal subunits and tRNA- and mRNA-bound 40S subunits. Further dissociation is catalyzed principally by eIF3, eIF1, and eIF1A (Pisarev et al., 2007; 2010). eIF1 mediates the release of the tRNA from the 40S subunit, while eIF3j ensures subsequent mRNA dissociation. The exact role of each factor and the mechanism are not fully understood (Pisarev et al., 2007).

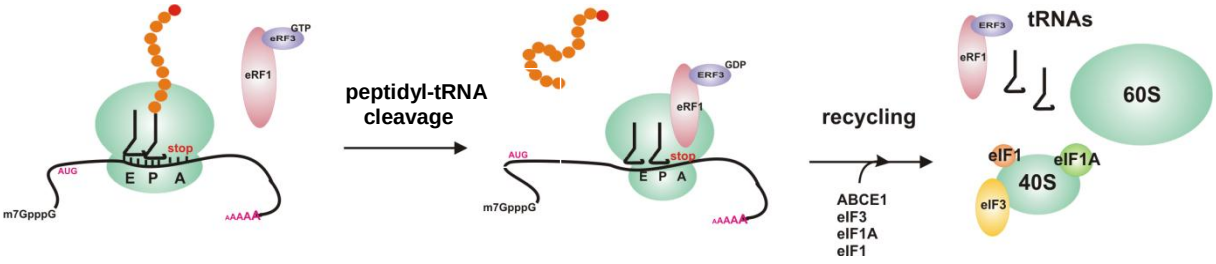

Figure 5. Termination of translation and recycling of the post-termination complexes. The scheme represents the factors involved and the order of events according to Pisarev et al., (2007; 2010). The stop codon is shown in pink and can be either UAA, UGA or UAG. A, P and E stand for ribosomal A-, P- and E-sites, respectively; eRF: eukaryotic release factor; eIF: eukaryotic initiation factor.

2.4 Stop codons - dual role in translation

Alternative ways of reading the genetic code can exist. They comprise programmed ribosomal frameshifting, ribosomal hopping and termination codon reassignment (for review, see Namy et al., 2004). I will focus on the latter case in this section.

UAG, UAA and UGA can possess alternative meanings other than signaling termination of translation. Table 4 represents known deviations from the universal genetic code. In some organelles or even organisms, stop codons have changed their meanings. For example, in mitochondria of many species, UGA encodes tryptophane (Osawa et al., 1992). UAA and UAG code for glutamine in *Acetabularia*, *Tetrahymena*, *Paramecium* and a few others (Preer et al., 1985).

Two cases are known where stop codons can be read alternatively in response to signals present in the mRNA. First, UAG can code for the rare amino acid pyrrolysine in some methanogenic archaea and in one bacterium, *Desulfitobacterium hafniense* (reviewed in Zhang et al., 2005; Zhang and Gladyshev, 2007). Second, in all three domains of life but not in all species, UGA can encode selenocysteine. Apart from its dual stop or selenocysteine meaning, UGA can also specify the insertion of two different amino acids within the same gene: this is the case in the ciliate *Euplotes crassus* where it can encode selenocysteine or cysteine. It was shown that the mRNA structure and the location of UGA within the gene influence the choice of the encoded amino acid in this case (Turanov et al., 2009).

Codon	Universal code	Unusual code	Occurrence
UGA	stop	Trp	Mitochondria, yeast
		Sec, stop	All three domains of life, but not in all species
		Cys, Sec, stop	*Euplotes crassus*
UAA, UAG	stop	Gln	Ciliates
UAG	stop	Pyl, stop	Methanogenic archaea

Table 4. Unusual stop codon usage. Abbreviations used: Trp: tryptophane; Sec: selenocysteine; Cys: cysteine, Gln: glutamine; Pyl: pyrrolysine.

3. The selenocysteine biosynthesis pathway

Unlike the vast majority of other amino acids, selenocysteine does not exist in the free form in vivo. It is synthesized from the seryl residue on the Ser-tRNA^{Sec}. In bacteria, a one step conversion catalyzed by selenocysteine synthase (SelA) converts the seryl moiety to

selenocysteine on the Ser-tRNA^{Sec} using monoselenophosphate as the selenium donor (reviewed in Böck, 2006). The latter is produced from selenide (its actual chemical form is poorly characterized) by a reaction catalyzed by selenophosphate synthetase (SelD) (Lacourciere, 1999).

A more complex mechanism is involved for selenocysteine synthesis in eukaryotes. It will be described here.

3.1. Structural features of the eukaryotic tRNA^{Sec}

The tRNA^{Sec} plays a dual role in selenoprotein synthesis. On the one hand, it serves as the carrier molecule for the synthesis of selenocysteine and, on the other hand, it delivers the Sec residue to the ribosome in response to a UGA Sec codon. The tRNA^{Sec} from different organisms share common structural features. Moreover, the tRNAs^{Sec} from eukaryotes or archaea are functional in bacterial systems (Baron et al., 1994; Rother et al., 2000).

The bacterial, archaeal and eukaryal tRNA^{Sec} differ from canonical tRNAs in several aspects. It is the longest known tRNA (90-100 nt compared to 76 nt in the canonical one) (Diamond et al., 1981; Hatfield et al., 1982; Amberg et al., 1993; Diamond et al., 1993; Bock, 2006). The eukaryal molecule will be discussed here. Chemical and enzymatic probing experiments revealed that it can adopt a distinctive secondary structure shown in Figure 6 (Sturchler et al., 1993).

Compared to canonical tRNAs (Figure 6), the tRNA^{Sec} has a longer D-stem (6 bp instead of 3-4 bp). This feature is important for selenocysteine biosynthesis (discussed in 3.2). The A- and T-arms are also longer – 9 bp for A and 4 bp for T (forming the 9/4 coaxial arm in the L shape) compared to 7/5 in canonical tRNAs. The tRNA^{Sec} adopts a unique 3D structure as was first revealed by chemical/enzymatic probing and sequence comparisons (Sturchler et al., 1993; Hubert et al., 1998) and later confirmed by X-ray analysis (Itoh et al., 2009). Aminoacylation of tRNA^{Sec} with the serine residue by the conventional Seryl-tRNA synthetase (SerRS) is the first step in the tRNA^{Sec}-dependent selenocysteine biosynthesis (summarized in Figure 7). Docking experiments showed that the variable arm shares the same orientation as in tRNA^{Ser}, suggesting that the SerRS interacts with both tRNAs in similar manner (Itoh et al., 2009).

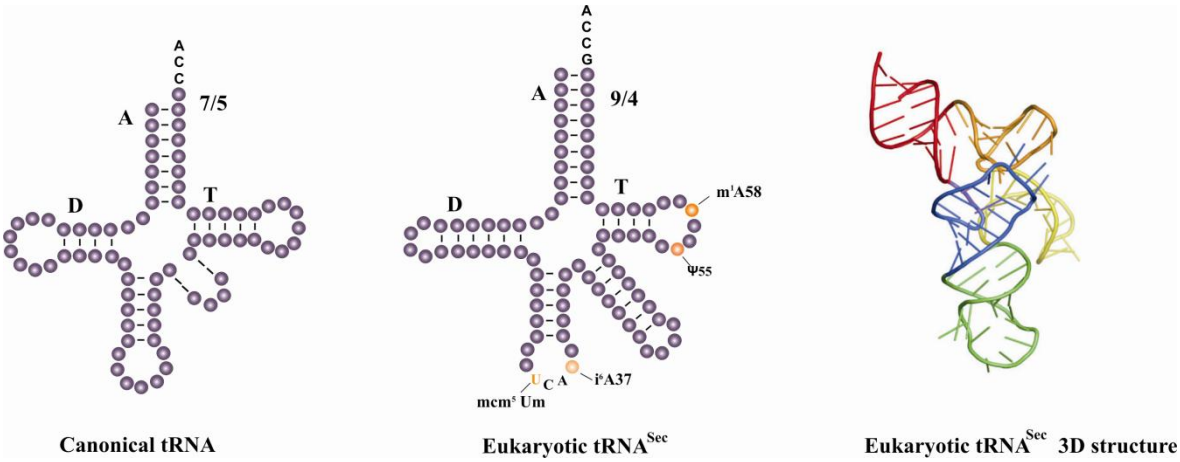

Figure 6. Secondary structure models for the eukaryotic canonical tRNA and $tRNA^{Sec}$. The right panel displays the human $tRNA^{Sec}$ crystal structure. A, T, AC and D stands for acceptor, T-, anticodon and D-stems, respectively. 7/5 and 9/4 show the number of base pairs forming the coaxial A-T arm. Modified bases in $tRNA^{Sec}$ are highlighted in orange. Indication of modified bases in the canonical tRNA was omitted. The length of the variable arm in canonical tRNAs varies (indicated by a dashed line). On the 3D structure, the acceptor stem, AD linker, D stem, anticodon arm, variable stem and T stem are colored red, purple, blue, green, yellow and orange, respectively. The 3D structure is taken from Itoh et al. (2009).

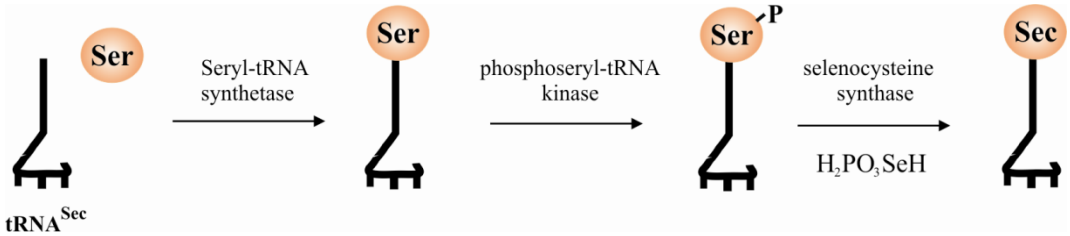

Figure 7. Summary of the selenocysteine synthesis pathway in eukarya. The $tRNA^{Sec}$ is charged with serine by the Seryl-tRNA synthetase (SerRS). Next, the seryl residue is phosphorylated by the O-phosphoseryl-$tRNA^{Sec}$ kinase, and then converted to selenocysteine by the Selenocysteine synthase (SecS) using monoselenophosphate. The monoselenophosphate (H_2PO_3SeH) selenium donor is generated from selenite or more likely selenide by a reaction catalyzed by Selenophosphate synthetase 2 (SPS2).

The $tRNA^{Sec}$ contains only 4 modified nucleotides. These nucleotides are 5-methylcarboxymethyluridine (mcm^5U) at the wobble position 34 (or its 2'-O-methylated version mcm^5Um), N6-isopentenyladenosine at position 37, pseudouridine at position 55 and 1-methyladenosine at position 58, as shown for the *X.laevis* $tRNA^{Sec}$ (Diamond et al., 1993; Sturchler et al., 1994). Strikingly, the $tRNA^{Sec}$ lacks the dihydrouridine (in D-loop) and ribothymidine (in T-loop) modifications occurring in almost all tRNAs.

3.2 From serine to selenocysteine residue

After $tRNA^{Sec}$ aminoacylation, the seryl residue is phosphorylated by the O-phosphoseryl-$tRNA^{Sec}$ kinase (PSTK) to yield the O-phosphoseryl-$tRNA^{Sec}$ (Sep-$tRNA^{Sec}$) (Carlson et al., 2004). PSTK discriminates between the Ser-$tRNA^{Ser}$ and Ser-$tRNA^{Sec}$ by recognizing the length and unique secondary structure of the Ser-$tRNA^{Sec}$ D-stem and the discriminatory base G73 (Wu and Gross, 1994; Sherrer et al., 2008).

The final step of selenocysteine formation is the conversion of the Sep-$tRNA^{Sec}$ to Sec-$tRNA^{Sec}$, catalyzed by the O-phosphoseryl-tRNA:selenocysteinyl-tRNA synthase (SepSecS) (Yuan et al., 2006; 2007). As revealed from the crystal structure of the human SepSecS-(Sep-$tRNA^{Sec}$) complex (Figure 8), this enzyme forms a tetramer and binds two Sep-$tRNA^{Sec}$ molecules via their unique 13-bp coaxial acceptor and T-stems, forming the 9/4 arm (Palioura et al., 2009). SecSepS catalyzes the conversion of the phosphoseryl moiety to a selenocysteyl group by a pyrodoxal phosphate dependent mechanism. Selenophosphate is used as the selenium donor (Xu et al., 2007). It is synthesized from selenite or selenide by selenophosphate synthetase 2 (SPS2) which, interestingly, is a selenium-containing protein itself (Guimaraes et al., 1996).

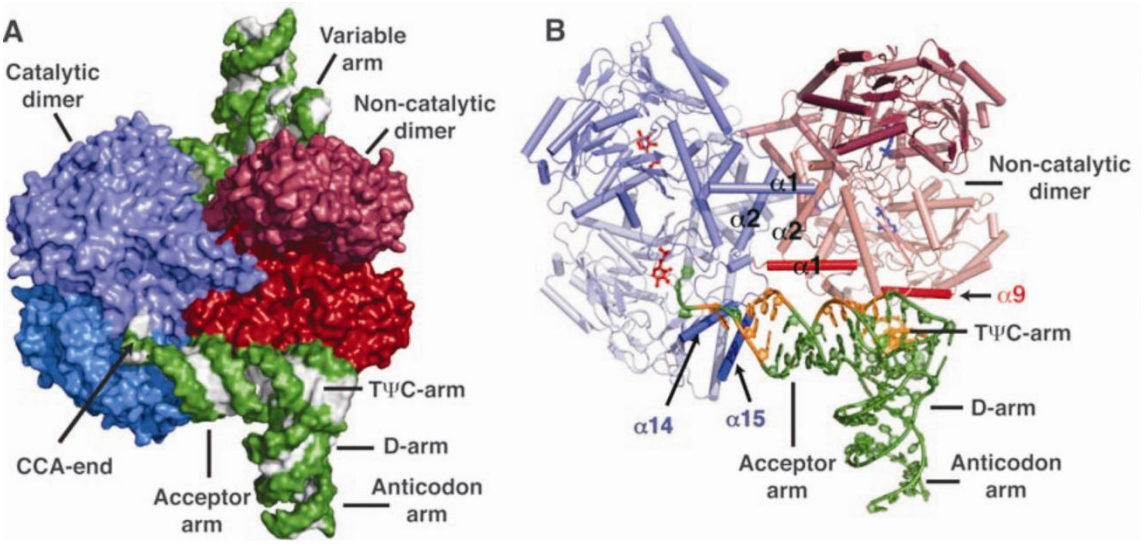

Figure 8. Structure of the human SepSecS in complex with the unacylated $tRNA^{Sec}$. (A) Surface representation of the physiological complex of SepSecS with $tRNA^{Sec}$. The subunits of the catalytic dimer are dark and light blue, those of the noncatalytic dimer are dark and light red; the backbone and the bases of $tRNA^{Sec}$ are green and gray, respectively. (B) The regions of $tRNA^{Sec}$ that interact with SepSecS are shown in orange; the rest of the molecule is in green. One $tRNA^{Sec}$ molecule is shown for clarity. The figure is taken from Palioura et al. (2009).

4. Selenoprotein synthesis

As selenocysteine is encoded by a UGA codon, normally read as a stop, the translation machinery has to be modified to incorporate this amino acid. The bacterial selenoprotein synthesis mechanism has been extensively studied and well established by the Böck's group (reviewed in Böck, 2006). However, this is not the case for archea and eukaryotes. All the known *cis*- and *trans*- acting factors involved in this process in eukaryotes will be described and an overview on hypothetical mechanisms for selenoprotein synthesis will be given here.

4.1. The SECIS element

All eukaryotic selenoprotein mRNAs contain a stem-loop in the 3'-UTR, called SECIS for **SE**leno**C**ysteine **I**nsertion **S**equence (Berry et al., 1991) that is mandatory for selenoprotein synthesis. The distance between UGA and SECIS can vary from 50 to several thousands nucleotides, depending on the identity of the selenoprotein mRNA. Almost all selenoproteins contain a single selenocysteine. The unique case of the selenoprotein P is striking, containing up to 28 Sec residues (10 in humans) (Lobanov et al., 2009). Its mRNA contains only two SECIS elements: the 3' proximal one is responsible for decoding the first UGA while the 3' distal one enables decoding of the others (Stoycheva et al., 2006). Always harbored by the 3'UTR, a SECIS element was surprisingly identified in the coding frame and at the very 3'end of the mRNA in the avian fowlpox virus-encoded GPx4 selenoprotein. It was shown experimentally to support synthesis of a full-length Sec-containing GPx4 in mammalian cells (Mix et al., 2007).

The sequence of SECIS elements is not highly conserved among selenoprotein mRNAs and across species, but its 2D structure is. Structure probing and site-directed mutagenesis experiments resulted in the conserved secondary structure model presented in Figure 9 (Walczak et al., 1996; 1998, Grundner-Culemann et al., 1999, Fagegaltier et al., 2000a). In fact, two forms of SECIS elements exist (Figure 9). SECIS form 2 has a shorter apical loop and an additional helix compared to SECIS form 1. Using structure based sequence alignments and an in-house developed bioinformatic tool, Chapple and colleagues showed form 2 to be more widespread than form 1 - only 62 form 1 versus 224 form 2 SECIS elements were found among all the selenoprotein mRNAs known at the time of the analysis

(Chapple et al., 2009). Despite the difference, they both appear to function similarly (Grundner-Culemann et al., 1999).

Two conserved regions that are absolutely necessary for Sec incorporation can be distinguished (Figure 9, highlighted in red),. The first one is located at the foot of helix 1 and is composed of a quartet of non-Watson-Crick base pairs where the central GA/AG tandem adopts the "sheared" conformation (Walczak et al, 1996; 1998).

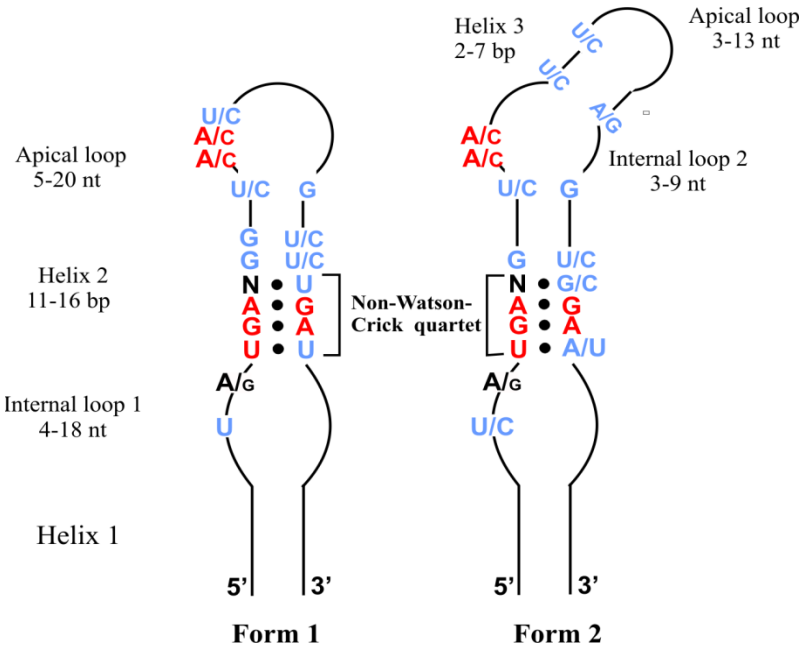

Figure 9. Secondary structure of the two forms of SECIS elements according to Walczak et al. (1996; 1998); Grundner-Culemann et al. (1999); Fagegaltier et al. (2000a). Nucleotides essential for Sec incorporation are shown in red. Novel conserved residues found by SECISaln are shown in blue (Chapple et al., 2009). Figure adapted from Chapple et al., (2009). The lengths of helices and loops are indicated.

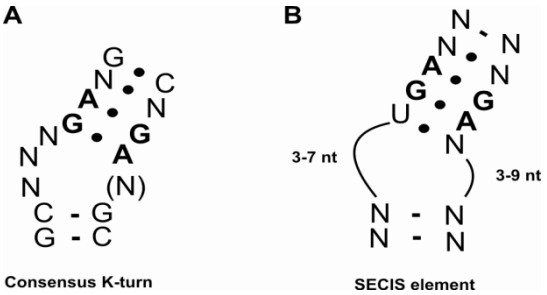

Figure 10. Secondary structures of a consensus K-turn motif (A) and K-turn like motif present in the SECIS RNA (B). The non-Watson-Crick base pairs GA/AG are in bold. The K-turn motif in (A) is adapted from Klein et al. (2001).

The quartet forms a binding site for the SECIS binding protein 2 (SBP2) described in detail in paragraph 4.3.2. The GA/AG tandem can also be found in other RNAs, for example in mRNAs, ribosomal RNAs, ribozymes, snRNAs and snoRNAs (Vidovic et al., 2000; Chao and Williamson, 2004; Moore et al., 2004). It is part of a recurrent motif called the "kink-turn" or K-turn motif since it causes a sharp kink in the phosphodiester backbone (Vidovic et al., 2000; Klein et al., 2001). This RNA structural motif mediates RNA tertiary structure interactions and is known to form RNA-protein binding sites (Vidovic et al., 2000; Chao and Williamson, 2004; Moore et al., 2004). It can be described as a two-stranded, helix-internal loop-helix motif comprising about 15 nucleotides, characterized by base stacking, the presence of a tandem of GA/AG sheared base pairs, and a protruding residue accommodated by a protein pocket (Klein et al., 2001). By comparing the motif present in the SECIS element with the consensus K-turn motif (Figure 10), it was proposed that the SECIS RNA adopts a "K-turn like" motif (Allmang and Krol, 2006a). The importance of this region for selenoprotein synthesis was confirmed by the observation that a patient, carrying the single U to C mutation in the conserved U of the SECIS quartet in the selenoprotein N mRNA, presents with a severe muscular dystrophy (Allamand et al., 2006). Since the K-turn is a widespread RNA-protein interaction motif, there should exist additional factors to define the specificity of interaction for each RNA-protein complex. In fact, it was shown that SBP2 cannot interact with other K-turn containing RNAs such as the U4 snRNA or the U3 snoRNA (Cléry et al., 2007). SELEX and site-directed mutagenesis experiments established that the large internal loop 1, that is not found in other K-turn RNAs, and a series of Watson-Crick base-pairs stabilizing helix 2 (Figure 9), are of high importance for the SBP2-SECIS interaction (Fletcher et al., 2001; Cléry et al., 2007).

The second conserved region is the single-stranded AAR sequence in the apical loop of SECIS form 1 or in internal loop 2 of SECIS form 2. As an exception, some selenoprotein mRNAs contain the sequence AAC or only Cs, or variations around the theme, as in the mammalian selenoprotein M mRNA (Korotkov et al., 2002). Site-directed mutagenesis experiments showed that the A/C nucleotides were important for Sec incorporation *in vivo* (Berry et al., 1993). To date, the function and the interaction partner remain unknown and it is clearly not SBP2 (Fletcher et al., 2001; Cléry et al., 2007; Beniaminov and Krol, unpublished data). More recently, new SECIS regions important for Sec incorporation *in vitro* and *in vivo* were determined. Having analyzed all the 26 human SECIS elements, Latrèche and

colleagues concluded that the presence of a G-C base pair in helix 2 (on top of the quartet) and a U in the 5' side of the internal loop of the SECIS (Figure 9) are predictive of a weak UGA decoding activity. These nucleotides were only found in the SECIS elements of glutathione peroxidases 3 and 6, selenoprotein H, selenoprotein O, selenoprotein S and thioredoxin reductase 3 mRNAs (Latrèche et al., 2009). Thus, SECIS elements could modulate the efficiency of selenocysteine incorporation. Additionally, conservation of these nucleotides was independently predicted by a bioinformatic approach using SECISaln, a web-based tool designed to provide structure-based alignments of SECIS elements (Chapple et al., 2009).

4.2. The SRE element

The Selenocysteine Redefinition Element (SRE) is the second RNA motif used for UGA decoding. It was found by comparative sequence analysis by two independent research groups (Howard et al., 2005; Pedersen et al., 2006). The SRE is a stem-loop located at the conserved distance of 6 nucleotides downstream of the UGA Sec codon. SREs were found only in the mRNAs of selenoprotein N, selenoprotein H, selenoprotein O, selenoprotein T mRNAs, and in that of SPS2, the selenophosphate synthetase. Since the SRE is contained in the open reading frame, there is little possibility for sequence conservation between different selenoprotein mRNAs. The 2D structure is maintained between SREs of the same selenoprotein mRNA across species, sometimes through 1-2 compensatory base changes, but the 2D structure is clearly not strictly identical between SREs of different selenoprotein mRNAs (Howard et al., 2005). Experimental data confirmed the functionality of this element: (i) it was shown that it has a stimulatory effect on UGA Sec readthrough in an *in vitro* translation system from rabbit reticulocyte lysate. Nevertheless, efficient selenocysteine insertion was still dependent on the presence of the SECIS (Howard et al., 2007); (ii) four point mutations leading to the SEPN1-related myopathy were found in the selenoprotein N SRE element. One of them weakens the secondary structure of the SRE by abolishing a G-C base pair, leading to a decrease in Sec incorporation and selenoprotein N levels (Maiti et al., 2008).

Interestingly, selenoproteins H and O contain a SECIS element with the lowest efficiency in Sec incorporation (Latrèche et al., 2009). Thus, the presence of an additional

stimulating element such as the SRE may help to increase the selenoprotein synthesis level if the SECIS is weak.

4.3. Protein factors involved in selenoprotein synthesis

Selenoprotein synthesis is regulated by a number of protein factors. The elongation factor EFSec, the SECIS binding protein SBP2 and ribosomal protein L30 described below were shown to be direct participants in UGA Sec decoding. This paragraph also briefly describes three proteins, eIF4a3, nucleolin and SECp43, that may act as regulators of selenoprotein synthesis.

4.3.1. The elongation factor EFSec

UGA Sec decoding in bacteria, archaea and eukarya employs a specialized elongation factor that binds specifically the Sec-tRNA^{Sec}. The bacterial factor is called SelB. It is the homologue of the general elongation factor EF-Tu but it possesses a C-terminal extension (ref Böck). SelB interacts by its N-terminal domain with the Sec-tRNA^{Sec} (in a manner similar to the EF-Tu-tRNA interaction) while the interaction with the bacterial SECIS element is provided by the C-terminal domain (reviewed in Böck, 2006). In eukaryotes, only the murine elongation factor was characterized to date and called EFSec (Fagegaltier et al., 2000b; Tujebajeva et al., 2000). Two functional domains of these factor can be distinguished. The N-terminal part is homologous to the canonical elongation factor EF1A and binds the Sec-tRNA^{Sec} (Fagegaltier et al., 2000b). However and in sharp contrast with SelB, EFSec does not interact specifically with the SECIS element (Fagegaltier et al., 2000b; Tujebajeva et al., 2000). Another group showed the existence of an EFSec-SBP2-SECIS tertiary complex formation *in vitro* using electrophoretic mobility shift essay (Donovan et al., 2008). Moreover, it was shown that EFSec co-immunoprecipitated with SBP2 from mammalian cells overexpressing both proteins, and that the EFSec-SBP2 complex formation *in vivo* was dependent on the presence of the Sec-tRNA^{Sec} (Tujebajeva et al., 2000). Using a co-immunoprecipitation approach, the C-terminal part of EFSec was found to interact with SBP2 (Zavacki et al., 2003). Thus, the N-terminal domain of EFSec provides tRNA binding, the C-terminal part establishing protein-protein interactions in either a Sec-tRNA^{Sec}- or SECIS-dependent manner.

