RECONSTITUTION OF PROCESSIVE SELENOCYSTEINE INCORPORATION IN WHEAT GERM LYSATE PROVIDES INSIGHT INTO SELENOCYSTEINE INSERTION SEQUENCE BINDING PROTEINS

by

MARK HURST PINKERTON JR

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Microbiology and Molecular Genetics

Written under the direction of

Dr. Paul R. Copeland

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2019

ABSTRACT OF THE DISSERTATION

Reconstitution of Processive Selenocysteine Incorporation in Wheat Germ Lysate Provides Insight into Selenocysteine Insertion Sequence Binding Proteins

by MARK PINKERTON

Dissertation Director:
Dr. Paul R. Copeland

A UGA stop codon is recoded to accommodate the incorporation of the 21st amino acid selenocysteine (Sec). For a UGA to be recoded a specialized set of cis and trans factors are required and consist of: an mRNA with an in frame UGA codon, a selenocysteine insertion sequence (SECIS) in the 3' untranslated region (3' UTR), a SECIS binding protein 2 (SBP2), a specific translation elongation factor (eEFSec), and a selenocysteine tRNA (Sec-tRNA[Ser]Sec). The N-terminus of SBP2 is believed to have no direct role in Sec incorporation because the C-terminus of SBP2 is sufficient for the incorporation of Sec into selenoproteins that have one Sec codon. Selenoprotein P (SELENOP) is an essential selenoprotein for male fertility and proper neuron function. SELENOP is also unique in that it contains 10 selenocysteines and is involved in selenium uptake and transport. Interestingly, an in vitro translation system using wheat germ lysate, has no endogenous selenocysteine incorporation factors, cannot synthesize full length SELENOP even when all the known factors are added. The aim of this thesis is to gain insight into the mechanism of processive eukaryotic selenocysteine incorporation by *in vitro* reconstitution and bioinformatics.

Acknowledgements

The work contained in this thesis would not have been possible without the hard work and support from my mentor and fellow lab members both past and present: Dr. Paul Copeland, Dr. Sumangala Shetty, Michael Vetick, Supriya Sinha, and Kevin Gurrero. I wish you all success in your future endeavors, and thank you for your friendship, criticism, and commiseration. Paul, thank you for your guidance and leadership, which has taught me valuable lessons at and away from the bench. I will always appreciate the risk you took to support me transferring from the master's program into the PhD program. Sorry for all of the headaches, next round is on me.

Our collaborators from the Hatfield Lab and Simonović Lab, and were generous in offering their resources and feedback throughout my research. Thanks to Marco Mariotti for help in bioinformatic analysis and technical support. I would also like to thank my committee members: Dr. Tracy Anthony, Dr. Gary Brewer, and Dr. Kiran Madura for their constructive feedback and help with direction in my research. I would like to also acknowledge Dr. Terri Kinzy and her lab for early feedback and for the ten dollars. I also thank Dr Patrick Nosker, Dr. Janet Alder, and Jessica Trusiani for their help and encouragement during the toughest time in my research.

I would like to thank my parents Mark and Angela Pinkerton for their encouragement and support throughout the years. A special thanks to my partner, Jennifer Csatari for her monumental efforts and immeasurable support over the past 4 years. Finally, my dog Ghost, for being a good boy.

Table of Contents

List of Illustrations

List of Tables

Chapter 1: Introduction

This chapter introduces selenium, selenocysteine, and selenoproteins as a basis for the following chapters and provides a general overview of the history of the field of study. Sections of Chapter 1 appeared as a book chapter titled "Eukaryotic Mechanisms of Selenocysteine Incorporation and Its Reconstitution In Vitro" in *Selenium: Its molecular biology and role in human health*: Fourth edition, published 2016 by Springer, Cham, New York, NY.

Pinkerton, M. H., & Copeland, P. R. (2018). In Vitro Translation Assays for Selenocysteine Insertion. Methods in Molecular Biology (Clifton, N.J.), 1661, 93–101.

Historical and Biological Significance of Selenium and Selenocysteine

Selenium (Se) is an essential trace element found and utilized in organisms across all domains of life and is an essential micronutrient for humans. Originally discovered by in 1817 by a Swedish chemist Jöns Jacob Berzelius. Berzelius was investigating an illness of workers in a chemical factory producing sulphuric acid, which he was a partial owner of. When the source of ore was moved to one closer to the factory, the workers fell ill and presented with a garlic odor, hair loss, brittle nails, skin rashes, and diarrhea (Oldfield, 1995; Oldfield, 2002). In the early 19th century sulphuric acid was produced by oxidizing sulphur dioxide in lead chambers, which left a residual red-brown sediment. He suspected the sediment contained an impurity and used a blowpipe to assay and isolate a new element.

This new element was described by Berzelius as having the characteristics of both tellurium and sulfur, and he stated, "This substance has the properties of a metal, combined with that of sulfur to such a degree that one would say it is a new kind of

sulfur. The similarity to tellurium has given me occasion to name the new substance selenium." (Trofast, 2011; Weeks, 1932). Berzelius was inspired by the origin for the naming of tellurium, which had been named after the Greek word for earth goddess. Due to the similarities between tellurium and this new element and due to both elements being found in the same ores. Berzelius name his discovery selenium, after the Greek word for moon goddess Selene.

Like the two faces of the moon, selenium levels in living organisms have a dichotomy: Too much selenium becomes toxic to the organism, and too little causes major essential processes of organisms to shut down (Oldfield, 1987). Selenium was first recognized as being a toxin for incidents such as the Swedish factory of Berzelius and overaccumulation of selenium in livestock grazing on accumulator plants (*Astragalus*) in selenium rich soils. Overaccumulation can causes alkali disease which presents as a loss of hair, deformed and sloughing hooves, joint erosion, anemia, and cardiovascular, hepatic and renal complications (Franke and Painter, 1936). Selenium deficiency diseases include a type of muscular dystrophy in large livestock known as white muscle disease and mulberry heart disease which results in hemorrhaging of the heart. In humans the most notable selenium deficiency diseases are Keshan disease, an underlying viral cardiomyopathy and Kashin-Beck disease which results in severe osteochondropathy and results in the destruction of cartilage in the joints. (Anonymous, 1979).

A study conducted by Schwarz and Foltz in 1957 provided the first insight on dietary requirements of Selenium. Schwartz and Foltz had been studying the origin of

dietary liver necrosis in rats when they discovered that types of yeast impacted the disease. These diseases were precipitated by a diet consisting of torula yeast, a kind of yeast which typically grows on wood liquor, a byproduct of paper production. Schwarz initially found out that switching from torula yeast to brewer's yeast prevented the disease. This finding led him to look for the missing factors in torula yeast that were present in brewer's yeast. The factors were able to be found via fractionation of proteins, one which was called "factor 3." Later, with Foltz he identified the "factor 3" as selenium (Schwarz et al., 1957). That same year both Schwarz, et al., and Patterson, et al., would independently publish in the same issue of the same journal their research on chicks developing exudative diathesis after being fed torula yeast as well (Schwarz et al., 1957; Patterson et al., 1957). Both arrived at the same conclusion: selenium supplementation of the yeast in various forms of would prevent the disease.

Selenium continued to be researched further, with different studies focusing on its unique and unknown aspects. In 1936 Painter and Franke had two significant results for the biology of selenium after their study of disease in cattle: (1) selenium was present in the protein fractions of grain sulfuric acid hydrosylates, and (2) selenium in the hydrosylates were in a compound with similar properties as cysteine (Franke and Painter, 1936). Based on those findings, Franke and Painter theorized that selenium could take the place of sulfur in proteins. In 1967 Huber and Criddle characterized the amino acid Sec and selenium was found to be essential for the enzymes formate dehydrogenase and glycine reductase in *Escherichia coli* and *Clostridium sticklandii* respectively (Pinsent, 1954; Turner and Stadtman, 1973). It was not until 1976 that

Stadtman's group was able to determine the selenium requirement for glycine reductase function in *Clostridium sticklandii* was because of a selenocysteine residue (Cone et al., 1976; Forstrom et al., 1978). After the discovery of Selenocysteine being incorporated into proteins, it was later designated as the twenty-first amino acid.

The twenty-first amino acid Selenocysteine is structurally similar to the amino acids (Ser) and cysteine (Cys) and is synthesized from a seryl acylated precursor on the Sec-tRNA$^{[Ser]Sec}$. Selenoproteins have a higher enzymatic efficiency and faster chemical reaction rates with electrophiles than their cysteine counter parts, which gives selenoproteins high redox potentials (Steinmann et al., 2008; Nauser et al., 2006; reviewed in Reich and Hondal, 2016). In addition, enzymes harboring Sec instead of Cys in their active sites are much more resistant to oxidative inhibition (Snider et al., 2013). Due to these redox characteristics it is no surprise that Sec is found in antioxidant enzymes such as glutathione peroxidases (GPXs), thioredoxin reductases (TXNRDs), and iodothyronine deiodinases (DIOs) in humans and other eukaryotes. The importance of selenium as a nutritional requirement and selenocysteine utilization varies among the species in eukaryotes. For example, fungi and higher plants do not utilize Sec and completely lack the capacity for Sec incorporation. Other organisms, notably *Drosophila*, have the required machinery but only use three selenoproteins that are not essential for survival, fertility or protection from oxidative stress (Hirosawa-Takamori et al., 2004). Even in mammals, the naked mole rat *Heterocephalus glaber*, has a reduced utilization of selenium due to a defect in GPX1 expression (Kasaikina et al., 2011).

Mechanism of Sec Incorporation in Eukaryotes

Translation occurs in three main steps**:** initiation, elongation, and termination. Generally, the process of initiation is a complex assembly of factors to recruit the ribosome subunits and Met-tRNAi to the initiator codon AUG. Elongation starts with a peptidyl-tRNA in the ribosomal P-site, an empty A-site, and a deacylated tRNA in the E-site. The complex of eukaryotic elongation factor 1A (eEF1A) and GTP delivers an amino acylated tRNA (aa-tRNA) to the A-site of the cognate codon to the anti-codon of the mRNA. GTP is hydrolyzed which reduces affinity of eEF1A to the aa-tRNA and eEF1A is released from the ribosome along with the deacylated tRNA in the E-site. The ribosome, utilizing peptidyl transferase activity will bond the amino acid of the aa-tRNA to the elongating peptide chain and then conclude the cycle when eEF2 will move the ribosome to the next codon upon hydrolysis of GTP, restarting the process until the ribosome reaches a stop codon. When the ribosome reaches one of the stop codons (UAA, UAG, or UGA) the codons are recognized by eukaryotic release factor 1 (eRF1), which begins to release the nascent polypeptide and is joined by eRF3 to enhance the reaction in a GTP dependent manner. Sec incorporation could be considered a special cycle of elongation during the process of translation which also interferes with canonical termination.

The molecular machinery for Sec incorporation in eubacteria is well characterized in both synthesis and translation, with the eukaryotic mechanism only having been deciphered within the past few decades. Sec incorporation in eukaryotes is also dependent on *cis* acting factors within the mRNA and *trans* acting factors involved

in bringing a Sec charged tRNA (Sec-tRNA$^{[Ser]Sec}$) to the ribosome to be added to the elongating nacent peptide chain. The Sec-tRNA$^{[Ser]Sec}$ recognizes the UGA stop codon, thus representing one of the exceptions to the canonical genetic code.

Specifically, recoding will occur when a selenoprotein mRNA contains a stem loop structure in the 3' untranslated region (3' UTR). This element was originally discovered to be required for the synthesis of type 1 iodothyronine deiodinase (DIO1) and was named the selenocysteine insertion sequence or SECIS (Berry et al., 1991). In a search for SECIS binding proteins (SBPs), a 120 kDa factor was found to specifically interact with the GPX4 SECIS element (Lesoon et al., 1997). This factor, eventually named SECIS Binding Protein 2 (SBP2) was shown to be required for Sec incorporation *in vitro* (Copeland et al., 2000). Soon after the discovery of SBP2 a Sec-specific elongation factor (eEFSec) was also found to be a specific Sec-tRNA$^{[Ser]Sec}$ binding factor (Fagegaltier et al., 2000; Tujebajeva et al., 2000). Recently, these core factors were shown to be the minimum requirements for Sec incorporation into proteins a reconstituted cell free in vitro system (Gupta et al., 2013). There are still underlying fundamental questions that remain about factors that may govern the efficiency and processivity of the Sec incorporation machinery. Additionally, other factors have been discovered to be involved in regulating selenoprotein expression. These factors will be discussed later in this chapter.

Recoding the UGA Stop Codon

Recoding is a translational event where the ribosome incorporates an amino acid different than that specified by the cognate codon and the canonical genetic code. This

process fundamentally describes the mechanism by which Sec is incorporated. The programmed recoding of codons for different amino acids was once unknown in biology, but studies have discovered numerous ways organisms have altered the canonical genetic code for translation.

UAA, UAG, or UGA are the standard translation termination codons for ribosomes, but only UAG and UGA appear to be the codons that are recoded in organisms (Ivanova et al., 2014). The stop codon is recoded in mycoplasma and mitochondria, both of which use UGA for tryptophan rather than for translation termination (Inamine et al., 1990; Inagaki et al., 1998). Another recoded stop codon is pyrrolysine, which is coded by UAG primarily in methanogenic archaea (Hao et al., 2002). Typically, stop codon redefinition results from a difference of tRNA specificity or an error in stop codon recognition by termination factors (Bertram et al., 2001) or suppression due to lack of tRNA-Trp in E. coli (Engelberg-Kulka et al., 1981). In Sec incorporation, however, the mechanism is specific. It requires the combination of specific factors to work in harmony with the canonical translation elongation system to recode UGA for Sec. This system is unique across all domains of life in the recoding of stop codons (Turanov et al., 2013; Donovan et al., 2010a).

Selenocysteine Incorporation Sequence (SECIS)

The translational machinery for Sec incorporation requires a mechanism to ensure that only selenoprotein mRNAs have their UGA codons recoded. Selenoprotein mRNAs contain the SECIS cis element in the 3' UTR, which is the only required cis acting element for Sec incorporation. SECIS was first reported after sequence alignments of type I iodothyronine deiodinase (DIO1) located in the 3' UTRs containing a loosely conserved sequence that was required for Sec incorporation both in vitro and in Xenopus oocytes (Berry, et al., 1991). It was elaborated further in a follow up study that the SECIS elements have been found exclusively the 3' UTR of mRNAs belonging to Sec containing proteins (Berry et al., 1993a).

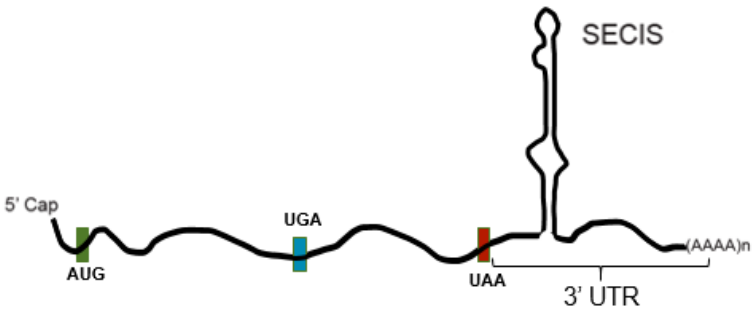
SECIS elements are variable in their sequence across species and different mRNA 3' UTRs but share certain rules of the secondary structure motifs consisting of 2 helices, an internal loop, and an apical loop. The helices and loops of the SECIS elements are used to define the two forms of the SECIS elements. Form 1 contains a relatively large apical loop and a single internal loop separated by a 12-14bp helix, while Form 2 has another short 2-7bp helix, an additional internal bulge, and a smaller apical loop (Grundner-Culemann et al., 1999). The apical loop contains an AAR residue in the apical loop, and unpaired AUGA and UGR in the 5' and 3' sides of the internal loop designated the SECIS core, both of which are essential to for translation of Sec proteins (Berry et al., 1993a; Berry et al., 1993b). While the AAR residue is required in Sec incorporation, its specific role has not been elucidated and is further confounded by the SECIS elements

SelM and SelO which contain CCX residues in place of the AAR. (Korotkov et al., 2002; Kryukov et al., 2003).

The conserved sequences of the SECIS element are imperative for the proper recruitment of factors for Sec incorporation. The AUGA of the SECIS forms non-Watson-Crick base pairs, and a central tandem of GA/AG pairs which forms a kink turn (K-turn) motif and creates a 120 degree bend in the helix (Figure 1-1). The motif is more commonly found in ribosomal RNA. The kink turn structures are found to bind to SBP2, another indispensable factor for Sec incorporation, via the kink-turn binding motif designated L7Ae (Koonin et al., 1994; Walczak et al., 1998; Klein et al., 2001; Copeland et al., 2001; Caban, et al., 2007). K-turns appear to be flexible in their conformation and are thought to act like a hinge to support biological movements (Rázga et al., 2005). In the presence of Mg^{2+} the hinge is found to be in its fully bent confirmation which can explain the phenomenon of Mg^{2+} inhibition of SECIS/SBP2 binding (Copeland and Driscoll, 2001). There also seems to be an optimal amount of Mg^{2+} for each SECIS element to work at maximum efficiency, typically around 2mM (Shetty et al., 2018). The importance of the AUGA region is emphasized in a later study of a single homozygous point mutation of the AUGA to ACGA in the SECIS element of selenoprotein N (SEPN1) in vivo causes congenital myopathies because the mutation prevents SBP2 binding to the SECIS element (Allamand et al., 2006).

The SECIS element itself demonstrates intrinsic regulation hard coded into its sequence. This phenomenon is observed by differences in UGA recoding efficiency both in vitro and in vivo of various SECIS elements in chimeric constructs (Latreche et al.,

A



B

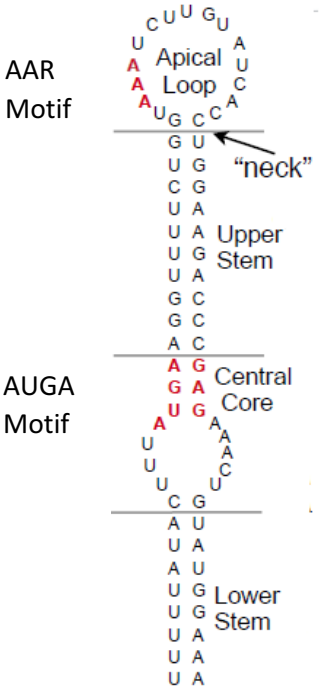


Figure 1-1. Diagram of Selenoprotein cis Factors. (A) Depiction of a typical Selenoprotein mRNA and the essential cis factors, in frame UGA (blue) and the SECIS element in the 3' UTR, required for eukaryotic Sec incorporation (B) Depiction of the SECIS element. The SECIS core is sectioned off with the AUGA motif on the 5' strand with the conserved GA(G) in red, is essential for SBP2 binding SBP2 binding. The AAR motif in the apical loop is also marked in red.

2009; Gupta et al., 2014; Shetty et al., 2018). Predictors of translation efficiency included the presence of a GC base pair in helix 2 of SECIS and a U in the 5'-side of the internal loop (Latreche et al., 2009). SECIS elements are also quite variable in the distance from the UGA, anywhere from 104-5200 nt away in known transcripts, however it also appears to be a minimal spacing requirement of 51nt downstream of the UGA (Martin et al., 1996). In addition, it does not appear that the SECIS element binding affinity to SBP2 correlates to translation efficiency (Donovan et al., 2008) nor does concentration of factors EFSec or SBP2, which makes differential translational efficiency somewhat of a mystery (Weiss and Saude 1998; Latreche et al., 2012). Currently, it is known that the SECIS element can modulate UGA recoding, yet little is known about the exact reasons why.

SECIS Binding Protein 2

SECIS elements are made functional, at least in part, by the binding of SBP2, which was originally identified as a 120 kDa protein that specifically cross-linked to the Gpx4 SECIS element (Lesoon et al., 1997). Subsequent purification and characterization led to the discovery that it is essential for Sec incorporation (Copeland et al., 2000). The known functions of SBP2 include SECIS binding, ribosome binding and transient interaction with eEFSec. Truncation and site-specific mutagenesis experiments on SBP2 were initially used to demonstrate that SBP2 contains three distinct domains: An N-terminal domain of ~400 amino acids with no known function, a central Sec incorporation domain (SID) of about 100 amino acids; and a C-terminal ~300 amino acid RNA binding domain (RBD) containing the L7Ae RNA motif (Copeland et al., 2000; Caban

et al., 2007; Copeland et al., 2001, Krol, 2002; Allmang et al., 2002; Donovan et al., 2008).

Interestingly, the SID and RBD are sufficient for all known functions of SBP2, thereby demonstrating that the SBP2 N-terminal domain is not essential for Sec incorporation *in vitro* (Copeland et al., 2000; Mehta et al 2004). *In vitro* experiments have also demonstrated that full length protein does not increase efficiency of Sec incorporation. Evidence for the possibility of a function for the N-terminal domain comes from discrete highly conserved stretches of amino acid sequence across metazoans (Donovan and Copeland, 2009). Also, mutations in humans with truncations in the N-terminal domain present with symptoms of primarily thyroid hormone deficiency (reviewed in Dumitrescu and Refetoff, 2013). Most of the protein is highly disordered, especially in the N-terminal domain, which has made structural research and full length purification difficult (Olieric et al., 2009).

The N-terminus of SBP2 could function as a regulatory domain instead of enhancement of Sec incorporation efficiency or needs to undergo modification before the N-terminal domain is in a functional conformation. The N-terminal domain is predicted to contain a lysine-rich nuclear localization sequence (NLS) (Copeland et al., 2000; Papp et al., 2006). The presence of a functional NLS was confirmed in a study of liver cells under conditions of high oxidative stress and when nuclear export was blocked by leptomycin B resulted in reduced selenoprotein levels. The reduction of Sec incorporation suggested that SBP2 may get shunted to the nucleus during oxidative stress perhaps to avoid oxidative damage (Papp et al., 2006). More evidence for the N-

terminus to have a regulatory role is SBP2 has been due to the discovery of N-terminal mutations of SBP2 introns that cause exon frame shifting. The exon frame shift mutations delete part or all of the N-terminus of SBP2 result in congenital hypothyroidism (Schoenmakers, 2010; Dumitrescu et al., 2010). Also, patients with this disorder appear express less GPX3 and SELENOP in the plasma. More research in this area would need to be performed in order to determine the mechanistic roles of the N-terminus of SBP2.

The functional relationship between the SID and RNA binding domains is complex. Even when the SID and RBD are expressed separately as individual proteins, they form a stable SECIS-dependent complex and retain all their functions in Sec incorporation in vitro apart from stable ribosomal binding (Donovan et al., 2008). SID has been described as an extension of the RNA binding domain, but does not make direct contact with the SECIS element (Bubenik et al., 2007). It does however seem to play a role in increasing SECIS binding affinity of the RNA binding domain (Donovan et al., 2008). The complexity of the SID/RBD interplay is well illustrated by the fact that mutation of the conserved IILKE 526–530 to alanine eliminated Sec incorporation and stable interaction with the RNA binding domain, but it did not affect high affinity SECIS binding (Donovan et al., 2008). This same mutation in the intact C-terminal half of SBP2 results in complete inactivation of SECIS binding. From this, the authors concluded that physical linkage between the SID and RNA binding domain constrains conformational options, but the structural significance of this awaits high resolution structure determination.

As alluded to above, one role of SBP2 that is not well understood is the ability to bind to the ribosome. Recent work that followed the initial characterization of the interaction between SBP2 and the large ribosomal subunit included mapping ribosome conformational changes by selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) and mapping of the ribosomal binding sites of SBP2 on 28S rRNA to an expansion segment 7L (reviewed in Caban and Copeland, 2006; Caban and Copeland, 2012; Kossinova et al., 2014). As tantalizing as these two findings have been, little is still known about the role that SBP2 plays when bound to the ribosome. It has been proposed that SBP2 binding to the ribosome could cause a conformational change in the A site to accommodate the larger Sec-tRNA$^{[Ser]Sec}$ or prevention of release factors to the A site (Gupta and Copeland 2007; Caban and Copeland, 2012). Resolution of this mechanism would require a substantial effort integrated with structural studies.

Sec Specific Elongation Factor

eEFSec is central to the Sec incorporation process as the factor that delivers Sec-tRNA$^{[Ser]Sec}$ to the ribosome. eEFSec was discovered based on sequence similarity to its archaeal counterpart SelB. Much like SelB, eEFSec is a GTP binding protein with roughly equal affinity for GTP and GDP thus lacks the requirement for action of a guanine nucleotide exchange factor (GEF) for functionality (Fagegaltier et al., 2000, Tujebajeva et al., 2000). Unlike SelB, eEFSec does not bind to the SECIS element. Instead, experiments suggest eEFSec forms a transient complex with SECIS bound SBP2 and delivers Sec-tRNA$^{[Ser]Sec}$ to the ribosomal A-site during translation when the ribosome encounters a UGA codon in selenoprotein mRNA.

In contrast to eEFSec, the canonical translation elongation factor 1a (eEF1A) is the workhorse during protein elongation as it carries aminoacylated tRNAs to the ribosome to allow for translation elongation. eEF1A contains 3 domains that are phylogenetically related to eEFSec. Domain I is required for binding to the ribosome and guanosine-5'-triphosphate phosphatase (GTPase) activity, which drives binding of aminoacyl-tRNA. Domain II, and Domain III are specifically involved in binding the acceptor arm of tRNA as well as interacting with its GEF, eEF1B (Noble and Song, 2008). This characterization has provided some insight to the function of the similar domains in eEFSec. However, it does not help explain the functional differences between the two elongation factors. The main functional difference between eEFSec and eEF1A begins with tRNA specificity, as eEFSec can only bind to Sec-tRNA$^{[Ser]Sec}$, while eEF1A binds to the 20 canonically charged tRNAs. Analysis of the structure of archaeal SelB with X-ray crystallography revealed a chalice like structure, which has only been previously reported in IF2/eIF5B and not eEF1A (Roll-Mecak et al., 2000). While much is known about eEF1A, the precise roles for eEFSec domains remain largely unstudied and speculative based on similarity to eEF1A and prokaryotic SelB. The relatively recent crystal structure of eEFSec demonstrated the structure was closer to the prokaryotic SelB chalice structure (Dobosz-Bartoszek et al., 2016)

The most conspicuous difference between the two elongation factors is the additional unique domain on the C-terminal end of eEFSec, Domain IV, which has been implicated in all of the known functions for eEFSec: SBP2/SECIS binding, Sec-tRNA$^{[Ser]Sec}$ binding, and GTP hydrolysis (Gonzalez-Flores et al., 2012). The interaction between

eEFSec and the SBP2/SECIS complex has only been observed in cells when $tRNA^{[Ser]Sec}$ was overexpressed, or when an electrophoretic mobility shift assay was used to capture the transient complex (Zavacki et al., 2003; Donovan et al., 2008). The fact that this interaction requires the presence of Domain IV may suggest that the SBP2/SECIS complex induces a stable conformational change in eEFSec that allows recognition of the UGA codon. This hypothesis remains to be tested.

