RNA-protein interaction in the selenoprotein synthesis machinery

Akiko Takeuchi

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

Akiko TAKEUCHI

RNA-protein interaction in the selenoprotein synthesis machinery

Soutenue publiquement le 1 juillet 2009

Membres du jury

Rapporteur externe:	Mme Christiane BRANLANT, Directeur de Recherche du CNRS, Nancy
Rapporteur externe:	M Jean-Pierre ROUSSET, Professeur à l'Université Paris-Sud, Orsay
Rapporteur interne:	M Mario KELLER, Professeur à l'Université de Strasbourg
Examinateur:	M Rémy BORDONNE, Directeur de Recherche du CNRS, Montpellier
Directeur de thèse:	M Alain KROL, Directeur de Recherche du CNRS, Strasbourg
Directeur de thèse:	Mme Christine ALLMANG-CURA, Chargée de Recherche du CNRS, Strasbourg

Acknowledgements

I wish to express my appreciation to Dr. Christiane Branlant (ARN, RNP, Structure-Fonction-Maturation Enzymologie Moléculaire et Structurale, Nancy), Pr. Jean-Pierre Rousset (Institut de Génétique et Microbiologie, Orsay) and Pr. Mario Keller (Institit de Biologie Moléculaire des Plantes, Strasbourg) for having accepted to evaluate my PhD studies. I am also grateful to Dr. Rémy Bordonné (Institut de Génétique Moléculaire de Montpellier, Montpellier) for having examined my studies.

I would like to express profound gratitude to my advisor, Dr. Alain Krol for having accepted me as a laboratory member and for his supervision.

I am also highly thankful to Christine Allmang-Cura for her supervision and many advices on science and the daily life in France.

I would also like to thank everyone in the laboratory for their kindness. I will never forget my three years in Strasbourg.

Table of contents

TABLE OF CONTENTS **1**

TABLE OF FIGURES **3**

LIST OF ABBREVIATIONS **5**

RESUME DE LA THESE EN FRANÇAIS **7**

PART 1. INTRODUCTION **13**

1. SELENIUM AND ITS BIOLOGICAL FUNCTION 15
1.1. Selenium *15*
1.2. Selenocysteine *15*
1.3. Selenoproteins *16*
2. SELENOPROTEIN SYNTHESIS 20
2.1. Selenocysteine biosynthesis *20*
2.1.1. $tRNA^{Sec}$ 20
2.1.2. From serine to phosphoserine (O-phosphoseryl-$tRNA^{Sec}$ kinase / PSTK) 22
2.1.3. From phosphoserine to selenocysteine 24
2.1.3.1. Generation of the selenium donor (SPS1/2) 24
2.1.3.2. From Sep-$tRNA^{Sec}$ to Sec-$tRNA^{Sec}$ (SecS) 24
2.1.4. SECp43 25
2.2. Sec incorporation *25*
2.2.1. Cis-acting factors 26
2.2.1.1. SElenoCysteine Incorporation Sequence (SECIS) 26
2.2.1.1.a. Location in mRNA 26
2.2.1.1.b Secondary structure 27
2.2.1.2. SRE 31
2.2.2. Trans-acting factors 32
2.2.2.1. EFSec 32
2.2.2.2. SBP2 33
2.2.2.2.a Domain structure of SBP2 34
2.2.2.2.b. SECIS binding 35
2.2.2.2.c. EFSec-SBP2 interaction 38
2.2.2.2.d. Ribosomal binding 38
2.2.2.2.e. Expression and localization 39
2.2.2.3. L30 41
2.2.2.4. Other proteins 41
2.3. Sec incorporation model *42*
3. SELENOPROTEIN MRNP ASSEMBLY 43
3.1. Nuclear assembly *43*
3.2. Assembly of selenoprotein mRNAs - similarities with sn/snoRNP assembly *44*
4. OBJECTIVES AND OUTLINE OF THIS THESIS 47

PART 2. RESULTS **49**

1. FUNCTIONAL CHARACTERIZATION OF *DROSOPHILA MELANOGASTER* SBP2 51
1.1. Selenoproteome in Drosophila *51*
1.2. Objective *53*
1.3. Summary of Article 1 ..... *56*
1.4. Article 1 ..... *58*
1.5. Additional results and discussion *75*
SBP2 and the selenoprotein synthesis machinery in Drosophila willistoni 75
2. TOWARD CRYSTALLIZATION OF THE SBP2/SECIS COMPLEX 81
2.1. Objective *81*
2.2. Results *82*
2.2.1. cDNA cloning of SBP2 from various organisms 82
2.2.2. Expression of various SBP2 cDNAs using the Baculovirus expression system 83
2.2.3. Biophysical analysis of SBP2 87
2.3. Article 2 (in press) *89*
2.4. SBP2 is an Intrinsically Disordered Protein *99*
3. TOWARD IDENTIFICATION OF SBP2 PARTNERS 101

3.1. Objective ... *101*
3.2. Results ... *103*
3.3. Discussion ... *105*

PART 3. GENERAL CONCLUSION ... **109**

PART 4. ANNEX/ METHODS ... **117**

1. cDNA CLONING USING THE GATEWAY TECHNOLOGY ... 117
2. BACULOVIRUS EXPRESSION SYSTEM ... 119
2.1. Bacmid preparation ... *121*
2.2. Mini expression test ... *121*
2.3. Titration of viral particles and insect cell culture ... *121*

REFERENCES ... **125**

Table of figures

Table 1. Selenoproteins identified in eukaryotes and their functions. 19
Table 2. Selenoproteins identified in representative eukaryotic organisms. 53
Table 3. List of primers used for PCR amplification. 119

Figure 1. Chemical structures of cysteine and selenocysteine 16
Figure 2. Secondary structure models of canonical tRNAs and $tRNAs^{Sec}$ 22
Figure 3. The selenocysteine biosynthesis pathway 23
Figure 4. Secondary structure models of form 1 and 2 SECIS. 28
Figure 5. The secondary structure of SECIS RNA and various K-turn RNAs. 29
Figure 6. The SECIS and SRE elements of SEPN1 mRNAs 30
Figure 7. Schematic representations of the selenocysteine specialized translation elongation factors compared to general elongation factors 32
Figure 8. Schematic representation of protein factors involved in selenoprotein synthesis 35
Figure 9. RNA-protein interfaces at various L7Ae protein-K turn RNA complexes. 36
Figure 10. SECIS RNA determinants for SBP2 binding 37
Figure 11. Proposed model for the regulation of SBP2 subcellular localization and function after oxidative stress. 40
Figure 12. Selenocysteine incorporation models 42
Figure 13. Nuclear assembly of the selenoprotein synthesis machinery 44
Figure 14. Selenoprotein mRNP assembly model. 45
Figure 15. Eukaryotic selenoproteomes 52
Figure 16. Selenoproteins in *D. melanogaster* SBP2. 54
Figure 17. Shematic representation of the human and *D. melanogaster* SBP2 proteins. 55
Figure 18. One amino acid insertion in *D. willistoni SBP2*. 76
Figure 19. Aspargine insertion into the Glu 679-Glu 699 conserved spacing in human SBP2. 77
Figure 20. Disrupting the conserved Glu 679-Glu 699 spacing in hSBP2 by one amino acid insertion had no significant effect on SECIS-binding. 78
Figure 21. Schematic drawings of SBP2 proteins designed and generated for crystallization purposes. 83
Figure 22. Mini expression assays in baculovirus infected insect cells. 85
Figure 23. Expression tests from large-scale culture of baculovirus infected insect cells. 86
Figure 24. Examples of X-ray structures of IDPs bound to their targets. 100
Figure 25. Composition and organization of C/D sno(s)RNPs and U4 snRNP 102
Figure 26. GST pull-down experiments. 104
Figure 27. Proposed model for SBP2 functions during the selenoprotein mRNP formation. 111
Figure 28. Principle of the GATEWAY cloning 118
Figure 29. PCR amplification strategy to generate the SBP2 cDNA fragment harboring the AttB recombination sites 119
Figure 30. Generation of recombinant baculoviruses and gene expression. 120

List of abbreviations

A, C, G, T, U	adenine, cytosine, guanine, thymine, uracil
ATP	adenosine 5'-triphosphate
bp	base pair
CD	circular dichroism
C. elegans	Caenorhabditis elegans
C-terminal	carboxy-terminal
DIO	iodothyronine deiodinase
DLS	dynamic light scattering
D. melanogaster	Drosophila melanogaster
DNA	deoxyribonucleic acid
cDNA	complementary deoxyribonucleic acid
E. coli	Escherichia coli
EFSec	selenocysteine-specific elongation factor
EST	expressed sequence tag
GFP	green fluorescent protein
GPx	glutathione peroxidase
GST	glutathione S-transferase
His	hexahistidine tag
Hsp	heat shock protein
IDP	intrinsically disordered protein
kD	kilo dalton
Msr	methionine sulfoxide reductase
NES	nuclear export signal
NLS	nuclear localization signal
NMD	nonsense-mediated decay
NMR	nuclear magnetic resonance
nt	nucleotide
N-terninal	amino-terminal
ORF	open reading frame
PCR	polymerase chain reaction
PSTK	phosphoseryl-tRNA kinase
RNA	ribonucleic acid

mRNA	messenger ribonucleic acid
sRNA	small ribonucleic acid
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 nuclear ribonucleoprotein
SAXS	small-angle-X-ray scattering
SBP2	SECIS binding protein 2
Sec	selenocysteine
SECIS	selenocysteine insertion sequence
SecS	selenocysteine synthase
Sel	selenoprotein
SELEX	systematic evolution of ligands by exponential enrichment
SEPN1	selenoprotein N gene
Ser	serine
SerRS	seryl-tRNA synthetase
SPS	selenophosphate synthetase
SRE	selenocysteine codon redefinition element
TR	thioredoxin reductase
UTR	untranslated region
ψ	pseudouridine

Résumé de la thèse en français

Interactions ARN-protéines dans le mécanisme de biosynthèse des sélénoprotéines

Le sélénium est un oligo-élément essentiel. Sa forme biologique majeure est l'acide aminé sélénocystéine (Sec) que l'on retrouve essentiellement dans le site actif des sélénoprotéines. La sélénocystéine est incorporée dans les sélénoprotéines de façon co-traductionnelle en réponse à un codon UGA habituellement reconnu comme l'un des trois codons de terminaison. Chez les eucaryotes, la biosynthèse et l'incorporation de sélénocystéine requièrent la participation d'une machinerie moléculaire complexe qui implique, entre autres, une structure en tige-boucle située dans la région 3'UTR de l'ARNm des sélénoprotéines (élément SECIS), l'ARNt^{Sec} spécifique, le facteur d'élongation spécialisé EFSec ainsi que la protéine SBP2 (SECIS-binding protein). SBP2 joue un rôle majeur dans le mécanisme de synthèse des sélénoprotéines.

Chez les mammifères, le domaine de liaison à l'ARN de SBP2 est situé dans la région C-terminale de la protéine. Celui-ci comprend un module conservé, présent chez d'autres protéines de liaison à l'ARN mais possédant d'autres fonctions, appelé motif L7Ae dans les banques de données. La région N-terminale est dépourvue de toute similitude avec des protéines connues et n'est pas nécessaire à la synthèse des sélénoprotéines *in vitro*. De façon intéressante, une recherche bioinformatique dans les génomes de drosophile nous a permis d'identifier des séquences potentielles portant toutes les signatures d'une vraie protéine SBP2 mais plus courte et ne possédant pas de domaine N-terminal homologue à celui des mammifères. Avant que je n'entreprenne ce travail, SBP2 n'avait été caractérisée fonctionnellement que chez le rat et l'homme. Au cours de cette thèse, j'ai cloné l'ADNc et caractérisé fonctionnellement la protéine SBP2 de Drosophila melanogaster (dSBP2) à l'aide de tests de liaison à l'ARN et d'expression de sélénoprotéines dans des lysats de réticulocytes de lapin. Malgré sa taille plus courte, dSBP2 a montré la même capacité à promouvoir la synthèse de sélénoprotéines que son homologue mammifère. Il n'en va pas de même en ce qui concerne la liaison à l'ARN SECIS : en effet, alors que la protéine SBP2 humaine (hSBP2) est capable de lier deux formes distinctes d'ARN SECIS (appelées type 1 et 2) avec des

affinités similaires, dSBP2 ne présente d'affinité forte que pour le type 2, qui est d'ailleurs le seul présent chez la drosophile. Par ailleurs, nous avons identifié un domaine additionnel riche en lysines (K-rich domain), différent du module de liaison à l'ARN L7Ae, mais essentiel à la liaison à l'ARN SECIS. L'échange de seulement cinq acides aminés entre dSBP2 et hSBP2 au sein du domaine K-rich a permis d'inverser les propriétés de liaison à l'ARN SECIS des deux protéines, révélant ainsi l'existence d'un penta-peptide important pour la liaison aux SECIS de type 1.

Dans cette étude, nous avons également montré que la protéine SBP2 était capable d'interagir avec la sous-unité 60S du ribosome et que le domaine K-rich était essentiel pour cette interaction. Le fait que les mêmes acides aminés sont requis à la fois pour la liaison à l'ARN SECIS et au ribosome suggère que SBP2 est incapable de se lier simultanément à ces deux cibles et que des mécanismes d'échange dynamiques ont lieu au cours de la synthèse des sélénoprotéines.

Publication 1 :

A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins. Akiko Takeuchi, David Schmitt, Charles Chapple, Elena Babaylova, Galina, Karpova, Roderic Guigo, Alain Krol and Christine Allmang (2009). *Nucleic Acids Research*, *37(7):2126-41*.

Afin d'obtenir plus d'informations sur la nature des interactions SBP2-ARN SECIS au niveau atomique, nous avons entrepris l'analyse structurale de SBP2, en collaboration avec l'équipe de Philippe Dumas dans notre unité (UPR 9002 du CNRS). Des essais de cristallisation de SBP2 avaient été réalisés au laboratoire. Malheureusement, malgré de nombreux essais, aucun cristal n'a pu être obtenu, ni avec la protéine seule, ni en complexe avec l'ARN SECIS. Nous avons proposé que ceci soit vraisemblablement dû à la présence de l'extrémité N-terminale de la protéine qui ne semblait pas structurée. Nous avons donc utilisé des versions plus courtes de SBP2, dépourvues du domaine N-terminal. Dans ce but, j'ai construit des clones codant pour des protéines SBP2 de différents organismes, fusionnés à des étiquettes différentes et les ai exprimés dans des cellules d'insecte infectées par baculovirus afin d'améliorer le niveau d'expression des protéines. J'ai bénéficié pour cela de l'aide de la Plateforme de Génomique et Biologie Structurales au CEBGS et du service baculovirus de l'IGBMC à Illkirch.

Ceci ne nous a cependant pas permis d'obtenir de cristaux. En fait, l'analyse biophysique par différentes techniques (RMN 1D, centrifugation analytique, dichroïsme circulaire, diffusion dynamique de la lumière) a permis d'établir que SBP2 était globalement non-structurée, à

l'exception de son domaine L7Ae. Cette observation est cohérente avec nos résultats de prédictions informatiques de régions désordonnées qui indiquaient que 70% de la séquence de SBP2 était non-structurée, ainsi qu'avec les mêmes analyses biophysiques réalisées avec la protéine SBP2 produite dans *E. coli.* Ces résultats renforcent notre hypothèse selon laquelle SBP2 fait partie de la famille des protéines intrinsèquement désordonnées (Intrinsically Disordered Proteins ou IDP). Il est vraisemblable que SBP2 ne se structure qu'en présence de ses partenaires. Cette hypothèse est en accord avec des résultats récents de notre laboratoire montrant que le repliement et l'assemblage de la protéine SBP2 sur l'ARN SECIS étaient dépendants d'un complexe d'assemblage conservé lié au chaperon protéique Hsp90. La résolution de la structure du complexe SBP2-ARN SECIS ne semble donc envisageable que sous réserve de l'identification de partenaires de la protéine SBP2 capables d'induire son repliement stable.

Publication 2 :

Vincent Oliéric, Philippe Wolff, Akiko Takeuchi, Guillaume Bec, Catherine Birck, Marc Vitorino, Bruno Kieffer, Artemy Beniaminov, Giorgio Cavigiolio, Elizabeth Theil, Christine Allmang, Alain Krol and Philippe Dumas. SECIS-binding protein 2, a key player in selenoprotein synthesis, is an intrinsically disordered protein, *Biochimie (2009) 91 (8): 1003-1009*.

Le domaine de liaison à l'ARN de SBP2 contient le module L7Ae présent chez d'autres protéines de la même famille mais assurant des fonctions variées, telles que les protéines ribosomiques L7A et L30, la protéine 15.5kD/Snu13p de la snRNP U4 (épissage) et Nhp2p des snoRNP (biogenèse des ribosomes). Les protéines L7Ae se lient à des ARN de structure commune et leur fixation est requise pour l'assemblage des autres protéines core du complexe RNP auquel elles appartiennent. Notre laboratoire avait établi que l'assemblage correct des mRNP de sélénoprotéines, pré-requis à leur traduction, obéit aux mêmes règles que celui des sno/snRNP. Cet assemblage fait appel à un complexe supramoléculaire lié au chaperon protéique Hsp90, conservé de la levure à l'homme et d'importance fondamentale pour la cellule. Cette machinerie moléculaire est associée la protéine Nufip qui joue le rôle d'adaptateur. Nufip interagit avec toutes les protéines L7Ae, y compris SBP2, et est capable de promouvoir l'interaction avec les protéines core des sn/snoRNP en cours de synthèse. Nufip joue probablement le même rôle lors de l'assemblage des mRNP de sélénoprotéines. Les protéines core des mRNP de sélénoprotéines restent cependant largement inconnues. Un autre aspect de mon projet a consisté à déterminer si certaines protéines core majeures des

complexes sn/snoRNP pouvaient être des partenaires potentiels de SBP2. Cette hypothèse se confirme puisque mon travail a permis de montrer que SBP2 interagissait *in vitro* avec au moins l'une des protéines core des sn(o)RNP à boîte C/D, la protéine Nop58, et que cette interaction est directe. A notre grande surprise, ces résultats révèlent que l'assemblage de la catégorie particulière des ARNm de sélénoprotéines présente de nouvelles similitudes avec celui des sn- et snoRNP.

L'ensemble de ces résultats a permis de mieux comprendre comment se forme le complexe SBP2-ARN SECIS lors de la synthèse des sélénoprotéines, un processus au cœur du mécanisme de recodage du codon UGA.

Part 1. Introduction

1. Selenium and its biological function

1.1. Selenium

The non-metal element selenium was discovered by the Swedish chemist Jacob Berzelius in 1817. It was named after Sêlenê, the Greek goddess of the moon, in reference to the previously discovered and chemically related chalcogen element tellurium (tellus, earth in Latin). Selenium was considered a poison for a long time, especially to livestock eating selenium accumulator plants of the genus *Astragalus* during periods of drought in western USA and China. Later, selenium was defined as an essential micronutrient that exerts significant health benefits. In the 1970's, its biological activity could be attributed to the newly identified amino acid selenocysteine (Sec). In humans, selenium deficiency has been implicated as a factor for the emergence of the Keshan disease, an endemic cardiomyopathy in certain regions of eastern China, where dietary selenium is very low because the soil is deprived of this element. Selenium has also been implicated in the prevention of viral infections, cancer, infertility; it has been shown as an important factor for thyroid hormone maturation, the immune system as well as muscle development and function. However, molecular evidence is missing for most of these pathologies with the exception of infertility, thyroid maturation and muscle development. (See 1.3. Selenoproteins)

Selenium may also have a protective effect against inflammatory diseases (reviewed in Hatfield & Gladyshev, 2002; Hatfield et al, 2009; Lescure et al, 2009; Rederstorff et al, 2006). Selenium is mostly found at the catalytic site of most of the selenium-containing proteins which are called selenoproteins.

1.2. Selenocysteine

Selenocysteine is the major biological form of selenium in eukaryotes and is mostly found in the active site of selenoproteins. Selenocysteine is called the 21st amino acid. Its chemical structure differs from cysteine only by the presence of selenium in place of the sulfur atom (see Figure 1). Even though selenium and sulfur belong to the same family, selenocysteine exhibits distinct chemical properties versus cysteine. Selenocysteine has a lower pKa (5.2

versus 8.5 for cysteine) and is deprotonated under the physiological pH range. It thus exhibits a stronger nucleophilicity and reactivity than cysteine. Cysteine homologues of selenoenzymes are generally weaker catalysts, and Sec-to-Cys mutations result in a 100- to 1000-fold decrease in the catalytic activity (reviewed in Muttenthaler & Alewood, 2008). Selenocysteine is encoded by a UGA codon which is usually recognized as a translational stop signal, and is co-translationally incorporated into nascent peptide chains by a mechanism that will be described below (described in 2.2. Sec incorporation).

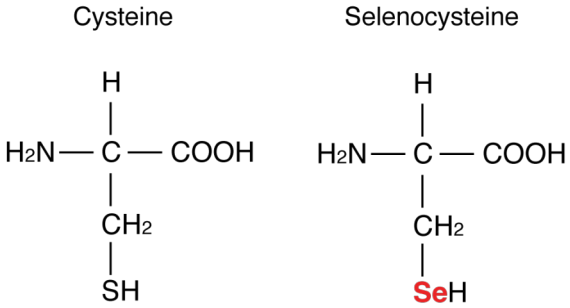


Figure 1. Chemical structures of cysteine and selenocysteine.
Selenocysteine differs form cysteine by a single atom: the selenium (Se, in red) instead of the sulfur (S) atom.

1.3. Selenoproteins

Selenoproteins have been found in the three domains of life but not in all species of bacteria, archaea and eukaryotes. For example, neither fungi nor higher plants possess selenoproteins. Vertebrates encode up to 25-26 selenoproteins, but larger selenoproteomes can be found in aquatic organisms (Lobanov et al, 2007). Selenoproteins are generally involved in catabolic pathways in bacteria and archaea, whereas eukaryotic selenoproteins participate rather in anabolic and antioxidant processes (Herbette et al, 2007). Based on the location of the Sec residue, mammalian selenoproteins can be classified into two groups (Kryukov et al, 2003). One group of selenoproteins contains Sec in the C-terminal region. This group includes thioredoxin reductases, selenoproteins S, R, O, I and K. The second group, that includes the remaining selenoproteins, contains the Sec residue in the N-terminal region. Some selenoproteins of the second group possess a CXXU motif (C and U stand for cysteine and selenocysteine, respectively, X for any amino acid) which is similar to the thioredoxin active-site CXXC motif (Dikiy et al, 2007; Ferguson et al, 2006; Lu & Holmgren, 2009; Novoselov

et al, 2007b). Such sequence signatures suggest that selenoproteins have redox-related functions. Indeed, some of the selenoproteins are involved in oxidation-reduction reactions to protect cells from oxidative damage; there is good reason to believe that the majority of the still functionally uncharacterized selenoproteins participate in such mechanisms as well. Selenoproteins with identified redox activity include five glutathione peroxidases (GPx), three thioredoxin reductases (TR), three iodothyronine deiodinases (DIO) and selenophosphate synthetase 2 (SPS2). Selenoproteins participate in thyroid hormone metabolism, muscle formation, selenocysteine synthesis and in sperm maturation (Rederstorff et al, 2006). Eukaryotic selenoproteins and their functions are summarized in Table 1.

Thioredoxin reductases regulate the thioredoxin system that participates in many cellular signaling pathways by controlling the activity of transcription factors. Therefore, thioredoxin reductases are involved in various cellular functions such as cell proliferation, antioxidant defense and redox-regulated signaling cascades (reviewed in Arner, 2009; Lu & Holmgren, 2009).

Glutathione peroxidase (GPx, Enzyme Commission number 1.11.1.9; now GPx1) was the first mammalian selenoprotein identified in 1973 (Flohe, 2009). There are seven isoenzymes identified in humans, and five of them are selenoproteins (GPx1, 2, 3, 4 and 6). GPxs reduce hydrogen peroxide and organic hydroperoxides, thus protecting cells from oxidative damage.
GPx1 is a cytosolic enzyme that is abundant in liver and erythrocytes. Its major function is the detoxification of hydroxyperoxides to protect cells from oxidative stress that could result in DNA damage. The GPx1 polymorphisms are also associated with cancer risk (reviewed in Flohe, 2009; Gromer et al, 2005; Zhuo & Diamond, 2009).
Glutathione peroxidase 4 (GPx4; Enzyme Commission number 1.11.1.12) is also known as phospholipid hydroperoxide GPx (PHGPx) because of its role in detoxification of lipid peroxides. GPx4 transforms into a structural component of the midpiece of mature spermatozoa by using hydroperoxides (Ursini et al, 1999). GPx4 is therefore involved in sperm maturation and male fertility (reviewed in Flohe, 2009; Lu & Holmgren, 2009).

Iodothyronine deiodinases (DIOs) cleave specific iodine carbon bonds in the thyroid hormones thyroxin (T4), bioactive 3,5,3'-tri-iodothyronine (T3) and 3'3'5' reverse tri-iodothyronine (rT4) which is less active than T3. Thereby DIOs regulate the hormonal activity of the thyroid. DIO 1 and 2 convert T4 to T3, and DOI 3 converts T4 to rT3. DIO 1

can also convert T4 to rT3 (Reviewed in Gromer et al, 2005; Lu & Holmgren, 2009; Pappas et al, 2008).

SPS2 is the selenophosphate synthetase which is involved in selenocysteine biosynthesis. This selenoprotein will be further described in 2.1.3.1.

Selenoprotein N (SelN) was the first selenoprotein shown to be involved in a genetic disorder (Moghadaszadeh et al, 2001). SelN was discovered in the laboratory using a computational screen based on the search of a conserved RNA structural motif that acts as a signature for selenoprotein mRNAs, the selenocysteine insertion sequence (SECIS) (Lescure et al, 1999). The pathology was known before SelN was identified. A large number of mutations in the coding region of the SelN gene (*SEPN1)* are associated with a wide range of early-onset muscular disorders now referred to as *SEPN1*-related myopathies. However, its catalytic function still remains unknown. SelN was characterized as a glycosylated transmembrane protein of the endoplasmic reticulum (ER). In addition to the transmembrane domain, SelN contains a predicted domain consisting in a calcium binding EF-hand motif which may contribute to the overall structure of the protein, and a SCUG catalytic site, reminiscent of a thioredoxin reductase motif, suggesting a reductase activity (reviewed in Lescure et al, 2009).

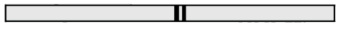
seleno-protein	Function	Schematic representation
DIO1	thyroid hormone maturation (Conversion of T4 to T3 and T4 to rT3)	
DIO2	thyroid hormone maturation (Conversion of T4 to T3)	
DIO3	thyroid hormone catabolism (Conversion of T4 to rT3)	
GPx1	Antioxidant protection	
GPx2	Antioxidant protection	
GPx3	Maintenance of cellular redox status	
GPx4	Detoxification of lipid hydroperoxides	
GPx6	Antioxidant protection	
TR1	Part of the thioredoxin system. Antioxidant defense, redox regulation, cell signaling	
TR2	Part of the thioredoxin system. Antioxidant defense, redox regulation, cell signaling	
TR3	Part of the thioredoxin system. Antioxidant defense, redox regulation, cell signaling	
SPS2	Selenocysteine synthesis	
15kDa	Role in cell apoptosis and mediation of chemopreventive effects of Se	
SelN	Associated with muscular diseases	
SelW	Antioxidant protection Associated with cardiac calcification	
SelP	Plasmatic selenium transport Antioxidant defense	
SelH	Unknown	
SelI	Unknown	
SelJ	Present in fish and sea urchin	
SelK	Unknown	
SelL	Unknown, present in diverse aquatic organisms	
SelM	Unknown	
SelO	Unknown	
SelS	Unknown, involved in cellular redox balance	
SelT	Role in regulation of Ca2+ homeostasis and neuroendocrine secretion	
SelU	Unknown Present in fish, birds, sea urchin, green algae and diatoms only	
SelV	Unknown	
MsrA	Reduction of oxidized methionine residues Selenocysteine in C. reinhardtii only	
MsrB	Reduction of oxidized methionine residues	

Table 1. Selenoproteins identified in eukaryotes and their functions.

Data taken from (Pappas et al, 2008; Rederstorff et al, 2006; Reeves & Hoffmann, 2009; Shchedrina et al. 2007).

The relative position of the selenocysteine residue is indicated by a black box. DIO: iodothyronine deiodinase, GPx: glutathione peroxidase, TR: thioredoxin reductase, SPS2: selenophosphate synthetase, Sel: selenoprotein, Msr: methionine sulfoxide reductase

2. Selenoprotein synthesis

Because selenocysteine is encoded by a UGA codon, one of the translational termination signals in the universal genetic code, discriminating UGA Sec from the stop codon requires a specialized translational machinery. This reprogramming mechanism is called UGA recoding. The bacterial selenoprotein synthesis mechanism has been extensively studied and well established by the Böck's group (reviewed in Böck, 2006). Recent important progresses have been made toward the elucidation of this mechanism in eukaryotes. This will be described here in more details (reviewed in Allmang et al, 2009; Papp et al, 2007; Squires & Berry, 2008). Selenoprotein synthesis comprises two steps, selenocysteine biosynthesis and its co-translational incorporation.

2.1. Selenocysteine biosynthesis

Selenocysteine does not occur as a free amino acid. Its biosynthesis occurs in two steps by conversion of serine to selenocysteine directly on the selenocysteine tRNA.

2.1.1. tRNA^{Sec}

tRNA^{Sec} is the selenocysteine specialized tRNA harboring anticodon complementary to UGA. Although tRNA^{Sec} species in bacteria differ in sequence from eukaryal and archaeal homologs, structure probing and computer modeling proposed similarities at the three-dimentional structures (Baron 1993; Sturchler 1993). They also show functional conservation since both eukaryotic and archaeal tRNA^{Sec} can function in bacterial systems (Baron et al, 1994; Lee et al, 1989; Rother et al, 2000).

Eukaryotic tRNA^{Sec} was initially discovered as a serine acceptor suppressing the UGA opal codon (Hatfield & Portugal, 1970). Later it was shown that this tRNA exists in the form of selenocysteyl-tRNA^{Sec} (Lee et al, 1989; Mizutani, 1989). Heterozygous knockout mice retain selenoprotein synthesis ability despite the reduced level of tRNA^{Sec}, implicating that it is not limiting for selenoprotein synthesis. Homozygous knockout mice are embryonic lethal demonstrating that selenoprotein synthesis is essential to mammalian development (Bosl et al, 1997).

tRNA^{Sec} has characteristic features in its secondary structure and a post-transcriptional modification pattern that distinguish it from canonical tRNAs (reviewed in Allmang & Krol,

2006b). $tRNA^{Sec}$ is the longest tRNA, with 95 nucleotides in E.coli and 90 nucleotides in eukaryotes (Amberg et al, 1993; Böck, 2006; Diamond et al, 1981; Diamond et al, 1993; Hatfield et al, 1982).

Secondary structure models for the eukaryotic $tRNA^{Sec}$ were proposed based on enzymatic and chemical probing and structure-based sequence alignments (Hubert et al, 1998; Sturchler et al, 1993). Compared to canonical tRNAs, $tRNA^{Sec}$ has a longer D-stem and an extended amino acid acceptor arm (consisting of the A and T-stems). The length of the D-stem is 6 bp, whereas it only has 3-4 bp in other tRNAs. While the amino acid acceptor arm of canonical tRNAs is 12 bp long, comprising a 7 bp A-stem and a 5 bp T-stem, that of $tRNA^{Sec}$ is 13 bp. Archaea and eukaryotes have a 9 bp A-stem and 4 bp T-stem, called 'the 9/4 model', and bacteria have an 8 bp A-stem and a 5 bp T-stem, called 'the 8/5 model' (Figure 2). In bacteria, the extra length of the acceptor arm is the determinant for binding to the specific elongation factor SelB. It is required for the serine to selenocysteine conversion in eukaryotes (Baron & Bock, 1991; Sturchler-Pierrat et al, 1995), which does not exclude the possibility that it also participates in recognition of the homologous factor in eukaryotes. The long variable arm and the discriminatory base G73 are the major identity elements for the serylation of $tRNA^{Sec}$ and $tRNA^{Ser}$ (Wu & Gross, 1993, Figure 2).

Post-transcriptional modification of the Xenopus $tRNA^{Sec}$ has been investigated (Diamond et al, 1993; Sturchler et al, 1994). Compared to canonical tRNAs which contain 15-17 modified bases, eukaryotic $tRNA^{Sec}$ contains only 4 post-transcriptionally modified nucleotides: pseudo-U55 (pseudouridine) and m^1A58 (1-methyladenosine) in the T-loop, i^6A37 (6-isopentenyladenosine) and mcm^5Um34 (5-methylcarboxymethyluridine-2'-O-methylribose) in the anticodon loop. There are two major isoforms of $tRNA^{Sec}$ differing by the methylation state of the ribose at U34, mcm^5U34 and mcm^5Um34. The relative amounts and distribution of these two isoforms vary in different cells and tissues. Efficient methylation of the U34 ribose to yield mcm^5Um34 requires the prior modification of other bases and an intact tertiary structure (Kim et al, 2000). Furthermore, methylation of the U34 ribose appears to be enhanced in the presence of selenium (Diamond et al, 1993). Transgenic mice, overexpressing a mutant $tRNA^{Sec}$ gene lacking i^6A (consequently also lacking Um34), display reduced expression of several stress-related selenoproteins such as GPx1, GPx3 SelR and SelT (Carlson et al, 2005). These results suggest that the isoforms may have different functions. In addition, the Um34 modification appears to have a greater influence than that of i^6A37 in regulating the expression of various mammalian selenoproteins.

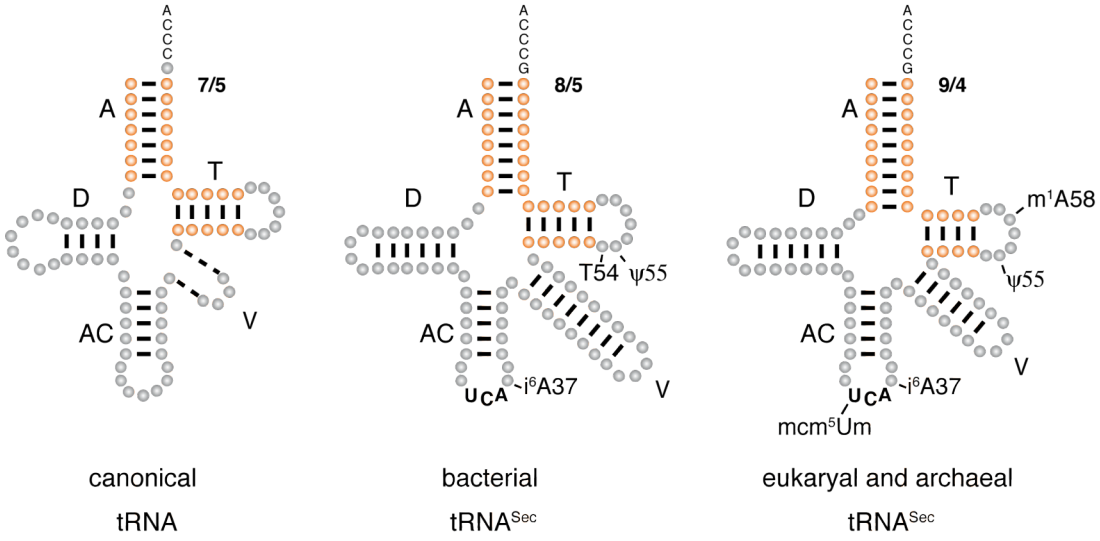


Figure 2. Secondary structure models of canonical tRNAs and tRNAs^{Sec}.

The acceptor arms are shown in orange: 7/5, 8/5 and 9/4 indicate the number of base pairs constituting the amino acid- and T-stems, respectively. i^{6}A37, T54, ψ55, m^{1}A58 and mcm^{5}Um are the base or ribose modifications in the bacterial and eukaryotic tRNAs^{Sec}. The secondary structure elements are indicated by abbreviations (A: the amino acid-stem, D: D-stem, AC: anticodon-stem V: variable arm, T: T-stem). The length of the variable arm in the canonical tRNA varies according to tRNAs^{Sec}. The figure is taken from (Allmang & Krol, 2006b).

2.1.2. From serine to phosphoserine (O-phosphoseryl-tRNA^{Sec} kinase / PSTK)

Since selenocysteine does not occur as a free amino acid, the biosynthesis of selenocysteine begins with the charge of serine on the tRNA^{Sec} by the conventional seryl-tRNA synthetase. The seryl to selenocysteine conversion occurs on the tRNA.

In bacteria, selenocysteine synthase (SelA), a pyridoxal phosphate enzyme, converts directly the seryl moiety to selenocysteine on the tRNA^{Sec} using monoselenophosphate as the substrate (reviewed in Böck, 2006). The monoselenophosphate selenium donor is produced from selenide by a reaction catalyzed by selenophosphate synthetase (SelD).

Unlike the bacterial selenocysteine biosynthesis where the Ser-tRNA^{Sec} is converted directly to Sec-tRNA^{Sec}, selenocysteine biosynthesis in eukaryotes and archaea occurs in two steps postcharging: phosphorylation of the Ser-tRNA^{Sec} by the O-phosphoseryl-tRNA^{Sec} kinase

(PSTK) and conversion of the phosphoseryl-tRNA^{Sec} (Sep-tRNA^{Sec}) to Sec-tRNA^{Sec} by Selenocysteine synthase.

The presence of a kinase activity to convert the Ser-tRNA^{Sec} to Sep-tRNA^{Sec} was reported in 1970 (Maenpaa & Bernfield, 1970), but the O-phosphoseryl-tRNA^{Sec} kinase (PSTK) enzyme was identified only recently by using a comparative genomics approach (Carlson et al, 2004). This enzyme phosphorylates the serine moiety of Ser-tRNA^{Sec} to yield Sep-tRNA^{Sec} by using ATP. In contrast to SerRS that recognizes both the tRNA^{Ser} and tRNA^{Sec}, PSTK discriminates Ser-tRNA^{Sec} from Ser-tRNA^{Ser}. In eukaryotes, the length and secondary structure of the D-stem of tRNA^{Sec} are the major determinants for phosphorylation (Wu & Gross, 1994), whereas the archaeal enzyme recognizes the acceptor stem of the tRNA^{Sec} (Sherrer et al, 2008). Interestingly, the archaeal PSTK can efficiently phosphorylate a chimeric Thr-tRNA^{Sec}, providing evidence that this enzyme does not recognize the amino acid (Figure 3).

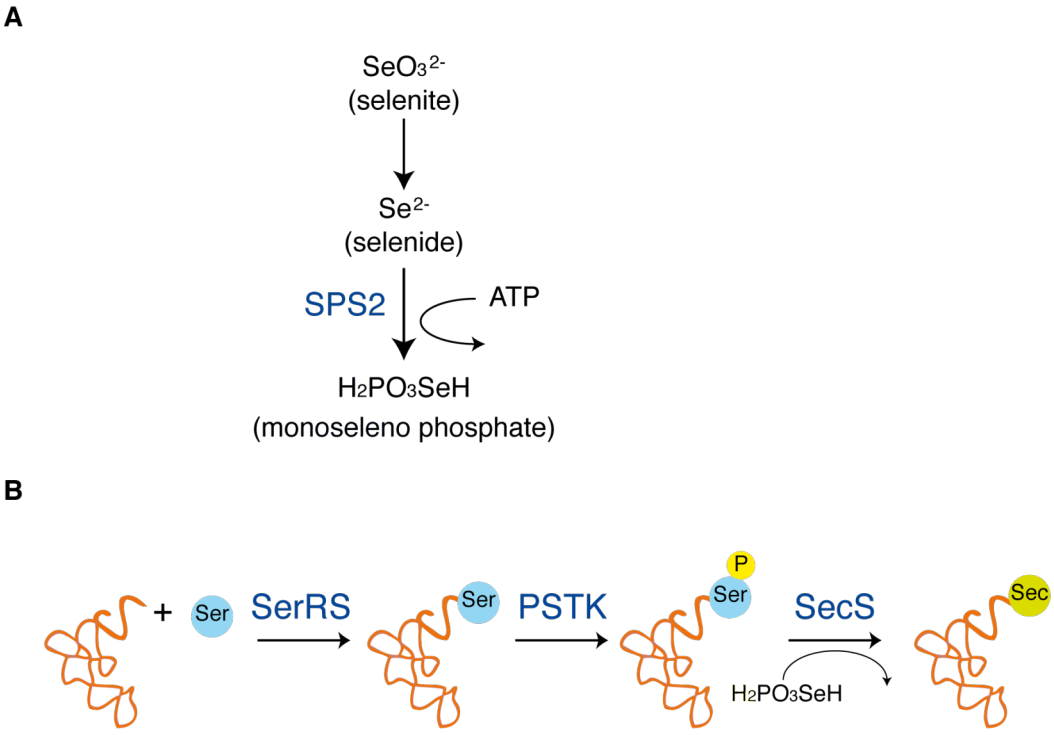


Figure 3. The selenocysteine biosynthesis pathway.

A. The selenium donor, monoselenophosphate (H_2PO_3SeH), is generated from selenite or more likely selenide by a reaction catalyzed by Selenophosphate synthetase 2 (SPS2). **B.** The tRNA^{Sec} is charged with serine by the conventional Seryl-tRNA synthetase (SerRS). In archaea and eukaryotes, the seryl residue is phosphorylated by the phosphoseryl-tRNA kinase (PSTK), and then converted to selenocysteine by Selenocysteine synthase (SecS) using monoselenophosphate.

2.1.3. From phosphoserine to selenocysteine

2.1.3.1. Generation of the selenium donor (SPS1/2)

Selenophosphate synthetase (SelD) in bacteria generates monoselenophosphate which is the selenium donor for selenocysteine biosynthesis. Selenophosphate synthetase 1 (SPS1) and later selenophosphate synthetase 2 (SPS2) were identified as the eukaryotic homologues (Guimaraes et al, 1996; Low et al, 1995). SPS2 is itself a selenoprotein in most organisms (Guimaraes et al, 1996). Recent studies demonstrated that SPS2 but not SPS1 can synthesize monoselenophosphate *in vitro*, and only SPS2 is essential for selenoprotein synthesis *in vivo* (Xu et al, 2007a; Xu et al, 2007b). In addition, SPS1 is present in insects that have lost selenoproteins, indicating that one of its major role is unrelated to selenoprotein synthesis (Chapple & Guigo, 2008).

2.1.3.2. From Sep-tRNA^{Sec} to Sec-tRNA^{Sec} (SecS)

Soluble Liver Antigen/Liver Pancreas (SLA/LP) was initially identified as a 48kDa protein co-immunoprecipitated with tRNA^{Sec} by autoantibodies from a subgroup of patients with a severe form of autoimmune chronic active hepatitis, and implicated in the selenocysteine pathway (Costa et al, 2000; Gelpi et al, 1992; Kernebeck et al, 2001). Later, two research teams identified independently SLA/LP as the eukaryotic and archaeal selenocysteine synthetase (Xu et al, 2007a; Yuan et al, 2006). The human and archaeal (*Methanococcus. maripaludis*) enzymes were named SepSecS (Yuan et al, 2006), whereas the mouse homolog was called mSecS (Xu et al, 2007a) according to the authors ('SecS' will be used in this thesis for reason of convenience.). Human and archaeal SecS were shown to complement *in vivo* an E.coli SelA null-strain and to convert the Sep-tRNA^{Sec} to Sec-tRNA^{Sec} in the presence of sodium selenite and recombinant *Escherichia.coli* SelD *in vitro* (Yuan et al, 2006). In addition, SecS exhibits higher affinity for the Sep-tRNA^{Sec} than for the tRNA^{Sec} and Ser-tRNA^{Sec} (Xu et al, 2007b). These studies provided evidence that, in contrast to bacterial SelA, eukaryotic and archaeal selenocysteine biosynthesis has an intermediate step where Sec-tRNA^{Sec} is generated, using Sep-tRNA^{Sec} and monoselenophosphate as substrates.

The crystal structures of the *M.maripaludis* and mouse SecS were solved and showed that both enzymes are members of the fold Type 1 pyridoxal phosphate (PLP)-dependent enzyme family, as is bacterial SelA (Araiso et al, 2008; Ganichkin et al, 2008).

2.1.4. SECp43

SECp43 was reported to interact with the $tRNA^{Sec}$ and to be involved in selenocysteine incorporation mechanism (Ding & Grabowski, 1999). SECp43 is predominantly present in the nucleus (Xu et al, 2005) and can interact with Sec-$tRNA^{Sec}$-EFSec complex *in vitro* (Small-Howard et al, 2006). SECp43 interacts with SecS and SPS1 *in vivo*, and redistributes these proteins to the nucleus (Small-Howard et al, 2006). Knockdown of SECp43 by siRNA demonstrated that SECp43 is required for ribose methylation at Um34 of $tRNA^{Sec}$, and increases selenoprotein expression at both mRNA and protein levels. A role for SECp43 has also been proposed in the orchestration of the interactions and localization of other selenoprotein synthesis factors (Small-Howard et al, 2006; Xu et al, 2005).