4.3.2. SBP2

SBP2 is one of the key players in the synthesis of selenoproteins. It binds the SECIS element and interacts with the ribosome (Copeland et al., 2001; Takeuchi et al., 2009). Indeed, patients carrying mutation(s) in the SBP2 gene display numerous pathologies such as abnormal thyroid metabolism, delayed bone maturation, photosensitivity and/or congenital myopathy since the synthesis of all selenoproteins is disordered (Dumitrescu et al., 2005; Azevedo et al., 2010; Schoenmakers et al., 2010).

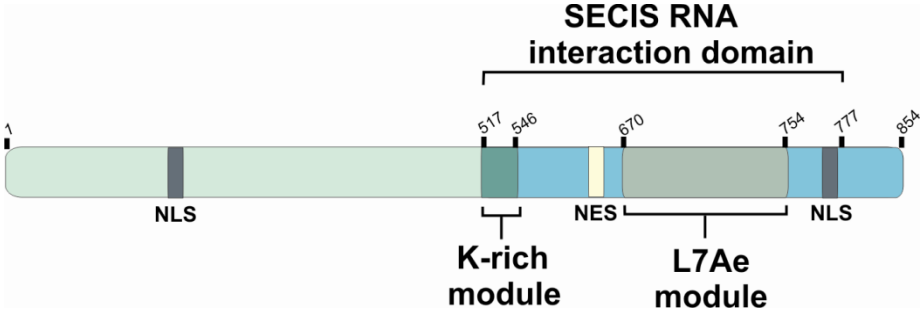

Figure 11. Schematic representation of the mammalian SBP2. The SECIS-interaction domain is shown by a bracket. The K-rich and L7Ae modules are depicted in grey-blue and light grey, respectively (Takeuchi et al., 2009). The Nuclear Localization Signal (NLS) and Nuclear Export Signal (NES) are shown in dark grey and beige, respectively, according to Papp et al. (2006).

SBP2 was only characterized in humans, rat and *D. melanogaster* (Copeland and Driscoll, 1999; Copeland et al., 2000; Lescure et al., 2002; Takeuchi et al., 2009). The mammalian SBP2 is approximately 850 amino acid long and can be roughly divided into two domains. The N-terminal two-thirds (aa 1-513) form the N-terminal domain whose function is still unknown. Moreover, it was shown to be dispensable for selenoprotein synthesis *in vitro* (Copeland et al., 2000). The SBP2 amino acid sequence was predicted to contain Nuclear Localization (NLS) and Export (NES) Signals (de Jesus et al., 2006; Papp et al., 2006)). One NLS resides in the N-terminal domain, the other NLS and the single NES being located in the C-terminal domain. SBP2 shuttling between the cytoplasm and the nucleus was observed (Papp et al., 2006).

The C-terminal domain of SBP2 is necessary for the interaction with the SECIS RNA and the ribosome (Copeland et al., 2001; Caban et al., 2007; Takeuchi et al., 2009). It comprises a conserved module called L7Ae (Figure 11) also found in other proteins such as ribosomal proteins L7Ae (hence the name) and L30, and in the core proteins of sn- and snoRNPs (reviewed in Boulon et al., 2008). Moreover, the C-terminal domain contains a conserved lysine-rich module called either the bipartite RNA binding domain, selenocysteine insertion domain (SID) or K-rich, according to authors (Bubenik et al., 2007; Donovan et al., 2008; Takeuchi et al., 2009). Both the L7Ae and K-rich motifs contribute to the SECIS-SBP2 interaction. In the absence of the crystal structure of the SBP2-SECIS complex, the SBP2-SECIS interaction was modeled based on a structure-guided mutagenesis strategy (Allmang et al., 2002) using the crystal structure of the 15.5kD-U4 snRNA complex **(**Vidovic et al., 2000**),** since high amino acid sequence similarity exist between the RNA binding domains of SBP2 and the 15.5kD proteins, and both RNAs contain K-turn motifs (Figure 12).

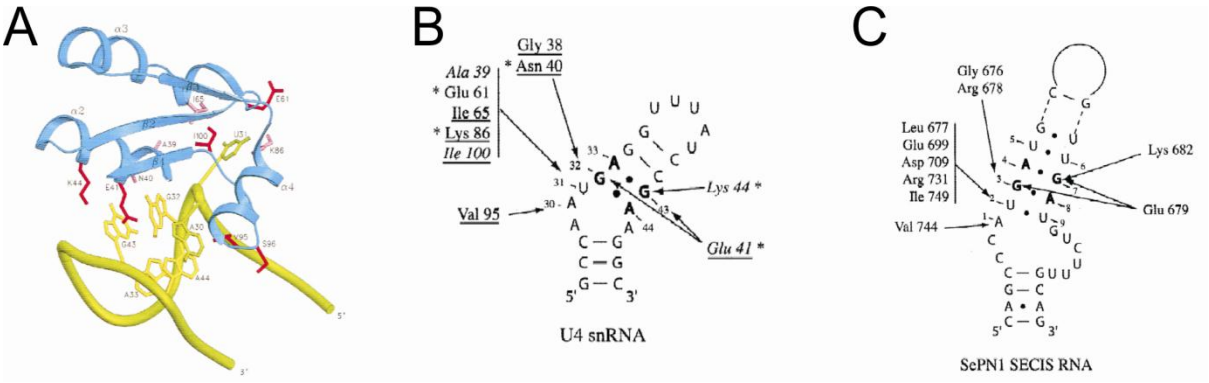

Figure 12. SBP2-SECIS interaction model. (A) The 3D structure of the 15.5kD-U4 snRNA complex, according to Vidovic et al. (2000). (B) Scheme of the RNA-protein interactions in the 15.5kD-U4 snRNA complex, adapted from the 3D structure. (C) Putative contacts between SBP2 and the selenoprotein N SECIS element (Allmang et al., 2002). Only a part of helices 1 and 2 is displayed. The G-A sheared base pairs are in bold.

The second module, called K-rich by our laboratory because of the relative abundance of lysine residues (amino acid residues 516-544 plus the 535-IILKE-539 sequence), is essential for SBP2 to recognize both forms of SECIS elements. In fact, *Drosophila* SBP2 (dSBP2) that lacks the IILKE sequence in the K-rich module but has SVRVY instead, cannot bind SECIS form 1. This is consistent with *Drosophila* selenoprotein mRNAs having form 2 SECIS exclusively. Swapping the 5 amino acids led to the ability of dSBP2 to bind both

forms of SECIS but prevented the mammalian protein to bind form 1, in vitro (Takeuchi et al., 2009).

Despite several attempts using various truncated forms of SBP2 and different types of SECIS elements, the laboratory has been so far unable to obtain crystals of the SBP2-SECIS complex, or even of the individual components. One likely interpretation arose from both computer predictions and biophysical studies that established that a high part of the recombinant SBP2 protein (both the E.coli and the insect-cell produced forms) is intrinsically disordered, with the exception of the L7Ae module (Oliéric et al., 2009).

Another function of SBP2 is to communicate with the ribosome during Sec incorporation (Copeland et al., 2001; Caban et al., 2007). In fact, SBP2 was shown to co-sediment with ribosomes in glycerol gradients and, furthermore, it co-sedimented with the purified 28S rRNA. It was thus proposed that SBP2 could interact with the ribosome via one of the seven K-turns of the 28S rRNA (Copeland et al., 2001). The L7Ae module and/or the SBP2 region 520-545 were predicted to contribute to the ribosome-SBP2 interaction (Caban et al., 2007). More recently, the human SBP2 was shown to bind the human 60S and not the 40S ribosomal subunit *in vitro* (Takeuchi et al., 2009). Moreover, the use of SBP2 mutants established that amino acid residues 529-PTSLKK-534 are of prime importance for this interaction (Takeuchi et al., 2009). However, the binding site of SBP2 on the 60S subunit was unknown at the time when I started my thesis work. This constituted part of my experimental work that will be described in "Results", section 7.

4.3.3. L30

Ribosomal protein L30 was proposed to be a participant in selenoprotein synthesis (Chavatte et al., 2005). Based on: (i) the SECIS element crosslinked to L30 in rat tissues; and (ii) L30 competition with SBP2 for SECIS binding *in vitro*, L30 was proposed to displace SBP2 on the SECIS element during selenocysteine incorporation (Chavatte et al., 2005).

In fact, L30 is a ribosomal protein specific to eukaryotes and archaea. L30 was found located at the interface of the 60S subunit and involved in formation of intersubunit bridges, according to the X-ray structure of the yeast 80S ribosome, and cryo-EM structures of the wheat and canine 80S ribosome, and the yeast and wheat 60S subunits (Halic et al., 2005;

Chandramouli et al., 2009; Ben-Shem et al., 2010; Arnache et al., 2010). L30 was predicted to belong to the L7Ae protein family (Koonin et al., 1994), which was later confirmed by NMR spectroscopy and crystal structures of the protein alone or in an RNA-protein complex (Chen et al., 2003; Chao and Williamson, 2004). It can be found not only in a ribosome-bound form but also in the free state, and it is known that yeast L30 can regulate its own translation by binding to a kink-turn at the 5' splice site of its pre-mRNA, thus inhibiting splicing (Macias et al., 2008).

5. Putative regulators of selenoprotein synthesis

Studies using crosslinking of SECIS RNA probes with proteins contained in eukaryotic cell extracts led to identification of two new SECIS-interacting proteins, eIF4a3 and nucleolin. These proteins possibly act as regulators of selenoprotein expression.

5.1. eIF4a3

eIF4a3 bears 70% amino acid identity with initiation factor eIF4a1, but it does not take part in translation initiation (Li et al., 1999). It was shown to be a component of the exon junction complex (EJC) and to participate in the non-sense mediated mRNA decay (NMD) (Chan et al., 2004; Ballut et al., 2005). Recently, eIF4a3 was shown to interact with the internal loop of the SECIS (form 1) of glutathione peroxidase 1 (GPx1) but not with that of the glutathione peroxidase 4 (GPx4) which is form 2 (Budiman et al., 2009). Moreover, eIF4a3 can suppress expression of GPx1 under selenium-limiting conditions. The same authors showed that the binding site of this factor on the SECIS element is located in the vicinity or even overlaps the SBP2 binding site. eIF4a3 can thus prevent the SECIS-SBP2 interaction and therefore the expression of a subset of selenoproteins.This work is consistent with previous findings showing that the GPx1 activity was reduced to 1% in the liver and 4-9% in heart, kidney and lung tissues in rats under a low selenium diet compared to selenium-sufficient rats. In contrast, the GPx4 activity was 25-50% lower in these tissues and absolutely not effected in testes (Lei et al., 1995). Interestingly, it is not only the SECIS form that can have an influence on selenoprotein synthesis. A tissue-specific selenoprotein expression has also been described. During selenium deprivation in the diet of rodents, the amount of Se is reduced first in liver and kidney, while brain and testes retained most of the selenium content

(Behne et al., 1988; Hill et al., 1992). This phenomenon was called selenoprotein hierarchy (Sunde et al., 2009). Further studies are needed to understand if eIF4a3 is indeed the factor that determines the hierarchy of selenoprotein expression.

5.2. Nucleolin

Nucleolin is a multifunctional protein participating in various cellular processes such as transcription, small RNA modification, ribosome biogenesis, etc. (for review see for example, Ginisty et al., 1999; Mongelard and Bouvet, 2007). Nucleolin was also shown to bind a subset of selenoprotein mRNAs *in vitro* and *in vivo* (Wu et al., 2000; Squiers et al., 2007). It was proposed to regulate selenoprotein synthesis as an enhancer of the expression of the subset of selenoproteins (Miniard et al., 2010). Futher studies are needed to identify the mechanism of nucleolin action.

5.3. SECp43

SECp43 was identified as a $tRNA^{Sec}$ interactant in a pull-down assay (Ding and Grabowski, 1999). Knockdown of SECp43 by siRNA demonstrated that this protein is required for ribose methylation at Um34 of $tRNA^{Sec}$ and it acts as enhancer of selenoproteins expression. Ananlysis of the subcellular localization of SECp43 suggests that it may regulate the nucleocytoplasmic shuttling of the SepSecS-Sec-$tRNA^{Sec}$ complex (Xu et al., 2005; Small-Howard et al., 2006).

6. Selenoprotein mRNP assembly and selenoproteins synthesis mechanism

6.1.Selenoprotein mRNP assembly

Selenoprotein mRNAs contain an in-frame UGA codon that makes them potential targets for NMD. Indeed, according to authors, these mRNAs are subjected to NMD under low selenium conditions (Moriarty et al., 1998; Weiss and Sunde, 1998; Sun et al., 2001). Moreover, as discussed in section 3.3.2, SBP2 contains an L7Ae domain much as the core protein (15.5 kD) of some of the sn- and snoRNPs, and possibly undergoes nucleocytoplamic shuttling (Allmang et al., 2002; Papp et al., 2006). These observations led to the hypothesis that, to circumvent the NMD pathway, selenoprotein mRNPs should be assembled in the

nucleus (de Jesus et al., 2006; Small-Howard et al., 2006), obeying the same rules as for sno- and snRNP formation (Boulon et al., 2008). Indeed, a study performed in our laboratory identified a complex assembly machinery linked to the protein chaperone Hsp90 that assembles RNPs of the L7Ae family, such as box C/D and H/ACA snoRNPs, U4 snRNP and selenoprotein mRNPs (Boulon et al., 2008; Zhao et al, 2008). This machinery is composed of a co-chaperone complex comprising Rvb1-Rvb2, Spagh and Pih1 (the R2TP complex) associated to the Hsp90 chaperone, and the adaptor protein Nufip (Figure 13). Nufip binds the L7Ae family of proteins including SBP2 and can tether them to other proteins of immature RNPs. It also links them to the Hsp90 chaperone complex. These associations are very likely necessary for the stable association of SBP2 on the SECIS RNA (Boulon et al., 2008). Altogether, these results imply that SBP2 has functional similarities with other L7Ae proteins during RNP assembly, and that the interaction with Nufip leads to recruit additional core protein(s) not identified yet.

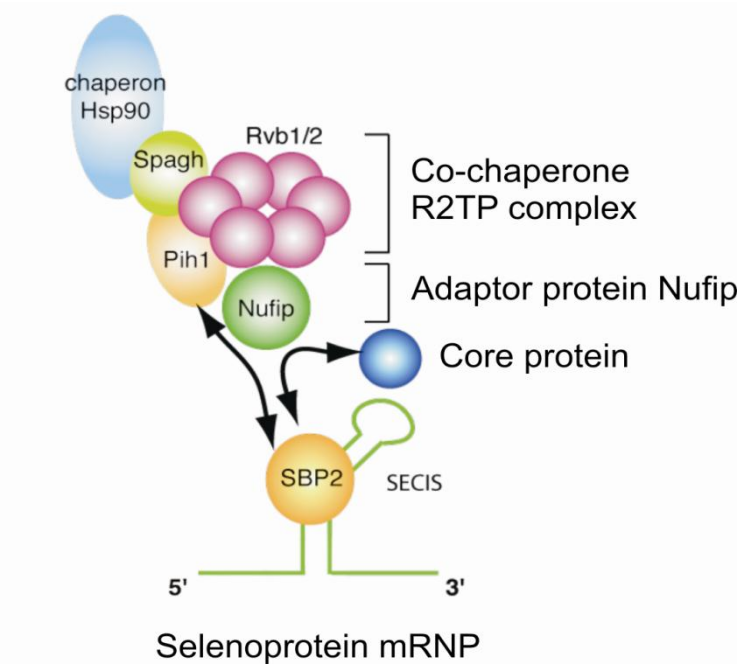

Figure 13. Assembly of selenoproteins mRNP, according to Boulon et al. (2008). The Hsp90 co-chaperone components are represented in blue, orange, mustard and violet. The adaptor protein Nufip is shown in green. An arrow between SBP2 and Pih1 indicates an experimentally demonstrated interaction. The curved arrow points to interaction with putative core proteins yet to be discovered. The figure is taken from Allmang et al. (2009).

6.2. Two models describing the selenocysteine insertion mechanism

So far, no unifying model could be proposed for the molecular mechanism of selenocysteine insertion in eukaryotes. Two models describing this process have however been proposed, based on the current experimental knowledge, but still they do not explain in detail the mechanistic issues.

To answer the question of how can the complex that assembles at the SECIS element facilitate Sec incorporation, two models were proposed (Figure 14). In the first one (Figure 14A), SBP2 travels with the ribosome and, once the ribosome has reached a UGA Sec codon, SBP2 interacts with the SECIS via its L7Ae domain. Subsequntly, SBP2 recruits the EFSec-Sec-tRNA^{Sec}-GTP complex to bring it the ribosomal A-site (Donovan et al., 2008).

According to the second model (Figure 14B), SBP2 is already bound to the SECIS element. Prior to UGA Sec decoding, it binds the EFSec-Sec-tRNA^{Sec}-GTP complex. During UGA Sec decoding, L30 displaces SBP2 on the SECIS element, thus promoting a slight change in the SECIS K-turn conformation which in turn leads to delivery of the Sec-tRNA^{Sec} to the ribosomal A-site.

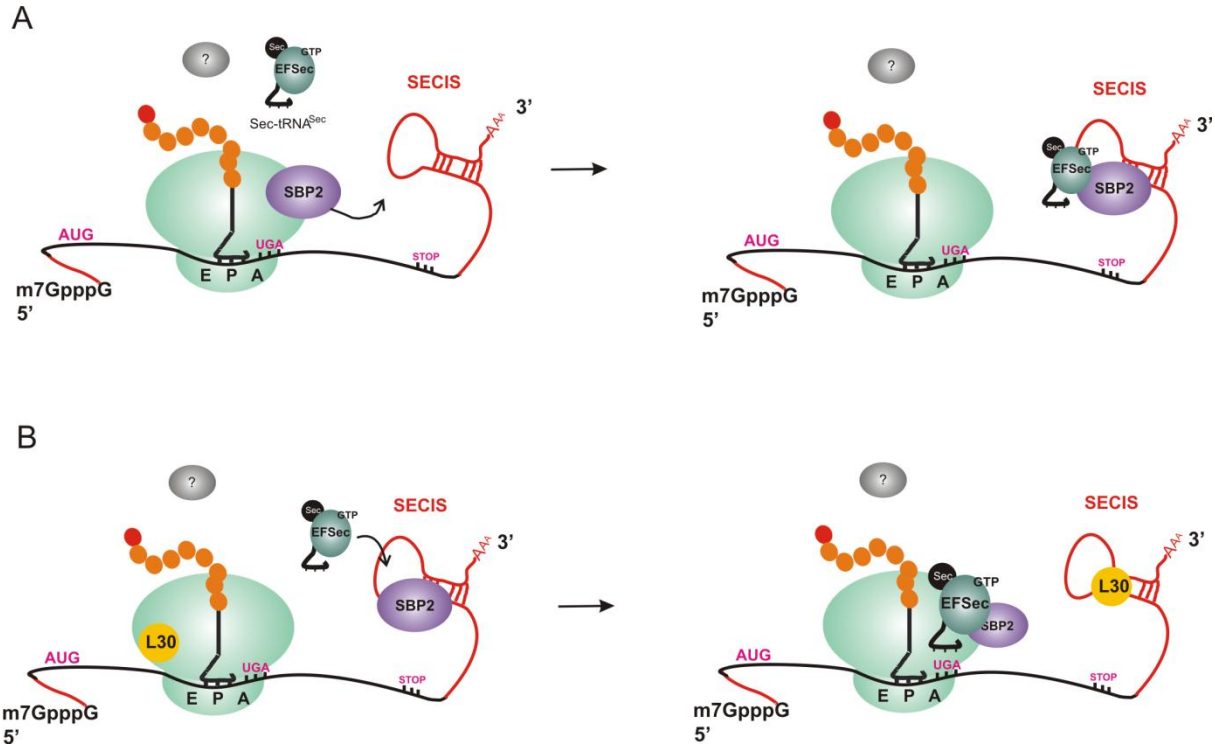

Figure 14. Two models of UGA Sec decoding. (A) according to Donovan et al. (2008). (B), according to Chavatte et al. (2005). The mRNA coding region is shown in black, the 5' and 3' non-coding regions in red. Protein factors that take part in UGA Sec decoding are indicated. A, P and E stand for the ribosomal A-, P- and E-sites, respectively. The question mark indicates that more factors can be implicated in this process.

Nevertheless, many aspects are still to be unveiled. First of all, it is unknown whether there are more participants in selenoprotein synthesis. Also, it still remains elusive of what prevents the termination factors eRF1/3 to bind the ribosomal A-site at the UGA Sec codon. How does decoding occur when several selenocysteine codons (up to more than 20) are present in an mRNA, as is the case for selenoprotein P? Obviously, not all the secrets of selenoprotein synthesis are revealed.

7. Objectives of the thesis.

Despite a lot of effort devoted to the investigation of the selenocysteine insertion mechanism, many aspects are still unknown. Gaining information on this mechanism is complicated by the absence of crystal structures of SBP2 and the SECIS element. Therefore, the aim of the project was to bring new insight into the mechanism of selenocysteine insertion in mammals. More precisely, we were interested in obtaining a detailed understanding of how and where the SECIS and SBP2 interact with the ribosome, with particular focus on their molecular environment. Several strategies were taken to attain this goal:

(1) designing short, synthetic selenoprotein mRNAs, containing all elements necessary to support selenocysteine insertion in *in vitro* translation system; derivatize them for cross-linking experiments;

(2) studying the components of the translation machinery which interact with the SECIS element in the course of selenoprotein mRNA translation, using the set of synthetic mRNAs;

(3) studying the molecular environment of SBP2 on the human ribosome using cross-linking approaches.

The results obtained are described in the following sections “Results” and “General discussion”.

Part II

Results

Results

1. Experiments strategy

In the first place, we aimed at studying the interactions between the SECIS and the components of the translational apparatus at different steps of translation. This can be done in a reconstructed ribosome•selenoprotein mRNA complex mimicking a particular translation step. In the synthetic mRNA to be used, a selenocysteine UGA codon is placed 3' adjacent to the start codon AUG. With the use of different inhibitors of translation, it will be possible to direct the UGA Sec codon to the A-site of the ribosome and to obtain various types of ribosomal complexes with the occupied A-site. The whole RNA, or the SECIS only, may contain photoreactive group(s) introduced either statistically or at a chosen position. This enables to crosslink the SECIS to the possible interaction partners at the time of UGA decoding (Figure 15). Rabbit reticulocyte lysate was chosen as an *in vitro* translation system as it contains all necessary factors to incorporate Sec in response to UGA codon, except for SBP2 that should be added to the reaction mixture (Mehta et al., 2004).

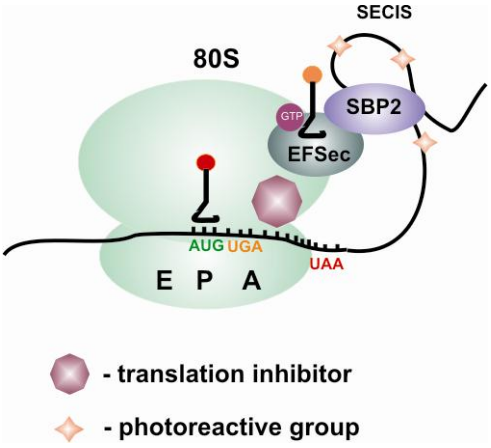

Figure 15. Schematic representation of a ribosomal complex with a synthetic selenoprotein mRNA. AUG (start), UGA (Sec) and UAA (stop) codons are shown in green, orange and red, respectively. Known factors that take part in Sec incorporation are indicated. A, P and E: tRNA binding centers.

2. Synthetic mRNA design

Two sets of synthetic selenoprotein mRNAs were obtained, differing in the composition of their 5'- and partly 3'-UTRs. Such an mRNA consists of a short 5' UTR (a natural one taken from the ribosomal protein S14 mRNA (Genbank reference BC001126.1) or an A-rich 5'UTR (Bulygin et al., 2005)) followed by the AUGUGAUUCUUCUUA sequence encoding the tetrapeptide Met-Sec-Phe-Phe (or Met-Phe-Phe-Phe for the control mRNA), a

spacer sequence (either from the rat GPx1 mRNA or an A-rich sequence), and a SECIS element (arising from either the rat Gpx1, rat GPx4 or human SelN mRNAs) at the 3' end.

To prepare this mRNA, a DNA template corresponding to the mRNA lacking the SECIS was first synthesized, and the corresponding RNA was obtained by T7 transcription (this RNAs are called 5'RNA in general, or 5'Sec mRNA when containing the Sec codon and 5'Phe mRNA with a Phe codon instead of Sec) (Figure 16A, B). The full-length SECIS-containing mRNA was obtained by splint-aided ligation of the mRNA to the SECIS element (Figure 16C). The SECIS element was taken from either the rat GPx1 or the human SelN genes. The resulting full-length mRNA (flRNA, or flSec mRNA when containing the Sec-codon, flPhe mRNA for the control flRNA with Phe instead of Sec) was used to synthesize cDNAs by RT-PCR. This full-length template is used to synthesize preparative amounts of flRNA by T7 transcription; this mRNA was used in toe-printing experiments, for binding experiments and also for synthesizing photoreactive derivatives with the $-N_3$ group introduced at a chosen position.

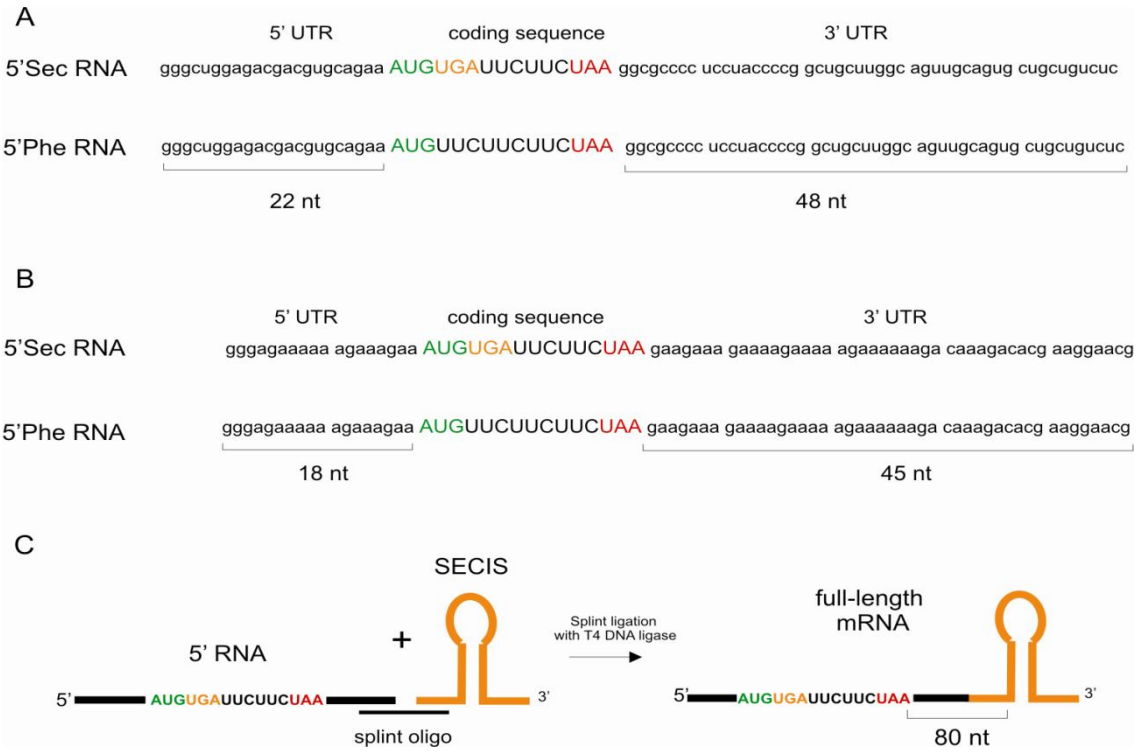

Figure 16. Synthetic mRNA synthesis. (A) Sequences of the 5'RNA, which contain natural 5' and 3' UTRs; (B) 5'RNA sequences with A-rich 5' and 3' UTRs; (C) Schematic representation of full-length mRNA synthesis (flSec in this case). The start, Sec and stop codons are shown in green, orange and red, respectively. The size of the untranslated region is indicated (nt stands for nucleotide).