GTP hydrolysis plays an important role in the proper function of canonical eEF1A and eEF2, however, the role GTP hydrolysis plays in eEFSec seems to be different. In eEF1A and eEF2, GTP hydrolysis is critical for the conformational changes, which is required for stepwise progression through the elongation cycle. In eEF1A and eEF2, a GEF is used after hydrolysis to exchange GDP for GTP to allow for a conformational change in Domain I to promote tRNA binding. The same change does not appear to occur based on the most recent x-ray crystallography data. Instead, GTP causes Domain IV of eEFSec to swing (Dobosz-Bartoszek et al., 2016). The prokaryotic homolog SelB has a million-fold higher affinity for Sec-$tRNA^{[Ser]Sec}$ than the GDP bound or apo form. Upon binding GTP, SelB undergoes a conformational change and then a stabilization of the SelB/GTP/Sec-$tRNA^{[Ser]Sec}$ complex occurs (Paleskava et al., 2012). Sec-$tRNA^{[Ser]Sec}$ is delivered to the ribosomal A site by SelB in the presence of the SECIS element, and then SelB hydrolyzes GTP, which causes the rapid release of Sec- $tRNA^{[Ser]Sec}$ from eEFSec (Hüttenhofer et al., 1996; Hüttenhofer and Böck, 1998). While bacterial SelB and eukaryotic eEFSec have many differences, it is likely that the fundamental mechanism of Sec-$tRNA^{[Ser]Sec}$ accommodation in the ribosomal A-site is conserved.

Sec-tRNA$^{[Ser]Sec}$

Sec-tRNA$^{[Ser]Sec}$ is the last essential component in Sec insertion for the synthesis of proteins, it is bound by eEFSec and is brought to the ribosome during Sec insertion. It has its own unique pathway of synthesis (reviewed in Xu et al., 2015; Carlson et al., 2016). In eukaryotes, the tRNA$^{[Ser]Sec}$ backbone is synthesized by the transcription of the TRU-TCA1-1 gene (formally known as TRSP and TRNAU1) and is 90 nucleotides in length, making it longer than other tRNAs, which are around 76 nucleotides long (Diamond et al., 1981). The extra nucleotide length results in Sec-tRNA$^{[Ser]Sec}$ having a modified cloverleaf 9/4 acceptor stem T stem structure, which is absent in the traditional 7/5 clover leaf form. These changes result in a longer variable and acceptor arm, as well as a longer stem helices than other typical tRNAs. The changes are believed to be important for recognition of the tRNA$^{[Ser]Sec}$ by the aminoacylation pathway components serine synthetase (SerRS), Phosphoseryl-TRNA Kinase (PSTK), Selenocysteine Synthase (SecS) as well as eEFSec. The synthesis of Sec-tRNA$^{[Ser]Sec}$ occurs on the tRNA$^{[Ser]Sec}$ by first serylation of Sec-tRNA$^{[Ser]Sec}$ by SerRS, resulting in Ser-tRNA$^{[Ser]Sec}$. tRNA$^{[Ser]}$ is also larger than most other tRNAs, this is likely why SerRS is shared by both tRNA$^{[Ser]}$ and tRNA$^{[Ser]Sec}$ and therefore recognition is not sequence dependent, but instead dependent based on the orientation of the tRNA based on the longer variable and acceptor arms (Achsel and Gross 1993; Wu and Gross, 1994). Next, PSTK will phosphorylate Ser-tRNA$^{[Ser]Sec}$ to PSer-tRNA$^{[Ser]Sec}$. Selenophosphate, which synthesized by selenophosphate synthetase 2 (SPS2) will then exchange with the added phosphate group by reaction with Selenocysteine Synthase (SecS). The final result is

Sec-tRNA$^{[Ser]Sec}$, which can undergo a unique methylation depending on selenium levels at the wobble position of the uridine on the anticondon ribose 2'-O-methylation of 5-methylcarboxymethyluridine (Diamond et al., 1993). The methylation is shown to contribute to efficient Sec incorporation in a subset of selenoproteins both *in vivo and in vitro* (Carlson et al., 2005; Carlson et al., 2017).

SECIS binding protein 2 like (SBP2L)

While the core essential factors involved in Sec incorporation have been shown to be sufficient for Sec incorporation, there are many other factors that are implicated in Sec incorporation either by direct experimental evidence or phylogenetic relationships (Gupta et al., 2013). The most striking example of the latter is the SECIS binding protein 2 like protein (SBP2L). SBP2L was identified via BLAST searches based on its similarity to SBP2 in the C-terminal domain where it shares a 46% amino acid identity (Copeland et al., 2000). Like SBP2, SBP2L has both an RNA binding domain and Sec incorporation domain and can specifically bind to the AUGA core. Human SBP2L is not functionally active in Sec incorporation in vitro, and there currently is no direct evidence of any function despite sharing many motifs with SBP2 (Donovan and Copeland, 2009).

The lack of a discernable function for SBP2L is not fully clarified by phylogenetic analysis of SBP2L's its origins. It appears that SBP2 and 2L are paralogs that diverged after a gene duplication event during early evolution of vertebrates (Donovan and Copeland, 2009). Additionally, vertebrate SBP2L seems to be more closely related to invertebrate SBP2 than vertebrate SBP2 based on the number of conserved regions

found between them. Many deuterostomes do not have SBP2, but still retain SBP2L as the SECIS binding protein, suggesting in some organisms SBP2L is active in Sec incorporation (Donovan and Copeland, 2009). Currently, it is believed that the divergence of SBP2 and SBP2L in vertebrates caused SBP2L to lose its ability to support Sec incorporation. While not functional for Sec incorporation *in vitro*, it is still possible SBP2L might serve some other undiscovered function *in vivo*. Despite the lack of direct evidence, it is notable that mice and cells lacking SBP2 retain substantial selenoprotein synthesis capacity, which is possibly supported by an as-yet, undetermined function for SBP2L (Seeher et al 2014; Dubey and Copeland, 2016).

Another identified non-essential SECIS binding protein identified is eukaryotic initiation factor 4AIII (eIF4A3). eIF4A3 is an RNA dependent ATPase, ATP dependent RNA helicase, and a DEAD-box protein family member that was found to mediate regulate Sec incorporation (Budiman et al., 2009). While eIF4A3 is similar to the two other isoforms of eIF4A (I and II), it appears that is functionally distinct. In addition to its function in nonsense mediated mRNA decay, eIF4A3 binds to the GPX1 SECIS element at both the internal and apical loop (Gehring et al., 2005). Binding to the SECIS by eIF4A3 prevents SBP2 binding and therefore inhibits Sec incorporation in vitro. It also appears to have a differential binding affinity, specifically to the GPX1, but not the GPX4 SECIS element, giving it a role in selenoprotein regulate selenoprotein synthesis (Budiman et al., 2011). In McArdle 7777 rat hepatoma cells, eIF4A3 expression becomes upregulated in the absence of Se, and in turn reduced Gpx1 levels while Gpx4 levels were unaffected (Budiman et al., 2009). Thus, eIF4A3 has emerged as a potentially key factor in

determining the hierarchy of selenoprotein expression when Se becomes limiting (reviewed in Sunde and Raines., 2011).

A SECIS binding ribosomal protein L30 (RPL30) is another SECIS binding protein implicated in regulation of Sec incorporation (Chavatte et al., 2005). It is a small (14.5 kDa) protein that is part of the large ribosomal subunit (Vilardell et al., 2000). Like SBP2, RPL30 also contains an L7Ae RNA binding motif and competes with SBP2 binding to the SECIS element *in vitro* (Chavatte et al., 2005; Bifano et al., 2013). Although RPL30 was shown to stimulate Sec incorporation in transfected cells, it is not yet known whether it is essential for Sec incorporation if it can bind to all SECIS elements. The SECIS binding activity of RPL30 is consistent with the phylogenetic analysis, which concluded that the L7Ae motif of SBP2 arose from RPL30 as well (Donovan and Copeland, 2009). RNase footprinting assays show the binding of SBP2 and RPL30 have some overlap on the SECIS element, but also have their own individual sites as well [Bifano et al., 2013]. The current model of RPL30 activity is in promoting dissociation of the SBP2/SECIS complex by binding to SECIS allowing for canonical ribosomal elongation to continue.

Selenoprotein P and Processive Sec Incorporation

The plasma selenoprotein, selenoprotein P (SELENOP), is unique because it possesses multiple UGA codons in the coding mRNA. In humans and rats, there are 10 Sec residues, but it is highly variable between other eukaryotic species. Most of the Sec residues tend to be concentrated at the C-terminal end of the protein, while the position of the first UGA codon is highly conserved. Another unique feature of the SELENOP mRNA is a long 800 nucleotide 3' UTR containing both types of SECIS elements,

Form 1 and Form 2. *In vitro* translation experiments in rabbit reticulocyte lysate have demonstrated the efficiency of Sec incorporation into single UGA transcripts is relatively low at about 5–8% (Mehta et al., 2004). With such a low efficiency, it would seem to be impossible to efficiently create proteins with multiple UGAs, but the 26 μg/ml of SELENOP protein observed in plasma and the high SELENOP levels found in tissue culture suggest that the efficiency is much higher *in vivo* (Yang et al., 1989). Recent work has indicated that the efficiency of Sec incorporation at UGA codons downstream of the first UGA is much more efficient (Fixsen and Howard, 2010). Analysis of UGA redefinition in vivo using ribosomal profiling also supports a higher than 10 % UGA redefinition efficiency in hepatic selenoprotein biosynthesis (Howard et al., 2013). Two questions emerge when considering the requirements for SELENOP synthesis: how is a protein with 10 Sec codons made efficiently, and how are the multiple Sec residues incorporated processivity in order to generate full length SELENOP protein?

It has been proposed that SELENOP translation efficiency and processivity is regulated by cis acting factors in the 3' UTR of SELENOP involved with recoding of UGAs for Sec incorporation. Some of these *cis* elements were hypothesized in initial sequence analysis of SELENOP looking for conserved regions across species where two regions in the 3' UTR surrounding and including the two SECIS elements were found to be highly conserved (Hill et al., 1993). It was hypothesized that the two SECIS elements were involved in processive Sec incorporation; however, *in vitro* tests showed neither Form II SECIS nor most of the 3' UTR is required for synthesis of full length SELENOP (Stoytcheva et al., 2006; Fixsen and Howard, 2010; Shetty et al., 2014). Swapping the SELENOP

3'UTR with other selenoprotein 3' UTRs resulted in reduced efficiency, but processive production of full length SELENOP protein in vitro was still observed, indicating clear separation of efficiency and processivity (Fixsen and Howard, 2010, Shetty et al., 2014). Recent, experiments of SECIS I and II chimeras demonstrated that the apical loop and the upper stem of SECIS I was another *cis* requirement for processive Sec incorporation. It seems that the SECIS elements from other Selenoprotein mRNAs when added to a Selenoprotein P construct had highly variable capacities for full length Selenoprotein P synthesis and could be modulated by Mg^{2+} (Shetty et al., 2018).

In vitro Translation Systems for Studying Sec Incorporation

Cell free in vitro translation systems have been key tools for investigators in determining the factors involved with Sec incorporation. The identification, validation and characterization of both SBP2 and eEFSec has in large measure taken place in two commercially available *in vitro* translation systems: rabbit reticulocyte and wheat germ lysates. Despite both cell free systems having limitations, they have become a reliable means to answer some of the fundamental questions in translation and Sec incorporation. Prior to the discovery of SBP2, the rabbit reticulocyte lysate system was used to explore the ability of $tRNA^{[Ser]Sec}$ to suppress translation termination [68, 69]. The use of rabbit reticulocyte lysate expanded significantly when it was found to be nearly devoid of all SBP2 in the system This finding paved the way for the formal proof that SBP2 is necessary for Sec incorporation (Copeland et al., 2000), and extensive characterization of its domain functions (Caban et al., 2007; Latrèche et al., 2009; Donovan et al., 2008; Mehta et al 2004; Howard et al 2007). The physiological basis for

the lack of SBP2 in this system is an interesting topic that has not been investigated. In the case of eEFSec, there is no evidence that the factor is limiting in rabbit reticulocyte lysate, and its biochemical characterization required the development of a new system.

The commercially available wheat germ lysate system was used to attempt to reconstitute Sec incorporation with only the known factors, taking advantage of the fact that higher plants do not utilize Sec and do not have any of the Sec incorporation factors. Initially it was found that the addition of SBP2, eEFSec, Sec-tRNA$^{[Ser]Sec}$and mammalian ribosomes was sufficient to promote robust Sec incorporation into a luciferase reporter construct harboring SELENOP 3' UTR (Gupta et al., 2013). Subsequent work with a different batch of wheat germ lysate demonstrated that mammalian ribosomes were not required (Shetty et al., 2014), thus establishing that SBP2, eEFSec and Sec-tRNA$^{[Ser]Sec}$ are sufficient to promote Sec incorporation even with ribosomes that did not evolve to work with these factors.

The wheat germ lysate (WGL) system has tremendous potential in terms of providing a test bed for all of the fundamental mechanistic questions surrounding both Sec incorporation and Sec-tRNA$^{[Ser]Sec}$ synthesis. For example, three different SECIS elements showed markedly different Sec incorporation efficiencies (Gupta et al., 2013), consistent with those previously observed in rabbit reticulocyte lysate (Latrèche et al., 2009). In addition, it was recently found that while full length SELENOP protein can readily be made in rabbit reticulocyte lysate, only a single Sec incorporation event was observed in wheat germ lysate thus providing strong evidence for the existence of a yet unidentified factor required for processive Sec incorporation (Shetty et al., 2014). This

wheat-based system is primed to answer key questions regarding the mechanism of Sec incorporation. The assay should provide insight to fundamental questions regarding the requirements for SELENOP synthesis and eEFSec Sec-tRNA$^{[Ser]Sec}$ specificity, the extent to which our knowledge of these two processes have been elusive.

Although much progress in understanding the mechanism and regulation of Sec incorporation questions remain. Understanding of the essential Sec incorporation factors and their fundamental roles provides insight to the complexity of selenoprotein regulation. Over the years expanded knowledge of the mechanism of all the components of Sec incorporation has provided a complete picture of the sources of diseases previously unknown and will continue to do so. Many questions remain in the basic mechanism of Sec incorporation involving eEFSec complex formation with SBP2/SECIS and Sec-tRNA$^{[Ser]Sec}$ delivery to the ribosome. It is also unknown Sec-tRNA$^{[Ser]Sec}$ is accommodated into the ribosome or the role and mechanism of GTP hydrolysis for eEFSec, nor the temporal sequence of events. It is unknown if all the factors required for the regulation of Sec incorporation in vivo have been discovered due to the difference in efficiency between cell free extracts and tissue culture. *In vitro* translation systems have helped immensely in shedding light on fundamental questions of Sec incorporation so that the molecular mechanism of Sec incorporation can be deciphered.

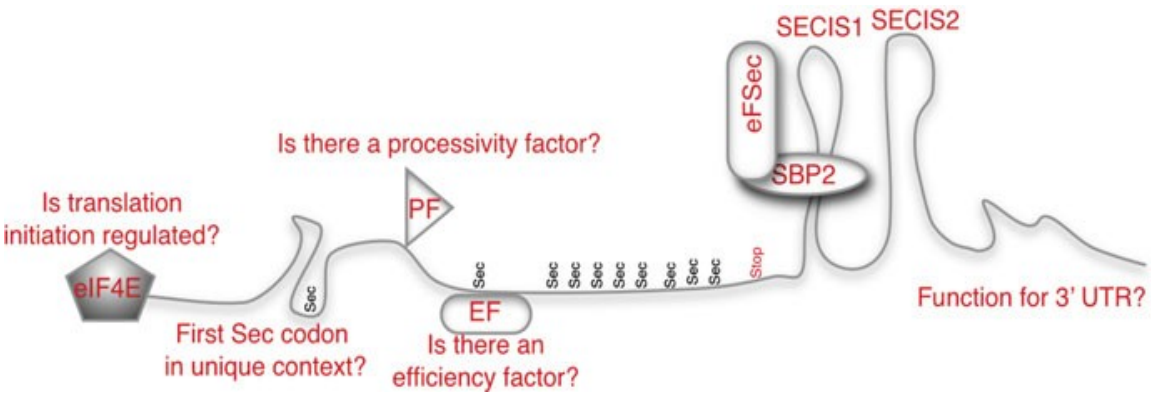


Figure 1-2. Mechanistic Questions of SELENOP Synthesis. An illustration of the many mechanistic questions that can be answered about the mechanism of SELENOP synthesis using a wheat-germ lysate based in vitro translation system. The figure shows the SELENOP mRNA containing ten Sec codons and two SECIS elements along with the known (SBP2, eEFSec) and a proposed processivity (PF) and efficiency (EF) factors. The potential role of codon context, regulated translation initiation and a function for the conserved non-SECIS portions of the SEPP1 3’ UTR are also depicted.

Chapter 2: Reconstitution of SELENOP Synthesis in Wheat Germ Lysate

Part of the Materials and Methods section appears in a book chapter from Selenoproteins; Methods and Protocols. Methods in Molecular Biology, Vol. 1661):

Pinkerton, M. H., & Copeland, P. R. (2018). In Vitro Translation Assays for Selenocysteine Insertion. Methods in Molecular Biology (Clifton, N.J.), 1661, 93–101.

Introduction

In vitro Sec incorporation and Selenoprotein P

For Sec to be incorporated into proteins, an in frame UGA is recoded from a canonical stop codon. *In vitro* lysates have been important in understanding the minimal essential components for this process. The process of recoding UGA from a translational stop codon to a Sec codon requires both *cis*-acting elements *and trans*-acting factors. The essential *cis* element required for recoding is a conserved RNA hairpin loop called the Sec insertion sequence (SECIS) element in the 3' UTR of a selenoprotein mRNA downstream of the in frame UGA. One of the essential *trans* factors consist of a SECIS binding protein (SBP2), which can be split by its a C-terminal domain and N-terminal domain. The C-terminal domain (CTSBP2) confers all the currently known functions for Sec incorporation *in vitro* and can be split into two functional domains: the RNA binding domain (RBD) and the Sec incorporation domain (SID). The purpose of the RBD is to bind to the SECIS element and the SID is required for transient binding to the specialized Sec elongation factor (eEFSec) to deliver a selenocysteine tRNA (Sec-tRNA$^{[Ser]Sec}$) to the ribosome during translation of an in frame

UGA codon. These factors have been elucidated, in part, by *in vitro* translation assays such as rabbit reticulocyte lysate (RRL) and wheat germ lysate (WGL).

Cell free *in vitro* translation systems have been utilized for decades in studying the mechanism and the components of Sec incorporation. The commonly used systems are RRL, which contains native Sec incorporation factors and the plant-based WGL which contains no Sec incorporation factors. The commercially available RRL is a useful system to study the functional domains of SBP2 because RRL lacks a sufficient amount to translate selenoproteins (Copeland et al., 2000; Metha et al., 2004; Caban et al., 2007; Howard et al., 2007; Donovan et al., 2008; Latrèche et al., 2009). Since, RRL still contains an abundance of eEFSec and other mammalian Sec incorporation factors it is mainly limited to the functional study of SBP2 and *cis* factors. Experiments in RRL, discovered that only the C-terminus of the protein was required for full Sec incorporation activity (Copeland et al., 2000). After it was discovered that CTSBP2 performs all of the known SBP2 functions it has become standard practice to use only CTSBP2 for the translation of selenoproteins *in vitro*. Another reason for the exclusion of the N-terminal domain was in part due to purification difficulty involved in the full length of recombinant protein from low efficiency and degradation.

The extracellular glycoprotein SELENOP is of interest in Sec incorporation because it contains 10 Sec codons in humans as well as two SECIS elements (Berry et al., 1993; Hill et al., 1993). The 10 Sec codons create a conundrum because single UGA Sec incorporation in RRL and WGL has been determined to have a low efficiency of about 5-10%, however translation of SELENOP in RRL results in primarily full-length product

(Mehta et al., 2004; Gupta et al., 2007; Shetty et al., 2014). *In vivo* SELENOP is also abundant in plasma and in tissue culture media suggesting that synthesis is much more efficient than suggested *in vitro* (Yang et al., 1989). Later work has added some insight into this observation, demonstrating the efficiency of Sec incorporation after the first UGA is greatly increased (Fixsen and Howard, 2010; Shetty et al., 2014). Yet, the translation of SELENOP mRNA in wheat germ lysate with the known Sec incorporation factors results in a primarily truncated product corresponding to the second Sec codon (Shetty et al., 2014). This provides substantial evidence for the requirement of a mammalian factor or a set of factors present in RRL for processive Sec incorporation.

Speculation of a Processive Sec incorporation Factor

Previous research suggests the possible involvement of the aminoacylation components for processive Sec incorporation. Most notably, Shetty et al., 2014 observed a six-fold increase in full length SELENOP synthesis upon selenium supplementation of RRL. This effect was unique to processivity and SELENOP as the same effect was not observed in single UGA constructs. Previous *in vitro* experiments also saw no increase with the addition of added total tRNA from HepG2 cells, which are rich source of Sec-tRNA$^{[Ser]Sec}$. They speculated that the effect could be related to a channeling system of the Sec aminoacylation components for processive Sec incorporation. Some previous studies on the mammalian translation systems suggested a high level of organization of the aminoacylation components *in vivo*, where they were found to be organized along the cytoskeleton (Sivaram and Deutscher, 1990; Negrutskii and Deutscher, 1991; Stapulionis

and Deutscher, 1995). Taken a step further, this could also explain why single UGA synthesis is inefficient *in vitro.*

The aminoacylation components for Sec-tRNA$^{[Ser]Sec}$ synthesis according to the current knowledge seems to be a likely candidate for the requirement for processive Sec incorporation. Plants have lost the components of Sec incorporation at some point early in their evolution, and a key difference between WGL and RRL is the presence of the Sec-tRNA$^{[Ser]Sec}$ aminoacylation components. Indeed, Sec-tRNA$^{[Ser]Sec}$ aminoacylation is a major difference between mammalian and plant cell free lysates, but several other factors may be also be involved. One difference could be something simple, such as the differences in important ions such as magnesium. Mg^{2+} is an important component of SBP2 binding to the SECIS element and in the correct conformation of the SECIS element. The extent of the importance of Mg^{2+} on processive Sec incorporation *in vitro*, has been established in RRL with (Shetty et al., 2018). The main implications from this report showed the SECIS elements all have differential UGA recoding efficiencies and capabilities for processive Sec incorporation. The mechanism for such differential in Sec incorporation efficiency and processivity would presumably stem from a conformational change within SBP2 upon binding to different SECIS structures. The result of the conformational change could be tighter binding to the SECIS element, eEFSec, or the ribosome. More research would need to be conducted to see if this is indeed the case.

SBP2 is one major factor that could contribute to SELENOP synthesis due to the addition of the recombinant C-terminal SBP2 instead of the endogenous protein. It does not seem like the N-terminal domain is required for processivity, but SBP2 is intrinsically

disordered and may need post translational modifications or chaperone interaction in order to be folded in the proper conformation (Oleric et al., 2009). Evidence for post translational modifications of SBP2 come from *in vitro* translation experiments comparing *in vitro* translated SBP2 and bacterial recombinant SBP2 in the translation of a UGA + SECIS luciferase reporter construct. SBP2 *in vitro* translated in a mammalian construct appears to be 50 to 100 fold more efficient in Sec incorporation than with a recombinant protein (Bubenik et al., 2014). Previous studies have shown it interacts with the chaperone HSP90 as well (Boulon et al., 2008). There could also be differences between wheat ribosomes and mammalian ribosomes as well at the binding sites of SBP2. The reported binding sites of SBP2 on mammalian ribosomes are H89, ES31, and ES7L-E (Caban et al., 2012; Kossinova et al., 2014). It is unknown if the paralog SBP2L binds at these same segments of the ribosome, however it appears to be found in more abundance in RRL.

A cell free translation system devoid of Sec incorporation components is necessary to independently study the mechanistic questions revolving around processive Sec incorporation. We first wanted to determine the minimum factors required for processive Sec incorporation, which made the WGL assay necessary as it has no Sec incorporation factors. It is also unknown if the translational machinery contained in WGL is capable of processive Sec incorporation. Therefore, our first goal for this chapter was to reconstitute processive Sec incorporation in WGL, and the second goal was to determine the existence of a *trans* factor or set of factors required.

Materials and Methods

Constructs and mRNA synthesis - All constructs used for mRNA synthesis for use in vitro was linearized with NotI (New England Biolabs) and capped mRNA was synthesized by T7 in vitro transcription (mMessage, Ambion). The Rat SELENOP construct containing the coding region, full 3' UTR, C-terminal FLAG tag, was made by isolating SELENOP from McArdle 7777 cells and cloned by Topo-TA ligation into PCDNA 3.1 (Invitrogen) as described in Shetty et al., 2014. SecS, SPS2$^{(U24C)}$, and PTSK were synthesized based on *Drosophila melanogaster* sequences by GENEWIZ. The coding sequences from the constructs were then Topo-TA ligated into PCDNA 3.1 (Invitrogen).

Recombinant Protein expression and purification: Recombinant Xpress/ His-tagged C-terminal SBP2 (XH-CTSBP2) was purified as previously described in (Kinzy et al., 2005) and (Pinkerton and Copeland, 2017). For recombinant eEFSec Preparation His tagged human eEFSec was a gift from the Simonovic Lab and was prepared using the methods described in (Dobosz-Bartoszek et al., 2016). Quantitation of all recombinant proteins were performed via densitometry using Imagequant (GE Healthcare) of a coomassie stained gel using a bovine serum albumin or ovalbumin standard.

75Se labeling of HAP1 Cell lines: HAP1 cell lines were purchased from Horizon and contain a 181bp insertion in exon 13 of coding exon of SECISBP2 (NM_024077) or a 290bp insertion in exon 12 or SBP2L (NM_001193489). Labeling of the HAP1 cell was performed in serum free media (RPMI 1640) for 24 hours with supplemented 100nM ($Na_2[^{75}Se]O_3$) Specific activity ~1000 Ci/mmol (University of Missouri Research Reactor). After 24 hours the cells were washed in PBS then lysed in a NP40 lysis buffer (1% NP-40,

50 mM Tris-HCl (pH 8.0), 150mM sodium chloride, and 1x Roche cOmplete protease inhibitor).