2.2. Sec incorporation

The general Sec incorporation mechanism is different in bacteria and eukaryotes. In bacteria, bSECIS (bacterial SElenoCysteine Insertion Sequence, a stem-loop structure immediately downstream of the in-frame UGA codon in selenoprotein mRNAs) and SelB, a translation elongation factor specialized for selenocysteine incorporation, play central roles for Sec incorporation. The N-terminal region of SelB is highly-sequence similar and functionally homologous to EF-Tu, the general translation elongation factor. Its C-teminal domain binds to bSECIS. SelB binds specifically and uniquely Sec-$tRNA^{Sec}$. The Sec-$tRNA^{Sec}$ harbored by SelB, is brought directly to the UGA Sec codon through the bSECIS-SelB interaction, allowing the incorporation of selenocysteine into the nascent polypeptide chain.

In eukaryotes, the SECIS element is located in the 3'UTR of selenoprotein mRNAs. Eukaryotic SECIS elements have conserved helix-loop structures and differ from the bSECIS structure. Sec incorporation requires the SECIS Binding Protein 2 (SBP2) and the specialized translation elongation factor EFSec (reviewed in Allmang & Krol, 2006b; Allmang et al,

2009). Ribosomal protein L30 has also been implicated in this mechanism (Chavatte et al, 2005).

2.2.1. Cis-acting factors

2.2.1.1. SElenoCysteine Incorporation Sequence (SECIS)

The SECIS is an RNA stem-loop structure that is mandatory for selenocysteine incorporation. Depending on the kingdom, it varies in sequence, structure and localization in the mRNA.

2.2.1.1.a. Location in mRNA

In bacteria, the SECIS RNA is located in the coding region immediately downstream of the in-frame UGA codon of selenoprotein mRNAs (reviewed in Böck, 2006). Unlike in bacteria, the SECIS is found in the 3'UTR of selenoprotein mRNAs in eukaryotes and archaea, suggesting similarities in the selenocysteine incorporation mechanism between archaea and eukaryotes. The advantage of having the SECIS element in the 3'UTR rather than in the coding region is that the RNA sequence is not constrained to maintain both the coding capacity and the base-pairing ability of the SECIS element. The localization of the SECIS element in the 3'UTR introduces flexibility by looping-out the intervening sequence between the UGA codon. It can thus interact with distant UGA Sec codons. In addition, its residence in the 3'UTR also enables selenoprotein mRNAs to harbor more than one UGA Sec codon. Indeed, the SECIS element in the 3'UTR relieves the necessity for stem-loop structures in the coding region, therefore providing complete flexibility in UGA codon position (Berry et al, 1993). Also, the SECIS element in the 3'UTR provides eukaryotes with a different Sec incorporation mechanism than in bacteria, for example, it enables rapid and efficient exchange of empty Sec-specific elongation factors (EFSec, see 2.2.2.1. EFSec) for Sec-tRNA^{Sec}/EFSec complexes, which seems to be essential in the case of multiple UGA codons (Tujebajeva et al, 2000). This is examplified for selenoprotein P (SelP). While most selenoprotein mRNAs contain a single UGA codon and a single SECIS element, SelP contains 10 to 18 UGA Sec codons, depending on animals, and 2 SECIS elements. In addition to the full-length protein, rat SelP has three isoforms resulting from termination at the second, third and seventh UGAs (Ma et al, 2002). However, it is possible that the isoforms of various

length could arise from experimental conditions and not from abortive synthesis. It was shown that the first UGA Sec is decoded by the second SECIS, and the first SECIS is required for decoding the downstream UGA Sec codons (Stoytcheva et al, 2006). Another surprising exception was found in the Fowlpox virus. The Fowlpox virus GPx4 mRNA contains a SECIS element at the 3'end of the coding region and not in the 3'UTR. Surprisingly also, this in-frame SECIS is able to support selenoprotein synthesis when the virus GPx4 is expressed in mammalian cells (Mix et al, 2007).

2.2.1.1.b Secondary structure

Although there is little sequence similarity between SECIS RNAs, the SECIS 2D structure is well conserved within each kingdom.

Bacterial SECIS is an approximately 40 nucleotide-long stem-loop structure. Although SECIS sequence vary depending on species, the structure is grossly conserved in different organisms and the apical loop is important for binding to the specialized translational factor SelB (Böck, 2006).

In eukaryotes, the secondary structure models of the SECIS were proposed based on structure probing experiments (Walczak et al, 1998; Walczak et al, 1996). There are two types of functional SECIS RNAs in eukaryotes, called form 1 and form 2. Both forms have conserved structures, composed of internal loops, helices and four consecutive non-Watson-Crick base pairs, called the quartet (Figure 4). They also present conserved As in the apex, and A/G 5' to the quartet. Form 2 SECIS possesses an additional helix 3 but a shorter apical loop (Fagegaltier et al, 2000b; Grundner-Culemann et al, 1999). Structure-based sequence alignments of SECIS elements from the currently available eukaryotic selenoproteome resulted in a collection of 62 form 1 and 224 form 2 SECIS sequences, showing that form 2 SECIS are more widespread than form 1 (Chapple et al, 2009). However, mRNAs encoding the same selenoprotein can harbor different forms of SECIS depending on the species. For example, SelM mRNA contains a form 2 SECIS element in mammals, whereas form 1 is present in zebrafish (Korotkov et al, 2002). Furthermore, introduction of mutations in the apex of forms 1 and 2 led to the conclusion that both types of SECIS can function equally well under the experimental conditions used (Grundner-Culemann et al, 1999). Why there are two forms of SECIS elements is still unclear.

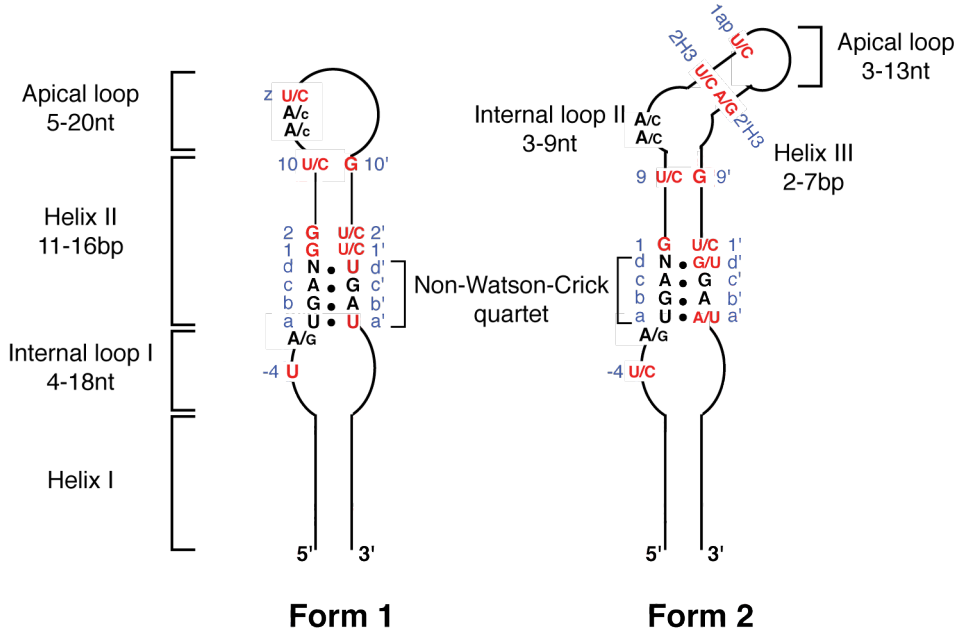


Figure 4. Secondary structure models of form 1 and 2 SECIS.

The conserved sequence and structural features are indicated. Novel conserved residues found by SECISaln are shown in red (Chapple et al, 2009). The positions of the conserved nucleotides are indicated in blue. abcd/a'b'c'd' indicates base-pairs forming the non-Watson-Crick quartet. Numberings (-4, 1, 2, 9, 10, 1' and 2') show the distance from the quartet. Position "z" is the first nucleotide after the conserved A/Cs, positions 2H3/2'H3 are the second base pair of the Helix III abd 1ap is the first nucleotide of the apical loop. The structures are from (Fagegaltier et al, 2000b; Grundner-Culemann et al, 1999; Walczak et al, 1998; Walczak et al, 1996).

The non-Watson-Crick quartet is essential to selenocysteine incorporation *in vivo*, and constitutes the binding site for SECIS Binding Protein 2 (the function of this key protein will be detailed in paragraph 2.2.2.2.). This motif contains a central tandem of sheared G.A/A.G base pairs (Fagegaltier et al, 2000b; Walczak et al, 1998; Walczak et al, 1996). Such a tandem of G.A/A.G base pairs is also found in other RNAs such as ribosomal RNAs, snRNAs and snoRNAs, and it constitutes a conserved structure, called the K (kink)-turn motif (Klein et al, 2001). The K-turn is an RNA structural motif that binds proteins in most of the cases and mediates RNA tertiary structure interactions. The K-turn is a two-stranded, helix-internal loop-helix motif comprising about 15 nucleotides, characterized by base stacking, the presence of a tandem of G-A sheared base pairs, and a protruding residue accommodated by a protein pocket. As a result, the structure has a kink of 120° in the phosphodiester backbone that causes a sharp turn in the RNA helix (Klein et al, 2001). A K-turn was also found in the

crystal structure of U4 snRNA-15.5kD, L30e RNA-L30e and sRNA-L7Ae complexes (Chao & Williamson, 2004; Moore et al, 2004; Vidovic et al, 2000)(Figure 5).

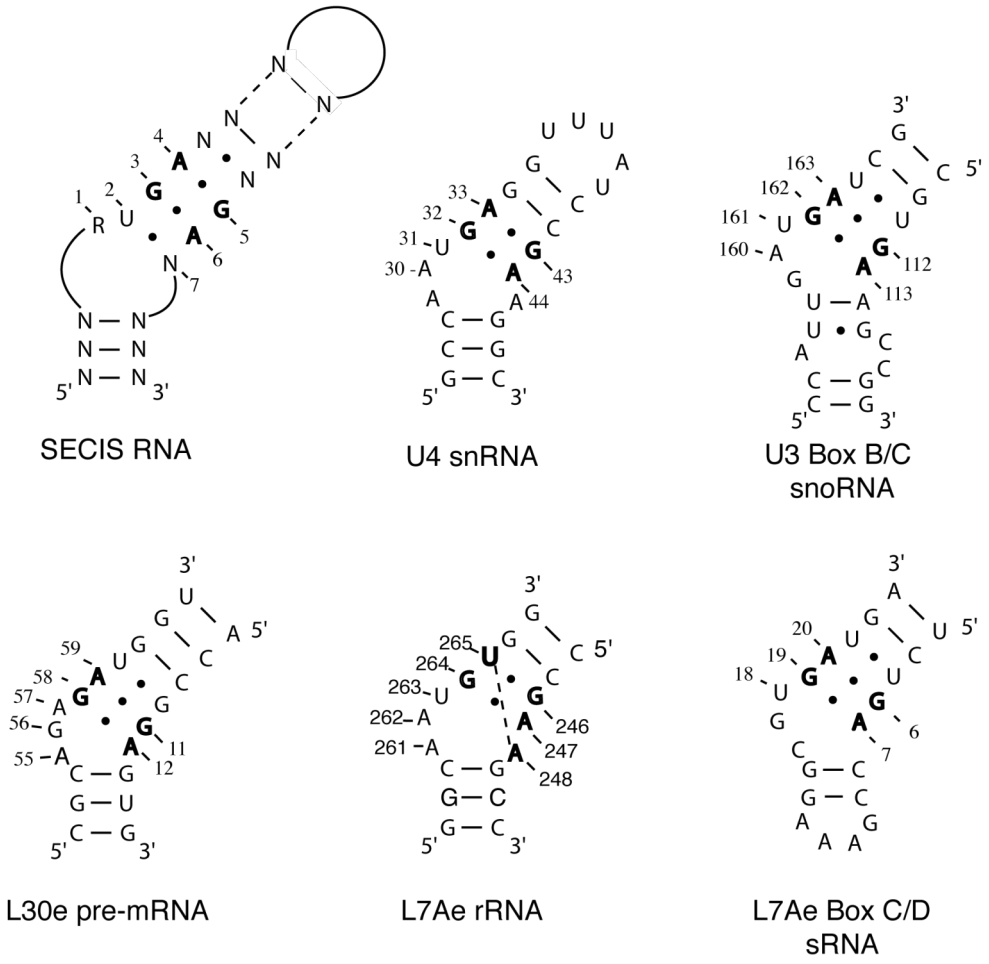


Figure 5. The secondary structure of SECIS RNA and various K-turn RNAs.
The secondary structures of the U4 snRNA, L30e pre-mRNA, L7Ae rRNA, L7Ae box C/D sRNA were taken from the crystal structures of the corresponding RNA-protein complexes (Chao & Williamson, 2004; Moore et al, 2004; Vidovic et al, 2000). Those of SECIS RNA and U3 Box B/C snoRNA were determined by structure probing analyses (Fagegaltier et al, 2000b; Marmier-Gourrier et al, 2003; Walczak et al, 1998; Walczak et al, 1996). The sheared G.A/A.G base pairs are indicated in bold. The figure is taken from (Allmang & Krol, 2006a).

Because of these secondary structure similarities, we have proposed that the SECIS RNA is a K-turn like motif (Allmang & Krol, 2006a). Furthermore, this is supported by previous findings where structure probing and mutagenesis data allowed a 3D model for the SECIS RNA to be proposed by computer modeling. In this model the phosphodiester backbone is

indeed bent at the internal loop (Walczak et al, 1996). Compared to canonical K-turn RNAs, SECIS elements have larger internal loops. This larger internal loop and a long helix 2 provide specificity for SBP2 binding to the SECIS (Cléry et al, 2007). The nucleotide 5' to the quartet is A in most of the cases, but G can be found, and the replacement by G does not affect SECIS activity *in vivo* (Buettner et al, 1999; Fagegaltier et al, 2000b; Taskov et al, 2005). Interestingly, in a patient suffering from a SEPN1-related myopathy, a mutation in the non-Watson-Crick quartet of the SEPN1 SECIS element that prevents the interaction with SBP2, was found to be responsible for the pathology (Allamand et al, 2006)(Figure 6 A).

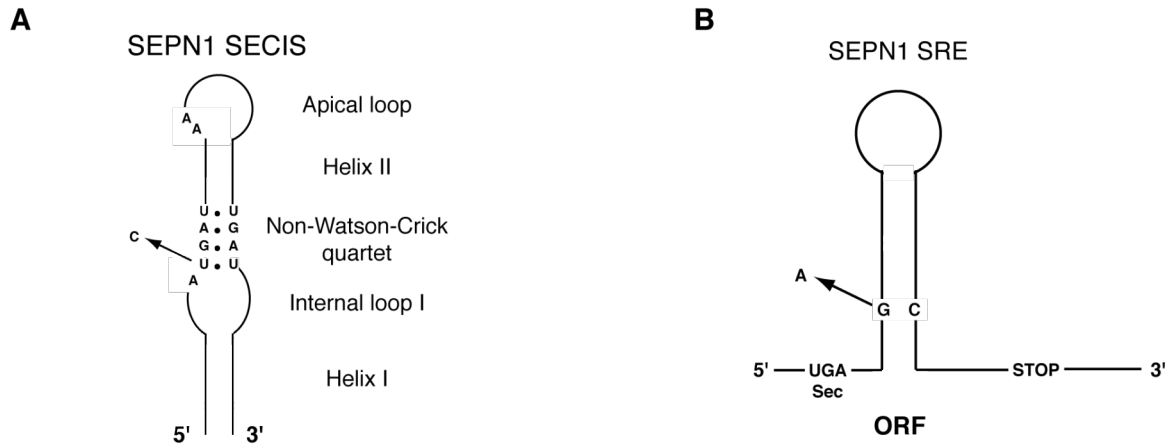


Figure 6. The SECIS and SRE elements of SEPN1 mRNAs

A. Secondary structure of the SEPN1 SECIS RNA. The conserved U in the non-Watson-Crick quartet is essential for the recognition by SBP2 and the U to C mutation abolishes SBP2 binding (arrow). This mutation was initially found in Selenoprotein N (SEPN) SECIS element from a patient with a *SEPN1*-related myopathy (Maiti et al, 2008). **B.** Secondary structure model of the SRE RNA. The SRE hairpin structure is located within the open reading frame (ORF) of certain selenoprotein mRNAs, here the selenoprotein N (SEPN1) (Howard et al, 2005). The G to A mutation was found in a patient with SEPN1-related myopathy (Allamand et al, 2006).

RNA structure probing experiments indicated that the conserved As in the apical loop (form 1) or the internal loop 2 (form 2) are single stranded and well accessible (Fagegaltier et al, 2000b). However, some exceptions to the invariant presence of As were reported. For example, mammalian SelM SECIS and some of Chlamydomonas form 2 SECIS contain Cs without altering Sec incorporation activity (Korotkov et al, 2002; Novoselov et al, 2002). Other examples of Cs or a combination of As and Cs or even Gs were later found in

eukaryotes (Chapple et al, 2009; Lobanov et al, 2006a; Lobanov et al, 2007; Lobanov et al, 2006b). Although this apical A/C rich loop is not necessary for SBP2 binding, site-directed mutagenesis showed that the unpaired As/Cs are important for selenoprotein synthesis *in vivo* (Berry et al, 1993).

Such detailed knowledge of the secondary structure of SECIS element was used in computational analysis to identify novel selenoprotein mRNAs (Kryukov et al, 1999; Lescure et al, 1999) and to establish the whole mammalian selenoproteome with the help of SECISearch, a computer program for analyzing structural and thermodynamic features of SECIS elements (Kryukov et al, 2003). Recently, the well-defined secondary structure of the SECIS RNA and the increased size of the eukaryotic selenoproteome allowed the establishment of a web-based tool, SECISaln, providing extensive structure-based sequence alignments of SECIS elements (Chapple et al, 2009). Analyzing the structural alignments produced by SECISaln highlighted a few previously undetected conserved residues (see Figure 4). There is an overrepresentation of G at position 1 (3' to the quartet) and a corresponding overrepresentation of C or U at position 1' (see Figure 4). SECISaln also found differences between form1 and form2 SECISes. The most striking one is a well-conserved U 4 nucleotides upstream of the quartet (at position -4 in Figure 4) in form1 SECIS, whereas C can also be found in the form 2 SECIS (Chapple et al, 2009).

2.2.1.2. SRE

Another cis-acting element reported recently is the Selenocysteine codon Redefinition Element (SRE). SRE is a phylogenetically conserved stem-loop structure located within the coding region of selenoprotein mRNAs, adjacent to the UGA Sec codon. This element is sufficient to stimulate readthrough of the UGA Sec codon in the absence of a SECIS element in the 3'UTR in a synthetic mRNA, although higher readthough efficiency is observed in its presence. SelN SRE inserted in a dual-luciferase system had a stimulatory effect on the UGA Sec decoding *in vitro* (Howard et al, 2005; Howard et al, 2007). SRE was experimentally analyzed in SelN mRNA, but bioinformatic approaches predicted found SREs in a few other selenoprotein mRNAs such as SPS2, SelH, SelO and SelT (Howard et al, 2005; Pedersen et al, 2006). Their 2D structure, however, is not conserved. The presence of an SRE in some but not all selenoprotein mRNAs implies a differential role in regulating selenoprotein expression at the translation level. Four point mutations leading to the SEPN1-related myopathy were

found in the SelN SRE element. One of them weakens the secondary structure of SRE by abolishing a G-C base pair, leading to a decrease in Sec incorporation and SelN levels (Maiti et al, 2008). This data supports the importance of the SRE structure for selenoprotein synthesis (Figure 6B).

2.2.2. Trans-acting factors

2.2.2.1. EFSec

In bacteria, SelB is the translation elongation factor specialized for selenocysteine incorporation. The N-terminal domain of SelB is highly-sequence similar and functionally homologous to EF-Tu (see Figure 7), the general translation elongation factor, and the C-teminal domain binds to SECIS. SelB binds specifically and uniquely the Sec-tRNA^{Sec} (Böck, 2006).

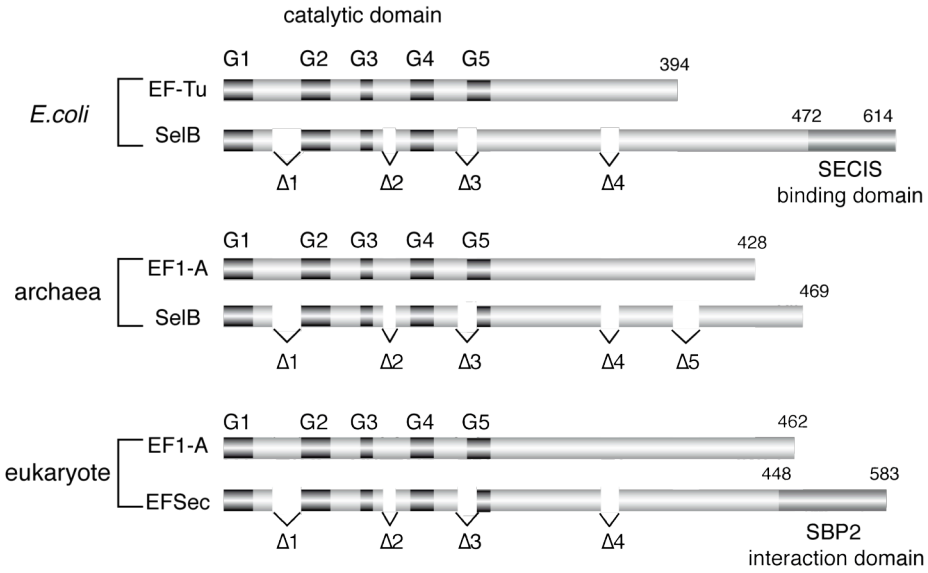


Figure 7. Schematic representations of the selenocysteine specialized translation elongation factors compared to general elongation factors.

The specialized translation elongation factors in *E. coli,* archaea and eukaryotes (SelB or EFSec) are shown, in comparison with the general elongation factors EF-Tu or EF1-A. The GTP-binding domains are indicated by G1-G5. Δ1-Δ5 are the deletion regions relative to EF-Tu or EF1-A. The C-terminal extensions in *E. coli* and eukaryotes contain the SECIS binding domain and the SBP2 interaction domain, respectively. The figure is taken from (Allmang & Krol, 2006b).

EFSec is the mammalian homolog of SelB. It was independently characterized in mouse by our laboratory and by Berry's group (Fagegaltier et al, 2000a; Tujebajeva et al, 2000). EFSec binds specifically to the Sec-tRNA^{Sec} but not to Ser-tRNA^{Sec}. Like for bacterial SelB, the N-terminal domain of EFSec has sequence similarities with the general elongation factor EF1A and contains homologies to the G1-G4 GTP-domain (Fagegaltier et al, 2000a). The length of the C-terminal extension varies in different organisms. In contrast to SelB, EFSec cannot bind specifically the SECIS RNA, indicating another role than in bacteria. EFSec co-immunoprecipitates SBP2 from mammalian cell extracts, and the SBP2 interaction domain resides in the C-terminal extension (Tujebajeva et al, 2000). Thus, it is likely that EFSec is recruited to the selenocysteine incorporation machinery by SBP2.

EFSec contains putative nuclear export and nuclear localization signals, in the N-terminal domain and the C-terminal SBP2 interaction domain, respectively. The EFSec subcellular localization varies depending on the cell line and may be influenced by SBP2 levels and localization (de Jesus et al, 2006).

Archaeal EFSec (called SelB) was identified in *Methanococcus jannaschii* (Rother et al, 2000), and it possesses sequence features characteristic of bacterial SelB and EFSec (Fagegaltier et al, 2000a; Rother et al, 2000). Furthermore, crystal structure of the *Methanococcus maripaludis* EFSec revealed that its overall shape resembles a 'chalice' observed so far in translational initiation factor IF2/eIF5B (Leibundgut et al, 2005). This raises the interesting question of whether mechanistic similarities could exist between Sec incorporation and translational initiation.

2.2.2.2. SBP2

SBP2 (SECIS Binding Protein 2) is a trans-acting factor that plays a central role in eukaryotic Sec incorporation. SBP2 was isolated and functionally characterized in rat and humans (Copeland & Driscoll, 1999; Lescure et al, 2002). Its known functions are SECIS binding, ribosomal binding and Sec incorporation. The importance of SBP2 for selenoprotein synthesis was shown by SBP2 depletion which results in decreased Sec incorporation in cells and *in vitro* (Copeland et al, 2000; Papp et al, 2006). Additionally, patients carrying mutations in SBP2 display abnormal thyroid hormone metabolism leading to reduction of DIO2 activity (Dumitrescu et al, 2005).

2.2.2.2.a Domain structure of SBP2

Mammalian SBP2 is about 850 amino acid long. The domain structure of SBP2 can be roughly divided into two parts, the N-terminal and C-terminal domains. The N-terminal domain is not essential for selenoprotein synthesis *in vitro* (Copeland et al, 2000). This domain has no sequence similarity with any other known protein, thus its function still remains unknown except for the presence of an NLS (Nuclear Localization Signal) (Papp et al, 2006).

The C-terminal domain is essential and sufficient for Sec incorporation *in vitro*. It contains the RNA-binding domain in a region lying between positions 516 and 777 in rat SBP2 (Copeland & Driscoll, 2001, see also Figure 8). This RNA-binding domain includes a conserved motif, called the L7Ae motif. The L7Ae motif was originally identified as a putative RNA binding motif by a computational study (Koonin et al, 1994). It is shared by other functionally unrelated proteins such as 15.5kD/Snu13, Nhp2 and ribosomal proteins L7Ae and L30, all of which bind to a K-turn motif and trigger RNP complex formation. Later, point mutation analysis showed that the L7Ae motif in SBP2 is essential for SECIS RNA binding (Allmang et al, 2002). In addition to the L7Ae motif, SBP2 specific sequences upstream from the L7Ae motif also play an important role for the interaction with the SECIS RNA. The RNA binding domain of SBP2 is thus bipartite (Bubenik & Driscoll, 2007; Donovan et al, 2008; Takeuchi et al, 2009). The characterization of the additional RNA binding module represents an important contribution to my thesis and will be detailed in Chapter1 of Part 2. The C-terminal domain contains two functional NES (Nuclear Export Signal) (Papp et al, 2006).

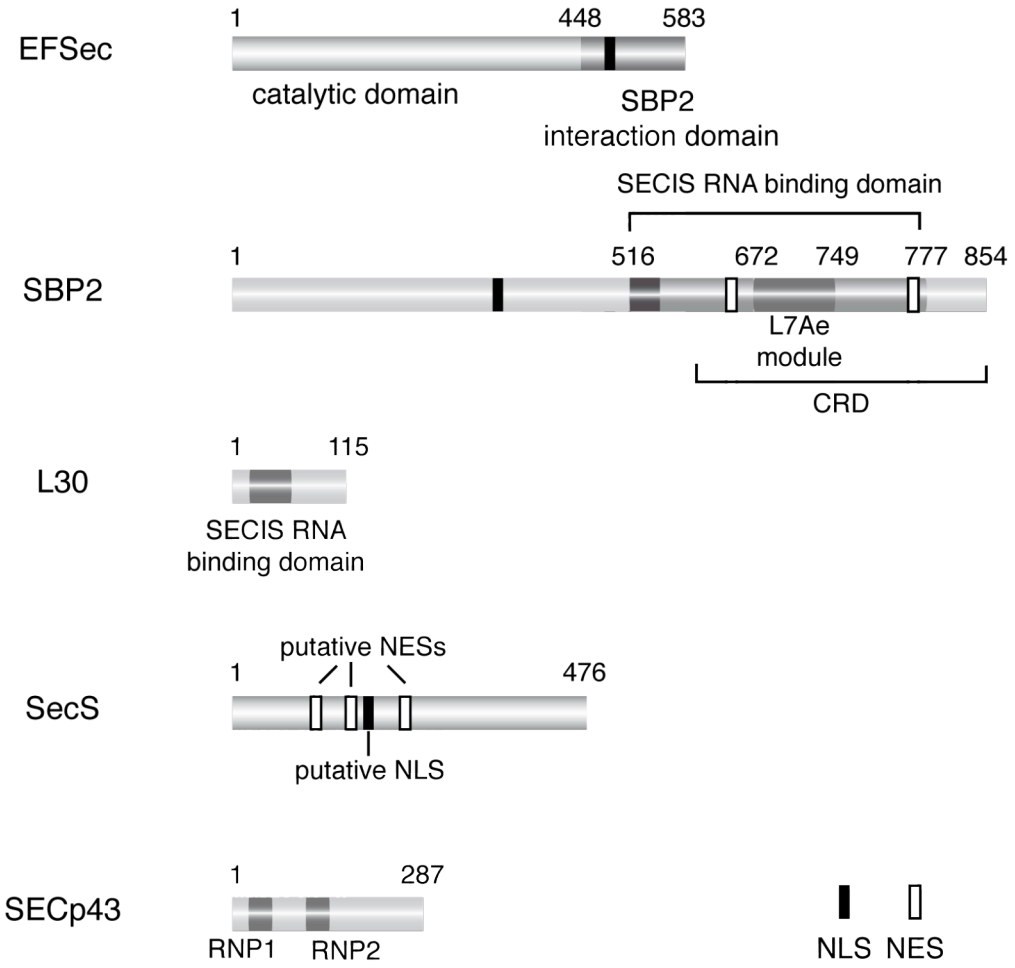


Figure 8. Schematic representation of protein factors involved in selenoprotein synthesis.
NLS: nuclear localization signal, NES: nuclear export signal, CRD: cysteine rich domain, RNP: RNP consensus RNA binding sequence. This was described in (Papp et al, 2007; Small-Howard & Berry, 2005)

2.2.2.2.b. SECIS binding

SBP2 specifically binds to the SECIS RNA via the non-Watson-Crick quartet. Multiple sequence alignments revealed that the RNA binding domains of SBP2 and 15.5kD/Snu13p have higher sequence similarity between each other than with other members of the L7Ae family. 79 amino acids in the human SBP2 RNA binding domain (position 672-750) possess 47% similarity (26% identity) with the RBD of 15.5kD/Snu13p, 43% similarity (20% identity) with Nhp2p and 30% similarity (16% identity) with the yeast L30 and human L7A proteins. A structure-guided strategy based on the SBP2 and 15.5kD similarities, and the crystal structure of the 15.5kD-U4 snRNA complex (Vidovic et al, 2000), predicted amino acids in the L7Ae motif of SBP2 that should be critical for SECIS RNA binding (Allmang et

al, 2002). Alanine scanning mutagenesis identified 8 important amino acids in SBP2. Four crucial amino acids are postulated to recognize characteristic bases in the non-Watson-Crick quartet of the SECIS RNA, the opposite G residues in the sheared G.A/A.G base pairs and the U residue. These findings led to the proposal that the SBP2-SECIS RNA interaction principles are indeed similar to those of the 15.5kD-U4 snRNA complex (Allmang et al, 2002) (Figure 9 A and B).

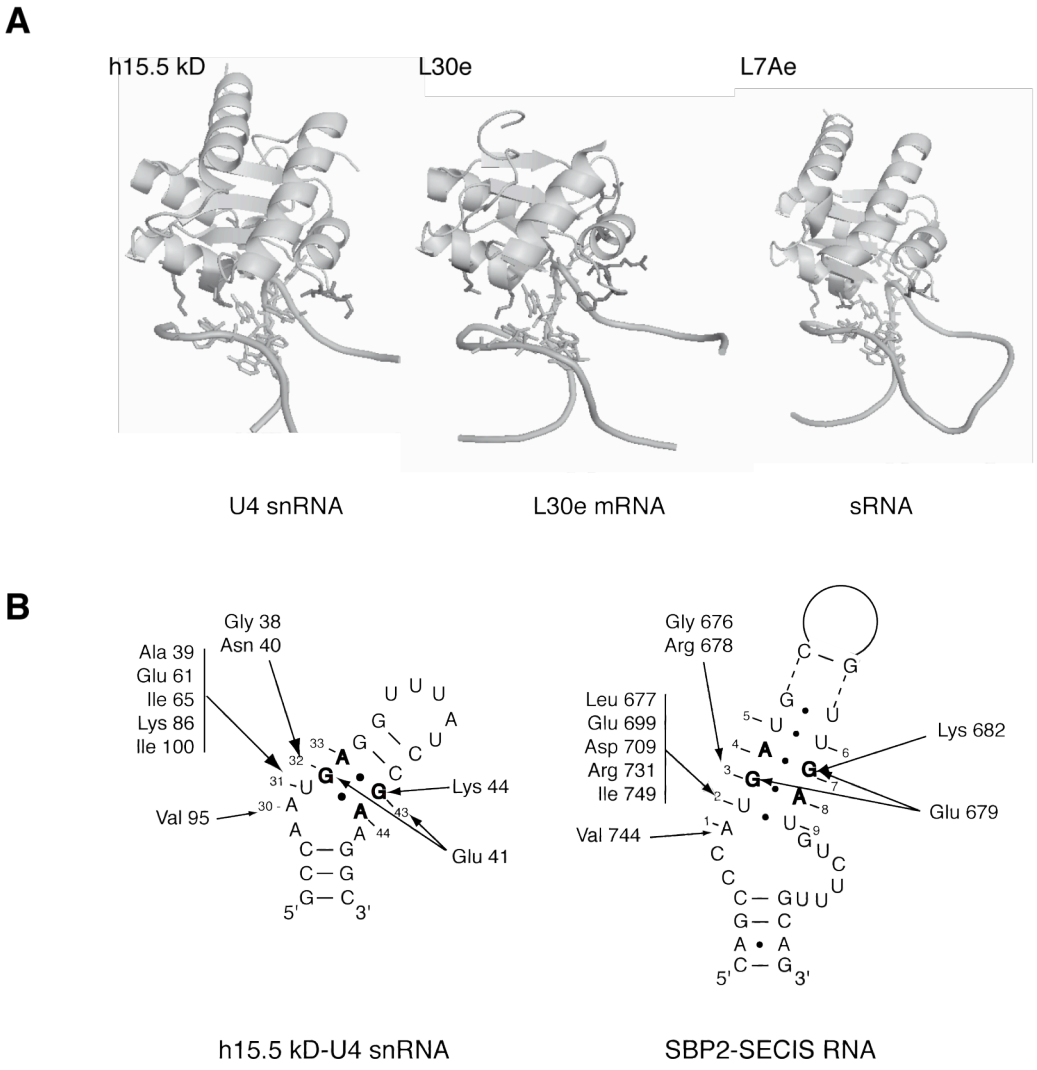


Figure 9. RNA-protein interfaces at various L7Ae protein-K turn RNA complexes.

A. Overall crystal structures of human 15.5kD protein-U4 snRNA, L30e protein-L30e mRNA and L7Ae protein-sRNA complexes (Chao & Williamson, 2004; Moore et al, 2004; Vidovic et al, 2000). The figure is taken from (Allmang & Krol, 2006a). **B.** Scheme of the RNA-protein interactions in the h15.5kD-U4 snRNA derived from the crystal structure (Vidovic et al, 2000) and SBP2-SECIS RNA (inferred from a structure-guided strategy (Allmang et al, 2002) complexes. Similar interaction principles govern both complexes, as described in (Allmang et al, 2002).

This raised the question of whether SBP2 can bind the RNA targets of other L7Ae proteins. To answer this question, SELEX and site-directed mutagenesis were performed and showed that SBP2 is capable to bind K-turn motifs with a protruding U residue (Figure 5). However, SBP2 exhibits higher affinity for the RNA motif when the bulged loop is converted to a large internal loop. Helix 1 and internal loop 1 are also important for SBP2 binding. The SBP2-SECIS interaction is therefore similar to that of L7Ae proteins/K-turn RNAs but requires additional RNA-protein contacts (Cléry et al, 2007) (Figure 10).

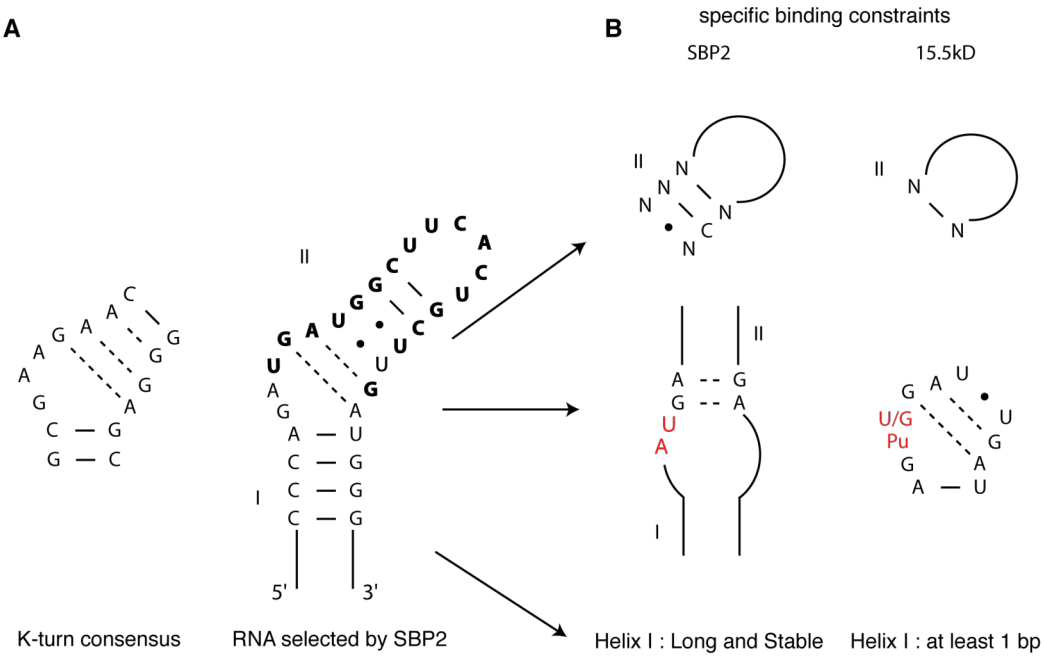


Figure 10. SECIS RNA determinants for SBP2 binding.

The sequence and structural determinants for SBP2 binding were identified by SELEX and mutagenesis. Stronger recognition constraints were identified for SBP2 than for 15.5kD(Cléry et al, 2007). **A.** K-turn consensus is shown on the left, the secondary structure of the RNA selected by SBP2 with the highest affinity on the right. The selected RNA can adopt a K-turn structure. Nucleotides in bold were initially degenerated in the SELEX experiment, the selected RNA with the highest affinity for SBP2 is represented. **B.** Specific binding constraints for SBP2 and 15.5kD. While 15.5kD binds the K-turn motif with a small internal loop and a short helix I (at least one base pair, right), SBP2 requires long and stable helices I and II (left, upper and bottom panels) and a rather large internal loop (bottom panel) for SECIS RNA recognition. SBP2 also requires strict sequence conservations compared to 15.5kD (shown in red). The higher specificity of SBP2 for RNA recognition suggests the existence of additional protein-RNA contacts in SBP2-SECIS RNA complex, compared to other L7Ae proteins.

Furthermore, the SECIS-binding affinity of SBP2 differs between different selenoprotein mRNAs and is suggested to play a major role in determining the differential selenoprotein mRNA translation and sensitivity to nonsense-mediated decay (Squires et al, 2007).
In addition to mutations in the coding frame of the SelN protein causing SEPN1-related myopathies, a single homozygous point mutation in the SelN mRNA SECIS element was also shown to be responsible for the pathology. This mutation was found in the non-Watson-Crick quartet, preventing the interaction with SBP2 (Allamand et al, 2006)(Figure 6A).

2.2.2.2.c. EFSec-SBP2 interaction

An interaction between SBP2 and EFSec was observed in co-immunoprecipitation assays using mammalian cell extracts (Tujebajeva et al, 2000), requiring the $tRNA^{Sec}$ (Zavacki et al, 2003). This interaction occurred via the C-terminal 64 amino acid sequence of EFSec and the C-terminal domain of SBP2 (Donovan et al, 2008; Zavacki et al, 2003). However, no interaction could be detected in rabbit reticulocyte lysate. The interaction could not be reconstituted *in vitro* unless a masking region of EFSec was removed (Zavacki et al, 2003). Surprisingly, however, a recent study reported the EFSec-SBP2 interaction *in vitro* in the sole presence of the SECIS RNA in the reaction mixture (Donovan et al, 2008). The discrepancies observed by the various investigators may be caused by differences in the experimental conditions, a co-immunoprecipitation assay using cell extracts on the one hand, EMSA using recombinant proteins on the other. A 6xHis tagged EFSec may also be detrimental to the interaction (Donovan et al, 2008). Furthermore, co-expression of SECp43 was shown to promote the interaction between EFSec and SBP2 (Small-Howard et al, 2006), explaining why it might be difficult to observe it *in vitro*.

2.2.2.2.d. Ribosomal binding

Since eukaryotic selenoprotein mRNAs contain the SECIS element in the 3'UTR, selenocysteine incorporation requires factor(s) that connect the ribosome with the SECIS element to tell not to stop at the UGA Sec codon. Indeed, SBP2 plays an important role in this process. Glycerol gradient centrifugation established that SBP2 quantitatively associates with ribosomes through its RNA binding domain (Kinzy et al, 2005). SECIS RNA can compete with the ribosome for SBP2 binding, indicating that SBP2 is not able to simultaneously interact with the ribosome and the SECIS RNA (Kinzy et al, 2005). Like SECIS binding, the

ribosome binding activity of SBP2 is essential for Sec incorporation (Caban et al, 2007; Donovan et al, 2008; Kinzy et al, 2005). We have analyzed ribosomal binding in more detail using purified ribosomal subunits. The results will be described in Chapter 1 of Part 2.

2.2.2.2.e. Expression and localization

SBP2 mRNA is expressed in most tissues as revealed by Northern blotting analyses, with higher levels in testis (Copeland et al, 2000; Lescure et al, 2002). SBP2 was detected predominantly in the cytoplasm, in stable association with ribosomes (Copeland et al, 2001; Kinzy et al, 2005; Papp et al, 2006). SECp43, which promotes the interaction between EFSec and SBP2, interacts *in vivo* with the Sec-tRNA^{Sec}/EFSec in a high molecular weight complex (Small-Howard et al, 2006), implying that SBP2 is also present in the high molecular weight complex comprising EFSec and SECp43. However, recent studies showed that SBP2 can undergo nucleocytoplasmic shuttling via intrinsic nuclear localization (NLS) and nuclear export signals (NES) that are located in the N-terminal part and the C-terminal cysteine-rich domain (CRD), respectively (Papp et al, 2006). The nuclear export of SBP2 is dependent on the CRM1 pathway. Indeed, the best characterized pathway for nuclear export of proteins from nucleus to cytoplasm involves the nuclear export receptor CRM1, which binds to NES. Inhibition of CRM1 induces nuclear sequestration of SBP2 and decreases selenoprotein synthesis. Interestingly, oxidative stress induces nuclear accumulation of SBP2 through the formation of disulfide (S-S) and/or glutathione-mixed disulfide (S-SG) bonds in the redox sensitive cysteines of the CRD, which masks the NES. These modifications are efficiently reversed *in vitro* by thioredoxins and glutaredoxins. These antioxidant systems might regulate the redox state of SBP2. The nuclear retention of SBP2 after oxidative stress reduces Sec incorporation, suggesting a mechanism to regulate selenoprotein expression (Papp et al, 2006)(Figure 11).

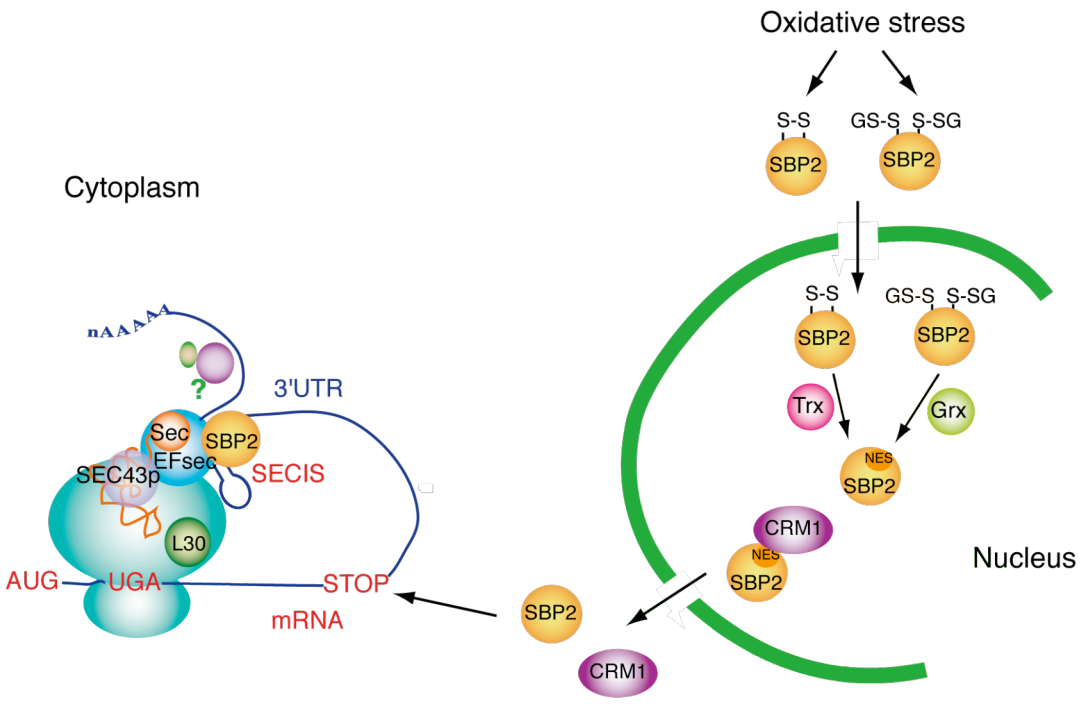


Figure 11. Proposed model for the regulation of SBP2 subcellular localization and function after oxidative stress.