Selenoprotein mRNAs are known to have a short 5'UTR (6-200 nt). The distance between the Sec codon and the SECIS element can vary from 40 to several thousands nucleotides. The distance between the Sec codon and the stop codon can also vary in the range from 3 nt to several thousands (Castellano et al., 2008). In the case of the distances between UGA/SECIS and UGA/stop codon shown in Table 5, the synthetic mRNA sequence resembles that of the human selenoprotein O (SelO) encoding mRNA (Table 5). Thus, the selenoprotein mRNAs obtained should have the appropriate UGA/SECIS and UGA/stop codon distances to support Sec incorporation. We have also taken into account sequence contexts: a G 3' to UAA creating a strong context for stop codon recognition by eRF1; a weak context for UGA recognition as a stop codon by placing a U 3' to UGA.

	Stop codon	Distance UGA-stop, nt	Distance UGA-SECIS, nt
Synthetic mRNA	UAA	6	89
SelO mRNA	UAA	6	49

Table 5. Characteristics of the synthetic selenoprotein mRNA obtained compared with the natural selenoprotein mRNA encoding the human SelO protein.

3. Assaying the synthetic mRNAs in an in vitro translation system

The applicability of the synthetic selenoprotein mRNAs to study the selenocysteine insertion mechanism was assayed by two approaches.

3.1. Toe-printing technique

The toe-printing technique was applied to ask whether the synthetic mRNAs were phased correctly, e.g. if the AUG codon was in P-site and the UGA Sec codon in the A site (the method is described in detail in Materials and Methods and in Sachs et al., 2002). We have used the following translation inhibitors: the non-hydrolysatable GTP analog GMPPNP, which inhibits translation at the stage of 48S pre-initiation complex formation; (ii) anisomycin, which provides 80S complex where transpeptidation is blocked; and (iii) emetine, which enables peptide bond formation but blocks translocation by preventing EF2-ribosome binding (Anthony and Merrick, 1992;) (Figure 17).

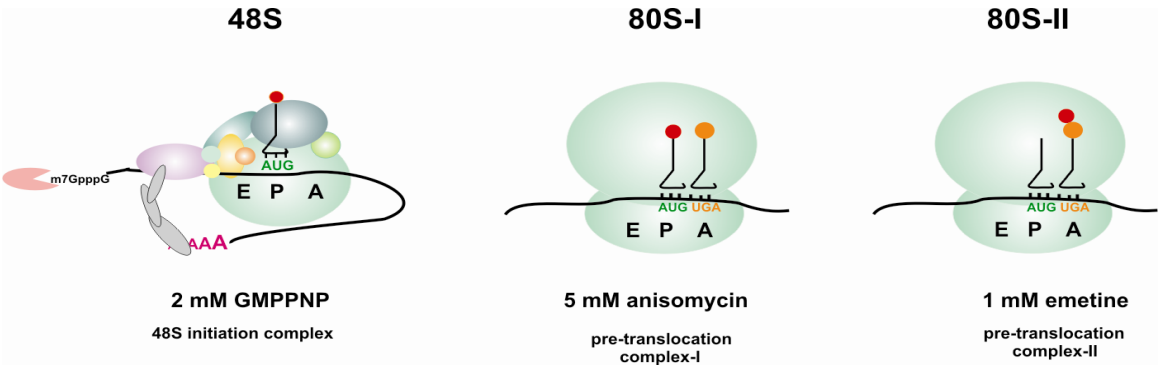

Figure 17. Schematic representation of complexes formed in rabbit reticulocyte lysate. The type of translation inhibitor and the concentration used are indicated.

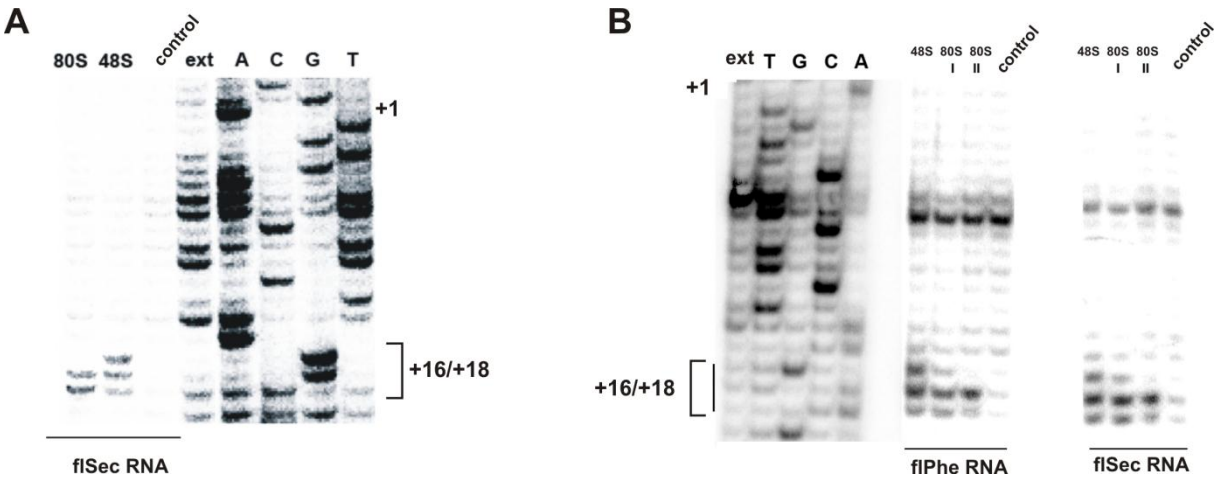

Figure 18. Toe-printing assay of 48S and 80S complexes formed in rabbit reticulocyte lysate (RRL) with the (A) flSec mRNA (natural UTRs) or (B) flPhe (lanes 6-9) and flSec (lanes 10-13) (A-rich UTRs) in the presence of SBP2. Lane "80S-II" corresponds to the complex formed in the presence of 1 mM emetine; lane "48S" corresponds to 2 mM GMPPNP; "80S-I" corresponds to the complex formed in the presence of 5 mM anisomycin; no transaltion inhibitor was added in lane "control". A,C,G,T are sequencing lanes, in (B) sequencing lanes are shown for flPhe mRNA only; "ext" no ddNTP was added.

The toe-printing assay results are shown in Figure 18. As expected, when RRL was supplied with GMPPNP, the translation was stopped at the stage of 48S pre-initiation complex formation, giving a toe-print at positions +16/+17 from the first nucleotide of the triplet positioned in the A-site (the adenosine of AUG in the particular case) (Figure 18A, lane 1; Figure 18B, lanes 6 and 10). When the lysate was supplemented with the elongation inhibitors emetine or anisomycin, the toe-print occurred at positions +17 and +18 from the adenosine of the AUG in the P-site (Figure 18A, lane 2; Figure 18B, lanes 7, 8, 11, 12). These results indicate that both sets of synthetic mRNAs can be translated in RRL and that the start

codon AUG was correctly positioned in the P-site, meaning that the UGA Sec codon was in the A-site.

3.2. *Efficiency of mRNA-ribosome complex formation*

Secondly, we have checked the efficiency of mRNA-ribosome complex formation. RRL was supplemented with SBP2 and 1 mM emetine, the ^{32}P-labeled selenoprotein mRNA being added subsequently. Ribosomal complexes were then isolated by sucrose gradient centrifugation. The gradients were fractionated, the radioactivity of each fraction was measured and the ratio of the bound mRNA (mol) per mol of ribosomes was determined (Figure 19).

The flPhe mRNA and flSec mRNAs showed the same efficiency of mRNA•ribosome complex formation (Figure 19A). The 5'UTR A-rich mRNA turned out to possess a higher ribosome binding efficiency (35%) compared to that of the mRNA with a short natural 5' UTR (10%) (Figure 19B).

Thus, taking into account these results, we have chosen mRNAs with an A-rich 5'- UTR and a partly A-rich 3'- UTR for further use.

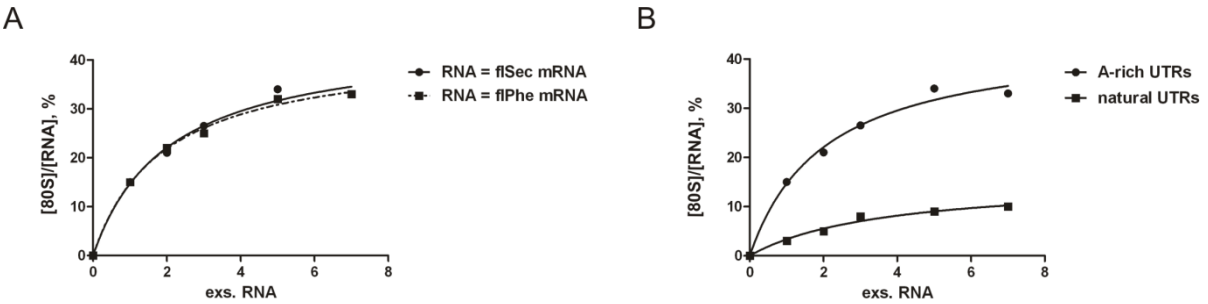

Figure 19. Isotherms of flRNA binding to the ribosomes in RRL. (A) flSec mRNA (circles) and flPhe mRNA (squares) binding to ribosomes. (B) flSec mRNA with A-rich 5'UTRs (circles) and flSec mRNA with natural 5'UTRs (squares) binding to the ribosomes. "exs. RNA" corresponds to the excess of RNA loaded in RRL relative to the amount of ribosomes in RRL. The relative error was about 10% in average.

4. Detection of SBP2 in the mRNA-ribosome complexes formed in RRL

The flSec mRNA was used to determine the localization of SBP2 during selenoprotein mRNA translation. The flSec mRNA (Figure 20A) was added to RRL supplemented with SBP2 and 3 types of translation inhibitors. Ribosomes and a pre-translocation complex formed with the 5'Phe mRNA (Figure 20A) isolated from RRL supplemented with SBP2

were used as a control. After formation of the mRNA•ribosome complex in RRL, it was isolated by sucrose gradient centrifugation, and the peaks corresponding to the complex were precipitated. Subsequently, the presence of SBP2 in the complexes was checked by western blotting. Figure 20 shows that SBP2 was present in all complexes formed with flSec mRNA (lanes 2, 3, 4, Figure 20B). This finding allowed us to suggest that SBP2 is bound to the SECIS element in the 48S complex since it is known that SBP2 does not bind the 40S ribosomal subunit (Takeuchi et al., 2009). SBP2 was not detected in the complex formed in RRL with the 5'Phe mRNA (which does not contain a SECIS element) (lane 7, Figure 20B) and with the ribosomes isolated from micrococcal nuclease treated RRL to which the mRNA was not added (lane 8, Figure 20B). Our data differ from an earlier finding by others establishing that SBP2 binds to the ribosomes in a binary complex (Kinzy et al., 2005). The absence of SBP2 on ribosomes isolated from RRL could be explained by the different experimental conditions used in the present (a complex multi-component cell lysate) and the other work where a mixture contained only ribosomes and SBP2 (Kinzy et al., 2005). Thus, RRL may contain proteins and/or RNAs to which SBP2 has a higher affinity than to the ribosome.

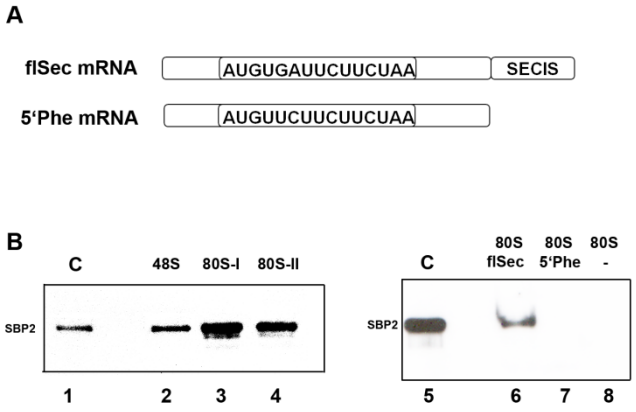

Figure 20. SBP2 content in ribosomal complexes formed in RRL. (A) Synthetic selenoprotein mRNAs used in the experiment. (B) The presence of SBP2 in complexes shown by western blotting. Left panel: lane 48S, complex formed in the presence of 2 mM GMPPMP; lane 80S-I, 80S complex formed in RRL in the presence of 5 mM anisomycine; lane 80S-II, pre-translocation complex formed if RRL was supplemented with 1 mM emetine; lane C, control corresponding to the lane with purified SBP2. Right panel: lane 80S flSec, mRNA•ribosome complex formed with flSec mRNA; lane 80S 5'Phe, mRNA•ribosome complex formed with the 5'Phe mRNA; lane 80S, ribosomes isolated from RRL; lane C, control corresponding to the lane with purified SBP2.

5. Photoreactive derivatives of synthetic selenoprotein mRNAs

The synthetic mRNAs obtained were used to synthesize various photoreactive derivatives. These derivatives will be further used to study the interaction between the SECIS element and the components of the translational apparatus.

Three sets of photoreactive derivatives were obtained:

- 4-thiouridine-containing derivatives;
- Perfluorophenylazide – containing derivatives, in which a perfluorophenylazide group was introduced statistically at 5-(3-aminoallyl)-uridines of mRNAs;
- Site-specifically introduced perfluorophenylazide groups.

5.1. 4-thiouridine-containing derivatives

Three types of 4-thiouridine containing synthetic selenoprotein mRNA derivatives were obtained.

The first one is represented by the 5'RNA and flRNA in which 4-thiouridines are statistically distributed. Additionally, these mRNAs are statistically labeled with [α-^{32}P]ATP. The 5'RNA is used as a control, to exclude proteins crosslinked to both the flRNA and 5'RNA (Figure 21).

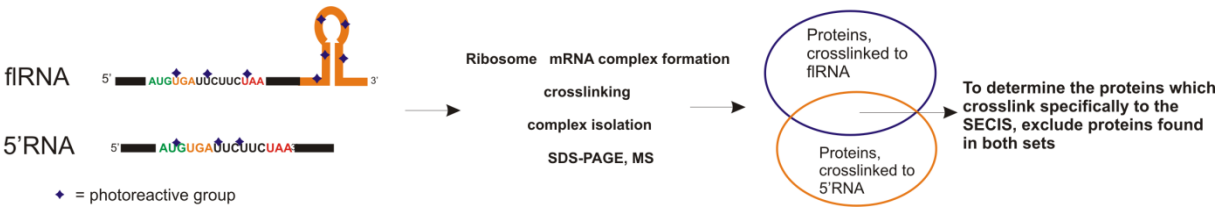

Figure 21. The strategy of using mRNAs with statistically distributed photoreactive groups. The photoreactive group can be either 4-thiouridine or azido-aUMP.

The second type was obtained by splint-aided ligation of the unlabeled 5' part of the synthetic mRNA (5'RNA) and a ^{32}P labeled SECIS-element containing statistically distributed 4-thiouridines. This mRNA is intended to provide a general overview of rRNA or proteins that are crosslinked to the SECIS element (Figure 22A). As the synthesis of this mRNA involves a ligation step with T4 DNA ligase, the 5'RNA part must possess a homogenous 3' end to form an efficient substrate for the enzyme. To this end, we have used a method described in (Kao et al., 1999) to generate RNA molecules with homogeneous 3' ends. According to (Kao et al., 1999), the modification of the last two nucleotides at the 5'

terminus of the DNA template with methoxy moieties at the ribose C2' position significantly reduces N+1 activity by the T7 RNA polymerase. The resulting DNA fragment was used to produce the 5'RNA by T7 transcription. Using this 5'RNA resulted in a 15% yield of ligation reaction.

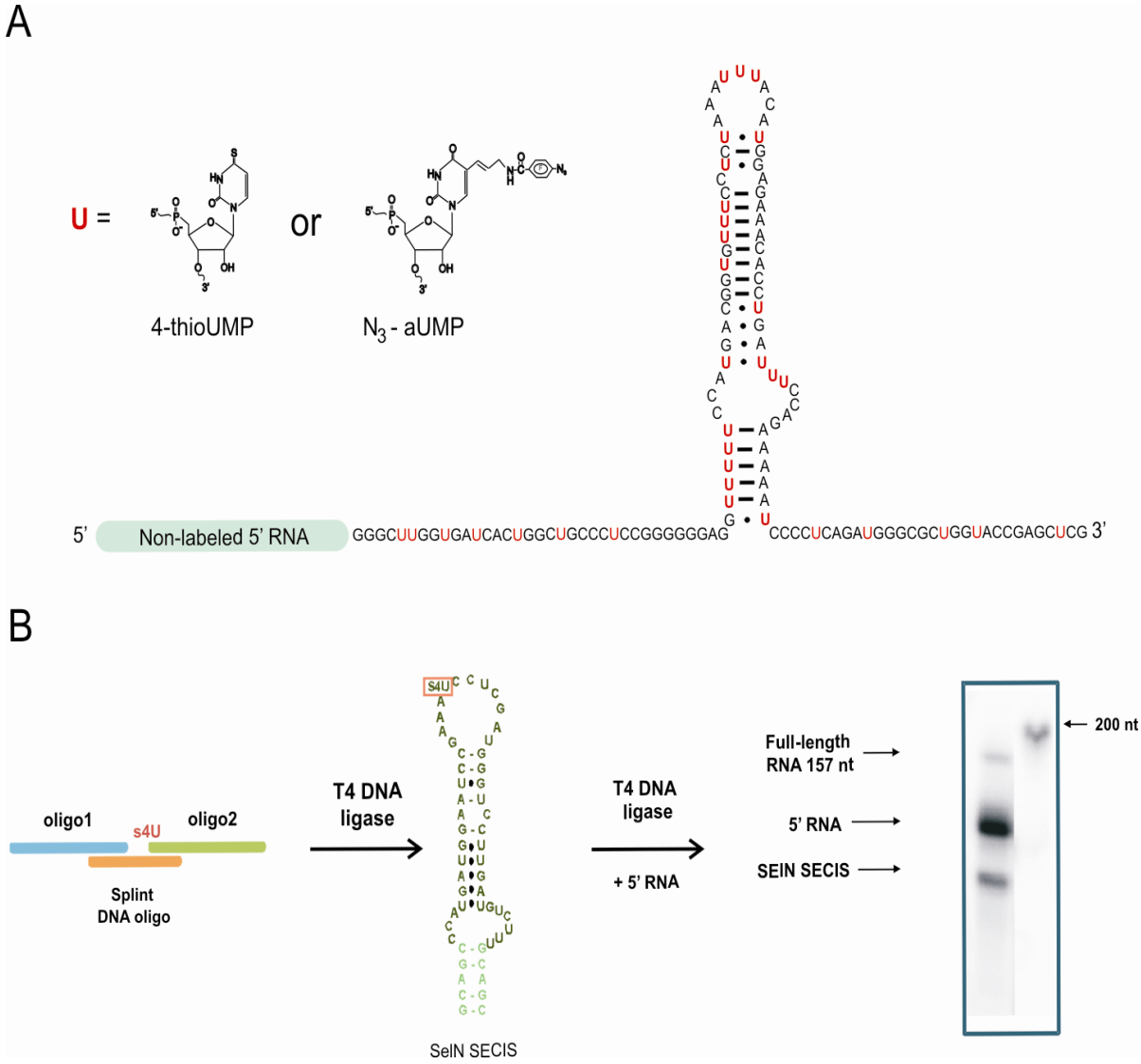

Figure 22. Example of photoreactive derivatives of synthetic selenoprotein mRNAs. They contain 4-thio-UMP or modified UMP containing a perfluophenylazide group that can be statistically introduced in the SECIS GPx 1 element (A). This SECIS RNA is subsequently added to the 5'RNA resulting in the flRNA with photoreactive groups introduced statistically into the SECIS. Possible positions of the photoreactive group are shown in red. (B) The 2 step ligation scheme to obtain the flRNA containing a 4-thiouridine at a selected position.

The third type of mRNA was synthesized by a two step splint-assisted ligation in order to obtain flRNA containing a single 4-thiouridine at a defined position (Figure 22B).

This set of photoreactive derivatives gives zero-length RNA-RNA and/or protein-RNA crosslinks when exposed to mild UV irradiation (wavelength > 290 nm).

5.2. *RNA-derivatives containing statistically distributed perfluorophenylazide-derivatized uridines*

Two types of photoreactive derivatives can be obtained using 5-(3-aminoallyl)-UTP (aaUTP); aaUTP can be introduced statistically in the whole mRNA (Figure 21) or in the SECIS-element only (Figure 22A). Prior to synthesizing the mRNA, the perfluorophenyl azide moiety was added to the $-NH_2$ group of aaUTP. The resulting N_3-aUTP was used for RNA synthesis by T7 transcription. This type of photoreactive mRNA derivatives leads to formation of RNA-RNA and/or protein-RNA crosslinks at a 10Å radius when irradiated with UV (wavelength > 290 nm).

5.3. mRNA derivatives with site-specifically introduced perfluorophenylazide groups

The scheme for introduction of a photoreactive perfluorophenylazide group at a selected position in the synthetic selenoprotein mRNA transcript is presented in Figure 23. This approach was successfully employed earlier to obtain IRES HCV mRNA photoreactive derivatives and to study its molecular environment on the 40S ribosomal subunit (for example, see Malygin et al., 2003; Laletina et al., 2006; Babaylova et al., 2009). Site-specific alkylation of the RNA was performed using 4-(N-2 chloroethyl-N-methylamino)benzyl-5' (or 3')-phosphoramides derivatives of oligodeoxyribonucleotides complementary to positions 26-46, 14-34 and 35-53 of the GPx1 SECIS (numbering starts at the first nucleotide of the SECIS) (Figure 24). Sequences 26-46 and 14-34 were chosen to find new interacting partners of the SECIS element, while positions 35-53 yield an mRNA derivative bearing a crosslinking group at the SBP2 interaction site. This mRNA can thus be used as a control for the validity of the crosslinking procedure.

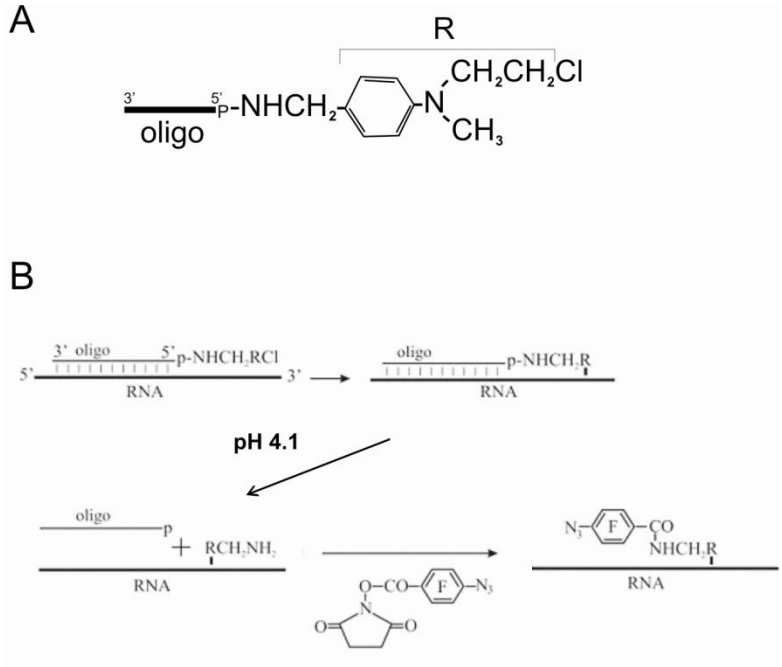

Figure 23. Site-specific introduction of a photoreactive group into specific RNA sites. (A) Structure of the alkylating reagent ([4-(N-2-chloroethyl-N-methylamino)benzyl]-phosphoramides of oligodeoxyribonucleotide) (B) Chemistry of the site-specific introduction of a photoreactive group into specific RNA sites based on the complementary-addressed alkylation of the RNA with [4-(N-2-chloroethyl-N-methylamino)benzyl]-phosphoramides of oligodeoxyribonucleotides. (B) is taken from (Laletina et al., 2006).

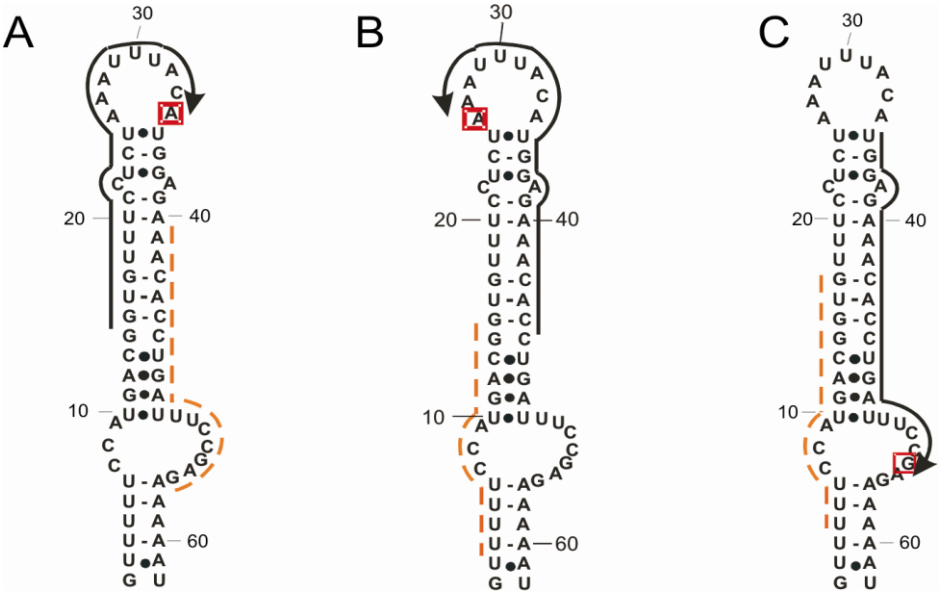

Figure 24. GPx1 SECIS derivatives bearing photoreactive moieties at positions 34 (A), 26 (B) and 54 (C). Sequences complementary to the alkylating oligomer are marked with solid lines. Arrow indicates the 5' or 3' terminus of the oligonucleotide bearing the alkylating group. Sequences complementary to helper oligomers, that are used to destabilize the secondary structure of the RNA, are marked with dashed lines. Alkylated nucleotides are in boxed in red.