Aminoacyl tRNA Purification: Rat testis total aminoacylated tRNA was purified from rat testicles from Pel-Freez Biologicals (Rogers, AR) and prepared via the method by Gupta et al., 2013.

Preparation of ^{75}Se Sec-tRNA$^{[Ser]Sec}$: ^{75}Se radiolabeled Se used in this chapter was obtained in the form sodium selenite ($Na_2[^{75}Se]O_3$), Specific activity ~1000 Ci/mMol, purchased from the University of Missouri Research Reactor. All ^{75}Se Sec-tRNA$^{[Ser]Sec}$ was examined by scintillation counter reading, absorbance at 260nm and acid urea gel electrophoresis followed by PhosphorImager analysis. RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$ tRNA (~6000 cpm/pmol) was a generous gift from the Hatfield Lab and was prepared via the method described by (Hatfield D, et al., 1991; Carlson, et al., 2018).

^{75}Se Sec-tRNA$^{[Ser]Sec}$ purification from HepG2 cell lines: 6 T-75 flasks of HepG2 cells were grown in HepG2 cells that were grown in Eagle's minimal essential medium (EMEM) supplemented with Earle's balanced salt solution, L-glutamine, 10% FBS, 50nM cold sodium selenite to approximately 70-80% confluence. When cells reached the desired confluence, the media was removed then the plates were washed with 10ml 1x PBS. 6mL DMEM medium supplemented with 10% FBS and 100nm $Na_2[^{75}Se]O_3$ (^{75}Se) was added and the cells were incubated in a tissue culture incubator for 16 hours before harvesting. After 16 hours cells were washed with PBS and then lysed in 0.5mL per flask of a 1% NP40 tRNA lysis buffer (10 mM NaOAc pH 4.8, 650 mM NaCl, 10 mM MgOAc, 2mM DTT, 1 mM EDTA). Cells were scraped and transferred to a 15mL conical tube and

spun at 10,000 x g for 15min at 4°C. Supernatant was removed and transferred to an ultracentrifuge tube and spun at 300,000 x g at 4°C for 30 min in a Beckman 70ti rotor (65,000 rpm). The supernatant was transferred to a new tube and spun again at 300,000 x g at 4°C for 30 min in the same rotor. The supernatant was collected and measured. 1/5th volume of a tRNA buffer (50mM NaOAc pH 4.8, 650mM NaCl, 10mM MgOAc, 1mM EDTA) was added so that the final concentration was equal to 10 mM NaOAc pH 4.8, 650 mM NaCl, 10 mM MgOAc, 1 mM EDTA. The solution was then phenol extracted in low pH 4.8 and vortexed. Then the solution was spun at 10,000 x g for 30min at 4°C. The aqueous layer was removed then extracted again with pH 4.8 phenol, vortexed and spun again at 10,000x g for 30min at 4°C. The aqueous layer was removed and 3x volumes of -20°C 200 proof ethanol was added then the entire solution was incubated on dry ice for 15min. After the dry ice incubation, the solution was spun at 10,000x g for 30min at 4°C. The ethanol was removed and washed with 70% ethanol and respun at 10,000x g for 30min at 4°C. After the ethanol was removed the pellet was resuspended in nuclease free H2O. Our specific activity for the prep used in this chapter was 4297cpm/ul.

Fractionation of Mammalian Lysate for addition to Wheat Germ Lysate: RRL fractions were made from translational grade micrococcal nuclease treated RRL (Promega). A diagram of the process is shown in Figure 2-6. A Sorvall MX120 centrifuge (Thermo Scientific) was used with a S80AT2 rotor was used for all spins. 200µl of RRL was centrifuged at 300,000x g for 30 min at 4°C to separate crude ribosomes from the supernatant. The supernatant was saved for later use immediately after the second

spin. The crude ribosomes were washed in ribosome buffer A (0.25M sucrose, 2mM $MgCl_2$, 0.05M Tris-HCL pH 7.5, 1mM DTT 0.1mM EDTA) then resuspended in one fifth the original volume of 80 mM HEPES-KOH, pH 7.5, 100 mM KCl, 2 mM DTT, and 0.5 mM $MgCl_2$ or resuspended in a high salt buffer (0.5M KCl, .25M sucrose, 2mM MgCl2, 50mM Tris-HCL pH 7.5, 1mM DTT 0.1mM EDTA) and spun again 2x 300,000x g spins for 30min at 4°C. The supernatant was buffer exchanged 3x via a Bio-Rad Bio-Spin® Columns with Bio-Gel® P-30 equilibrated with 80 mM HEPES-KOH, pH 7.5, 100 mM KCl, 2 mM DTT, and 0.5 mM $MgCl_2$. The pellet was washed briefly in ribosome buffer A and resuspended in 1/5 the original volume of 80 mM HEPES-KOH, pH 7.5, 100 mM KCl, 2 mM DTT, and 0.5 mM $MgCl_2$. Preparation of crude fractions from HepG2 cell lines were lysed via hypotonic solution (10 mM Tris-HCl, pH 7.5) then passed through an 18g needle then spun at 16,000x g for 30min. The supernatant was fractionated in the same way as Rabbit Reticulocyte lysate. A diagram of the purification strategy for RRL is pictured in Figure 2-6 and a coomassie stain of the fractions are shown in Figure 2-7.

WGL In vitro assay (Selenoprotein P): Using a rat selenoprotein P construct containing the coding region and the full 3' UTR. Each 12.5 µl reaction contained 6.25 µl wheat germ extract (Promega), 320 nM recombinant SBP2, 320 nM wild-type His-tagged eEFSec, 20 µM amino acid mix, 200 ng of SELENOP mRNA and 75ng/µl (6000cpm) RPC-5 pool of methylated and unmethylated (mcmUm and mcmU) ^{75}Se Sec-tRNA$^{[Ser]Sec}$. For ^{35}S labeling 1.25 µl of 11 mCi/mL ^{35}S Cys was used (Perkin Elmer Waltham, MA) with 1.25ug total amino acylated tRNA from rat testis. The reactions were incubated at 25°C for 2 h.

RRL In vitro assay (Selenoprotein P): 12.5µl Rabbit reticulocyte lysate reactions contained 320 nM recombinant SBP2, 320 nM wild-type His-tagged eEFSec, 20 µM amino acid mix, 200 ng of SELENOP mRNA and 75ng/µl (6000cpm) RPC-5 pool of methylated and unmethylated ^{75}Se Sec-tRNA$^{[Ser]Sec}$ (Carlson et al., 2017). Samples were analyzed using SDS PAGE electrophoresis, the entire reaction was loaded for WGL while 4µl was used for RRL. The gels were dried and exposed to a phosphoimage screen and scanned on a Typhoon Phosphoimager (GE Healthcare).

Invitro synthesis of aminoacylation components in WGL: We used a fraction of each of the whole WGL translation reactions of each aminoacylation component and combined it together with a fresh WGL reaction for the translation of another selenoprotein mRNA. 250ng of SecS, SPS2$^{(U24C)}$, and PSTK mRNA was used to translate each protein. These reactions consisted of half pretranslated WGL reactions of SecS, SPS2$^{(U24C)}$, and PSTK and a standard reaction for SELENOP.

Results

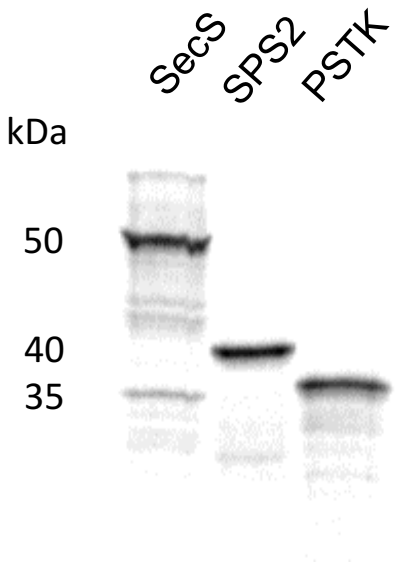
Previously published results have demonstrated the importance of selenite for processive Sec incorporation in RRL. With increasing amounts of selenite added to RRL up to 500nM selenite increased full length SELENOP synthesis approximately 4-fold compared to no addition of selenite (Shetty et al., 2014). This effect was not seen with the addition of HepG2 total tRNA, a rich source of Sec-tRNA$^{[Ser]Sec}$ (Shetty et al., 2014). The results from the addition of selenite to RRL along with the lack of WGL to suggests the possibility of selenite being channeled. To test this, we attempted to reconstitute an *in vitro* aminoacylation system in WGL. Our collaborators from the Simonović lab

have been unsuccessful in recombinant purification of functional proteins for all of selenocysteine aminoacylation components (personal communications). Therefore, we attempted to pretranslate SecS, SPS2$^{(U24C)}$, and PSTK constructs in separately in WGL (Figure 2-1, A). We then used a fraction of each of the whole WGL translation reactions of each aminoacylation component and combined them together with a fresh WGL reaction for the translation of another selenoprotein mRNA.

We wanted to first test the ability of the *in vitro* translated Sec aminoacylation components to be able to aminoacylate tRNA$^{[Ser]Sec}$. To check for Sec aminoacylation of tRNA$^{[Ser]Sec}$ we used rat testis total tRNA (another rich source of Sec-tRNA$^{[Ser]Sec}$), which was deacylated in 1 M Tris (pH 8.0) at 37°C for 1 h to ensure it was completely deacylated and *in vitro* transcribed tRNA$^{[Ser]Sec}$. These reactions contained half of the pretranslated WGL reactions of SecS, SPS2$^{(U24C)}$, and PSTK and the other half of the reactions contained rat testis deacylated total tRNA or *in vitro* transcribed tRNA$^{[Ser]Sec}$, and 100nM ^{75}Se (Figure 2-1, B). We found only a slight increase in ^{75}Se Sec-tRNA$^{[Ser]Sec}$ compared to the negative control reactions (Figure 2-1, B, lanes 1 and 3). We also did not find *in vitro* transcribed tRNA$^{[Ser]Sec}$ to have been aminoacylated either (Figure 2-1, B, lane 4).

Since we only observed a slight difference between ^{75}Se Sec-tRNA$^{[Ser]Sec}$ purified from our negative control compared to the reaction containing *in vitro* translated Sec aminoacylation components we wanted to test if tRNA$^{[Ser]Sec}$ was being serylated first by wheat SerRS. This concern was because of the first step of Sec aminoacylation requires the charging of tRNA$^{[Ser]Sec}$ by seryl-tRNA synthetase with serine to yield Ser-tRNA$^{[Ser]Sec}$.

A



Wheat Germ 35S Cys

B

Lane	1	2	3	4	5	6	7
Lysate	Wheat	Retic	Wheat	Wheat	Wheat	Wheat	Wheat
Aminoacylation Components	-	-	+	+	-	-	-
Rat Testis deacylated total tRNA	+	+	+	-	+	-	-
in vitro Sec-tRNA	-	-	-	+	-	+	-
^{75}Se Sec-tRNA	-	-	-	-	-	-	+

Acid Urea Gel ^{75}Se Wheat Germ reaction

Figure 2-1. *In vitro* Translation of Sec Aminoacylation Components in Wheat Germ Lysate and Testing of *in vitro* Aminoacylation. A. *in vitro* translation of SecS (53kDa), SPS2 (41kDa), and PSTK (32kDa) in WGL **B.** ^{75}Se was added to WGL or RRL with and without the presence of *in vitro* pretranslated total WGL reactions of the Sec aminoacylation components (SecS, SPS2, and PSTK) as indicated. Reactions were pH 4.3 phenol extracted and ran on an 8% acid urea gel for autoradiography analysis. tRNA source of each reaction is indicated.

It could be possible that the difference between mammalian $tRNA^{[Ser]Sec}$ and plant $tRNA^{[Ser]}$ could cause inefficient Ser charging by plant SerRS. We felt this would be unlikely, however, as $tRNA^{[Ser]}$ is generally known to have a longer stem arm than other tRNAs and the stem and variable arm and is found to be not nucleotide dependent for binding to the active site even in *Arabidopsis* (Wu and Gross., 1993; Asahara et al. 1994; Lenhard et al., 1999; Rokov-Plavec et al. 2004; Kekez et al., 2016). To determine if $tRNA^{[Ser]Sec}$ was being aminoacylated with Ser we used ^{14}C Ser labeling in a WGL reaction with deacylated tRNA and pretranslated aminoacylation components in WGL. The reactions were acid phenol extracted after the reactions were complete and resolved on an acid urea gel. We found that the addition of the deacylated rat testis tRNA did increase ^{14}C tRNA labeling, but with the further addition of the pretranslated aminoacylation components there was not a significant increase (Figure 2-2, A, compare lanes 3-4 to lanes 5-6). The increase of ^{14}C Ser with the addition deacylated tRNA is indicative of the ability of wheat SerRS to be able to aminoacylate mammalian tRNA. Due to the relatively close size between $tRNA^{[Ser]}$ and $tRNA^{[Ser]Sec}$, however, we are unable to tell which tRNA is being aminoacylated. Furthermore, the ^{75}Se and ^{35}S labeled WGL extracted tRNAs all seemed to be smaller than the Ser-tRNA fraction, suggesting a smaller tRNA is picking up ^{75}Se instead of $tRNA^{[Ser]Sec}$ (Figure 2-2, A, compare lanes 1 and 2).

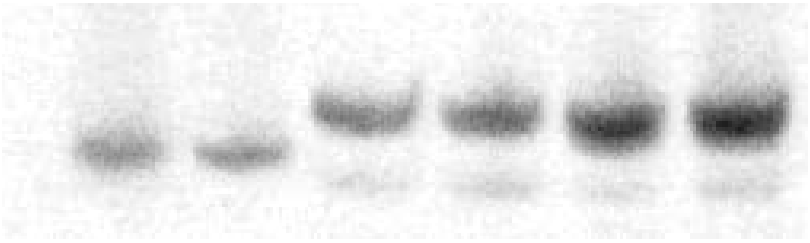
There was also the chance of ^{35}S Cys labeled tRNA being incorporated into $tRNA^{[Ser]Sec}$ as well. Sulfide is able to replace selenide in a SPS2 reaction to generate thiophosphate ($H_2SPO_3^-$) to yield $Cys\text{-}tRNA^{[Ser]Sec}$ (Xu et al., 2010). To determine if this

was the case in our WGL system, we added ^{75}Se, ^{14}C Ser, ^{35}S Cys along with pretranslated aminoacylation components in RRL and WGL. This would allow us to see if there was a size difference between the ^{75}Se tRNA fraction from each reaction on an acid urea gel. We found that the size of the ^{75}Se labeled tRNA extracted from WGL with was smaller than the ^{75}Se labeled tRNA band found in retic (Figure 2-2, B, compare lanes 2 to 3 and 4). The addition of the aminoacylation components was able to increase the intensity of the ^{14}C Ser band in retic (compare lane 6 to lane 8). However, the addition of the aminoacylation components did not seem to change the intensity of ^{75}Se Sec-tRNA$^{[Ser]Sec}$ (compare lanes 4 and 9). In this experiment we were unable to reconstruct Sec amino aminoacylation in WGL by *in vitro* synthesis of the required components.

Since aminoacylation factors pretranslated in WGL did not appear to be fully functional, we wondered if the aminoacylation components could still enhance processive Sec incorporation thought another interaction with the translational machinery or Sec incorporation components. We tested our hypothesis by adding pretranslated aminoacylation components from a WGL reaction to a WGL reaction with SELENOP mRNA and the minimal Sec incorporation components. We also wanted to add Sec-tRNA$^{[Ser]Sec}$ in an attempt see if high Sec-tRNA$^{[Ser]Sec}$ concentrations would make a significant difference in processivity. Currently, the only way to purify ^{75}Se Sec-tRNA$^{[Ser]Sec}$ in high concentrations is by purification of a large amount of radioactively labeled cells and extracting the total aminoacyl-tRNA pool. The total aminoacyl-tRNA pool is then further purified using reverse phase chromatography using a

A

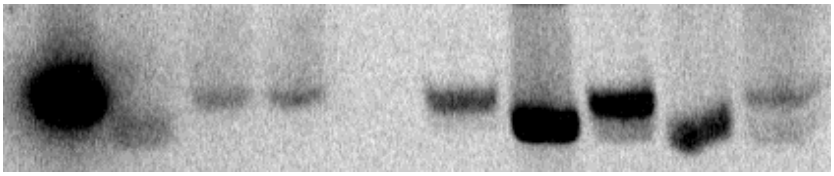
Lane	1	2	3	4	5	6
Aminoacylation Components	+	-	-	+	-	+
Rat Testis total tRNA deacylated	+	+	-	-	+	+
Radioactivity	^{75}Se	^{35}S Cys	^{14}C Ser	^{14}C Ser	^{14}C Ser	^{14}C Ser



Wheat Germ Reaction Acid Urea Gel

B

Lane	1	2	3	4		5	6	7	8	9
Lysate	None	Wheat	Retic	Retic		Retic	Retic	Retic	Retic	Retic
Aminoacylation Components		+	-	-		-	-	+	+	+
Rat Testis total tRNA deacylated		+	-	+		+	+	+	+	+
Radioactivity	^{75}Se	^{75}Se	^{75}Se	^{75}Se		^{14}C Ser	^{35}S Cys	^{14}C Ser	^{35}S Cys	^{75}Se



Acid Urea Gel

Figure 2-2. ***In vitro*** **aminoacylation testing in Wheat Germ Lysate.** **A.** 1.25ug of total HepG2 deacylated tRNA incubated in a WGL reaction with the indicated radioactivity with or without added *in vitro* translated Sec aminoacylation components. After the reactions were complete, they were pH 4.8 phenol extracted and loaded onto an 8% acid urea polyacrylamide gel. **B.** 1.25ug of total HepG2 deacylated tRNA incubated with a WGL reaction as a control and RRL with the indicated radioactivity with or without added *in vitro* translated Sec aminoacylation components. In both experiments the reactions were pH 4.8 phenol extracted and resolved by an 8% acid urea polyacrylamide gel for autoradiography analysis.

polychlorotrifluoroethylene resin support (RPC-5; reviewed in Carlson et al., 2018). For the following experiments we used RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$ for labeling due to its high specific activity and higher overall concentration without additional aminoacyl tRNAs, which was provided to us as a generous gift from the Hatfield Lab. With the addition of the aminoacylation components we saw an inhibitory effect compared to no addition (Figure 2-3, lane 1 and 4). Interestingly, the addition of RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$ showed UGA termination products further downstream of the 2nd UGA, which has not been previously reported in WGL. This result was somewhat surprising, because previous reports show additional Sec-tRNA$^{[Ser]Sec}$ into an RRL luciferase assay did not enhance UGA readthrough efficiency (Mehta et al., 2004). This suggests an increase in Sec-tRNA$^{[Ser]Sec}$ concentration could be an important requirement for processive Sec incorporation specifically. WGL would be more reliant on the addition of high amounts of aminoacylated Sec-tRNA$^{[Ser]Sec}$ than RRL, because it has no pre-existing endogenous Sec-tRNA$^{[Ser]Sec}$ nor Sec aminoacylation components to recycle deacylated tRNA$^{[Ser]Sec}$.

To see if Sec-tRNA$^{[Ser]Sec}$ concentration was important for processive Sec incorporation we added increasing amounts of ^{75}Se HepG2 tRNA or RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$ into a WGL assay for SELENOP synthesis. We also utilized a new method of HepG2 total tRNA purification, which uses ultra-centrifugation to remove the ribosomal pellet. We have found it to be an effective way to rapidly obtain a high specific activity ^{75}Se Sec-tRNA$^{[Ser]Sec}$. With the addition of this method of preparing a ^{75}Se labeled HepG2 total amino acylated tRNA we were able to detect 3 major bands

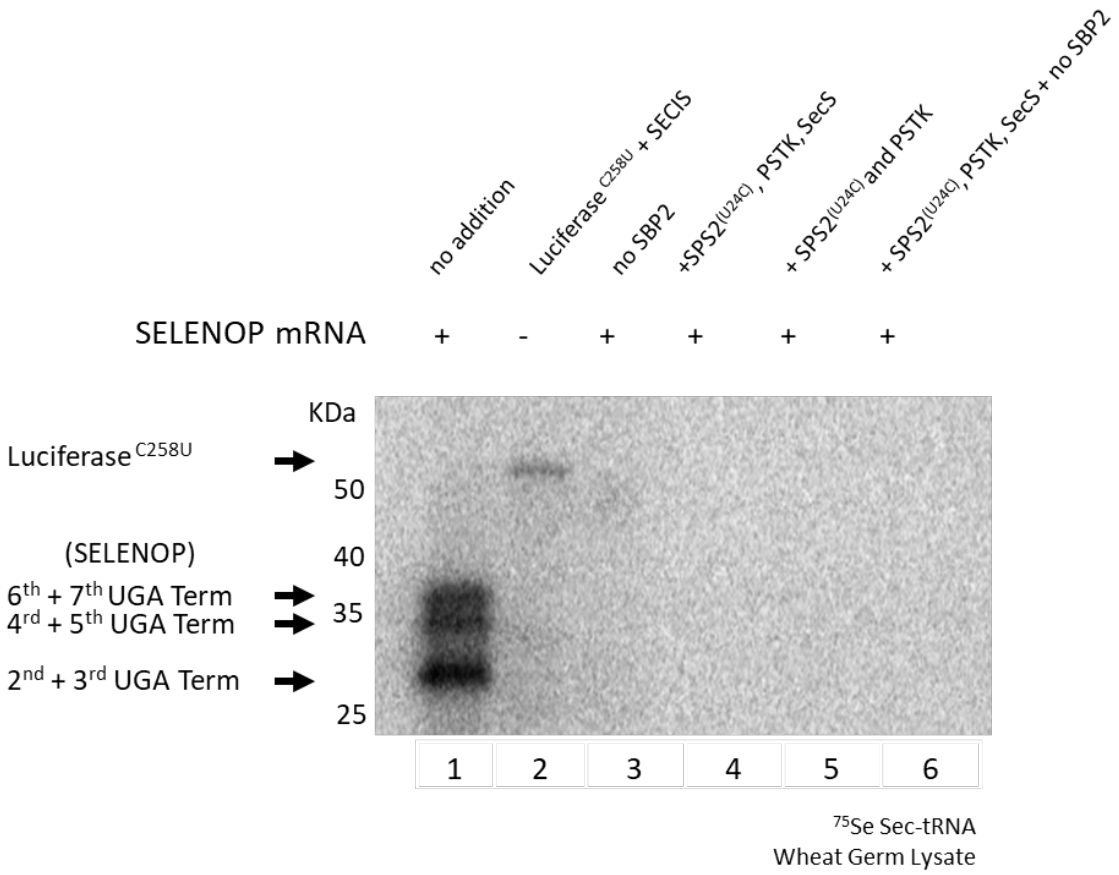


Figure 2-3. ***In vitro*** **translated aminoacylation components in Wheat Germ with ^{75}Se Sec-tRNA inhibits synthesis of SELENOP.** SELENOP or a Luciferase^{C258U} mRNA was translated in WGL with 75ng RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$ with and without the addition of Sec aminoacylation components. The lower arrow set represents SELENOP products from early termination at the specified UGA based on size. The upper arrow marks full length Luciferase ^{C258U}. Reactions were resolved on a 12% SDS page gel for autoradiography analysis.

corresponding to different SELENOP termination products similar to what we saw with RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$ (Figure 2-4, A, compare left panel to the right panel). By increasing the amount of tRNA by either the use of HepG2 total tRNA or RPC-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$ the intensity of the bands past the 2nd UGA are enhanced (Figure 2-4, A), however adding too much of either causes inhibition of overall Sec incorporation (Figure 2-4, B). It seems that Sec-tRNA$^{[Ser]Sec}$ concentration is only part of the reason WGL is unable to synthesize full length SELENOP, as we are not able to fully restore processive Sec incorporation as seen in RRL when increasing Sec-tRNA$^{[Ser]Sec}$ concentration alone.

Stimulation of Processivity in Wheat Germ Lysate by Mammalian Lysate Supplementation

We also wanted to determine if WGL was a suitable lysate to look for a processivity factor. In order to accomplish this, we tested if supplementation of a small amount of mammalian lysate could be enough to stimulate processive Sec incorporation. We tested the addition of both RRL or McArdle lysate extracts to a WGL reaction with SELENOP mRNA with its native 3' UTR, eEFSec, C-terminal SBP2, and 75ng per reaction of RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$. WGL seemed to be sensitive to any addition of RRL (Figure 2-5, lanes 3-5), which eliminated all detectable Sec incorporation products presumably due to total translation inhibition. We saw similar inhibition with the addition of McArdle extract added to WGL (Figure 2-5, lanes 6-8). It seems the addition whole mammalian lysate to WGL was inhibitory, which is likely due to the overabundance of protein from mammalian lysates.