The model is from (Papp et al, 2006). Oxidative stress oxidizes redox-sensitive SBP2 cysteine residues to disulfide (S-S) or glutathione-mixed disulfide (S-SG) bonds, triggering its nuclear translocation. An oxidized state masks the NES, inhibiting its nuclear export and leading to nuclear retention. During cell recovery, the thioredoxin (Trx) and the glutaredoxin (Grx) reduce the cysteine residues, leading to exposure of the NES and CRM1-dependent nuclear export. (Figure adapted from (Allmang et al, 2009))

In silico and *in vivo* studies established a complex alternative splicing pattern in the 5' region of the human SBP2. There are at least eight splice variants encoding five isoforms with N-terminal sequence variation. One of them, the most abundant variant after the full-length SBP2, contains a mitochondrial targeting sequence (MTS), perhaps functioning in the translation of selenoproteins targeted to mitochondria on mitochondria-bound polysomes (Papp et al, 2008). In the same report, it was shown that full-length and some alternatively spliced variants are subject to a coordinated transcriptional and translational regulation in response to UVA irradiation-induced stress.

It was recently reported that 3'UTR sequences in the SBP2 mRNA are highly conserved and two RNA binding proteins, CUG-BP1 and HuR, interact with this region. Both CUG-BP1 and HuR are involved in mRNA stability and translational regulation of their mRNA targets,

suggesting that the SBP2 mRNA is regulated at the post-transcriptional level (Bubenik et al, 2009).

2.2.2.3. L30

L30 is specific to eukaryotes and archaea, although not all archaeal ribosomes possess it. It belongs to the L7Ae protein family. L30 is part of the large ribosomal subunit but exists also as a free protein. L30 binds its own pre-mRNA and regulates its splicing and expression (MacÌas et al, 2008). However, its function during translation is still elusive. The rat L30 protein was reported to be a component of selenocysteine incorporation machinery. It binds the SECIS RNA *in vitro* and *in vivo* and competes with SBP2 for SECIS RNA binding, especially under high Mg^{2+} concentration, because these metal ions induce kink-turn conformation in the SECIS RNA and increase the L30-SECIS interaction (Chavatte et al, 2005). The ribosomal-bound L30 has a higher affinity for the SECIS RNA than the recombinant protein. This allowed to suggest a model in which L30 displaces transiently SBP2 during selenocysteine incorporation (Chavatte et al, 2005).

2.2.2.4. Other proteins

Nucleolin is a multifunctional protein described to function in many pathways including transcriptional regulation and maturation of ribosomal RNA (reviewed in Mongelard & Bouvet, 2007). cDNA library screening identified this protein as a SECIS RNA binder (Wu et al, 2000). In contrast to SBP2, nucleolin is less discriminatory for SECIS RNA binding (Squires et al, 2007).

Nuclease Sensitive Element Binding Protein 1 (NSEP1) is also called DNA-binding protein 1B (dbpB) or Y-box binding protein 1 (YB-1) in accordance with its DNA binding properties and its role as a transcription factor. This protein was also shown to bind the SECIS RNA and to be functionally involved in selenoprotein synthesis in mammalian cells. (Shen et al, 1998, Shen et al, 2006)

.

2.3. Sec incorporation model

How the selenocysteine incorporation machinery tells the ribosome not to stop at the UGA codon still remains controversial. However, two distinct models have been proposed describing the selenocysteine incorporation mechanism. (Figures 12)

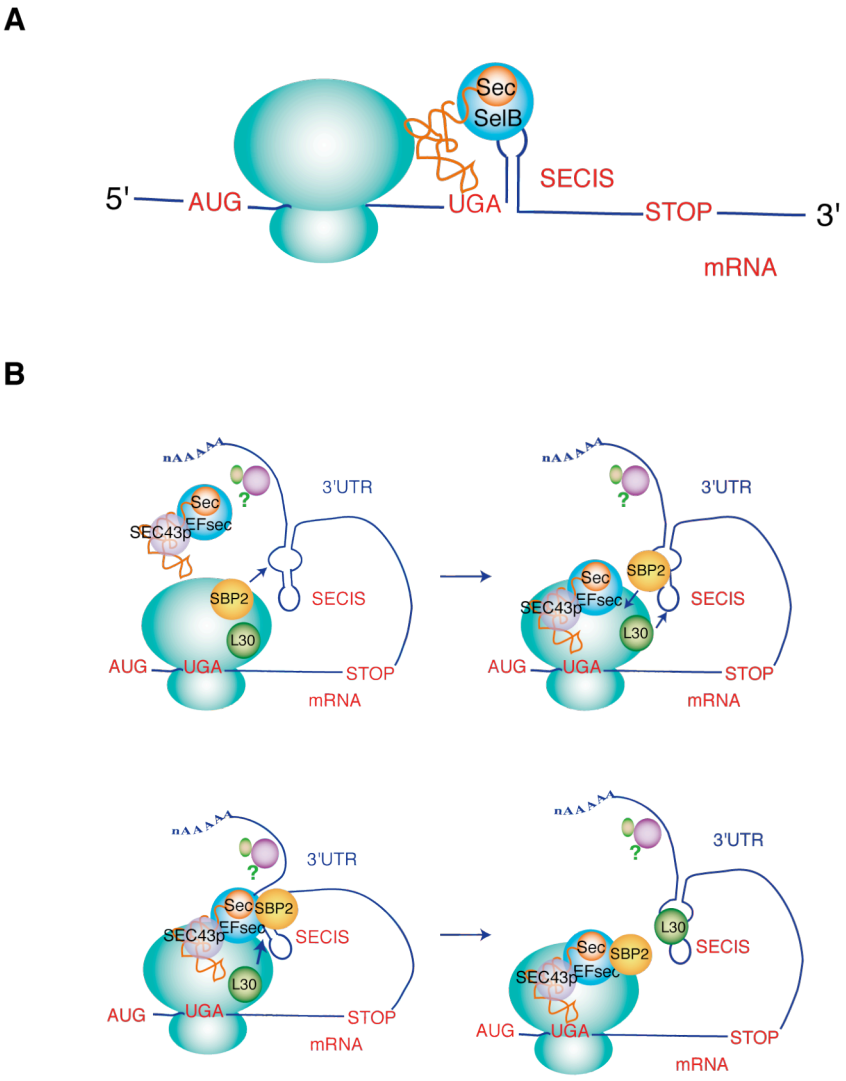


Figure 12. Selenocysteine incorporation models

A. In bacteria, SelB binds both the Sec-tRNA^{Sec} and bSECIS element immediately downstream of the UGA Sec codon. The Sec-tRNA^{Sec}, harbored by SelB, is brought directly to the UGA codon through the bSECIS-SelB interaction, allowing the incorporation of selenocysteine into the nascent polypeptide chain. **B.** In eukaryotes, two models were proposed. (1, upper panel) SBP2 travels with the ribosome, interacts with the SECIS RNA and the EFSec/Sec-tRNA^{Sec} to deliver this complex to the A site of the ribosome. Then, L30 displaces SBP2 from the SECIS RNA; (2, bottom panel) SBP2 is bound to the SECIS RNA and recruits the EFSec/Sec-tRNA^{Sec} complex to the SECIS RNA. Then,

ribosome-bound L30 displaces SBP2 so that the SBP2/ EFSec/Sec-tRNA^{Sec} complex can be delivered to the ribosome. Figures are taken from (Allmang & Krol, 2006b).

In the first model, SBP2 travels with the ribosome. The ribosome stalling at a UGA Sec codon allows the SECIS element to interact with SBP2, bringing the Sec-tRNA^{Sec} bound EFSec to the ribosomal A-site (Donovan et al, 2008; Kinzy et al, 2005). This model was proposed based on the findings that SBP2 quantitatively interacts with the ribosome, but cannot bind simultaneously to the SECIS RNA. In the second model, SBP2 is bound to the SECIS element prior to translation, and recruits the EFSec/Sec-tRNA^{Sec} complex prior to UGA decoding. Then, an approaching ribosome leads L30 to displace SBP2 by competing for the SECIS element. As a consequence, the SBP2/EFSec/Sec-tRNA^{Sec} complex is delivered to the ribosomal A-site (Chavatte et al, 2005).

3. Selenoprotein mRNP assembly

3.1. Nuclear assembly

Nonsense-mediated decay (NMD) is a consequence of premature termination codon (PTC) recognition during a pioneer round of translation. This pathway is important for cells because it functions as a quality control mechanism to eliminate aberrant transcripts. In mammalian cells, NMD occurs when an mRNA contains a non-sense codon located more than 25 nucleotides upstream of the post-splicing exon junction complex (EJC) (Lejeune & Maquat, 2005). Because selenoprotein mRNAs contain in-frame UGA codons, they can be targets for NMD. Analyses of the genome structure suggest that 14 of the 25 human selenoprotein mRNAs are potentially NMD sensitive (Squires et al, 2007). Indeed, selenoprotein mRNAs are subjected to NMD under low selenium condition (Moriarty et al, 1998; Sun et al, 2001; Weiss & Sunde, 1998). One mechanism that could allow selenoprotein mRNAs to circumvent NMD is the early assembly on selenoprotein mRNAs, in the nucleus, of the selenocysteine incorporation machinery (therefore before initiation of the first round of translation). This is suggested by the nucleocytoplasmic shuttling ability of SBP2 (de Jesus et al, 2006) and supported by the supramolecular complex formation model of the selenoprotein synthesis machinery (Small-Howard et al, 2006). (Figure 13)

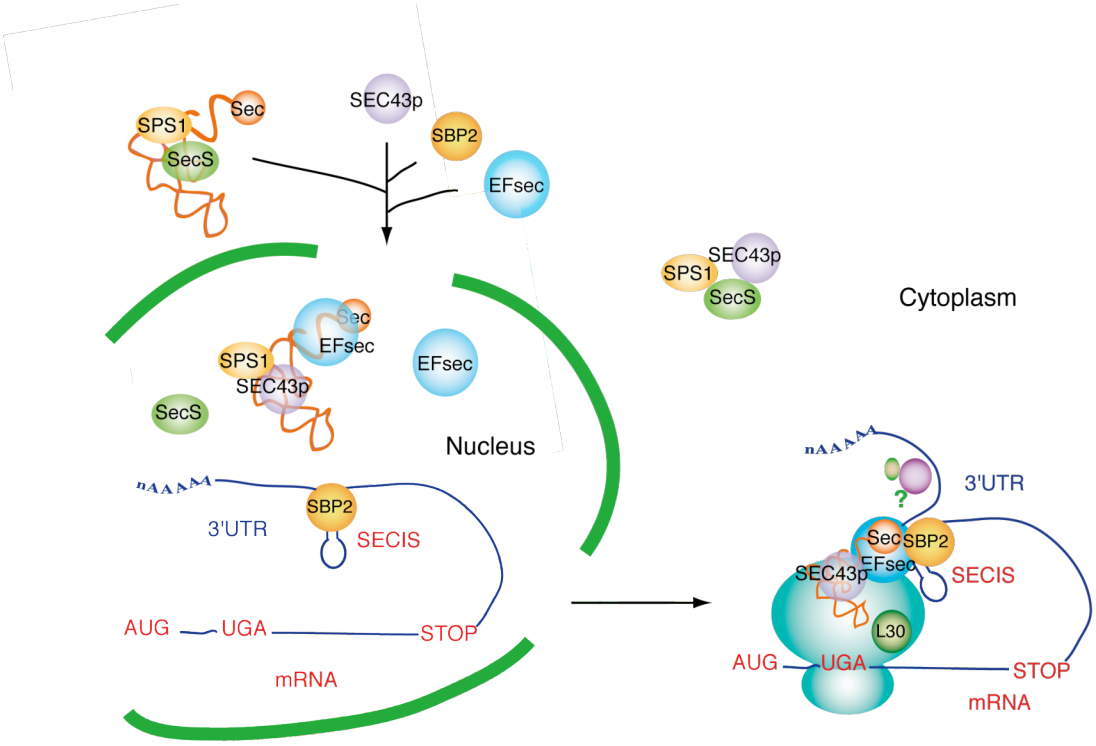


Figure 13. Nuclear assembly of the selenoprotein synthesis machinery

This model incorporates experimental evidence regarding selenocysteine and selenoprotein synthesis according to (de Jesus et al, 2006; Small-Howard et al, 2006). Shuttling of the Sec-tRNA^{Sec}/EFSec/SPS1/SECp43 complex into the nucleus and association with SBP2 and the SECIS element are depicted. Cytoplasmic export of the selenocysteine incorporation machinery is shown on the left. The figure is adapted from (Allmang et al, 2009).

3.2. Assembly of selenoprotein mRNAs - similarities with sn/snoRNP assembly

The RNA binding domain of SBP2 belongs to the L7Ae family (see 2.2.2.2). The L7Ae family proteins have been found in many essential ribonucleoproteins (RNPs). For example, the 15.5kD/ Snu13p, with which the SBP2 L7Ae motif has high sequence similarity, plays a central role in the formation of the U4 snRNP, box C/D snoRNPs and the B/C structure of the U3 snoRNP (Watkins et al, 2000). The L7Ae proteins all bind to a common RNA structure, the K-turn (See also 2.2.1.1.b). Binding of the L7Ae proteins to the target RNA exposes further contact surfaces on both RNA and protein; this is required for recruiting additional factors (core proteins) to the RNPs. The K-turn like motif of the SECIS RNA and the

nucleocytoplasmic shuttling ability of SBP2 are consistent with a possible nuclear assembly of SECIS mRNPs obeying to the same rules as the sn/snoRNP assembly. Indeed, a recent study in the laboratory identified a complex assembly machinery linked to the protein chaperone Hsp90 and that assembles RNPs of the L7Ae family, such as box C/D and H/ACA snoRNPs, telomerase, 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 (R2TP complex) associated to the Hsp90 chaperone, and the adaptor protein Nufip. Nufip binds the L7Ae family proteins including SBP2 and can tether them to other core proteins of immature RNPs. It also links them to the Hsp90 chaperone complex. These associations are necessary for the stability of SBP2 and its assembly on the SECIS RNA (Boulon et al, 2008). Altogether, these results implied 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) yet to be identified (Figure 14).

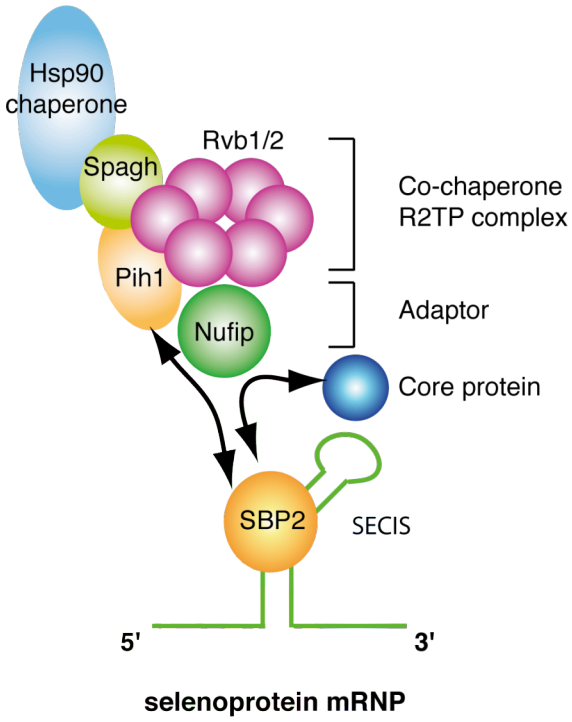


Figure 14. Selenoprotein mRNP assembly model.

The Hsp90 co-chaperone complex is composed of the AAA+ ATPases Rvb1 and Rvb2 (Rvb1/2), Spagh and Pih1. Hsp90 is required for the stability of SBP2. It may control the folding of SBP2 during SECIS RNP formation. Nufip acts as an adaptor between the chaperone complex and SBP2. The interaction between Pih1 and SBP2 has been experimentally demonstrated. The curved arrow points to interactions with putative core proteins yet to be discovered. The figure is taken from (Allmang et al, 2009).

4. Objectives and outline of this thesis

SBP2 plays a central role in the eukaryotic Sec incorporation mechanism. However, SBP2 was characterized only in rat and humans when I started my PhD work. We were therefore interested in SBP2 from other species for evolutionary and crystallographic purposes. During my PhD studies, I cloned the *Drosophila melanogaster* SBP2 and analyzed its function. We identified a new domain which is important for SECIS RNA binding and exhibits selective affinities toward SECIS RNAs. This new domain, found in all SBP2s, is also responsible for the binding of SBP2 to the ribosome. This work will be described in Chapter 1 of Part 2.

The laboratory recently identified a common molecular machinery for the assembly of selenoprotein mRNPs with sn/snoRNPs. Since most of the core proteins of selenoprotein mRNPs are still unknown, I investigated whether major core proteins of the sn/snoRNP complex could also be potential partners of SBP2. This wok will be described in Chapter 2 of Part 3

Finally, I also participated in the structural analysis of SBP2, in collaboration with the group of Philippe Dumas in the same UPR. This will be described in Chapter 2 of Part 2.

Altogether, these studies provided important insight into how the SBP2-SECIS RNA complex is formed during selenoprotein synthesis, a process which is at the heart of the UGA reprogramming mechanism.

Part 2. Results

1. Functional characterization of *Drosophila melanogaster* SBP2

1.1. Selenoproteome in Drosophila

Selenoproteins are found in the three domains of life, but not in all species of each taxa. Selenoproteins were even found to be encoded by two viruses (Mix et al, 2007; Shisler et al, 1998). About 25 Sec-containing proteins have been identified in mammals (Castellano et al, 2004; Kryukov et al, 2003), but the distribution among taxa varies greatly. For instance, no selenoprotein has been found in yeast and land plants, only one in nematodes and three in flies. A few selenoproteins have homologues in which Sec is replaced by cysteine (Cys). This is for example the case in genomes completely devoid of Sec-containing genes. The phylogenetic distribution of selenoproteins shows that Sec utilization in eukaryotes is sporadic and that the mammalian selenoproteome is not the recapitulation of that of eukaryotes (reviewed in Castellano, 2009, Figure 15). While vertebrates have a large number of selenoprotein genes, non-vertebrate species can possess even larger selenoproteomes. Interestingly, aquatic organisms have larger selenoproteomes compared to terrestrial organisms, likely due to the higher availability of dissolved organic selenium in oceans, leading to the occurrence of large selenoproteomes in aquatic organisms. This is the case for example of the algae Ostreococcus which contains about 30 selenoprotein genes (Table 2). On the contrary, the higher content of oxygen in the air, which may make selenoproteins more susceptible to oxidative damage due to the side reaction of the selenol group of Sec with O_2, led to the reduction of selenoproteins in terrestrial organisms (Lobanov et al, 2007; Lobanov et al, 2008b). In invertebrates, three selenoproteins were found in *Drosophila melanogaster* and a single one in *Caenorhabditis elegans* (Castellano et al, 2001; Taskov et al, 2005). The only selenoprotein in *C. elegans* is thioredoxin reductase. Protozoa also possess selenoproteins and often have specific selenoproteins. Trypanosoma and Leishmania have three selenoproteins, that are distant homologs of mammalian SelK and SelT, and a kinetoplastida-specific selenoprotein (SelTryp) (Lobanov et al, 2006b). Plasmodia have four selenoproteins that are not homologues of previously known selenoproteins (Lobanov et al, 2006a). Although showing some specificities that set it slightly apart from canonical SECISes, the Plasmodium SECIS element is also able to support selenoprotein synthesis in mammalian cells (Lobanov et al, 2006a). These findings support the conservation of the Sec incorporation mechanisms from selenoprotein-containing lower eukaryotes to mammals. We

were therefore interested in the selenoprotein synthesis machinery in other species than mammals where it had not been investigated at all, especially SBP2 in *D. melanogaster* because of its shorter length that will be described in the following chapter (1.2. Objective).

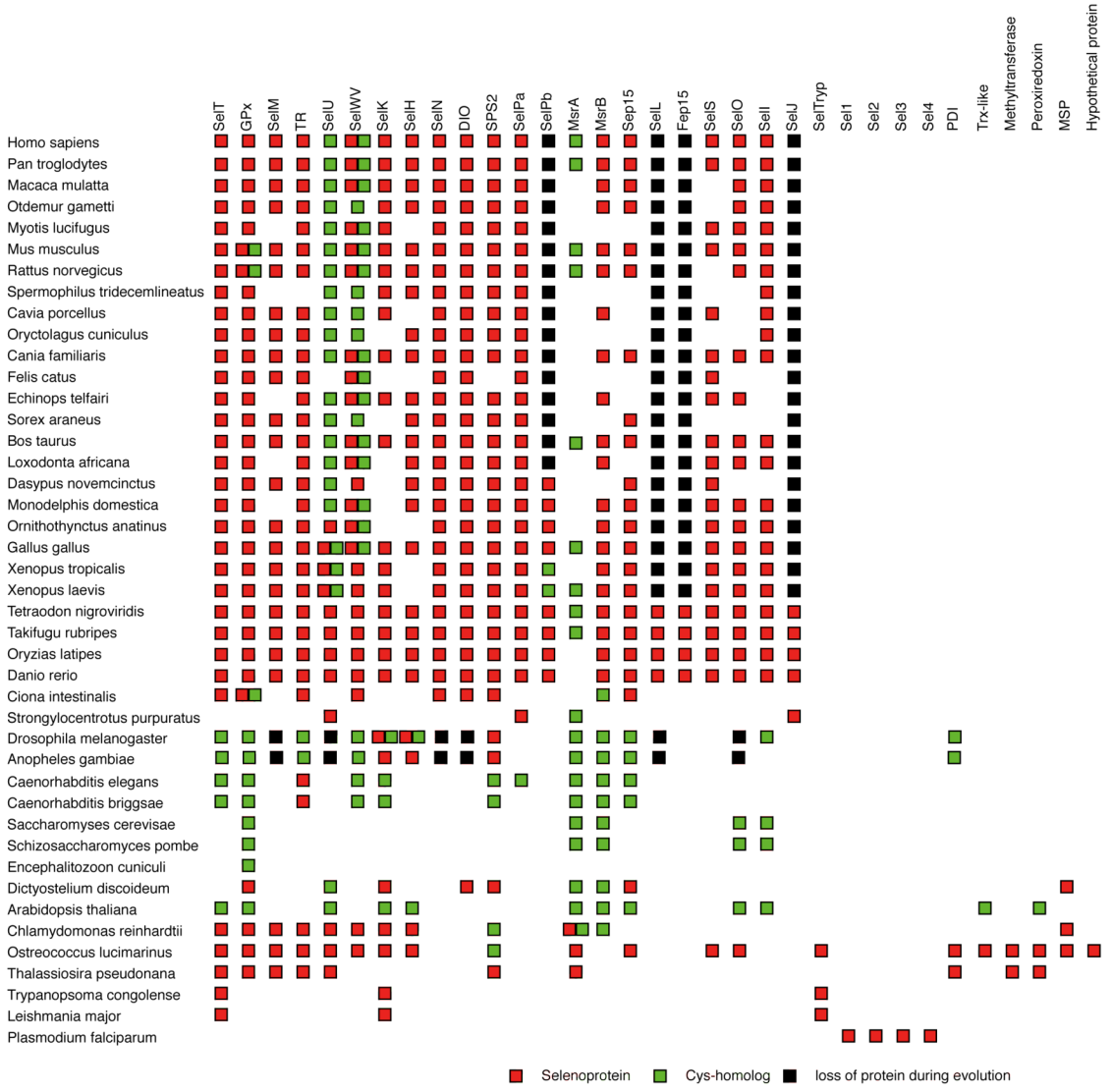


Figure 15. Eukaryotic selenoproteomes

The eukaryotic selenoproteins, as identified in (Castellano et al, 2008; Castellano et al, 2005; Castellano et al, 2001; Castellano et al, 2004; Chapple & Guigo, 2008; Dayer et al, 2008; Kryukov et al, 2003; Lobanov et al, 2006a; Lobanov et al, 2007; Lobanov et al, 2006b; Lobanov et al, 2008a; Lobanov et al, 2008b; Taskov et al, 2005). Red, green and black boxes indicate selenoproteins, cysteine homologs and proteins lost during evolution, respectively.

GPx: Glutathione peroxidases, TR: Thioredoxin reductases, DIO: Iodothyronine deiodinases, SPS2: Selenophosphate synthetase, PDI: Protein disulfide isomerase, MSP: Predicted membrane selenoprotein. Msr: methionine sulfoxide reductase.

	SelT	GPx	SelM	TR	SelU	SelWV	SelK	SelH	SelN	DIO	SPS2	SelP	MsrA	MsrB	Sep15	SelL	Fep15	SelS	SelO	SelI	SelU	PDI	Trx-like	Methyltransferase	Peroxiredoxin	MSP	Hypothetical protein	Total
H.sapiens	1	5	1	3		2	1	1	1	3	1	1		1	1			1	1	1								25
M.musculus	1	4	1	3		2	1	1	1	3	1	1		1	1			1	1	1								24
R.norvegicus	1	4	1	3		2	1	1	1	2	1	1		1	1				1	1								22
G.gallus	1	3	1	3	1	1	1	1	1	3	1	2		1	1			1	1	1								24
X.laevis	2	4	1	2	1	2	1		1	3	1	1		1	1			1	1	1								24
D.rerio	3	4	1	2	3	3	1	1	1	3	1	2		2	1	1	1	1	1	1	1							34
D.melanogaster							1	1			1																	3
C.elegans				1																								1
D.discoideum							1			1	1				1											1		5
C.reinhardtii	1	2	1	1	1	2	1	1					1													1		12
O.lucimarinus	1	5	1	1	1	2	1	1					1		1			1	1			3	1	1	3	1	3	29
T.pseudonana	1	2	2	1	2						1		1									2	1	1	2			16

Table 2. Selenoproteins identified in representative eukaryotic organisms.

Data taken from (Lobanov et al, 2007; Lobanov et al, 2008b). GPx: Glutathione peroxidase, TR: Thioredoxin reductase, DIO: Iodothyronine deiodinase, SPS2: Selenophosphate synthetase, Msr: methionine sulfoxide reductase, PDI: Protein disulfide isomerase, MSP: Predicted membrane selenoprotein.

1.2. Objective

The three selenoproteins in *D. melanogaster* are dSPS2, dSelK and dSelH (Castellano et al, 2001; Hirosawa-Takamori et al, 2000) (dSelK and dSelH were initially identified as dSelG and dSelM, respectively, but renamed later (Kryukov et al, 2003)). dSelK has one cysteine homolog and dSelM has two (Castellano et al, 2001; Martin-Romero et al, 2001). While dSPS2, the selenophosphate synthetase, belongs to a known family of selenoproteins, both dSelK and dSelH are poorly characterized functionally but are likely involved in antioxidative reactions (Dikiy et al, 2007; Lu et al, 2006; Morozova et al, 2003). dSelK has a cysteine homolog in *C. elegans* of unknown function (Taskov et al, 2005), while dSelH appears to belong to a class of selenoproteins widely distributed across the phylogenetic spectrum, as dSelH was found in zebrafish, human and mouse EST databases (Dikiy et al, 2007; Novoselov et al, 2007a). dSPS2, dSelK and dSelH all contain mammalian-type SECIS elements, composed of the characteristic helix and loop structures as well as the non-Watson-Crick quartet. Strikingly, only form 2 SECIS were found in the 3'UTR of their mRNAs (Figure 16).

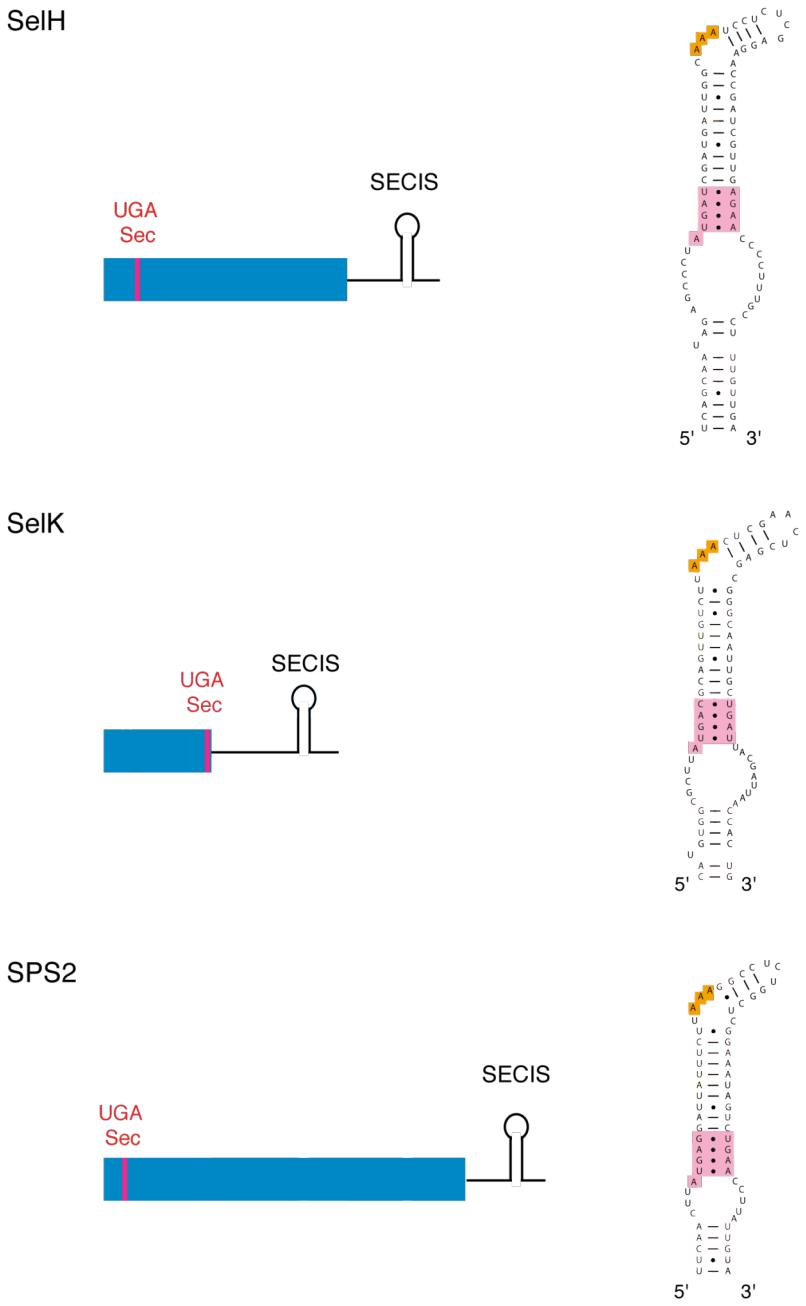


Figure 16. Selenoproteins in *D. melanogaster* SBP2.

Schematic representations of SelH, SelK and SPS2 mRNAs (left) and secondary structure of the corresponding SECIS elements (right) described in (Castellano et al, 2001; Fagegaltier et al, 2000b; Hirosawa-Takamori et al, 2000). The coding regions are shown by blue boxes with the UGA Sec codon in red. The non-Watson-Crick quartet and the conserved As in the internal loop II are shown in pink and orange, respectively (right).

Although a number of SBP2 sequences from mammals, non-mammalian vertebrates, invertebrates or unicellular organisms were annotated in databases, SBP2 had been so far isolated and functionally characterized only in rat and humans.

Interestingly, a database search for Drosophila SBP2 sequences yielded a candidate bearing all the signatures to be a *bona fide* SBP2 (our work, and Chapple & Guigo, 2008). Compared to mammalian SBP2 that is about 850 amino acid long, Drosophila SBP2 lacks homology to the mammalian N-terminal domain and is only 313 amino acid long (Figure 17).

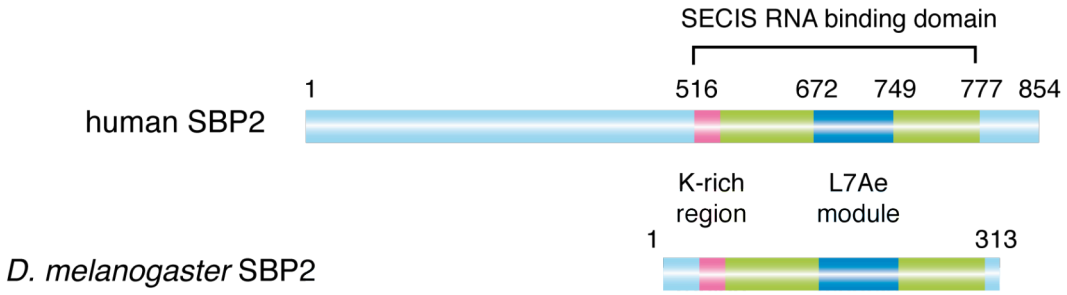


Figure 17. Shematic representation of the human and *D. melanogaster* SBP2 proteins.

The K-rich region and the L7Ae module are shown in pink and blue, respectively. The *D. melanogaster* SBP2 lacks the sequence corresponding to the N-terminal domain of human SBP2. The C-terminal domain of *D. melanogaster* SBP2 contains the conserved RNA binding domain.

As mentioned in part 1 (Introduction), the N-terminal domain of mammalian SBP2 is dispensable for selenoprotein synthesis *in vitro*. Furthermore, the existence of selenoproteins in Drosophila implied that Drosophila SBP2 can function in selenoprotein synthesis. Thus, we postulated that this shorter SBP2 would be a good model to study SBP2 function. In the course of my work, I cloned the cDNA of the *Drosophila melanogaster* SBP2 protein (dSBP2) and tested whether this shorter SBP2 is functional in selenoprotein synthesis. I compared it to mammalian SBP2. This allowed the identification and characterization of a novel K (lysine)-rich domain in the *Drosophila* and also all SBP2s, essential for SECIS and 60S ribosomal subunit binding. This additional, novel domain differs from the well-characterized L7Ae RNA-binding domain. We have analyzed its specificities. Altogether, this work is described in the following Article 1.

1.3. Summary of Article 1

In this work, I have isolated and functionally characterized a *bona fide D. melanogaster* SBP2 that specifically lacks the region homologous to the N-terminus of vertebrate SBP2. While human SBP2 (hSBP2) binds both form 1 and 2 SECIS RNAs with similar affinities, dSBP2 exhibits high affinity toward form 2 only. In addition, with a homology search, we identified a lysine-rich (K-rich) domain in human SBP2 that is essential for SECIS binding. Both domains differ by their lysine content. They read SVRVY in Drosophila and IILKE in humans. We showed, by introducing point mutations into hSBP2, that the differential binding affinities to SECIS RNAs are attributed to a short amino acid sequence in the K-rich domain. Exchanging the sequence of dSBP2 (SVRVY) with that of hSBP2 (IILKE) enabled dSBP2 to bind form 1 SECIS RNAs but impaired binding of hSBP2. We therefore concluded that this five amino acid sequence (IILKE) in the hSBP2 K-rich domain confers the ability to bind form 1 SECIS RNAs.

It was previously shown that SBP2 binds the 80S ribosome (Copeland & Driscoll, 2001). In this study, we showed that SBP2 binds in fact to the 60S subunit and that the K-rich domain is essential for binding. This is consistent with the proposal that this association with the ribosome may occur through binding to one of the K-turn motifs in the 28S ribosomal RNA (Caban et al, 2007; Copeland & Driscoll, 2001; Kinzy et al, 2005). Furthermore, our finding that the same amino acids in the K-rich region that are crucial to SECIS recognition are also involved in ribosomal binding supports the notion that SBP2 cannot bind the SECIS element and the ribosome simultaneously (Copeland & Driscoll, 2001).

Although SBP2 belongs to the L7Ae family, it differs from other L7Ae family proteins. While the L7Ae family requires the single L7Ae domain to bind specifically the cognate RNA (Chao & Williamson, 2004; Marmier-Gourrier et al, 2003; Moore et al, 2004; Nottrott et al, 2002; Vidovic et al, 2000), SBP2 needs both the L7Ae and the K-rich domain for SECIS specific binding. These findings were in fact already suspected from the following experiments. While the 15.5kD and L7Ae ribosomal proteins have a rather broad specificity and can recognize the SECIS RNA, SBP2 displayed a higher selectivity since it could not bind either the U3 snoRNA or the archaea sRNA (Cléry et al, 2007). This specificity is also reflected by the presence of nucleotide determinants in the SECIS RNA that are unique to SBP2 recognition (See Figure 10). Indeed, helix I and internal loop 1 which are required for SBP2 recognition, are found only in the SECIS RNAs and not in other K-turn RNAs (Cléry et al, 2007).

These observations led us to propose the following model rationalizing the necessity of two domains in SBP2 for complex formation with the SECIS RNA. The L7Ae motif recognizes the guanine bases of the G.A/A.G base pairs in a similar manner to other L7Ae proteins, and the conserved U in the SECIS non-Watson-Crick quartet. The positively charged lysines in the K-rich domain likely increase the affinity of SBP2 for SECIS RNAs through electrostatic interactions with the phosphates of the SECIS-specific structural features. Indeed, earlier experiments showed that SBP2 required the phosphates of helix 1 for binding (Fletcher et al, 2001). In the K-rich domain, IILKE sequence could mediate form 1 SECIS recognition by either a direct effect on SECIS binding or indirectly by inducing a conformational change of the L7Ae motif. Validation of these models has to await resolution of the crystal structure of the SBP2-SECIS RNA complex.

1.4. Article 1

A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins,

Akiko Takeuchi, David Schmitt, Charles Chapple, Elena Babaylova, Galina, Karpova, Roderic Guigo, Alain Krol and Christine Allmang (2009) *Nucleic Acids Research, 37(7):2126-41*.

Nucleic Acids Research Advance Access published February 17, 2009

*Nucleic Acids Research, 2009, **1–16***
doi:10.1093/nar/gkp078

A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins

Akiko Takeuchi[1], David Schmitt[1], Charles Chapple[2], Elena Babaylova[1,3], Galina Karpova[3], Roderic Guigo[2], Alain Krol[1,*] and Christine Allmang[1]

[1]Architecture et Réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC, 15 Rue René Descartes, 67084 Strasbourg, France, [2]Centre de Regulacio Genomica, Genome Bioinformatics Research Group, Parc de Recerca Biomedica de Barcelona, c/Dr. Aiguader 88, 08003 Barcelona, Spain and [3]Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, Prospekt Lavrentjeva 8, Novosibirsk, 630090, Russia

Received December 17, 2008; Revised and Accepted January 27, 2009

ABSTRACT

Selenoproteins contain the amino acid selenocysteine which is encoded by a UGA Sec codon. Recoding UGA Sec requires a complex mechanism, comprising the *cis*-acting SECIS RNA hairpin in the 3′UTR of selenoprotein mRNAs, and *trans*-acting factors. Among these, the SECIS Binding Protein 2 (SBP2) is central to the mechanism. SBP2 has been so far functionally characterized only in rats and humans. In this work, we report the characterization of the *Drosophila melanogaster* SBP2 (dSBP2). Despite its shorter length, it retained the same selenoprotein synthesis-promoting capabilities as the mammalian counterpart. However, a major difference resides in the SECIS recognition pattern: while human SBP2 (hSBP2) binds the distinct form 1 and 2 SECIS RNAs with similar affinities, dSBP2 exhibits high affinity toward form 2 only. In addition, we report the identification of a K (lysine)-rich domain in all SBP2s, essential for SECIS and 60S ribosomal subunit binding, differing from the well-characterized L7Ae RNA-binding domain. Swapping only five amino acids between dSBP2 and hSBP2 in the K-rich domain conferred reversed SECIS-binding properties to the proteins, thus unveiling an important sequence for form 1 binding.

INTRODUCTION

Selenoproteins are a diverse family of proteins characterized by the presence of the 21st amino acid selenocysteine (Sec). This amino acid is co-translationally incorporated into the growing peptide chain in response to a UGA Sec codon, otherwise read as a signal for termination of translation. In eukaryotes, the correct recoding event of UGA stop to UGA Sec relies on specific, conserved RNA structures and proteins. The $tRNA^{Sec}$ and the SECIS element, an RNA hairpin in the 3′UTR of selenoprotein mRNAs, and two *trans*-acting proteins, the specialized translation elongation factor eEFSec and the SECIS Binding Protein 2 (SBP2), are the key players of the recoding machinery (1). Specialized protein complexes that involve SECp43, the Phosphoserine $tRNA^{Sec}$ Kinase (PSTK) and the Sec synthase are recruited to the $tRNA^{Sec}$ to ensure proper selenocysteine synthesis (2–4). Ribosomal protein L30 has also been implicated in this mechanism and shown to compete with SBP2 for SECIS binding (5).

There are two types of functional SECIS RNAs, forms 1 and 2, classified according to their different apex: form 2 has an additional helix, and its apical loop is shorter than in form 1 (6,7). Structure-based alignments in the currently available eukaryotic selenoproteome identified form 2 SECIS as the most widespread element (8). Except for the apex, SECIS RNA hairpins share common structural features, in particular four consecutive non-Watson–Crick base pairs (the quartet) composed of a central tandem of sheared G.A/A.G base pairs (7–10).

*To whom correspondence should be addressed. Tel: +33 3 88 41 70 50; Fax: +33 3 88 60 22 18; Email: a.krol@ibmc.u-strasbg.fr
Present address: David Schmitt, Novartis, Basel, Switzerland.

Such non-canonical base pairs are characteristic of K (kink)-turn motifs which are recurrent in a variety of RNAs (11–13). The SECIS RNA has been proposed to contain a K-turn like motif (14) that is essential for SBP2 interaction and selenoprotein incorporation *in vivo* (10). A number of proteins fulfilling different functions such as snRNPs, snoRNPs or ribosomal proteins bind K-turn RNA motifs (15,16). They all carry the L7Ae RNA-binding domain (or module) (17) that contains a restricted set of amino acids that establish base-specific contacts with the sheared G.A/A.G base pairs (11,18,19).

SBP2 also has the L7Ae module in its RNA-binding domain (20–22). In an earlier work, we predicted the human SBP2 (hSBP2) amino acids that contact the SECIS RNA at the K-turn like motif (20). However, while sharing some RNA-binding properties with other proteins of the L7Ae family, SBP2 possesses its own specificities (23). The known functions of SBP2, comprising SECIS and ribosome binding, and Sec incorporation, reside in the C-terminal two-thirds of the protein (21,22,24). However, no function has been attributed to the remaining N-terminal section which has been shown to be dispensable for Sec incorporation in rabbit reticulocyte lysate (25).

Selenoproteins exist in the three domains of life. Vertebrate genomes encode up to 25–26 selenoproteins but surprisingly larger selenoproteomes can be found in aquatic unicellular organisms (26). Only three selenoprotein genes have been discovered in *Drosophila melanogaster*, SPS2, SelH and SelK (27,28). SPS2 is the selenophosphate synthetase involved in selenocysteine biosynthesis. SelH and SelK are poorly characterized functionally but they seem nevertheless to play an antioxidant role (29,30). In each case however, only form 2 SECIS RNAs were found in the 3′UTR of the selenoprotein mRNAs.

Some of us have recently published the annotation and multiple sequence alignments of insect selenoprotein synthesis factors, especially in 12 *Drosophila* genomes (31). Among these factors, our attention was attracted by the putative *Drosophila* SBP2 because they lack the sequence homologous to the N-terminus of hSBP2. Although a number of SBP2 sequences from mammals, non-mammalian vertebrates or even unicellular organisms are annotated in databases, only the rat and hSBP2 have been so far isolated and functionally characterized (20–25,32,33). This prompted us to study *Drosophila* SBP2s and in particular *Drosophila melanogaster*. In this work, we report that despite its shorter length, *D. melanogaster* SBP2 (dSBP2) retains functional properties similar to its mammalian counterpart. However, dSBP2 exhibits selective affinities toward SECIS RNAs, being almost unable to bind form 1 SECIS. We determined that the discriminating amino acids reside in a K (lysine)-rich region that we also identified in hSBP2 as essential for SECIS RNA binding. In addition we showed that, in hSBP2, mutating the K-rich region affected form 1 and form 2 SECIS interaction differently, and that this region also plays a crucial role in 60S ribosomal subunit binding.

MATERIALS AND METHODS

Strains and growth conditions

The *Escherichia coli* TG2 strain was used as the host strain for plasmid construction. Growth was performed at 37°C in LB medium, complemented with 100 μg/ml ampicillin. The *E. coli* strain BL21(DE3)-star was used for production of *Drosophila* SBP2 proteins at 25°C in ZYM-5052 auto-induction medium as described by Studier *et al*. (34). The *E. coli* strain BL21(DE3)RIL (Novagen) was used for production of hSBP2 proteins at 18°C in LB medium.

Bioinformatic analyses

Alignment of human/pig/rat/insect SBP2s. Annotated SBP2 sequences from human (gb|AAK57518.1|AF38 0995), rat (sp|Q9QX72.1|SEBP2_RAT), pig (ef|XP_0019 28402.1) and chicken (ref|XP_424425.2|) were aligned against the putative SBP2 sequences found in three *Drosophila* species, *D. melanogaster*, *D. pseudoobscura* and *D. sechelia* (31) by Kalign (35).