To increase the yield of the covalent adduct with the oligonucleotide derivative, a helper oligonucleotide was used in all cases. This helper oligonucleotide is complementary to the opposite strand of the helix, thus helping its unwinding and facilitating base-pairing between the alkylating oligomer and the synthetic mRNA (Figure 24). After alkylation, the modified RNA is separated from the unmodified version by 8% denaturating PAGE. The extent of alkylation was 40±5% in all cases. It is known that the $ClRCH_2NH$- moiety attached to the 5' or 3' terminal nucleotide of the oligonucleotide will crosslink to the nucleotide to which it base pairs or to the neighboring unpaired nucleotide (Zenkova et al., 1995; Buligyn et al., 1998). A, G and C can be modified, except for uracil that is unable to react with aromatic 2-chloroethyl amines under the conditions used (Karpova, 1987). The SECIS nucleotides modified by the alkylating derivatives used in the present work were identified by reverse transcription. These experiments were carried out in the laboratory of Prof Galina Karpova by Dr. Elena Babaylova (data not shown). It was found that nucleotides 26, 34 and 54 (starting from the first nucleotide of the SECIS) were alkylated by these reagents (Figure 24).

After the alkylation step, the phosphoramide bond in the covalent adducts was hydrolyzed under mild acidic conditions, liberating the aliphatic amine group $–RCH_2NH_2$ linked to the RNA. The modified RNA is purified by 8% denaturating PAGE. Next, the N-oxysuccinimide ester of the 4-azidotetrafluorobenzoic acid is added to the $-RCH_2NH_2$ containing RNA, to selectively introduce photoreactive moieties (Figure 23B). The mRNA bearing a crosslinking group at positions 26, 34 and 54 were thus obtained.

The same approach was used to synthesize mRNA derivatives of the flRNAs containing the GPx4 and SelN SECIS.

* * *

In the further functional study, we finally set out to use mRNA derivatives bearing statistically distributed 4-thiouridines because these derivatives were obtained with the best yield of synthesis and we could show they were appropriate to study the molecular contacts of the SECIS in general. In perspective, mRNA derivatives bearing selectively introduced photoreactive groups can be used to study which region of the SECIS RNA interacts with a particular protein and/or rRNA.

6. SECIS-ribosome interactions during selenoprotein mRNA translation.

To identify the components of the translation machinery that interact with the SECIS element in the course of selenoprotein mRNA translation, we have used the flSec mRNA harboring a statistically ^{32}P-labeled SECIS element and bearing statistically distributed 4-thiouridines. These mRNAs were obtained by splint-aided ligation as described in section 5.1 and shown in Figure 22A. The flPhe mRNA was used as a control for unspecifically bound ligands.

Using flRNAs (either flSec or flPhe mRNAs) and RRL supplemented with SBP2 and translation inhibitors (GMPPNP, anisomycin and emetine), three types of complexes were formed (Figure 25A). To generate crosslinks, reaction mixtures were irradiated with mild UV light (>290 nm).

To identify the SECIS-protein crosslinks, ribosomal complexes were isolated by sucrose gradient centrifugation. Ribosomes were next precipitated, the RNAs were hydrolyzed by RNase A, and half of the samples were treated with proteinase K as a control to prove that the signal observed after RNase A treatment is indeed due to the labeled protein. Samples were then resolved on 14% SDS-PAGE and the dried gel was exposed to a PhosphorImager plate. The resulting autoradiogram is shown in Figure 25C. With the 48S pre-initiation complex (lane 2) and in the 80S pre-translocation complex (lane 6) formed in the presence of emetine, the only protein crosslinked to the SECIS migrates at the same level as SBP2. In the case of the complex formed in the presence of anisomycin (lane 15), one can see two groups of proteins migrating at the level of ribosomal proteins (10-20 kDa range). These crosslinks were considered as specific because we did not observe the same bands in the control lanes (lanes 11-14, 19-22).

We conclude from these experiments that SBP2 is bound to the SECIS RNA in both the 48S pre-initiation complex and 80S pre-translocation complex. In the case of the complex where trans-peptidation was blocked, SBP2 did not crosslink to the SECIS RNA; this strongly suggests that SBP2 does not bind to the SECIS RNA in this complex. However, SBP2 was present in this type of complex (as shown by western blot). Thus, SBP2 is very likely to be bound to the ribosome in this case.

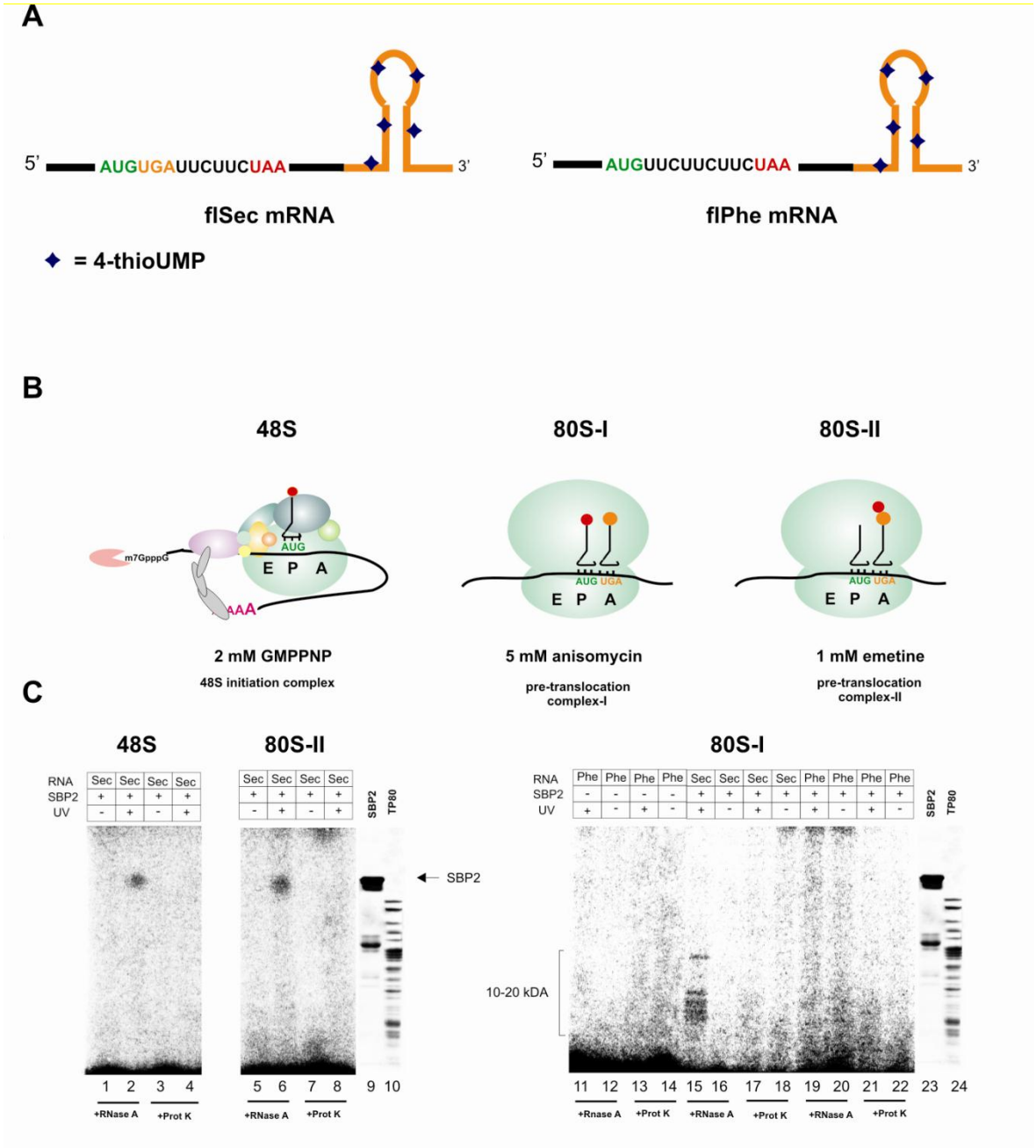

Figure 25. (A) Schematic representation of flSec and flPhe mRNAs used in the experiment (B) Schematic representation of the complexes which were isolated from RRL. The antibiotic concentration used is indicated. (C) Autoradiograms showing proteins crosslinked to the ^{32}P-labelled 4-thio-U-containing SECIS. The information about the type of l mRNA used, whether SBP2 was added and whether the reaction mix was exposed to UV, is indicated above each lane. Below the gel, it is indicated whether the complex in the lane were treated with RNase A or proteinase K (ProtK). Lanes SBP2 and TP80 (total protein from 80S ribosomes) were Coomassie stained.

Unfortunately, low yields of SECIS-proteins crosslinks did not allow us to identify proteins by mass spectrometry. This low yield may indicate a very short interaction time

between the ribosome and the SECIS element. Such contacts may occur when the SECIS is recruiting the SBP2/EFSec/tRNA^{Sec}/GTP complex to the ribosome.

In a further step, we wished to investigate the possible SECIS-rRNA crosslinks. Proteins in RRL were hydrolyzed by proteinase K and rRNAs were extracted. The resulting RNAs were resolved by denaturing 5% PAGE-8M urea. After autoradiography, only the signal from the mRNA could be seen. No SECIS-rRNA crosslink could be observed (data not shown).

The results discussed in this section were the reason why we wished to investigate the molecular environment of SBP2 on the ribosome. Moreover, in a collaborative work between the Novosibirsk and Strasbourg labs, SBP2 was shown to bind *in vitro* the human 60S and not to the 40S subunit, and the SBP2 domain important for this interaction was determined (Takeuchi et al., 2009).

7. SBP2-ribosome interaction studies

To study SBP2-ribosome interactions, we have used purified human ribosomes from placenta and a recombinant human SBP2 carrying a deletion of the 344 N-terminal amino acids. The full-length SBP2 and its truncated form have the same affinity to the ribosome (Takeuchi et al., 2009), but the truncated version gives a higher yield of production in *E.coli* and is more stable.

7.1. Choosing SBP2•ribosome binding conditions

Prior to setting up crosslinking experiments, we have chosen a buffer that is suitable to studying SBP2-ribosome interactions. It is known that above 2mM Mg^{2+}, SBP2 looses its ability to interact with the SECIS RNA and, possibly, with the ribosome (Copeland et al., 1999). Usually, the PBS buffer, which contains no Mg^{2+}, or PBS containing 0.5 mM Mg^{2+} (PBS-Mg) were used to study SECIS-SBP2 and ribosome-SBP2 interactions (Kinzy et al., 2005; Takeuchi et al., 2009). On the other hand, ribosomal subunits do not associate *in vitro* to form the 80S ribosome under low Mg^{2+} concentration. Formation of the SBP2•ribosome complex was tested under 4 different buffers to find which one would enable both 80S ribosome association and SBP2 binding to the 80S (Figure 26A, B). As seen in the Figure, 80S ribosomes associated neither in PSB nor in PBS-Mg. A 4 mM Mg^{2+} concentration enabled 80S ribosome formation but not the ribosome-SBP2 interaction. The only buffer

fulfilling both conditions was 20 mM HEPES-KOH pH 7.8, 100 mM KCl, 2 mM DTT, 2 mM $MgCl_2$. As shown in (Shenvi et al., 2005), in vitro translation in RRL is optimal at this Mg^{2+} concentration which maintains the ribosome structure. Also, as expected, SBP2 binds to 80S and 60S, but not to 40S ribosomal subunits in this buffer (Figure 26C).

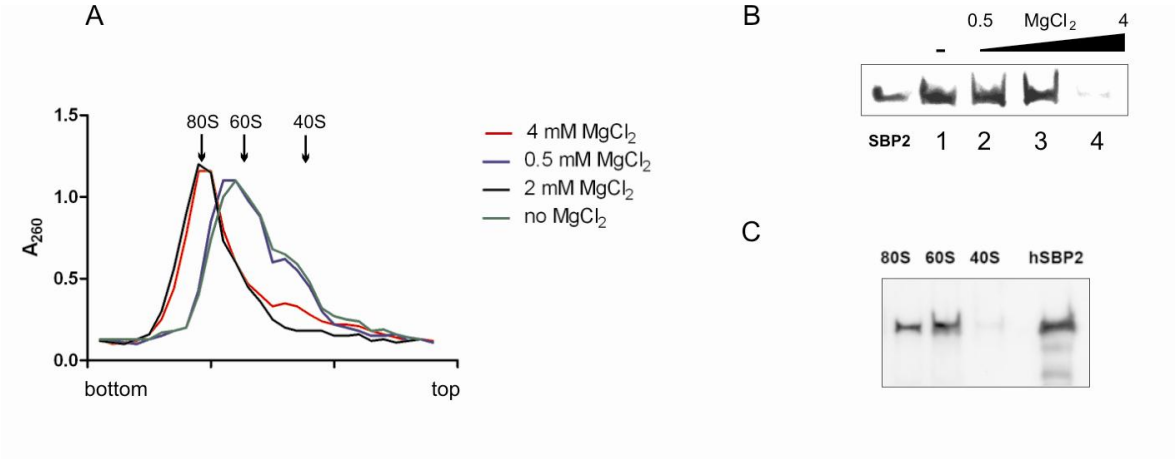

Figure 26. SBP2•80S binding depending on Mg^{2+} concentration. (A) Sucrose gradient sedimentation profiles of 80S ribosomes. (B) Western blotting analysis of 80S•SBP2 binding depending on Mg^{2+} concentration. Lanes 1, 2, 3, 4 correspond to buffers with Mg^{2+} concentration of 0, 0.5, 2 and 4 mM, respectively. (C) Western blotting analysis of SBP2 binding to the 40S, 60S and 80S in the buffer with 2 mM Mg^{2+}.

Furthermore, we have tested SBP2•ribosome complex formation depending on the excess of SBP2 added. We determined that a maximum of 80% SBP2•ribosome complex can be obtained when SBP2 is in a 3-fold excess to ribosomes (data not shown). This SBP2/ribosome proportion was used in all experiments described below.

7.2. Bifunctional reagents crosslinking

To determine whether the 40S subunit contributes to the SBP2 binding to the 80S ribosome, we treated 40S•SBP2, 60S•SBP2 and 80S•SBP2 complexes with two bifunctional reagents – diepoxybutane and 2-iminothiolane. Diepoxybutane is highly reactive towards – NH_2 and –SH groups and is known to form RNA-protein and protein-protein crosslinks at a 4Å radius (Figure 27A, B) (Baumert et al., 1978). 2-iminothiolane can act either as an RNA-protein crosslinker (crosslinking radius 5Å), or as a protein-protein crosslinker (crosslinking radius 14Å), depending on the reaction conditions used to generate the crosslink (Figure 27C) (Traut et al., 1973).

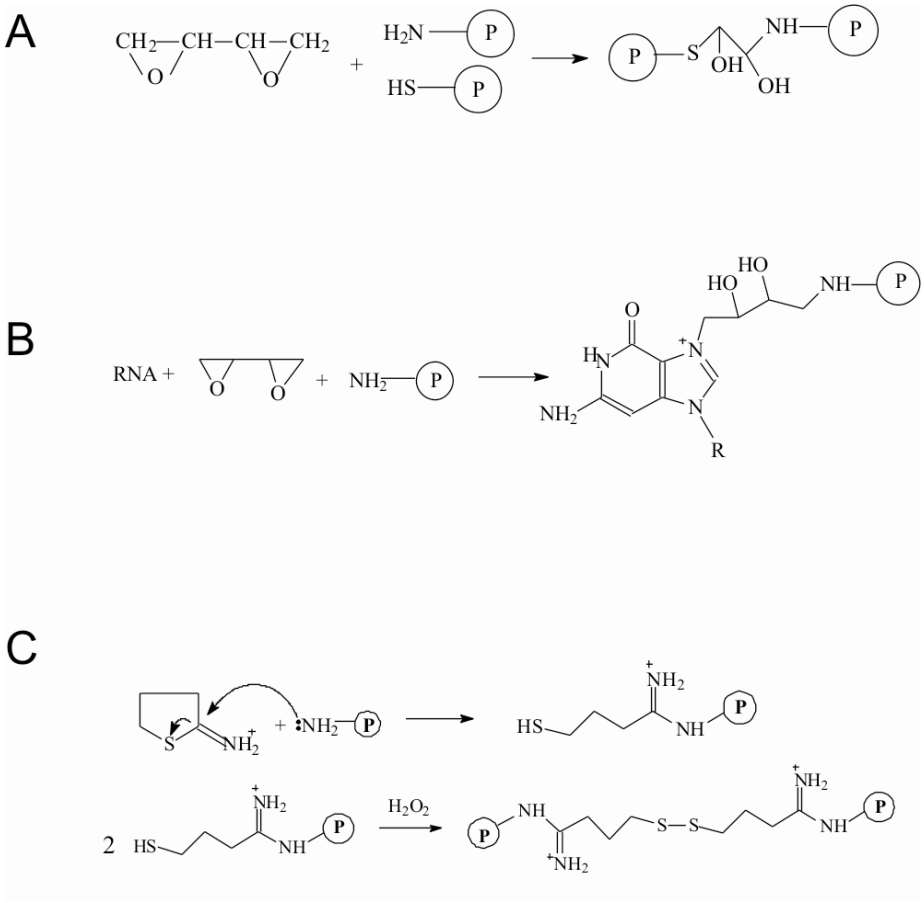

Figure 27. Reactions describing cross-link formation generated by diepoxybutane (A), (B), and 2-iminothiolane (C).

After treatment with the crosslinking reagents, SBP2•ribosome complexes were purified by sucrose gradient ultracentrifugation under dissociating conditions. This procedure enables the removal of the uncrosslinked SBP2 and dissociates the 80S ribosome into 60S and 40S subunits. Protein-protein crosslinks can then be analyzed; alternatively, a second centrifugation in the presence of SDS and EDTA can be performed to isolate rRNA and to subsequently analyze the SBP2 – rRNA crosslinks (Figure 28).

Using diepoxybutane that forms rRNA-proteins and protein-protein crosslinks under the same reaction conditions, we have first analyzed the distribution of the crosslinked SBP2 among subunits (Figure 29A). Our data showed that SBP2 crosslinks only to the 60S subunit, whether isolated or in the 80S ribosome. The 40S subunit does not contribute to the crosslink with SBP2 when part of the 80S ribosome.

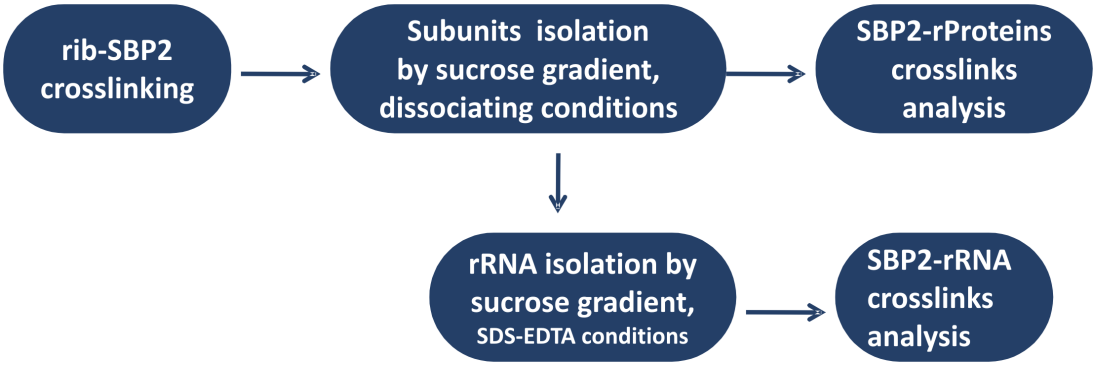

Figure 28. Experimental scheme for studying ribosomal components that interact with SBP2, using crosslinking assays with bifunctional reagents.

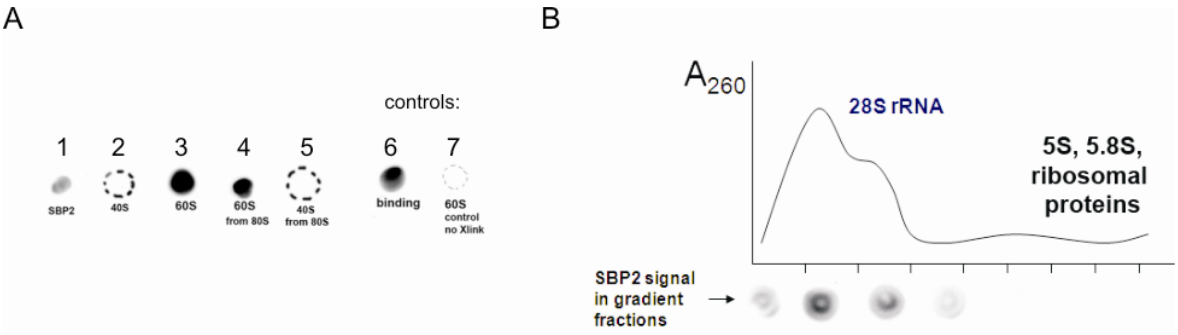

Figure 29. Diepoxybutane crosslinking assay. (A) Dot-blot analysis of SBP2-ribosome crosslink distribution among ribosomal subunits. 1: signal from SBP2; 2: 40S•SBP2 complex; 3: 60S•SBP2 complex; 4: 80S•SBP2 complex, 60S subunit, 5: 80S•SBP2 complex, 40S subunit; control experiments. 6: 60S•SBP2 complex that was purified by sucrose gradient centrifugation under non-dissociating conditions; 7: 60S•SBP2 complex that was not treated with diepoxybutane but purified by sucrose gradient under dissociating conditions. (B) Dot-blot analysis of SBP2-rRNA crosslink after centrifugation in SDS-EDTA sucrose gradient.

Furthermore, the distribution of crosslinks of SBP2 between rRNA and ribosomal proteins with diepoxybutane showed that the efficiency of crosslinking to rRNA was 10-fold higher than to ribosomal proteins. This strongly suggests that the rRNA contributes mainly to the formation of the SBP2 binding site. Crosslinking performed with diepoxybutane and 2-iminothiolane under conditions where only RNA-protein crosslinks are formed, revealed that SBP2 crosslinks to the 28S rRNA (the diepoxybutane crosslinking data is presented in Figure 28B; the 2-iminothiolane crosslinking data is not shown).

From the above experiments, we can conclude that the SBP2 binding site is located on the 60S subunit only.

7.3. SBP2 protects a discrete domain of expansion segment 7 of the human 28S RNA

Since the 28S rRNA contributes mainly to the formation of the SBP2 binding site on the ribosome, we wished to delineate the area of binding of SBP2 on this rRNA. In a first attempt, we decided to obtain an idea of the region of the 28S rRNA that is protected by SBP2. To this end, we set out to use hydroxyl radical footprinting of 28S rRNA in the SBP2•60S complex. Hydroxyl radicals, generated by the Fenton reaction, lead to RNA cleavage of the phosphodiester bond regardless of the RNA sequence, by reacting with solvent accessible 2'OH of the sugar moieties. After reaction, the cleavage sites were mapped by primer extension.

Almost all of the 5000 nucleotides of the human 28S rRNA were scanned, except for 50 nt at the very 5' end and 80 nt at the very 3' end, as well as 230 nt in expansion segment (ES) 27 due to the difficulties to find the proper primer that would hybridize to this highly GC-rich region. However, we obtained very interesting data. Nucleotides protected from hydroxyl radical cleavage were found in only two regions, residing in ES7. Footprints were located at nucleotides 1111-1122, 1133-1137, 1169-1179 and 1182-1184 (Figure 30A, B). Nucleotides with increased sensitivity to hydroxyl radical attack, positions 4951-4955, were found in ES39 (Figure 30C)

ES7 in humans is very large (856 nt). Since no secondary structure is currently available for most of its sequence, Mfold (Zuker, 2003) was used to predict secondary structure models of ES7. One such model is presented in Figure 31. According to this structure, nucleotides protected from hydroxyl radicals are located on 2 adjacent helices, helix 1 and helix 2 (Figure 31). It should be noted that we do not possess structure-based sequence alignments of the 100 or so 28S rRNA sequences available so far, rendering the vailidity of this model only putative. Besides, the current 3D structures of the eukaryotic ribosomes, obtained either by cryo-EM or X-ray crystallography (Chandramouli et al., 2008; Arnache et al., 2010ab), did not resolve the entirety of ES7 whose length vary anywhere from 100 nt to 856 nt, according to the species. Therefore, the secondary structure prediction for this section of the human ES7 may not be accurate. Further studies are underway (in collaboration with Fabrice Jossinet in the Westhof group) to generate structure-based sequence alignments of the whole set of available 28S rRNA sequences, and thus to more precisely map the area protected by SBP2.

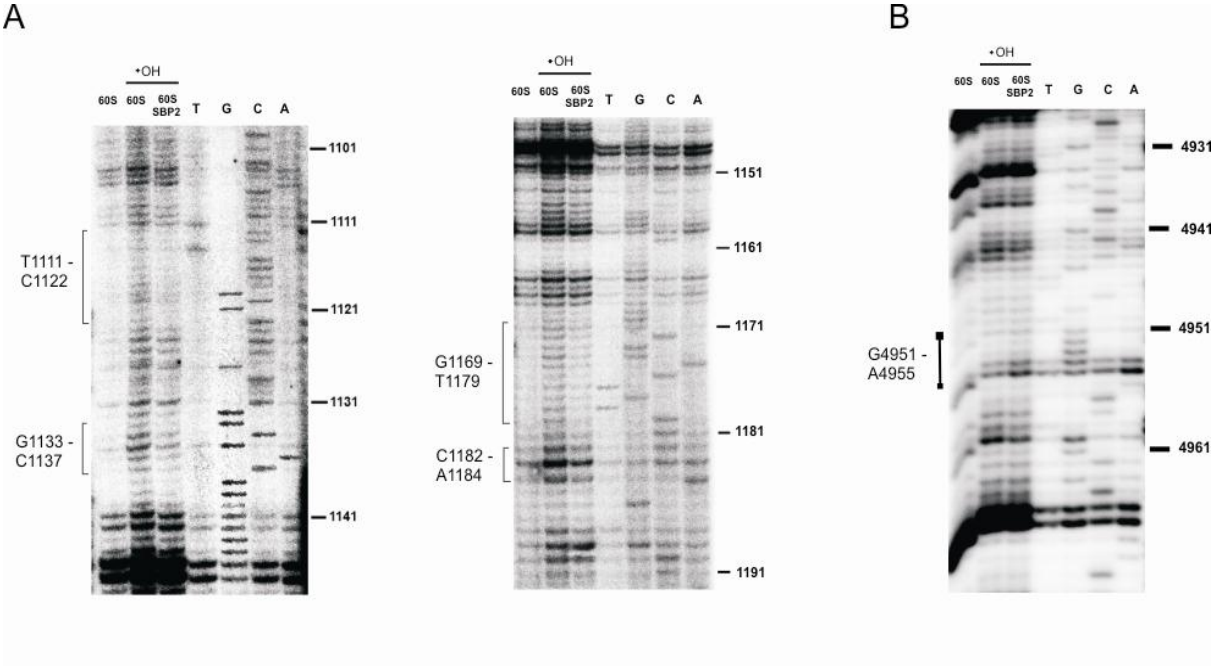

Figure 30. Hydroxyl radical probing of the 28S rRNA in the 60S subunit complexed with SBP2. (A) and (B) Fragments of the gel containing the footprints of SBP2 on the 28S rRNA. Lane "60S" corresponds to the isolated 60S subunit without hydroxyl radical treatment; lanes •OH "60S" and "60S-SBP2" correspond to the 60S subunit either free or in complex with SBP2 treated with hydroxyl radicals; T,G,C,A, sequencing lanes. Positions of footprints are marked by brackets. Those in B indicate the region that becomes more accessible to hydroxyl radicals on SBP2 binding.