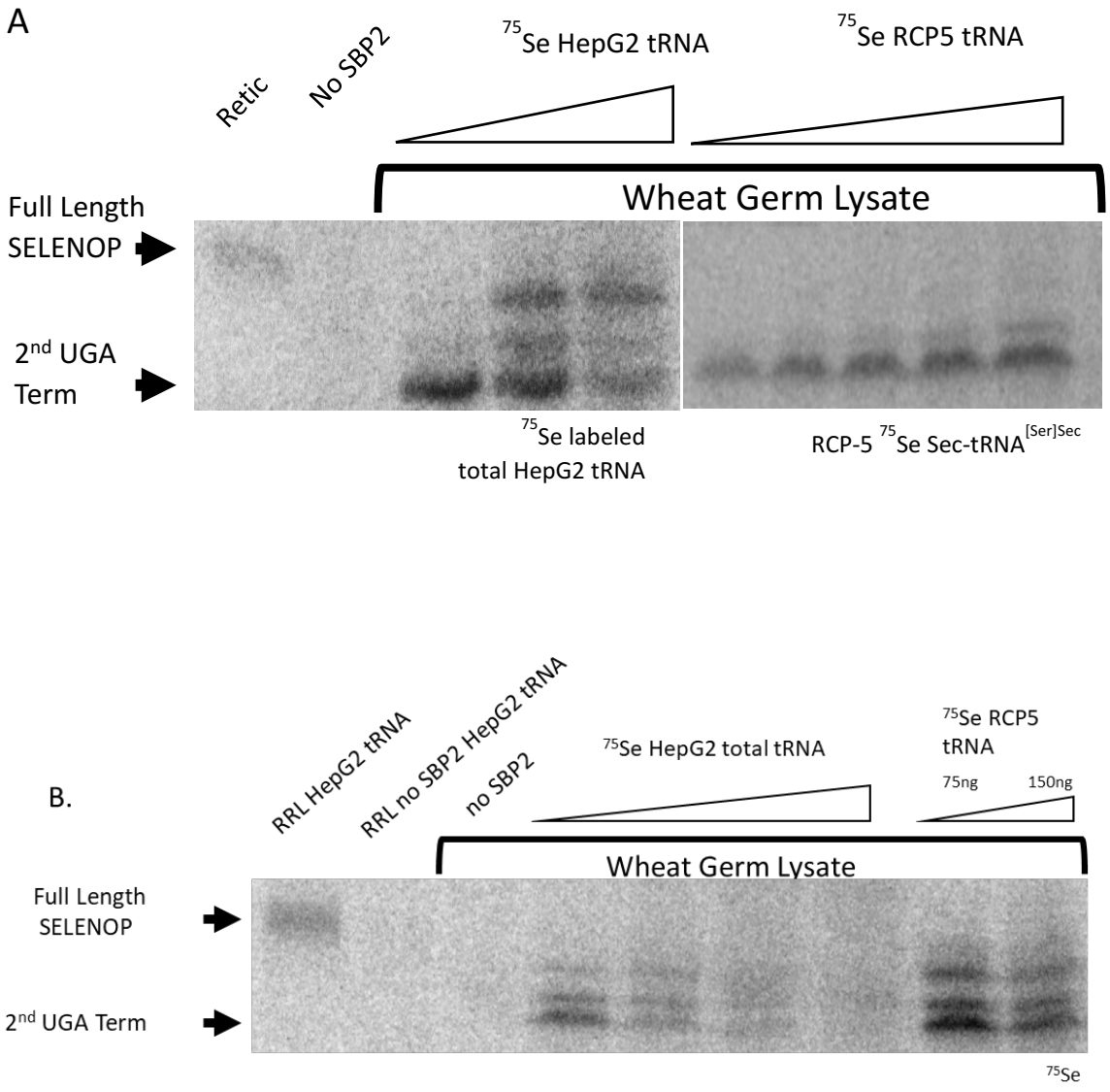


Figure 2-4. Sec-tRNA$^{[Ser]Sec}$ concentration is Important for Synthesis of SELENOP in WGL. A. SELENOP mRNA was translated in RRL as a control and with WGL with doubling amounts of ^{75}Se labeled total HepG2 tRNA in each lane (left gel, 500ng-1.5ug) or doubling amounts of RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$ (right gel, 10-160 ng) The lower arrow represents SELENOP product from termination at the second UGA, and the upper arrow marks full length SELENOP. **B.** SELENOP mRNA was translated in RRL as a control and with WGL with ^{75}Se labeled total HepG2 tRNA in each lane (1ug,2ug, 3ug, 4ug) or 75ng and 225ng of RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$. The lower arrow represents SELENOP product from termination at the second UGA, and the upper arrow marks full length SELENOP. Arrows mark full length and 2nd UGA termination products of SELENOP. All gels in this figure were analyzed by phosphorimaging.

Due to our results with whole lysates we wanted to see if the addition of mammalian ribosomes or ribosomal proteins were a requirement for processive Sec incorporation. The rationale for this approach was based on the first paper to use WGL to assay Sec incorporation, which required the use of mammalian ribosomes for efficient selenoprotein synthesis (Gupta et al., 2013). There was some discrepancy, however, because a later publication was able to see similar Sec incorporation efficiency in another batch of WGL without the addition of mammalian ribosomes (Shetty et al., 2014). While the discrepancy was never fully resolved, it still highlights a case where selenoprotein synthesis in WGL required the supplementation of mammalian ribosomes. We began to question if there was a difference between ribosomes in WGL and RRL, which could be the basis of the difference in observed processivity. SBP2 interacts with the ribosome at Helix 89 (H89) and expansion segment 31 (ES31). It could be possible on wheat ribosomes these regions are different enough that SBP2 stability is compromised (Caban and Copeland, 2012). Since eRF1 also contacts H89 of the ribosome as well, SBP2 binding to the ribosome seems like it would be a likely requirement for processive Sec incorporation by preventing eRF1 binding in the presence of a UGA (Bulygin et al., 2016). Another possibility could also be an unknown ribosomal protein unique to mammals is allowing processive Sec incorporation to occur.

Using the working hypothesis of mammalian ribosomes or a ribosomal protein being required for processive Sec incorporation, we wanted to fractionate RRL. Translation grade RRL was chosen because it already demonstrated processive Sec incorporation and micrococcal nuclease treatment would eliminate the possibility of

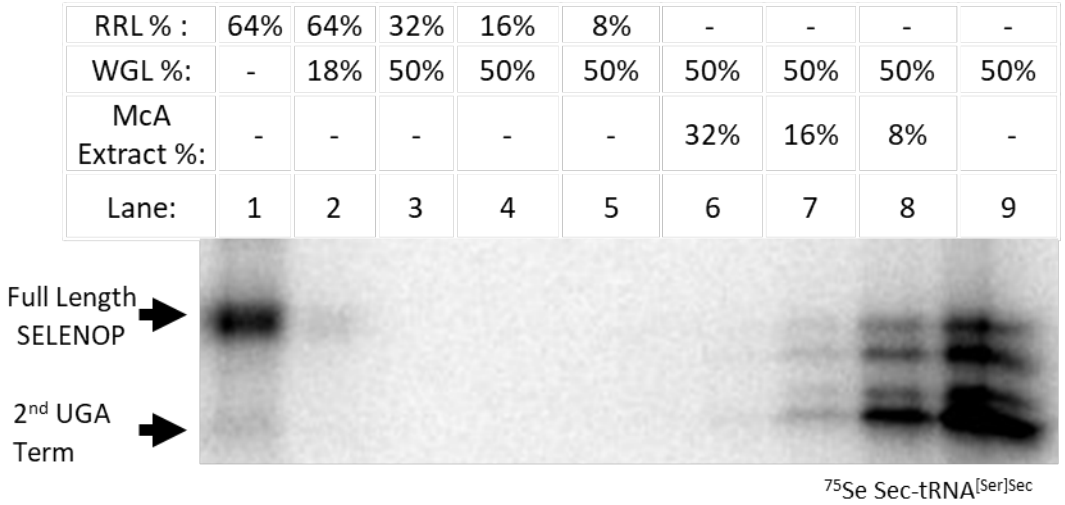


Figure 2-5. Addition of whole mammalian lysates to wheat germ lysate does not enhance processivity. SELENOP mRNA was translated in RRL or WGL supplemented with 75ng RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$ with either no addition (–) or differing amounts of mammalian lysate from RRL or McA (% of lysate is indicated). RRL was supplemented with CTSBP2, and WGL reactions were supplemented with CTSBP2 and eEFSec. The lower arrow represents SELENOP products from early termination at the 2nd UGA based on size while the upper arrow represents full length SELENOP. The reactions were resolved on a 12% SDS page gel for autoradiography analysis.

contaminating mRNAs after fractionation. We fractionated RRL by ultra-centrifugation, salt washing, and sucrose filtration to create separate ribosomes and ribosomal protein fractions. This process allows us to fractionate in a relatively gentle way and provides the highest probability of success in isolating a fraction with an intact processivity factor(s). Figure 2-6 illustrates a diagram of the process used to make each fraction used in the following experiments in this chapter. RRL was first spun down at 300,000 x g for 30min at 4°C and the pellet was saved, and the supernatant was spun again at 300,000 x g for 30min at 4°C. The supernatant from the second spin (S300) was saved and stored at -80°C for later use. The pellets from both spins (P300) were washed gently in 2mM $MgCl_2$, 0.05M Tris-HCL pH 7.5, 1mM DTT, 0.1mM EDTA then pooled together and resuspended in one fifth volume in translation buffer (80 mMHepes-KOH, pH 7.5, 100 mM KCl, 2 mM DTT, and 0.5 mM $MgCl_2$) and stored at -80°C. We also made a ribosomal salt wash (RSW) by resuspending the P300 in a high salt buffer (0.25M sucrose 0.5M KCL, 2mM $MgCl_2$, 0.05M Tris-HCL pH 7.5, 1mM DTT, 0.1mM EDTA) instead of translation buffer and spun again at 300,000 x g for 30 min at 4°C. The resulting fractions consisted of a pellet of salt washed ribosomes (SWR) and a ribosomal salt wash (RSW) from the supernatant. Both the SWR and RSW was then buffer exchanged into translation buffer on a BioRad p30 column. We also sucrose filtered the ribosomes after the ribosome salt wash. To make the sucrose filtered ribosomes the SWR fraction was layered on top of a sucrose cushion in a 2:1 ratio (1M Sucrose, 0.5M KCL, 2mM $MgCl_2$, 0.05M Tris-HCL pH 7.5, 1mM DTT, 0.1mM EDTA) and spun again at 300,000 x g. The resulting pellet was resuspended in translation buffer and buffer exchanged in

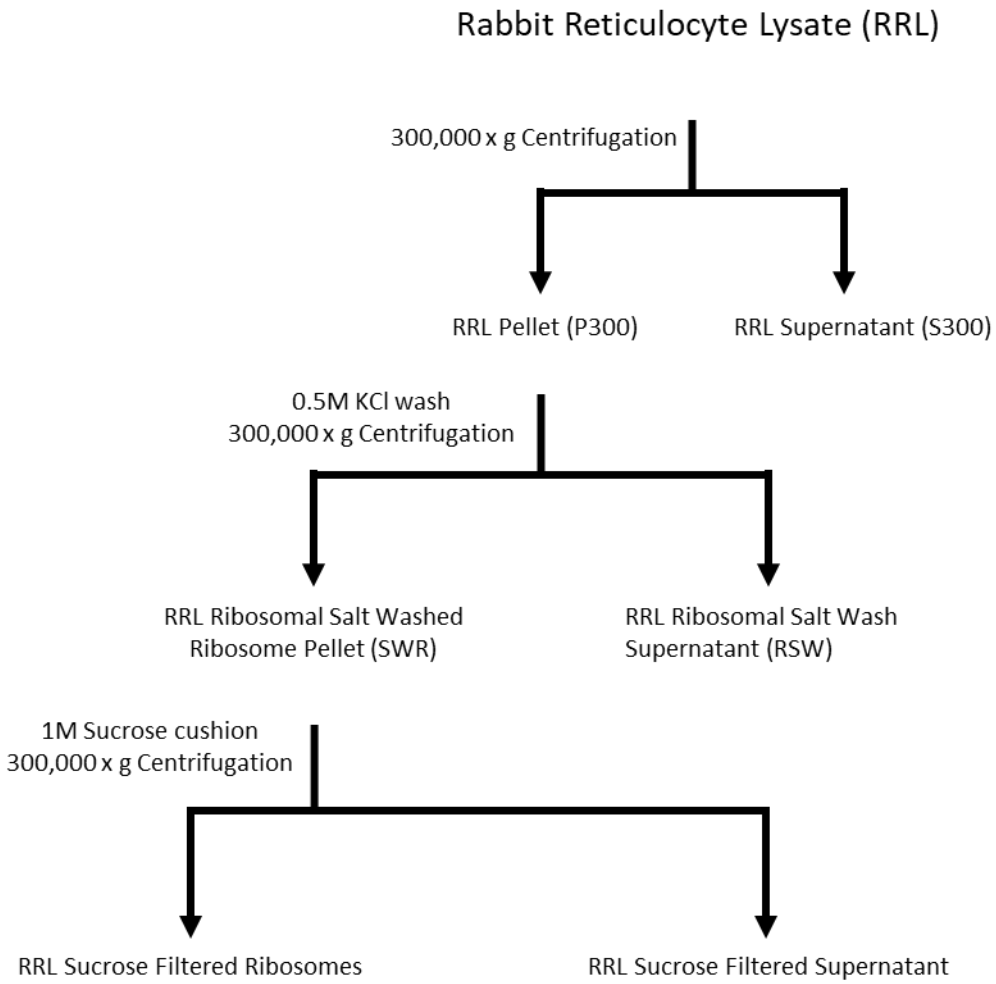


Figure 2-6. Diagram of Rabbit Reticulocyte Lysate fractionation strategy. Diagram of RRL fractionation for *in vitro* translation: RRL was centrifuged twice at 300,000xg to separate the crude pelleted ribosomes from the rest of the lysate (RRL Pellet) and resuspended in a translation buffer. Resulting crude ribosomes were washed in high salt buffer and centrifuged, or salt washed then sucrose filtered. The intention was to use these RRL fractions to add to a WGL reaction with SELENOP and the minimum Sec incorporation components to see if WGL can synthesize full length SELENOP.

translation buffer to remove excess sucrose. Fractions were quantified via BCA assay and A260 absorption, and equal amounts of each fraction were loaded on an SDS-PAGE gel for visualization (Figure 2-7). Each fraction contained numerous proteins, with some variation in the band pattern and reduced overall proteins. We observe primarily a difference in the banding patterns between the supernatants but there appeared to be few differences in the banding pattern of the ribosomal pellets under 50 kDa. To check if the fractions were still functional and the ribosomes were depleted from retic, the P300, SWR and sucrose filtered ribosomes were added back to S300 to see if they could still synthesize SELENOP. The S300 alone was unable to make SELENOP indicating that the supernatant had all the ribosomes removed (Figure 2-7, B, lane 2). When the P300 and salt washed ribosomes were added back to the S300, the ability to synthesize full length SELENOP was restored (Figure 2-7, B, lane 3 and 4 respectively). Sucrose filtered ribosomes appeared to be far less efficient in the synthesis of SELENOP (Figure 2-7, B, lane 5).

Since our fractions of retic appeared to retain full functionality for SELENOP synthesis we wanted to see if the addition of the fractions could increase processive Sec incorporation in WGL. We added the 300,000xg ribosomal pellet (P300) and salt washed ribosomes (SWR) in increasing amounts to a WGL reaction with SELENOP mRNA, eEFSec, SBP2, and RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$. Addition of the P300 increased processivity in WGL in a dose dependent manner (Figure 2-8, A, compare lane 2 with lanes 3-9). The salt washed ribosomes (SWR), however, did not increase processive Sec incorporation (Figure 2-8, B, compare lanes 2 to lanes 3-7). The ribosomes were still

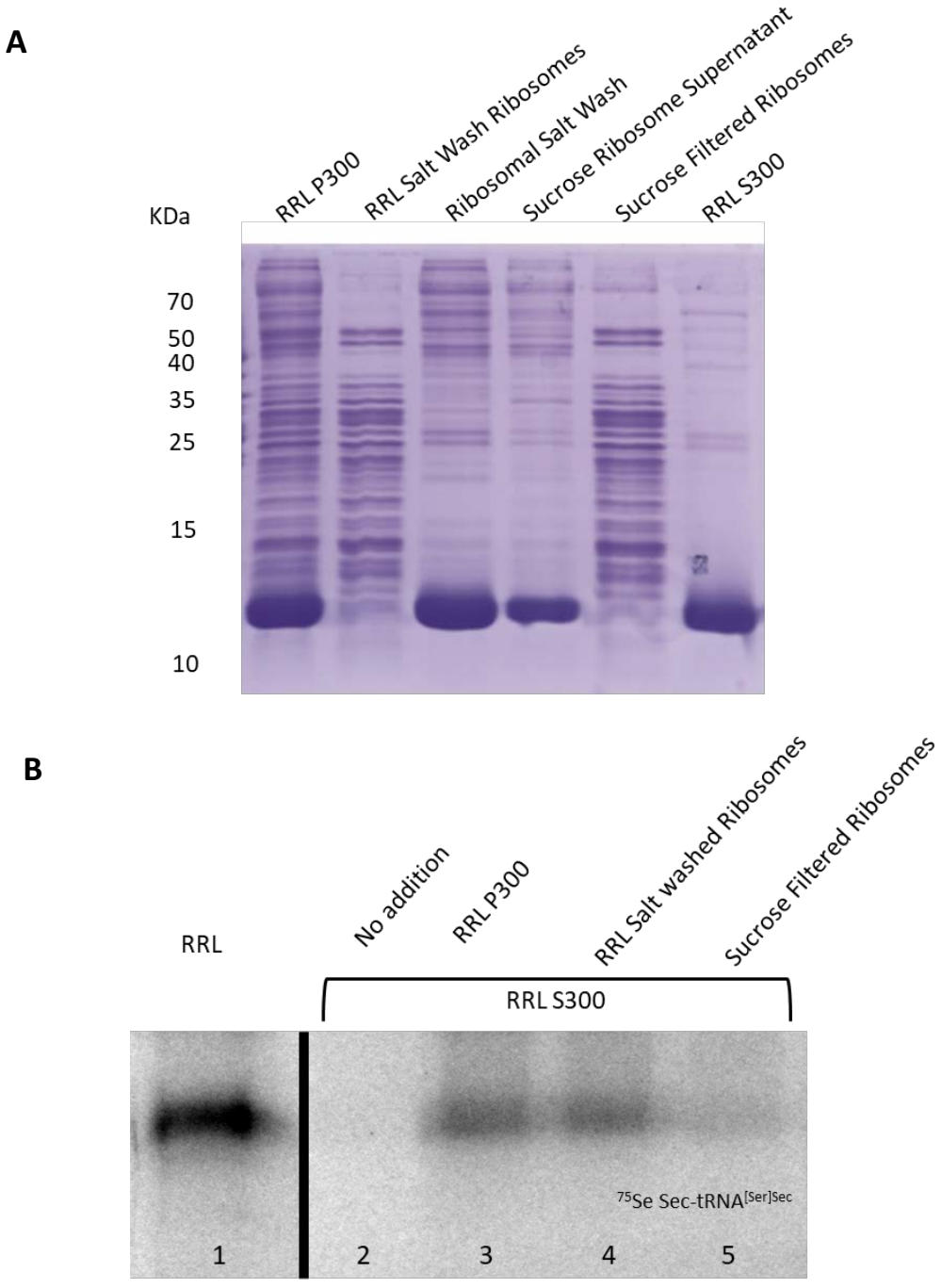


Figure 2-7. Coomassie Stain and Evaluation of Retic Ribosomal Fractions for SELENOP Synthesis. A. RRL fractions were prepared as described in Figure 2-6. 5ug of RRL fractions were loaded on a 15% SDS-PAGE gel and coomassie stained. **B.** Ribosome fractions of P300, salt wash ribosomes, and sucrose filtered ribosomes were added back to a ribosome depleted lysate with SELENOP mRNA and supplemented with SBP2 to test for translation and Sec processivity function and run on a 12% SDS-PAGE gel autoradiography analysis.

functional when added back to the ribosome depleted RRL S300 and only saw a slight reduction of full length SELENOP signal compared to whole RRL (Figure 2-8, B, compare lanes 1 to 8). This result provides evidence a factor(s) exists necessary for processive Sec incorporation and this factor is compatible with WGL. The results also indicated the factor(s) had to be a part of the ribosomal salt wash supernatant (RSW) and not the ribosomes themselves.

We then wanted to test if the addition of RSW and the S300 would stimulate processive Sec incorporation activity in WGL and added RRL RSW in increasing amounts and the S300 at a fixed amount. The RSW and S300 seemed to have the greatest effect on processive Sec incorporation resulting in an increase in termination products corresponding to UGAs 7-10 of SELENOP, which would create products between 41.2-42.6 kDa (Figure 2-9 A, compare lane 2 with lanes 6-10). In the same lanes we also see a decrease in the amount of early termination products as well. We also observe the increase in full length SELENOP repeatable with the addition of ribosomal salt washes prepared from HepG2 lysates when added to WGL. Although, the increase in processive Sec incorporation appears in a lesser extent than our original RRL RSW experiment (Figure 2-9 B, lanes 4 and 5). We also attempted to try the same fractionation technique to bulk RRL, as it would have allowed us to have enough RRL to use gel filtration to further fractionate RSW to purify the processivity factor(s). This RRL was not treated with micrococcal nuclease and has no other additions such as an energy-regenerating system, hemin, etc. When we added RSW made from bulk RRL we did not observe the same increase in processive Sec incorporation as we saw with commercial

translation grade RRL (Figure 2-9, B, lane 8). We also observe different batches of RRL do not transfer the same increase in processivity to WGL, which is apparent in our later experiments using a different batch of RRL (Figure 2-9, B, lanes 2 and 3).

We also wanted to see if we could observe processive Sec incorporation with ^{35}S Cys labeling in order to avoid using RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$. This is due to the high costs and lack of availability of ^{75}Se and polychlorotrifluoroethylene resin, along with wanting to avoid the radioactivity exposure that occurs using this method. Using ^{35}S Cys for labeling with HepG2 total tRNA as a source of Sec-tRNA$^{[Ser]Sec}$ we translated SELENOP mRNA in WGL. In anticipation of a weak signal we also included two Cys mutant SELENOPs as a size control: one with the first UGA changed to UGC (SELENOP$^{(U59C)}$) and a SELENOP mRNA where all 10 UGAs were changed to UGC (CysSELENOP). In WGL we were able to see bands past the 2nd UGA in the wild type SELENOP, but the signal was faint (Figure 2-10, lane 5). The SELENOP$^{(U59C)}$ construct did appeared to help to increase the signal after the 2nd UGA, as expected since the 1st UGA is a known bottleneck for SELENOP synthesis and is the least efficient *in vitro* (Figure 2-10, lane 8). We were unable to see an increase processivity with the addition of RRL RSW in this assay (Figure 2-10, lanes 6 and 9)(Gupta and Copeland, 2007). Using ^{35}S Cys also has a drawback of having high background bands, and we also observed a nonspecific band occurring just above the 2nd UGA termination product even in our no mRNA control (Figure 2-10, lane 3). Taken together, were felt that we were unable to use this assay for the detection of factors that increase processivity. Despite this, it was

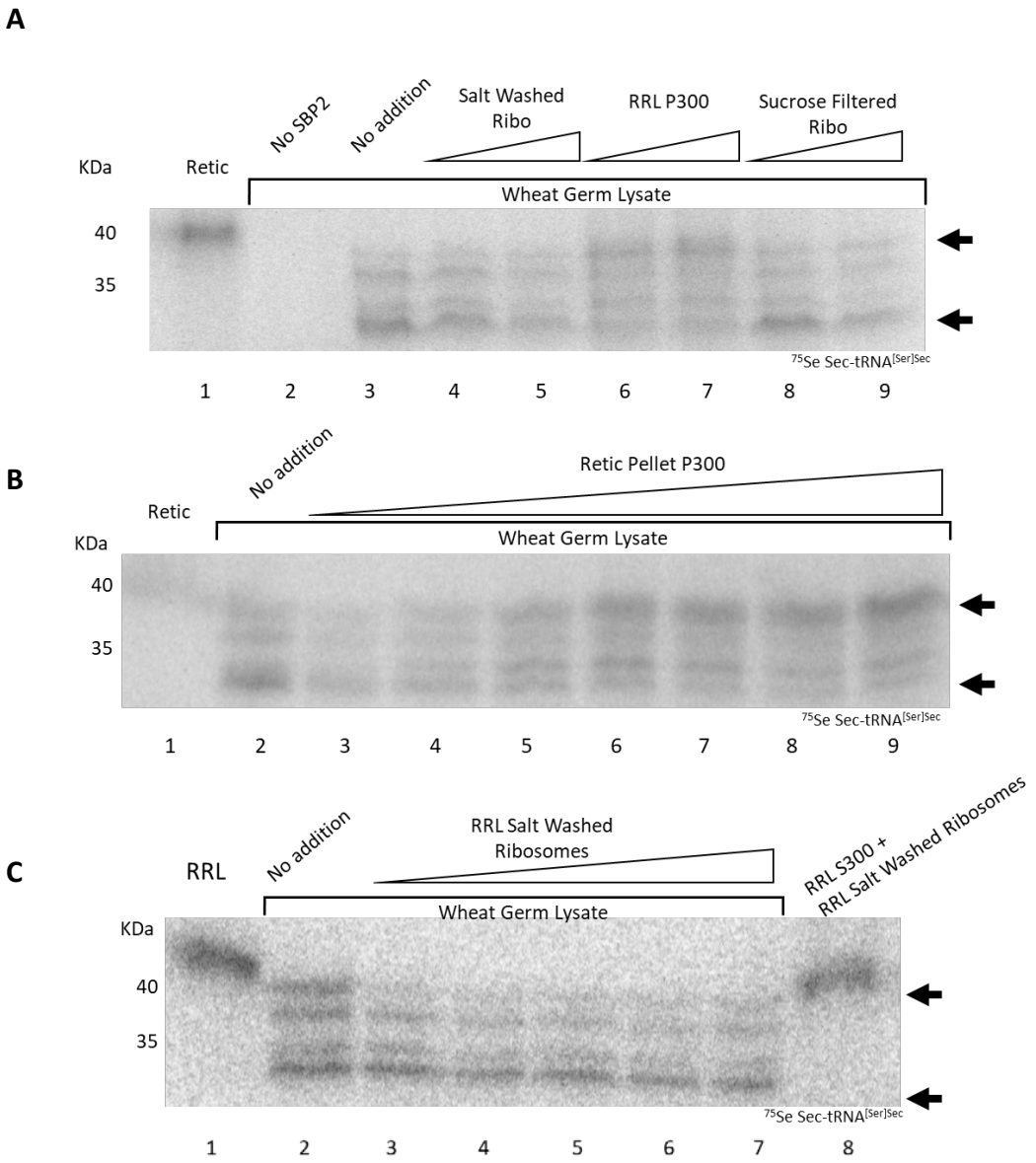


Figure 2-8. Addition of RRL 300,000 x g pellet, but not RRL salt wash ribosomes increases processivity in wheat germ lysate. A. Each WGL reaction contained SELENOP mRNA and was supplemented with recombinant CTSBP2, recombinant eEFSec, and RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$. The WGL reactions were supplemented with retic 300,000 x g pellets after buffer exchange as described in Figure 1. Pellets were obtained by centrifugation of retic lysate at 300,000 x g to separate crude ribosomes (P300, lanes 6 and 7) from the supernatant (RRL S300). The P300 was salt washed with 0.5M KCL to obtain salt wash ribosomes (SWR, lanes 4 and 5), which were further sucrose filtered in 1M sucrose (lanes 8 and 9). **(b)** A larger range of P300 was added to a standard WGL reaction with SELENOP mRNA. **(c)** A larger range of SWR (20nM to 200nM) was added to a standard WGL reaction with SELENOP mRNA (lanes 3 to 7). The SWR was also added back to a ribosome depleted retic lysate (RRL S300) with SELENOP mRNA and SBP2, and to test for function (lane 8). Upper and lower arrows representing full length SELENOP and from early termination at the 2nd UGA are marked in a, b and c.

still interesting to visualize termination products past the 2nd UGA of SELENOP using this assay.