Alignment of the K-rich domain. All annotated members of the SBP2 family in Ensembl (ENSF00000007674) were extracted and all those with no 'X' in the relevant region were kept. The search was extended by blasting the *D. melanogaster* SBP2 against the nr database of NCBI. Of the resulting hits, only those containing the IHSRRF motif (positions 624–629 in hSBP2) characteristic of SBP2 proteins (C.A and A.K., unpublished data) were kept. We subsequently used the L7Ae RNA-binding module of SBP2 (33) to query the nr database using Hmmer (36). Finally, we also added the insect SBP2 sequences (31). The resulting 40 sequences were aligned using mafft (37). The alignment images shown in Figures 1 and 7 were produced by Jalview (38).

cDNA cloning and recombinant DNA

Drosophila melanogaster SBP2 ORF (Genebank accession # AI062219) was amplified from pOT2 cDNA clone GH01354 (Research Genetics) in a two-step PCR reaction and introduced into the pHMGWA vector (39) using the Gateway Technology (Invitrogen). The resulting pHMdSBP2 vector contains a 6× His-tag and Maltose-Binding Protein coding-frame upstream of the dSBP2 ORF and allows *E. coli* expression of the protein. Human SBP2 ORF was amplified by PCR from plasmid pA11 (33) and subsequently cloned into pET32b (Novagen), generating phSBP2-FL (full-length), as well as plasmids encoding the N-terminal truncated proteins phSBP2 344–854, phSBP2 399–854, phSBP2 515–854, phSBP2 525–854, phSBP2 545–854, phSBP2 625–854 and phSBP2 674–854, and C-terminal truncated proteins phSBP2 344–674, phSBP2 344–790 and phSBP2 344–820. Alanine scanning mutants in hSBP2 were generated in phSBP2 344–854 using the Kunkel mutagenesis method (40). Amino acids swapping mutants exchanging hSBP2 aa535–539 for dSBP2 aa95–99 (phSBP2-SVRVY) and dSBP2 aa95–99 for hSBP2 aa535–539 (pdSBP2-IILKE), were generated by site-directed mutagenesis of phSBP2 344–854 and pHMdSBP2, respectively,

using the QuickChange XL Site-Directed Mutagenesis kit (Stratagene) according to the manufacturer's instructions.

Plasmids pT7SelN, pT7GPx1 and pT7PHGPx were used for T7 transcription of human SelN, rat GPx1 and PHGPx SECIS RNAs, respectively (7,9). To allow *in vitro* transcription of *Drosophila* SECIS RNAs, *D. melanogaster* SelK and SelH SECIS elements were PCR amplified from pOT2 cDNA clones GH03581 and SD09114, respectively (generously provided by M. Corominas, University of Barcelona) and introduced into the BclI-EcoRI sites of pT7-Bck vector (9) to create pT7dSelK and pT7dSelH. A point mutant in dSelH SECIS (dSelHmut), and the SECIS RNA apex-swapped mutants PHGPx-ApSelN and SelN-ApPHGPx SECIS RNAs, were generated by site-directed mutagenesis of pT7dSelH, pT7PHGPx and pT7SelN, respectively, using the QuickChange XL Site-Directed Mutagenesis kit.

To generate selenoprotein mRNA reporter constructs for *in vitro* translation assays, *D. melanogaster* SelH ORF and 3′UTR (Genebank accession #AI542675) were amplified from pOT2 cDNA clone SD09114 and cloned into the HindIII-KpnI sites of the pXJ(HA)3 eukaryotic expression vector (41) to create pHAdSelH. Rat selenoprotein reporter constructs pGPx1-GPx1SECIS and pGPx1-ΔSECIS were as described in (10). To create pGPx1-PHGPxSECIS and pGPx1-SelNSECIS, the GPx1 SECIS element of the pGPx1-GPx1SECIS reporter construct was exchanged for PHGPx and SelN SECIS elements from pT7PHGPx and pT7SelN, respectively, using BclI-EcoRI sites.

Oligonucleotides used for PCR and mutagenesis are listed in Supplementary Data.

Recombinant protein preparation

Drosophila SBP2 recombinant proteins expressed in *E. coli* were purified using Amylose Resin column (NEB). 6× His and MBP tags were cleaved from the dSBP2 protein by thrombin (Sigma) when used for electrophoretic mobility shift assays. hSBP2 recombinant proteins expressed in *E. coli* were purified using Ni-NTA agarose (Qiagen). Elution buffer was exchanged to dialysis buffer containing 20 mM Tris–HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 20% glycerol, 1 mM DTT and Cocktail inhibitor (Sigma).

Electrophoretic mobility shift assay

Plasmids yielding PHGPx, SelN, GPx1, dSelK, dSelH, dSelHmut, PHGPx-ApSelN and SelN-ApPHGPx SECIS RNAs were linearized by EcoRI. Internally labeled SECIS RNAs were obtained by T7 transcription using 80 μCi of [α-^{32}P]-ATP (3000 Ci/mmol). SECIS RNA-SBP2 complexes were formed as described in (20). Routinely, 30 000 cpm of ^{32}P-labeled SECIS RNA were incubated for 30 min at 30°C with various concentrations of purified SBP2 protein (from 0 to 2000 nM), in 7.5 μl of phosphate buffer saline pH 7.4, 2 mM DTT. RNA–protein complexes were separated on 6% non-denaturing polyacrylamide gel electrophoresis in 0.5× TBE, 5% glycerol buffer. The intensities of free and bound RNAs were quantitated with the Fujifilm FLA-5100 Imaging system. Kds were determined from three independent experiments.

In vitro selenoprotein synthesis assays

In vitro translation of *Drosophila* (dSelH) or rat GPx1 from selenoprotein encoding plasmids carrying (or lacking) wild-type SECIS elements were performed using TNT Coupled Reticulocyte Lysate Systems (Promega). One microgram of each of the selenoprotein plasmid DNA was used as the template in 50 μl *in vitro* transcription/translation reactions in the presence of 25 μl rabbit reticulocyte lysate, 20 μCi of ^{35}S-methionine and 120–240 nM of purified SBP2 protein. *In vitro* translated HA-tagged dSelH and hGPx1 proteins were purified using microMACS Epitope Tag Protein Isolation Kits (Miltenyi Biotec), resolved by 10% SDS–PAGE and detected with the Fujifilm FLA-5100 Imaging system. Quantification of selenoprotein synthesis was performed from two to three independent experiments.

Ribosome-binding assays

60S and 40S ribosomal subunits were isolated from full term human placenta according to ref. (42). Human recombinant ribosomal protein p40 was a kind gift of Dr Alexey Malygin (ICBFM, Novosibirsk, Russia). Monoclonal antibodies against human p40 were provided by Dr Valery Loktev. Binding mixtures (50 μl) containing 30 pmol of 60S or 40S subunits were reactivated at 37°C for 10 min in PBSD buffer (150 mM NaCl, 27 mM KCl, 8 mM Na_2HPO_4, 1.7 mM KH_2PO_4 and 2 mM DTT) containing 0.5 mM $MgCl_2$. Then 3.5 μg SBP2 (or SBP2 mutants) or 2 μg ribosomal protein p40 were added and incubated at 22°C for 20 min. The mixtures were loaded onto a 15–30% linear sucrose gradient in PBSD with 0.5 mM Mg^{2+} and centrifuged in a SW41 rotor at 23 000 rpm for 15 h. Fractions corresponding to 60S and 40S subunits were precipitated by 10% trichloroacetic acid, and the pellet content loaded onto 10% SDS–PAGE which was blotted onto nitrocellulose membranes. SBP2 was detected with an anti-SBP2 polyclonal antibody (1/5000–1/10 000 dilution) in PBST (1× PBS containing 0.1% Tween 20, 3% dry milk), p40 with a monoclonal anti-p40 (1/3000 dilution). Membranes were treated with anti-rabbit HRP-conjugated secondary antibody (1/10 000 dilution), revealed with the ECL plus kit (GE HealthCare) and exposed to either X-ray film or ChemDoc XRS.

RESULTS

Identification and functional characterization of dSBP2

Recent comparative analysis of insect genomes have identified putative SBP2 proteins (31). In this work we set out to clone and characterize dSBP2 and compare it to hSBP2. tBlastn searches were performed at NCBI using the hSBP2 amino acid sequence (33) and one significant hit (AI062219) was obtained with a *D. melanogaster* sequence predicted to encode a 313 amino-acid protein. Figure 1 shows a multiple sequence alignment of this protein as well as other putative SBP2 from *D. pseudoobscura*

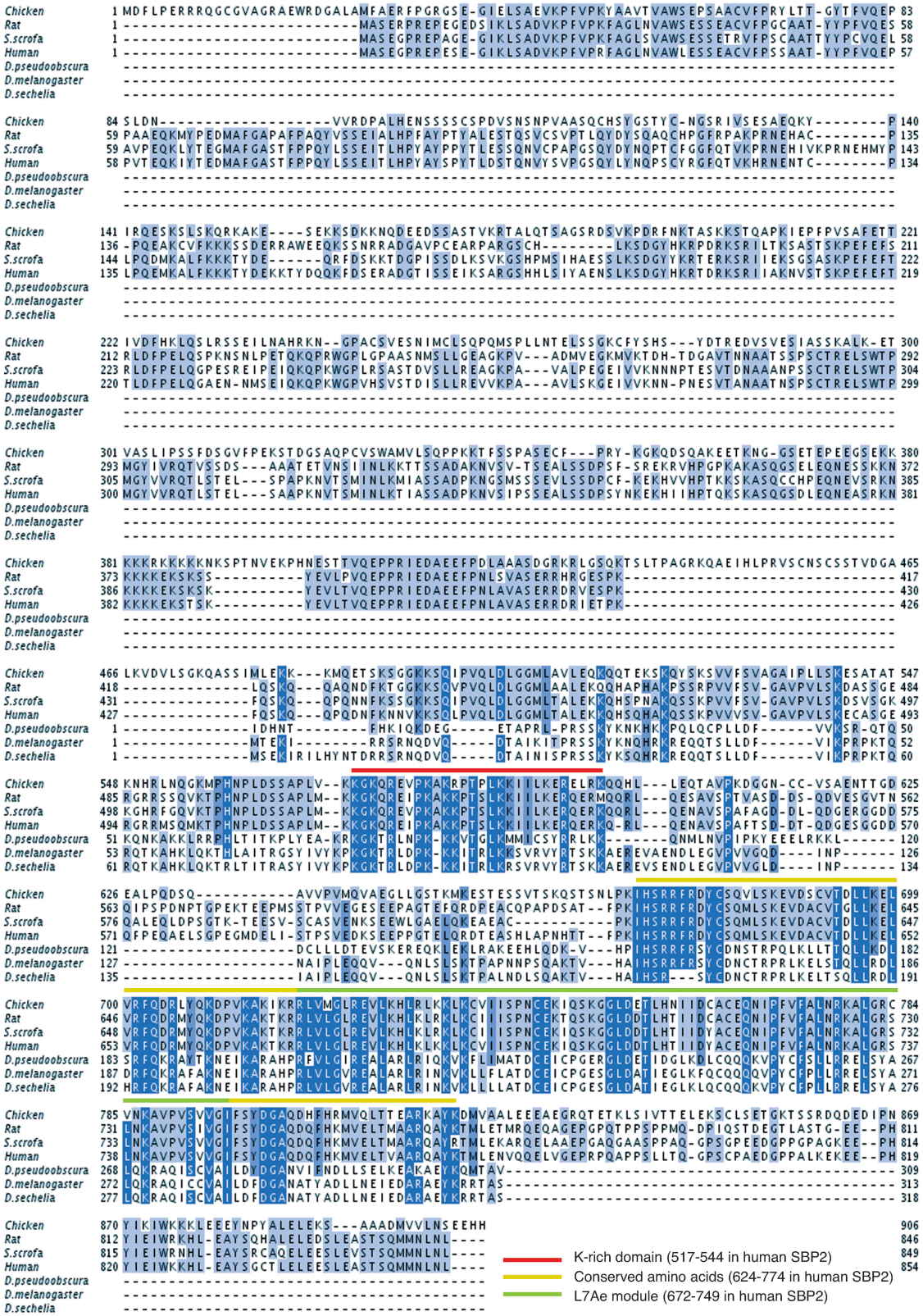


Figure 1. Alignment of SBP2 proteins. *Drosophila* (*pseudoobscura*, *melanogaster* and *sechelia*) SBP2 sequences were from (31). Accession numbers for vertebrate SBP2 are XP_424425.2| (chicken), Q9QX72.1 (rat), XP_001928402.1 (*Sus scrofa*), AF380995 (human). The color scheme above the sequences is explicited at the bottom of the figure.

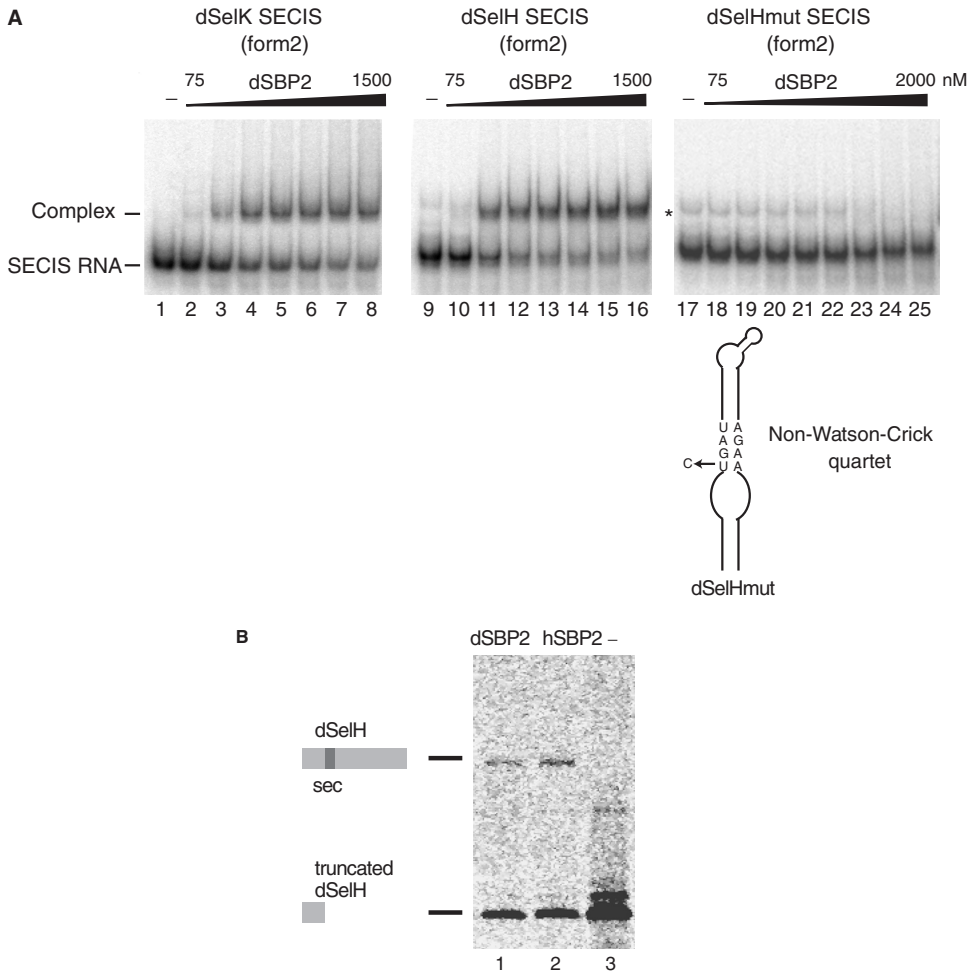


Figure 2. Functional characterization of dSBP2. (**A**) Electrophoretic mobility shift assays were performed between the purified dSBP2 and the *in vitro* transcribed ^{32}P-labeled form 2 SECIS RNAs (dSelK SECIS and dSelH SECIS) from *Drosophila* SelK and SelH selenoprotein mRNAs. Increasing concentrations (75–2000 nM) of dSBP2 were added. A dSeH mutant SECIS RNA (dSelHmut), carrying a U to C mutation in the non-Watson-Crick quartet, abolished the binding of dSBP2. dSBP2 was omitted in the control lanes 1, 9 and 17 (−). The asterisk denotes the position of a band that is also present in the control lanes 9 and 17. It often occurs and presumably contains an RNA conformer arising from T7 transcription. (**B**) dSBP2 can support selenoprotein H (dSelH) synthesis in rabbit reticulocyte lysate. Synthesis of full-length (^{35}S)-Met-labeled selenoprotein dSeH was obtained by adding purified dSBP2 (lane 1) or hSBP2 (lane 2) to the lysate. Lane 3: absence of SBP2 led to premature termination of translation (Truncated). Translation products were immunopurified and resolved by SDS–PAGE.

and *D. sechelia* (31) with the characterized rat and human SBP2 (25,33) and the annotated chicken and *Sus scrofa* SBP2s. Strikingly, the three putative *Drosophila* SBP2s lack homology to the vertebrate SBP2 N-terminal region. Homologies to vertebrate SBP2s were found throughout the rest of the sequence, starting from hSBP2 residue 427. Two blocks of high sequence identity were detected, one extending from K517 to K544, the other from I624 to K774 (hSBP2 numbering), with 50% and 36% identity, respectively. The latter block contains the previously identified L7Ae RNA-binding module spanning residues R672 to I749 (21,22).

To verify that the *D. melanogaster* cDNA encodes a bona fide SBP2, it was cloned and expressed in *E. coli* and assayed for SECIS RNA binding. Gel shift assays were performed with the form 2 SECIS RNAs of the dSelK and SelH mRNAs (dSelK and dSelH SECIS RNAs), showing formation of dSBP2–dSECIS RNA complexes (Figure 2A, lanes 2–8 and 10–16). Kd values were 260 nM and 170 nM for dSelK SECIS and dSelH SECIS RNAs, respectively, indicating a slightly higher affinity for dSelH SECIS RNA. To determine whether the binding was specific, we used a mutant SECIS RNA (dSelHmut SECIS, Figure 2A) in which the conserved U in the non-Watson–Crick quartet was replaced by a C, a mutation that impedes hSBP2 binding (43). Figure 2A (lanes 18–25) showed that no retarded complex was obtained, the band marked by an asterisk containing an RNA conformer also present in the control lane 17 (see also lane 9) but not an RNA–protein complex. We therefore concluded that dSBP2 bound specifically to the cognate SECIS RNAs.

We next investigated whether dSBP2 could support selenoprotein synthesis. To this end, rabbit reticulocyte lysate was used to synthesize *D. melanogaster* selenoprotein SelH (dSelH) from a reporter construct. The advantage of such

an extract is that it contains all the components required for selenoprotein mRNA translation but lacks SBP2 which must be added to obtain a full-length selenoprotein (25). Figure 2B (lane 3) showed that omission of SBP2 resulted in the synthesis of a truncated dSelH selenoprotein arising from premature termination of translation at the UGA Sec codon. Full-length dSelH could be obtained only upon addition of dSBP2 (Figure 2B, lane 1), in much the same way as hSBP2 (hSBP2; Figure 2B, lane 2). The truncated form of dSelH ending at the UGA Sec codon (lanes 1 and 2) very likely originated from saturation of the selenoprotein synthesis machinery in the reticulocyte lysate.

Overall, our data established that we have identified and functionally characterized the dSBP2 protein which possesses the same selenoprotein synthesis capacities as the human counterpart.

dSBP2 manifests distinctive SECIS RNA-binding activities

The abilities of dSBP2 to bind various SECIS RNAs were further investigated and compared to hSBP2. Form 2 SECIS, the only type of SECIS found in *Drosophila* (31), were tested first. Figure 3 and Table 1 show that complexes were obtained between the SECIS of the rat PHGPx (phosphohydroxylipid glutathione peroxidase) mRNA and dSBP2 (Figure 3A, lanes 2–7; K_d = 220 nM) or hSBP2 (Figure 3A, lanes 23–29; K_d = 320 nM). Additionally, hSBP2 also bound dSelH and dSelK SECIS RNAs (data not shown). Major differences appeared when form 1 SECIS was tested (a schematic drawing of SECIS forms 1 and 2 is shown in the inset of Figure 4). While hSBP2, as expected, bound readily the SECISes of the rat GPx1 (glutathione peroxidase 1) and human SelN (selenoprotein N) mRNAs (Figure 3A, lanes 31–37 and 39–45), dSBP2 only weakly recognized the SelN SECIS and was unable to bind the GPx1 SECIS at all (Figure 3A, lanes 16–21 and 9–14). These findings were corroborated by selenoprotein synthesis experiments in reticulocyte lysate in which dSBP2 enabled synthesis of a selenoprotein reporter (rat GPx1) from an mRNA carrying a form 2 PHGPx but not a form 1 GPx1 SECIS in the 3′UTR (Figure 3B, compare lanes 2, 3 with lanes 7, 8). Synthesis from the form 1 SelN SECIS-containing reporter was weak (Figure 3B, lanes 12, 13), in keeping with the low affinity to SelN SECIS. In contrast, hSBP2 led to synthesis of GPx1 regardless of the type of SECIS harbored by the reporter mRNA (Figure 3B, lanes 4, 5, 9, 10, 14, 15). Control experiments with a GPx1-SECIS lacking construct (Figure 3B, lanes 17–20) indicated that selenoprotein synthesis was indeed SECIS-dependent.

This unexpected result raised the possibility that the distinctive binding affinities of dSBP2 for SECIS RNAs could originate from the different apical secondary structures classifying form 1 and form 2 SECIS RNAs [Figure 4; (6,7)]. To answer the question, we decided to swap the apexes between both forms of SECIS RNAs to yield chimeric SECIS RNAs: a PHGPx-ApSelN SECIS with the apex of SelN SECIS, and a SelN-ApPHGPx SECIS with that of PHGPx SECIS (Figure 4). The binding of dSBP2 and hSBP2 was tested by gel-shift assays (data expressed as K_d values in Table 1). Surprisingly, the apex of the form 1 SECIS did not lead to abolition of dSBP2 binding to the chimeric PHGPx-ApSelN SECIS while the SelN-ApPHGPx SECIS, carrying the stem of SelN and the apex of the form 2 PHGPx, did not confer to dSBP2 the ability to bind. Indeed, the apex is not responsible for the distinctive binding of dSBP2 since replacing the genuine apexes by an ultrastable UUCG tetraloop in PHGPx and SelN SECIS did not modify the effects induced by apex swapping (data not shown). These results suggest that the difference between forms 1 and 2 SECIS may not rely solely on distinct apical structures (6) but that other determinants/discriminants could also exist elsewhere in the SECIS RNA 2D and/or 3D structures.

Point mutations in a newly mapped lysine-rich domain in hSBP2 affect binding to form 1 but not to form 2 SECIS RNAs

We hypothesized that a domain(s) in SBP2 might be responsible for the differential SECIS binding. First, we re-examined in-depth the boundaries of the hSBP2 RNA-binding domain. A series of N- and C-terminal deletions was performed in the construct encoding hSBP2 344–854 (Figure 5A). The resulting 6× His-tagged hSBP2 proteins were expressed in *E. coli*, purified and their binding abilities assessed by gel-shift assays with SelN (form 1) and PHGPx (form 2) SECIS RNAs (Figure 5A). Lack of binding of the 344–674 hSBP2 protein resulted from amputation of the major part of the previously mapped L7Ae RNA-binding module (20–22,25). However, the loss of binding manifested by hSBP2 proteins 545–854, 625–854 and 674–854 revealed that a region outside of the L7Ae module is also crucial. More precisely, this other domain should lie between positions 525 and 545 since hSBP2 525–854 retained SECIS binding activity. Interestingly, this region contains the conserved block of amino acids of highest sequence identity revealed by our alignments (Figure 1, residues 517–544). To characterize this domain with more accuracy, we carried out alanine scanning mutations in construct hSBP2 344–854 from positions K516 to R546, in groups of three amino acids except K516 which was singly mutated (Figure 5B). The mutant proteins were expressed in *E. coli*, purified and screened by gel-shift for their ability to bind the GPx1 and SelN form 1 and the PHGPx form 2 SECIS RNAs. Mutations that yielded the most significant impacts on K_d values are given in Figure 5C. While the mutations did not have a major effect on binding to the form 2 PHGPx SECIS (K_d values ranging from 480 to 580 nM, not deviating too much from the 350 nM wt value), the affinity dropped dramatically for form 1 SECIS and more particularly for GPx1 SECIS: a gradual augmentation of the K_d values was observed which peaked to a maximum for the Ala535–537 mutation, and dropped to near wt values for Ala541–543. For Ala535–537, the Kd increase for GPx1 SECIS was 4.5-fold the wt value. In conclusion, the most severely affected sequence was 526-AKKPTSLKKIILKER-540 and major effects were

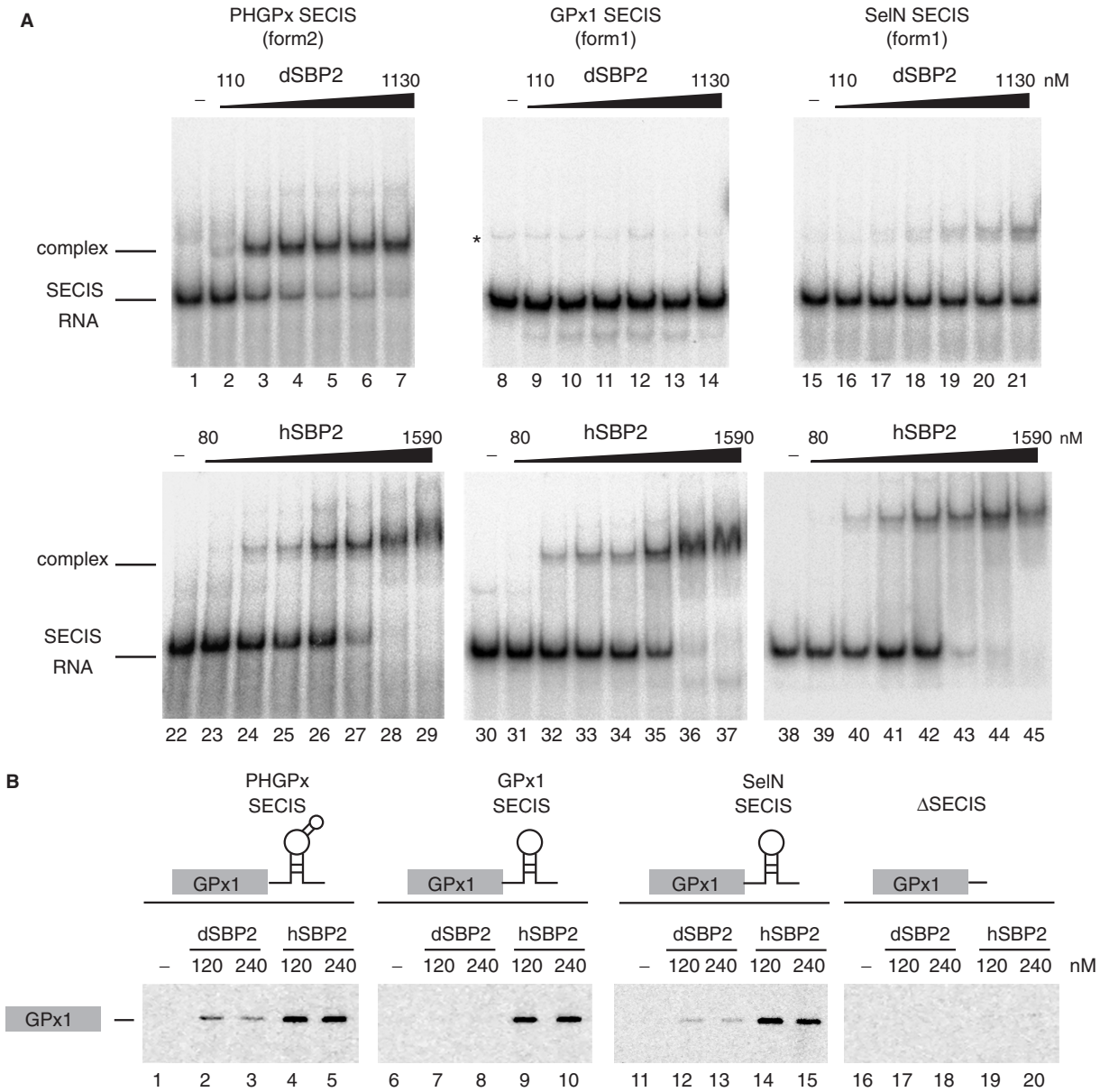


Figure 3. dSBP2 has distinct affinities for form 1 and 2 SECIS RNAs. (**A**) Gel shift assays were performed between *in vitro* transcribed ^{32}P-labeled PHGPx (form 2), GPx1 and SelN (form 1) SECIS RNAs and either purified dSBP2 or hSBP2, with the range of protein concentration (nM) indicated above each gel. Proteins were omitted in lanes 1, 8, 15, 22, 30 and 38. The asterisk in lanes 8–14 is as in Figure 2A. (**B**) *In vitro* translation assay (reticulocyte lysate) of (^{35}S)-Met-labeled rat glutathione peroxidase 1 (GPx1) from reporter constructs carrying the GPx1 ORF without SECIS RNA (ΔSECIS) or with either the PHGPx, GPx1 or SelN SECIS RNAs in the 3′UTR. Purified dSBP2 or hSBP2 were omitted (lanes 1, 6, 11 and 16) or added at the indicated concentrations (nM). Translation products were treated as in Figure 2B.

Table 1. K_d values of hSBP2 and dSBP2 for form 1, form 2 and chimeric SECIS RNAs

SECIS RNA	SBP2 protein	
	dSBP2	hSBP2
PHGPx (form2)	220 ± 50	320 ± 5
GPx1 (form1)	No binding	400 ± 50
SeIN (form1)	nd	340 ± 15
PHGPx-APSeIN	230 ± 30	530 ± 120
SeIN-ApPHGPx	No binding	470 ± 40
		K_d ± SD (nM)

provoked by mutations of 532-LKK-534 and 535-IIL-537, the latter providing a culminating effect.

Taken together, sequence comparisons, deletion and alanine scanning mutagenesis enabled us to identify a domain in hSBP2, residing between K516 and K544, that differs from the already known L7Ae module. We named it the K-rich domain because of its relatively high content in lysine residues (34%; 10 lysines out of 29 residues). This domain contains amino acids 526–540 essential for binding to form 1 SECIS RNAs but its mutation did not affect form 2 SECIS recognition.

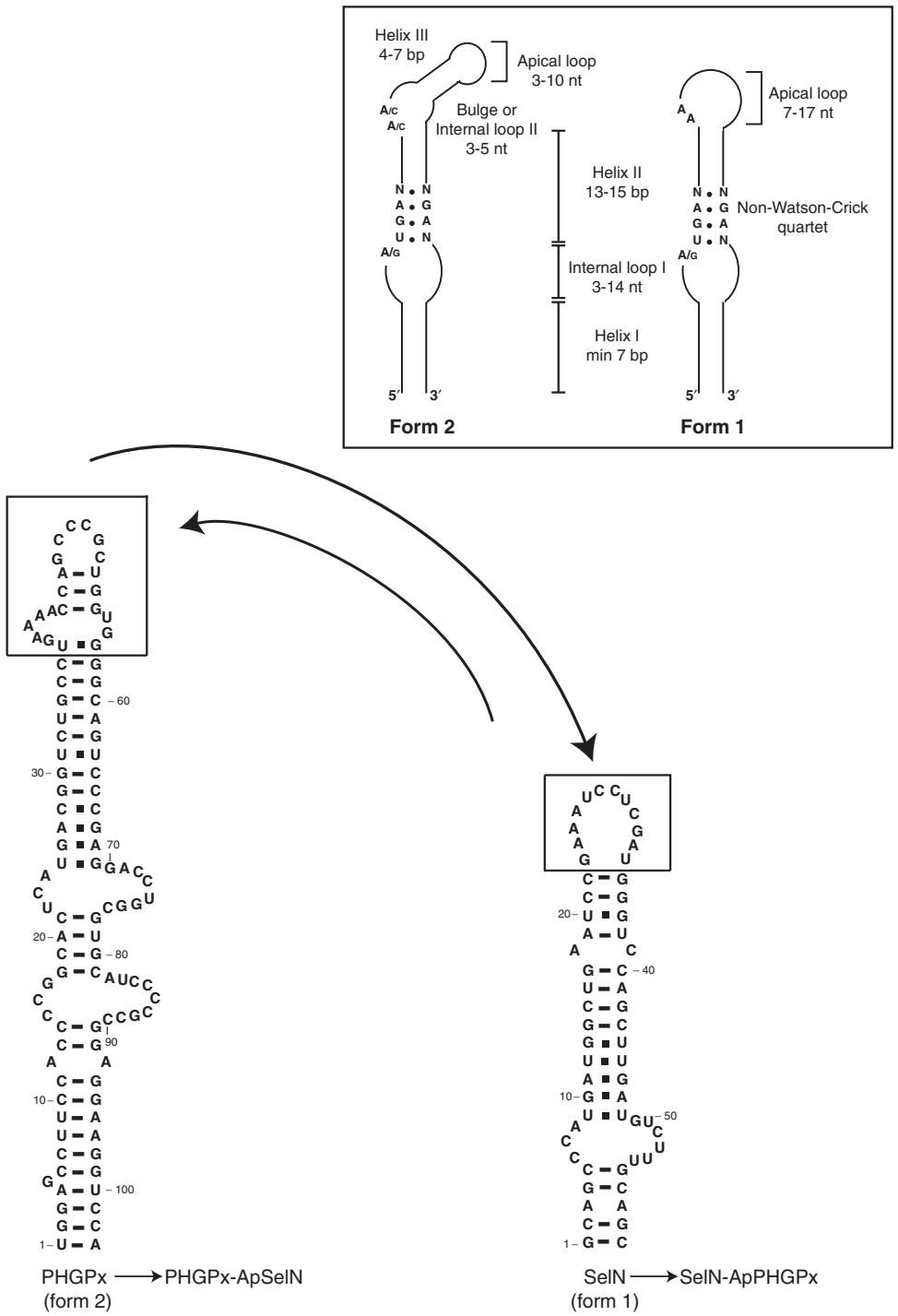


Figure 4. Chimeric SECIS RNAs containing swapped apexes. The apical regions (boxed) of form 2 PHGPx and form 1 SelN SECIS were exchanged (curved arrows) to yield the chimeric PHGPx-ApSelN and SelN-ApPHGPx SECIS harboring the SelN and PHGPx apexes (Ap), respectively. The drawing in the inset schematizes the conserved and variable structural features occurring in forms 1 and 2 SECIS RNAs.

The same mutations in the K-rich domain affecting SBP2 binding to form 1 SECIS also impede interaction with the 60S ribosomal subunit

Rat SBP2 has been shown to bind purified rat 80S ribosomes (21,24,44). The down-effects that we observed upon mutating the K-rich domain prompted us to investigate whether the same mutations would affect ribosome binding. More specifically, we first determined which of the ribosomal subunits, 60S or 40S, was the target for the full-length hSBP2 (hSBP2 FL). In these and further experiments, after incubation with hSBP2, the purified ribosomal subunits were loaded onto sucrose gradients in a buffer containing 0.5 mM Mg^{2+}, and the fractions were analyzed by western blotting with an anti-SBP2 antibody.

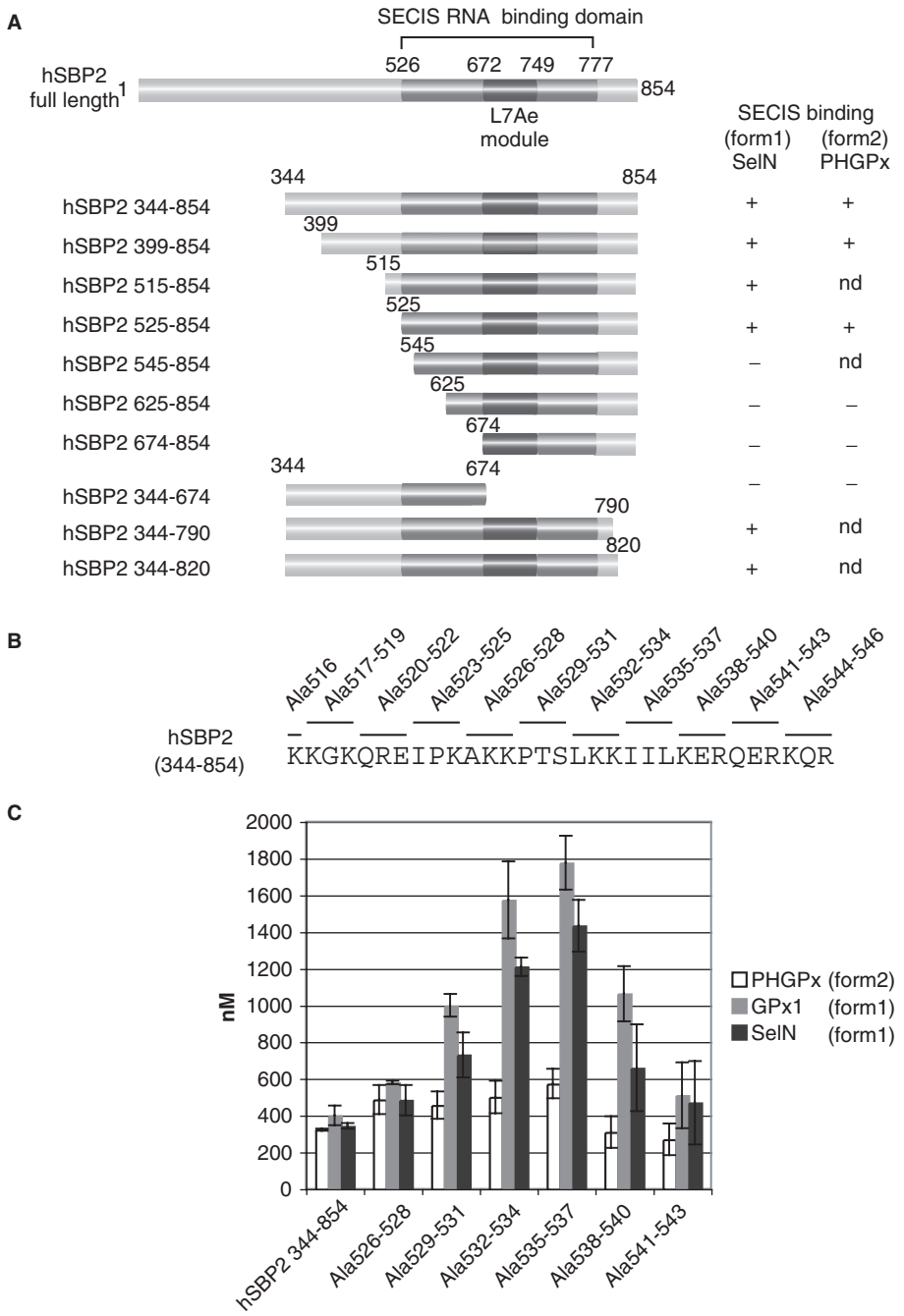


Figure 5. Identification of the K-rich additional RNA binding domain in hSBP2. (**A**) N- and C-ter deletion mutants were engineered in the full-length hSBP2 cDNA to yield the displayed constructs. hSBP2 proteins were purified and assayed by EMSA for their binding to *in vitro* transcribed ^{32}P-radiolabeled SelN (form 1) or PHGPx (form 2) SECIS RNAs. (+), binds; (–) no binding; nd, not determined. (**B**) Alanine scanning mutagenesis. Residues 516–546 in the hSBP2 344–854 cDNA were mutated to alanines in groups of three amino acids (with the exception of the single K516A), as shown above the sequence. Proteins were purified and assayed by gel shifts with *in vitro* transcribed ^{32}P-labeled PHGPx, GPx1 or SelN SECIS RNAs. (**C**) K_d values derived from gel shift experiments between the indicated hSBP2 alanine scanning mutants and the form 2 PHGPx (white bars), and form 1 GPx1 (grey bars) and SelN (black bars) SECIS RNAs. Values are given in nM (mean ± SD).

Figure 6A shows a high intensity signal with the 60S but a very faint one with the 40S subunit (compare lane 2 with lane 1). Similar results were obtained with 80S ribosomes incubated with hSBP2 FL under these conditions (data not shown). Human recombinant ribosomal protein p40 was used as a control since it was recently shown to be capable of binding 40S subunits (45): it effectively bound 40S but not 60S subunits (Figure 6A, compare lane 4 with lane 3). Therefore, we can conclude that under the conditions used, hSBP2 FL bound to 60S subunits in a specific manner. To identify the hSBP2 domain required for 60S binding, we tested the effects of the

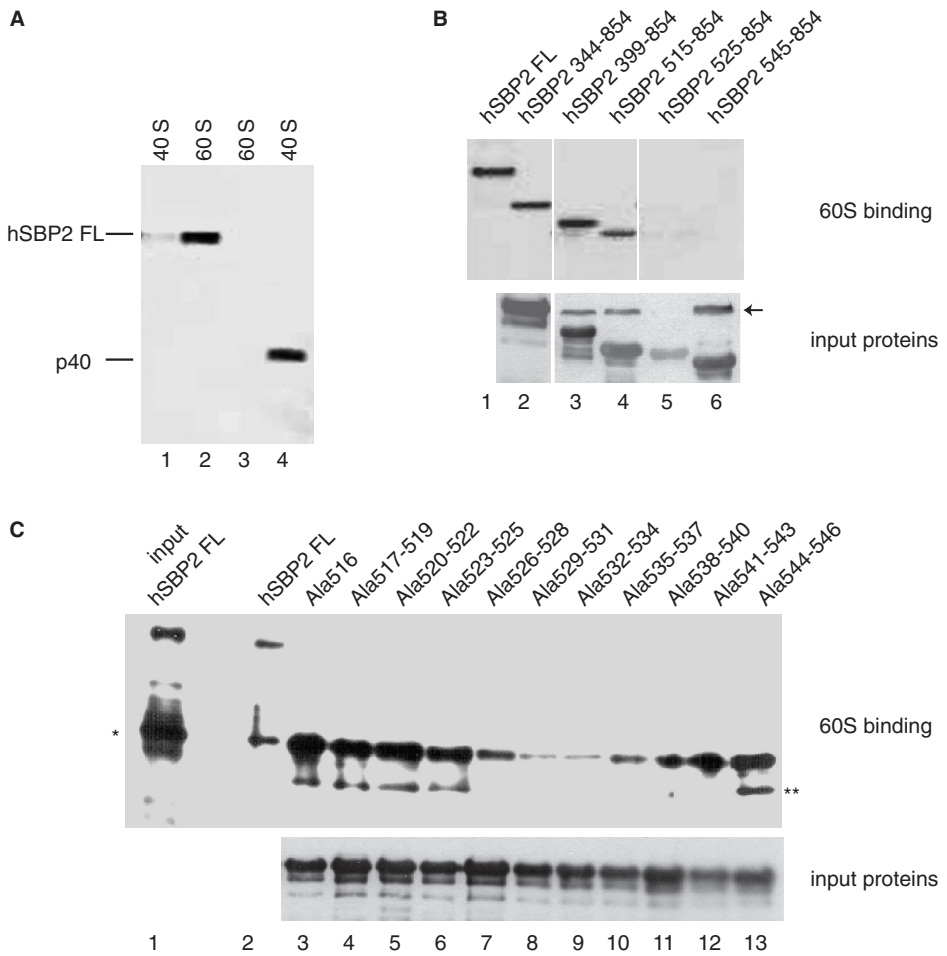


Figure 6. Human SBP2 interacts with the 60S ribosomal subunit. (**A**) Detection of hSBP2 in sucrose gradient fractions of 60S and 40S ribosomal subunits. TCA-precipitated fractions were resolved by SDS–PAGE and assayed by western blot with an anti-hSBP2 antibody (lanes 1 and 2). hSBP2 FL is the full-length human SBP2, p40 (lane 4) a 40S ribosomal subunit protein used as the control. (**B**) hSBP2 deletion mutant proteins were assayed for ribosome binding. Fractions containing 60S bound hSBP2 were treated as in (A). Nomenclature as in (A) and Figure 5A. The lower panel (input proteins) is a western blot showing that the mutant proteins were indeed expressed and detectable by the anti-SBP2 antibody. The arrow points to a cross-reaction product arising from an *E.coli* protein. (**C**) The 60S-bound alanine scanning mutant proteins (shown in Figure 5B) were analyzed as in (A). Lane 1: hSBP2FL detected by western blot before gradient loading. The single and double asterisks denote frequently encountered hSBP2 proteolysis fragments. The lower panel (input proteins) serves the same purpose as in (B).

hSBP2 deletion mutants ending at 854 and starting at 344, 399, 515, 525 and 545. Truncations until position 515 were innocuous (Figure 6B, lanes 2–4, compare with lane 1) whereas removing sequences downstream from 515 (hSBP2 525–854 and hSBP2 545–854; Figure 6B, lanes 5 and 6, respectively) led to complete inhibition of 60S binding. Next, the alanine scanning mutants were used to identify the important amino acids. None of the alanine replacements, from 516 to 525 or 541 to 546, significantly altered the binding of hSBP2 (Figure 6C, compare lanes 3–6 and 12, 13 with lane 2). In contrast, alanine scanning mutants Ala526–528 to Ala538–540 (Figure 6C, lanes 7–11) provoked a marked drop in binding, mutants Ala529–531 and Ala532–540 inducing an almost complete abolition of binding (lanes 8 and 9).

The two sets of mutagenesis experiments identify hSBP2 amino acids located between A526 and R543 that are crucial for both 60S ribosomal subunit interaction and to provide affinity to form 1 SECIS RNAs.