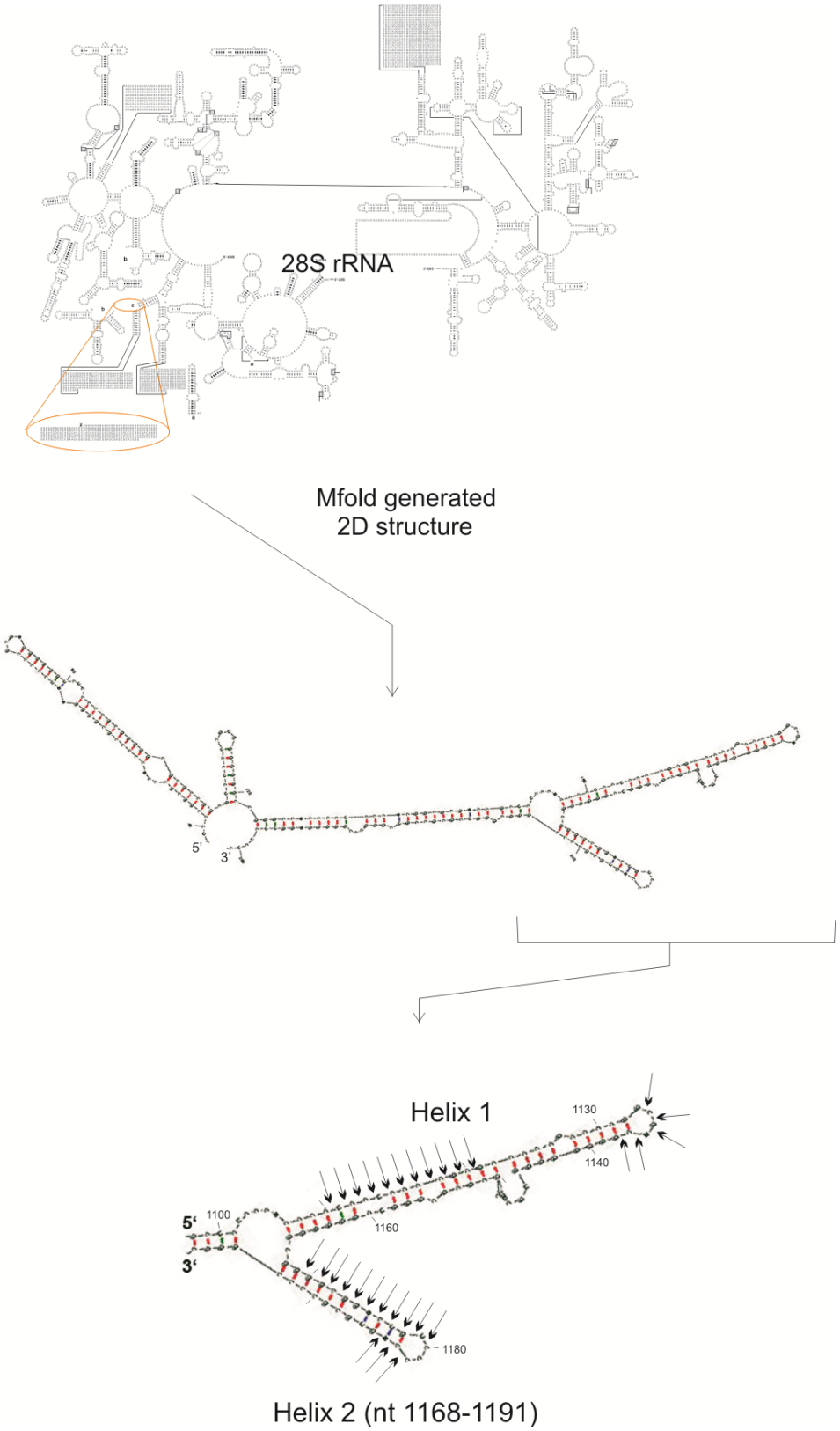

Figure 31. Mapping of the footprinting results on the ES7 secondary structure model. Secondary structure model of the human 28S rRNA and numeration according to (Chandramouli et al., 2008). Arrows indicate the position of the obtained footprint.

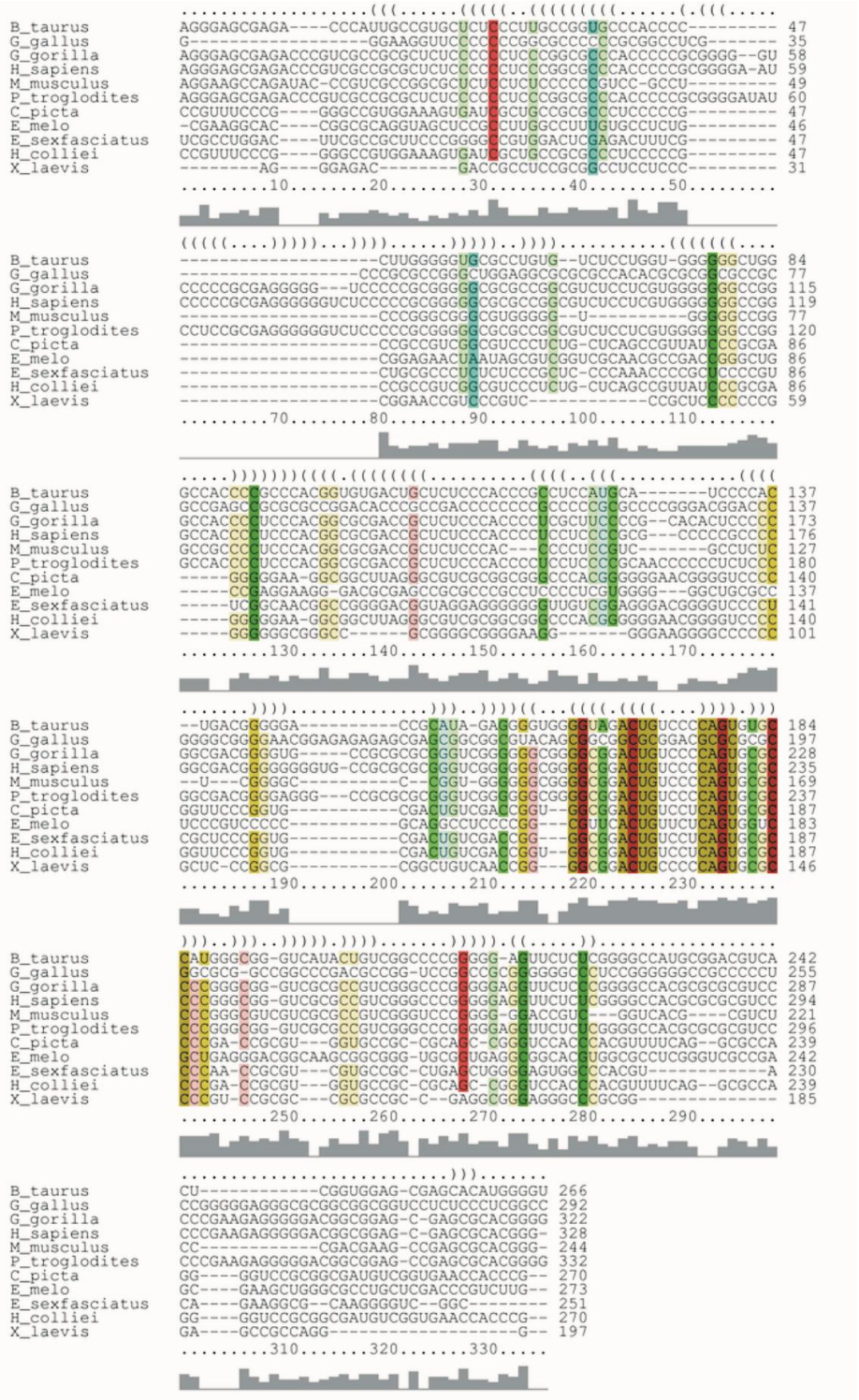

Figure 32. Structure-based sequence alignment of part of ES7 in 10 organisms: humans, chimpanzee, gorilla, cow, mouse, chicken, frog, turtle, shark and aquarium fish. Numbering of nucleotides: A1 corresponds to A956 in the 28S rRNA of *H.sapiens*. Color code: hue shows sequence conservation and saturation decreases with the number of compatible base pairs, thus showing structural conservation (generated automatically by LocARNA) (Will et al., 2007).

Today, we can only present a structure-based sequence alignment of part of ES7 from 10 organisms (nt 956-1284 according to the numbering of the human 28S rRNA, Genbank reference NR_003287) performed with LocARNA, a web-based tool which simultaneously folds the input sequences using the RNAfold algorithm and aligns them (Will et al., 2007) (Figure 32). Also, a consensus secondary structure was predicted by this program, showing the conservation of helix 2 (nt 1168-1191) (Figure 33A). A number of compensatory base changes provides strong evidence for the existence of helix 2 (Figure 33B).

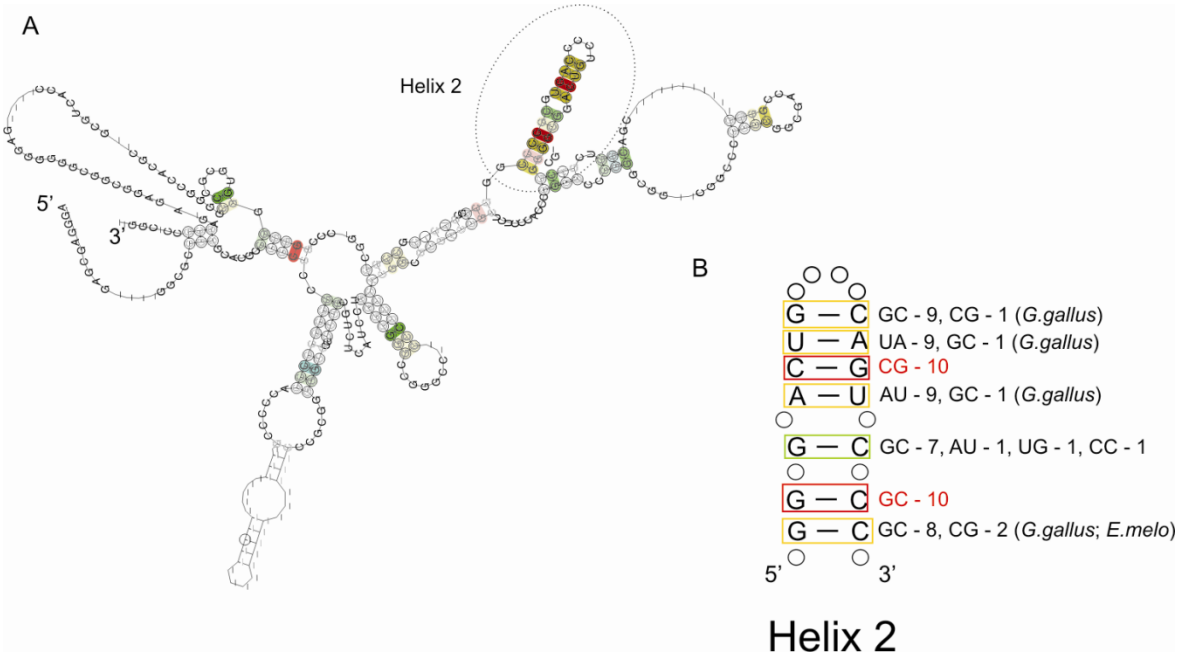

Figure 33. (A) Consensus secondary structure of the ES7 fragment (nt 956-1284) predicted by locARNA. The color code corresponds to that of the alignment (Figure 32). Helix 2 is circled by a dashed line. (B) Covariations occurring in helix 2 (nt 1170-1193 in the human 28S rRNA) are shown for the 28S rRNA sequences of the 10 organisms aligned in Figure 32. The color code is the same as in Figure 32.

An ES39 (the region of increased hydroxyl radical attack) secondary structure model was proposed for mouse in (Nygard et al., 2006). Our data can be mapped on this structure because the sequence conservation between human and mouse in ES39 is almost 100% (Figure 34). Based on the structure-based alignment of ES39 from wheat, fungi and mammals done in (Nygard et al., 2006), the corresponding hairpin in ES39 from wheat was found. Footprinting data could be mapped on the available three-dimensional model of the 80S wheat ribosome (Armache et al., 2011a,b).

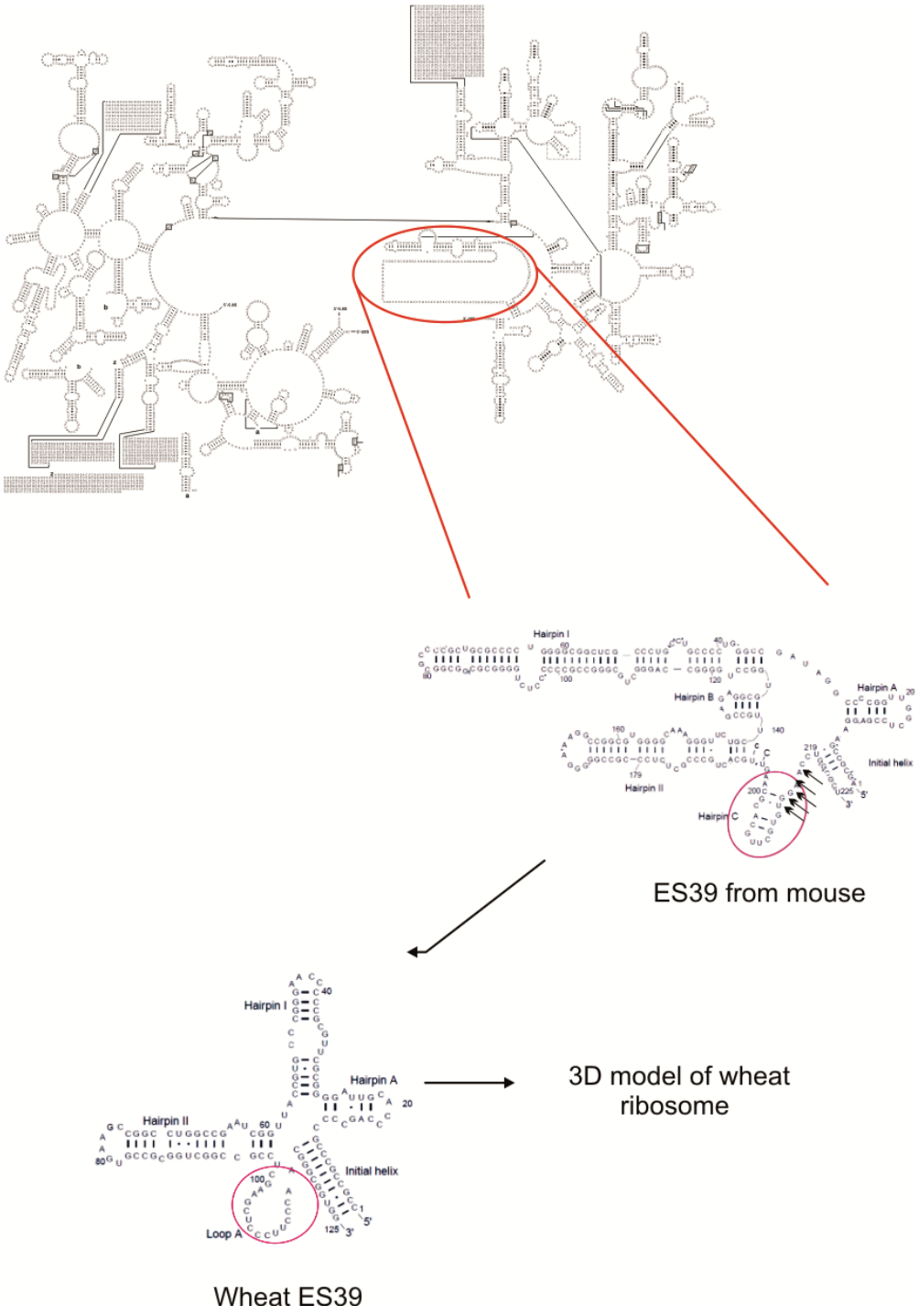

Figure 34. Mapping of the footprinting results on the ES39 secondary structure model. Secondary structure model of the human 28S rRNA according to (Chandramouli et al., 2008). ES39 is highlighted, nucleotides that become more accessible to hydroxyl radical attack upon SBP2 binding are indicated by arrows. Numbering, names of hairpins and helices are as in (Nygard et al., 2006).

The ES7 helices of interest for us are not present on the available cryo-EM models of the eukaryotic ribosome. Nevertheless, a part of this expansion segment was modeled for the wheat 60S ribosomal subunit (in wheat 28S rRNA this part constitutes 30 nt while in human this part is 328 nt long) (Armache et al., 2010a). Taking into account this model, it is possible to suggest that the SBP2 binding site is located near the L7/L12 stalk (Figure 35).

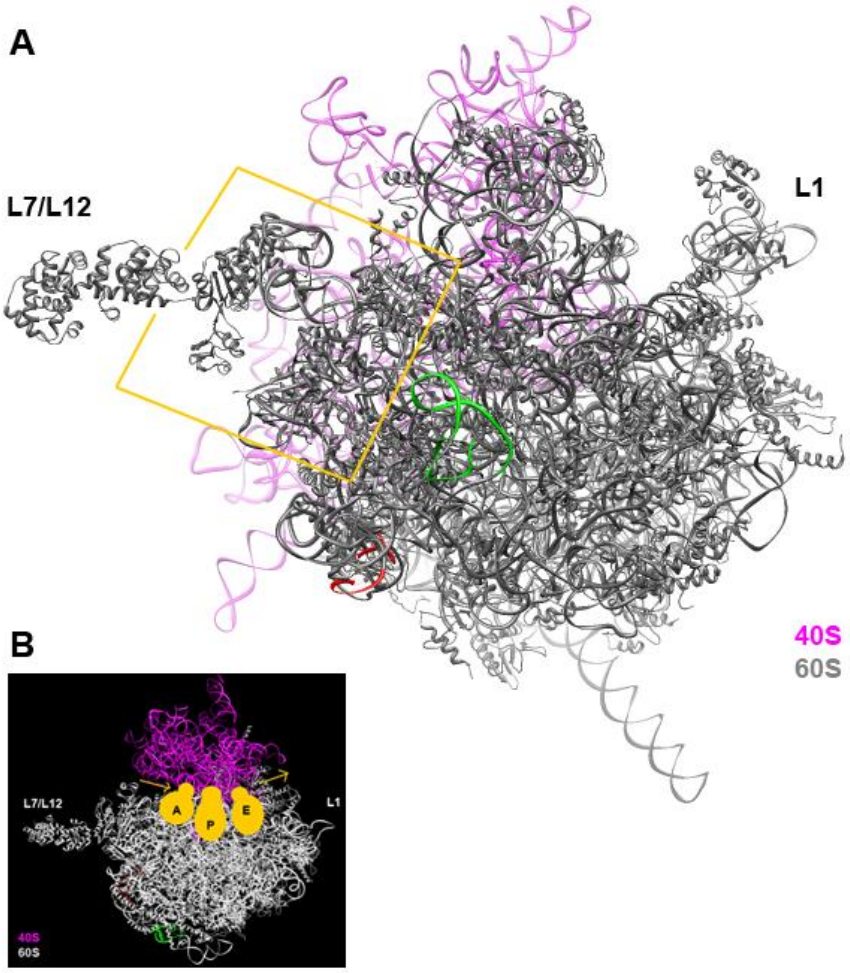

Figure 35. Possible SBP2 binding site on the ribosome. (A) 3D model of the wheat 80S ribosome, 60S solvent side view. The 60S (proteins and rRNAs) is shown in grey, the 40S is in pink (only the 18S rRNA backbone is displayed). The L7/L12 and L1 stalks are indicated. The part of wheat ES7 (nt 580-610, numeration as in Armache et al., 2010a) is in green; the helix from ES39, containing nucleotides that become more accessible to hydroxyl radicals upon SBP2 binding, is shown in red. Possible SBP2-ribosome interaction area circled in orange. (B) Wheat 80S ribosome, top view. Orange arrows indicate the entry and the exit of mRNA, orange circles "A", "P" and "E" show shcematical location of the A-, P- and E-site, reaspectively. The 3D structure was taken from (Armache et al., 2010a,b; PDB accession numbers 3IZR, 3IZ7 and 3IZ9). This figure was generated by Chimera ver. 1.5.2 (Pettersen et al., 2004).

Part III

General discussion

General discussion

Selenocysteine-containing proteins constitute a small but extremely important protein family. They are implicated in cell protection against free radicals (Gromer et al., 2003), male fertility maintenance (Foresta et al., 2002), metabolism of thyroid hormone (Beckett and Arthur, 2005) and muscle development (Rederstorff et al., 2006). The translational machinery requires a number of protein and RNA factors to enable decoding of UGA to a Sec codon and thus to synthesize active full-length selenoproteins (Allmang and Krol, 2006; Allmang et al., 2009). Up to date, information on the molecular interactions underlying the events occurring during translation of selenocysteine mRNAs is scarce. The work done in the course of my thesis helped bring some important novel findings:

1. In the cell, SBP2 is bound to the SECIS of selenoprotein mRNAs rather than to ribosomes.
2. The affinity of SBP2 to ribosomes is much lower than to the SECIS. These two issues agree with each other.
3. The SECIS contacts the ribosome only at the time of binding of the Sec-tRNA^{Sec} in the A site (i.e., before transpeptidation, in other words, before insertion of Sec into the nascent polypeptide). This follows from the results that SBP2 is present in the respective complex but it does not crosslink to the SECIS. The latter implies that SBP2 binds to the ribosome.
4. The SECIS most probably does not bind to the ribosome as a usual ligand (this could be deduced from very the low yields of cross-links) but has forced contacts with the ribosome when SBP2 moves from the SECIS to the 60S subunit.

We have chosen a crosslinking approach to study the interactions between selenoprotein mRNAs (their SECIS elements in particular) and components of the translation machinery. The crosslinking approach where the RNA molecule bears photoreactive groups has been widely used to study RNA-ribosome interactions in various translation complexes (Graifer et al., 2004; Laletina et al., 2006; Pisarev et al., 2008; Babaylova et al., 2009).

In this study, mRNAs bearing SECIS elements with statistically distributed 4-thiouridine residues were used to determine at which translation step the SECIS RNA interacts with the ribosome. The results obtained allowed us to propose the following model for selenoprotein mRNA translation (Figure 36). SBP2 is already bound to the SECIS element at the step of formation of the 48S initiation complex. The SECIS RNA delivers the SBP2-

bound EFSec/Sec-tRNA^{Sec}/GTP complex at a ribosome approaching the UGA Sec codon. After binding of the ternary complex EFSec/Sec-tRNA^{Sec}/GTP or accommodation of the Sec-tRNA^{Sec} to the A-site, SBP2 dissociates from the ribosome. Once the peptidyl moiety of the P-site-bound peptidyl-tRNA and the Sec-residue of the A-site-bound Sec-tRNA^{Sec} (before translocation of the Sec codon to the P-site), SBP2 binds again to the SECIS.

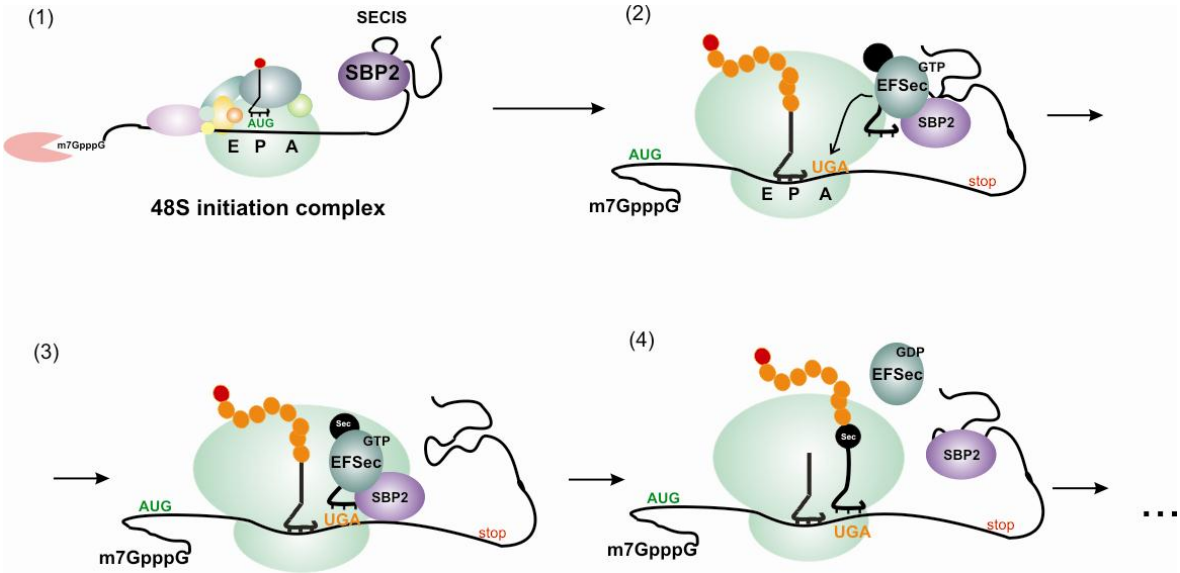

Figure 36. A model for eukaryotic selenocysteine incorporation. SBP2 travels bound to the SECIS element (1). Once the ribosome has reached the UGA Sec codon, the SECIS element delivers the SBP2/EFSec/Sec-tRNA^{Sec}/GTP complex to the ribosomal A-site (2), during which step the SECIS may be contacting the ribosome in a rather transient manner. (3) After tRNA accommodation or after peptide bond formation, SBP2 leaves the ribosome and (4) binds back to the SECIS, owing to its high affinity for this motif.

However, this model still leaves certain aspects unclarified. First of all, what is the nature of the SECIS-ribosome interaction? In the complex formed in the presence of anisomycin, one can envisage that the ribosome interacts both with SBP2 (shown by western blotting) and the SECIS (shown by crosslinking). Taking into account the very low cross-linking yield, one can suggest that the SECIS-ribosome interaction is rather fast, presumably occurring during the binding of the SBP2/EFSec/Sec-tRNA^{Sec}/GTP complex to the ribosome. A question remains opened as to whether SBP2 can interact at this stage with both the ribosome and the SECIS. We can not exclude this possibility although we could not see crosslinks between the SECIS and SBP2 in the complex formed in the presence of

anisomycin. The absence of crosslinks can be explained in different ways: (i) there were no possible target for 4-thiouridines on SBP2 or (ii) all the available 4-thiouridine moieties reacted with ribosomal proteins.

We have shown that the SECIS element is bound to SBP2 in 48S complex, but it is still unclear whether SBP2 binds to the SECIS element when the selenoprotein mRNA is already recruited to the 40S subunit, or the selenocysteine mRNA is always complexed with SBP2. The latter possibility is more likely since it was shown that SBP2 co-localized with selenoprotein mRNA in the nucleus (unpublished data obtained by Dr. L. Wurth from the laboratory of Dr. A. Krol). Besides, SBP2 can shuttle between the nucleus and the cytoplasm (Papp et al., 2008), carrying EFSec when traveling from one compartment to the other. This suggests that the SECIS/SBP2/EFSec/Sec-tRNA^{Sec}/GTP complex may already form in the nucleus (de Jesus et al., 2006; Boulon et al., 2008). Together, all these findings suggest that SBP2 is always complexed with the selenoprotein mRNA.