RRL Ribosomal Salt Wash Stimulates SBP2 Independent Sec Incorporation

The next question we asked about RRL RSW, was if it could stimulate Sec incorporation independently. Since SBP2 and SBP2L are capable of ribosome binding we believed that either of the factors could have been present in the salt wash and added to WGL. Without the addition of exogenous SBP2, RRL is devoid of any Sec incorporation activity in our assays. SBP2, however, should still present in RRL in trace amounts. The ribosomal salt wash we made from RRL is a 5x concentrate of the ribosomal bound proteins, which could have increased concentration of endogenous SBP2 to make full length SELENOP visible in WGL. It is also reported that CTSBP2 *in vitro* translated in RRL has a 50-100 fold difference in UGA readthrough efficiency in a luciferase^{C258U} construct compared to a recombinant equivalent (Bubenik et al., 2014). Based on this data it is probable the amount of SBP2 needed from the RSW to enhance Sec incorporation would be significantly less than that from a recombinant source.

Recombinant C-terminal SBP2L is not known to be functional *in vitro*, but reports of SBP2 KOs have shown selenoprotein synthesis still occurs at reduced expression levels (Dubey and Copeland, 2016; Seeher, et al., 2014). This finding was repeated with our own labeling of HAP1 cell lines, which were CRISPR deleted for SBP2 and SBP2L. The cells were labeled with 100nM ^{75}Se, then lysed and ran on SDS-PAGE gel for analysis of selenoprotein expression. Our result confirms other observations that even when SBP2

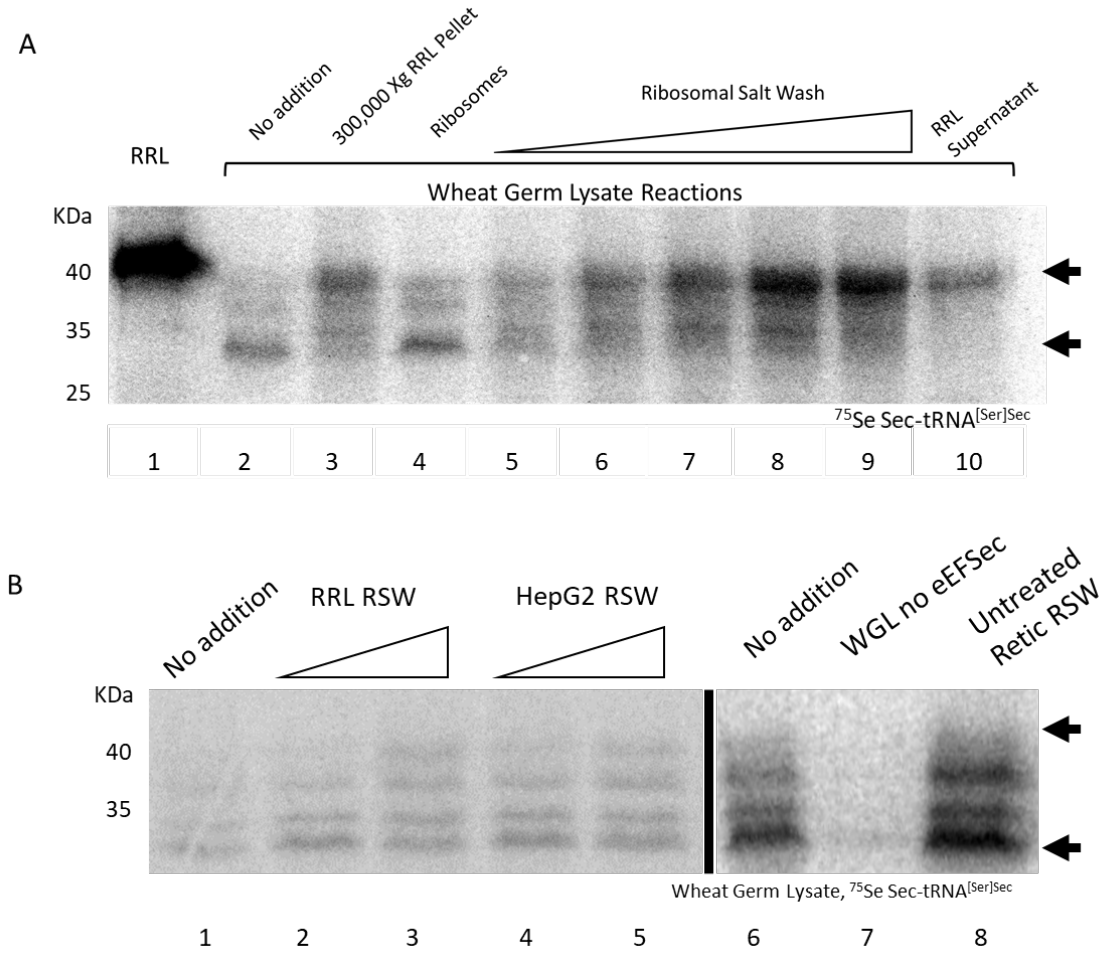


Figure 2-9. Analysis of Processive Sec Incorporation of SELENOP mRNA in WGL with the addition of RRL components. A. SELENOP mRNA was translated in RRL or WGL reactions was performed with the minimum required components for Sec incorporation. For RRL the reaction was supplemented with CTSBP2 RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$. For WGL CTSBP2, eEFSec, CTSBP2, and RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$ were added unless stated. The standard reaction (No addition) was supplemented with salt washed ribosomes (Ribosomes), increasing amounts of ribosomal salt wash (1ug to 3ug total in 0.5ug increments) or 3ul of 300,000 x g supernatant from RRL (described in Figure 2-6). **B.** SELENOP mRNA was translated in WGL supplemented with CTSBP2, eEFSec, CTSBP2, and RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$. The standard reactions (No addition) was then further supplemented with either RSW from RRL or HepG2 cells (1ug or 3ug total left gel) or 3ug of RRL not treated for translation.

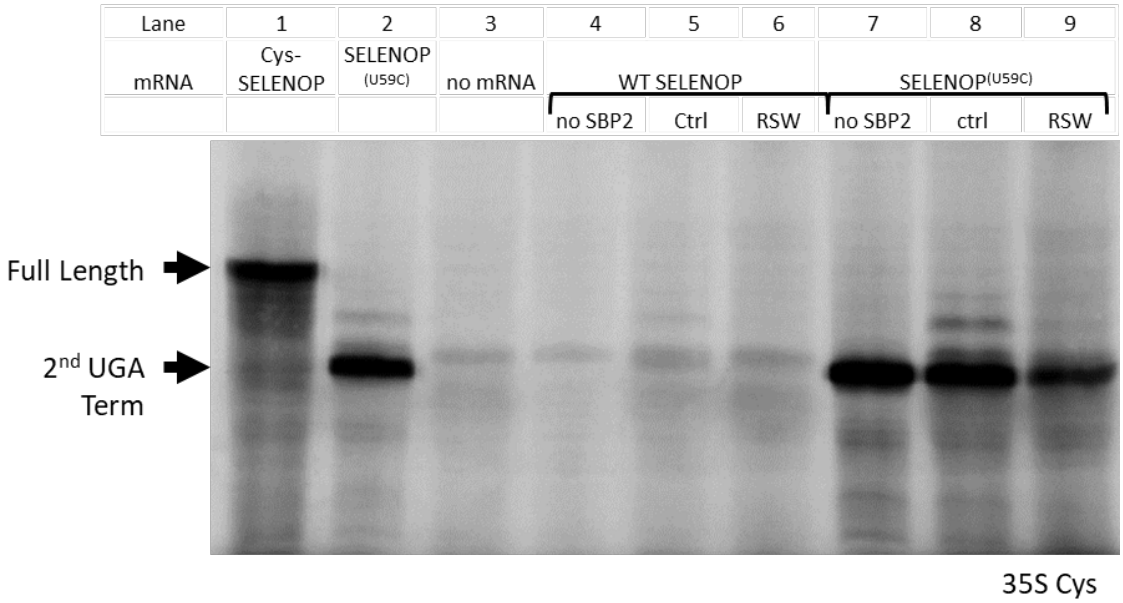


Figure 2-10. ^{35}S labeling of SELENOP synthesis in WGL. Wild type SELENOP mRNA, a mutant SELENOP$^{(U59C)}$ with the first UGA changed to UGC (Cys), and SELENOP with all UGAs changed to UGC (Cys-SELENOP) were translated in WGL using ^{35}S Cys labeling supplemented with 1.25 µg rat testis total tRNA, CTSBP2, and eEFSec. RRL RSW was also added to see if enhancement of processivity is detectable in this assay. The lower arrow represents SELENOP product from termination at the second UGA, and the upper arrow marks full length SELENOP.

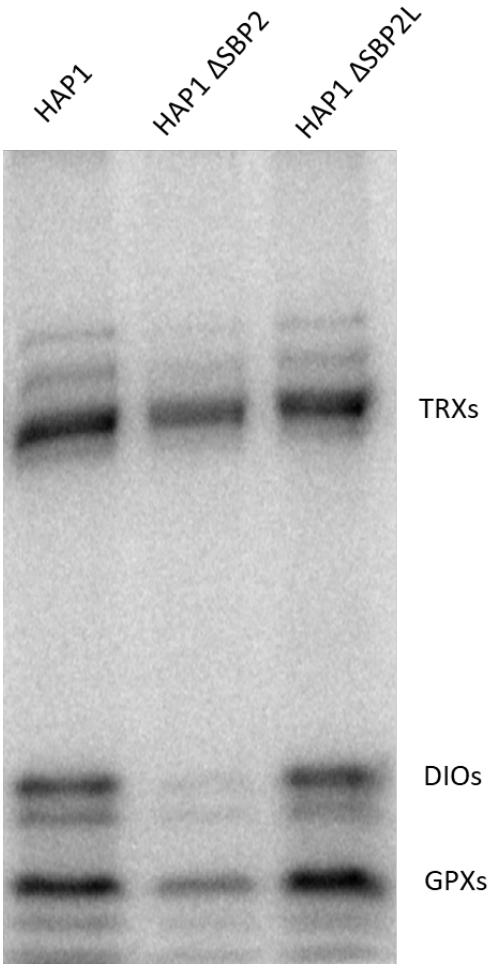


Figure 2-11. 24 hour ^{75}Se labeling of HAP1 cells and SBP2 and 2L knock outs. HAP1 cells and SBP2 and SBP2L CRISPR knock out cells were incubated for 24 hours in 100nM ^{75}Se for labeling and then lysed and resolved on a 15% SDS-PAGE gel for autoradiography analysis. A global reduction, but not ablation of selenoprotein synthesis in the SBP2 null cell line is observed. TRXND, DIO and GPX bands are marked based on approximate molecular weights.

is knocked out selenoprotein expression still occurs, but at a lower level (Figure 2-11). Yet, deletion of SBP2L in the same cells do not appear to change expression of selenoproteins. It could be possible SBP2L has a secondary role in Sec incorporation under certain conditions and becomes activated by a post translational modification or after chaperone mediated folding into the active form. SBP2L may also require the N-terminal domain for full functionality, or it may function solely for processive Sec incorporation. More support for the function of SBP2L in processivity is we observe association of SBP2L with the ribosome pellet in RRL as well and is find it in higher abundance in RRL than SBP2.

To test if the RSW is capable of Sec incorporation, we added ribosomal salt wash to RRL with SELENOP mRNA without any added SBP2 with either ^{75}Se Sec-tRNA$^{[Ser]Sec}$ (Figure 2-12, A) or ^{75}Se Sec-tRNA$^{[Ser]Sec}$ + 200nM ^{75}Se for labeling (Figure 2-12, B). While not as robust as with the addition of 2 pmol CTSBP2, there is a clear product of full length SELENOP being synthesized, suggesting that RSW contained enough SBP2 or SBP2L to incorporate Sec in our WGL assays.

Discussion

In vitro reconstitution of Sec incorporation in WGL is a relatively new technique for the evaluation of Sec incorporation components and has established itself as being a necessity for the study of the minimal requirements for Sec incorporation *in vitro.* The minimal requirements for processive Sec incorporation are still an open question in the

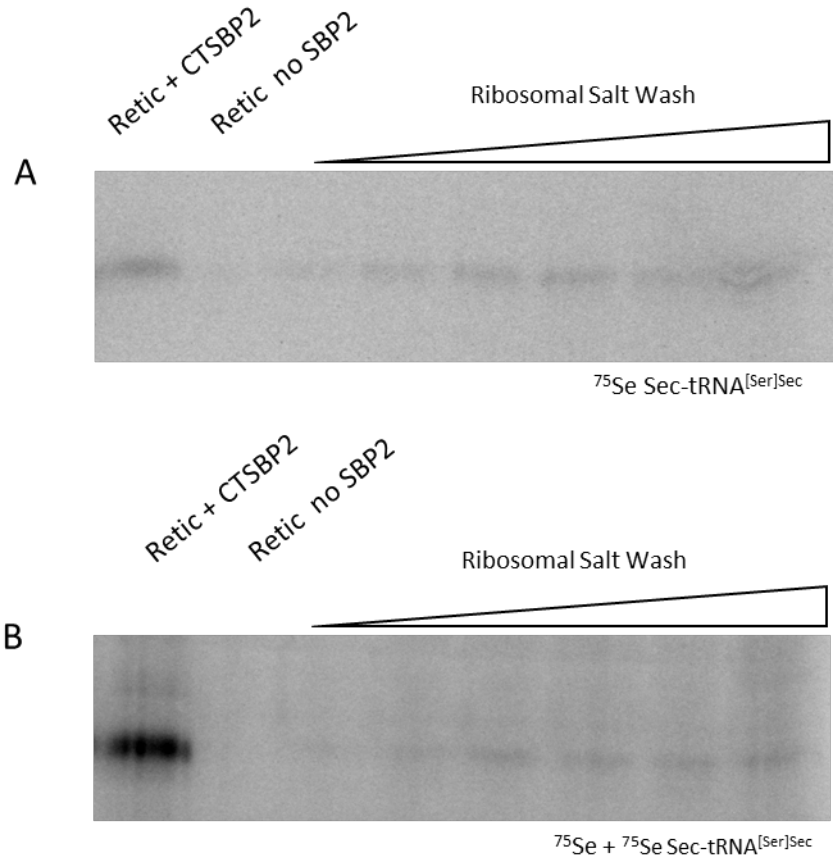


Figure 2-12. SBP2 independent processive Sec incorporation of SELENOP with RRL with the addition of ribosomal salt wash. A. Ribosomal pellet from a 300,000 x g spin was washed with 0.5M KCl and spun again to separate ribosomes from the ribosomal protein rich supernatant (ribosomal salt wash) then dialyzed in a translation buffer. SELENOP mRNA was translated in RRL supplemented with 75ng RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$ with either CTSBP2 as a positive control, no addition, or increasing amounts of RRL ribosomal salt wash (3ug to 7.5ug in 1.5ug increments). **B.** Experiment in **A** was repeated with the additional supplementation of 200 uM ^{75}Se to enhance signal.

field. In our experiments we tested the theory of channeling being a driver of processive Sec incorporation by attempting to reconstitute the Sec aminoacylation pathway *via in vitro* translation of SecS, SPS2$^{(U24C)}$, and PSTK. While we were able to synthesize the proteins, we were unable to aminoacylate tRNA$^{[Ser]Sec}$ in WGL. The addition of the *in vitro* pre-translated components also inhibited translation of SELENOP as well, which was most likely due to the sensitivity of WGL to handling. We also observed a nonspecific ^{75}Se labeled tRNA from our phenol extracted WGL experiments, which was smaller than that that from mammalian cells. We believe this to be ^{75}Se Selenomethionine-tRNA^{Met} or ^{75}Se Sec-tRNA$^{[Cys]}$ as plants are known to interchangeably use Se for S by the same sulfur assimilation and metabolism pathways (reviewed in Pilon-Smits and Quinn, 2010; Gupta and Gupta, 2017). Use of free ^{35}S in comparison to ^{75}Se aminoacylation would give us a more information on this pathway in WGL.

Instead of using *in vitro* translated Sec aminoacylation factors we attempted to use a relatively high abundance of high specific activity RPC-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$ and ^{75}Se labeled total tRNA purified from HepG2 cells to observe changes in processivity. In using a higher concentration of Sec-tRNA$^{[Ser]Sec}$ we were able to see termination products past the first UGA for the first time in WGL. We were also able to see this when the first UGA was deleted from SELENOP in ^{35}S Cys labeling as well to a lesser extent. It seemed that RPC-5 purified ^{75}Se Sec-tRNA$^{[Ser]Sec}$ was the most sensitive in detection of changes in processive Sec incorporation. We believe this is due to a combination of the high specific activity of the ^{75}Se Sec-tRNA$^{[Ser]Sec}$ and the total amount of Sec-tRNA$^{[Ser]Sec}$. Without the high purity of labeled Sec-tRNA$^{[Ser]Sec}$ the amount of

protein produced after the 2nd UGA is likely inefficient as depicted by ^{35}S Cys labeling using HepG2 purified total tRNA. This could potentially be because of the concentration of Sec-tRNA$^{[Ser]Sec}$ is too low to support processive Sec incorporation. The amount of tRNA that can be added is limited as we have shown that too much added tRNA from either source can also inhibit SELENOP synthesis. It is not certain if this is an actual requirement *in vivo*, or due to the lack of any Sec aminoacylation factors in WGL.

Our results from the addition of mammalian ribosomal fractions were able to show two important pieces of information about processive Sec incorporation: a specific factor for processive Sec incorporation exists and WGL translation machinery is compatible with processive Sec incorporation. We observed processivity increase in the ribosomal salt wash and the RRL supernatant indicates the existence of a processivity factor or set of factors that are bound transiently to the ribosome. While this could include many ribosomal bound proteins, it is more likely a factor that does not have a plant ortholog. We also were able to clearly demonstrate that functional mammalian ribosomes are not necessary for processive Sec incorporation.

Based on our results we believe that SBP2 or SBP2L is the most likely factor increasing processivity in WGL, because it has been previously reported to associate with ribosomes after centrifugation and previous *in vitro* experiments have only evaluated the C-terminal of SBP2 (Kinzy et al., 2005; Gupta et al., 2013). The SBP2 N-terminal domain has remained relatively uncharacterized except for the known nuclear localization sequence (Papp et al., 2006). Since RRL only has trace amount of SBP2 we needed to determine if there was enough SBP2 in ribosomal salt wash by adding it to

RRL without recombinant SBP2 added. The successful synthesis of SELENOP in RRL with RSW added instead of recombinant SBP2 suggests SBP2 was present in enough quantity in WGL as well. The variation we saw between batches of RRL in RSW stimulation of processivity in WGL might be due to the variation of SBP2 between batches. It is also tempting to conclude that the missing function for the N-terminal domain of SBP2 is for processive Sec incorporation based on the result of full-length synthesis of SELENOP based on all of our results. This has corroborated *in vivo* by N-terminal genetic mutations in patients with congenital hypothyroidism (Schoenmakers et al., 2010; Discussed in Chapter 4). However, SBP2L cannot fully be ruled out because SBP2 conditional KOs in mice and CRISPR knockouts in our own HAP1 cells can still incorporate Sec (Seeher et al., 2014).

If we are under the assumption that SBP2L is not functional in this assay based on previous experiments (Donovan, et al., 2009), this means that the ribosomal salt wash has a high enough concentration of SBP2 to stimulate Sec incorporation. Our next steps would need to determine if SBP2 and SBP2L are detectable in whole RRL and subsequent ribosomal fractionation. Our inability to yield the same result from untreated RRL could be due to degradation of protein in this lysate or contamination of RNAs both of which could inhibit a WGL reaction. Previous experiments support this hypothesis as others have shown that the full length SBP2 is an unstable protein due to its intrinsic disorder (Oliéric et al., 2009). Overall, the results in this chapter indicate the existence of a processivity factor or set of factors that are bound transiently to the ribosome, and WGL requires a high level of Sec-tRNA$^{[Ser]Sec}$. Once the essential factors

have been determined for processive Sec incorporation the mechanism will begin to be elucidated.

Chapter 3: Relation of the N-terminus of SECIS Binding Proteins with C-terminal Selenocysteine Rich Selenoprotein P in Metazoans

Introduction

The evolutionary history of selenocysteine is a unique field of study due to variation in its utilization across the three domains of life. The lineages of many selenoproteins have been thoroughly researched (reviewed in Gladyshev, 2016). Multiple species representing model organisms in evolution have had their selenoproteome bioinformatically identified from prokaryotes to eukaryotes. The two key *cis* factors for selenoprotein mRNAs are also used in their identification bioinformatically; an in frame UGA and a hairpin loop in the 3' UTR called the SECIS element. The in frame UGA is not easily discerned from a stop codon, but the SECIS element has unique properties that bioinformatics can leverage to discover selenoproteins. The SECIS element contains a kink turn motif, an AUGA core, a AAA motif of the apical loop, and makes non-canonical A-G base pairs. While the sequences of the SECIS element of prokaryotes and eukaryotes share a general harpin structure they vary significantly enough that they are not interchangeable in bioinformatic searches, in other words conserved eukaryotic SECIS elements could not be used to look for prokaryotic SECIS elements.

The existence of the $tRNA^{[Ser]Sec}$ gene is generally the first marker of a selenoprotein producing organism, while the detection and annotation of specific selenoproteins in an organism is more complicated. The initial strategy for discovering

theoretical selenoproteins in mammals was based on discovering SECIS elements, typically from cDNA libraries at first. After finding a SECIS element the next step would be to determine the upstream coding region and presence of an in frame UGA. The selenoprotein would then be verified via tissue culture labeling with ^{75}Se (Kryukov et al., 1999, Lescure et al., 1999). Modern bioinformatic tools have also evolved to aid in identification based on newer algorithms, sequencing techniques, and pipelines to leverage the available databases. Some of these publicly available tools specific to the field of selenoprotein bioinformatics include SECISsearch, Seblastian, SelenoDB, SelGenAmic and Selenoprofiles (Castellano et al., 2004; Mariotti et al., 2013; Romagne et al., 2014; Santesmasses et al., 2017; Jiang et al., 2012). Programs such as the ones listed previously are necessary for the field as they are tuned specifically for selenoproteins. The general genome annotation prediction programs, such as Gnomon, typically cause a UGA contained in an exon to be mis annotated as an intron without cDNA or EST data (Castellano et al., 2004). Researchers can also use these programs to detect cysteine residue homologs by examining the conserved coding sequences surrounding the UGA (Santesmasses et al., 2017). This allows researchers to make inferences regarding the gain or loss of Sec residues in specific protein families. These newer approaches to selenoprotein discovery reduced the difficulty in rapid scanning of entire genomes for the identification of selenoproteins.

SELENOP is a well-studied selenoprotein due to having two SECIS elements and multiple UGAs (two or more) in the coding region. In organisms that have been studied so far it seems that all of them produce primarily the full-length product, indicating that

in vivo the protein is synthesized efficiently and processively. SELENOP has a variable prevalence in the metazoans as well. The protein seems to be primarily found in deuterostomes, while only a few protostomes have been found so far to have a Sec rich C-terminal SELENOP (Jiang et al., 2012). SELENOP also has a related gene, Selenoprotein Pb (SELENOP2). SELENOP2 only contains the N-terminus of SELENOP with a conserved thioredoxin fold (UXXC) motif, without a Sec rich C-terminal domain. Some organisms have both SELENOP and SELENOP2 genes while others only have one of them. The SELENOP2 seems to be found in more organisms and is frequently found as the cysteine form. The origin of SELENOP appears to be related to prokaryotic thiol oxidoreductases at least regarding the N-terminus (Lobanov et al., 2008). It is believed accumulations of Sec residues were acquired *de novo* later in the evolutionary timeline by separate gene duplication events in the metazoan lineage (Mariotti et al., 2012). Aquatic invertebrate organisms such as *Strongylocentrotus purpuratus* (the purple sea urchin) are predicted to contain 20 or more in frame UGAs (Lobanov et al., 2008). Many aquatic organisms like zebrafish also tend to contain more Sec residues in the C-terminus of SELENOP (17 UGAs in zebrafish), but land animals have a reduction in the number Sec in their SELENOP (10 UGAs in humans) as well as other selenoproteins (Kryukov and Gladyshev, 2001; Lobanov et al., 2008). This phenomenon has been attributed to the reduction in environmental selenium as many of the UGA Sec codons have changed to a UGC Cys instead in mammalian evolution (Lobanov et al., 2008).

Most of the bioinformatic research in the selenium field to date has focused on selenoproteins and their evolution, with only two papers about the evolution of SECIS

binding proteins. One of the SBP2 phylogeny papers traced its evolution in insects, which revealed a reduced usage of selenoproteins and their SBP shortened to only include the SID and RBD (Chapple and Guigó, 2008). Another phylogenetic paper on SBPs compared the relationship between SBP2L and SBP2 in eukaryotes (Donovan and Copeland, 2009). SBP2 and SBP2L appear to have high conservation of the L7ae motif in their RNA binding domain (RBD) and Selenocysteine Incorporation Domain (SID; Donovan and Copeland, 2009). These two domains are required for binding to the SECIS element, ribosome, and eEFSec (Copeland et al., 2000; Fletcher et al., 2001; Caban et al., 2007; Donovan et al., 2008). The L7ae motif is a requirement for binding to SECIS elements and other kink-turn RNA motifs and the motif is found in the ribosomal proteins NHP2, RPL7ae, and RPL30 (Koonin et al., 1994; Donovan and Copeland, 2009). The motif is also be found in other proteins such as eukaryotic release factor 1 (eRF1) (Koonin et al., 1994; Copeland et al., 2000; Donovan and Copeland, 2009). A common ancestral protein of RPL30 and eRF1 is believed to be the progenitor of eukaryotic SBPs based on phylogenic relationships (Donovan and Copeland, 2009). eRF1 and SBP2 also seem to bind the Helix 89 region of the eukaryotic ribosome as well, suggesting a competition of SBP2 and eRF1 for readthrough of UGAs (Bulygin et al., 2016). The origin of another conserved motif in the RBD is upstream of the L7ae and appears to be from an ancestral gene CBF5, a pseudouridine synthase (Donovan and Copeland, 2009). Other SBP domains and motif origins remain a mystery, specifically, where the shift from archaeal ribosomal proteins to a full C-terminal SBP2 with a SID and RBD in eukaryotes and metazoans occurred. Another mystery is the origin of the N-terminus of

SBP2, as there are no well conserved domains or motifs of the n-terminus in any other proteins and is currently believed to be unique to only SECIS binding proteins.