Swapping five amino acids in the lysine-rich domain between *D. melanogaster* and hSBP2 reversed the affinity toward form 1 or form 2 SECIS RNAs

Multiple sequence alignments of 26 vertebrate, one echinoderm (sea urchin) and 13 insect SBP2 (12 *Drosophila* and the *Anopheles gambiae* mosquito) amino acid sequences were performed to obtain information about the evolutionary conservation of the K-rich domain, aligning sequences corresponding to hSBP2 positions 515–547 (Figure 7). This highlighted the sequence conservation 517-KGKXXRXXPKXKKXTXLKXXI/$_{V}$XXXR-540 (numbering of hSBP2). However, inspection of non-conserved residues revealed a dichotomy. In particular, the sequence 535-I/$_{V}$ILKE-539 is replaced by SVRVY in dSBP2, and shows less conservation across the 12 *Drosophila* species and the *Anopheles* mosquito than between other SBP2s. The only similar amino acid is I536 in hSBP2 which is replaced by a valine residue

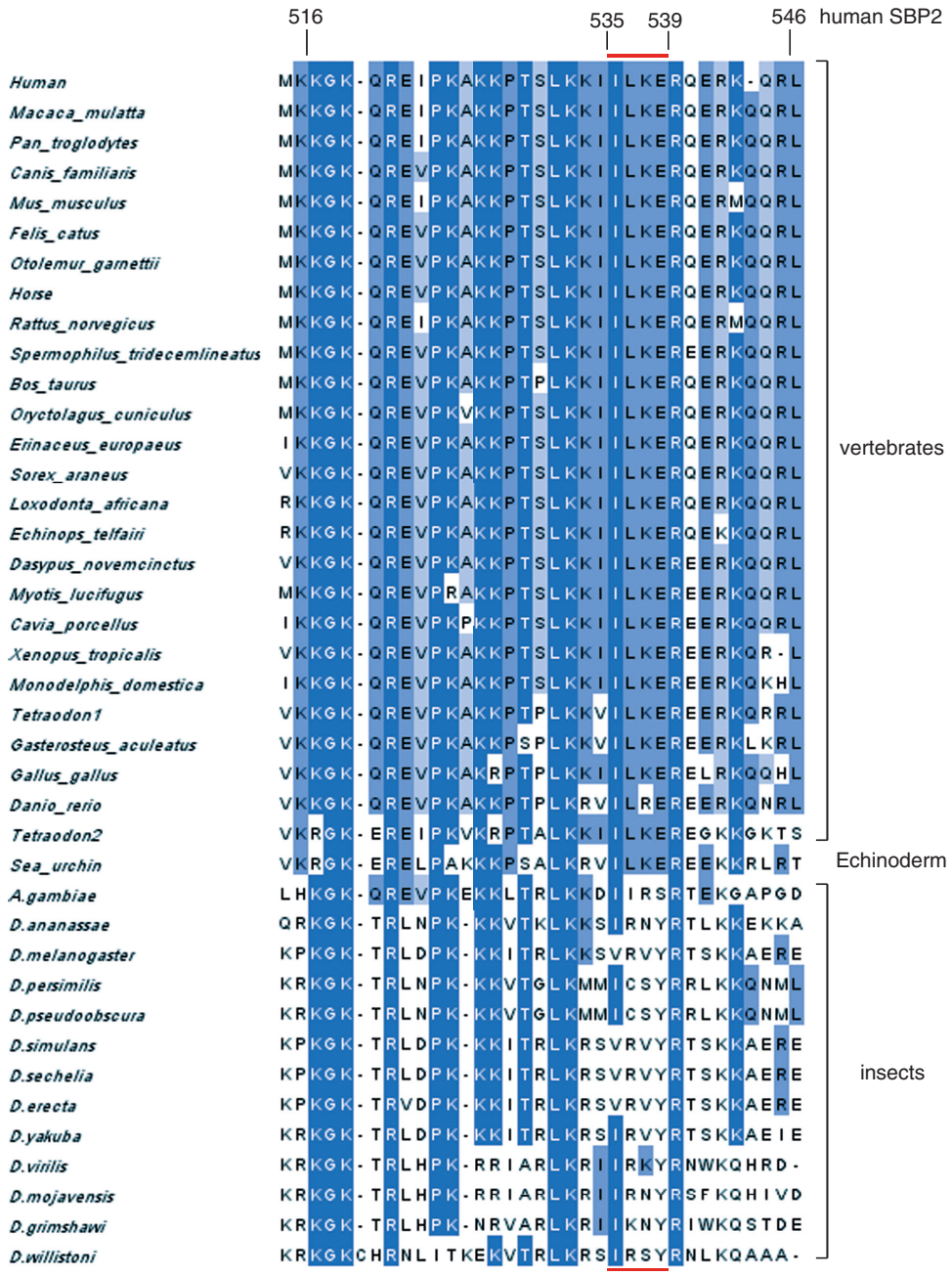


Figure 7. Alignment of the K-rich domain in 26 vertebrate, one echinoderm (sea urchin) and 13 insect (12 *Drosophila* and the mosquito *Anopheles gambiae*) SBP2 sequences. The numbering is that of hSBP2. The red bar above and below sequences refers to the IILKE and SVRVY pentapeptides mentioned in the text.

in *Drosophila*. We asked whether this five amino acid change (residues 535–539) between hSBP2 and dSBP2 could be responsible for the inability of dSBP2 to bind form 1 SECIS RNAs. To test this possibility, we swapped the hSBP2 IILKE for the corresponding dSBP2 SVRVY sequence, and vice versa, to yield dSBP2-IILKE and hSBP2-SVRVY. The binding activities to form 1 and 2 SECIS RNAs of the purified chimeric SBP2 proteins were measured by gel-shift assays. Verifying our prediction, dSBP2-IILKE gained the ability to bind the GPx1 and SelN form1 SECIS RNAs (Figure 8A, lanes 10–16 and 18–24, respectively), the affinity for SelN SECIS being higher than for GPx1 (K_d = 200 nM versus K_d > 1500 nM for GPx1; Table 2). The binding affinity to form 2 PHGPx (Figure 8A, lanes 2–8) even increased by a factor of five with a K_d = 40 nM (Table 2) versus 220 nM for the wt dSBP2 (Table 1). Conversely, the five amino acid swapping led hSBP2-SVRVY to completely loose its ability to bind GPx1 and SelN SECIS (Figure 8B, lanes 10–16 and 18–24, respectively), and modified moderately (by a factor

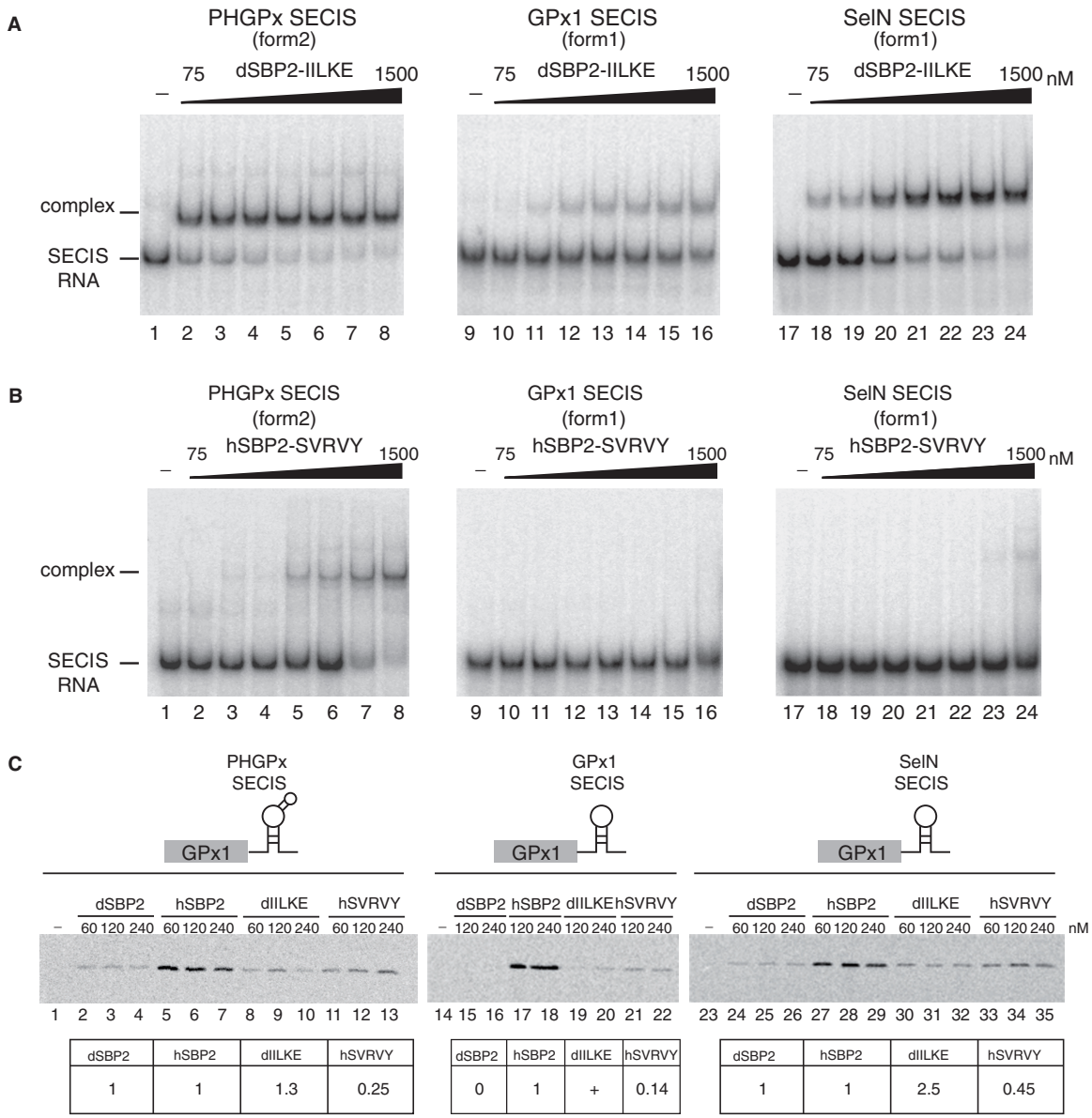


Figure 8. Reversed SECIS RNA binding properties provided by swapping five amino acids between *Drosophila* and human SBP2s. (**A**) Increasing concentrations (75–1500 nM) of purified dSBP2 carrying the human SBP2 IILKE sequence (dSBP2-IILKE) were assayed with the indicated form 2 or form 1 ^{32}P-labeled SECIS RNAs. No protein was added in lanes 1, 9, 17. (**B**) The same experiments were carried out with the hSBP2 containing the *D.melanogaster* SVRVY sequence (hSBP2-SVRVY). (**C**) *In vitro* translation assay (reticulocyte lysate) of (^{35}S)-Met-labeled rat glutathione peroxidase 1 (GPx1) from reporter constructs carrying the GPx1 ORF with either the PHGPx, GPx1 or SelN SECIS RNAs in the 3′UTR. Rabbit reticulocyte lysate was supplemented with purified dSBP2, dSBP2-IILKE (dIILKE), hSBP2 or hSBP2-SVRVY (hSVRVY) at the indicated concentrations (nM); SBP2 proteins were omitted in lanes 1,14,23. Translation products were treated as in Figure 2B. Values (expressed relative to dSBP2 or hSBP2 set as 1, or 0 for dSBP2 in lanes 15, 16) resulted from the mean of the values obtained with two to three protein concentrations in two to three independent experiments.

lower than 2) its binding to form 2 PHGPx SECIS (Figure 8B, lanes 2–8) with a K_d of 600 nM (Table 2) versus 320 nM for the wt hSBP2 (Table 1).

Next, the abilities of the chimeric SBP2 proteins to support synthesis of the GPx1 selenoprotein, carrying distinct SECIS in the 3′UTR of its mRNA, were verified in rabbit reticulocyte lysate. With the form 2 PHGPx SECIS, Figure 8C shows that the amount of the synthesized reporter was not significantly altered whether wt dSBP2 or dSBP2-IILKE was employed (value = 1.3; Figure 8C, compare lanes 8–10 with lanes 2–4). However, dSBP2 acquired the ability of supporting GPx1 selenoprotein synthesis by the simple replacement of SVRVY by the homologous human IILKE sequence in the swapping mutant dSBP2-IILKE: this was the case for GPx1 (Figure 8C, lanes 19, 20; compare with the complete inability of wt dSBP2, lanes 15, 16) and SelN SECIS (Figure 8C, compare lanes 30–32 with lanes 24–26) which

Table 2. K_d values of chimeric *Drosophila* and human SBP2 for form 1 and 2 SECIS RNAs

SECIS RNA	SBP2 protein	
	dSBP2-IILKE	hSBP2-SVRVY
PHGPx (form2)	40 ± 6	600 ± 45
GPx1 (form1)	>1500	No binding
SeIN (form1)	200 ± 10	nd
		$K_d \pm$ SD (nM)

showed a 2.5-fold gain. In contrast, introduction of the *Drosophila* SVRVY sequence instead of IILKE in hSBP2 (hSBP2-SVRVY) provided an almost total (14% residual) or partial (45% residual) inhibition of selenoprotein synthesis when assayed with the form 1 SelN and GPx1 SECIS, respectively (Figure 8C; compare lanes 33–35 to lanes 27–29; lanes 21–22 with lanes 17, 18). Tested with the form 2 PHGPx SECIS, hSBP2-SVRVY led to a 0.25-fold drop of the value obtained with wt hSBP2 (Figure 8C, compare lanes 11–13 with lanes 5–7). This was unexpected but this finding nevertheless correlated with the data shown in Figure 8A and B and Table 2 establishing a loss of affinity between this chimeric SBP2 and the PHGPx SECIS (the K_d increased by a factor 2). At present, we are unable to explain why hSBP2-SVRVY did not retain close to wild-type properties in gel-shift and selenoprotein synthesis assays with form 2 SECIS RNA. Notwithstanding, this set of experiment was globally in line with the *in vitro* EMSA data shown in Figure 3B and Figure 8A and B and the K_d values in Tables 1 and 2.

We conclude from these experiments that the IILKE sequence confers the ability to bind form 1 SECIS RNAs.

DISCUSSION

SBP2 plays a central role in selenoprotein synthesis by binding to SECIS hairpins in the 3′UTR of selenoprotein mRNAs. Earlier domain dissection of the human and rat SBP2 established that the C-terminal 2/3 are involved in SECIS and ribosome binding as well as selenocysteine incorporation while no function has been attributed so far to its N-terminal part, which is dispensable for selenoprotein synthesis in reticulocyte lysate (20–22,24). In this work, we have isolated and functionally characterized a bona fide dSBP2 that specifically lacks the region homologous to the N-terminus of vertebrate SBP2. In addition, we report the identification in human SBP2 of a lysine-rich (K-rich) domain that is essential for SECIS binding, point mutations therein affecting form 1 but not form 2 SECIS binding. Sequence comparisons established that the K-rich domain is encountered in all the SBP2s analyzed (this work and data not shown). In the *D. melanogaster* K-rich domain, a five amino acid sequence difference renders *Drosophila* SBP2 incapable of binding form 1 SECIS RNAs with high affinity. Exchanging this sequence with that of hSBP2 enabled binding of the insect SBP2 to form 1 SECIS but impaired that of hSBP2. Moreover, we found that the K-rich sequence is also crucial for the binding of hSBP2 to the 60S ribosomal subunit.

The insect SBP2 is 313 amino acids long, a little less than a third of the mammalian counterpart. *Drosophila* SBP2 is shorter at the C-terminus and a small internal deletion removed positions corresponding to hSBP2 565–689 (Figure 1). However, the lack of the N-terminal region corresponding to hSBP2 1–427 accounts for most of its reduction in size. That selenoprotein synthesis can be achieved in an organism lacking this domain corroborates the finding that it is dispensable in mammalian SBP2 under the experimental conditions used (25). Although one cannot exclude the possibility that this segment of SBP2 is encoded by a separate gene in insects, this appears unlikely because a search in the *Drosophila* genomes failed to find significant sequence similarity to the vertebrate N-terminal region (data not shown). What might then be the function of the N-terminal extension in higher eukaryotes? Several possibilities exist that would pertain to a more complex selenoprotein synthesis mechanism in higher eukaryotes: (i) the N-terminal extension could participate in fine-tuning selenoprotein expression with a direct or indirect role in the SBP2 nuclear/cytoplasmic shuttling, as it contains a nuclear localization signal (46); (ii) we established by structure prediction and experimental data that SBP2 is an Intrinsically Disordered Protein and that the N-terminal extension is widely unstructured (Olieric e*t al.*, manuscript in preparation). It is very possible that this region acquires its proper folding in the presence of yet to be discovered protein partners, consistent with the role of the Hsp90 chaperone and co-chaperones in the folding and assembly of proteins bearing an L7Ae RNA-binding module (47) and (iii) finally, the N-terminal domain could be involved in an SBP2 function different from selenocysteine incorporation, as inferred very recently from the finding that several SBP2 isoforms arise from splice variants in the N-terminal region (48).

We found a region in hSBP2, the K-rich (lysine-rich) domain 516-KKGKQREIPKAKKPTSLKKIILKERQ ER-543, that is essential for SECIS binding. This sequence is highly conserved across vertebrates and is distinct from the L7Ae RNA-binding module which is therefore insufficient on its own to provide SECIS recognition. Surprisingly, however, alanine scanning mutagenesis of sequence 516–543 manifested adverse effects on SECIS binding. While it did not significantly prohibit binding to form 2, dramatic effects were observed on form 1, especially when altering the sequence 526-AKKPTSLKKII LKER-540 and most prominently 532-LKKIIL-537 which profoundly disabled SBP2.

A multiple sequence alignment identified KGKTRLD PKKKITRLKKSVRVYR (*D. melanogaster* sequence) as the homolog of the human K-rich domain. This sequence analysis highlighted a characteristic feature distinguishing the insect and other K-rich domains, i.e. the five amino-acid substitution $I/_V$ILKE in vertebrates to SVRVY in *D. melanogaster*. dSBP2 can bind *Drosophila* and mammalian form 2 SECIS RNAs but was unable—or with very low affinity—to bind form 1 SECIS RNAs, consistent with *Drosophila* having selenoprotein mRNAs with form 2 SECIS exclusively (8,31). The IILKE/SVRVY

swapping experiment enabled dSBP2 to bind form 1 SECIS but led hSBP2 to loose this ability. It is very unlikely that the SVRVY sequence per se is inhibitory to form 1 binding since replacement of IIL and KER by alanines in hSBP2 yielded the same down effect as introduction of SVRVY in hSBP2. It seems more plausible that the ability to bind form 1 SECIS requires the occurrence of IILKE because of its different amino-acid composition versus SVRVY, giving rise to a different charge and hydrophobicity. The SBP2 functional RNA-binding domain (526–777 in humans) consists therefore of two subdomains, the K-rich (this work) and the previously characterized conserved L7Ae module extending from Arginine 672 to Isoleucine 749 (20,49). The intervening sequence exhibits less conservation with the exception of the 624-IHSRRFR-630 block (positions relative to human SBP2). Bubenik and Driscoll (22) reported in rat SBP2 the existence of a bipartite RNA binding domain in which R531 (R540 in hSBP2) appeared important for discriminating forms 1 and 2 SECIS hairpins. Indeed, this arginine is universal in SBP2 (Figure 7) and a mutation to glutamine, described in SBP2 patients with thyroid dysfunctions, impaired binding to the form 1 SECIS of the iodothyronine deiodinase mRNA (50). However, we did not observe such a dramatic effect of the R540 mutation in human SBP2 since the Ala 538–540 mutation only mildly affected form 1 binding.

Additionally, and very interestingly, we showed the importance of that same 526-AKKPTSLKKIILKER-540 sequence for binding to the 60S ribosomal subunit. These findings establish the ability of SBP2 to stably bind the 60S subunit and the identification of the PTSLKK motif that contributes in a very important manner to the binding. While this work was in progress, Donovan *et al.* (51) reported in rat SBP2 the identification of the additional domain SID, required for SECIS and 80S ribosome binding. The boundaries of the SID are similar to those we delineated for the K-rich domain but the authors did not report a distinctive effect of the SID on form 1 or form 2. Moreover, the mutation of the IILKE sequence by these authors significantly affected binding to form 2 SECIS, in contrast to our alanine scanning and swapping experiments. Also, we determined by sucrose gradient centrifugation that SBP2 interacts with the 60S but not the 40S ribosomal subunit. This is consistent with a proposal that SBP2 could bind the 28S ribosomal RNA (24). Lastly, our finding that the same amino acids in the K-rich domain are involved in SECIS RNA and 60S ribosomal subunit recognition strengthens the model establishing that SBP2 cannot bind both simultaneously (44).

Worthy of note is that SBP2, with two domains crucial for SECIS binding, is set apart from the other proteins of the L7Ae family which are shorter and for which the L7Ae domain is itself sufficient to ensure specific interactions with the cognate RNA (11,18,19,52,53). In a previous work, we provided a very precise definition of the RNA-binding specificity of hSBP2 (23). In particular, we identified nucleotide determinants in the SECIS RNA that are unique to SBP2 among the L7Ae family members: while the 15.5 kD and L7Ae ribosomal proteins have a rather broad specificity and can recognize the SECIS RNA, the converse does not hold true since SBP2 is unable to bind the U3 snoRNA or the archaeal sRNA (23,54). As a matter of fact, footprinting experiments and interference of binding established the requirement of helix I and internal loop 1 in SECIS RNAs for SBP2 recognition (23,55). These structural features are idiosyncratic to SECIS RNAs and therefore not found in other Kink-turn containing RNAs, leading to our proposal that SECIS hairpins contain a Kink-turn like rather than a canonical Kink-turn motif (14). We therefore propose a model rationalizing the necessity of two domains in SBP2 for complex formation with the SECIS RNA: (i) a restricted number of amino acids in the L7Ae module establish contacts with the guanine bases of the G.A/A.G base pairs as in other L7Ae proteins, and the conserved U in the SECIS non-Watson-Crick quartet (11,20); (ii) the lysines in the K-rich domain contribute positive charges for electrostatic interactions with the phosphates of the SECIS-specific structural features and thus increase the affinity of SBP2 for SECIS RNAs. In this domain, the amino acid composition of the IILKE sequence could have a direct impact on SECIS binding or indirectly lead to an L7Ae conformational change allowing form 1 SECIS recognition.

Validation of these models definitely requires that the crystal structure of the SECIS RNA-SBP2 complex be solved.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to A. Beniaminov for providing plasmid GPx1-SelN SECIS, to S. Baudrey and A. Schweigert for skillful technical assistance. We thank A. Beniaminov, P. Carbon, M. Leichter, E. Myslinski and L. Wurth for reading the manuscript and helpful discussions.

FUNDING

Action Concertée Incitative (BCMS 226) and Programme InterOrganismes (Tox.Nuc-E) [to A.K.]; the Spanish Ministry of Education (BIO2006-03380) and Biosapiens LSHG-CT-2003-503265 (FP6 Programme of the European Commission) [to R.G.]; the Russian Foundation for Basic Research (grant #08-04-00508) [to G.K.]; Pre-doctoral fellowship from the Japanese Ministry of Education, Culture, Sports, Science and Technology [to A.T.]; Pre-doctoral fellowship of the Spanish Ministry of Education and Science [to C.C.]; EMBO short-term fellowship (ASTF 91-2007) [to E.B.]. Funding for open access charge: CNRS.

Conflict of interest statement. None declared.

REFERENCES

1. Allmang,C. and Krol,A. (2006) Selenoprotein synthesis: UGA does not end the story. *Biochimie*, **88**, 1561–1571.

2. Carlson,B.A., Xu,X.-M., Kryukov,G.V., Rao,M., Berry,M.J., Gladyshev,V.N. and Hatfield,D.L. (2004) Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. *Proc. Natl Acad. Sci. USA*, **101**, 12848–12853.
3. Xu,X.-M., Mix,H., Carlson,B.A., Grabowski,P.J., Gladyshev,V.N., Berry,M.J. and Hatfield,D.L. (2005) Evidence for direct roles of two additional factors, SECp43 and soluble liver antigen, in the selenoprotein synthesis machinery. *J. Biol. Chem.*, **280**, 41568–41575.
4. Small-Howard,A., Morozova,N., Stoytcheva,Z., Forry,E.P., Mansell,J.B., Harney,J.W., Carlson,B.A., Xu,X.M., Hatfield,D.L. and Berry,M.J. (2006) Supramolecular complexes mediate selenocysteine incorporation *in vivo*. *Mol. Cell Biol.*, **26**, 2337–2346.
5. Chavatte,L., Brown,B.A. and Driscoll,D.M. (2005) Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat. Struct. Mol. Biol.*, **12**, 408–416.
6. Grundner-Culemann,E., Martin,G.W.3rd, Harney,J.W. and Berry,M.J. (1999) Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. *RNA*, **5**, 625–635.
7. Fagegaltier,D., Lescure,A., Walczak,R., Carbon,P. and Krol,A. (2000) Structural analysis of new local features in SECIS RNA hairpins. *Nucleic Acids Res.*, **28**, 2679–2689.
8. Chapple,C.E., Guigo,R. and Krol,A. (2009) SECISaln, a web-based tool for the creation of structure-based alignments of eukaryotic SECIS elements. *Bioinformatics*, [Epub ahead of print].
9. Walczak,R., Westhof,E., Carbon,P. and Krol,A. (1996) A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA*, **2**, 367–379.
10. Walczak,R., Carbon,P. and Krol,A. (1998) An essential non-Watson-Crick base pair motif in 3'UTR to mediate selenoprotein translation. *RNA*, **4**, 74–84.
11. Vidovic,I., Nottrott,S., Hartmuth,K., Lührmann,R. and Ficner,R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell*, **6**, 1331–1342.
12. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
13. Lescoute,A., Leontis,N.B., Massire,C. and Westhof,E. (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.*, **33**, 2395–2409.
14. Allmang,C. and Krol,A. (2006) SECIS RNAs and K-turn binding proteins. A survey of evolutionary conserved RNA and protein motifs. In Hatfield,D.L. (ed.), *Selenium, its Molecular Biology and Role in Human Health 2nd edition*, Vol. 5. Kluwer Academic Publishers, New York, USA, pp. 51–61.
15. Watkins,N.J., Segault,V., Charpentier,B., Nottrott,S., Fabrizio,P., Bachi,A., Wilm,M., Rosbash,M., Branlant,C. and Luhrmann,R. (2000) A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, **103**, 457–466.
16. Rozhdestvensky,T.S., Tang,T.H., Tchirkova,I.V., Brosius,J., Bachellerie,J.P. and Hüttenhofer,A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.
17. Koonin,E.V., Bork,P. and Sander,C. (1994) A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Res.*, **22**, 2166–2167.
18. Chao,J.A. and Williamson,J.R. (2004) Joint X-ray and NMR refinement of the yeast L30e-mRNA complex. *Structure*, **12**, 1165–1176.
19. Moore,T., Zhang,Y., Fenley,M.O. and Li,H. (2004) Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure*, **12**, 807–818.
20. Allmang,C., Carbon,P. and Krol,A. (2002) The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *RNA*, **8**, 1308–1318.
21. Caban,K., Kinzy,S.A. and Copeland,P.R. (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. Cell Biol.*, **27**, 6350–6360.
22. Bubenik,J.L. and Driscoll,D.M. (2007) Altered RNA binding activity underlies abnormal thyroid hormone metabolism linked to a mutation in selenocysteine insertion sequence-binding protein 2. *J. Biol. Chem.*, **282**, 34653–34662.
23. Cléry,A., Bourguignon-Igel,V., Allmang,C., Krol,A. and Branlant,C. (2007) An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery. *Nucleic Acids Res.*, **35**, 1868–1884.
24. Copeland,P.R., Stepanik,V.A. and Driscoll,D.M. (2001) Insight into mammalian selenocysteine insertion: domain structure and ribosome binding properties of Sec insertion sequence binding protein 2. *Mol. Cell Biol.*, **21**, 1491–1498.
25. Copeland,P.R., Fletcher,J.E., Carlson,B.A., Hatfield,D.L. and Driscoll,D.M. (2000) A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J.*, **19**, 306–314.
26. Lobanov,A.V., Fomenko,D.E., Zhang,Y., Sengupta,A., Hatfield,D.L. and Gladyshev,V.N. (2007) Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome Biol.*, **8**, R198.
27. Hirosawa-Takamori,M., Jackle,H. and Vorbruggen,G. (2000) The class 2 selenophosphate synthetase gene of Drosophila contains a functional mammalian-type SECIS. *EMBO Rep.*, **1**, 441–446.
28. Castellano,S., Morozova,N., Morey,M., Berry,M.J., Serras,F., Corominas,M. and Guigó,R. (2001) In silico identification of novel selenoproteins in the Drosophila melanogaster genome. *EMBO Rep.*, **2**, 697–702.
29. Lu,C., Qiu,F., Zhou,H., Peng,Y., Hao,W., Xu,J., Yuan,J., Wang,S., Qiang,B., Xu,C. *et al.* (2006) Identification and characterization of selenoprotein K: an antioxidant in cardiomyocytes. *FEBS Lett.*, **580**, 5189–5197.
30. Dikiy,A., Novoselov,S.V., Fomenko,D.E., Sengupta,A., Carlson,B.A., Cerny,R.L., Ginalski,K., Grishin,N.V., Hatfield,D.L. and Gladyshev,V.N. (2007) SelT, SelW, SelH, and Rdx12: genomics and molecular insights into the functions of selenoproteins of a novel thioredoxin-like family. *Biochemistry*, **46**, 6871–6882.
31. Chapple,C.E. and Guigo,R. (2008) Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. *PLoS ONE*, **3**, e2968.
32. Copeland,P.R. and Driscoll,D.M. (1999) Purification, redox sensitivity, and RNA binding properties of SECIS- binding protein 2, a protein involved in selenoprotein biosynthesis. *J. Biol. Chem.*, **274**, 25447–25454.
33. Lescure,A., Allmang,C., Yamada,K., Carbon,P. and Krol,A. (2002) cDNA cloning, expression pattern and RNA binding analysis of human selenocysteine insertion sequence (SECIS) binding protein 2. *Gene*, **291**, 279–285.
34. Studier,F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.
35. Lassmann,T. and Sonnhammer,E.L. (2005) Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*, **6**, 298.
36. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
37. Katoh,K., Misawa,K., Kuma,K. and Miyata,T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059–3066.
38. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.
39. Busso,D., Delagoutte-Busso,B. and Moras,D. (2005) Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in Escherichia coli. *Anal. Biochem.*, **343**, 313–321.
40. Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.*, **154**, 367–382.
41. Lescure,A., Gautheret,D., Carbon,P. and Krol,A. (1999) Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif. *J. Biol. Chem.*, **274**, 38147–38154.
42. Matasova,N.B., Myltseva,S.V., Zenkova,M.A., Graifer,D.M., Vladimirov,S.N. and Karpova,G.G. (1991) Isolation of ribosomal subunits containing intact rRNA from human placenta: estimation of functional activity of 80S ribosomes. *Anal. Biochem.*, **198**, 219–223.

43. Allamand,V., Richard,P., Lescure,A., Ledeuil,C., Desjardin,D., Petit,N., Gartioux,C., Ferreiro,A., Krol,A., Pellegrini,N. *et al.* (2006) A single homozygous point mutation in a 3'untranslated region motif of selenoprotein N mRNA causes SEPN1-related myopathy. *EMBO Rep.*, **7**, 450–454.
44. Kinzy,S.A., Caban,K. and Copeland,P.R. (2005) Characterization of the SECIS binding protein 2 complex required for the co-translational insertion of selenocysteine in mammals. *Nucleic Acids Res.*, **33**, 5172–5180.
45. Kossinova,O.A., Malygin,A.A., Babailova,E.S. and Karpova,G.G. (2008) Binding of human ribosomal protein p40 and its truncated mutants to the small ribosomal subunit. *Molecular Biology*, **42**, 911–916.
46. Papp,L.V., Lu,J., Striebel,F., Kennedy,D., Holmgren,A. and Khanna,K.K. (2006) The redox state of SECIS binding protein 2 controls its localization and selenocysteine incorporation function. *Mol. Cell Biol.*, **26**, 4895–4910.
47. Boulon,S., Marmier-Gourrier,N., Pradet-Balade,B., Wurth,L., Verheggen,C., Jady,B.E., Rothe,B., Pescia,C., Robert,M.C., Kiss,T. *et al.* (2008) The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J. Cell Biol.*, **180**, 579–595.
48. Papp,L.V., Wang,J., Kennedy,D., Boucher,D., Zhang,Y., Gladyshev,V.N., Singh,R.N. and Khanna,K.K. (2008) Functional characterization of alternatively spliced human SECISBP2 transcript variants. *Nucleic Acids Res.*, **36**, 7192–7206.
49. Lescure,A., Fagegaltier,D., Carbon,P. and Krol,A. (2002) Protein factors mediating selenoprotein synthesis. *Curr. Protein Pept. Sci.*, **3**, 143–151.
50. Dumitrescu,A.M., Liao,X.H., Abdullah,M.S., Lado-Abeal,J., Majed,F.A., Moeller,L.C., Boran,G., Schomburg,L., Weiss,R.E. and Refetoff,S. (2005) Mutations in SECISBP2 result in abnormal thyroid hormone metabolism. *Nat. Genet.*, **37**, 1247–1252.
51. Donovan,J., Caban,K., Ranaweera,R., Gonzalez-Flores,J.N. and Copeland,P.R. (2008) A novel protein domain induces high affinity selenocysteine insertion sequence binding and elongation factor recruitment. *J. Biol. Chem.*, **283**, 35129–35139.
52. Nottrott,S., Urlaub,H. and Lührmann,R. (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.*, **21**, 5527–5538.
53. Marmier-Gourrier,N., Clery,A., Senty-Segault,V., Charpentier,B., Schlotter,F., Leclerc,F., Fournier,R. and Branlant,C. (2003) A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA*, **9**, 821–838.
54. Charron,C., Manival,X., Clery,A., Senty-Segault,V., Charpentier,B., Marmier-Gourrier,N., Branlant,C. and Aubry,A. (2004) The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J. Mol. Biol.*, **342**, 757–773.
55. Fletcher,J.E., Copeland,P.R., Driscoll,D.M. and Krol,A. (2001) The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA*, **7**, 1442–1453.

1.5. Additional results and discussion

SBP2 and the selenoprotein synthesis machinery in Drosophila willistoni

Although dietary selenium has been thought to be involved in normalizing the life span and fertility in Drosophila (Martin-Romero et al, 2001), selenoproteins do not seem to be essential for viability in this fly. Mutant flies that lack EFSec failed to decode the UGA Sec codon, but were viable and fertile (Hirosawa-Takamori et al, 2004). Furthermore, *in silico* comparative genomic approaches including 12 Drosophila genomes revealed that several insect species such as *Drosophila willistoni*, honey bee (*Apis mellifera*), wasp (*Nasonia vitripennis*), beetle (*Tribolium castaneum*) and silkworm (*Bombyx mori*), do not possess selenoproteins and have lost some of the factors involved in selenoprotein synthesis (Chapple & Guigo, 2008). In *D. willistoni*, SelK and SelH are cysteine orthologs. Moreover and very interestingly, the SelK mRNA contains a SECIS relic. EFSec is also absent in *D. willistoni*, and the sequence of the $tRNA^{Sec}$ is degenerated. An SBP2 homolog can be found, but it contains an amino acid variation in the conserved RNA binding region (Chapple & Guigo, 2008). SPS1 and Secp43 are present and highly conserved in *D. willistoni* but also in other Drosophila and insect species irrespective of their ability to encode selenoproteins. This could either mean an ongoing genetic drift or imply that these proteins play additional function(s) unrelated to selenoprotein synthesis (Chapple & Guigo, 2008). These data suggest the dispensability of selenoproteins in insects, and selenoprotein may therefore harbor important but non vital functions in insects, at least under the experimental conditions tested. We have mentioned the existence of an SBP2 homolog in *D. willistoni*. Its C-terminal region, containing the SECIS RNA-binding domain, is conserved as in the other Drosophila species (Chapple & Guigo, 2008) (Figure 18A). However, *D. willistoni* SBP2 contains one amino acid insertion in a critical region of the L7Ae SECIS RNA binding domain (Figure 18B). Indeed, a conserved 19 amino acid spacing is always observed in this domain of SBP2 between two acidic amino acid residues (Glu 679 and Glu 699 Figure 18 B and C). Position 679 is always occupied by a glutamic acid while position 699 is sometimes replaced by an aspartic acid. Surprisingly, the conserved spacing is not maintained in *D. willistoni* where insertion of an asparagine was observed (Figure 18B). The two glutamic acids are essential for SECIS recognition: it has been proposed that the first glutamic acid (Glu 679 in Figure 18C) contacts the conserved guanines of the sheared base-pairs and the second one (Glu 699 in Figure 18C) recognizes the conserved U in the non-Watson-Crick quartet (Allmang et al, 2002).

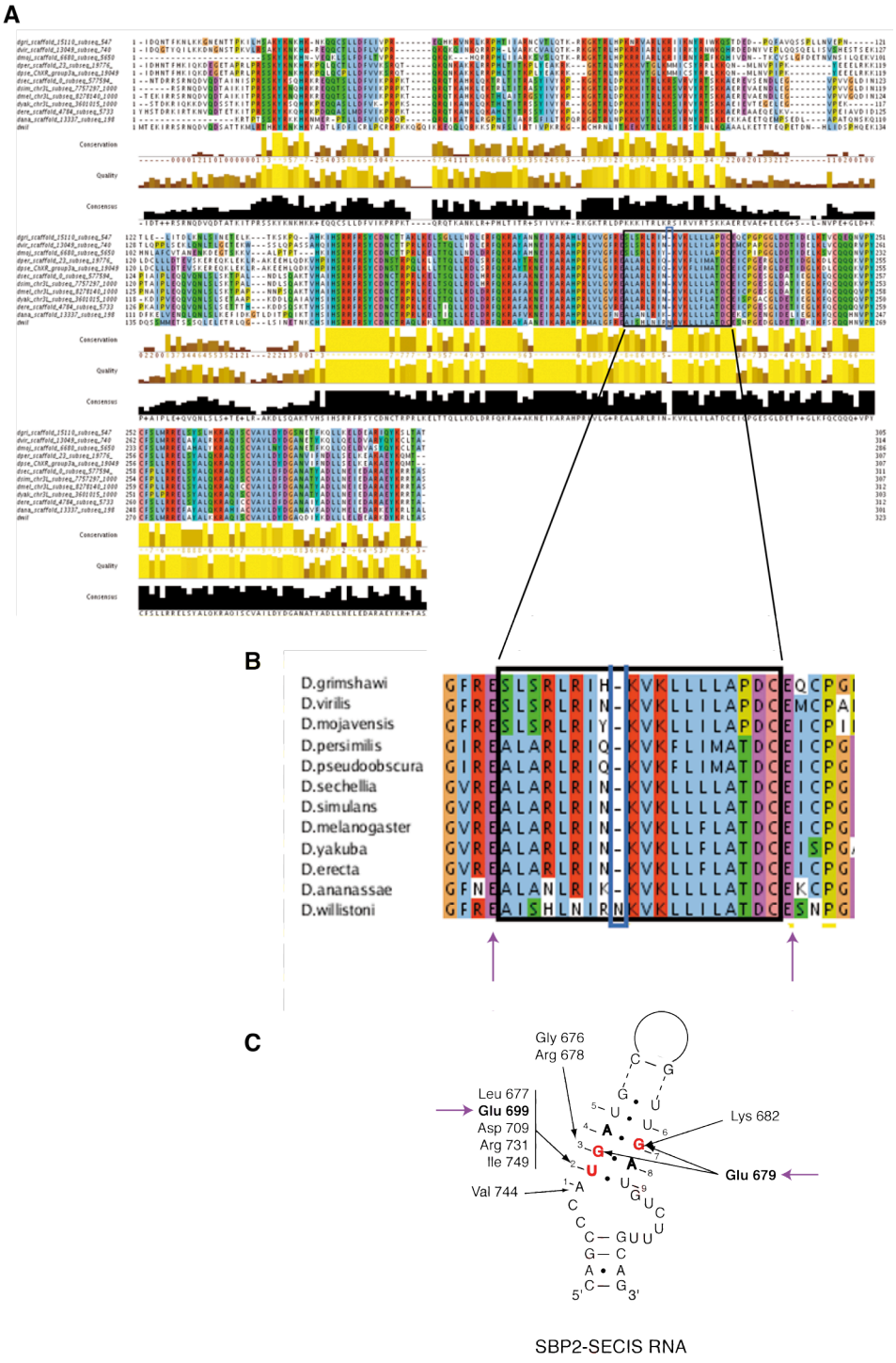


Figure 18. One amino acid insertion in *D. willistoni SBP2*.

A. Multiple sequence alignment of *Drosophila* SBP2 proteins. The conserved region containing one amino acid insertion in *D. willistoni* is boxed and shown in a blow up. **B.** Alignment of the conserved region in *Drosophila* SBP2. A 19 amino acid spacing is conserved between the two glutamic acids (purple arrows) except for *D. willistoni*. **C.** The two glutamic acids (Glu 679 and Glu 699, purple arrows) shown in A and B were predicted to recognize the conserved bases (shown in red) in the non-Watson-Crick quartet (Allmang et al, 2002). The alignments were taken from (Chapple & Guigo, 2008).

We therefore speculated that the single amino acid insertion between the two glutamic acids may affect SECIS RNA recognition. To test this hypothesis, we introduced an asparagine at the corresponding position of SBP2 and assayed the effect of the mutation on SECIS binding by gel-shift. Since this study was initiated before the isolation and functional characterization of *D. melanogaster* SBP2, the insertion was performed in human SBP2 (see Figure 19A). I generated two mutant constructs in human SBP2, inserting either an asparagine (as in *D. willistoni* SBP2*)* or an alanine (Figure 19B).

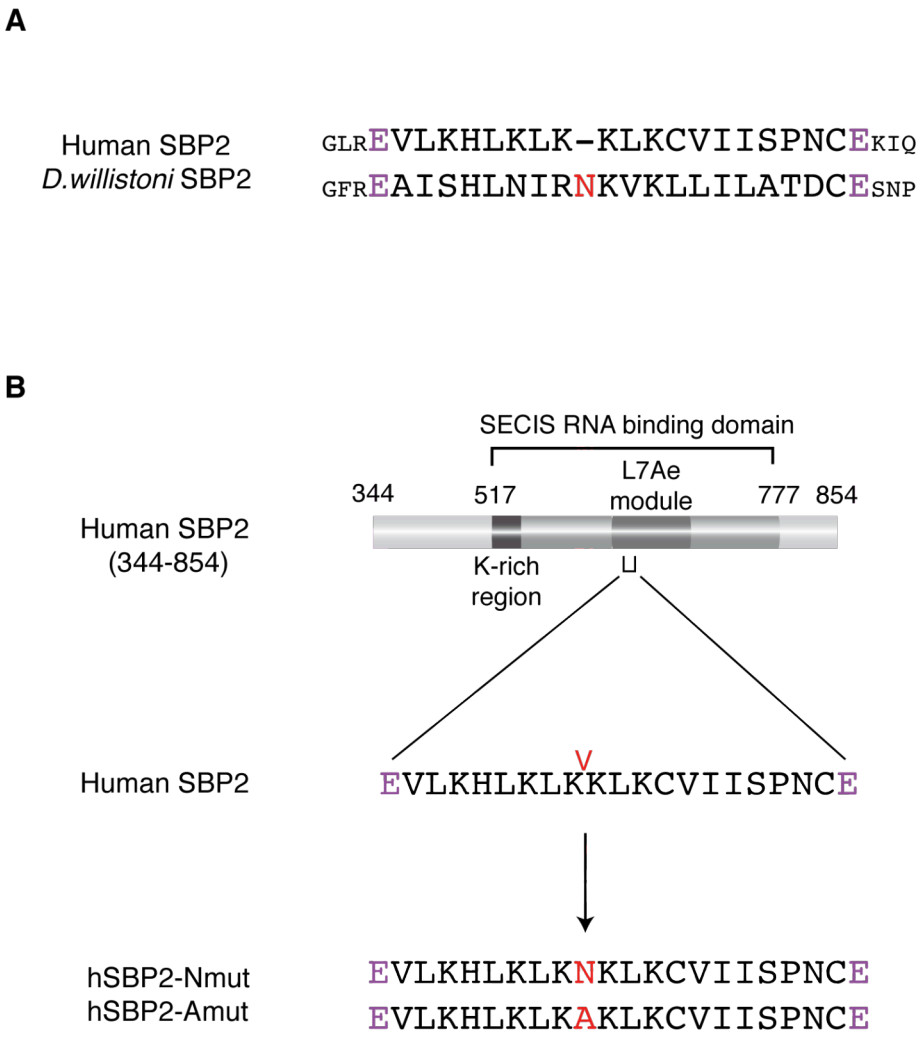


Figure 19. Aspargine insertion into the Glu 679-Glu 699 conserved spacing in human SBP2.

A. Sequence alignment of the Glu 679-Glu699 region between human and *D. willistoni* SBP2s. The conserved glutamic acids and the inserted asparagine in *D. willistoni* are shown in purple and red, respectively. **B.** Engineering of a Willistoni-like mutation in human SBP2. Either an asparagine or an alanine (shown in red) were inserted at the corresponding position of human SBP2.

The recombinant proteins were expressed in *E.coli* and purified using Ni-NTA agarose (Qiagen). The SECIS RNA binding abilities were tested by the electrophoretic mobility shift assay (EMSA) using three SECIS RNAs: two form 1 SECIS RNAs (GPx1 and SelN) and the PHGPx form 2 SECIS RNA (Figure 20).