Furthermore, it is of prime importance to understand at which steps the SBP2-ribosome interaction occurs. Based on our data, we suggest that it may occur when the SBP2/EFSec/Sec-tRNA^{Sec}/GTP complex binds to the ribosome, or past this step, during Sec-tRNA^{Sec} accommodation to the ribosomal A-site.

Finally, it was still unknown when SBP2 leaves the ribosome. We propose that SBP2 dissociates during or after trans-peptidation due to conformational changes occurring in the ribosome since SBP2 was found to bind to the SECIS element in the pre-translocation complex obtained in RRL in the presence of emetine.

It was reported earlier by the Strasbourg and Driscoll laboratories that the C-terminal 2/3 of SBP2 are involved in SECIS and ribosome binding (Copeland et al., 2001; Allmang et al., 2002; Kinzy et al., 2005). In addition, the region of SBP2 important for 60S subunit-SBP2 interaction was determined by the Copeland and Karpova/Krol laboratories (Donovan et al., 2008; Takeuchi et al., 2009). However, the molecular environment of SBP2 on the ribosome had not been identified when I started my thesis work. There are currently two commonly used biochemical methods to study the topography of ribonucleoprotein complexes: (i) crosslinking by means of bifunctional reagents and (ii) hydroxyl radical footprinting. In our case, the crosslinking data revealed that the 28S rRNA plays a key role in the formation of the ribosomal binding site of SBP2. The data are in agreement with an earlier suggestion based on co-sedimentation of the 28S rRNA and SBP2 in glycerol gradient (Copeland et al., 2001).

The hydroxyl footprinting data established that SBP2 can be placed on the solvent side of the 60S subunit, at proximity of the L7/L12 stalk and the ribosome-EF1A interaction site. The EFSec interaction site on the ribosome is unknown but one can assume that EFSec and eEF1A should share the same binding site since they possess the same function. The footprinting data also support our suggestion that SBP2 may leave the ribosome during ribosome conformational changes after aa-tRNA accommodation. Interestingly, the position of SBP2 on the ribosome that we proposed seems to contradict the SBP2-L30 exchange at the SECIS element, as discussed in (Chavatte et al., 2005). According to a number of publications, L30 is located at the interface between the 40S and 60S subunits where it takes part in intersubunit bridge B4 formation (Figure 37) (Halic et al., 2005; Chandramouli et al., 2008; Arnache et al., 2010a,b; Ben-Shem et al., 2010). Therefore, dissociation of L30 from the ribosome seems unlikely since this would impair subunit interaction. Comparing the distance between the SBP2 binding site and the L30 position, it is difficult to imagine how L30 could displace SBP2 from the SECIS element. Additionally, SBP2 and L30 share the same L7Ae RNA binding domain, and bind RNA targets containing Kink-turn (or Kink-turn like) motifs via the L7Ae domain (L30 is described in 4.3.3 section of Introduction). Chavatte et al. (2005) observed an interaction between the SECIS and L30. Thus L30, when not bound to the ribosome, could unspecifically interact with the Kink-turn-like motif of the SECIS element. This is especially likely taking into account that L30 exists in the free state in addition to being bound to the ribosome.

With regard to the footprinting data, one should keep in mind that the 28S rRNA region protected from hydroxyl radicals upon SBP2 binding may represent a region larger than the actual SBP2 binding site. Intuitively, one would expect SBP2 to bind one of the many K-turn motifs of the 28S rRNA since this protein has affinity for such RNA structural motifs. However, the helices we found protected by SBP2 do not contain any K-turn. In fact, the ribosome binding domain (lysine-rich region) and the SECIS binding domain of SBP2 are physically separated and only partially overlap. Meaning that SBP2 may not necessarily bind to a ribosomal RNA K-turn motif. In the 3D structure of the ribosome, these helices may just be located in the vicinity of the 28S rRNA fragment that actually interacts with SBP2. Further studies are needed to map it precisely.

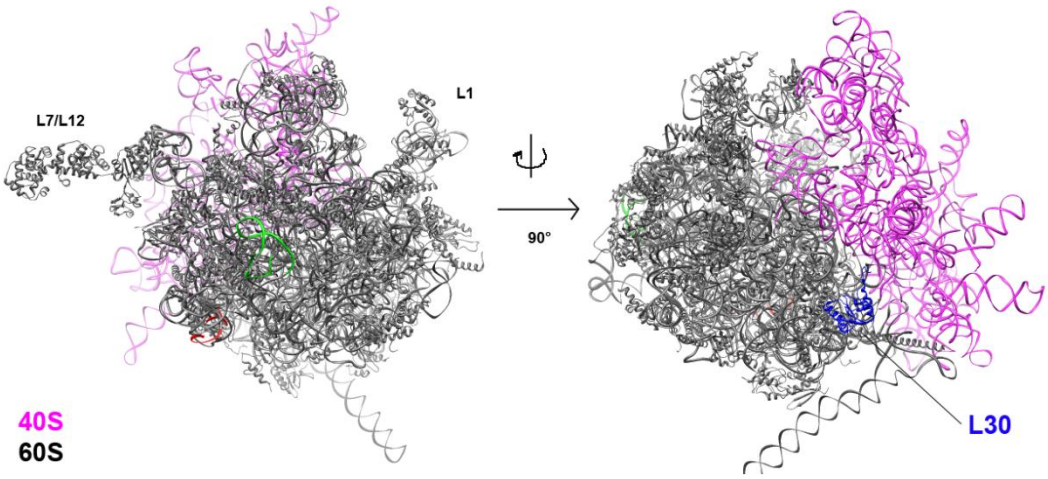

Figure 37. Position of L30 on the eukaryotic ribosome. The structure of the 80S ribosome from *T.aestivum* is presented (Armache et al., 2010a,b). The 60S subunit is shown in grey (both proteins and rRNAs), the 40S subunit in cyan (only 18S rRNA backbone is shown). The putative interaction site of SBP2 is circled. L30 is shown in dark blue. This figure was generated by Chimera ver. 1.5.2 (Pettersen et al., 2004).

* * *

The mechanism of selenocysteine biosynthesis and insertion is indeed very complex and requires several additional factors (Allmang et al., 2009). In humans, about 25 selenoproteins are synthesized while in *C.elegans* the whole selenocysteine-insertion machinery has been kept to ensure synthesis of one single selenoprotein. Why did this machinery appear or why is it still kept?

When selenoproteins were first discovered and characterized, it was widely discussed that the highly oxidizable Sec could be a relic of an earlier form of genetic code when no or very little O_2 was present in the Earth atmosphere (Leinfelder et al., 1988). Later, Sec was considered as an innovation that occurred since Sec-containing enzymes might have a catalytic advantage over the Cys-containing versions (Osawa et al., 1992). Nowadays, it is believed that the advantage of Se over S in catalyzing reactions is less evident. One of the most discussed differences between the two amino acids is the lower pK_a of –SeH compared to –SH (5.2 versus 8.5, respectively, in the free state). At physiological pH, most of the Cys would be found in the protonated thiol rather than in the thiolate form, while the selenol group of Sec would be deprotonated and, thus, more nucleophilic (Gromer et al., 2003). However,

this theoretical situation is unlikely to happen in proteins where the polarisability of functional groups in amino acids is influenced by their neighbors. This issue was studied in detail for the thioredoxin reductase. In this enzyme, the active site is formed by either Gly-Sec-Cys-Gly (mammalian Trx) or Ser-Cys-Cys-Ser in selenium-devoid enzymes (Trx of *D.melanogaster*, for example) (Kanzok et al., 2001). In the latter, the flanking Ser residues led to a decrease of the pK_a of the Cys thiol group and to a concomitant increase of its nucleophilicity. The Ser residues are thus crucial for the enzyme activity (Gromer et al., 2003). Because of the inherent pause of the ribosome at the UGA Sec codon, the yield of selenoprotein synthesis is rather poor. However, it was measured that the catalytic activity of Sec-containing enzymes is roughly 5-fold higher in vitro than that of their Cys-containing counterparts, thus, the apparent lower catalytic activity of Cys-containing enzymes may be compensated by the higher yield of their synthesis, thereby a higher enzymatic concentration (Gromer et al., 2003). The burning question of why has Sec appeared in enzymes remains still opened. Recently, another possible explanation was put forward. Hondal and Ruggles (2010) compared the possibility of oxidation of –SeH and –SH, and it appeared that selenoenzymes are able to resist to inactivation by oxidation whereas their Cys counterparts are not. Even though the selenol group –SeH is much more sensitive to oxidation to the seleninic version $–SeO_2^-$ than the –SH thiol, $–SeO_2^-$ displays the superior ability to be recycled back to its parent form –SeH compared to the sulfinic version $–SO_2^-$. This interesting feature of the selenol group in selenoenzymes may explain why the components of the UGA-decoding machinery are maintained in the evolution at the expenses of a high energetic cost. (Hondal and Ruggles, 2010).

Part IV

Conclusions and perspective

Conclusion and perspectives

To conclude, I would like to summarize the main results of the thesis. The translation of selenocysteine mRNA was studied in rabbit reticulocyte lysate using photoreactive mRNA derivatives. Based on the results obtained, the mechanism for UGA decoding was proposed. Moreover, the binding site of SBP2, a key factor of the selenocysteine insertion process, was mapped on the human 60S ribosomal subunit. We proposed that the SBP2 binding site is located on the solvent side of the 60S subunit, in the vicinity of the L7/L12 stalk. Lastly, a set of synthetic selenoprotein mRNAs was developed that can be used in future studies of the selenocysteine insertion mechanism.

Altogether, our studies have provided novel important insights into the UGA selenocysteine reprogramming mechanism, giving also new detailed information on the molecular principle of the SECIS-ribosome and SBP2-ribosome interactions.

In perspective, and to proceed further to obtaining a deeper understanding of the mechanism, it would be necessary to study it in an *in vitro* system assembled from purified components. It would be interesting to recapitulate the steps of the SBP2/EFSec/Sec-tRNA^{Sec}/GTP binding to the 80S ribosome and the subsequent accommodation of the tRNA^{Sec} in the A-site. To this end, the set of synthetic mRNAs bearing crosslinking groups in the SECIS, either statistically distributed or at defined locations, and added translation factors (eIFs 1, 2, 3, 5) may be used and the SECIS contacts in the reconstructed complex could be studied. By recapitulating the stages of binding of the Sec-tRNA^{Sec} in the A-site, it would be possible to answer at which step the SBP2-ribosome interaction occurs. Further steps worth of investigation are: (i) the 80S initiation complex, where the UGA Sec codon of the synthetic mRNA is placed in the A-site; (ii) the 80S elongation complex with SBP2/EFSec/Sec-tRNA^{Sec}/GDPPNP; (iii) the 80S complex obtained with the participation of SBP2/EFSec/Sec-tRNA^{Sec}/GTP, in which trans-peptidation is blocked by anisomycin; (iv) the same complex as in (iii) but without antibiotic and with eEF2-GTP, e.g. the post-translocational complex with Met/Sec-tRNA^{Sec} in the P-site. Only if conditions can be found to assemble these complexes in sufficient amounts, then it will become interesting to validate by cryo-electron microscopy the information obtained by the crosslinking technique. Alternatively, it is worth trying to purify the flRNA•ribosome complex formed in RRL in the presence of anisomycin using 3’

biotinylated flRNA and to study it with cryo-EM (a collaboration is envisaged with the Bruno Klaholz group from IGBMC).

Meanwhile, we planned to confirm our finding of the localization of SBP2 on the 60S subunit by other approaches. A work has already been initiated with the group of Eric Westhof (IBMC, Strasbourg) to derive a secondary structure model of expansion segment 7 of the 28S RNA, by structure-based sequence alignments of about 50 ribosomal RNA species, ranging from protozoa to humans.

Part V

Materials and methods

Materials and methods.

1. Materials

Chemicals used in the study were from Sigma-Aldrich, Fluka, Merck, Roth or Euromedex. 4-thio-UTP and Biotin-11-CTP were from Jena Bioscience (Germany). 5-(3-aminoallyl)UTP was kindly provided by Dr. Alexey Malygin (Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia). Enzymes were purchased from New England Biolabs, MPBio, Epicentre, Promega, Fermentas. Nucleosil 5C18 was from Macherey-Nagel (Germany), Sephadex G25 and G15 superfine were from GE Healthcare.

Oligodeoxyribonucleotides were purchased from Sigma-Aldrich or were synthesized in the Laboratory of Biotechnology, Institute of Chemical Biology and Fundamental Medicine (Novosibirsk, Russia).

[α-^{32}P]ATP, [α-^{32}P]GTP and [γ-^{32}P]-ATP (3000 Ci/mmol) were purchased from Hartmann Analytic (Germany) or synthesized in the Laboratory of Biotechnology, ICBFM. ^{35}S methionine (1000 Ci/mmol) was bought from Hartmann Analytic.

4-(N-2-chloroethyl-N-methylamino)benzylamine ($ClRCH_2NH_2$) and N-oxysuccinimide ester of 4-azidotetrafluorobenzoic acid were synthesized in the Laboratory of Biothechnology (ICBFM).

Table 6. Oligonucleotides list.

	Oligoribonucleotides
name	**Sequence (5' to 3')**
5'SelN	gggugcagcccaugauggcugaauccgaaa
3'SelN	(s4u)ccucgauggguccagcuugaugucuuugcagcaucc
	Oligodeoxyribonucleotides
name	**Sequence (5' to 3')**
5'UTR-1	aaattaatacgactcactatagggctttccagccagcgccgagcg
3'UTR-1	gagacagcagcactgcaactgccaagcagccggggtaggaggggcgcc
Sec-1	ggtaggaggggcgccttagaagaatcacattcgctcggcgctgg
Phe-1	ggtaggaggggcgccttagaagaagaacattcgctcggcgctgg
5'UTR-2	aaattaatacgactcactatagggagaaaaaagaaagaaatg
3'UTR-2	(2'OMe-G)(2'OMe-C) - -ttccttcgtgtctttgtctttttttcttttttctttttctttcttct
Sec-2	ttcttttctttcttcttagaagaatcacatttctttctttttttc
Phe-2	ttcttttctttcttcttagaagaagaacatttctttctttttttc
Link-1	gtgatcaccaagcccgagacagcagcactg
Link-2	gtgatcaccaagcccgcttccttcgtgtct
Link_SelN1	tgggctgcaccctcttccttcgtgtct
Link_SelN2	gacccatcgaggatttcggattcagccatc
GPx-26	gtgtttctccatgtaaat*t*
GPx-34	tgtaaatttagaggaaacac
GPx-54	cggaaatcaggtgtttctcca
Helper-1	ctcggaaatcaggtgtt
Helper-2	caccgtcatggtttt
SelN-24	tcggattcagccatcatgg
SelN-33	tcgaggatttcggattcag
SelN-51	gacatcaagctggacccat
Helper-A33	aaagacatcaagctggacc
Oligo_s	cgagctcggtacctgcg
TP oligo-1	gagacagcagcactcgaac
TP oligo-2	gcttccttcgtgtctt

name	Sequence (5’ to 3’)
28S_196	cctcgatcagaagga
28S_394	ctcttgaactctctcttc
28S_508	cgggaaagatccgccg
28S_693	gtggaaatgcgcccg
28S_735	ccttcccagccgtcccg
28S_948	gggattcggcgagtgctgc
28S_1178	cggggcgcactggggac
28S_1221	gtggccccgagagaacctc
28S_1298	acgtcgccgcccgaccc
28S_1435	ctcggcggactggagaggc
28S_1589	cgatttgcacgtcagg
28S_1792	ggcttcttacccattta
28S_1991	cacactccttagcggat
28S_2160	tccgacgcacaccacacg
28S_2313	taggcttcaaggctcacc
28S_2406	tgctgttcacatggaa
28S_2618	caaagaaaagagaactctccc
28S_2795	accttggagacctgct
28S_2889	ccaatccttatcccgaag
28S_3047	gagggagcgagcggcgc
28S_3142	cgccccgacccttctcc
28S_3260	ggcgatccacgggaagg
28S_3626	attaacagtcggattccc
28S_3809	agtgggaatctcgttcatcc
28S_4036	gtggtatttcaccggcggc
28S_4184	gccccagtcaaactcc
28S_4361	ttctgacacctcctgct
28S_4409	ccacaagccagttatccct
28S_4540	cagtagggtaaaactaac
28S_4721	gcgctgccgtatcgttcc
28S_4880	ccgtttcccaggacgaag
28S_5010	acgtacgaaaccccgacc

Buffer name	Buffer composition
T7	120 mM HEPES-KOH pH 7.8, 22 mM $MgCl_2$, 1 mM spermidine, 20 mM DTT, 0.1 mg/ml BSA
buffer A	20 mM Tris-HCl pH 7.5, 200 mM KCl, 20 mM $MgCl_2$, 0.5 mM EDTA
elution buffer B	0.3 M NaOAc pH 5.5, 1% SDS, 0.5 mM EDTA
loading buffer C	0.01% BP, 0.01% XC in formamide
buffer D	20 mM HEPES-KOH pH7.8, 100 mM KOAc, 2 mM $Mg(OAc)_2$, 0.25 mM spermidine, 10 mM sodium creatine phosphate, 1 mM ATP, 0.5 mM GTP
buffer RT	50 mM Tris-HCl pH 8.3, 40 mM KCl, 6 mM $MgCl_2$, 10 mM DTT
buffer E	4M urea, 1 mM EDTA-KOH pH 7.5, 5 mM DTT
buffer F	0.3M NaOAc, 1% SDS, 0.5 mM EDTA
buffer G	20 mM Tris-HCl pH 7.5, 125 mM KCl, 10 mM $MgCl_2$, 0.5 mM EDTA, 4 mM β-mercaptoethanol
buffer H	20 mM Tris-HCl pH 7.5, 100 mM NH_4Cl, 3 mM $MgCl_2$, 0.125 mM EDTA, 4 mM β-mercaptoethanol
buffer I	20 mM Tris-HCl pH 7.5, 100 mM NH_4Cl, 3 mM $MgCl_2$, 0.125 mM EDTA, 10 mM β-mercaptoethanol
buffer J	20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 500 mM KCl 0.5 mM EDTA, 10 mM β-mercaptoethanol
buffer K	20 mM HEPES-KOH pH 7.8, 100 mM KCl, 2 mM $MgCl_2$
buffer L	20 mM HEPES-KOH pH 7.8, 100 mM KCl, 4 mM $MgCl_2$
buffer M	20 mM Tris-HCl pH 7.5, 2 mM EDTA, 0.1% SDS
buffer N	20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 500 mM KCl 0.5 mM EDTA
PBS	8 mM Na_2HPO_4, 1.7 mM $KH_2PO_{4,,}$ 150 mM NaCl, 2.7 mM KCl

Table 7. Buffer solutions used.

2.1. flRNAs and synthesis of its photoreactive derivatives

2.1.1. 5' RNA synthesis

The DNA matrix for 5'RNA was obtained by PCR. A typical reaction contained 1μM of primers "5'UTR" and "3'UTR" and 0.05 μM of "Sec" or "Phe" primer in 50 μl of DyNAzyme buffer (Finnzymes) supplemented with 0.2 mM dNTPs and 1 U of DNA polimeraze DyNAzyme EXT (Finnzymes). Amplification was carried out in Progene

amplificator (Techne), using the following program: 30 cycles at 94˚C for 30 sec, 55˚C for 30 sec, and 68˚C for 30 sec. The PCR products were analyzed on 1% agarose gel. The DNA matrices obtained were used to synthesize 5'Phe and 5'Sec mRNAs by T7 transcription *in vitro*.

2.1.2. 3' RNA synthesis by T7 transcription (5' monophosphate containing RNA)

The 3' RNA was either the GPx1, GPx4 or SelN SECIS elements (plasmids used were pT7GPx1, pT7GPx4 or pT7SelN (Walczak et al., 1996; Fagegaltier et al., 2000)). Typically, the plasmid was linearized with EcoRI and served as the DNA template in T7 transcription reaction. A typical reaction contained 10 µg of DNA template, 4 mM ATP, CTP and UTP, 2 mM GTP and 8 mM GMP, 20U of RNasin (Promega) and 1 µl of home-made T7 RNA polymerase (10 mg/ml) in 100 µl of T7 buffer.

2.1.3. RNA ligation using T4 DNA ligase

The 5'RNA (either with natural UTRs or A-rich), splint oligodeoxyribonucleotide (link1 or link2) and 3'RNA (the SECIS element), containing guanosinmonophosphate at the 5' end, were mixed in a 2:2:1.5 ratio and incubated for 5 min at 85˚C, then for 5 min at 4˚C. The reaction mix was subsequently supplemented with T4 DNA ligase buffer (Epicentre), 40 U of Rnasin (Promega) and T4 DNA ligase (Epicentre), at 0.5 U enzyme/pmol of 3'RNA. The reaction was carried out at 16˚C for 15-20 h. Two volumes of buffer F was added and the RNA was phenol-chloroform extracted. The full-length synthetic RNA was gel purified (10% polyacrylamide denaturation gel, 16 cm long).

For preparative ligation of the full-length mRNA, containing the SECIS element with statistically distributed 4-thiouridines, a typical reaction contained 150 pmol of 5'RNA, 150 pmol of splint oligonucleotide and 120 pmol of ^{32}P-internally labeled SECIS (1000 cpm/pmol), 40 U of RNasin and 100 U of T4 DNA ligase (Epicentre) in 100 µl of T4 DNA ligase buffer (Epicentre). After gel purification and elution, the yield of full-length RNA was 12-15%.

2.1.4. Full-length DNA amplification

The DNA matrix corresponding to the full-length synthetic mRNA was obtained by reverse transcription and subsequent PCR amplification. For reverse transcription, 10 µl of the ligation reaction mix was supplemented with 30 pmol of oligo_s primer and 12.5 µl of H_2O. The reaction mix was incubated at 85˚C for 3 minutes and then placed on ice. The reaction mix was next supplemented with MMLV RT buffer (Promega), 0.5 µl of 10 mM dNTP mix, 10 U of RNasin and 200 U of MMLV reverse transcriptase (Promega). The reaction was incubated for 1 h at 37˚C. 0.2 µl of RT reaction was added to the PCR reaction mix. The DNA was amplified using oligo_s and 5'UTR primers. The following PCR program was used (Alkami biosystem, 1999):

1 cycle: 94˚C, 2 sec;
20 cycles: 94˚C for 30 sec, 65˚C to 45˚C for 30 sec, 72˚C for 35 sec;
25 cycles: 94˚C for 30 sec, 45˚C for 30 sec, 72˚C for 35 sec;
1 cycle: 72˚C for 7 min.

The PCR products were analyzed on 1% agarose gel.

2.1.5. ^{32}P-labeled mRNA synthesis

A typical reaction (100 µl) contained 10 µg of DNA template, 4 mM of CTP, UTP, 0.4 mM ATP and GTP, 50 µCi of [α-^{32}P]ATP , 50 µCi of [α-^{32}P]GTP , 20 U of RNasin (Promega) and 1 µl of T7 RNA polymerase (10 mg/ml, homemade) in 100 µl of T7 buffer. The mixture was incubated for 60 min at 37°C; GTP and ATP were next added to 4 mM and the reaction was incubated for 3 h at 37 °C. RNA was phenol-chloroform extracted, precipitated with ethanol and either desalted using G15 mini spin columns (Thermo Scientific) or gel purified.

2.1.6. Synthesis of RNA with statistically distributed 4-thiouridines

To obtain ^{32}P-labeled RNA with statistically distributed 4-thiouridines, the UTP concentration was reduced to 2 mM and s4-UTP was added to 2 mM in the typical T7 transcription reaction (described in 2.1.5).

2.1.7 Synthesis of RNA with statistically distributed –N_3 group

N_3-aUTP was synthesized from aa-UTP and the N-oxysuccinimide ester of 4-azidotetrafluorobenzoic acid. Prior to reaction, 6 OD units of aaUTP were mixed with 20 µl of 8% cetavlon, dried in SpeedVac (Univapo) and dissolved in 50 µl "dry" DMSO. An equal volume of 0.5M N-oxysuccinimide ester of the 4-azidotetrafluorobenzoic acid in DMSO was added to aaUTP in DMSO and the reaction was incubated for 1 h at room temperature in the dark. The resulting N_3-aUTP compound was recovered by precipitation with 2% $LiClO_4$ in acetone and dissolved in H_2O.

To obtain ^{32}P-labeled RNA with statistically distributed N_3-aUMPs, the UTP concentration was reduced to 2 mM and the N_3-aUTP was added to 1 mM in the typical T7 transcription reaction (described in 2.1.5).

2.1.8. Site-directed introduction of -N_3 to flRNA

The 4-(N-2-chloroethyl-N-methylamino)benzyl-5' (or 3') -phosphoramides of oligodeoxyribonucleotides complementary to the SECIS GPx1, GpX4 or SelN sequences were synthesized and purified as described in (Malygin et al., 2003). 1.5 OD_{260} of oligonucleotide (listed in Table 6, oligonucleotides Gpx-26, 34, 54 and SelN-24, 33, 51), bearing a phosphate group at the 5' or 3' end, were dissolved in 10 µl of H_2O and 3 µl of 8% cetavlon solution in H_2O was added. The mixture was then dried and the pellet was dissolved in 20 µl of "dry" DMSO. A typical reaction contained 0.15 mM oligonucleotide, 0.25 M $ClRCH_2NH_2$, 0.5M condensing agents (triphenylphosphine and dipyridildisulfide) and 0.9 M N-methylimidazole as the catalyst in 60 µl of DMSO. The reaction was incubated at 20 $^{\circ}$C for 1 h. The oligonucleotide was then precipitated with 2% $LiClO_4$ in acetone, and the pellet dissolved in water and precipitated again with 2% $LiClO_4$ in acetone. The modified oligonucleotide was isolated by reverse phase HPLC (micro column volume 200 µl, sorbent – Nucleosil 5C18). The oligonucleotide was eluted by 10-20-30-40-90% methanol in TEA-OAc pH 7.4 (elution speed 100 µl/min). The modified oligodeoxyribonucleotide eluates later than the unmodified one. The fractions containing the modified oligodeoxyribonucleotide were precipitated with 2% $LiClO_4$ in acetone, dried on air, dissolved in H_2O and kept in liquid nitrogen prior to use. The yield of modified oligonucleotide was 50-60%. The oligonucleotide obtained was further used for site-specific alkylation of flRNA. To do this, 300 pmol of ^{32}P-labeled RNA (1000 cpm/pmol) was incubated with 3 nmol of alkylatying oligonucleotide and

3 nmol of helper oligonucleotide (Table 6, helper-1, 2 or A33) in 150 µl of buffer A for 18 h at 25 °C. Further, the RNA was ethanol precipitated, the pellet dissolved in RNA loading buffer C and the modified RNA was separated from unmodified one using 8% denaturating PAGE. The covalent adduct-containing RNA migrated slower than the unmodified RNA. Bands corresponding to the adduct containing RNA were cut out and the RNA was eluted overnight in buffer B. RNA was next ethanol precipitated and dissolved in H_2O.

To hydrolyze the phosphoramide bond in the oligonucleotide derivatives attached to the RNA and, therefore, to liberate the site-specifically introduced $–NH_2$ group, the adduct containing RNA was incubated in 0.3M NaOAc pH 4.4 for 6 h at 50 °C. The RNA was then ethanol precipitated, dissolved in RNA loading buffer C and the modified RNA was purified by 8% denaturating PAGE and eluted from the gel. The gel mobility of the resulting modified RNA was similar to that of the unmodified flRNA, as an evidence of the removal of the oligodeoxyribonucleotide moiety and liberation of the RCH_2NH_2 group attached to RNA.