Gene duplication events have played a key role in the development of SELENOP and SBPs. While only a small subset of invertebrates are identified to make a SELENOP with multiple UGAs to date (Jiang et al., 2012; Lobanov et al., 2008). The poor conservation of these proteins in the C-terminus suggests different gene duplications events in SELENOP resulted in an analogous Sec rich C-terminal domain for SELENOP. Both SELENOP and SELENOP2 are found in fish, but SELENOP2 is not found in other vertebrates and appears to have been eventually phased out (Kryukov and Gladyshev, 2000; Lobanov et al., 2008). For vertebrates SBP2 and SBP2L seem to have been the result of gene duplication somewhere around the emergence of Gnathostomata (jawed vertebrates) (Donovan and Copeland, 2009).

The function of the N-terminal domain of SBP2 and SBP2L is not fully understood, however prior research has suggested it as possible regulatory domain. The N-terminus is highly disordered and contains multiple predicted sites of potential post translational modification (Oliéric et al., 2009). Some previous work aligning the N-terminal domain of both SBP2 and SBP2L in vertebrates and invertebrates found two short conserved motifs, $LSAD^{15-26}$ and $PFVQ^{44-56}$ to be conserved (human SBP2 aa notation). The rest of the N-terminal domain appeared to be highly variable across the metazoans. Other motifs identified in the N-terminus include $DFPE^{216-226}$ and $QEPP^{380-405}$ in SBP2, and $QTDF^{210-252}$, $DSGY^{265-279}$, and $SEIS^{442-474}$ in SBP2L (Donovan and Copeland, 2009). Within the N-terminus of SBP2 is also a nuclear localization sequence (NLS) just

before the QEPP motif and allows SBP2 to be shuttled via the chromosome region maintenance 1 CRM1 pathway into the nucleus under high ROS conditions (Papp et al., 2006).

The N-terminal domain of SBP2, while not necessary for *in vitro* Sec incorporation, appears to be necessary for function *in vivo*. Reports of SBP2 truncation mutations in the N-terminal domain in humans results in a reduction of serum selenium and SELENOP, as well as thyroid dysfunction (Schoenmakers, et al., 2010; reviewed in Dumitrescu and Refetoff, 2011; reviewed in Schoenmakers, et al., 2016, Summarized in Table 3-1). The symptoms resulting from SBP2 partial deficiency, can range from mild to severe depending on the defect. Typically, individuals present with short stature, abnormal thyroid hormone tests, decreased selenium, and lower selenoprotein levels. The removal of large parts of the N-terminal domain has shown to reduce SELENOP, DIO2, and SELENON specifically.

SECIS binding protein 2-like (SBP2L) is a paralogue of SBP2, but currently has no known direct role in Sec incorporation *in vitro* (Donovan and Copeland, 2012). SBP2L, was first discovered by a BLAST search of the C-terminus of SBP2, which it shares 46% identity with in humans (Copeland et al., 2000). SBP2L appears to have a conserved RBD and SID, but no direct role in Sec incorporation has been attributed to SBP2L *in vitro* (Copeland et al., 2001; Donovan and Copeland, 2012). Despite having no direct role in Sec incorporation SBP2L is found to be expressed in multiple tissues based on global transcriptome data, which suggests it may function for something else *in vivo.*

Mutation	SBP2 type; Phenotype	Reference
Q79X/M515fsX563, compound heterozygous	Single functional SBP2, reinitiation at M139; intellectual disability altered thyroid hormone metabolism, hearing loss	Hamajima et al., 2012
C691R/fs65X+fs76X, compound heterozygous	Single functional SBP2; Short Stature, developmental delay, hearing loss, myopathy, hypoglycamia	Schoenmakers et al., 2010
R120X/R770X, compound heterozygous	SBP2 N-terminal truncation reinitiation at M139 and early stop codon at C-term end of SBP2; altered thyroid hormone metabolism, stunted growth, muscle weakness, bone defects	Azevedo et al., 2010
R128X, homozygous	Premature stop codon with reinitiation at M139; altered thyroid hormone metabolism	DiCosmo et al., 2009
R197X/E679D, compound heterozygous	Single functional SBP2, reinitiation at M139; altered thyroid hormone metabolism	Fu., et al. 2014
Fs295x+fs302/F223fs255X, compound heterozygous	Reinitiation at M233 and M300; Azoospermia, muscular dystrophy, Photosensitivity, abnormal immune response, insulin sensitivity, Raynaud's disease, hearing loss, stunted growth	Schoenmakers et al., 2010
K438X/fs431X, compound heterozygous	No N-terminus and missing part of SID, altered thyroid hormone metabolism and stunted growth	Dumitrescu et al., 2005

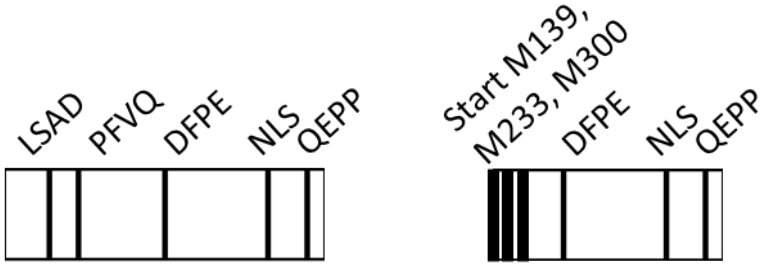


Table 3-1. Relative positions of known SPB2 N-terminal motifs and known human mutations of the N-terminus of SBP2 partial deficiency. Known instances of SBP2 mutations of the N-terminal domain, all mutations are known to cause decreased SELENOP and Se levels in serum. Fate of SBP2 mutations are listed and a diagram of conserved motifs in the N-terminus of SBP2 are depicted in a diagram of WT on the left and downstream initiation sites of mutants (Donovan and Copeland, 2009). Sources for each mutation are listed.

In certain animals such as the purple sea urchin, *Strongylocentrotus purpuratus,* along with other invertebrates are predicted to synthesize a Sec rich SELENOP with SBP2L as the sole SECIS binding protein. Furthermore, *Strongylocentrotus purpuratus* also encodes a SELENOP protein containing 27 Sec codons, which implies that SBP2L is functional and incorporates Sec in highly processive manner as well (Lobanov et al., 2008). The currently known invertebrates which make a Sec rich SELENOP only also only have a single SBP2L gene. Further *in vitro* translation testing of invertebrate SBP2L and SELENOP mRNA could shed light into the matter, too. Previous testing with C-terminal SBP2L from *Capitella* demonstrated Sec incorporation activity *in vitro*, which shows that SBP2L can be a full functioning protein in other invertebrates (Donovan and Copeland, 2012). Due to SBP2L being the only SBP found in the organisms producing SELENOP, we hypothesized that the N-terminal domain could be a requirement for processivity. With a more widespread bioinformatic analysis of SBPs from organisms who produce SELENOP, we would be able to see if there is a link between the N-terminal domain of SBPs and SELENOP. The data would also allow us infer what regions of the N-terminus of SBPs are involved in processivity by determining what motifs are highly conserved in organisms that produce SELENOP. We would also expect to be able to predict SELENOP producing animals based on the conservation of SBP N-terminal sequences.

Materials and Methods

Identification of SBP2 and SBP2L homologues and Identification of N-terminal Motifs:

The search to look for SBP2 and SBP2L homologs utilizing an initial amino acid alignment

via the **MU**ltiple **S**equence **C**omparison by **L**og-**E**xpectation (MUSCLE) algorithm in Geneious Prime of a mix of vertebrate and invertebrate SBP2 and 2L EST sequences to build a profile (Figure 3-1 and 3-2). The results from that alignment were combined with a PSIBlast using the same initial alignment to search via the most recent BLAST database v5 using BLAST+ 2.8.1. Sequences were aligned and filtered to only select for sequences with a N-terminal domain after the conserved SID and RBD. Some sequences had gaps in sequence coverage which included significant stretches of the gene, in these cases the sequences were not used unless otherwise noted. Only primary transcripts were used in this analysis. The search for SBP homologs *de novo* was conducted by Selenoprofiles v 3.5b using the same initial alignment to set up a profile (Mariotti and Guigó, 2010; Santesmasses et al., 2018). Version 3.5b of Selenoprofiles is updated with the SBP profile used in this search (https://github.com/marco-mariotti/selenoprofiles). In some cases Selenoprofiles and BLAST would provide sequences from different initiation sites, due to multiple ATGs and splice variants of SBP2. In these cases, the sequences were manually verified. Sequences already annotated and *de novo* predictions used in this chapter can be found online at: https://pastebin.com/CSnCMMcp.

Identification of SELENOP homologs: A SELENOP homolog search was conducted by BLAST 2.7.1 search of organisms which had any well conserved motif in the N-terminal domain of their SECIS binding protein. Organisms which we were unable to find a conserved SELENOP sequences for via BLAST were double checked using raw scaffold and EST data in Selenoprofiles v3.5b (Mariotti and Guigó, 2010; Santesmasses et al.,

2018). Manual verification of each gene was also conducted via SECISSearch to determine the presence of at least one SECIS element in relative proximity to the in frame UGA. Sequences were verified manually verified with EST data when possible, which allowed for easier determination of the full-length mRNA for SELENOPs.

SELENOPs from genomic sequences were frequently misannotated due to the nature of annotation programs typically determining a UGA as an intron. We also found that some organisms the full SELENOP sequence was not detected by Selenoprofiles, likely due to poor conservation of the C-terminus of SELENOP in invertebrates. In such cases we typically were able to identify SELENOP2 then obtained the downstream exonic sequence by alignment with the nearest neighboring organism with transcript or EST data. We also utilized GenomeScan and Gnomon to assist in exon prediction of genomic data by changing in frame UGAs to UGCs before prediction (Burge and Karlin, 1997; Burge and Karlin 1998; Yeh et al., 2001; Souvorov et al., 2010). TGAs were then manually verified by searching for a UxxC or CxxU motif in the N-terminal domain along with any other in frame TGAs.

Construction of Phylogenetic Trees:

Eukaryotic SBPs found by Selenoprofiles: The tree of all eukaryotic SECIS binding proteins was created using the ETE3 tool kit the prebuilt workflow: full_fast_modeltest_bootstrap and visualized using ETE3 (Huerta-Cepas et al., 2016).

Sorting of the initial SBP2 and SBP2L sequences and for the Vertebrate Sequence alignment: A Neighbor-Joining tree was created in Geneious Prime using the Jukes-

Cantor Genetic Distance Model with 500 bootstrap replicates to organize the initial alignments of vertebrate and invertebrate sequences used to search for SBPs and the initial alignment of vertebrate sequences (Saitou and Nei, 1987; Gascuel, 1997).

Phylogenetic tree of all vertebrates: A Phylogenetic tree of vertebrate SECIS binding proteins was generated using vertebrate sequences and constructed using MEGAX Evolutionary analysis by using the Maximum Likelihood method and JTT matrix-based model (Jones et al., 1992; Kumar et al., 2018). The tree with the highest log likelihood (-50636.73) was shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The rate variation model allowed for some sites to be evolutionarily invariable ([+I], >6.50% sites) representing the conserved SID and L7AE motifs. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 409 amino acid sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There was a total of 446 positions in the final dataset. The final tree was imported into Geneious Prime 2019.1.1 for visualization.

Phylogenetic tree of all chordates: SECIS binding proteins was generated using all chordate sequences and constructed using MEGAX Evolutionary analysis by using the Maximum Likelihood method and Jones Taylor Thornton (JTT) substitution matrix-based model (Jones, et al., 1992). The tree with the highest log likelihood (-50636.73) was

shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The rate variation model allowed for some sites to be evolutionarily invariable ([+I], >6.50% sites) representing the conserved SID and L7AE motifs. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 409 amino acid sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 654 positions in the final dataset. The final tree was imported into Geneious Prime 2019.1.1 for visualization.

Phylogenetic tree of all chordates and N-terminal containing invertebrates: The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. The tree with the highest log likelihood (-363822.11) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 0.25% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 687 amino acid sequences. There were a total of 3142 positions in the final dataset and the final tree was imported into Geneious Prime 2019.1.1 for visualization.

Results

Protein prediction of Eukaryotic SECIS binding proteins de novo via Selenoprofiles

Our first goal was to obtain SBP2 sequences from vertebrates and invertebrates. To do this we first wanted to make an initial SBP2 and SBP2L from invertebrate and vertebrates EST data to build a profile to search for SBP2 and SBP2L homologs using the Selenoprofiles pipeline program. We selected sequences from multiple taxa in metazoans and made an amino acid alignment using MUSCLE algorithm in Geneious Prime of SBP2 and 2L. As expected, we see that a high conservation of the SID and RBD as well as invertebrates which do not express SELENOP (such as insects) were missing large regions of the N-terminal Domain (Figure 3-1). We used Selenoprofiles with an alignment profile for SBP2 and 2L to find SBPs *de novo* that may not have been annotated. The pipeline workflow for Selenoprofiles is shown in Figure 3-2. We also used the profile for a PSI-BLAST search to find already annotated sequences. Most of the sequences we acquired were in vertebrates, arthropods, and single cell eukaryotes. We used the acquired *de novo* sequences to make a phylogenic tree using the ETE toolkit. Based on our phylogenic tree, the origin SBP2 and SBP2L appears to branch off a

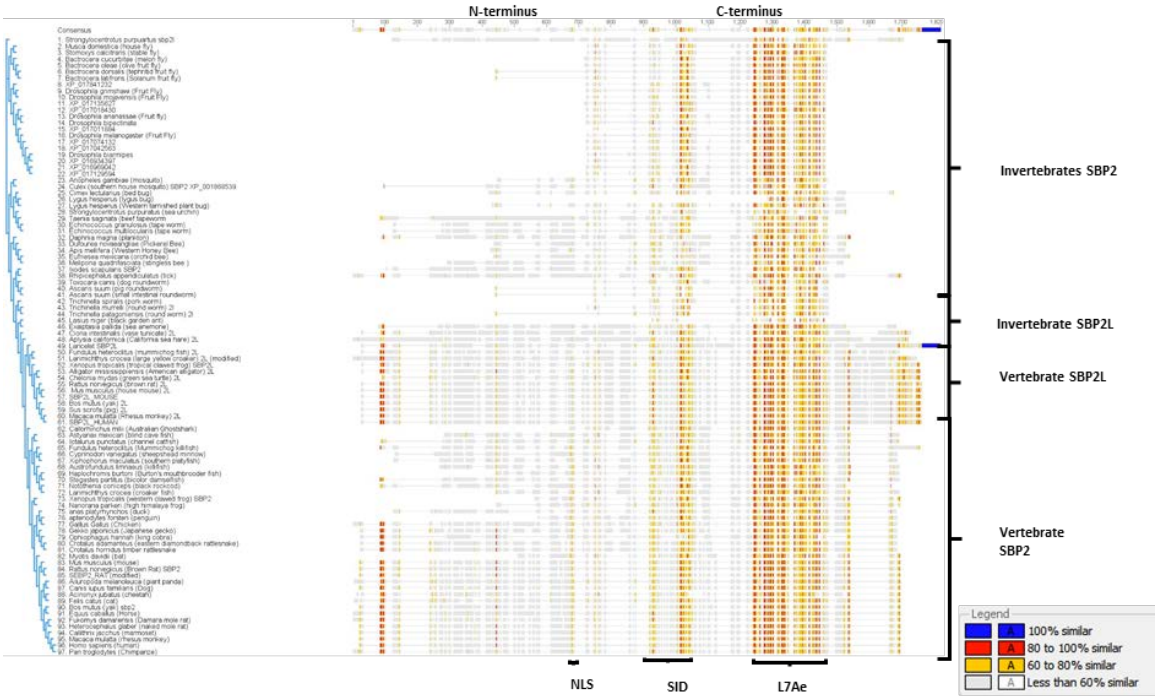


Figure 3-1. Initial Alignment of SBP2 and its paralogue SBP2L. Initial Alignment with MUSCLE algorithm and grouped by neighbor joining in Geneious Prime using annotated ESTs from invertebrate SBP2 and 2L and vertebrate SBP2 and 2L. The alignment was then grouped together by Neighbor-Joining using the Jukes-Cantor genetic distance model with 500 bootstrap replicates. Colors are representative of % similarity: Blue 100%, red 80% similarity, yellow 60% and grey <60% (lower right).

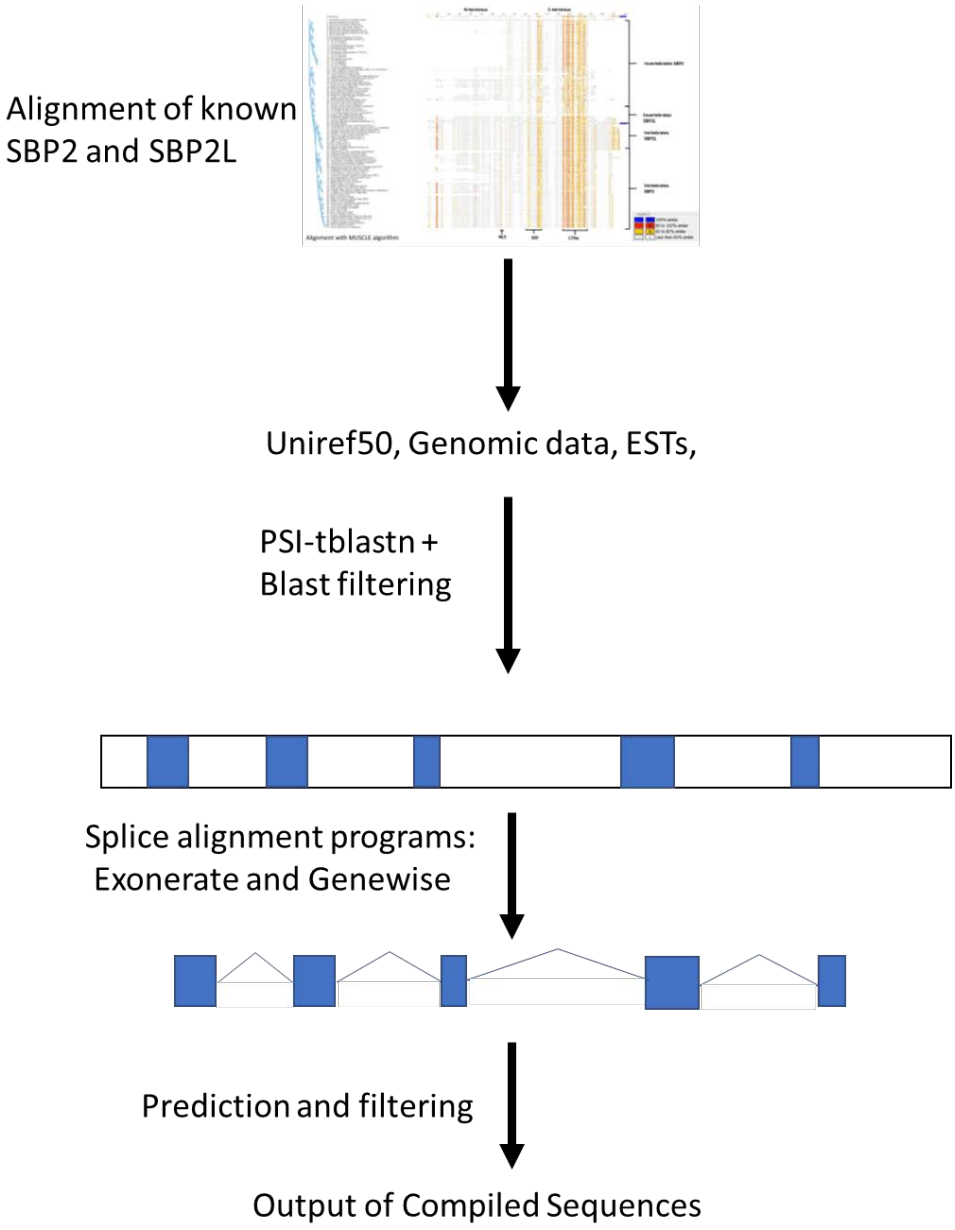


Figure 3-2. Pipeline used by Selenoprofiles 3.5b to search for SBP homologues *de novo*. This flowchart depicts the pipeline used by the Selenoprofiles program to search for homologues of genes. An initial alignment from (Figure 3-2) was used for our search of SBP2 sequences.

common ancestor early in metazoans (Figure 3-3). We also found SBP2 in several organisms of early plants and fungi, two kingdoms previously thought to not utilize selenium until recently (Mariotti et al., 2019; Figure 3-3). We were unable to find anything resembling an early eukaryotic SBP2 in eubacteria and archaea, but found SBP2 in a number of single celled eukaroyotes such as algaes, amoebazoans, and early flagellates. Our comprehensive alignment of SBP2 and 2L sequences across eukaryotes demonstrated little sequence conservation of the N-terminus, and therefore no motifs were identified in the N-terminus (Figure 3-3). The issue of poor sequence conservation is exacerbated by the variation of distance between each of the motifs as well as difference in the length of SBPs in eukaryotes.

Phylogeny and Conservation of Motifs in SECIS binding Proteins in Vertebrates

Since the N-terminal domain was so poorly conserved in our alignment and previous alignment data in invertebrates we wanted to narrow the range of species to use for SBP alignments (Donovan and Copeland 2009). We changed our strategy to look for N-terminal motifs in vertebrates first due to the higher sequence similarity among SBP2 and SBP2L and because they are known to produce SELENOP. After obtaining a consensus of the conserved N-terminal motifs in vertebrates we then wanted to use that data to search metazoans SBPs. Additionally, we chose to use vertebrates because they are the only group that has *in vivo* observations linking N-terminal SBP2 mutations

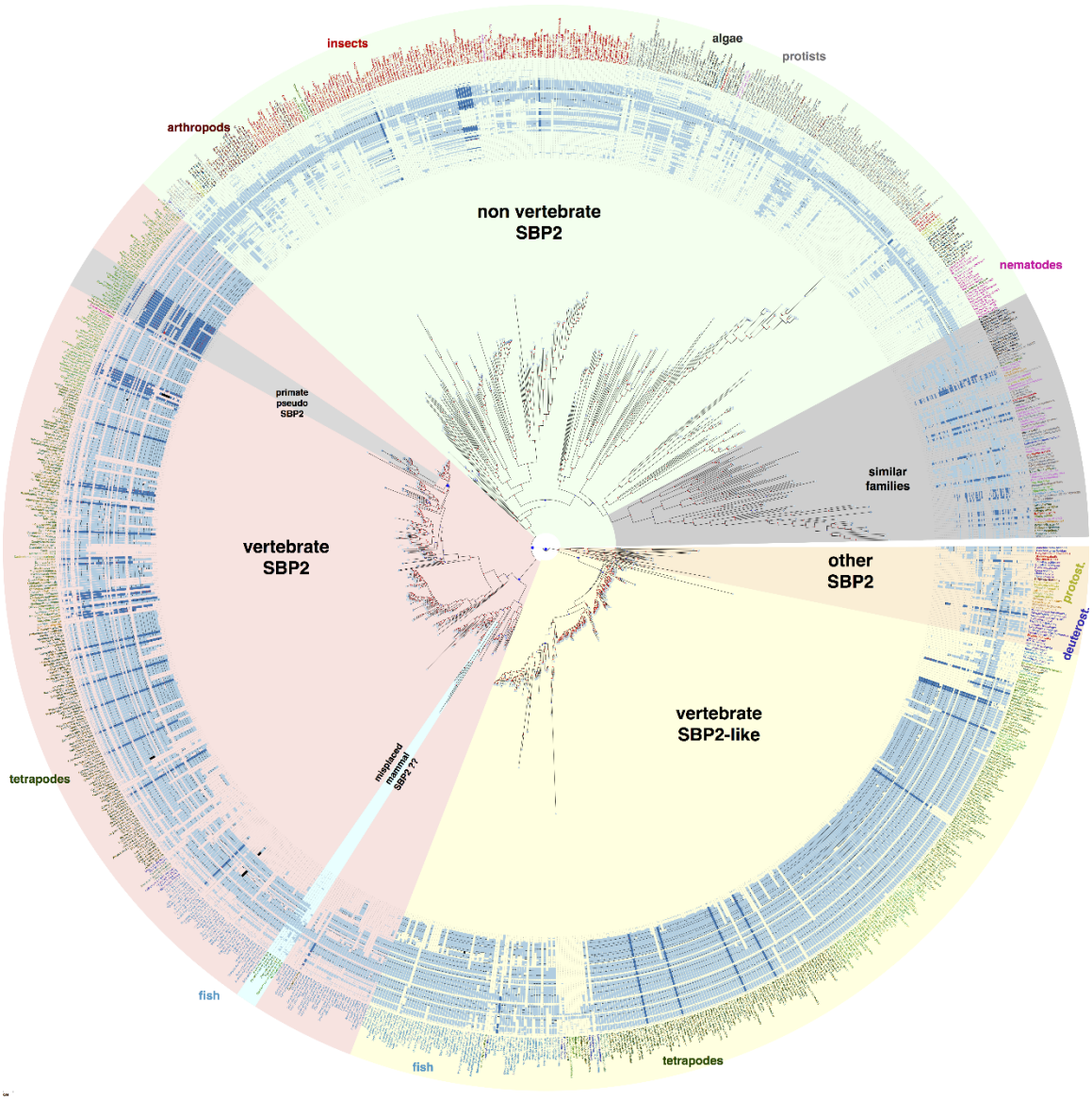


Figure 3-3. Phylogeny Tree of Eukaryotic *de novo* annotated SECIS binding proteins using Selenoprofiles. Selenoprofiles 3.5b was used to search for all SBP2 and SBP2L homologues and those sequences were used to create a phylogenic tree using the ETE3 tool kit. Major groups of animals are labeled and highlighted.