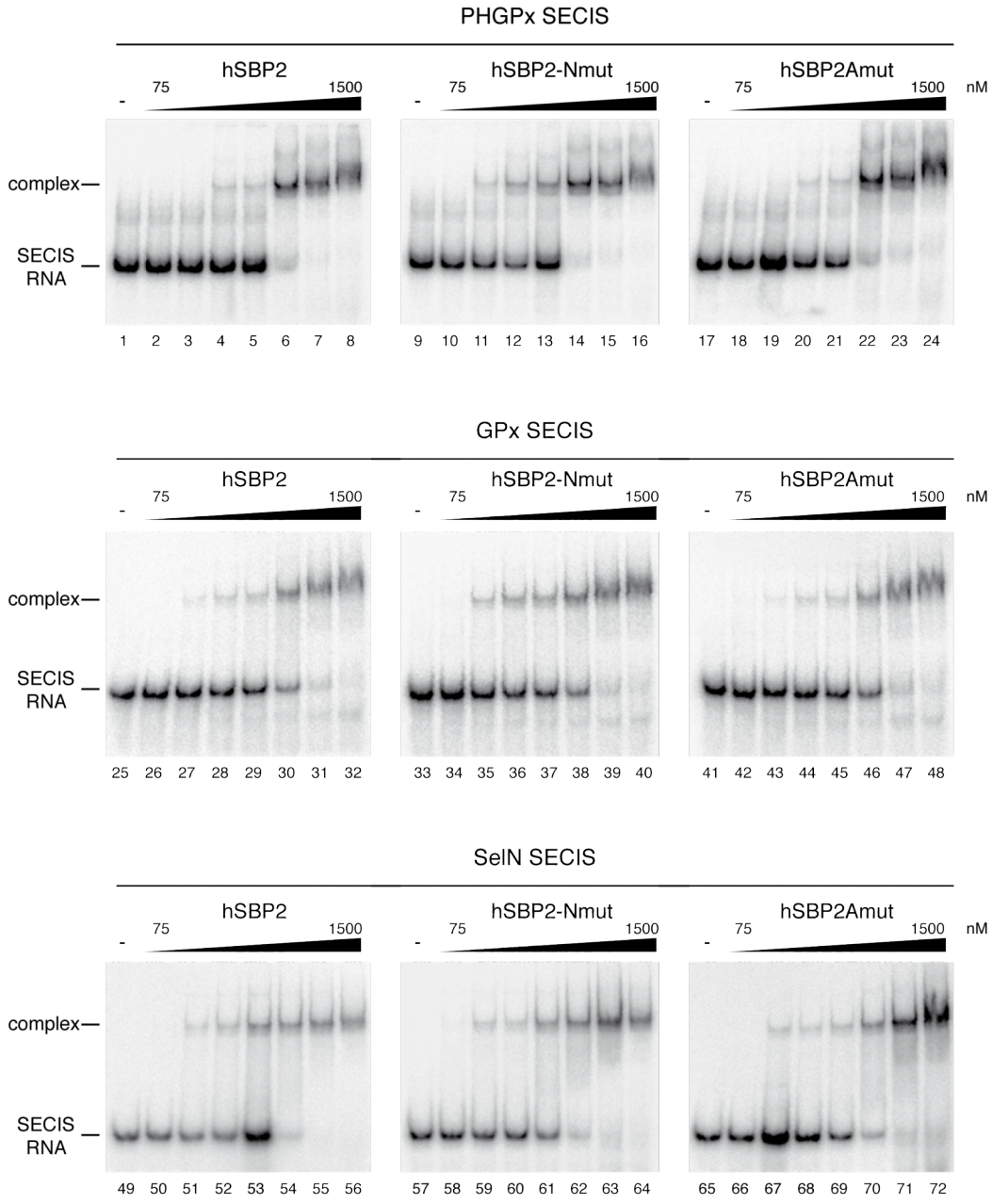


Figure 20. Disrupting the conserved Glu 679-Glu 699 spacing in hSBP2 by one amino acid insertion had no significant effect on SECIS-binding.

Electromobility shift assays were performed with purified hSBP2 mutant proteins and various *in vitro* transcribed ^{32}P-labeled SECIS RNAs (PHGPx, GPx and SelN). Increasing concentrations (75-1500nM) of SBP2 were added. SBP2 was omitted in the control lanes 1, 9, 17, 25, 33, 41, 49, 57 and 65 (-).

Surprisingly, both mutant proteins were able to bind all the SECIS RNAs tested, and no significant difference was observed with the wild-type SBP2. This result indicates that in contrast to our expectations, alteration of the spacing between the two glutamic acids has no major effect on hSBP2 SECIS RNA binding, under our experimental conditions. As shown in Article 1, however, the sequence of the K-rich domain differs substantially between Drosophila and all the other eukaryotes. It is thus possible that the K-rich domain of hSBP2 could compensate for the decreased affinity for SECIS RNAs caused by the one amino acid insertion. Other experiments will be required to determine whether the one amino acid insertion has a real impact on SECIS binding. The most straightforward would be to perform the insertion in the *D.melanogaster* SBP2 protein which displays a K-rich domain similar to that of *D. willistoni,* instead of the human SBP2. However, one cannot exclude the possibility that the insertion only reflects the genetic variety of the SBP2 gene in Drosophila without affecting its RNA binding ability *in vivo*. Also, the asparagine insertion could arise from an ongoing genetic drift concomitant to the selenoprotein gene disappearance in this Drosophila species. Interestingly*, D.willistoni* lives in a specific ecological niche in Brazil where the selenium content of the soil might be low. This could be an example of an adaptative evolution.

Except for this amino acid insertion, the C-terminal region of SBP2 is highly conserved across Drosophila species, independently of the selenoprotein coding capacity (Chapple & Guigo, 2008). Furthermore, the SBP2 gene was also found in the genomes of other selenoprotein-lacking insects, such as *A. mellifera* and *N. vitripennis* (Chapple & Guigo, 2008). SBP2 might also have other roles than selenoprotein synthesis so that SBP2 was retained in the genome of selenoprotein-lacking species (Chapple & Guigo, 2008). However, the conservation might also just reflect an ongoing genetic drift.

2. Toward crystallization of the SBP2/SECIS complex

2.1. Objective

The SBP2-SECIS interaction is at the heart of the selenocysteine incorporation mechanism. To get more insight into this mechanism, it is important to understand the specific features of this interaction at the atomic level. To this end, we started a crystallographic study of the SBP2-SECIS RNA complex in collaboration with the group of Philippe Dumas in the same research unit, UPR9002 of CNRS.

As discussed in Chapter 1 (in 1.4.), SBP2 is set apart from other L7Ae proteins: besides the L7Ae motif, SBP2 requires the additional domain (K-rich region) for SECIS recognition (Takeuchi et al, 2009). The SECIS element also presents unique functional and structural features that are different from canonical K-turn RNAs such as the long and stable helices 1 and 2 and the large internal loop (Cléry et al, 2007). Therefore, the SBP2-SECIS interaction principle has been proposed to be different from that of other L7Ae proteins/K-turn RNA complexes (Cléry et al, 2007 and unpublished results from the laboratory). Resolution of the crystal structure of the SBP2-SECIS complex is required for the validation of our model proposing that additional protein-RNA contacts must exist in the SBP2-SECIS complex, with respect to other complexes involving similar proteins and RNAs. Resolution of the structure would also be invaluable to understand the role of the K-rich region of SBP2 in SECIS RNA recognition as well as to establish whether the SECIS RNA adopts a K-turn like structure (Allmang et al, 2002).

Crystallization attempts were previously carried out in the laboratory and were the subject of the thesis of Vincent Olieric (team of Philippe Dumas). However, despite many trials, no crystal could be obtained, either of the protein alone or of the complex with the SECIS RNA. All the purification and crystallization attempts were performed using SBP2 proteins expressed in *E coli*. However, several problems were observed during protein expression and purification, such as a low level of protein production and sensitivity to proteolysis. We speculated that the unavailability of sufficient amounts of SBP2 may rely on the absence of post-translational modification. We therefore decided to clone the cDNA of other SBP2 proteins, fused with different tags and to express them in baculovirus-infected insect cells. This would hopefully improve protein expression and enable the obtention of proteins produced in a eukaryotic system for further crystallization assays. My participation in this

study was to generate various SBP2 constructs compatible with expression in insect cells using the baculovirus expression system. I also determined the best conditions for protein expression with the help of the Plateforme de Génomique et Biologie Structurales at the Centre Européen de Biologie et Génomique Structurales (CEBGS) and the baculovirus service at IGBMC in Illkirch. In collaboration with the group of Philippe Dumas, various biophysical techniques as well as sequence analysis were used in addition to the crystallographic attempts to better understand the structure of SBP2.

2.2. Results

2.2.1. cDNA cloning of SBP2 from various organisms

Because only human SBP2 had been used so far for crystallographic attempts, we set out to clone and express SBP2 from rat and Drosophila (see Figure 21). In addition to rat SBP2, that has been used by others for experimental studies (Bubenik & Driscoll, 2007; Caban et al, 2007; Chavatte et al, 2005; Copeland et al, 2000; Copeland et al, 2001; de Jesus et al, 2006; Donovan et al, 2008; Gupta & Copeland, 2007; Kinzy et al, 2005; Small-Howard et al, 2006) and that is highly similar to human SBP2 in both length and sequence (Lescure et al, 2002), we also selected the newly identified Drosophila SBP2 (Takeuchi et al, 2009). Because of its shorter length, we postulated that it would be produced more efficiently, and that it would be especially more soluble than the longer mammalian SBP2s, both criteria being important for the crystallographic study.

A previous work in the laboratory, using different constructs of human SBP2 bearing a thrombin cleavage site between the tag and the actual SBP2 sequence, revealed the occurrence of a sensitive cleavage site after Lys525 during the purification steps. Although this purification procedure included the treatment with thrombin to separate SBP2 from the tag, the cleavage after Lys525 did not seem to be provoked by thrombin because of the following observations. The sequence $EIPK_{525}AKK$ does not match the consensus thrombin cleavage site, and increasing concentrations of thrombin did not modify the cleaved/uncleaved ratio. In addition, the constructs used in the previous work led to low solubility. To avoid proteolysis, we resorted to a new construct starting at Ala526 (hereafter

called hSBP2Δ525), thus removing the cleavage site. In addition, this construct retains all the previously identified functional RNA binding motifs (see Figure 21).

I therefore generated eight constructs encoding SBP2 proteins from various organisms: human SBP2 full-length (hSBP2), human SBP2 526-854 (hSBP2Δ525), rat SBP2 and *Drosophila. melanogaster* SBP2 (dSBP2). These constructs were fused with either 6xHis or GST-tags and incorporated a thombin cleavage site (Figure 21). The constructs were engineered using the GATEWAY technology (Invitrogen) with the help of the Plateforme de Génomique et Biologie Structurales at the CEBGS. The principle of the GATEWAY technology and the methods for SBP2 cDNA cloning are described in Annex/Methods 1.cDNA cloning using the GATEWAY Technology.

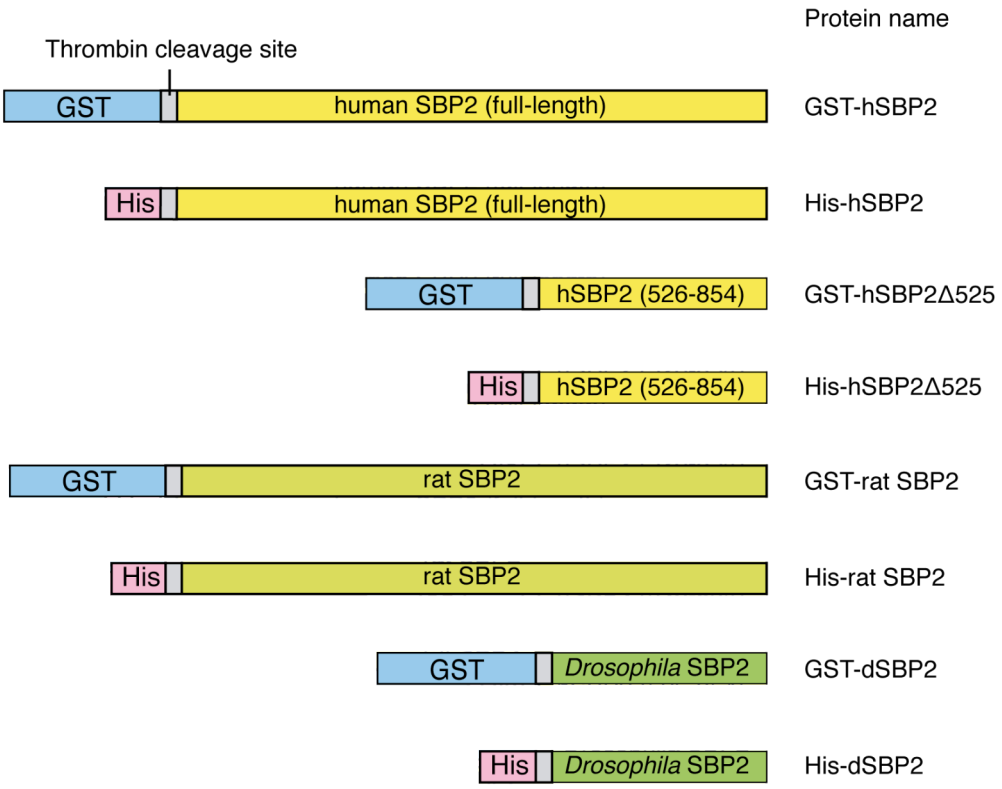


Figure 21. Schematic drawings of SBP2 proteins designed and generated for crystallization purposes.

The human, rat and *D. melanogaster* SBP2 sequences were fused with either a 6xHis or a GST-tag and incorporated a thrombin cleavage site (gray box).

2.2.2. Expression of various SBP2 cDNAs using the Baculovirus expression system

For expression of these SBP2 proteins, we used the Baculovirus expression vector system (Invitrogen). Baculoviruses are double-stranded, circular, supercoiled DNA molecules in a

rod-shaped capsid, and affect the insect population. Expressing proteins in baculovirus-infected insect cells has important advantages for protein production. First, the baculovirus expression system often achieves high levels of protein expression. Second, the insect cell lines used for this system can be used for large-scale culture. Finally, this system enables to express proteins that are post-translationally modified in a manner similar to that of mammalian cells (folding, disulfide bond formation, oligomerization, acylation and proteolytic cleavage), that are biologically active and functional. Protein expression in insect cells using the Baculovirus expression system consists of several steps: preparation of a Baculovirus shuttle vector (Bacmid), generation of virus particles, culture of baculovirus-infected insect cells and mini-expression tests. The methods used for these steps are described in Annex/Method 2. Baculovirus expression system. Ten recombinant bacmid clones (named GST-hSBP2a and b, GST-hSBP2Δ525, GST-ratSBP2a and b, GST-dSBP2, His-hSBP2, His-hSBP2Δ525, His-ratSBP2 and His-dSBP2) were obtained and they were transfected into insect cells to generate recombinant baculovirus particles.

Mini expression test

Recombinant baculoviruses were first tested using small-scale expression tests in order to assess their ability to trigger protein expression. None of the His-tagged SBP2 constructs tested yielded efficient expression. Only the His-ratSBP2 and His-hSBP2 proteins were detected in insoluble fractions by western blot analysis using anti-hSBP2 antibody (Figure 22 lanes 18 and 20). GST-tagged constructs allowed the synthesis of low levels of soluble protein. This was the case for GST-ratSBP2a and b, GST-hSBP2Δ525 and GST-hSBP2a and b (Figure 22 lanes 24, 26, 28, 30 and 32). Protein bands corresponding to GST-hSBP2Δ525, GST-ratSBP2a and b were further analyzed by mass spectrometry and confirmed to be SBP2 proteins (MS performed by P. Wolff, data not shown). We therefore selected the GST-hSBP2b, GST-hSBP2Δ525 and GST-ratSBP2b recombinant baculoviruses for large scale expression of the corresponding proteins.

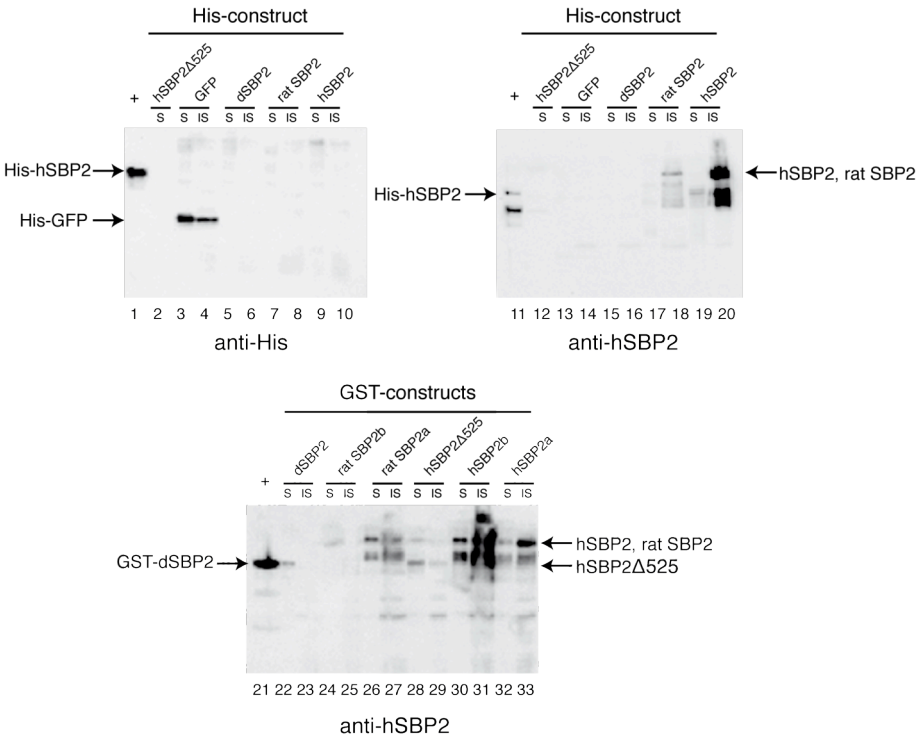


Figure 22. Mini expression assays in baculovirus infected insect cells.

Proteins (indicated above the lanes) were expressed in Sf9 insect cells. Cell extracts were incubated with either nickel or glutathione sepharose beads. Bound proteins were analyzed by Western blotting with anti-His or anti-SBP2 antibodies. S: Soluble fraction after incubation with beads. IS: Insoluble fraction. His-tagged GFP expressed in Sf9 cells (GFP), recombinant His-tagged human SBP2 (His-SBP2) or GST-tagged Drosophila SBP2 were used as positive controls.

Expression tests in medium volume culture

After amplification and titration of the viral particles, they were used for the final round of selection according to the protein expression levels achieved in a medium scale culture (25ml). Figure 23 (upper panels) shows the western blotting of soluble fractions using anti-hSBP2 antibody. While GST-hSBP2Δ525 was expressed and purified almost equally well under any condition tested (Figure 23 A and B lanes 7-12, 27-32 and 45-50), expression of GST-hSBP2b and GST-ratSBP2b was higher after 48 hours incubation than 72 hours (Figure 23 A, compare lanes 1-3 and 14-16 with lanes 4-6 and 17-19, respectively). However, neither GST-hSBP2b nor GST-ratSBP2b could be recovered in sufficient amount after GST-purification (Figure 23B middle and bottom panels, lanes 21-26, 33-38, 39-44 and 52-57). Therefore, the GST-hSBP2Δ525 bacmid-containing baculovirus was selected for large scale (5L) expression of SBP2 protein for crystallization purposes.

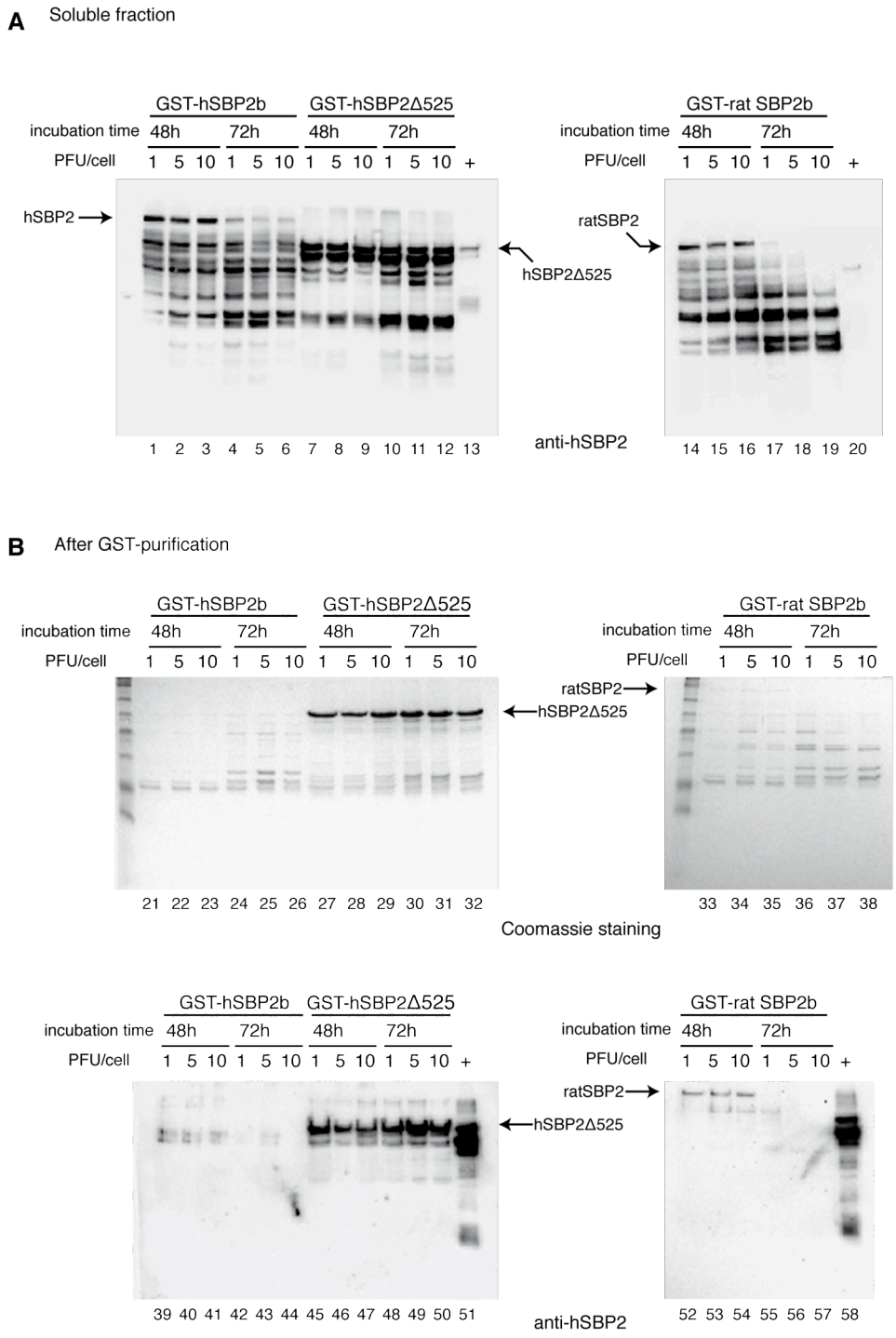


Figure 23. Expression tests from large-scale culture of baculovirus infected insect cells.

GST-hSBP2b, GST-hSBP2Δ525 and GST-ratSBP2b were expressed in Sf9 cells under different conditions. **A.** Cell extracts were analyzed by Western blotting with anti-SBP2 antibody. **B**. Cell extracts were incubated with glutathione sepharose beads. Bound proteins were analyzed by Coomassie staining (upper panels) and Western blotting with anti-SBP2 antibody (lower panels).

2.2.3. Biophysical analysis of SBP2

Our initial aim, when producing SBP2 in baculovirus-infected insect cells, was to improve protein expression and stability and therefore the ability of the protein to crystallize. However, no crystal could be obtained even with the protein produced under the conditions described in this chapter.

Interestingly, the use of structure prediction algorithms suggested the existence of disordered areas in SBP2. It appears now that a growing number of eukaryotic genes encode proteins that lack three-dimensional folding but become folded upon binding to their targets. Such proteins are called Intrinsically Disordered Proteins (IDPs).

Unstructured protein domains differ from structured proteins in many properties, such as amino acid compositional bias, high apparent molecular mass, heat resistance, acid resistance and extreme proteolytic sensitivity. For amino acid composition, unstructured domains are enriched in disorder-promoting amino acids (A, R, G, Q, S, P, E and K) and comprise less order-promoting amino acids (W, C, F, I, Y, V, L and N) (reviewed in Dyson & Wright, 2005; Rajkowitsch et al, 2007; Tompa, 2002; Wright & Dyson, 1999). The analysis of folded and unfolded proteins based on the normalized net charge and mean hydrophobicity revealed that IDPs are usually localized within a unique region of the charge-hydrophobicity phase space. Indeed, a combination of high mean net charge and low overall hydrophobicity represents a unique structural feature of IDPs (Uversky et al, 2000). A number of computer programs such as PONDR, DISOPRED2, DisEMBL GlobProt2 and IUPred allow the prediction of disordered regions based on the amino acid sequence (reviewed in Dyson & Wright, 2005; Rajkowitsch et al, 2007).

We therefore decided to test whether SBP2 is an IDP. Indeed, SBP2 does exhibit several features often observed with unfolded proteins. For example, the SBP2 sequence is rich in proline residues, one of the disorder promoting residues, and the percentages of proline contents are 7.3% for rat SBP2 and 8.6% for human SBP2, figures that are higher than for the average of the Swiss-Prot proteins (4.9%) (Linding et al, 2003; Tompa, 2002). This is consistent with the early experimental results in the laboratory that SBP2 exhibited unusually higher molecular mass in gel filtration experiments and SDS-PAGE than those measured by mass spectrometry. Computer predictions were performed using the various forms of SBP2 proteins that I described in this chapter. Interestingly, the disorder-prediction methods showed that ca. 70 % of the SBP2 sequence is disordered, whereas the L7Ae RNA binding domain

appears to be folded. We therefore decided to perform biophysical analyses to test this hypothesis. The results are described in the following article (2.3. Article 2)

Since no high-resolution X-ray structures of intrinsically disordered proteins are available due to the lack of stable three-dimensional folds, a number of other experimental methods are used to elucidate disordered domains. Nuclear magnetic resonance (NMR), circular dichroism (CD), differential scanning calorimetry and fluorescence anisotropy that are based on structure determination in solution, distinguish between structured and unstructured domains (reviewed in Rajkowitsch et al, 2007). Several hydrodynamic techniques can also detect unfolded conformations. Gel filtration (size-exclusion chromatography), Small-Angle-X-ray Scattering (SAXS), sedimentation analysis and dynamic light scattering provide information in hydrodynamic parameters that can detect characteristic behaviors to intrinsically disordered proteins (reviewed in Tompa, 2002).

We have used several of these biophysical techniques such as ^{1}H 1D-NMR, analytical centrifugation and dynamic light scattering (DLS). Results are described In Article 2.

They reveal that SBP2 shows an unusually high molecular weight in dynamic light scattering (DLS) experiments, supporting the non-globular character of the rat SBP2 and human SBP2Δ525. Ultracentrifugation analysis showed that a monomeric SBP2 protein is in part unstructured. ^{1}H 1D-NMR, a technique that can differentiate a folded from a non-folded protein in the resonance range characteristic of peptide bonds, showed the lack of peptide bonds and methyl resonances of the rat SBP2 and human SBP2Δ525 expressed in the bacterial and the eukaryotic systems. All these results show that SBP2 is largely unfolded, and is indeed a member of the family of Intrinsically Disordered Proteins. We postulated that the reasons for the previous unsuccessful crystallization may rely on the presence of disorder in the protein.

2.3. Article 2 (in press)

SECIS-binding protein 2, a key player in selenoprotein synthesis, is an intrinsically disordered protein

Vincent Oliéric, Philippe Wolff, Akiko Takeuchi, Guillaume Bec, Catherine Birck, Marc Vitorino, Bruno Kieffer, Artemy Beniaminov, Giorgio Cavigiolio, Elizabeth Theil, Christine Allmang, Alain Krol and Philippe Dumas (2009) *Biochimie 91(8): 1003-1009*.

Biochimie 91 (2009) 1003–1009

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

SECIS-binding protein 2, a key player in selenoprotein synthesis, is an intrinsically disordered protein

Vincent Oliéric [a,1], Philippe Wolff [a], Akiko Takeuchi [a], Guillaume Bec [a], Catherine Birck [b], Marc Vitorino [b], Bruno Kieffer [b], Artemy Beniaminov [a,c], Giorgio Cavigiolio [d], Elizabeth Theil [d,e], Christine Allmang [a], Alain Krol [a], Philippe Dumas [a,*]

[a] Architecture et Réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC, 67084 Strasbourg, France
[b] Institut de Génétique et de Biologie Moléculaire et Cellulaire, CEBGS, 1 rue Laurent Fries, BP 10142, 67404 Illkirch Cedex, France
[c] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov Street, Moscow 119991, Russia
[d] CHORI Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, USA
[e] Department of Nutritional Sciences and Toxicology, University of California, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 2 April 2009
Accepted 12 May 2009
Available online 23 May 2009

Keywords:
Selenocysteine
Recoding
SBP2
SECIS
Intrinsically disordered protein

ABSTRACT

Selenocysteine (Sec) is co-translationally incorporated into selenoproteins at a reprogrammed UGA codon. In mammals, this requires a dedicated machinery comprising a stem-loop structure in the 3′ UTR RNA (the SECIS element) and the specific SECIS Binding Protein 2. In this report, disorder-prediction methods and several biophysical techniques showed that *ca.* 70% of the SBP2 sequence is disordered, whereas the RNA binding domain appears to be folded and functional. These results are consistent with a recent report on the role of the Hsp90 chaperone for the folding of SBP2 and other functionally unrelated proteins bearing an RNA binding domain homologous to SBP2.

1. Introduction

Selenocysteine is found in the active site of selenoproteins that are involved in oxidation–reduction reactions (reviewed in Ref. [1]). Although selenocysteine is co-translationally incorporated into proteins, the insertion step is by no means standard since it requires the recoding of an in-frame UGA codon to mean selenocysteine instead of stop. In eukaryotes, this process is achieved by a complex machinery for which not all the partners have been identified yet (reviewed in Ref. [2]). The well-established facts are that (i) a stem-loop structure located in the 3′-UTR of selenoprotein mRNAs, the SElenoCysteine Insertion Sequence (SECIS), is mandatory for recognizing UGA as a selenocysteine codon and (ii), that protein SBP2 (for SECIS Binding Protein 2) specifically binds the SECIS element [3]. In addition, it is very likely that SBP2 recruits the $tRNA^{Sec}$-bound specialized translation elongation factor EFSec [4–6]. Ribosomal protein L30 has also been implicated in selenoprotein synthesis by recognizing the SECIS RNA as well [7], but the actual mechanistic issues of how the SECIS-bound complex reaches back to the UGA Sec codon at an approaching ribosome are far from being elucidated (reviewed in Ref. [8]). Interestingly, the RNA binding domain of SBP2 comprises a subdomain that belongs to the L7Ae/L30 family. The latter includes ribosomal proteins [9], but also protein components of spliceosomal small nuclear RNPs (snRNPs) and small nucleolar RNPs (snoRNPs) such as 15.5 kD/Snu13p and Nhp2p [10,11]. Most if not all of these proteins recognize K-turn RNAs, as first disclosed in the crystal structure of the complex formed by the U4snRNA and the spliceosomal 15.5 kD protein [12]. It has been proposed [13,14] that the SECIS RNA also adopts a K-turn like structure and interacts with SBP2. Additional contacts exist between SBP2 and the SECIS RNA compared to other members of the L7Ae family. Indeed, it was recently demonstrated that specific recognition is also provided by a second domain lying N-terminal to the L7Ae module. Called bipartite, SID or K-rich [15–17], it is crucial and part of the SBP2 RNA binding domain but is absent in other L7Ae proteins.

The 'structure–function relationship' has been a central dogma for decades in the protein field. As a result, an absence of folding has long been considered as an experimental artefact during the purification step and a cause of aggregation or proteolysis, rather than of biological significance. However, it is now clear that lack of

* Corresponding author. Tel.: +33 388 41 70 02; fax: +33 388 60 22 18.
E-mail address: p.dumas@ibmc.u-strasbg.fr (P. Dumas).
[1] Present address: Swiss Light Source at Paul Scherrer Institute, 5232 Villigen, Switzerland.

0300-9084/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2009.05.004

folding is not synonymous of lack of biological activity. This emerged from studies like that on a cyclin-dependent kinase (Cdk) inhibitor (p21) that is unstructured in absence of its target, but becomes structured upon binding to it [18]. Since then, a large body of other examples has been identified [19–22] and the role of structural disorder in the function of RNA and protein chaperones has been reviewed [23]. A search for sequence determinants preventing spontaneous folding led to a criterion as simple as considering the mean hydrophobicity of a sequence *vs.* its mean net charge thus reducing any protein to a particular point in a 2D diagram. This was shown to delineate a frontier between folded and natively unfolded proteins [21]. Independently, a 'sequence complexity' index was proposed, which showed that depletion in hydrophobic residues and enrichment in hydrophilic and charged residues are factors of disorder [24]. Other bioinformatics methods, as well as several experimental methods, are available to assess whether or not a protein is natively unfolded (for review, see Ref. [25]).

At the onset, we attempted to crystallize the SBP2–SECIS RNA complex to understand how different proteins possessing homologous RNA binding domains can discriminate the cognate K-turn containing RNA. Despite many efforts, we failed at obtaining crystals, be it of the complex or of SBP2 alone. In this paper, we show by using sequence analysis and biophysical techniques that SBP2, being largely unfolded, is a member of the family of Intrinsically Disordered Proteins (IDP, or Intrinsically Unfolded Protein according to another nomenclature).

2. Materials and methods

2.1. DNA constructs

The coding region of full length rat SBP2 (kindly provided by P. Copeland) was subcloned into STBlue-1 vector (Novagen), after adding NdeI and SalI restriction sites by PCR, followed by subcloning into pET-28a(+) to create plasmid FLSBP2. When the N-ter 6×His-tagged SBP2 protein was expressed from FLSBP2 in *Escherichia coli* Rosetta 2 (DE3) (Novagen), it was found in the soluble fraction, contrasting with earlier studies using other vector/host combinations [3,26,27].

The human SBP2 △525 (lacking the N-terminal 525 amino acids) was cloned into pET32b vector (Novagen). The resulting plasmid encodes an N-terminal thioredoxin-fusion protein with a thrombin cleavage site. The same construct was later subcloned using the Gateway Technology and Baculovirus Expression Vector System (Invitrogen) for expression in baculovirus-infected cells. The final vector pDEST20-SBP2 △525 encodes a GST-SBP2 fusion protein. A baculovirus shuttle vector (bacmid) was generated by transforming pDEST20-SBP2 △525 into *E. coli* DH10Bac.

2.2. SBP2 expression and purification

Bacteria grown in LB medium and expressing the full length 6xHis-tagged rat SBP2 were sonicated, and the supernatant fraction clarified by sedimentation before application to an anion-exchange column (PorosHQ, Applied Biosystems). The protein was then sequentially purified on hydroxyapatite (CHT-2, Biorad), heparin (GE Healthcare) and a final anion-exchange column (UnoQ, Biorad). The protein was kept in 50 mM sodium phosphate buffer pH 7.8, 200 mM NaCl, 1 mM DTT, 1 mM EDTA.

The human SBP2 △525 protein was purified by cobalt affinity (Talon, Clontech) and anion-exchange (PorosHQ) chromatography, followed by a 15 h thrombin digestion (Sigma) at 4 °C. It was next concentrated on a Centricon 30 K (Millipore) and size exclusion chromatography (Superdex 200, GE Healthcare). The final buffer was 20 mM HEPES-NaOH pH 7.8, 150 mM NaCl, 20% glycerol, 10 mM β-mercaptoethanol.

The human SBP2 △525 was also expressed using Sf9 (*Spodoptera frugiperda*) cells that were infected with recombinant baculoviruses. Infected cells were cultured in TNM-FH supplemented with 10% fetal calf serum and 50 μg/ml gentamycin at 27 °C for 72h. The GST-SBP2 △525 protein was purified on glutathione sepharose 4B (GE Healthcare) and heparin columns. After thrombin cleavage, the mixture of GST–lacking SBP2 △525 and uncleaved proteins was ultimately purified again on glutathione sepharose 4B followed by another heparin column. The final buffer was 25 mM sodium phosphate pH 8, 200 mM NaCl, 1 mM DTT.

2.3. Gel retardation assay

The ^{32}P-labeled SECIS RNA–SBP2 complex was formed and electrophoresed on a 5% non-denaturing polyacrylamide gel according to standard procedures [10].

2.4. DLS experiments

Dynamic Light Scattering (DLS) experiments were performed with a DynaPro-801 (Protein solutions Ltd, High Wycombe, UK) with the protein sample at *ca.* 1 mg ml^{-1} in the buffer of the last purification step (see Section 2.2) containing 1 mM DTT as the reducing agent.

2.5. Ultracentrifugaton analysis

Centrifugation experiments were performed at 4 °C by recording the absorbance profiles at 280 nm in an Optima XL-A analytical ultracentrifuge (Beckman-Coulter, Palo Alto, CA). The full-length protein concentration was 0.1 mg ml^{-1} in HEPES–NaOH 20 mM, pH 7.8, NaCl 200 mM, EDTA 2 mM, TCEP–HCl 2 mM (TCEP: Tris(2-CarboxyEthyl) Phosphine) as the reducing agent, and not DTT which absorbs at 280 nm. The calculations were made with the programs SEDPHAT and SEDNTERP for sedimentation equilibrium experiments and with SEDFIT for sedimentation experiments [28]; the viscosity of the solution and the protein specific volume were calculated with the program SEDNTERP [29].

2.6. NMR experiments

600 MHz ^{1}H spectra were collected at 10 °C on a Bruker DRX600 spectrometer equipped with a cryoprobe. The protein sample was at 100 μM in the conditions of the last purification step (see Section 2.2) supplemented with 8% D_2O.

2.7. Disorder prediction algorithms and sequence alignment

Several disorder predictions methods were used: IUPred [30], VL3H [31] and VSL2 [32]. Sequence alignment was performed with clustalw [33] and the figure with the program espript [34].

3. Results

Analysis of the human and rat SBP2 amino acid sequences led to a consistent pattern with *ca.* 70% of the sequence predicted as unfolded (Fig. 1A). It may be underlined that SBP2 contains 73 proline residues among 846 in total, which represents an unusually large fraction (8.6%) in comparison to the mean fraction in the Swiss-Prot database (4.9%). This is significant since proline is ranked first in one classification among the amino acids responsible for unfolding [35,36]. Nevertheless, the residue range 605–775, comprising the L7Ae/L30 module (part of the RNA-binding domain), is clearly

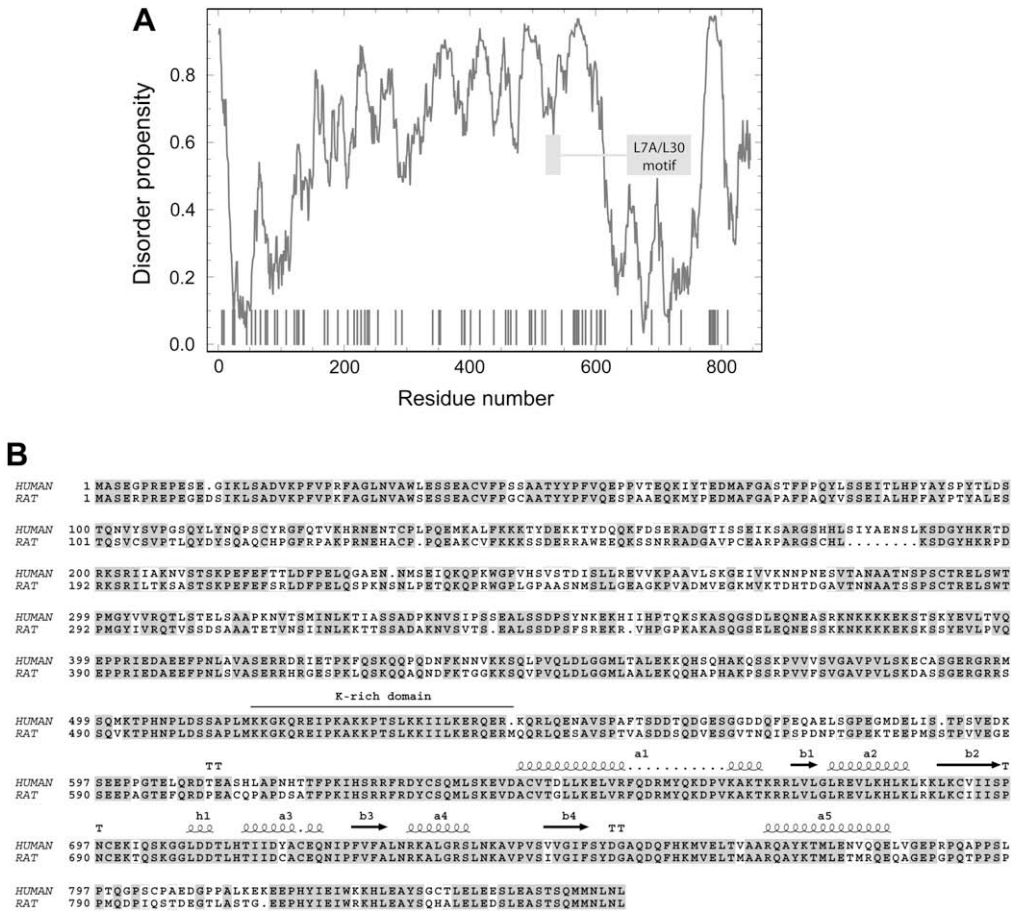


Fig. 1. Structured part of SBP2 and SECIS binding. (A) Result of folding prediction for the rat full-length SBP2 with IUPred [30] highlighting an essentially disordered protein with two ordered regions, one at the N-terminus, the second one between residues 605 and 775 containing the L7Ae/L30 RNA-binding domain. Note that the full RNA-binding domain is bipartite (grey rectangles) with a short N-term K-rich domain (see text) and that the prediction for human SBP2△525 is not shown since it is strictly comparable. Analogous disorder predictions were also obtained with other methods like VL3H [31] and VSL2 [32]. The vertical bars highlight proline positions that are in two-fold excess compared to the average content in proteins. (B) Sequence alignment of the human and rat SBP2 proteins. The K-rich domain is marked as well as the predicted secondary structure of the L7Ae RNA-binding domain derived from the crystal structure of the U4snRNA–15.5 kD complex (pdb code 1E7K) [12].

predicted as being structured; accordingly, the fraction of proline residues drops to 4.1% in this range. We set up to experimentally demonstrate these predictions. The experimental work was initiated by testing different constructs of human SBP2 bearing a thioredoxin N-terminal extension separated from the SBP2 sequence by a thrombin cleavage site. It was repeatedly observed that during the purification procedure, including the treatment with thrombin, a rather intense cleavage occurred after Lys525. However, thrombin was not responsible for this additional cleavage since (i) the sequence EIPK$_{525}$AKK does not match the consensus thrombin cleavage site and (ii) an important increase in thrombin concentration did not modify the cleaved/uncleaved ratio. In addition, the full-length human SBP2 led to low solubility. We thus turned our attention to a new construct starting at Ala526, immediately past the cleavage site (hereafter called hSBP2△525). A full-length soluble SBP2 could however be obtained with the rat construct that showed a cleavage too, but less pronounced (at least under our experimental conditions). This is somewhat surprising since the cleavage occurs in a highly conserved region (Fig. 1B). The functional analysis presented in this work was therefore performed with both the rat full-length SBP2 and the human SBP2△525 fragment.

In a first step, we demonstrated the ability of both the rat full-length SBP2 and the human SBP2△525 fragment (containing the L7Ae/L30 domain [10]) to bind the wild-type SECIS in a gel-shift assay (Fig. 2A). The rat SBP2, cleaved at residue 525, is responsible for the intermediate band marked with an asterisk. The binding was specific as it was severely impaired (Fig. 2B) by mutations in the non-Watson–Crick base pairs of the SECIS previously reported to inhibit SBP2 interaction [37]. This positive interaction data is consistent with the fact that both proteins contain the L7Ae module predicted to be properly folded. However, according to the predictions (Fig. 1A), the hSBP2△525 fragment still contains nearly 50% of disordered residues. It also contains the additional small patch of residues (from 525 to 545), named bipartite-SID or K-rich motif (Fig. 1B), located 120 amino acids upstream from the L7Ae/L30 domain and recognized as necessary for SECIS binding [15–17]. The intervening 120 amino acids are predicted as disordered. It may be hypothesized that (at least) the small patch of residues becomes ordered upon SECIS binding.