The RNA obtained after phosphoramide bond hydrolysis was dissolved in 50 µl of 50 mM HEPES-KOH, pH 9.0 and mixed with an equal volume of 20 mM N-oxysuccinimide ester of the 4-azidotetrafluorobenzoic acid in DMSO. The mixture was incubated in the dark at room temperature for 90 min. The low molecular mass organic compounds were then extracted with 3 volumes of ether. The modified RNA was ethanol precipitated, reprecipitated twice with ethanol from 50% DMSO in water, dissolved in water and stored in liquid nitrogen until used.

2.1.9. flRNA synthesis bearing site-specifically introduced 4-thiouridine residues

The flRNA, containing a single 4-thiouridine, was synthesized using T4 DNA ligase, three RNAs (5'RNA, oligoribonucleotides 5'SelN and 3'SelN) and 2 splint oligonucleotides, link-SelN1 and link-SelN2 (Table 6). The 5'RNA was obtained by T7 transcription. The 5'SelN RNA, and the 3'SelN RNA containing a 4-thiouridine at its 5' end, were synthesized by Dharmacon, USA. Prior to ligation, the 5'ends of the 3'SelN and 5'SelN were phosphorylated, with ATP in the case of the 5'SelN and with both ATP and [γ-^{32}P]ATP in the case of the 3'SelN. The technique was developed in the laboratory of A. Krol by Artemy Beniaminov (unpublished data).

Prior to ligation, 2 splint oligonuclotides and 2 RNAs (5'RNA and 5'SelN) were incubated at 85 °C for 5 min and then placed on ice. The 3'SelN RNA and other reaction

components were added subsequently. A typical ligation mix contained 100 pmol of 5'RNA, 100 pmol of link-SelN1, 100 pmol of 5'SelN RNA, 100 pmol of link-SelN2 and 75 pmol of 3'SelN, 40 U of RNasin and 100 U of T4 DNA ligase in 150 µl of T4 DNA ligase buffer (Epicentre). The mixture was incubated for 15 h at 16 °C. The RNA was phenol-chloroform extracted, ethanol precipitated and the pellets were dissolved in buffer C. The full-length RNA was isolated by 10% denaturating PAGE. The yield of flRNA was 5-8%. Alternatively, the synthesis was carried out in two steps: SECIS SelN was ligated first, gel purified and then ligated to 5'RNA.

2.1.10. mRNA capping

mRNAs were capped using the ScriptCap m7G capping system from Epicentre Biotechnologies. Prior to capping, the mRNA was heat-denaturated for 5 min at 75°C. A typical reaction was assembled on ice and contained 100 pmol of RNA, 1 mM GTP, 0.1 mM SAM, 20 U ScriptGuard RNase inhibitor and 10 U of ScriptCap capping enzyme in 20 µl of ScriptCap buffer (50 mM Tris-HCl pH 8.0, 6 mM KCl, 1.25 mM $MgCl_2$). The mix was incubated for 1 h at 37°C. After completion of the reaction, the RNA was purified using mini spin-columns (Thermo Scientific, USA) filled with 500 µl of Sephadex G15.

2.2. Treatment of rabbit reticulocyte lysate with micrococcal nuclease

To hydrolyze the endogenous mRNAs, rabbit reticulocyte lysate (RRL) was treated with micrococcal nuclease (Sigma). Typically, 100 µl of lysate were supplemented with 4 µl of 100 mM $CaCl_2$, 8 µl 1 mM hemin, 1 µl of 5 mg/ml creatine phosphokinase and 100 U of micrococcal nuclease. The mix was incubated for 20 min at 20°C, then placed on ice. The reaction was stopped by addition of 8 µl 100 mM EGTA-KOH pH 8.0. The treated lysate was either immediately used or frozen in liquid nitrogen and stored at -70°C before use.

The ability of RRL to support protein synthesis from exogenous mRNA and the protein synthesis from the endogenous mRNA were checked as described in (Rabbit reticulocyte system, Promega technical manual). The reaction mix contained 10 µl of lysate in buffer D, 0.5 mM amino acids mix, 2 µl of ^{35}S-methionine and either 1 µg of RNA T7luc (Promega, USA) or 1 µl of H_2O. 4 aliquots were taken at 0, 20, 60 and 90 min and kept on ice before the last one was taken. After the last aliquot was taken, they were diluted in 500 µl of H_2O, supplemented with 500 µl of 1M NaOH, thoroughly mixed and incubated for 15 min at

37°C. 500 µl of 50% cold TCA was added and the mixes were incubated for 30 min on ice. After incubation, the mixes were filtered through Whatmann GF/C glass fiber filters, each filter was washed 3x with 1 ml of 5% cold TCA, and then washed 2x with acetone and dried on air. Each filter was placed in 4 ml of scintillation liquid Ecoscint O (National diagnostics, USA) and the radioactivity signal was measured in the ^{14}C channel of a Rackbeta counter.

2.3. Toe-printing

The reaction mix contained 50% (v/v) of micrococcal nuclease treated RRL, 10 U of RNAsin, 0.25 mM amino acids mix in 10 µl of buffer D. Prior to adding mRNA, the reaction mixes were supplemented with either 2 mM GMPPNP (water solution) or 5 mM emetine (water solution) or 5 mM anisomycine (stock solution in DMSO) or 1 µl of H_2O for the control reaction. SBP2 was added in a 3-fold molar excess to the mRNA (1 pmol of SBP2 per reaction), when required. Then the mixtures were pre-incubated for 5 min at 30°C to allow antibiotic binding to the target. 0.3 pmol of capped mRNA was then added and the mix was incubated for an additional 5 min at 30°C. The mix was supplemented with either 1 µl of anisomycine (50mM) or emetine (50 mM) or H_2O (in case of control reaction and "GMPPNP" reaction), 1 µl of ^{32}P-labeled primer (5 pmol, 50000 cpm/pmol, TP oligo-1 or TP oligo-2) and 8 µl of reverse transcription mix (1.25 mM dNTPs and 5 U of AMV reverse transcriptase in RT buffer). Reverse transcription was carried out for 30 min at 30°C. The resulting cDNA products were phenol-chlorophorm extracted, ethanol precipitated, dissolved in loading buffer C and analyzed by 10% PAGE-8M urea. After electrophoresis, the gel was dried, exposed to a BioRad Phosphorimager plate and analyzed using the Quantity One program (BioRad, USA).

2.4. mRNA-ribosome binding in RRL

The reaction mix contained micrococcal nuclease treated RRL (50% v/v), 10 U of RNasin, 0.25 mM amino acids mix in 200 µl of buffer D. Prior to adding the mRNA, the reaction mixtures were supplemented with either 2 mM GMPPNP or either 5 mM of emetine or anisomycine, and incubated at 30 °C for 10 min. ^{32}P-labeled capped mRNA (either flRNA or 5'RNA) was renaturated prior to adding to the reaction mix. After mRNA addition, the reaction was incubated at 30 °C for 15 min. The mixture was then layered onto a 5-40% sucrose gradient in buffer D and centrifuged for 2.5 h at 52000 rpm (rotor TST60) on an L8

Beckman centrifuge. After centrifugation, the gradients were collected and the A_{260} of each fraction was measured.

To check the mRNA-ribosome binding, the radioactivity of each gradient fraction was measured and the binding efficiency was calculated (mol mRNA/mol of ribosomes). To check the SBP2 content in the 80S or 48S complexes, the fractions corresponding to either the 48S or 80S were ethanol precipitated and the pellet content was loaded onto a 10% SDS-PAGE which was blotted onto nylon membranes Immobilon (Millipore). The SBP2 signal was detected with rabbit polyclonal anti-SBP2 antibodies (1/5000 dilution in PBST). Membranes were treated with anti-rabbit HRP-conjugated secondary antibody (1/10000 dilution), revealed with the ECL plus kit (GE Healthcare) and exposed to either X-ray film or ChemiDoc XRS.

2.5. mRNA – ribosomes crosslinking in RRL

The mRNA-ribosome complexes were assembled in RRL as described in 2.4, with the only modification that, in this case, a photoreactive ^{32}P-labeled mRNA derivative was used. To generate a crosslink, the reaction mixtures were placed on ice in 96-well plates containing 25 µl aliquots and irradiated for 2 min with a SpotCure UV lamp (λ > 290 nm). Control reactions were not irradiated and kept on ice.

To identify crosslinks with ribosomal proteins, reaction mixtures were layered on 10-30% sucrose gradients in buffer D and centrifuged at 22000 rpm, at 4 °C for 17 hours in an SW40 rotor (L8 Beckman centrifuge). After centrifugation, fractions corresponding to 80S or 48S complexes were ethanol precipitated. The pellets were dissolved in 50 µl of buffer E and incubated at 37 °C for 10 min to dissociate the ribosomal particles into proteins and rRNA. 20 µg of RNase A was then added to hydrolyze the RNA. After 1 hour incubation with RNase A, reactions were devided into two part. The proteins from one half were precipitated with 6 volumes of acetone, while another half was supplemented with 10 µg of proteinese K and incubated for an additional 30 min at 37 °C and precipitated with acetone as well. Pellets were dissolved in SDS loading buffer and loaded onto 14% SDS-PAGE. After electrophoresis, gels were fixed in ethanol/AcOH mixture, dried, exposed to an PhosphorImager plate and analyzed using QuantityOne program (BioRad).

To identify the crosslinks with rRNA, 30 µl of each reaction mixture were supplemented after irradiation with 2 volumes of buffer F, incubated for 15 min at 37 °C. 30 µg of proteinase K was then added and reactions were incubated for 30 min at 37 °C. After

phenol-chloroform extraction, the rRNA was ethanol precipitated and the resulting pellets were dissolved in loading buffer C and loaded onto 5% denaturating PAGE. Gels were stained with EtBr and visualized by ChemiDoc. The gel was then dried, exposed to PhosphorImager plate and analyzed by QuantityOne (BioRad).

2.6. Isolation of human ribosomes from placenta

40S and 60S ribosomal subunits were isolated from human placenta, as described in (Matasova et al., 1991). The time lapse between delivery and tissue homogenization was less than 30 minutes. When ejected, the placenta was immediately placed in 500 ml of cold buffer G and transported to the laboratory on ice. The placenta was cut into pieces, the umbilical cord and the connective tissue were removed. To remove blood, pieces were washed in buffer G and the tissue was homogenized in buffer G, supplemented with 250 mM sucrose, 1 g/l heparin and 0.05 g/l cycloheximide. The resulting homogenate was centrifuged in a JA-14 rotor, in a Beckman J2-21M centrifuge (4°C, 30 min at 13000 rpm). The supernatant was brought to 1% in sodium deoxyhalate and mixed for 30 min at 4°C. It was then layered onto a 35% sucrose cusion in buffer H and centrifuged for 20 h at 4°C at 24000 rpm (Beckman L8 centrifuge, rotor type SW28). After centrifugation, the supernatant was carefully removed. Polysome pellets were homogenized in 10 ml of buffer I on ice. The particles that did not dissolve were removed by centrifugation in a JA-20 rotor, Beckman J2-21M centrifuge (4°C, 15 min at 17000 rpm). The resulting polysome solution was made 0.5 mM in puromicin and incubated for 10 min at 0 °C. Next, the solution was gradually brought to 0.5 M KCl, incubated for 40 min at 37 °C and layered onto a 10-30% sucrose gradient in buffer J . Centrifugation was carried out in a Beckman L8 centrifuge (SW28 rotor) for 20 h at 19000 rpm, at 4°C. Fractions corresponding to 60S or 40S subunits were collected, and KCl and $MgCl_2$ were added to 120 mM KCl and 15 mM, respectively. They were sedimented by ultracentrifugation (Beckman L8 centrifuge, rotor type SW28, 24000 rpm, 20 h, 4°C).

2.7. Recombinant SBP2 preparation

Rat (expression vector pET28a(+)rSBP2, kindly provided by Dr. G. Cavigiolio) and human (aa 344-843, phSBP2(344-854), (Allmang et al., 2002)) bearing N-terminal His-tag SBP2 recombinant proteins were expressed in *E.coli* (strain BL21(DE3)RIL) in 2YT medium supplemented with either 100ng/ml ampicillin (for hSBP2) or 25 ng/ml kanamycin (in the

case of rSBP2). Induction was with 0.45 mM IPTG at 18˚C for 3 h (rat SBP2) or 12 h (human SBP2). SBP2 was purified from soluble fraction using affinity chromatography on Ni-NTA agarose according to Qiagen standard protocol. The elution buffer (0.3M NaCl, 0.25 mM imidazole in PBS, pH 8.0) was exchanged to buffer K by dialysis. The purity of the protein was checked by 10% SDS-PAGE with subsequent CBB R250 staining.

2.8. SBP2•ribosome binding assay

To study the SBP2•ribosome complex formation, 0.6 μM 40S subunits, 60S subunits or 80S ribosomes were incubated with 1.8 μM SBP2 in 50 μl of 4 different buffers (PBS, PBS with 0.5 mM $MgCl_2$, buffer K and buffer L) at room temperature for 15 min. Each complex was isolated by centrifugation on 10-30% sucrose gradient in the corresponding buffer (centrifugation conditions were: rotor type SW41, 23000 rpm, 17 h, 4˚C). The SBP2 content in each complex was analyzed by western blotting with rabbit polyclonal anti-SBP2 antibodies (1/5000 dilution in PBST) and anti-rabbit HRP-conjugated secondary antibody (1/10000 dilution), revealed with the ECL plus kit (GE Healthcare) and exposed to ChemiDoc XRS.

2.9. SBP2-rRNA crosslinking assays

2.9.1. 2-iminothiolane crosslinking

Ribosomal subunits were reactivated in buffer K. 1.25 μM 60S or 80S ribosomes were then incubated with 3 μM SBP2 in 100 μl of buffer K at room temperature for 20 min to allow SBP2•ribosome complex formation. Subsequently, 4 μl of 0.3M 2-iminothiolane (freshly made solution) were added and the reactions were incubated at room temperature for 1 h. The excess of crosslinking reagent was removed by dialysis against buffer K for 2 h using Slide-a-lyzer mini dialysis units (ThermoScientific). The reactions were placed on a glass dish on ice and were irradiated using Stratalincker (Stratagene) UV lamps (λ = 254 nm) for 5 min at a distance of 3 cm to the surface of the solution. After irradiation, the mixtures were supplemented with 3% β-mercaptoethanol and incubated at 37˚C for 30 min. The control reaction had no 2-iminothiolane added and were not irradiated. Uncrosslinked SBP2 was removed by centrifugation in 10-30% sucrose gradient under dissociating conditions (buffer N, rotor SW41, L8 centrifuge, 23000 rpm, 16 h). Prior to centrifugation, $MgCl_2$ and KCl were

brought up to 10 mM and 0.5M, respectively. After centrifugation, the peaks corresponding to 60S or 40S subunits were ethanol precipitated and pellets were dissolved in 20 mM Tris HCl pH 7.5. The solutions were made 5 mM in EDTA and 0.5% in SDS and incubated at 37˚C for 30 min. The resulting mixtures were layered onto a 5-20% sucrose gradient in buffer M and centrifuged for 17 h at 27000 rpm in a L8 centrifuge (Beckman), rotor SW41. The gradients were fractionated and the A_{260} of each fraction was measured. The rRNA was ethanol precipitated, the pellets were dissolved in a minimal amount of water and 1/5 of each fraction was blotted onto a nitrocellulose membrane. The signal from SBP2 was detected with rabbit polyclonal anti-SBP2 antibodies (1/5000 dilution in PBST) and anti-rabbit HRP-conjugated secondary antibody (1/10000 dilution), revealed with the ECL plus kit (GE Healthcare) and exposed to ChemiDoc XRS.

2.9.2. Diepoxybutane crosslinking

The SBP2•ribosome complex was formed as described in 2.9.1. 0.2% (v/v) of diepoxybutane was added to each reaction mixture to generate crosslinks. The control reaction had no diepoxybutane added. The mixtures were incubated for 1 h at 37˚C, the reaction was stopped by addition of 2 µl of 1M Tris-HCl (pH 7.5). rRNA isolation and SBP2 detection were done as described in 2.9.1.

2.10. SBP2-rproteins crosslinking assays

2.10.1. 2-iminothiolane crosslinking

The SBP2•ribosome complex formation and the treatment with 2-iminothiolane were carried out as described in 2.9.1. The reactions were supplemented with 1 µl of 10% H_2O_2 and kept on ice for 30 min to promote crosslinking between adjacent sulfhydryl groups by disulfide bond formation. Iodoacetamide was added to a concentration of 40 mM and solutions were incubated for 30 min at room temperature to alkylate free sulfhydryl groups inaccessible to oxidation. Uncrosslinked SBP2 was removed by centrifugation in 10-30% sucrose gradient under dissociating conditions (buffer N, rotor SW41, L8 centrifuge, 23000 rpm, 16 h). Prior to centrifugation, $MgCl_2$ and KCl were brought up to 10 mM and 0.5M, respectively. Complexes that were not subjected to 2-iminothiolane treatment served as a control for uncrosslinked SBP2 dissociation. After centrifugation, fractions corresponding to

ribosomal subunits were ethanol precipitated, pellets were dissolved, ribosomal proteins were extracted with AcOH. To extract proteins $MgCl_2$ was brought up to 100 mM, placed on ice and two volumes of glacial AcOH were added in 6 portions during an hour. The solutions were then incubated on ice for 40 min to form a precipitate of rRNA. rRNA was pelleted by centrifugation and supernatant was taken. Ribosomal proteins were precipitated from the supernatant by six volumes of acetone, the pellet was dissolved in water and crosslinks were analyzed either by 6% SDS-PAGE with subsequent CBB G250 staining or silver staining, or by dot-blot analysis.

2.10.2. Diepoxybutane crosslinking

The SBP2•ribosome complex formation and the treatment with diepoxybutane were performed as described in 2.9.2. Complex isolation and crosslink analysis was performed as described in 2.10.1.

Diepoxybutane crosslinking assay was used to determine the distribution of crosslinked SBP2 between rRNA and ribosomal proteins. To do this, 40 pmol of 60S and 120 pmol of SBP2 were taken and the SBP2•ribosome complex was formed and treated with diepoxybutane as described in 2.9.2. Uncrosslinked SBP2 was removed by centrifugation in 10-30% sucrose density gradient under dissociating conditions (buffer N, rotor SW41, L8 centrifuge, 23000 rpm, 16 h, 4˚C). After centrifugation, fractions corresponding to 60S ribosomal subunits were ethanol precipitated, pellets were dissolved in H_20, and divided in two equal parts. Ribosomal proteins were extracted with AcOH from one part. rRNA was extracted by phenol-chloroform from another part. One half of resulting rRNA and ribosomal proteins was analyzed by dot-blot analysis. The signal from SBP2 was detected with rabbit polyclonal anti-SBP2 antibodies (1/5000 dilution in PBST) and anti-rabbit HRP-conjugated secondary antibody (1/10000 dilution), revealed with the ECL plus kit (GE Healthcare) and exposed to ChemiDoc XRS. The distribution of crosslinked SBP2 was analyzed by densitometry in QuantityOne (BioRad).

2.11. Hydroxyl radical footprinting of the 60S•SBP2 complex

Hydroxyl radical cleavage of 28S rRNA in 60S subunits and their complex with SBP2 was performed by addition of 1 µl of freshly made solution containing 12 mM

$Fe(NH_4)_2(SO_4)_2$, 62 mM ascorbic acid, 25 mM EDTA-KOH and 0.6% H_2O_2 to 10 µl of 60S subunit (0.5 µM) or their complex with SBP2 in buffer K. After 2 min incubation, the reaction was stopped by addition of 10 µl of 0.1 M thiourea. The 28S rRNA was isolated after phenol-chloroform treatment.

For reverse transcription, primers listed in Table 6 (32 primers, starting with 28S: 28S_196, 28S_394, etc) were 5'-labeled with [γ-^{32}P]-ATP (3000 Ci/mmol) and T4 polynucleotide kinase. Reverse transcription reaction mixtures contained 2 pmol of 28S rRNA, 5 pmol of labeled primer (60000 cpm/pmol), 0.25 mM dNTP and 1 U of AMV reverse transcriptase (MP Biomedicals, USA) in 10 µl of buffer RT. To obtain sequencing lanes, the reaction mixtures were supplemented additionally with 0.033 mM ddNTP (one of the four ddNTPs to each reaction). The reaction mixtures were incubated at 42˚C for 30 min. The products of the reaction were ethanol precipitated, dissolved in 5 µl of loading buffer C and analyzed through 10% sequencing PAGE. After the run, the gel was dried and exposed to a BioRad Phosphorimager plate and quantified using Quantity One software (BioRad, USA).

2.12. Structure-based alignment of ES7 by LocARNA

rRNA sequences corresponding to an ES7 part (nt 1170-1193 in the human 28S rRNA) from 10 organisms (9 vertebrates and 1 echinoderm) were aligned with a web-based tool "LocARNA" (Will et al., 2007). Sequences were taken from the following Genbank files: *B.taurus* NR_036644, turtle *C.picta* AY859626, aquarium fish *E.sexfasciatus* FJ872049, *X.laevis* x02995, *G.gallus* EF552813, *G.gorilla* M30951.1, sea urchin AF212171, *H.sapiens* NR_003287, *P.troglodites* M30950.1 , *H.colliei* AF061799).

References

References

Alkalaeva EZ, Pisarev AV, Frolova LY, Kisselev LL, Pestova TV (2006) In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eER3. *Cell* 125: 1125-1136.

Allamand V, Richard P, Lescure A, Ledeuil C, Desjardin D, Petit N, Gartioux C, Ferreiro A, Krol A, Pellegrini N, Urtizberea JA, Guicheney P (2006) A single homozygous point mutation in a 3'untranslated region motif of selenoprotein N mRNA causes SEPN1-related myopathy. *EMBO Rep* 7(4)**:** 450-454.

Allmang C, Carbon P, 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 (10): 1308-1318.

Allmang C, Krol A (2006) 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.

Amberg R, Urban C, Reuner B, Scharff P, Pomerantz SC, McCloskey JA, Gross HJ (1993) Editing does not exist for mammalian selenocysteine tRNAs. *Nucleic Acids Res* 21(24): 5583-5588

Armache JP, Jarasch A, Anger AM, Ville E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix S, Soding J, Westhof E, Wilson DN, Beckman R (2010a) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. *Prot. Natl. Acad. Sci. U.S.A.* 107(46): 19748-19753.

Armache JP, Jarasch A, Anger AM, Ville E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix S, Soding J, Westhof E, Wilson DN, Beckman R (2010b) Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. *Prot. Natl. Acad. Sci. U.S.A.* 107(46): 19754-19759.

Anthony DD and Merrick WC (1992) Analysis of 40S and 80S complexes with mRNA as measured by sucrose density gradients and primer extension inhibition. *J Biol Chem* 267(3): 1554-1562.

Azevedo MF, Barra GB, Naves LA, Velasco LFR, Castro PGG, de Castro LCG, Amato AA, Miniard A, Driscoll D, Schomburg L, Neves FAR (2010) Selenoprotein-related disease in a young girl caused by nonsence mutations in the SBP2 gene. *J. Clin. Endocrinol. Metab*. 95: 4066-4071.

Babaylova E, Graifer D, Malygin A, Stahl J, Shatsky I, Karpova G (2009) Positioning of subdomain IIId and apical loop of domain II of the hepatitis C IRES on the human 40S ribosome. *Nucleic Acids Res* 37(4): 2126-2141.

Ballut L, Marchadier B, Baguet A, Tomasetto C, Seraphin B, Le Hir H (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol* 12(10): 861-869.

Baron C, Sturchler C, Wu XQ, Gross HJ, Krol A, Bock A (1994) Eukaryotic selenocysteine inserting tRNA species support selenoproteins synthesis in Escherichia Coli. *Nucleic Acids Res* 22(12): 2228-2233.

Bäumert HG, Sköld SE, Kurland CG. (1978) RNA-protein neighbourhoods of the ribosome obtained by crosslinking. *Eur J Biochem.* 89(2):353-9.

Behne, D., H. Hilmet, S. Scheid, H. Gessner, and W. Elger. 1988. Evidence for specific selenium target tissues and new biologically important selenoproteins. *Biochim. Biophys. Acta* 996:12–21.

Bellinger FP, Raman AV, Reeves MA, Berry MJ (2009) Regulation and function of selenoproteins in human disease. *Biochem J* 422(1)**:** 11-22.

Ben-Shem A, Jenner L, Yusupova G, Yusupov M (2010) Crystal structure of the eukaryotic ribosome. *Science* 330(6008): 1203-1209.

Berry MJ, Banu L, Chen YY, Mandel SJ, Kieffer JD, Harney JW, Larsen PR (1991) Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. *Nature* 353(6341): 273-276.

Berry MJ, Banu L, Harney JW, Larsen PR (1993) Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *EMBO J* 12(8): 3315-3322.

Böck A (2006) Selenium metabolism in prokaryotes. *Selenium, its molecular biology and role in human health*, second edition. D L Hatfield (Ed). Springer Science + Business Media,: 9-28

Bosl MR, Takaku K, Oshima S, Nishimura S, Taketo MM (1997) Early embrionic lethality caused by targeted disruption of the mouse selenocysteine tRNA gene. *Prot. Natl. Acad. Sci. U.S.A.* 94: 5531-5534.

Boulon S, Marmier-Gourrier N, Pradet-Balade B, Wurth L, Verheggen C, Jady BE, Rothe B, Pescia C, Robert MC, Kiss T, Bardoni B, Krol A, Branlant C, Allmang C, Bertrand E, Charpentier B (2008) The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J Cell Biol* 180(3): 579-595.

Browne GJ and Proud CG (2002) Regulation of peptide-chain elongation in mammalian cells. *Eur. J. Biochem.* 269: 5360–5368.

Budiman ME, Bubenik JL, Miniard AC, Middleton LM, Gerber CA, Cash A, Driscoll DM (2009) Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation. *Mol Cell* 35(4): 479-489.

Buligyn K, Malygin A, Karpova G, Westermann P (1998) Site-specific modification of 4.5S RNA apical domain by complementary oligodeoxynucleotides carrying an alkylating group. *Eur J Biochem* 251: 175-180.

Bulygin K, Chavatte L, Frolova L, Favre A, Karpova G (2005) The first position of a codon placed in the A site of the human ribosome contacts nucleotide C1696 of the 18S rRNA as well as proteins S2, S3, S3a, S30 and S15. *Biochemistry* 44: 2153-2162.

Caban K, Kinzy SA, Copeland PR (2007) The L7Ae RNA Binding Motif Is a Multifunctional Domain Required for the Ribosome-Dependent Sec Incorporation Activity of Sec Insertion Sequence Binding Protein 2, *Mol Cel Biol*, 6350-6360.

Carlson BA, Xu X-M, Kryukov GV, Rao M, Berry MJ, Gladyshev VN, Hatfield DL (2004) Identification and characterization of phosphoseryl-tRNASerSec kinase. *Prot. Natl. Acad. Sci. U.S.A.* 101: 12848-12853.

Castellano S, Morozova N, Morey M, Berry MJ, Serras F, Corominas M, Guigo R (2001) In silico identification of novel selenoproteins in the Drosophila melanogaster genome. *EMBO Rep* 2(8): 697-702.

Castellano S, Novoselov S, Kryukov G, Lescure A, Blanco E, Krol A, Gladyshev V, Guigó R (2004) Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution. *EMBO Rep* 5(1): 71-77.