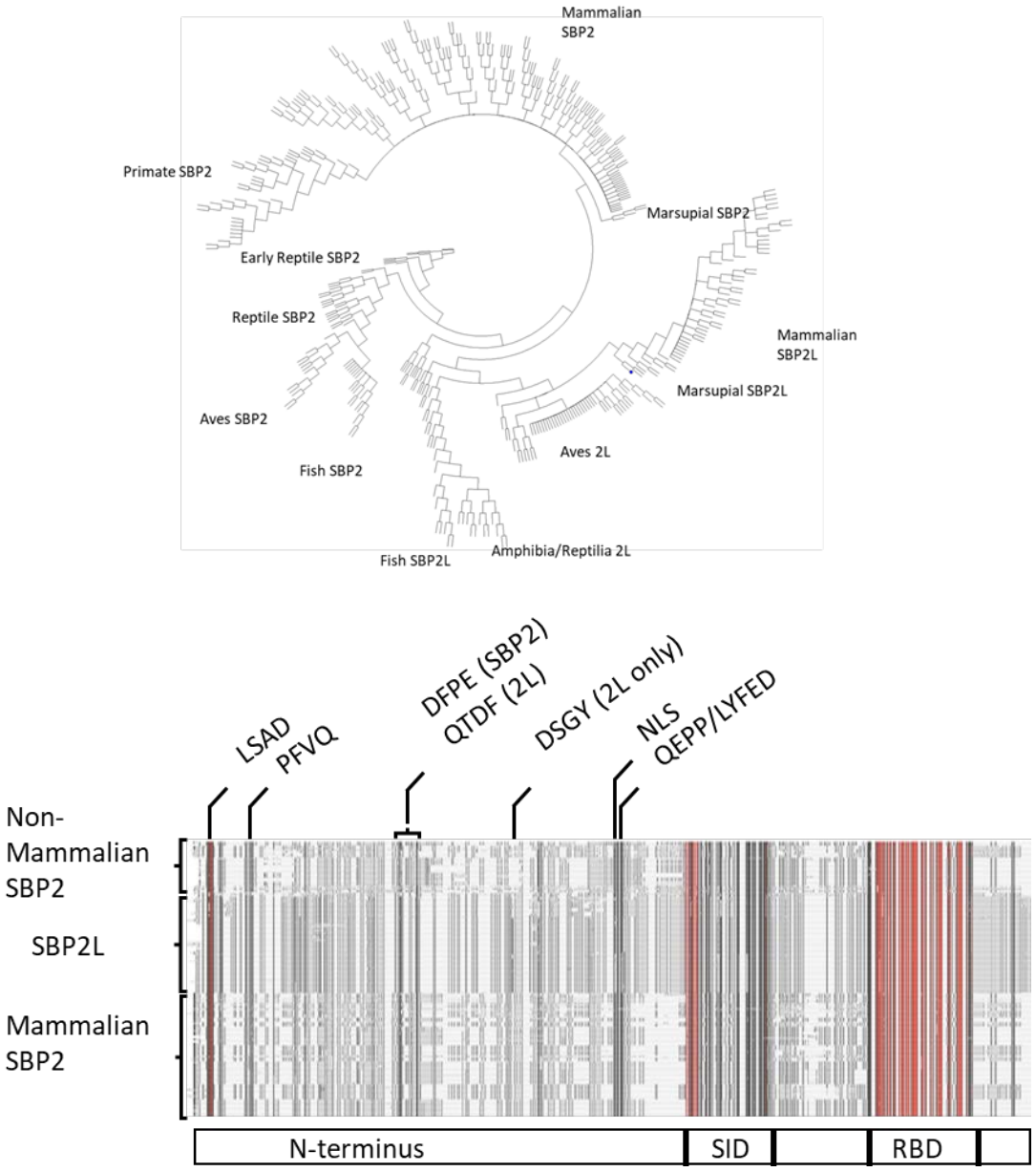


Figure 3-4. Phylogeny Tree and alignment of Vertebrate of annotated SECIS binding proteins. SECIS binding proteins from the various groups were aligned with the MUSCLE algorithm and the resulting alignment was used to construct a Neighbor Joining tree using the Jukes-Cantor model with 500 bootstrap replicates (top)**.** A compressed sequence alignment of all SECIS binding proteins found in Vertebrates (not to relative scale), and ordered based on a neighbor joining phylogenic tree (bottom). Conserved sequences are marked and the 3 major groupings of SBPs are also noted. Long stretches of non-conserved sequences of individual organisms were removed, and the ends of the alignment were trimmed for this figure. Percent similarity key is shown: red: 100%, black: >80%, and grey indicates 60% sequences conservation.

data that showed SBP2L and SBP2 was a duplication event somewhere between lancelets and early fish (Donovan and Copeland, 2009). We observed that SBP2 and SBP2L in vertebrates sorted themselves into 3 different groups of SBPs: SBP2L, non-mammalian SBP2, and mammalian SBP2 (Figure 3-4). Interestingly, we found the motifs DFPE from SBP2 and QTDF from SBP2L to align together. We also found the QEPP motif from SBP2 and a novel motif in SBP2L LYFED, align together as well (Figure 3-4, bottom). We wanted to take a closer look at the LYFED motif found in SBP2L by comparing it to the QEPP motif of SBP2 and looking at the percent similarity and the amino acid hydrophobicity as they appear to share a similar small stretch of basic amino acids towards the end of the motif (Figure 3-5). We believe the reason this was not observed previously was likely due to alignments including invertebrates and our inclusion of more sequences of both SBPs (Donovan and Copeland, 2009).

We wanted to look at the conserved sequences of the three groups of SBPs found in vertebrates separately, since SBP2L is not known to be functional in vertebrates. We aligned each SBP group separately, which was determined by our initial vertebrate alignment and phylogenetic tree from Figure 3-4. From the separate alignments we see that SBP2 in mammals have less overall sequence similarity than the other two SBP groups, but there was no change in the conserved N-terminal domains (Figure 3-6). We also found that a well conserved NLS was missing from SBP2L (Figure 3-7) (Papp et al., 2006). The N-terminal motifs in SBP2L were more conserved than SBP2, suggesting more divergences has occurred in SBP2 sequences than in SBP2L in

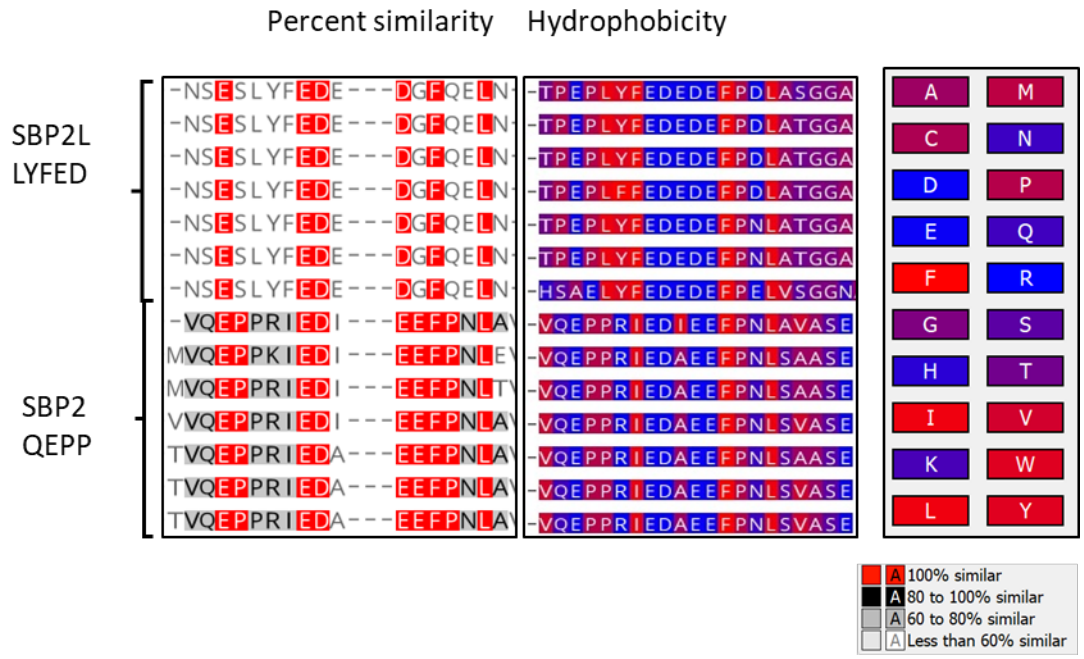


Figure 3-5. Alignment of the SBP2 N-terminal motif QEPP with SBP2L. A Representative sequence alignment extraction of SBP2 and SBP2L alignment of the SBP2 QEPP motif and novel SBP2L LYFED motif in percent similarity (left) and hydrophobicity (right). For percent similarity red indicates 100% sequence conservation, black indicates greater than 80% sequence conservation and grey indicates 60% sequences conservation. For hydrophobicity, red indicates more hydrophobic residues blue indicates more hydrophilic residues in the key on the far right.

vertebrates. Interestingly, we also saw that SBP2L appeared to have a high conservation of sequence around the SEIS motif, which is immediately upstream of the start of the SBP2L SID (Figure 3-7).

The highly conserved sequence surrounding the SEIS motif led us to wonder if the SID of SBP2L in vertebrates was further upstream than previously thought. Since the SID begins after QEPP motif in SBP2, it would make sense for the SID to start after analogous LYFED motif in SBP2L. To test this hypothesis, we looked at the predicted structure of the SID in humans using the QUARK webserver in SBP2 (amino acid 408-544) and the conventional start site of the SBP2L SID (amino acid 467-587). We found SBP2 contains five predicted helical motifs, while the conventional SBP2L $SID^{467-587}$ only contained four (Figure 3-7). Next, we predicted the structure of SBP2L SID starting from the LYFED motif (amino acid 387-587) and found a fifth helix motif corresponding to the one found in SBP2 (Figure 3-7). This led us to believe that the sequence upstream of SEIS motif is likely to be the actual start of the SID in SBP2L. The reason for the previous convention of the SBP2L SID starting at amino acid 467 was likely an artifact of pairwise alignment of SBP2 and SBP2L. Based on this data, we did not include SEIS in our search later search for N-terminal motifs in invertebrates.

We also wanted to get a better picture of when duplication and divergence of the three vertebrate SBP groups occurred, we gathered sequences from our vertebrate PSI-BLASTP search combined with *de novo* chordate sequences from our Selenoprofiles search. We obtained 334 SBP2L sequences and 324 SBP2 sequences and made an alignment and phylogenetic tree. Our phylogenic analysis confirmed SBP2 diverged

early from SBP2L of the lancelet and tunicate at the emergence of vertebrates (Figure 3-8, top). Unfortunately, we were unable to find any SBP2 sequence in the Agnatha genomes available, which would give us a more accurate idea of when the SBP2 and SBP2L split occurred. SBP2 in fish, reptiles, and birds appear more closely related to SBP2L. This is apparent in the N-terminal domains, which more closely resembles SBP2L in the SBP2 of those organisms (Figure 3-8, bottom). The loss of the NLS in SBP2L is represented in an alignment shown in Figure 3-9.

From our larger alignment were also able to identify why there was a divergence in the SBP2 in mammals by the conservation of a subset of mammalian SECIS binding protein pseudogenes. While there were some SBP2 pseudogenes detected in almost all the major class of chordates, only this subset of mammalian pseudogenes were grouped together by phylogenetics. We believe the other pseudo SBPs we found are likely to be left over from the two whole gene duplication events in vertebrates. We wanted to look at the conservation between the pseudo SBPs, SBP2 and SBP2L. In order to get a more accurate alignment we selected primate pseudo SBP2s and aligned them together with human SBP2L or SBP2 separately. Based on our alignment the pseudo SBP2 appeared be much more closely related to SBP2 than SBP2L and were well conserved even in the N-terminal domain (Figure 3-10). This provides evidence of the possibility of separate gene duplication occurring of SBP2 in the evolution of mammals. The SBP2 pseudogene is likely not translated into a functional protein as there are multiple stop sites within the L7Ae motif and SID. It is likely the second duplication of SBP2 could be

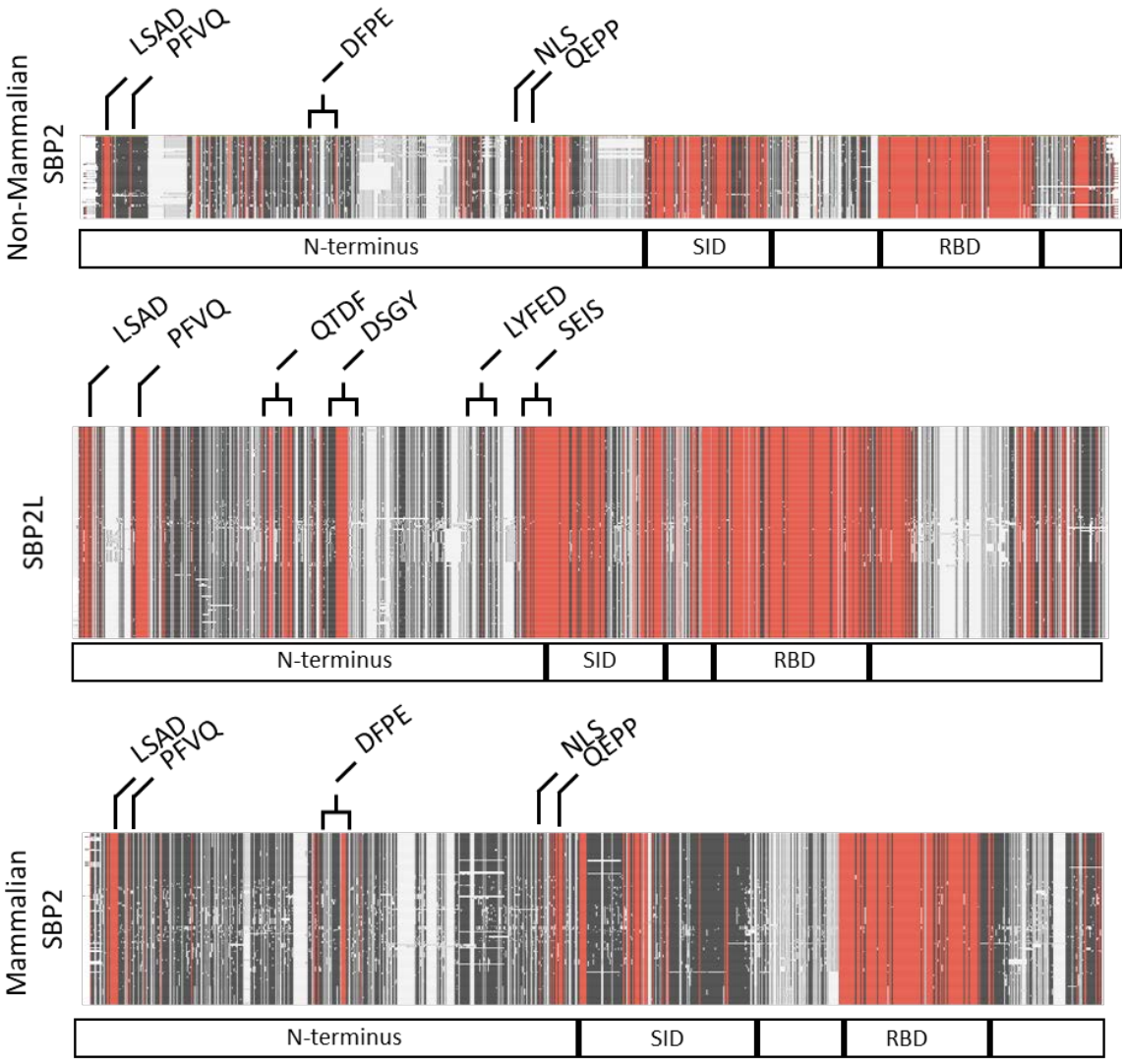


Figure 3-6. Alignment of Vertebrate SECIS binding proteins separate into three groups. A compressed sequence alignment of all SECIS binding proteins found in Vertebrates (not to relative scale), and ordered based on a neighbor joining phylogenic tree (top). Long stretches of non-conserved sequences of individual organisms were removed, and the ends of the alignment were trimmed for this figure. The sequences were then realigned based on their phylogenetic grouping from Figure 3-5, which separated into non-mammalian SBP2, SBP2L, and mammalian SBP2. Red indicated 100% sequence conservation, black indicates greater than 80% sequence conservation and grey indicates 60% sequences conservation.

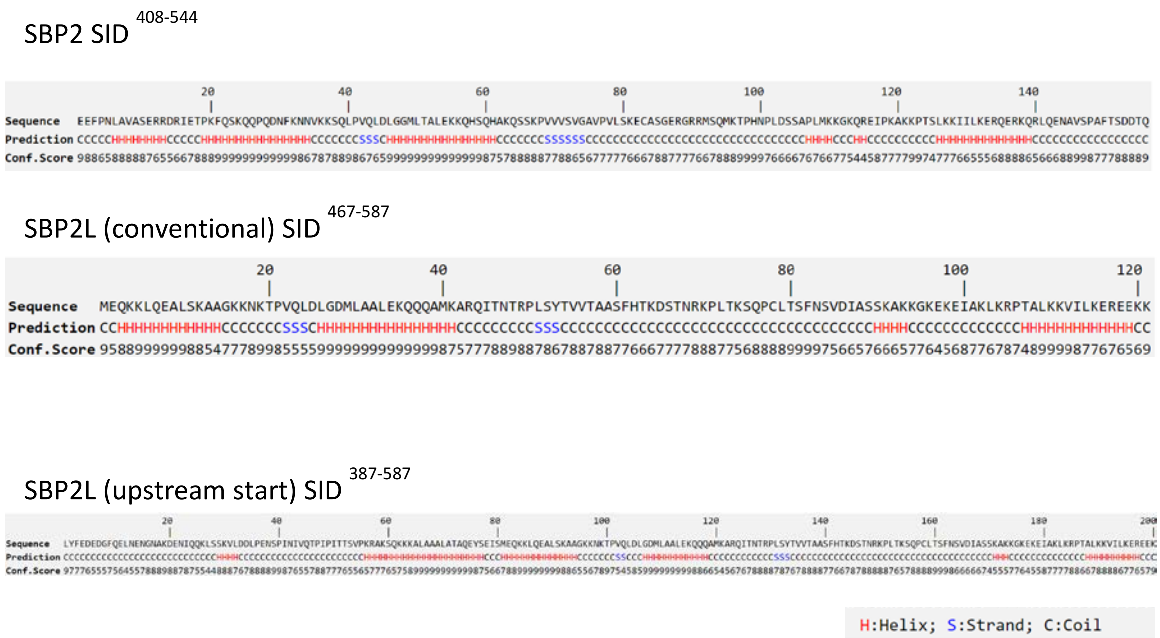


Figure 3-7. Predicted Structure of the SID of SBP2 and SBP2L from the Conventional SID and from the LYFED motif in Humans. The SID of SBP2 and SBP2L from either amino acid 467 or 387 were uploaded to the QUARK protein prediction webserver. To compare any potential structural differences between SBP2 and SBP2 at the conventional site of the SID and a possible upstream site based on alignment conservation of that area.

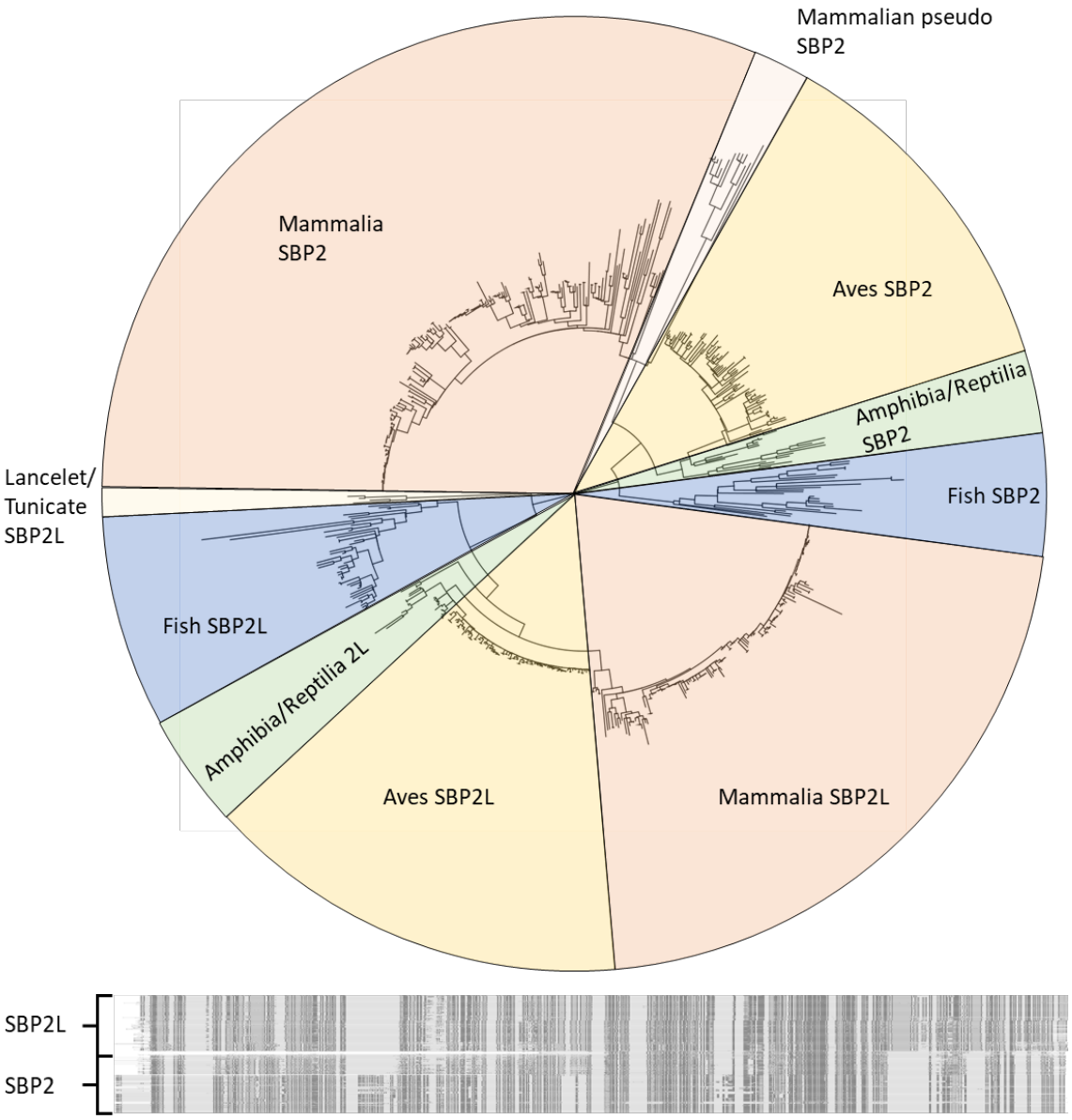


Figure 3-8. Phylogeny Tree of Chordate SECIS binding proteins. Annotated SECIS binding proteins from the Blast database of various groups were aligned with the MUSCLE algorithm and the resulting alignment was used to construct an inference tree of evolutionary history by using the Maximum Likelihood method and Jones Taylor Thornton (JTT) matrix-based model. The tree with the highest log likelihood (-50636.73) is shown of 500 bootstrap replicates. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 6.50% sites) representing the SID and RBD. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 658 amino acid sequences. All positions with less than 95% site coverage were eliminated. A compressed sequence alignment of the sequences used to make this tree is shown below.

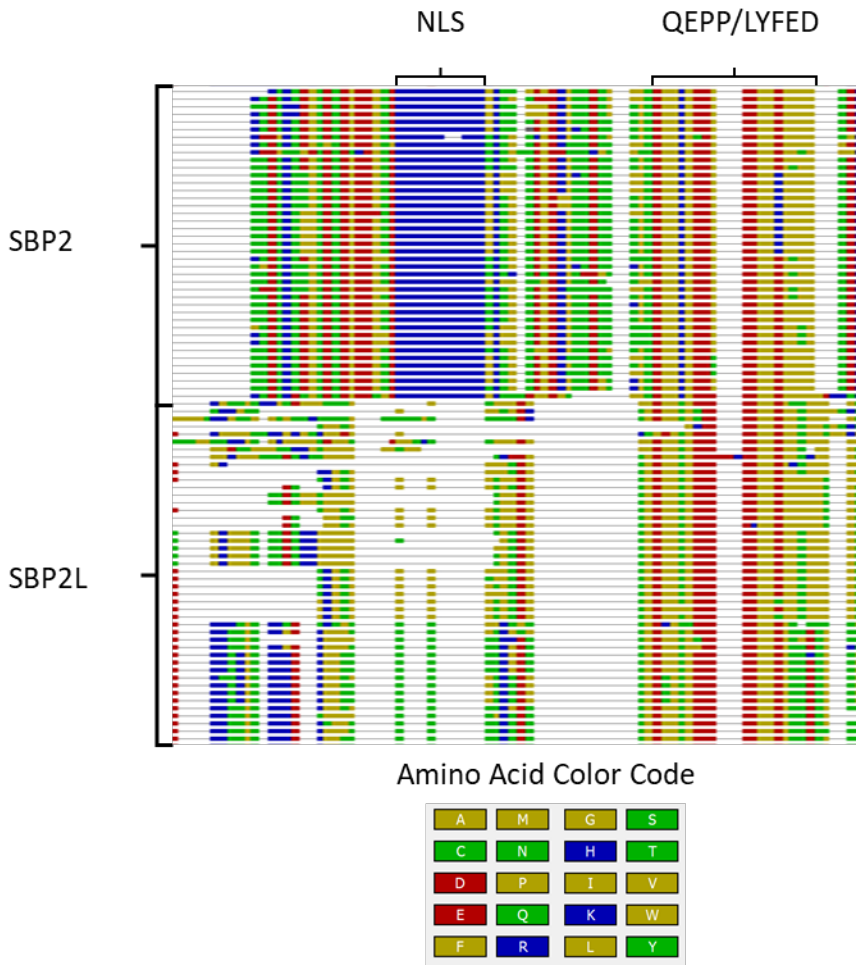


Figure 3-9. The conserved NLS is lost in vertebrate SBP2L. A. A compressed sequence alignment of vertebrate SBP2 and SBP2L N-terminus using the MUSCLE algorithm. Colors are representative of polarity, where blue represents the most basic amino acids . The NLS of SBP2 consists of a lysine and arginine rich stretch of amino acids which are colored in ^{382}KKKK^{385}.

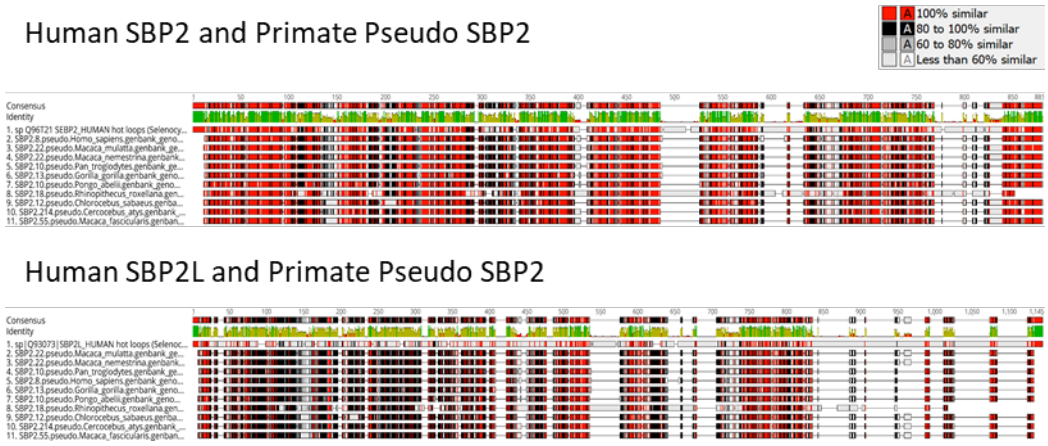


Figure 3-10. Alignment of Primate Pseudo SECIS binding proteins to human SBP2 and SBP2L. A sequence alignment using the MUSCLE algorithm in Geneious Prime of all Pseudo SECIS binding proteins found in Primates compared to human SBP2 (top image) or human SBP2L (bottom image) for comparison. Red indicated 100% sequence conservation, black indicates greater than 80% sequence conservation and grey indicates 60% sequences conservation.

the reason for the difference between SBP2 of mammals and the rest of the chordate lineage.