These computer predictions leading to conclude that SBP2 is an IDP were consistent with our early experimental results with the rat full-length protein (MW = 95.3 kD) and the human SBP2△525 fragment (MW = 37.3 kD). In both cases, protein solutions dialysed against ammonium acetate (10 mM, pH 6.8) yielded the expected MW by ESI-MS, but anomalous ones were obtained from gel filtration experiments on Superdex 200 (300 kD and 100 kD, respectively). We also observed an abnormal migration on SDS-PAGE corresponding to 125 kD and 50 kD, respectively, which had already been noticed for SBP2 and several of its fragments [26]. This is a feature often observed with IDP owing to their abnormal amino

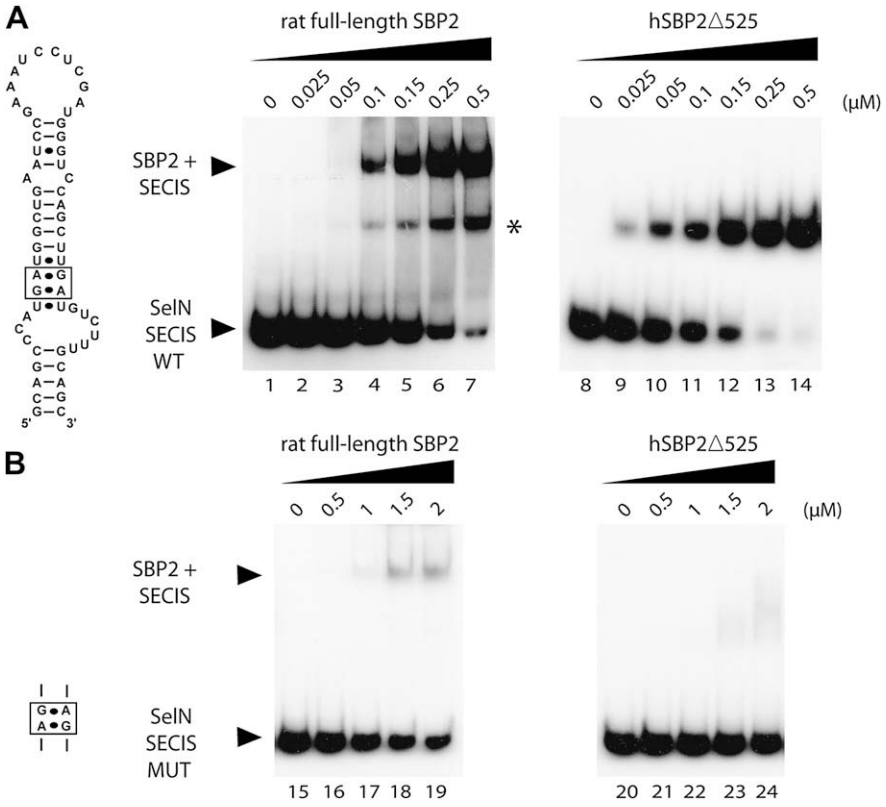


Fig. 2. Gel retardation assays. (A) Gel retardation of the complexes formed between the wild-type SelN SECIS RNA with either the full-length rat SBP2 (lanes 1–7) or the human SBP2△525 (lanes 8–14). The secondary structure of the wild-type SelN SECIS RNA is shown in full with the mutation site boxed. The asterisk represents a complex formed with a fragment resulting from a spontaneous cleavage of the rat SBP2, in good agreement with the shifted band obtained with the human SBP2△525 at the same level on the right. (B) A mutation of the SelN SECIS (shown on the left) severely impaired binding of the two proteins (lanes 15–19 and 20–24). Note the increase in protein concentrations necessary to induce a very faint binding of the mutated RNA.

acid composition [36]. Analogous results were obtained by Dynamic Light Scattering (DLS) since the raw diffusion coefficients led to 345 kD and 130 kD, respectively, with a rather low dispersity index in each case (12% and 9%). Therefore, these too high MW values from DLS are essentially those of single species in solution and not of polydisperse aggregates.

Ultracentrifugation analysis was performed in order to go beyond these simple pieces of evidence in favour of the non-globular character of the rat SPB2 and the human fragment hSBP2△525. We first performed sedimentation equilibrium experiments for which hydrodynamical properties are without importance. SBP2 was centrifuged for 22 h at 4 °C at three different velocities (9000, 12 000 and 15 000 rpm). Using a value of 0.74 ml g^{-1} for the protein specific volume, the three absorbance profiles were well fit with a common MW value taken equal to the calculated value of 95 326 Da (Fig. 3A). The same quality of agreement was observed with hSBP2△525 centrifuged at 20 000 rpm, but in that case it was necessary to assume that up to 17% of the protein molecules were engaged in multimers containing six monomers in average (not shown). Sedimentation velocity analysis was then used to evaluate hydrodynamic parameters. A distribution of sedimentation coefficients was obtained for the full-length protein (Fig. 3B) with a major peak (90%) around 1.8 S (1 Svedberg = 10^{-13} s), which would agree with a globular protein of roughly 30 kD, but certainly not 95 kD. Interestingly, a minor component (10%) was also present at 4.2 S, and this would agree with a molecular species of roughly 100 kD. This is not sufficient, however, to conclude firmly that this corresponds to folded monomers. The frictional ratio f/f_0, where f is the actual frictional coefficient of the protein and f_0 the frictional coefficient of a sphere of equal volume, was determined to be roughly equal to 2.28. For an elongated ellipsoidal rigid body this would correspond to an axial ratio of 25 (Ref. [38]). These results are in complete agreement with a monomeric protein being in part unstructured.

We also made use of ^{1}H 1D-NMR, which can differentiate easily a folded from a non-folded protein in the resonance range characteristic of peptide bonds. This appears as a spectral shrinkage from 6–11 ppm to 7.5–9 ppm and a suppression of the methyl signal around 0 ppm when going from a folded to a disordered protein with no secondary structures (see for example Ref. [39]). The results, again, were clearly in favour of a disordered state for SBP2 (compare Fig. 4a and b). The same was true for hSBP2△525 (Fig. 4c). However, a small amount of methyl and peptide peaks was indicative of a slight increase of folding, may be as a consequence of the lack of the first 525 residues that are mostly predicted as disordered. Notably, in both cases, addition of the SECIS RNA did not induce any additional ordering (not shown), which is consistent with the RNA binding domain being predicted as folded.

To eliminate the possibility that the unfolding of SPB2 and of the SBP2 fragment was dependent on the bacterial expression system, we made use of the eukaryotic insect cell/baculovirus expression system. However, only hSBP2△525 could be obtained in sufficient amount. The NMR experiment was then reproduced, which showed a pattern (Fig. 4d) comparable to that of the full-length *E. coli* expressed protein (Fig. 4b), in agreement with an important lack of folding. Therefore, the insect cell/baculovirus expression system yielded a short SPB2 even slightly less folded than its counterpart expressed in *E. coli* (Fig. 4c). At this stage, one cannot conclude on either a limited effect of the expression system or on some variations resulting from different purification steps.

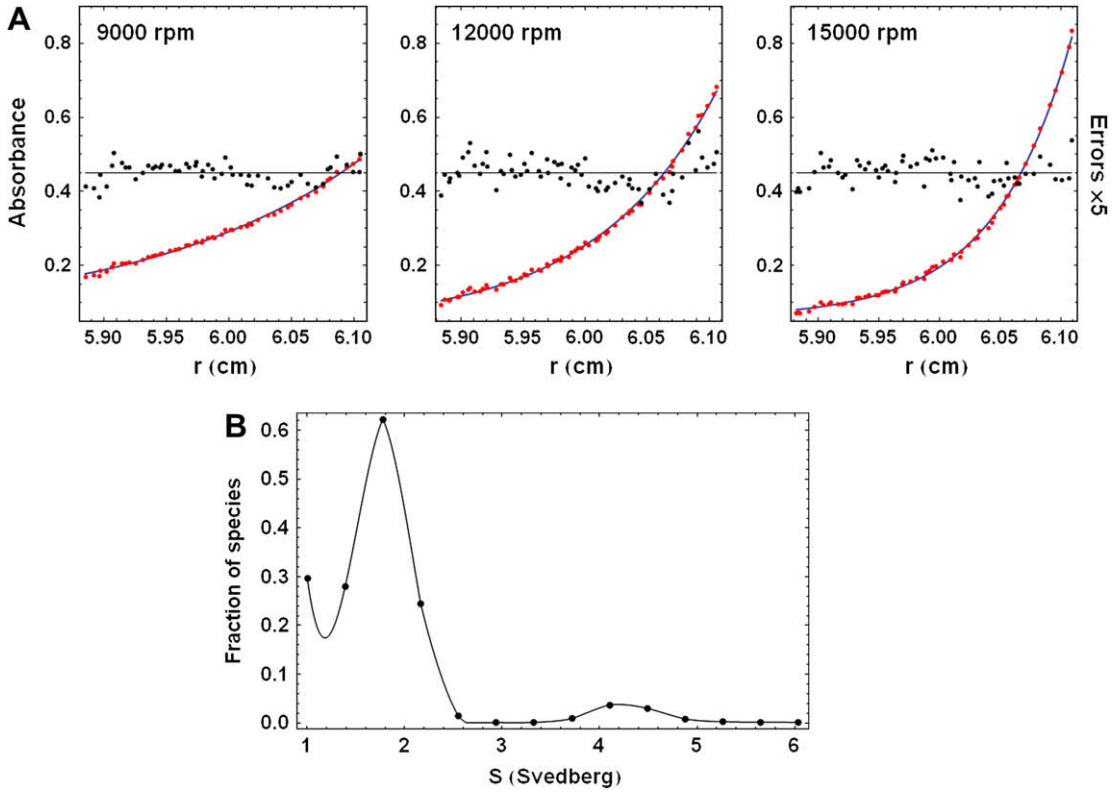


Fig. 3. Behaviour of SBP2 in solution assessed by ultracentrifugation analysis. (A) Sedimentation equilibrium experiment of rat SBP2 at 9000, 12 000 and 15 000 rpm. The theoretical curves shown here were obtained with the sequence-derived SBP2 molecular weight (95 326 Da), and with $\rho_{solvent} = 1.009\ g\ ml^{-1}$, $\bar{V}_{protein} = 0.74\ ml\ g^{-1}$. The zero level for the distribution of errors was offset by 0.45. (B) Distribution of sedimentation coefficients explaining the sedimentation profiles (not shown) obtained with rat SBP2 at 40 000 rpm. The continuous curve is merely an interpolation curve as a guide for the eye.

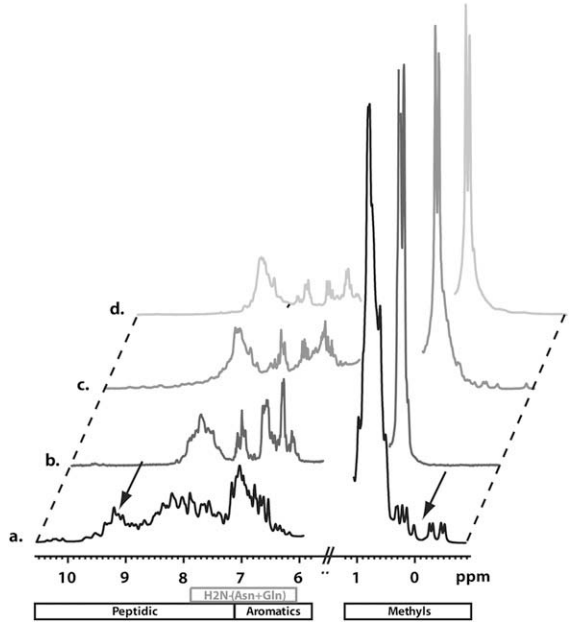


Fig. 4. Lack of SBP2 folding assessed by 1D-^{1}H NMR. (a) 600 MHz ^{1}H spectrum of a folded protein (24 kDa N-terminal fragment of *E. coli* gyrase B) compared to the spectra (b) of the full-length rat SBP2, (c) of hSBP2△525 proteins expressed in *E. coli*, and (d) of hSBP2△525 proteins expressed in baculovirus-infected cells. The folded protein (a) shows both the characteristic full-range peptide bond resonances between 7 and 10 ppm, and the characteristic methyl resonances around 0 ppm (arrows). In contrast, all three SBP2 samples (b,c,d) show either a complete lack of these resonances, or only small methyl peaks.

4. Discussion

Altogether, computer predictions and experimental results are in agreement with the rat SBP2 and human hSBP2△525 proteins being significantly disordered. Based upon glycerol gradient sedimentation and the presence of high molecular weight bands in gel-shift assays, it was originally thought that SBP2 formed a dimer, or even multimers [10,26], but later studies indicated the SBP2 state to be monomeric [40], which is fully accounted for by our present results. Altogether, these results rationalize our failure to crystallize SBP2, be it alone or in complex with several forms of the SECIS RNA.

It has been observed that EFSec co-immunoprecipitated [6] and colocalized with SBP2 [41]. *In vitro*, a direct interaction between both proteins could not be demonstrated in the absence of tRNASec (unless truncation of part of the N-terminal domain of EFSec was performed) [6], whereas a stable complex could be detected in the presence of SECIS [42]. The latter study implies that the SECIS element is important for SBP2–EFSec interaction. This is consistent with the region of interaction being mapped to the residue range 400–535 in SBP2 [42], which is right N-ter to the extended RNA binding domain (Fig. 1A). However, this does not imply that, *in vivo*, SECIS alone is sufficient and SBP2 might require additional folding before it can interact with EFSec. In this respect, it is of interest that SBP2 has been shown to interact with Nufip and the Hsp90 chaperone and co-chaperones, the components of an assembly machinery shared by other L7Ae/L30-containing ribonucleoprotein particles [43]. Sequence analysis with the same method as that used in Fig. 1A also indicated a significant degree of unfolding for Nufip (data not shown). It must be noticed that Hsp90 is not viewed as a regular Heat Shock Protein rescuing misfolded proteins

regardless of their function, but rather as a chaperone interacting with a set of various 'client proteins' [44]. A working hypothesis is thus that Nufip could present SBP2 to Hsp90 for controlling its folding, which would ensure efficient RNP formation. Interestingly, the tumor suppressor transcription factor p53 is one of the numerous substrates of Hsp90 and, like SBP2, it has been shown to be an IDP [45]. It should be stated clearly, however, that additional folding of SBP2, if any [46], could be limited to only a fraction of the whole protein.

Finally, one may raise the hypothesis that the spontaneous cleavage seen within a highly conserved region of both the rat and human SBP2 could be of biological significance. Such a hypothesis gains some support from the fact that, in other organisms like drosophila, SBP2 is fully active although it lacks most of the N-terminal domain liberated by the cleavage observed in the rat and human counterparts [17]. Interestingly, it was shown in [17] by systematic N-terminal deletions that fragments shorter than hSBP2△525 had lost the ability to bind the SECIS RNA, whereas only fragments longer than hSBP2△525 were able to bind specifically the ribosomal 60S subunit. Therefore, in agreement with the previous considerations, it may be suggested that this rat and human N-terminal domain could be involved in the recognition of other partner(s), possibly the ribosome. However, in the frame of the hypothesis, this might be a transient role due to subsequent cleavage of the N-terminal domain generating a short 'drosophila-like' form involved in the next step of selenocysteine insertion. In the present state of our knowledge going beyond these considerations would be too speculative.

Acknowledgments

We thank Bernard Lorber for his help with the DLS system, Didier Busso and Edouard Troesch of the CEBGS-Illkirch for use of the Gateway cloning system, Arnaud Poterszmann and the baculovirus/insect cell culture staff facility at IGBMC-Illkirch, and Anne Schweigert for excellent technical assistance. We also thank Véronique Receveur-Bréchot for valuable remarks on the manuscript. This work was supported in part by grants from the programs 'Action Concertée Incitative en Biologie Cellulaire Moléculaire et Structurale' #226 (to A.K. and P.D.) ToxNuc-E and ANR (to A.K.). V.O. received a PhD fellowship from MENRT, A.T. was supported by the Japanese government.

References

[1] L.V. Papp, J. Lu, A. Holmgren, K.K. Khanna, From selenium to selenoproteins: synthesis, identity, and their role in human health, Antioxid. Redox Signal. 9 (2007) 775–806.
[2] C. Allmang, L. Wurth, A. Krol, The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated, Biochim. Biophys. Acta doi:10.1016/j.bbagen.2009.03.003, in press, [Epub ahead of print].
[3] P.R. Copeland, J.E. Fletcher, B.A. Carlson, D.L. Hatfield, D.M. Driscoll, A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs, EMBO J. 19 (2000) 306–314.
[4] D. Fagegaltier, N. Hubert, K. Yamada, T. Mizutani, P. Carbon, A. Krol, Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation, EMBO J. 19 (2000) 4796–4805.
[5] R.M. Tujebajeva, P.R. Copeland, X.M. Xu, B.A. Carlson, J.W. Harney, D.M. Driscoll, D.L. Hatfield, M.J. Berry, Decoding apparatus for eukaryotic selenocysteine insertion, EMBO Rep. 1 (2000) 158–163.
[6] A.M. Zavacki, J.B. Mansell, M. Chung, B. Klimovitsky, J.W. Harney, M.J. Berry, Coupled tRNA(Sec)-dependent assembly of the selenocysteine decoding apparatus, Mol. Cell 11 (2003) 773–781.
[7] L. Chavatte, B.A. Brown, D.M. Driscoll, Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes, Nat. Struct. Mol. Biol. 12 (2005) 408–416.
[8] K. Caban, P.R. Copeland, Size matters: a view of selenocysteine incorporation from the ribosome, Cell. Mol. Life Sci. 63 (2006) 73–81.
[9] E.V. Koonin, P. Bork, C. Sander, A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? Nucleic Acids Res. 22 (1994) 2166–2167.
[10] C. Allmang, P. Carbon, A. Krol, The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding, RNA 8 (2002) 1308–1318.
[11] C. Allmang, A. Krol, Selenoprotein synthesis: UGA does not end the story, Biochimie 88 (2006) 1561–1571.
[12] I. Vidovic, S. Nottrott, K. Hartmuth, R. Luhrmann, R. Ficner, Crystal structure of the spliceosomal 15.5 kD protein bound to a U4 snRNA fragment, Mol. Cell 6 (2000) 1331–1342.
[13] C. Allmang, A. Krol, SECIS RNAs and K-turn binding proteins. A survey of evolutionary conserved RNA and protein motifs, in: H. DL, B. MJ, G. VN (Eds.), Selenium, its Molecular Biology and Role in Human Health, Springer, New York, USA, 2006, pp. 51–61.
[14] A. Clery, V. Bourguignon-Igel, C. Allmang, A. Krol, C. Branlant, 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 (2007) 1868–1884.
[15] J.L. Bubenik, D.M. Driscoll, Altered RNA binding activity underlies abnormal thyroid hormone metabolism linked to a mutation in selenocysteine insertion sequence-binding protein 2, J. Biol. Chem. 282 (2007) 34653–34662.
[16] K. Caban, S.A. Kinzy, P.R. Copeland, 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. Cell Biol. 27 (2007) 6350–6360.
[17] A. Takeuchi, D. Schmitt, C. Chapple, E. Babaylova, G. Karpova, R. Guigo, A. Krol, C. Allmang, A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins, Nucleic Acids Res. 37 (2009) 2126–2141.
[18] R.W. Kriwacki, L. Hengst, L. Tennant, S.I. Reed, P.E. Wright, Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 11504–11509.
[19] F. Bemporad, J. Gsponer, H.I. Hopearuoho, G. Plakoutsi, G. Stati, M. Stefani, N. Taddei, M. Vendruscolo, F. Chiti, Biological function in a non-native partially folded state of a protein, EMBO J. 27 (2008) 1525–1535.
[20] M. Sickmeier, J.A. Hamilton, T. LeGall, V. Vacic, M.S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V.N. Uversky, Z. Obradovic, A.K. Dunker, DisProt: the database of disordered proteins, Nucleic Acids Res. 35 (2007) D786–D793.
[21] V.N. Uversky, J.R. Gillespie, A.L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41 (2000) 415–427.
[22] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, J. Mol. Biol. 293 (1999) 321–331.
[23] P. Tompa, P. Csermely, The role of structural disorder in the function of RNA and protein chaperones, FASEB J. 18 (2004) 1169–1175.
[24] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, Proteins 42 (2001) 38–48.
[25] V. Receveur-Brechot, J.M. Bourhis, V.N. Uversky, B. Canard, S. Longhi, Assessing protein disorder and induced folding, Proteins 62 (2006) 24–45.
[26] P.R. Copeland, V.A. Stepanik, D.M. Driscoll, Insight into mammalian selenocysteine insertion: domain structure and ribosome binding properties of Sec insertion sequence binding protein 2, Mol. Cell Biol. 21 (2001) 1491–1498.
[27] J.E. Fletcher, P.R. Copeland, D.M. Driscoll, A. Krol, The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2, RNA 7 (2001) 1442–1453.
[28] P. Schuck, M.A. Perugini, N.R. Gonzales, G.J. Howlett, D. Schubert, Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems, Biophys. J. 82 (2002) 1096–1111.
[29] T. Laue, B. Shah, T. Ridgeway, S. Pelletier, Computer Aided Interpretation of Analytical Sedimentation Data for Proteins, The Royal Society of Chemistry, Cambridge, UK, 1992.
[30] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (2005) 3433–3434.
[31] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C.J. Brown, A.K. Dunker, Predicting intrinsic disorder from amino acid sequence, Proteins 53 (Suppl. 6) (2003) 566–572.
[32] K. Peng, P. Radivojac, S. Vucetic, A.K. Dunker, Z. Obradovic, Length-dependent prediction of protein intrinsic disorder, BMC Bioinform. 7 (2006) 208.
[33] J.D. Thompson, T.J. Gibson, D.G. Higgins, Multiple sequence alignment using ClustalW and ClustalX, Curr. Protoc. Bioinform. (2002) (Chapter 2) Unit 2 3.
[34] P. Gouet, E. Courcelle, D.I. Stuart, F. Metoz, ESPript: analysis of multiple sequence alignments in PostScript, Bioinformatics 15 (1999) 305–308.
[35] R. Linding, L.J. Jensen, F. Diella, P. Bork, T.J. Gibson, R.B. Russell, Protein disorder prediction: implications for structural proteomics, Structure 11 (2003) 1453–1459.
[36] P. Tompa, Intrinsically unstructured proteins, Trends Biochem. Sci. 27 (2002) 527–533.
[37] A. Lescure, C. Allmang, K. Yamada, P. Carbon, A. Krol, cDNA cloning, expression pattern and RNA binding analysis of human selenocysteine insertion sequence (SECIS) binding protein 2, Gene 291 (2002) 279–285.
[38] H.K. Schachman, Ultracentrifugation in Biochemistry, Academic Press, New York, 1959.
[39] Y. Wang, I. Filippov, C. Richter, R. Luo, R.W. Kriwacki, Solution NMR studies of an intrinsically unstructured protein within a dilute, 75 kDa eukaryotic protein assembly; probing the practical limits for efficiently assigning polypeptide backbone resonances, Chem. Biochem. 6 (2005) 2242–2246.

[40] S.A. Kinzy, K. Caban, P.R. Copeland, Characterization of the SECIS binding protein 2 complex required for the co-translational insertion of selenocysteine in mammals, Nucleic Acids Res. 33 (2005) 5172–5180.
[41] L.A. de Jesus, P.R. Hoffmann, T. Michaud, E.P. Forry, A. Small-Howard, R.J. Stillwell, N. Morozova, J.W. Harney, M.J. Berry, Nuclear assembly of UGA decoding complexes on selenoprotein mRNAs: a mechanism for eluding nonsense-mediated decay? Mol. Cell Biol. 26 (2006) 1795–1805.
[42] J. Donovan, K. Caban, R. Ranaweera, J.N. Gonzalez-Flores, P.R. Copeland, A novel protein domain induces high affinity selenocysteine insertion sequence binding and elongation factor recruitment, J. Biol. Chem. 283 (2008) 35129–35139.
[43] S. Boulon, N. Marmier-Gourrier, B. Pradet-Balade, L. Wurth, C. Verheggen, B.E. Jady, B. Rothe, C. Pescia, M.C. Robert, T. Kiss, B. Bardoni, A. Krol, C. Branlant, C. Allmang, E. Bertrand, B. Charpentier, The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery, J. Cell Biol. 180 (2008) 579–595.
[44] R. Zhao, W.A. Houry, Hsp90: a chaperone for protein folding and gene regulation, Biochem. Cell Biol. 83 (2005) 703–710.
[45] P. Tompa, The interplay between structure and function in intrinsically unstructured proteins, FEBS Lett. 579 (2005) 3346–3354.
[46] A.B. Sigalov, A.V. Zhuravleva, V.Y. Orekhov, Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form, Biochimie 89 (2007) 419–421.

2.4. SBP2 is an Intrinsically Disordered Protein

Our results established that ca. 70 % of the SBP2 sequence is disordered with the exception of the L7Ae RNA binding domain that appears to be folded, and supports our hypothesis that SBP2 is an Intrinsically Disordered Protein and will be folded only in the presence of its partners. This is consistent with a recent report, in which participated my laboratory, on the role of the Hsp90 chaperone for the folding of SBP2 and other L7Ae family proteins (Boulon et al, 2008). The flexibility of IDPs derived from the folding transition upon binding their targets could provide the ability to bind several different partners (reviewed in Tompa, 2002; Wright & Dyson, 2009). Thus, there may exist other partners of SBP2 yet to be identified. Furthermore, an IDP can play an important role in ensuring the correct assembly order of individual components of multimeric nucleoprotein complexes: indeed, the binding affinity of the IDP for its target can be regulated by its different structures induced by binding different partners (Wright & Dyson, 1999). The intrinsically disordered nature of SBP2 could also direct the recruitment of its several different partners in the correct order. Finally, since the reason of unsuccessful availability of the crystal may rely on the disordered nature of SBP2, we therefore conclude that the crystallization of SBP2 will be very difficult unless its partners are found.

A number of proteins have been reported whose native and functional states are intrinsically unstructured but adopt folding upon binding their targets, providing evidence that the unstructured state is essential for basic cellular functions (Reviewed in Dyson & Wright, 2005; Tompa, 2002; Wright & Dyson, 1999). For example, the N-terminal fragment of the cyclin-dependent kinase inhibitor $p21^{Waf1/Cip1/Sdi1}$ lacks stable structure in the free solution state, but adopts an ordered stable conformation when bound to its target, Cdk2 (Kriwacki et al, 1996). Indeed, many intrinsically disordered proteins are involved in cellular control mechanisms and signaling, by playing important roles in protein interaction networks (reviewed in Wright & Dyson, 2009). Intrinsically disordered domains have several advantages over rigid three-dimensionally structured proteins. Their higher flexibility enables them to bind numerous different targets and increases association and dissociation rates. One single protein can even perform different or opposing functions due to different active conformations induced upon binding their different targets. Moreover, intrinsically disordered regions can transit between different levels of three-dimensional organization (reviewed in

Rajkowitsch et al, 2007). Although unstructured domains are missing in three-dimensional structures determined by X-ray crystallography, the structure of an IDP bound to its target(s) was resolved by X-ray crystallography in a few cases (Figure 24) (reviewed in Tompa, 2002).

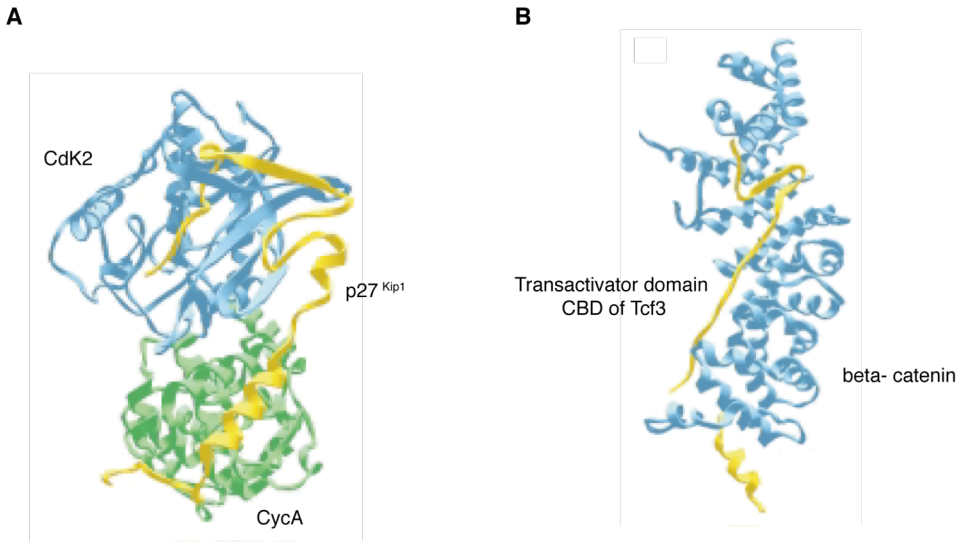


Figure 24. Examples of X-ray structures of IDPs bound to their targets.

A. The cyclin-dependent-kinase inhibitor $p27^{Kip1}$ (yellow) complexed with its targets, cyclin-dependent kinase 2 (Cdk2, blue) and cyclin A (CycA, green) (Russo et al, 1996). **B.** The transactivator domain β-catenin binding domain (CBD) of transcription factor Tcf3 (yellow) bound to β-catenin (blue) (Graham et al, 2000). Figures are taken from (Tompa, 2002).

Obviously, resolution of the crystal structure of the SBP2-SECIS complex requires the identification of other partners of SBP2 so that the intrinsically disordered domains of SBP2 become structured.

3. Toward identification of SBP2 partners

3.1. Objective

The expression mechanism of selenoprotein mRNAs involves a number of factors (see Part 1. Introduction). Among these, SBP2 plays an important role in the assembly of selenoprotein mRNPs. It notably recruits the Sec-tRNA^{Sec}/EFSec complex to the SECIS element during selenoprotein synthesis (Tujebajeva et al, 2000) but it also associates with a complex machinery linked to the Hsp90 chaperone that triggers proper SBP2 folding and subsequent SECIS RNP assembly (Boulon et al, 2008). This assembly machinery is conserved and involved in the assembly of several L7Ae RNPs and implies that SBP2 has functional similarities with other L7Ae proteins during RNP assembly (Boulon et al, 2008). Interestingly, the L7Ae proteins are primary binding proteins that participate in several RNP complexes by binding directly to the K-turn RNA motifs. This binding is a prerequisite to the assembly of the other core proteins of the RNP complexes to which they belong.

In archaea, formation of the L7Ae initiation complex enables the recruitment of Nop5 to the assembling archeal RNP, which in turn facilitates the association of fibrillarin to the catalytically active sRNP in vitro (Omer et al, 2002). In eukaryotes, 15.5kD (the L7Ae protein that exhibits the highest similarity with SBP2 (Allmang et al, 2002)) is at the heart of different sn/snoRNPs (Watkins et al, 2000) such as box C/D snoRNPs, U3 snoRNP and U4 snRNP. In box C/D snoRNPs, Snu13p/15.5kD (the eukaryotic homolog of archaeal L7Ae) is part of an RNP complex similar to the archaeal one containing Nop56, Nop58 (the eukaryotic orthologs of archaeal Nop5) and fibrillarin (Figure 25 A and B, Kuhn et al, 2002; Nottrott et al, 1999, for a review on sn(o)RNP structure, see Reichow et al, 2007). In the U3 box C/D snoRNPs, that plays essential roles in ribose 2'-O-methylation and pre-rRNA processing, 15.5kD is also present together with the methyltransferase fibrillarin, Nop56 and Nop58 (Watkins et al, 2002). In the case of the U4 snRNPs, one of the major components of the spliceosome, the interaction between 15.5kD and the U4 snRNA is required for the association of a different set of core proteins, PRP31 and the cyclophilin H-hPRP4-hPRP3 complex (Nottrott et al, 2002) (Figure 25 C).

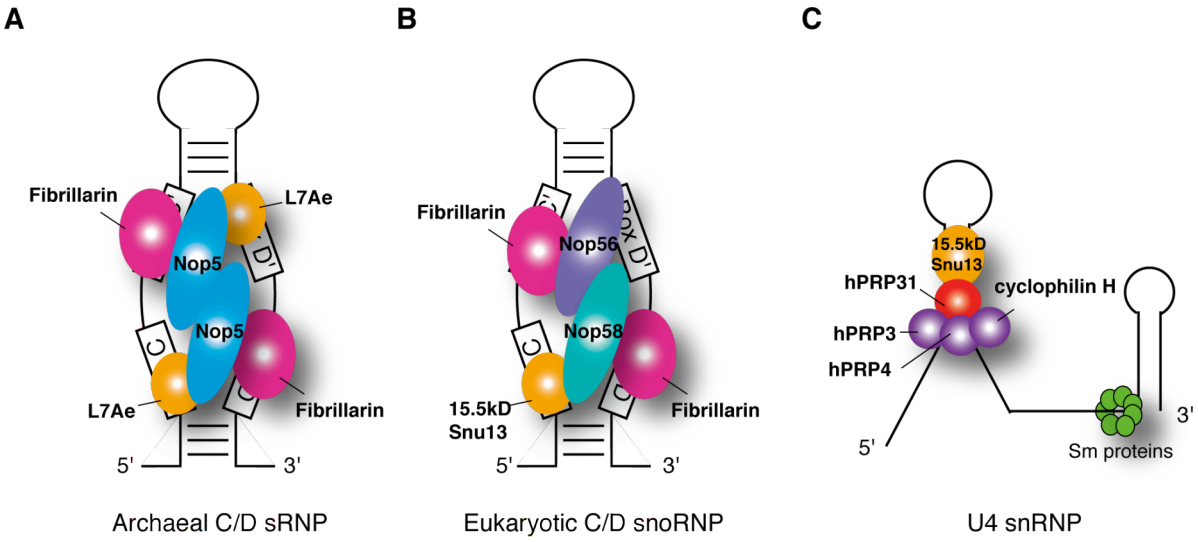


Figure 25. Composition and organization of C/D sno(s)RNPs and U4 snRNP

A. Core proteins of the archaeal C/D sRNP are L7Ae, fibrillarin and Nop5. Nop5 interacts with fibrillarin, and Nop5-fibrillarin complex dimerizes. **B.** The eukaryotic C/D snoRNP contains 15.5kD/Snu13 (L7Ae homolog), fibrillarin, Nop56 and Nop58 (Nop5 paralogs). In contrast to the archaeal L7Ae, 15.5kD appears to bind only at the C/D sites of the snoRNA. **C.** The U4 snRNP contains 15.5kD/Snu13, hPRP31, and the cyclophilin H-hPRP4-hPRP3 complex as well as seven common Sm proteins. hPRP31 contains the Nop domain that mediates binding to the 15.5kD-U4 snRNA complex (Liu et al, 2007).

We therefore speculated that, by analogy with 15.5kD, SBP2 may be able to recruit to selenoprotein mRNP core proteins that are common to those of sn/snoRNPs. As a starting hypothesis, we tested whether Nop56 and Nop58, the most common interactants of 15.5 kD, could also interact with SBP2. This work was initiated together with Laurence Wurth, another PhD student in the laboratory whose main project is to understand the assembly pathway of the selenoprotein mRNPs and identify new protein partners for SBP2. In this work, based on the encouraging preliminary results of her *in vivo* studies, I further analyzed the interaction between SBP2 and Nop56 and Nop58 *in vitro*, in order to determine whether these proteins could be potential partners of SBP2.

3.2. Results

Previous experiments in the laboratory detected interactions between SBP2 and Nop56/58 by co-transfection and co-immunoprecipitation assay (Laurence Wurth, unpublished data). To confirm these interactions and test whether SBP2 interacts directly with Nop58, we set out to test this interaction *in vitro*.

I performed GST pull-down assays using the recombinant *E. coli*-expressed GST-Nop58 protein and the *in vitro* translated ^{35}S-labelled hSBP2 protein. hSBP2 was first translated in the rabbit reticulocyte lysate that allows efficient production of eukaryotic proteins *in vitro*. A strong interaction between the two proteins was detected (Figure 26 lane 3). However, the rabbit reticulocyte lysate may contain eukaryotic cellular components that could mediate the interaction between Nop58 and SBP2. This possibility has to be envisaged as a protein called Nufip, that is part of the Hsp90 chaperone assembly machinery, was shown to bridge sn(o)RNP core proteins to 15.5kD (Boulon et al, 2008), even though these proteins were unable to interact directly with 15.5 kD. However, yeast two hybrid experiments did not detect any interaction between Nop58 and SBP2, even in the presence of Nufip (Laurence Wurth, unpublished data). To exclude completely the possibility that another protein may mediate the interaction between Nop58 and SBP2, hSBP2 was translated in *E.coli* S30 extract that does not contain Nop58 or SBP2 orthologues. Furthermore, *E.coli* also lacks components of snoRNPs and of the eukaryotic selenoprotein synthesis machinery that could bridge the interaction. When hSBP2 was produced in bacterial S30 extracts, the translation pattern was different from that of rabbit reticulocyte lysate, and more proteolytic fragments or internal termination products were detected (see Fig 26 lane 6). Nevertheless, a weak but positive signal of interaction between hSBP2 and Nop58 was detected in a GST-pull down assay using hSBP2 produced in the bacterial system (Figure 26 lane 5). This weak interaction appears to be due to the low expression efficiency of the full-length SBP2 (compare Figure 26 lanes 4 and 12).

Altogether, these results suggest that SBP2 is able to interact directly with Nop58. Nop58 is therefore a good candidate to be a real core component of the SECIS RNP. More experiments have to be done to confirm this preliminary observation and the proposal. This will be developed in the discussion.

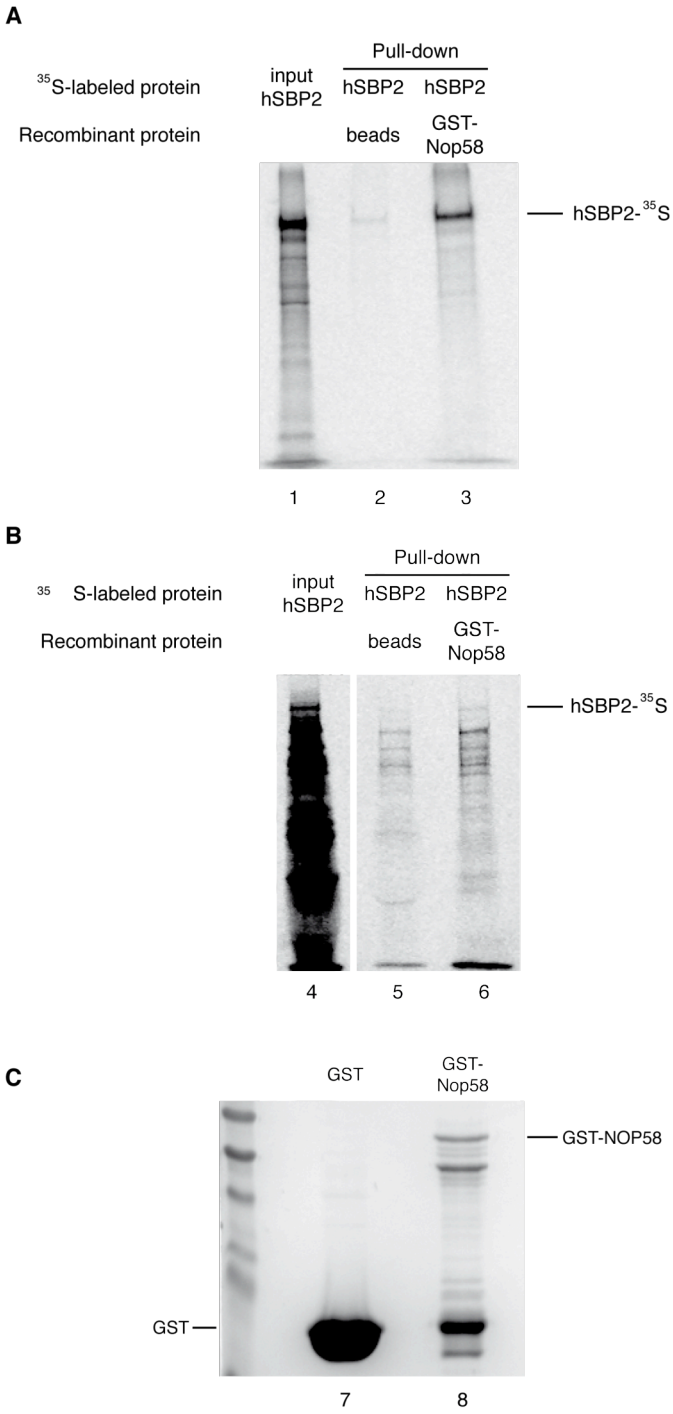


Figure 26. GST pull-down experiments.

E. coli cell extracts expressing GST-Nop58 or GST alone were incubated with glutathione agarose beads. **A.** hSBP2 translated *in vitro* in rabbit reticulocyte lysate in the presence of ^{35}S-Met were added to the beads and assayed for binding to the recombinant GST-Nop58. **B.** hSBP2 translated *in vitro* in E.coli S30 lysate in the presence of ^{35}S-Met were assayed for binding to the recombinant proteins (upper panel). The recombinant GST-Nop58 and GST proteins bound to the glutathione agarose beads are shown (lower panel).

3.3. Discussion

snoRNPs that localize to the nucleolus play essential roles in modification and processing of rRNAs, and contain a small nucleolar RNA (snoRNA) and a set of common snoRNP proteins. snoRNAs are categorized into two major classes according to the distinctive and conserved sequence elements that they have, the box C/D snoRNAs and box H/ACA snoRNAs. The box C/D snoRNA forms a complex with fibrillarin, Nop56, Nop58 and 15.5kDa/Snu13, which catalyzes the ribose 2'-O-methylation (see also 3.1.). The box H/ACA snoRNA is associated with dyskerin/Cbf5, Gar1, Nhp2 and Nop10, forming the box H/ACA snoRNP that functions in pseudouridine formation (reviewed in Charpentier et al, 2007; Kiss, 2002; Matera et al, 2007; Reichow et al, 2007). snRNPs that catalyze RNA splicing in the eukaryotic nucleus, comprise snRNAs, Sm core proteins and some other protein factors specific to a given species of snRNP. 15.5kD is also a component of the mature spliceosomal U4 snRNP (Nottrott et al, 2002; Patel & Bellini, 2008). During the maturation of sn/snoRNPs, sn/snoRNAs undergo dynamic processes such as association and dissociation of numerous factors, assembly of the core proteins and intracellular trafficking.

In the cases described above, the L7Ae proteins (15.5kD for box C/D snoRNPs and U4 snRNPs, and Nhp2 for the box H/ACA snoRNPs) that bind the K-turn motif recruit their protein partners to the sn/snRNAs and therefore play a central role in the formation of the mature RNPs (reviewed in Kiss, 2002; Matera et al, 2007; Reichow et al, 2007). Because of the sequence and functional similarity with other L7Ae proteins, SBP2 has been suggested to play the same role during the selenoprotein mRNP formation as that of other L7Ae proteins during sno/snRNP formation (Allmang et al, 2009).

In this study, we showed that SBP2 interacts *in vitro* with at least one of the core proteins of the box C/D snoRNPs, Nop58, and that this interaction is direct (Figure 26 lane 5). Together with the previous studies detecting this interaction *in vivo* (Laurence Wurth unpublished data), our results support the hypothesis that SBP2 may recruit to selenoprotein mRNPs core proteins that are common to snoRNPs. Further experiments will be required for validating this hypothesis. Indeed, it will be essential to test in gel-shift assays whether Nop58 is able to interact with the SECIS RNAs in the presence and absence of SBP2. *In vivo* immunoprecipitation assays of Nop58, followed by detection of selenoprotein mRNAs, will also allow us to define whether Nop58 belongs to the SECIS RNP. Cellular localization studies of these factors would also provide information whether Nop58 colocalizes with the SECIS RNPs.

In addition, to test whether more core proteins are common to sn(o)RNPs and selenoprotein mRNPs, the interaction between SBP2 and other core proteins of sn/snoRNPs such as Nop56, fibrillarin and PRP31 could also be tested.
Although the function of the interaction between SBP2 and Nop58 was not analyzed in this study, this interaction may contribute to the stabilization of SBP2 that we showed to be intrinsically disordered (see 2.4.); alternatively it may also contribute to the stability of the SBP2-selenoprotein mRNA complexes or the mechanism of assembly of the selenoprotein mRNA. Another possibility is that the interaction with Nop58 could direct localization of selenoprotein mRNAs to a particular cellular compartment so that they could circumvent the NMD pathway or be delivered to a specialized pool of ribosomes. Localization experiments will be required to test this hypothesis as well as functional analysis to determine the role of this interaction on selenoprotein synthesis.

Altogether, our results revealed that the molecular assembly of selenoprotein mRNPs has many similarities with that of sn- and snoRNPs.

Part 3. General conclusion

Part 3. General conclusion

Since selenocysteine is co-translationally incorporated into a growing peptide chain in response to a UGA Sec codon, otherwise read as a translational termination signal, the correct recoding of UGA stop to UGA Sec requires a specialized translational machinery. Recent important progresses have been made toward the identification of the involved factors and the elucidation of the UGA recoding mechanism in eukaryotes. Among the *cis*- and *trans*-acting factors, SBP2 plays a central role in the eukaryotic selenocysteine incorporation machinery. During my PhD studies, our results provided important insight into how the SBP2-SECIS RNA complex is formed during selenoprotein synthesis, a process that is at the heart of this recoding mechanism.

The first part of my PhD studies concerned the identification and functional characterization of the *Drosophila melanogaster* SBP2 protein (dSBP2) that lacks the region homologous to the N-terminus of vertebrate SBP2 (Chapple & Guigo, 2008; Takeuchi et al, 2009). Despite its shorter length, dSBP2 retained functional properties similar to the mammalian counterpart. However and interestingly, it exhibited differential SECIS recognition that was not observed in human SBP2 (hSBP2) and furthermore not expected at all. While hSBP2 binds both form 1 and 2 SECIS RNAs with similar affinities, dSBP2 exhibits high affinity toward form 2 only, the only form present in *Drosophila* selenoprotein mRNAs. In addition, we identified in SBP2 a K (lysine) rich-domain that is essential for SECIS and 60S ribosomal subunit binding (Fig 27). This domain thus constitutes an additional but different RNA binding domain from the L7Ae RNA binding module. Swapping five amino acids between dSBP2 (SVRVY) and hSBP2 (IILKE) in the K-rich domain conferred reversed SECIS binding properties to the proteins, thus unveiling a pentapeptide sequence important for form 1 binding.