Castellano S, Lobanov AV, Chapple C, Novoselov SV, Albrecht M, Hua D, Lescure A, Lengauer T,Krol A, Gladyshev VN, Guigo R (2005) Diversity and functional plasticity of eukaryotic selenoproteins: identification and characterization of the SelJ family. *Proc Natl Acad Sci U S A* 102(45): 16188-16193.

Castellano S, Lobanov A, Chapple C, Novoselov S, Albrecht M, Hua D, Lescure A, Lengauer T, Krol A, Gladyshev V, Guigo R (2005) Diversity and functional plasticity of eukaryotic selenoproteins: Identification and characterization of the SelJ family. *Prot. Natl. Acad. Sci. U.S.A.* 102 (45): 16188-16193.

Castellano S, Gladyshev VN, Guigo R, Berry MJ (2008) SelenoDB 1.0 : a database of selenoprotein genes, proteins and SECIS elements. *Nucleic Acids Res* 36(Database issue): D332-338.

Castellano S (2009) On the unique function of selenocysteine - insights from the evolution of selenoproteins. *Biochim Biophys Acta* 1790(11): 2031-2040.

Castets P, MAugenre S, Gartioux C, Rederstorff M, Krol A, Lescure A, Tajbakhsh S, Allamand V, Guicheney P (2009) Selenoprotein N is dynamically expressed during mouse development and detected early in uscle precursors. *BMC Dev* Biol 9: 46.

Chambers I, Frampton J, Goldfarb P, Affara N, McBain W, Harrison PR (1986) The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the 'termination' codon, TGA. *EMBO J* 5(6): 1221-1227.

Chan CC, Dostie J, Diem MD, Feng W, Mann M, Rappsilber J, Dreyfuss G (2004) eIF4A3 is a novel component of the exon junction complex. *RNA* 10(2): 200-209.

Chandramouli P, Topf M, Menetret JF, Eswar N, Cannone JJ, Gutell RR, Sali A, Akey CW (2008) Strucutre of the mammalian ribosome at 8.7 A resolution. *Structure* 16(4): 535-548.

Chapple CE, Guigo R, Krol A (2009) SECISaln, a web-based tool for the creation of structure-based alignments of eukaryotic SECIS elements. *Bioinformatics* 25(5)**:** 674-675.

Chavatte L, Brown BA, Driscoll DM (2005) Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat Sruct Mol Biol* 12(5): 408-416.

Cléry A, Bourguignon-Igel V, Allmang C, Krol A, 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(6)**:** 1868–1884.

Cone JE, Del Rio RM, Davis JN, Stadtman TC (1976) Chemical characterization of the selenoproteins component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. *Prot. Natl. Acad. Sci. U.S.A.* 73(8): 2659-2663.

Copeland PR, Driscoll DM (1999) Purification, redox sensitivity, and RNA binding properties of SECIS- binding protein 2, a protein involved in selenoprotein biosynthesis. *J Biol Chem* 274(36): 25447-25454.

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(2): 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(5): 1491-1498.

de Jesus LA, Hoffmann PR, Michaud T, Forry EP, Small-Howard A, Stillwell RJ, Morozova N, Harney JW, Berry MJ (2006) Nuclear Assembly of UGA Decoding Complexes on Selenoprotein mRNAs: a Mechanism for Eluding Nonsense-Mediated Decay? *Mol Cell Biol* 26(5): 1795-1805.

Dayer R, Fischer BB, Eggen RI, Lemaire SD (2008) The peroxiredoxin and glutathione peroxidase families in Chlamydomonas reinhardtii. *Genetics* 179(1): 41-57.

Diamond A, Dudock B, Hatfield D (1981) Structure and properties of a bovine liver UGA suppressor serine tRNA with a tryptophan anticodon. *Cell* 25(2): 497-506.

Diamond AM, Choi IS, Crain PF, Hashizume T, Pomerantz SC, Cruz R, Steer CJ, Hill KE, Burk RF, McCloskey JA (1993) Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA([Ser]Sec). *J Biol Chem* 268(19): 14215-14223.

Donovan J, Caban K, Ranaweera R, Gonzalez-Flores JN, Copeland PR (2008) A novel protein domain induces high affinity selenocysteine insertion sequence binding and elongation factor recruitment. *J Biol Chem* 283, 35129-35139.

Downey CM, Horton CR, Carlson BA, Parsons TE, Hatfield DL, Hallqrimsson B, Jirik FR. (2009) Osteo-chondroprogenitor-specific deletion of the selenocysteine tRNA gene, Trsp, leads to chondronecrosis and abnormal skeletal development: a putative model for Kashin-Beck disease. *PLoS Genet* 5(8): e1000616.

Fagegaltier D, Lescure A, Walczak R, Carbon P, Krol A (2000a) Structural analysis of new local features in SECIS RNA hairpins. *Nucleic Acids Res* 28(14): 2679-2689.

Fagegaltier D, Hubert N, Yamada K, Mizutani T, Carbon P, Krol A (2000b) Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J* 19(17): 4796-4805.

Fomenko DE, Marino SN, Gladyshev VN (2008) Functional diversity of cysteine residues in proteins and unique features of catalytic redox-active cysteines in thiol oxidoreductases. *Mol Cells* 26: 228-235.

Graifer DM, Molotkov M, Styazhkina V, Demeshkina N, Bulygin K, Eremina a, Ivanov A, Laletina E, Ven'yaminova A, Karpova G (2004) Variable and conserved elements of human ribosomes surrounding the mRNA at the decoding and upstream sites. *Nucleic Acids Res.* 32: 3282-3293.

Ginisty H, Sicard H, Roger B, Bouvet P (1999) Structure and functions of nuclelin. *Journal of Cell Science* 112: 761-772.

Gladyshev VN, Krause M, Xu XM, Korotkov KV, Kryukov GV, Sun QA, Lee BJ, WWootton JC, Hatfield DL (1999) Selenocysteine-containing thioredoxin reductase in C. elegans. *Biochem Biophys Res Commun* 259:244–249.

Gromer S, Johansson L, Bauer H, David Ascott L, Rauch S, Ballou D, Williams C, Schrimer H, ESJ Arner (2003) Ative sites of thioredoxin reductases: Why selenoproteins? *Prot. Natl. Acad. Sci. U.S.A.* 100(22): 12618-12623.

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

Guimaraes MJ, Peterson D, Vicari A, Cocks BG, Copeland NG, Gilbert DJ, Jenkins NA, Ferrick DA, Kastelein RA, Bazan JF, Zlotnik A (1996) Identification of a novel SelD homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism? *Proc Natl Acad Sci U S A* 93(26): 15086-15091.

Halic M, Becker T, Franck J, Spahn CM, Beckman R (2005) Localization and dynamic behavior of ribosomal protein L30e. Nat Struct Mol Biol 12(5): 467-468.

Hatfield D, Diamond A, Dudock B (1982) Opal suppressor serine tRNAs from bovine liver form phosphoseryl-tRNA. *Proc Natl Acad Sci U S A* 79(20): 6215-6219.

Hatfield D and Diamond A (1993) UGA: a split personality in the universal genetic code. *Trends Genet* 9(3): 69-70.

Hill, K. E., Lyons P. R., Burk R. F. (1992) Differential regulation of rat liver selenoprotein mRNAs in selenium deficiency. *Biochem. Biophys. Res. Commun.* 185:260–263.

Hoffmann PR, Hoge SC, Li PA, Hoffmann FW, Hashimoto AC, Berry MJ (2007) The selenoproteome exhibits widely varying, tissue-specific dependence on selenoprotein P for selenium supply. *Nucleic Acids Res* 35(12): 3963-3973.

Hondal JR and Ruggles RE (2010) Differing views of the role of selenium in thioredoxin reductase. *Amino Acids* DOI10.1007/s00726-010-0494-6.

Howard M.T. et al (2007) A recording element that stimulates decoding of UGA codons by Sec rRNA$^{[Ser]Sec}$, *RNA* 13: 912-920.

Howard M.T. et al, (2005) Recodnig elements located adjacent to a subset of eukaryotic selenocysteine—specifying UGA codons. *EMBO J.* 24: 1596-1607.

Hubert N, Sturchler C, Westhof E, Carbon P, Krol A (1998) The 9/4 secondary structure of eukaryotic selenocysteine tRNA: more pieces of evidence. *RNA* 4(9)**:** 1029-1033.

Itoh Y, Chiba S, Sekine SI, Yokoyama S (2009) Crystal structure of human selenocysteine tRNA. *Nucleic Acids Res* 37(18): 6259-6268.

Johannsson L, Gafvelin G, Arnér ESJ (2005) Selenocysteine in proteins – properties and biotechnological use. *Biochem Biophys Acta* 1726: 1-13.

Kanzok SM, Fechner A, Bauer H, Ulschmid JK, Muller HM, Botella-Munot J, Schneuwly S, Schrimer R, Becker K (2001) Substitution of the thioredoxin system for glutathione reductase in *Drosophila Melanogaster*. *Science*, 291: 643-646.

Kinzy SA, Caban K, Copeland PR (2005) Characterization of the SECIS binding protein 2 complex required for the co-translational insertion of selenocysteine in mammals. *Nucleic Acids Res* 33(16) 5172-5180.

Klein DJ, Schmeing TM, Moore PB, Steitz TA (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J* 20(15): 4214-4221.

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(11)**:** 2166-2167.

Korotkov KV, Novoselov SV, Hatfield DL, Gladyshev VN (2002) Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element. *Mol Cell Biol* 22(5)**:** 1402-1411.

Kozak M (1989) The scanning model for translation: an update. *J. Cell Biol* 18: 3078-3093.

Kryukov GV, Kryukov VM, Gladyshev VN (1999) New mammalian selenocysteine-containing proteins identified with an algorithm that searches for selenocysteine insertion sequence elements. *J Biol Chem* 274(48): 33888-33897.

Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigo R, Gladyshev V (2003) Characterization of mammalian selenoproteomes. *Science* 300: 1439-1443.

Lacourciere GM (1999) Biosynthesis of selenophosphate. *Biofactors* 10(2-3): 237-244.

Laletina E, Graifer D, Malygin A, Ivanov A, Shatsky I, Karpova G (2006) Proteins surrounding hairpin IIIe of the hepatitis C virus internal ribosome entry site on the human 40S ribosomal subunit. *Nucleic Acids Res* 34(7): 2027-2036.

Latreche L, Jean-Jean O, Driscoll DM, Chavatte L (2009) Novel structural determinants in human SECIS elements modulate the translational recoding of UGA as selenocysteine. *Nucleic Acids Res* 37(17): 5868-5880.

Lee BJ, Worland PJ, Davis JN, Stadtman TC, Hatfield DL (1989) Identification of a selenocysteyl-tRNA(Ser) in mammalian cells that recognizes the nonsense codon, UGA. *J Biol Chem* 264(17): 9724-9727.

Lee BC, Dikiy A, Kim HY, Gladyshev VN (2009) Functions and evolution of selenoprotein methionine sulfoxide reductases. *Biochem Biophys Acta*, 1790: 1471–1477.

Lee BC, Lobanov AV, Marino SM, Kaya A, Seravalli J, Hatfield DL, Gladyshev VN (2011) A four selenocysteine, two SECIS element methionine sulfoxide reductase from *Metridium senile* reveals a non-catalytic function of selenocysteines. *J Biol Chem* 286(21):18747-55.

Lei, X. G., J. K. Evenson, K. M. Thompson, and R. A. Sunde. 1995. Glutathione peroxidase and phospholipid hydroperoxide glutathione peroxidase are differentially regulated in rats by dietary selenium. *J. Nutr*. 125: 1438–1446.

Leinfelder W, Zehenlein E, Mandrand-Berthelot MA, Bock A (1988) Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* 331: 723-725.

Lescure A, Allmang C, Yamada K, Carbon P, Krol A (2002) cDNA cloning, expression pattern and RNA binding analysis of human selenocysteine insertion sequence (SECIS) binding protein 2. *Gene* 291(1-2): 279-285.

Lescure A, Gautheret D, Carbon P, Krol A (1999) Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif. *J Biol Chem* 274(53): 38147-38154.

Lescure A, Rederstorff M, Krol A, Guicheney P, Allamand V (2009) Selenoprotein function and muscle disease. *Biochem Biophys Acta - General Subjects* 90(11): 1569-74.

Li Q, Imataka H, Morino S, Rogers GW, Jr., Richter-Cook NJ, Merrick WC, Sonenberg N (1999) Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol Cell Biol* 19(11): 7336-7346.

Lobanov AV, Delgado C, Rahlfs S, Novoselov SV, Kryukov GV, Gromer S, Hatfield DL, Becker K, Gladyshev VN (2006a) The Plasmodium selenoproteome. *Nucleic Acids Res* 34(2): 496-505.

Lobanov AV, Gromer S, Salinas G, Gladyshev VN (2006b) Selenium metabolism in Trypanosoma: characterization of selenoproteomes and identification of a Kinetoplastida-specific selenoprotein. *Nucleic Acids Res* 34(14): 4012-4024.

Lobanov AV, Fomenko DE, Zhang Y, Sengupta A, Hatfield DL, Gladyshev VN (2007) Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome Biol* 8(9): R198.

Lobanov AV, Hatfield DL, Gladyshev VN (2008a) Reduced reliance on the trace element selenium during evolution of mammals. *Genome Biol* 9(3): R62.

Lobanov AV, Hatfield DL, Gladyshev VN (2008b) Selenoproteinless animals: selenophosphate synthetase SPS1 functions in a pathway unrelated to selenocysteine biosynthesis. *Protein Sci* 17(1): 176-182.

Lobanov AV, Hatfield DL, Gladyshev VN (2009) Eukaryotic selenoproteins and selenoproteomes. *Biochem Biophys Acta*, 1790: 1424-1428.

MacÌas S, Bragulat M, Tardiff DF, Vilardell J (2008) L30 Binds the Nascent RPL30 Transcript to Repress U2 snRNP Recruitment. *Mol Cell* 30(6): 732-742.

Maiti B, Arbogast S, Allamand V, Moyle MW, Anderson CB, Richard P, Guicheney P, Ferreiro A, Flanigan KM, Howard MT (2008) A mutation in the SEPN1selenocysteine redefinition element (SRE) reduces selenocysteine incorporation and leads to SEPN1 related myopathy. *Human Mutation,* doi: 10.1002/humu.20879.

Malygin AA, Graifer DM, Laletine ES, Shatskii IN, Karpova GG (2003) Approach to identifying the functionally important segments of RNA, based on complementation-addressed modification. *Mol Biol (Mosc)* 37(6):1027-34.

Merrick WC and Nyborg J (2000) The protein biosynthesis elongation cycle, *Translational control of gene expression*, 110-115.

Miniard AC, Middleton LM, Budiman ME, Gerber CA, Driscoll DM (2010) Nucleolin binds to a subset of selenoproteins mRNAs and regulates their expression. *Nucleic Acids Res*. 38(14): 4807-4820.

Mix H, Lobanov AV, Gladyshev VN (2007) SECIS elements in the coding regions of selenoprotein transcripts are functional in higher eukaryotes. *Nucleic Acids Res* 35(2): 414-423.

Moghadaszadeh B, Petit N, Jaillard C, Brockington M, Roy SQ, Merlini L, Romero N, Estournet B, Desguerre I, Chaigne D, Muntoni F, Topaloglu H, Guicheney P (2001) Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome. *Nat Genet* 29(1): 17-18.

Mongelard F, Bouvet P (2007) Nucleolin: a multiFACeTed protein. *Trends Cell Biol* 17(2): 80-86.

Moore T, Zhang Y, Fenley MO, 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(5): 807-818.

Moriarty PM, Reddy CC, Maquat LE (1998) Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA. *Mol Cell Biol* 18(5): 2932-2939.

Namy O, Rousset JJP, Napthine S, Brierley I (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol Cell* 13: 157-168.

Nygård O, Alkemar G, Larsson SL (2006) Analysis of the secondary structure of expansion segment 39 in ribosomes from fungi, plants and mammals. *J Mol Biol* 357(3):904-916.

Olieric V, Wolff P, Takeuchi A, Bec G, Birck C, Vitorino M, Kieffer B, Beniaminov A, Cavigiolio G, Theil E, Allmang C, Krol A, Dumas P (2009) SECIS-binding protein 2, a key player in selenoproteins synthesis, is an intrinsically disordered protein. *Biochimie* 91(8): 1003-1009.

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. *Microbiol Rev* 56: 229-264.

Palioura S, Sherrer RL, Steitz TA, Soll D, Simonovic M (2009) The human SepSecS-tRNASec complex reveals the mechanism of selenocysteine formation. *Science* 325(5938): 321-325.

Papp LV, Lu J, Striebel F, Kennedy D, Holmgren A, Khanna KK (2006) The redox state of SECIS binding protein 2 controls its localization and selenocysteine incorporation function. *Mol Cell Biol* 26(13), 4895-4910.

Pedersen J.S.et al (2006) Identification and classification of conserved RNA secondary structures in the human genome, *PLoS Comput Biol* Apr 2(4): e33.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13): 1605-1612.

Pisarev AV, Hellen CU, Pestova TV (2007) Recycling of eukaryotic posttermination ribosomal complexes. *Cell* 131:286-299.

Pisarev AV, Kolupaeva VG, Yusupov MM, Hellen CUT, Pestova TV (2008) Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* 27(11): 1609-21.

Pisarev AV, Skabkin MA, Pisareva VP, Skabkina OV, Rakotondrafara AM, Hentze MW, Hellen CU, Pestova TV (2010) The role of ABCE1 in eukaryotic posttermination riboosmal recycling. *Mol Cell* 37: 196-210.

Preer JR, Preer LB, Rudman BM, Barnett AJ (1985) Deviation from the universal code shown by the gene for surface protein 51A in Paramecium. *Nature* 314(6008): 188-190.

Rederstorff M, Krol A, Lescure A (2006) Understanding the importance of selenium and selenoproteins in muscle function. *Cell Mol Life Sci* 63(1): 52-59.

Rodnina M and Wintermeyer W (2009) Recent mechanistic insights into eukaryotic riboosmes. *Curr Op Cell Biol* 21, 435-443.

Rother M, Wilting R, Commans S, Bock A (2000) Identification and characterisation of the selenocysteine-specific translation factor SelB from the archaeon Methanococcus jannaschii. *J Mol Biol* 299(2)**:** 351-358.

Salas-Marco J and Bedwell DM (2004) GTP hydrolysis be eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol Cell Biol* 3: 601-609.

Schoenmakers et al, (2010) Mutations in the selenocysteine insertion sequence–binding protein 2 gene lead to a multisystem selenoproteins deficiency disorder in humans. Journal of clinical investigation 120: 4220-4235.

Shchedrina VA, Novoselov SV, Malinouski M, Gladishev VN (2007) identification and characterization of a novel selenoprotein family containing a diselenide bond in a redox motif. *Prot. Natl. Acad. Sci. U.S.A.* 104(35): 13919-13924.

Sherrer RL, Ho JM, Soll D (2008) Divergence of selenocysteine tRNA recognition by archaeal and eukaryotic O-phosphoseryl-tRNASec kinase. *Nucleic Acids Res* 36(6): 1871-1880.

Small-Howard A, Morozova N, Stoytcheva Z, Forry EP, Mansell JB, Harney JW, Carlson BA, Xu XM, Hatfield DL, Berry MJ (2006) Supramolecular complexes mediate selenocysteine incorporation in vivo. *Mol Cell Biol* 26(6): 2337-2346.

Sonenberg N and Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136(4): 731-745.

Stoytcheva Z, Tujebajeva RM, Harney JW, Berry MJ (2006) Efficient incorporation of multiple selenocysteines involves an inefficient decoding step serving as a potential translational checkpoint and ribosome bottleneck. *Mol Cell Biol* 26(24): 9177-9184.

Sturchler C, Westhof E, Carbon P, Krol A (1993) Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). *Nucleic Acids Res* 21(5): 1073-1079.

Sturchler C, Lescure A, Keith G, Carbon P, Krol A, Ibba M, Soll D (1994) Base modification pattern at the wobble position of Xenopus selenocysteine tRNA(Sec). *Nucleic Acids Res* 22(8): 1354-1358.

Sun X, Li X, Moriarty PM, Henics T, LaDuca JP, Maquat LE (2001) Nonsense-mediated decay of mRNA for the selenoprotein phospholipid hydroperoxide glutathione peroxidase is detectable in cultured cells but masked or inhibited in rat tissues. *Mol Biol Cell* 12(4): 1009-1017.

Takeuchi A, Schmitt D, Chapple C, Babaylova E, Karpova G, Guigo R, Krol A, Allmang C (2009) A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins. *Nucleic Acids Res* 37(7): 2126-2141.

Taskov K, Chapple C, Kryukov GV, Castellano S, Lobanov AV, Korotkov KV, Guigo R, Gladyshev VN (2005) Nematode selenoproteome: the use of the selenocysteine insertion system to decode one codon in an animal genome? *Nucleic Acids Res* 33(7): 2227-2238.

Traut RR, Bollen A, Sun TT, Hershey JWB, Sundbrg J and Pierce LR (1973) Methyl 4-mercaptobutyrimidate as a cleavable cross-linking reagent and its application to the Escherichia coli 30S ribosome. *Biochemistry* 12(17): 3266-3273

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(2): 158-163.

Turanov AA, Lobanov AV, Fomenko DE, Morrison HG, Sogin ML, Klobutcher LA, Hatfield DL, Gladyshev VN (2009) Genetic code supports targeted insertion of two amino acids by one codon. Science 323(5911): 259-261.

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

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(4): 367-379.

Walczak R, Carbon P, Krol A (1998) An essential non-Watson-Crick base pair motif in 3'UTR to mediate selenoprotein translation. *RNA* 4(1): 74-84.

Weiss SL, Sunde RA (1998) Cis-acting elements are required for selenium regulation of glutathione peroxidase-1 mRNA levels. *RNA* 4(7): 816-827.

Will S, Reiche K, Hofacker I, Stadlet P, Backofen R (2007) Inferring non-coding RNA families and classes by means of genome-scale structure-based clustering. *PLOS Comput Biol*, 3(4): e65.

Wu XQ, Gross HJ (1993) The long extra arms of human tRNA((Ser)Sec) and tRNA(Ser) function as major identify elements for serylation in an orientation-dependent, but not sequence-specific manner. *Nucleic Acids Res* 21(24): 5589-5594.

Wu XQ, Gross HJ (1994) 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(1): 241-248.

Xu X-M, Carlson BA, Mix H, Zhang Y, Saira K, Glass RS, Berry MJ, Gladyshev VN, Hatfield DL (2007) Biosynthesis of Selenocysteine on Its tRNA in Eukaryotes. *PLoS Biology* 5(1): e4.

Zavacki AM, Mansell JB, Chung M, Klimovitsky B, Harney JW, Berry MJ (2003) Coupled tRNA(Sec)-dependent assembly of the selenocysteine decoding apparatus. *Mol Cell* 11(3): 773-781.

Zenkova M, Ehresmann C, Caillet J, springer M, Karpova G, Ehresmann B, Romby P (1995) A novel approach to introduce site-directed specific crosslinks within RNA-protein complexes. Application to the E.Coli threonyl-tRNA synthetase/translational operator complex. *Eur J Biochem* 231(3): 726-735.

Zhang Y and Gladyshev VN (2007). High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis. *Nucleic Acids Res* 35 (15): 4952–4963.

Zhang Y, Baranov PV, Atkins JF, Gladyshev VN (2005) Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J Biol Chem* 280: 20740-20751.

Zhang Y and Gladyshev VN (2010) General trends in trace element utilization revealed by comparative genomic analyses of Co, Cu, Mo, Ni and Se. *J Biol Chem.* 285(5): 3393-3405.

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406-3415.

Insights into the selenocysteine incorporation mechanism in mammals

The biological form of the trace element selenium is the amino acid selenocysteine. It is encoded by a UGA triplet (Sec codon) which acts generally as a stop codon. Thus, a specialized machinery is used to incorporate this amino acid into selenoproteins during translation. This process involves a specific stem-loop located in the 3'-UTR of selenoprotein mRNAs, termed SelenoCysteine Insertion Sequence (SECIS), the specialized elongation factor EFSec and other protein factors. One of these is the SECIS Binding Protein 2 (SBP2), which is necessary for ribosomal recognition of the UGA triplet as the Sec codon.

To get new insight into the mechanism and, in particular, to find new interaction partners of the SECIS element in the course of selenoprotein mRNA translation, I have designed a set of mRNAs consisting of a short 5'-UTR (short natural 5' UTR or A-rich 5' UTR) followed by the sequence coding for the tetrapeptide Met-Sec-Phe-Phe (or Met-Phe-Phe-Phe for the control mRNA), a spacer sequence and the SECIS element at the 3'-end. We found that mRNAs containing A-rich 5'-UTR possess a higher ribosome binding efficiency compared to mRNAs with short natural 5' UTR (35-40% compared to 10%).

To identify the contacts of the SECIS element with the components of the translation machinery, mRNAs bearing 4-thiouridines in the SECIS element were used. Crosslinking experiments, together with western blotting (WB) analysis, revealed that SBP2 is bound to the SECIS-element in the 48S pre-initiation complex and in the 80S pre-translocation complex, in rabbit reticulocyte lysate. When formation of the peptide bond is blocked, SECIS does not crosslink to SBP2 anymore but rather to a set of ribosomal proteins. Nevertheless, SBP2 is present in the complex as shown by WB. This allowed us to propose the following interpretation. During the transpeptidation step of elongation, SBP2 is bound to the ribosome; however, after transpeptidation is accomplished, SBP2 leaves the ribosome (maybe due to a conformational change on the ribosome) not necessarily with the help of any factor and, because of its high affinity to SECIS, binds to SECIS in the pre-translocation step.

In an earlier collaboration between the two laboratories, SBP2 was shown to bind specifically to purified human 60S subunits but not to 40S subunits. Moreover, the SBP2 region necessary for this interaction was identified. However, the binding site of SBP2 on the 60S subunit was unknown. The SBP2•40S, SBP2•60S and SBP2•80S complexes were studied using crosslinking reagents. I have performed and crosslinks generated by bifunctional reagents (diepoxybutane, crosslink range 4Å) and 2-iminothiolane (crosslink range 7Å for rRNA-proteins crosslink, or 14Å for protein-protein crosslink) to investigate whether rRNA or ribosomal proteins form the molecular environment of SBP2 on the ribosome. SBP2 does not crosslink to 40S subunit in either the 40S-SBP2 or 80S-SBP2 complexes, correlating well with the binding data. Interestingly, SBP2 crosslinks to the 60S subunit in either the free state or in the 80S ribosome. Further, the yield of rRNA-SBP2 crosslink was more than 10 times higher than ribosomal proteins-SBP2 crosslinks, indicating that the 28S rRNA contributes more to the crosslink than ribosomal proteins. This allowed us to use hydroxyl radical probing to study the molecular environment of SBP2 on the ribosome. According to the probing data, the binding of SBP2 to the human 60S subunit protects 2 helixes in expansion segment 7. Unfortunately, these helices have not been fully modeled in the cryo-EM structures of eukaryotic ribosomes. Nevertheless, one can conclude that they are located on the solvent side of the 60S subunit near L7/L12 stalk.

Altogether, these studies have already provided novel important insights into the UGA selenocysteine reprogramming mechanism, giving also new detailed information on the molecular principle of the SBP2-SECIS and SBP2-ribosome interaction.

Keywords: selenocysteine, SBP2, SBP2-ribosome interactions, SECIS element.