Another part of my project consisted in the structural analysis of SBP2, in collaboration with the group of Philippe Dumas. Our results established that SBP2 is globally unstructured, with the exception of the L7Ae RNA binding domain (Oliéric et al, in press). This is consistent with recent results in the laboratory that showed that the stability (and most likely the folding) of SBP2 is dependent on the protein chaperone Hsp90. In that work, SBP2 has been shown to interact with the adaptor protein Nufip but also directly with some co-factors of Hsp90 (Figure 27, Boulon et al, 2008). Unlike other Heat Shock Proteins that function in folding

non-native proteins, Hsp90 acts as a chaperone interacting with target proteins (reviewed in Zhao et al, 2005). Therefore, Nufip could mediate the interaction between SBP2 and Hsp90 to ensure efficient RNP formation. This assembly machinery was shown to be conserved and involved in other L7Ae RNP complexes such as sn(o)RNPs. In these cases, the adaptor Nufip also played in important role in stimulating the interaction between the L7Ae proteins and other core proteins of the RNP during the assembly process. A similar mechanism may occur in the case of selenoprotein RNP assembly. The intrinsically disordered nature of SBP2 also implies the presence of several different partners. However, protein partners of SBP2 are poorly characterized and largely unknown apart from EFSec. An interaction between SBP2 and EFSec was detected by *in vivo* co-immunoprecipitation assays (Tujebajeva et al, 2000, see also 2.2.2.2.c in Part 1.). This interaction was later reported to be tRNA dependent unless a masking region of EFSec was removed (Zavacki et al, 2003), but more recently, it was shown that SBP2 and EFSec can form a stable complex *in vitro that is* SECIS dependent (Donovan et al, 2008). However, RNA binding may not be sufficient for the SBP2-EFSec interaction, and other protein factors may be required to reinforce this interaction. This is consistent with SECp43 being shown to promote the SBP2-EFSec interaction *in vivo* (Small-Howard et al, 2006). It appears therefore that many factors are involved in linking SBP2 and the Sec incorporation machinery.

The last part of my PhD studies was precisely to look for potential partners of SBP2. Previous studies in the laboratory established the functional similarities between SBP2 and other L7Ae proteins during RNP assembly (Allmang et al, 2002; Boulon et al, 2008). We therefore speculated that, by analogy with 15.5kD (the L7Ae protein that exhibits the highest similarity with SBP2 (Allmang et al, 2002)), SBP2 may be able to recruit to selenoprotein mRNPs core proteins that are common to those of sn/snoRNPs. Interestingly, I could show that SBP2 interacts directly with at least one of the core proteins of box C/D snoRNPs, Nop58. This corroborated *in vivo* results by Laurence Wurth. This result points to another similarity between the selenoprotein mRNP and sn/snoRNP assembly process. Future work will aim at clarifying and understanding the role of this interaction.

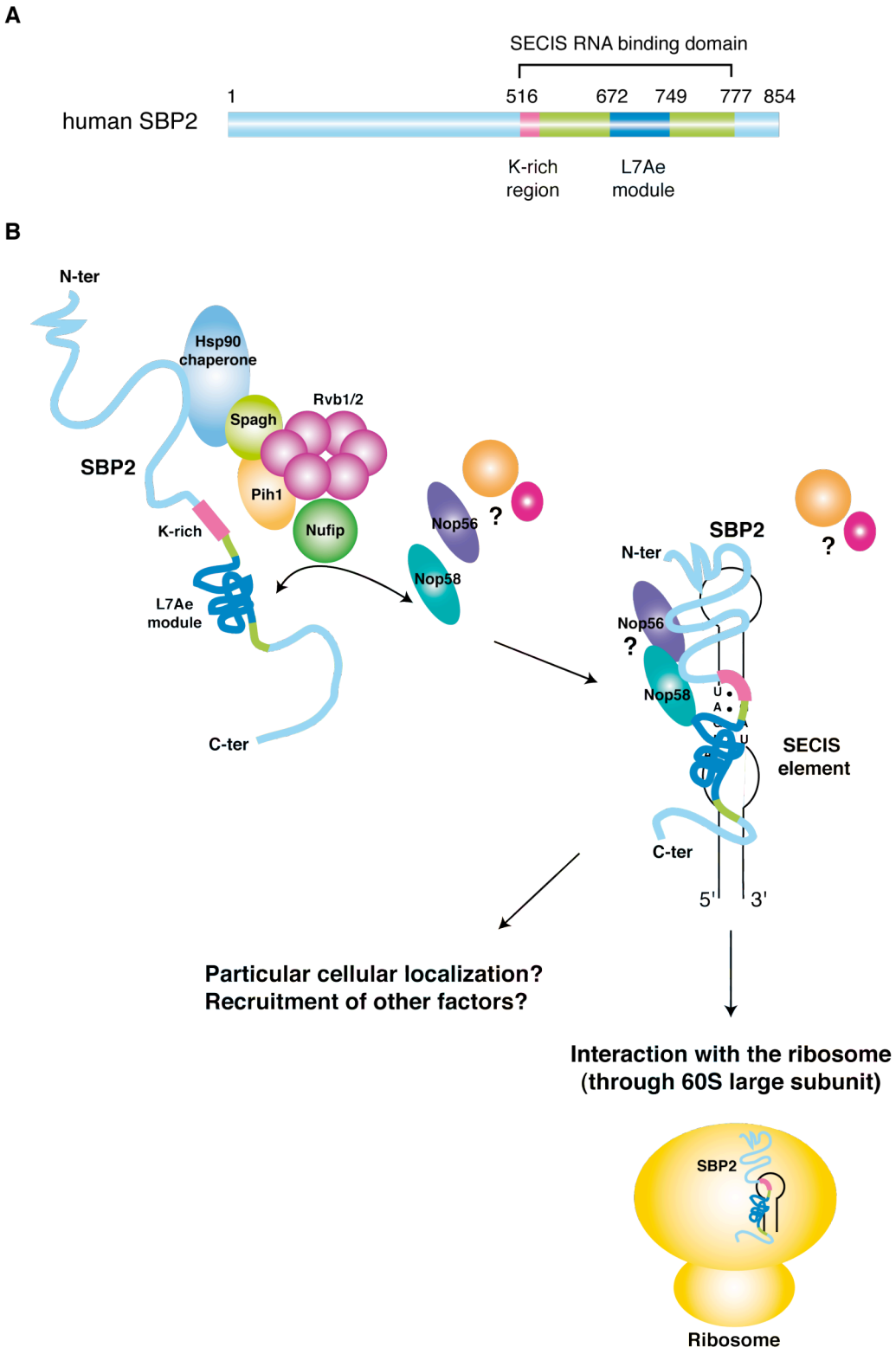


Figure 27. Proposed model for SBP2 functions during the selenoprotein mRNP formation.

A. Schematic representation of hSBP2. The K-rich region and the L7Ae module are shown in pink and blue, respectively. **B.** SBP2 is represented as a partially unfolded protein according to our disorder-prediction results described in (Oliéric et al, in press). The stability and likely proper folding of SBP2 is dependent on the association with the Hsp90 chaperone and its co-factors. The adaptor

protein Nufip stimulates the interaction between SBP2 and the chaperone complex but probably also core proteins. It also triggers SECIS RNP formation. Nop58 is likely to be one of the SECIS RNP core proteins. The cellular localization of the SECIS RNA-SBP2-Nop58/56 complex is unknown. Other core proteins may be recruited to the SECIS RNP. SBP2 also interacts with the 60S ribosomal subunit, very likely through binding to a stem-loop structure in the 28S rRNA. When associating with the ribosome, SBP2 cannot bind simultaneously SECIS RNA. The question mark represents potential interactions and interactants yet to be discovered. The stem-loop structure in the 60S ribosomal subunit is a possible SBP2 binding site in the 28S rRNA, yet to be identified.

Altogether, our observations allowed us to propose the following model for the role of SBP2 during selenoprotein mRNP formation (Figure 27). SBP2, that is globally unfolded except for the L7Ae domain, associates with the Hsp90 chaperone complex. This association contributes to the folding and therefore the stabilization of SBP2 (Boulon et al, 2008). This probably triggers SBP2 binding to SECIS RNA through the L7Ae module that establishes direct contacts with the non-Watson-Crick quartet. The K-rich domain in SBP2 may directly contact helix 1 of the SECIS RNA to increase the affinity of SBP2 for the SECIS RNA (Allmang et al, 2002; Takeuchi et al, 2009). The chaperone complex and the adaptor protein Nufip are also likely to mediate the recruitment of Nop58 to SBP2. Formation of the SECIS RNA-SBP2-Nop58 complex possibly induces further folding of SBP2, which could stabilize the complex and/or expose different interacting surfaces for other targets. Whether other core proteins common to sn/snoRNPs such as Nop56, fibrillarin and PRP31 are recruited to selenoprotein mRNPs remains to be tested.

The assembly of Nop58 to the selenoprotein mRNP may also help direct selenoprotein mRNAs to the nucleus to escape the NMD pathway or to deliver selenoprotein mRNPs to a specialized pool of ribosomes.

We showed that SBP2 interacts with the ribosome through binding the 60S subunit (Takeuchi et al, 2009). Very recently, SBP2 was shown to crosslink to the 28S rRNA (Olga Kossinova, PhD student in co-tutelle with our group and Galina Karpova, Novosibirsk). The exact SBP2-binding site in the 28S rRNA is still unknown but this result shows that there must be a helix in the 28S rRNA sharing a structure similar to the SECIS element. Our results revealed that the same region of SBP2 (K-rich domain) is essential for both SECIS and ribosomal binding. Therefore SBP2 cannot bind both targets simultaneously (see Chapter 1). This is consistent with previous models suggesting that SBP2 exchanges between the SECIS element and the ribosome during the recoding event and Sec incorporation (Allmang & Krol, 2006a; Caban &

Copeland, 2006; Chavatte et al, 2005; Kinzy et al, 2005). However, our results do not allow to discriminate whether SBP2 is pre-bound to the ribosome or the SECIS RNA prior to translation (see also 2.3. in Part 1.). They also do not explain if the complex formed at the SECIS in the 3'UTR could help to tether translation factors to the UGA Sec located in the coding region. More detailed information about the interaction between SBP2 and the ribosome will provide important insight into how the ribosome is told by the Sec incorporation machinery not to stop at the UGA Sec codon.

Altogether, our results revealed that the molecular assembly of selenoprotein synthesis machinery, that bears many similarities with that of sn- and snoRNPs, undergoes more dynamic processes than anticipated. Surprisingly, SBP2 plays important roles in regulating these events.

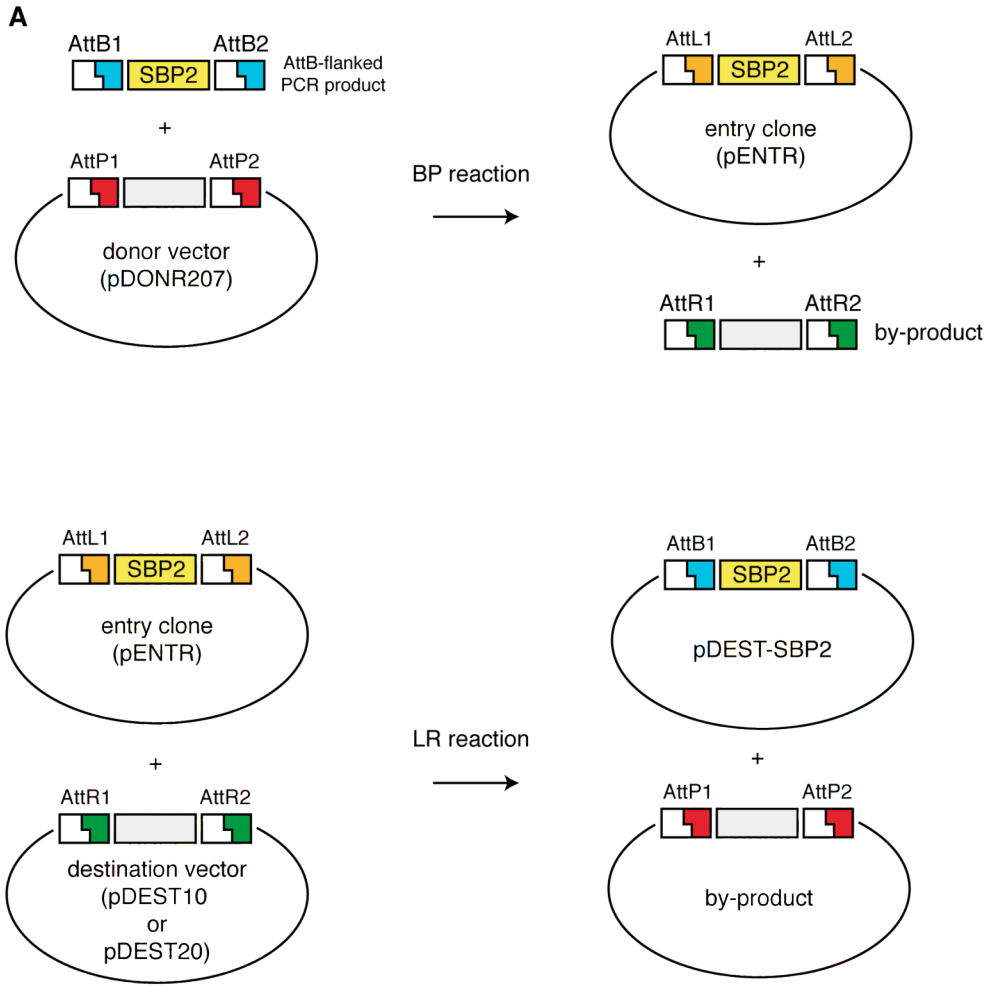
Part 4. Annex/ Methods

Part 4. Annex/ Methods

1. cDNA cloning using the GATEWAY Technology

The principle of the GATEWAY Technology is based on the conservative site-specific recombination of the λ phage between a DNA fragment harbouring specific recombination sites and GATEWAY-adapted vectors. Two recombination reactions are performed to create: (i) an entry clone; (ii) a destination vector. The first reaction is a recombination between an attB (the Bacterial attachment site) DNA segment and an attP (the Phage attachment site) donor vector to create an entry clone (also called the BP reaction because of the utilization of the attB and the attP sites, hence the names B and P), and the second recombination reaction occurs between an attL (the Left prophage attachment site) entry clone and an attR (the Right prophage attachment site) destination vector, hence the LR reaction (Figure 28).

The reading frame and orientation of the DNA fragment is maintained during recombination; attB1, attB2, attL1 and attL2 interact only with attP1, attP2, attR1 and attR2, respectively. To clone SBP2 cDNAs using this technology, I first added the recombination sequences at both ends of each SBP2 sequence. A two-step PCR was performed to generate SBP2 fragments harboring attB1 and attB2 recombination sites (Figure 29). The first PCR added the thrombin cleavage site at the 5' and the attB2 sequence at the 3' ends, the second one incorporating the attB1 sequence at the 5' end (Figure 29). Primers used for the PCR amplifications are listed in Table 3. The PCR products were then incubated with the pDONR207 (Invitrogen) vector in the presence of BP clonase mix (Invitrogen) containing the bacteriophage λ recombination protein Integrase (Int) and the *E.coli*-encoded protein Integration Host Factor (IHF), to generate entry clones (the BP reaction). The entry clones were next incubated with either the pDEST10 or pDEST20 (Invitrogen) destination vectors in the presence of LR clonase mix (Invitrogen) containing Int, IHF and Excisionase (Xis) (the LR reaction). pDEST10 and pDEST20 contain 6xHis tag and GST sequences, respectively, upstream of the attR1 recombination site.



B

	Reaction	Product
BP reaction	attB x attP → attL x attR	Entry Clone
LR reaction	attL x attR → attB x attP	Final Clone

Figure 28. Principle of the GATEWAY cloning

A. Overview of the BP and LR reactions with the resulting plasmid. The final clone, pDEST-SBP2 is used for generating a recombinant bacmid. **B.** Summary of reactions and nomenclature. The recombination occurring between the attB and the attP sites, and between the attL and the attR sites, are called the BP and the LR reactions, respectively. attB, attP, attL and attR are the Bacterial attachment site, the Phage attachment site, the Left prophage attachment site and the Right prophage attachment site, respectively.

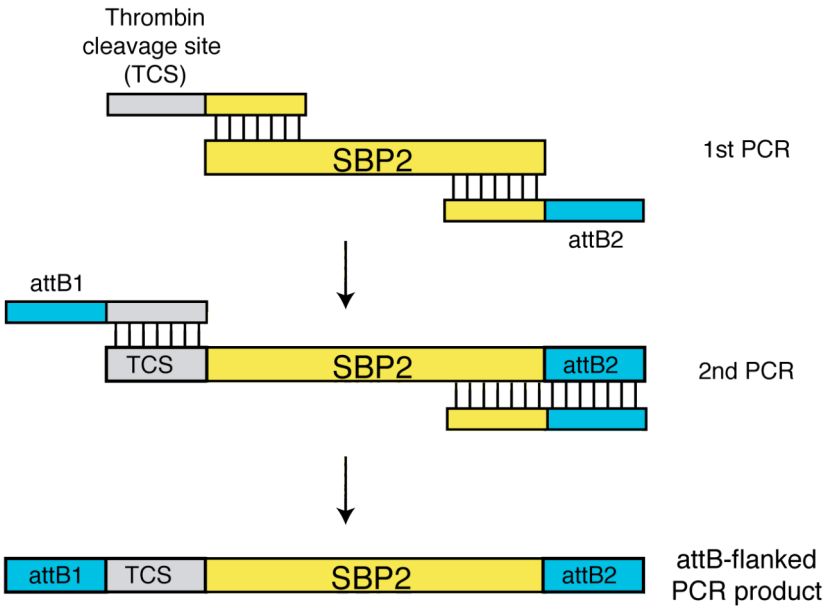


Figure 29. PCR amplification strategy to generate the SBP2 cDNA fragment harboring the AttB recombination sites.

Primer name	Sequence	use
P1610	ctggtgccacgcggttctgccaagaagccaacctcac	5' primer for hSBP2Δ525 1st PCR
P1613	ctggtgccacgcggttctatggcgtcggaggggcc	5' primer for hSBP2 1st PCR
P1615	ctggtgccacgcggttctatggcgtcggagcggcc	5' primer for ratSBP2 1st PCR
P1618	ctggtgccacgcggttctatgactgaaaaaataaggaggag	5' primer for dSBP2 1st PCR
Thrombin generic primer	ggggacaagtttgtacaaaaaagcaggcttcctggtgccacgcggttct	5' primer for 2nd PCR
P1612	ggggaccactttgtacaagaaagctgggtctcataaattcaaattcatcatttga	3' primer for hSBP2Δ525/hSBP2 2nd PCR
P1617	ggggaccactttgtacaagaaagctgggtcttataaattcaagttcatcatctg	3' primer for ratSBP2 2nd PCR
P1620	ggggaccactttgtacaagaaagctgggtctcacgaagcagttctgcgtt	3' primer for dSBP2 2nd PCR

Table 3. List of primers used for PCR amplification.

2. Baculovirus expression system

Protein expression in insect cells using the Baculovirus expression system requires multiple steps: preparation of the baculovirus shuttle vector (Bacmid), generation of virus particles,

determination of the viral titer, amplification of viral stocks, infection of insect cells and culture of baculovirus-infected insect cells. It also requires determination of optimized conditions for protein expression by mini and medium-scale expression tests before scaling-up the culture volume. The steps required for the generation of recombinant baculoviruses and gene expression are shown in Figure 30.

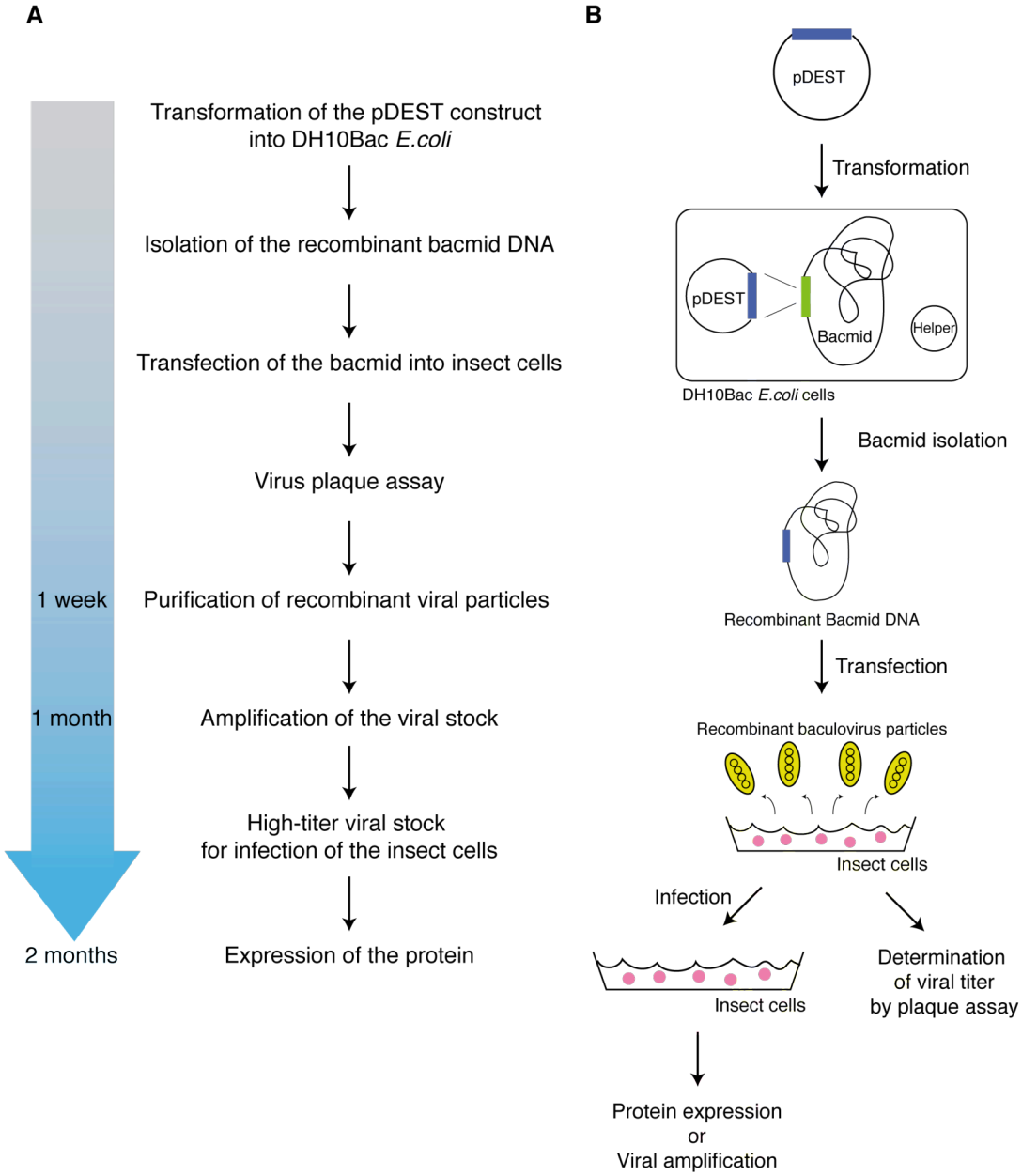


Figure 30. Generation of recombinant baculoviruses and gene expression.

A. Steps to generate recombinant baculoviruses. 1 week, 1 month and 2 months in the blue arrow indicate approximate time required. **B.** A recombinant bacmid is generated by site-specific transposition of an expression cassette into a host bacmid. The isolated recombinant bacmid is

transfected into insect cells to produce recombinant baculovirus particles. The viral particles are used for plaque assay or infection of the insect cells.

2.1. Bacmid preparation

The pDEST destination vectors cannot be used for protein expression in insect cells. The preparation of recombinant baculovirus shuttle vectors (Bacmid) is required. To generate recombinant bacmid DNAs, the pDEST destination constructs were transformed into the *E.coli* DH10Bac strain that contains a host bacmid and a helper plasmid encoding a transposase (Figure 30 B). This reaction is based on site-specific transposition of an expression cassette into a bacmid. After bacmid purification and PCR analysis to confirm insertion of the expression cassettes, ten bacmid clones were obtained (pGST-hSBP2a and b, pGST-hSBP2Δ525, pGST-ratSBP2a and b, pGST-dSBP2, pHis-hSBP2, pHis-hSBP2Δ525, pHis-ratSBP2 and pHis-dSBP2). They were used for generation of recombinant baculovirus particles by transfecting the recombinant bacmid DNA into insect cells (virus particles were generated at the baculovirus service at IGBMC). The recombinant viruses were used for expression tests.

2.2. Mini expression test

For initial screening of the recombinant baculovirus (before viral amplification and titration), the expression of each construct was tested. The insect cells (*Spodotera frugiperda*: Sf9 cells) from 2ml culture were collected and lysed in the $T_{20}N_{250}$ buffer containing 20mM of Tris-HCl pH8.0, 250mM NaCl. After sonication and centrifugation, soluble fractions were incubated with either Glutathione Sepharose (Amersham) or Ni-NTA (Qiagen) beads according to the tags. The purified proteins retained on the beads and the insoluble fractions were then loaded onto SDS-PAGE gels, followed by Western blotting. For the His-tagged constructs, Western blotting was performed using anti-His (SantaCruz) and anti-hSBP2 antibodies (NeoMPS).

2.3. Titration of viral particles and insect cell culture

The recombinant baculoviruses corresponding to the GST-tagged constructs (pGST-hSBP2b, pGST-hSBP2Δ525 and pGST-ratSBP2b) selected by the mini expression tests were amplified

and titrated. These titrated viral particles were used for infection of the insect cells. Sf9 cells in 25ml cell culture (20 x 10^6 cells), infected with different viral titers (from 1 to 10 PFU/cell: Plaque-forming unit per cell) were used for protein production The baculovirus infected-cells were incubated at 27 °C for either 48 or 72 hours and then collected. The cell extract was prepared as described in 2.2.2. in the paragraph of mini expression test, and the soluble fraction was incubated with Glutathione Sepharose (Amersham). The purified proteins were loaded onto SDS-PAGE gels, analyzed by Coomasie staining or Western blotting.

References

References

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 (2006a) SECIS RNAs and K-turn binding proteins. A survey of evolutionary conserved RNA and protein motifs. *Selenium, its molecular Biology and role in human Health 2nd edition DL Hatfield (ed) Kluwer Academic Publishers* **5:** 51-61

Allmang C, Krol A (2006b) Selenoprotein synthesis: UGA does not end the story. *Biochimie* **88**(11)**:** 1561-1571

Allmang C, Wurth L, Krol A (2009) The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated. *Biochim Biophys Acta,* **doi:10.1016/j.bbagen.2009.03.003.**

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

Araiso Y, Palioura S, Ishitani R, Sherrer RL, O'Donoghue P, Yuan J, Oshikane H, Domae N, Defranco J, Soll D, Nureki O (2008) Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic Acids Res* **36**(4)**:** 1187-1199

Arner ES (2009) Focus on mammalian thioredoxin reductases - Important selenoproteins with versatile functions. *Biochimica et Biophysica Acta (BBA) - General Subjects,* **doi:10.1016/j.bbagen.2009.01.014**

Baron C, Bock A (1991) The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of Escherichia coli is the determinant for binding to elongation factors SELB or Tu. *J Biol Chem* **266**(30)**:** 20375-20379

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

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

Bosl MR, Takaku K, Oshima M, Nishimura S, Taketo MM (1997) Early embryonic lethality caused by targeted disruption of the mouse selenocysteine tRNA gene (Trsp). *Proc Natl Acad Sci U S A* **94**(11)**:** 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

Bubenik JL, Driscoll DM (2007) Altered RNA binding activity underlies abnormal thyroid hormone metabolism linked to a mutation in selenocysteine insertion sequence-binding protein 2. *J Biol Chem* **282**(48)**:** 34653-34662

Bubenik JL, Ladd AN, Gerber CA, Budiman ME, Driscoll DM (2009) Known turnover and translation regulatory RNA-binding proteins interact with the 3' UTR of SECIS-binding protein 2. *RNA Biol* **6**(1)**:** 73-83

Buettner C, Harney JW, Berry MJ (1999) The Caenorhabditis elegans homologue of thioredoxin reductase contains a selenocysteine insertion sequence (SECIS) element that differs from mammalian SECIS elements but directs selenocysteine incorporation. *J Biol Chem* **274**(31)**:** 21598-21602

Caban K, Copeland PR (2006) Size matters: a view of selenocysteine incorporation from the ribosome. *Cell Mol Life Sci* **63**(1)**:** 73-81

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 Cell Biol* **27**(18)**:** 6350-6360

Carlson BA, Xu X-M, Kryukov GV, Rao M, Berry MJ, Gladyshev VN, Hatfield DL (2004) Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. *PNAS* **101**(35)**:** 12848-12853

Carlson BA, Xu XM, Gladyshev VN, Hatfield DL (2005) Selective rescue of selenoprotein expression in mice lacking a highly specialized methyl group in selenocysteine tRNA. *J Biol Chem* **280**(7)**:** 5542-5548

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

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, 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, 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

Chao JA, Williamson JR (2004) Joint X-ray and NMR refinement of the yeast L30e-mRNA complex. *Structure (Camb)* **12**(7)**:** 1165-1176

Chapple CE, Guigo R (2008) Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. *PLoS ONE* **3**(8)**:** e2968

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

Charpentier B, Fourmann JB, Branlant C (2007) Reconstitution of archaeal H/ACA sRNPs and test of their activity. *Methods Enzymol* **425:** 389-405

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

Cléry A, Bourguignon-Igel V, Allmang C, Krol A, 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

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, Driscoll DM (2001) RNA binding proteins and selenocysteine. *Biofactors* **14**(1-4)**:** 11-16

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.

Costa M, Rodríguez-Sánchez J, Czaja A, Gelpí C (2000) Isolation and characterization of cDNA encoding the antigenic protein of the human tRNP(Ser)Sec complex recognized by autoantibodies from patients withtype-1 autoimmune hepatitis. *Clin Exp Immunol* **121**(2)**:** 364-374

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

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

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, et al. (1993) Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA([Ser]Sec). *J Biol Chem* **268**(19)**:** 14215-14223

Dikiy A, Novoselov SV, Fomenko DE, Sengupta A, Carlson BA, Cerny RL, Ginalski K, Grishin NV, Hatfield DL, Gladyshev VN (2007) SelT, SelW, SelH, and Rdx12: genomics and molecular insights into the functions of selenoproteins of a novel thioredoxin-like family. *Biochemistry* **46**(23)**:** 6871-6882

Ding F, Grabowski PJ (1999) Identification of a protein component of a mammalian tRNA(Sec) complex implicated in the decoding of UGA as selenocysteine. *RNA* **5**(12)**:** 1561-1569

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**(50)**:** 35129-35139

Dumitrescu AM, Liao XH, Abdullah MS, Lado-Abeal J, Majed FA, Moeller LC, Boran G, Schomburg L, Weiss RE, Refetoff S (2005) Mutations in SECISBP2 result in abnormal thyroid hormone metabolism. *Nat Genet* **37**(11)**:** 1247-1252

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**(3)**:** 197-208

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

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

Ferguson AD, Labunskyy VM, Fomenko DE, Arac D, Chelliah Y, Amezcua CA, Rizo J, Gladyshev VN, Deisenhofer J (2006) NMR Structures of the Selenoproteins Sep15 and SelM Reveal Redox Activity of a New Thioredoxin-like Family. *J Biol Chem* **281**(6)**:** 3536-3543

Fletcher JE, Copeland PR, Driscoll DM, Krol A (2001) The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA* **7**(10)**:** 1442-1453.

Flohe L (2009) The labour pains of biochemical selenology: The history of selenoprotein biosynthesis. *Biochim Biophys Acta*

Ganichkin OM, Xu XM, Carlson BA, Mix H, Hatfield DL, Gladyshev VN, Wahl MC (2008) Structure and catalytic mechanism of eukaryotic selenocysteine synthase. *J Biol Chem* **283**(9)**:** 5849-5865

Gelpi C, Sontheimer EJ, Rodriguez-Sanchez JL (1992) Autoantibodies against a serine tRNA-protein complex implicated in cotranslational selenocysteine insertion. *Proc Natl Acad Sci U S A* **89**(20)**:** 9739-9743

Graham TA, Weaver C, Mao F, Kimelman D, Xu W (2000) Crystal structure of a beta-catenin/Tcf complex. *Cell* **103**(6)**:** 885-896

Gromer S, Eubel JK, Lee BL, Jacob J (2005) Human selenoproteins at a glance. *Cell Mol Life Sci* **62**(21)**:** 2414-2437

Grundner-Culemann E, Martin GW, 3rd, Harney JW, Berry MJ (1999) Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. *RNA* **5**(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

Gupta M, Copeland PR (2007) Functional analysis of the interplay between translation termination, selenocysteine codon context, and selenocysteine insertion sequence-binding protein 2. *J Biol Chem* **282**(51)**:** 36797-36807

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, Portugal FH (1970) Seryl-tRNA in mammalian tissues: chromatographic differences in brain and liver and a specific response to the codon, UGA. *Proc Natl Acad Sci U S A* **67**(3)**:** 1200-1206

Hatfield DL, Gladyshev VN (2002) How selenium has altered our understanding of the genetic code. *Mol Cell Biol* **22**(11)**:** 3565-3576

Hatfield DL, Yoo MH, Carlson BA, Gladyshev VN (2009) Selenoproteins that function in cancer prevention and promotion. *Biochim Biophys Acta*

Herbette S, Roeckel-Drevet P, Drevet JR (2007) Seleno-independent glutathione peroxidases. More than simple antioxidant scavengers. *FEBS J* **274**(9)**:** 2163-2180

Hirosawa-Takamori M, Chung HR, Jackle H (2004) Conserved selenoprotein synthesis is not critical for oxidative stress defence and the lifespan of Drosophila. *EMBO Rep* **5**(3)**:** 317-322

Hirosawa-Takamori M, Jackle H, Vorbruggen G (2000) The class 2 selenophosphate synthetase gene of Drosophila contains a functional mammalian-type SECIS. *EMBO Rep* **1**(5)**:** 441-446

Howard MT, Aggarwal G, Anderson CB, Khatri S, Flanigan KM, Atkins JF (2005) Recoding elements located adjacent to a subset of eukaryal selenocysteine-specifying UGA codons. *EMBO J* **24**(8)**:** 1596-1607

Howard MT, Moyle MW, Aggarwal G, Carlson BA, Anderson CB (2007) A recoding element that stimulates decoding of UGA codons by Sec tRNA[Ser]Sec. *RNA* **13**(6)**:** 912-920

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.

Kernebeck T, Lohse A, Grötzinger J (2001) A bioinformatical approach suggests the function of the autoimmune hepatitis target antigen soluble liver antigen/liver pancreas. *Hepatology* **34**(2)**:** 230-233

Kim LK, Matsufuji T, Matsufuji S, Carlson BA, Kim SS, Hatfield DL, Lee BJ (2000) Methylation of the ribosyl moiety at position 34 of selenocysteine tRNA[Ser]Sec is governed by both primary and tertiary structure. *RNA* **6**(9)**:** 1306-1315

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

Kiss T (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**(2)**:** 145-148

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.

Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* **93**(21)**:** 11504-11509

Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigo R, Gladyshev VN (2003) Characterization of mammalian selenoproteomes. *Science* **300**(5624)**:** 1439-1443.

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

Kuhn JF, Tran EJ, Maxwell ES (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res* **30**(4)**:** 931-941

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

Leibundgut M, Frick C, Thanbichler M, Bock A, Ban N (2005) Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors. *Embo J* **24**(1)**:** 11-22

Lejeune F, Maquat LE (2005) Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol* **17**(3)**:** 309-315

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. *Biochimica et Biophysica Acta (BBA) - General Subjects,* **doi:10.1016/j.bbagen.2009.03.002**

Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein disorder prediction: implications for structural proteomics. *Structure* **11**(11)**:** 1453-1459

Liu S, Li P, Dybkov O, Nottrott S, Hartmuth K, Luhrmann R, Carlomagno T, Wahl MC (2007) Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* **316**(5821)**:** 115-120

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, 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, 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, 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

Low SC, Harney JW, Berry MJ (1995) Cloning and Functional Characterization of Human Selenophosphate Synthetase, an Essential Component of Selenoprotein Synthesis. *J Biol Chem* **270**(37)**:** 21659-21664

Lu C, Qiu F, Zhou H, Peng Y, Hao W, Xu J, Yuan J, Wang S, Qiang B, Xu C, Peng X (2006) Identification and characterization of selenoprotein K: an antioxidant in cardiomyocytes. *FEBS Lett* **580**(22)**:** 5189-5197

Lu J, Holmgren A (2009) Selenoproteins. *J Biol Chem* **284**(2)**:** 723-727

Ma S, Hill KE, Caprioli RM, Burk RF (2002) Mass spectrometric characterization of full-length rat selenoprotein P and three isoforms shortened at the C terminus. Evidence that three UGA codons in the mRNA open reading frame have alternative functions of specifying selenocysteine insertion or translation termination. *J Biol Chem* **277**(15)**:** 12749-12754.

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

Maenpaa PH, Bernfield MR (1970) A specific hepatic transfer RNA for phosphoserine. *Proc Natl Acad Sci U S A* **67**(2)**:** 688-695

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**

Marmier-Gourrier N, Clery A, Senty-Segault V, Charpentier B, Schlotter F, Leclerc F, Fournier R, Branlant C (2003) A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA* **9**(7)**:** 821-838

Martin-Romero FJ, Kryukov GV, Lobanov AV, Carlson BA, Lee BJ, Gladyshev VN, Hatfield DL (2001) Selenium metabolism in Drosophila: selenoproteins, selenoprotein mRNA expression, fertility, and mortality. *J Biol Chem* **276**(32)**:** 29798-29804

Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8**(3)**:** 209-220

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

Mizutani T (1989) Some evidence of the enzymatic conversion of bovine suppressor phosphoseryl-tRNA to selenocysteyl-tRNA. *FEBS Lett* **250**(2)**:** 142-146

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

Morozova N, Forry EP, Shahid E, Zavacki AM, Harney JW, Kraytsberg Y, Berry MJ (2003) Antioxidant function of a novel selenoprotein in Drosophila melanogaster. *Genes Cells* **8**(12)**:** 963-971

Muttenthaler M, Alewood PF (2008) Selenopeptide chemistry. *J Pept Sci* **14**(12)**:** 1223-1239

Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, Luhrmann R (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J* **18**(21)**:** 6119-6133

Nottrott S, Urlaub H, Luhrmann R (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *Embo J* **21**(20)**:** 5527-5538

Novoselov SV, Kryukov GV, Xu XM, Carlson BA, Hatfield DL, Gladyshev VN (2007a) Selenoprotein H is a nucleolar thioredoxin-like protein with a unique expression pattern. *J Biol Chem* **282**(16)**:** 11960-11968

Novoselov SV, Lobanov AV, Hua D, Kasaikina MV, Hatfield DL, Gladyshev VN (2007b) A highly efficient form of the selenocysteine insertion sequence element in protozoan parasites and its use in mammalian cells. *Proc Natl Acad Sci U S A* **104**(19)**:** 7857-7862

Novoselov SV, Rao M, Onoshko NV, Zhi H, Kryukov GV, Xiang Y, Weeks DP, Hatfield DL, Gladyshev VN (2002) Selenoproteins and selenocysteine insertion system in the model plant cell system, Chlamydomonas reinhardtii. *EMBO J* **21**(14)**:** 3681-3693

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 selenoprotein synthesis, is an intrinsically disordered protein. *biochimie*

Omer AD, Ziesche S, Ebhardt H, Dennis PP (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci U S A* **99**(8)**:** 5289-5294

Papp LV, Lu J, Holmgren A, Khanna KK (2007) From selenium to selenoproteins: synthesis, identity, and their role in human health. *Antioxid Redox Signal* **9**(7)**:** 775-806

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

Papp LV, Wang J, Kennedy D, Boucher D, Zhang Y, Gladyshev VN, Singh RN, Khanna KK (2008) Functional characterization of alternatively spliced human SECISBP2 transcript variants. *Nucleic Acids Res* **36**(22)**:** 7192-7206

Pappas AC, Zoidis E, Surai PF, Zervas G (2008) Selenoproteins and maternal nutrition. *Comp Biochem Physiol B Biochem Mol Biol* **151**(4)**:** 361-372

Patel SB, Bellini M (2008) The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **36**(20)**:** 6482-6493

Pedersen JS, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, Lander ES, Kent J, Miller W, Haussler D (2006) Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol* **2**(4)**:** e33

Rajkowitsch L, Chen D, Stampfl S, Semrad K, Waldsich C, Mayer O, Jantsch MF, Konrat R, Blasi U, Schroeder R (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* **4**(3)**:** 118-130

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

Reeves MA, Hoffmann PR (2009) The human selenoproteome: recent insights into functions and regulation. *Cell Mol Life Sci*

Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G (2007) The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res* **35**(5)**:** 1452-1464

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

Russo AA, Jeffrey PD, Patten AK, Massague J, Pavletich NP (1996) Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* **382**(6589)**:** 325-331

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

Shen Q, Wu R, Leonard JL, Newburger PE (1998) Identification and molecular cloning of a human selenocysteine insertion sequence-binding protein. A bifunctional role for DNA-binding protein B. *J Biol Chem* **273**(10)**:** 5443-5446

Shen Q, Fan L, Newburger PE (2006) Nuclease sensitive element binding protein 1 associates with the selenocysteine insertion sequence and functions in mammalian selenoprotein translation. *J Cell Physiol* **207**(3)**:** 775-783

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

Shisler JL, Senkevich TG, Berry MJ, Moss B (1998) Ultraviolet-induced cell death blocked by a selenoprotein from a human dermatotropic poxvirus. *Science* **279**(5347)**:** 102-105

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

Small-Howard AL, Berry MJ (2005) Unique features of selenocysteine incorporation function within the context of general eukaryotic translational processes. *Biochem Soc Trans* **33**(Pt 6)**:** 1493-1497

Squires JE, Berry MJ (2008) Eukaryotic selenoprotein synthesis: mechanistic insight incorporating new factors and new functions for old factors. *IUBMB Life* **60**(4)**:** 232-235

Squires JE, Stoytchev I, Forry EP, Berry MJ (2007) SBP2 binding affinity is a major determinant in differential selenoprotein mRNA translation and sensitivity to nonsense-mediated decay. *Mol Cell Biol* **27**(22)**:** 7848-7855

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, 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. co-translationally inserted during protein synthesis, probably by a mechanism analogous to that operating during selenocysteine incorporation.

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-Pierrat C, Hubert N, Totsuka T, Mizutani T, Carbon P, Krol A (1995) Selenocysteylation in eukaryotes necessitates the uniquely long aminoacyl acceptor stem of selenocysteine tRNA(Sec). *J Biol Chem* **270**(31)**:** 18570-18574.

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

Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* **27**(10)**:** 527-533

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.

Ursini F, Heim S, Kiess M, Maiorino M, Roveri A, Wissing J, Flohe L (1999) Dual function of the selenoprotein PHGPx during sperm maturation. *Science* **285**(5432)**:** 1393-1396.

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**(3)**:** 415-427

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, 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

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

Watkins NJ, Dickmanns A, Luhrmann R (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol* **22**(23)**:** 8342-8352

Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R (2000) A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**(3)**:** 457-466

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

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**(2)**:** 321-331

Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* **19**(1)**:** 31-38

Wu R, Shen Q, Newburger PE (2000) Recognition and binding of the human selenocysteine insertion sequence by nucleolin. *J Cell Biochem* **77**(3)**:** 507-516

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, Carlson B, Irons R, Mix H, Zhong N, Gladyshev V, DL H (2007a) Selenophosphate synthetase 2 is essential for selenoprotein biosynthesis. *Biochem J* **404**(1)**:** 115-120

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

Xu X-M, Mix H, Carlson BA, Grabowski PJ, Gladyshev VN, Berry MJ, Hatfield DL (2005) Evidence for Direct Roles of Two Additional Factors, SECp43 and Soluble Liver Antigen, in the Selenoprotein Synthesis Machinery. *J Biol Chem* **280**(50)**:** 41568-41575

Yuan J, Palioura S, Salazar JC, Su D, O'Donoghue P, Hohn MJ, Cardoso AM, Whitman WB, Soll D (2006) RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *PNAS* **103**(50)**:** 18923-18927

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.

Zhao R, Davey M, Hsu Y-C, Kaplanek P, Tong A, Parsons AB, Krogan N, Cagney G, Mai D, Greenblatt J, Boone C, Emili A, Houry WA (2005) Navigating the Chaperone Network: An Integrative Map of Physical and Genetic Interactions Mediated by the Hsp90 Chaperone. *Cell* **120**(5)**:** 715-727

Zhao R, Kakihara Y, Gribun A, Huen J, Yang G, Khanna M, Costanzo M, Brost RL, Boone C, Hughes TR, Yip CM, Houry WA (2008) Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *J Cell Biol* **180**(3)**:** 563-578

Zhuo P, Diamond AM (2009) Molecular mechanisms by which selenoproteins affect cancer risk and progression. *Biochim Biophys Acta,* ***doi:10.1016/j.bbagen.2009.03.004***