N-terminal conservation of SECIS binding proteins in Invertebrates and Discovery of Novel SELENOP Genes

After filtering out organisms that lack an N-terminal domain, the number of invertebrate SBP2 and 2L sequences were reduced significantly (38 in total, summarized in Table 3-2). Of those, some stand as the only representation of the entire phyla of sequenced organisms. Surprisingly, an N-terminal extension, while not necessarily well conserved, is present in invertebrates across many clades. SBP2 sequences found lacking the N-terminus primarily belonged to the Ecdysozoa clade (Arthropods and round worms) and were not included in Table 3-2 along with other phyla that lacked any N-terminal conservation (except for Porifera). In order to see where each N-terminal motif emerged during evolution we combined our N-terminal containing invertebrate sequences with our chordate sequences for an alignment and phylogenetic tree. We found that SBPs from invertebrates were split into two groups which were either more closely related to SBP2 or SBP2L (Figure 3-11, top). This is likely due to the Cnidarian and Priapulida SBPs containing shorter sequences than the rest of the invertebrate SBPs, as both invertebrate groups are branched separately from the vertebrate SBP2 and SBP2L.

Sequence conservation overall in the N-terminus of invertebrates was relatively low and we noticed some of the N-terminal motifs misalign because of this (Figure 3-11, bottom). To solve this problem, we extracted N-terminal motifs were from vertebrate

Phylum Name	Sub Phylum	Sequences of SBPs	Sequences with N-terminus	Conserved SBP N-terminal motifs	SELENOP or SELENOP2
Porifera	N/A	1	1	0	0
Placozoa		3	1	1	0
Cnidaria		7	6	6	6
Priapulida		1	1	1	0
Panarthropoda	Chelicerata	10	6	5	5
	Myriapoda	1	1	1	1
	Crustacea	5	3	1	0
Annelida	N/A	1	1	1	1
Brachiopoda		1	1	1	1
Mollusca	Bivalves	3	3	3	2
	Gastropods	4	4	3	3
	Cephalopods	1	1	1	1
Echinodermata	N/A	3	3	3	3
Hemichordata		1	1	1	1
Chordate	Tunicata	3	3	3	0
	Cephalo-chordate	2	2	2	2

Table 3-2. Summary of N-terminal domain SBP2 Sequences found in Invertebrates by Phylum. The invertebrate sequences discovered with any conserved motif of SBP2 are summarized by phylum. Phyla with no conserved N-terminal motifs were not included with the exception of Porifera.

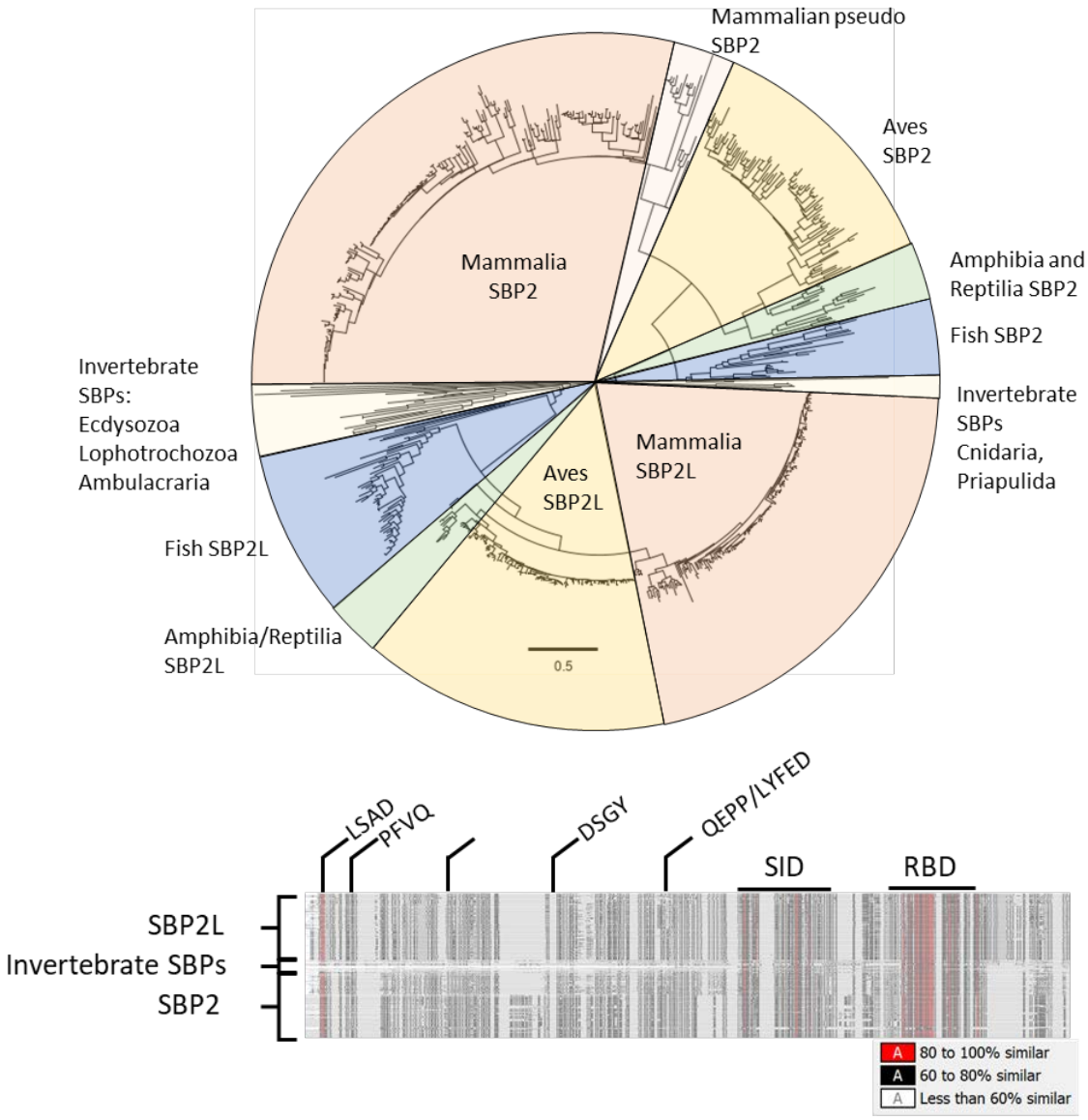


Figure 3-11. Phylogeny Tree of Chordate and Invertebrate N-terminal SBPs. Chordate and Invertebrate N-terminal SBPs were aligned with the MUSCLE algorithm and the resulting alignment was used to construct an inference tree of evolutionary he evolutionary history was inferred by using the Maximum Likelihood method and Jones Taylor Thornton (JTT) matrix-based model. The tree with the highest log likelihood (-363822.11) is shown of 500 bootstrap replicates. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 0.25% sites) representing conserved sequences of the RBD and SID. The tree is drawn to scale, with the scale bar represented at the bottom. This analysis involved 687 amino acid sequences, 1372 nodes, and a total of 3142 positions in the final dataset. Evolutionary analyses were conducted in MEGAX. A compressed sequence alignment of all the sequences used to make this tree is shown below, with the N-terminal motifs marked. The ends of the alignment were trimmed for this figure. Red indicated greater than 80% sequence conservation and black indicates 60%-80% sequences conservation.

alignments and matched by phyla or class of invertebrates. An example of how the Mollusk LSAD motif lined up to an entire SBP and alignment by individual extraction is represented in Figure 3-12. We used this method to find conserved N-terminal motifs of invertebrate SBP sequences, which is summarized in Figure 3-13. Based on our survey of the invertebrates we saw the most conservation of the LSAD motif even in Cnidarians and Placozoans, of which some of the Cnidarians had two LSAD motifs in place of the PFVQ motif (Figure 3-13). We begin to see less motifs conserved in some of the arthropods and overall the most conserved motifs throughout metazoans are LSAD, PFVQ, and the QEPP/LYFED motifs.

Correlation of the N-terminus of SBP2 and SELENOP in invertebrate Deuterostomes

While we found invertebrates with well conserved N-terminal motifs in their SBPs, most of them had no SELNOP gene published for these organisms. To test our theory of N-terminal domain conservation, we queried organisms with any of the conserved N-terminal motifs for a SELENOP gene first by a BLAST search, then by Selenoprofiles when needed. We found many of these sequences were annotated as Selenoprotein Pb (SELENOP2), which contains only the N-terminal domain of SELENOP including the conserved thioredoxin fold motif. On closer inspection it was discovered they were missing the first Sec of the thioredoxin fold due to incorrect gene prediction of exonic sequences from genomic or EST datasets or predicted termination. We then

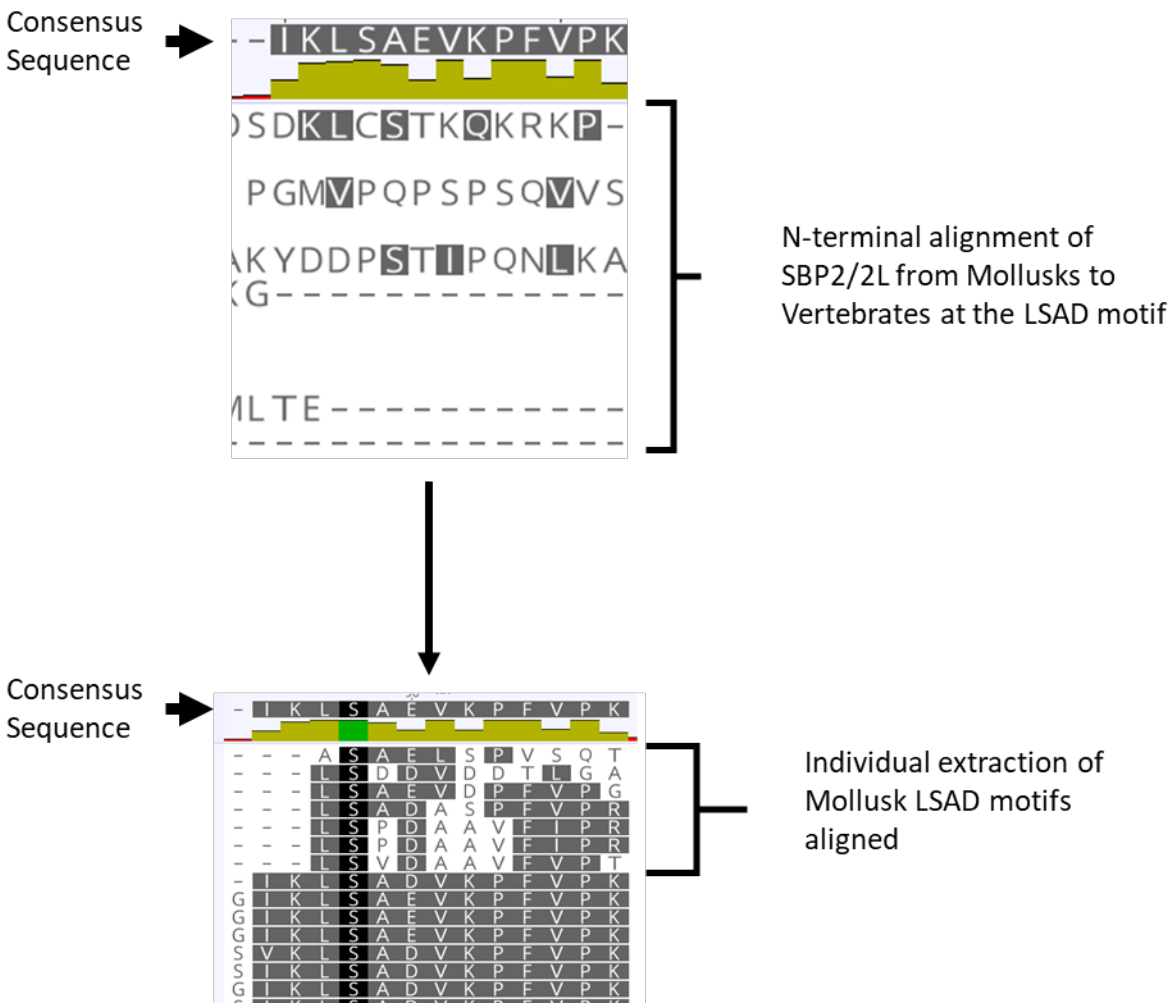


Figure 3-12. Search for individual N-terminal motifs in invertebrates by extraction from vertebrate consensus. Example of SBP2 and SBP2L sequences from invertebrates aligned to vertebrates before (above) and after (below) individual motif extraction from each sequence.

manually checked each species and related organisms in a phyla for SELENOPs and verified the presence of SECIS elements and an in frame UGA in a thioredoxin fold motif.

The N-terminus of SBPs was found to be well conserved in the invertebrate deuterostomes which included Tunicates, Lancelets, Echonoderms, and Hemichordates so we expected to be able to find Sec rich SELENOP sequences in these organisms. One of the invertebrate members of the chordates, *Branchiostoma floridae* (Florida lancelet) is already predicted to have three SELENOP2 genes and a SELENOP gene with 3N terminal CxxU motifs and 2 C-terminal in frame UGAs with 2 SECIS elements (Jiang et al., 2012). Another known SELENOP containing invertebrate deuterostome is *Strongylocentrotus purpuratus* (purple sea urchin), with 28 in frame UGAs two SECIS elements. Both SELENOP producing deuterostomes have high sequence conservation of the N-terminal motifs of SBP2, so we expected to find SELENOP in some of the related organisms. Sea urchin appears to have 2 SBPs and is predicted to have a second short C-terminal SBP2 gene just downstream of SBP2L, and we found an unreported SELENOP2 as well. In the remaining echinoderms we found SELENOP genes in *Acanthaster planci* (12 UGA, 2 SECIS elements) and *Apostichopus japonicus* (3 SELENOP2 sequences, one with 3 UGAs; Figure 3-14). In the hemichordate, *Saccoglossus kowalevskii* (acorn worm), we found a SELENOP with 5 in frame UGAs with two SECIS elements. We ended up finding no SELENOP in the tunicates despite high conservation of the SBP2 N-terminal motifs. A previous analysis of the tunicate *Ciona intestinalis* also did not find any SELENOP genes (Jaing et al., 2010). The new SELENOP sequences were only found in the

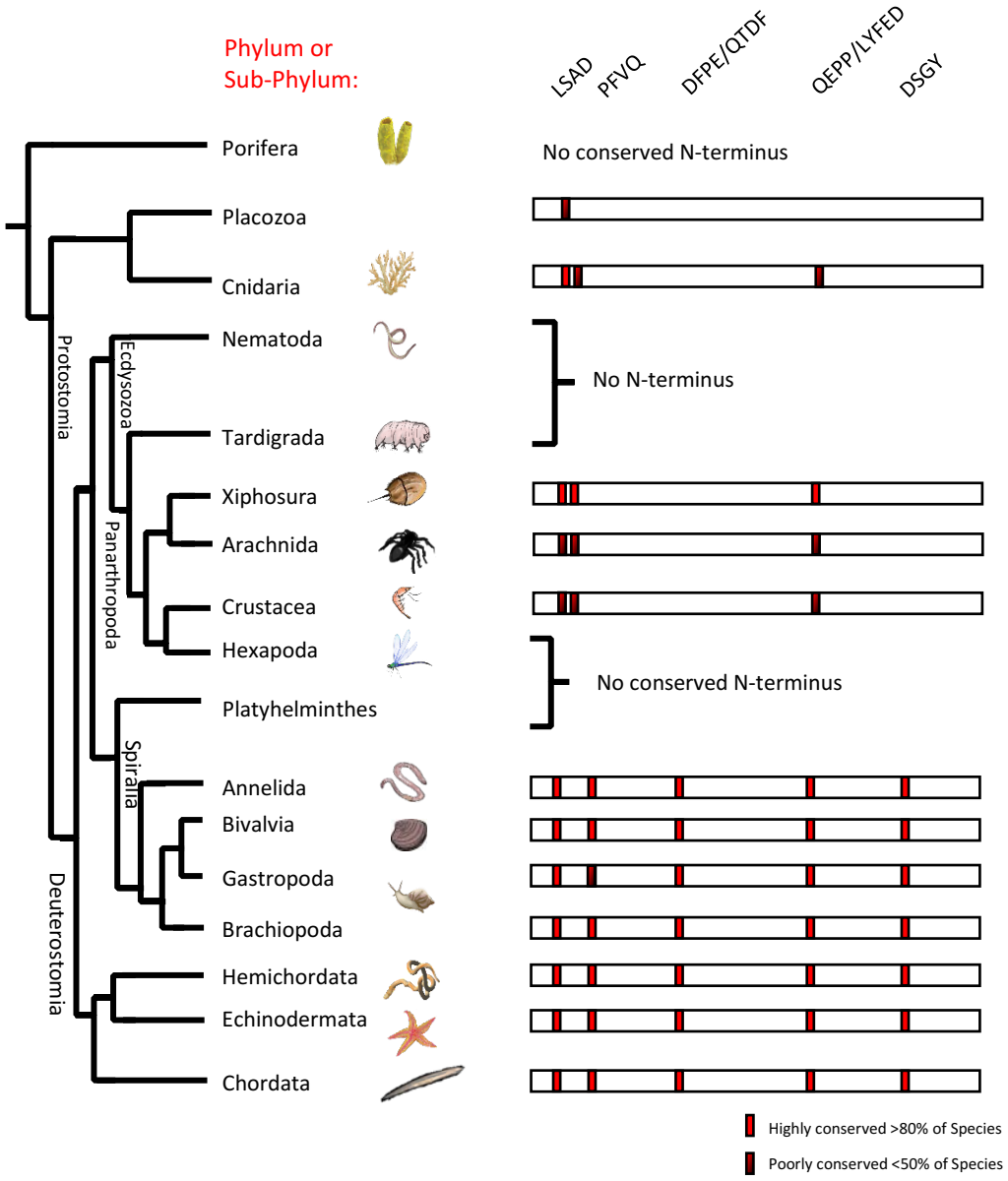


Figure 3-13. Representative diagram of SECIS binding protein N-terminal motif conservation in Metazoans. SBP2 or SBP2L from each phyla of metazoans were analyzed for the presence of N-terminal sequence upstream of the SID. The N-terminal domain of each sequence was aligned to consensus N-terminal motifs to find individual motifs, which are listed above. The diagram depicts evolutionary lineage for each phyla and the relative conservation of each SBP2 N-terminal motif of SBPs discovered in each phylum.

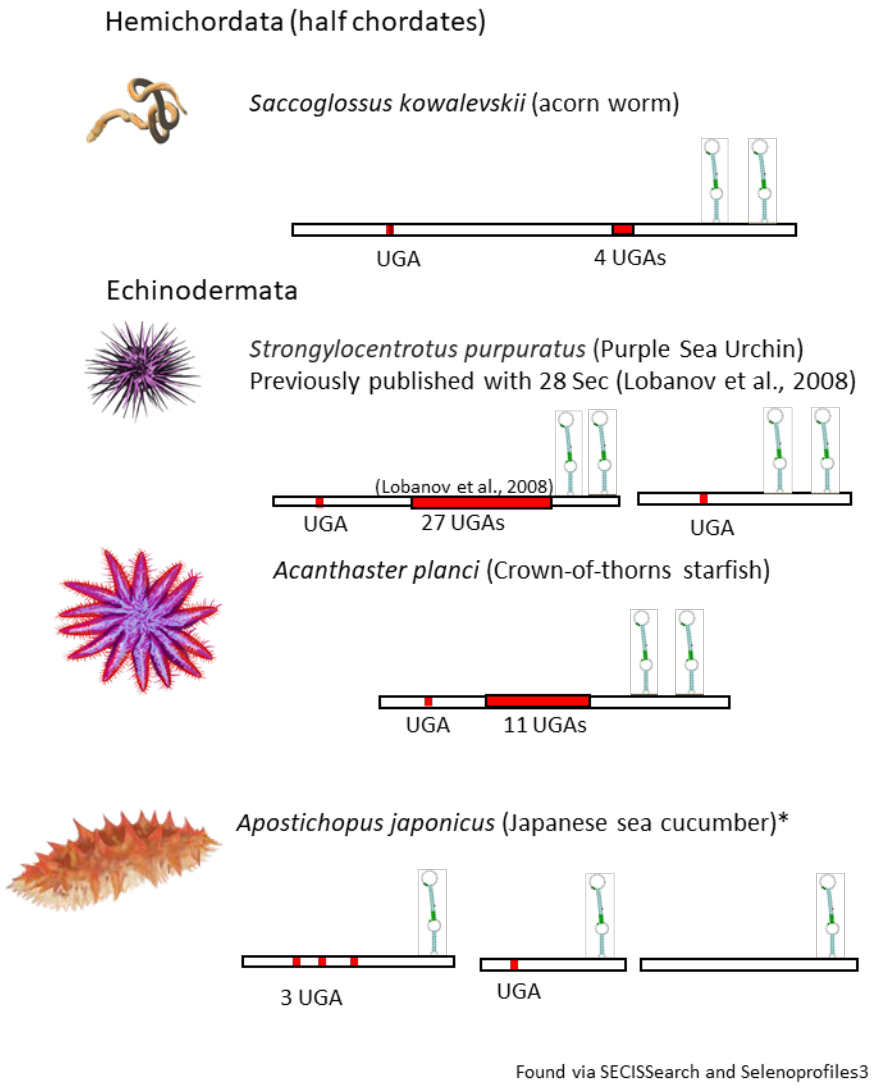


Figure 3-14. Representative diagram of Selenoprotein P found in Ambulacraria. A representative diagram was created from SELENOP mRNAs predicted in each species we found with a conserved motif in the SBP2 N-Terminal Domain. SELENOPs predicted previously are indicated.

Ambulacraria superphylum, which includes hemichordates and echinoderms, a summary of the sequences depicted in Figure 3-14.

Next, we searched Spiralia, a major clade of protostomes, which display spiral cleavage in early development stages. We found very little sequence homology of SECIS binding proteins in anything other than the Lophotrochozoan clade of Spiralia (annelids, mollusks, snails, etc) for SBP2 sequences with N-terminal domains (Figure 3-15). In Lophotrochozoans we expected high conservation of SECIS binding protein motifs due to previous alignments and because of the prediction of *Lottia gigantia* containing a multiple UGA selenoprotein P (Jiang et al., 2012). Of the organisms in this group of organisms we found N-terminal SBP2 sequences only in Gastropods, Bivavles, Brachiopods, and Polychaetes. There appears to be a duplication of SBP2 in *Crassostrea gigas* (Pacific Oyster) and *Crassostrea virginica* (Eastern Oyster). All the N-terminal motifs were well conserved, except for *Capitella telta*, *Aplysia californica*, and *Lottia gigantia. Aplysia californica*, specifically was missing most of the N-terminal SBP2 motifs and further investigation determined this was due to missing exonic sequences from an assembly gap of the sequencing data.

Looking for SELENOPs in the Lophotrochozoans, we succeeded in finding two SELENOPs in the oyster genus *Crassostrea*, both *C. gigas* and *C. virginica* have both a 47 and 17 UGA SELENOP with two SECIS elements. The single Brachiopoda sequence we were able to obtain for SBP2 and SELENOP was from *Lingula anatina,* which had 59 in

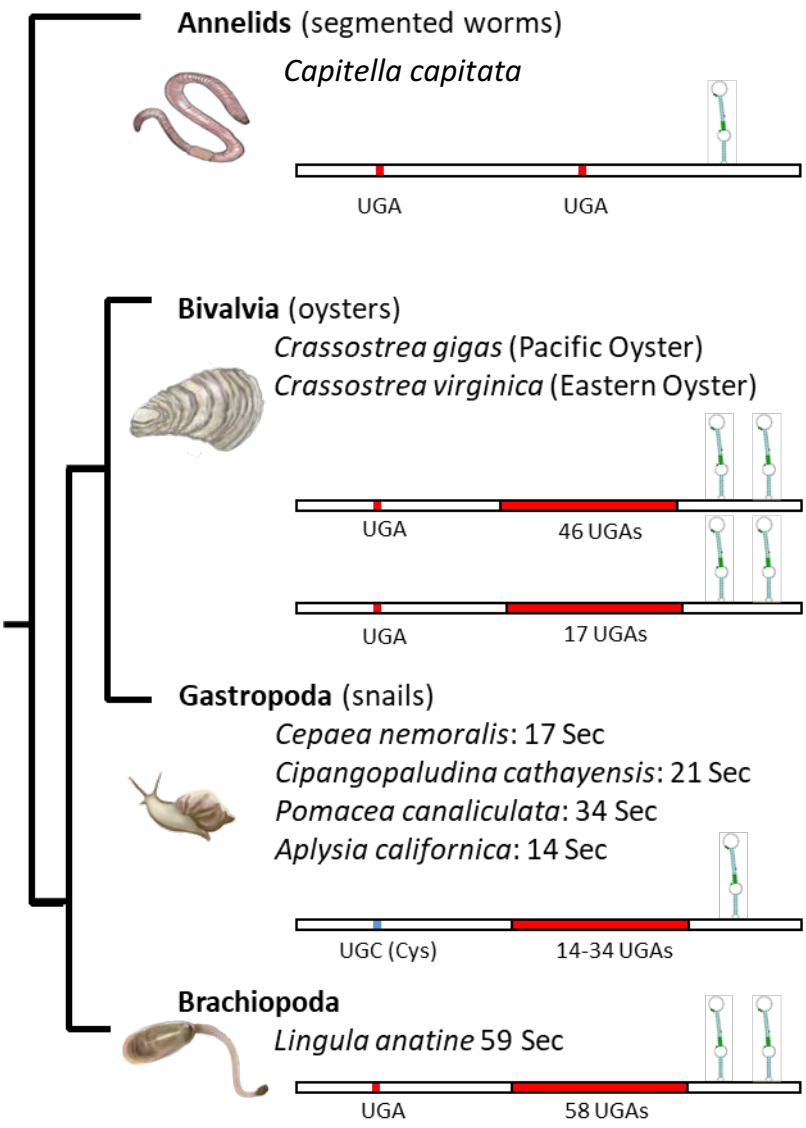


Figure 3-15. Representative diagram of Selenoprotein P found in Lophotrocozoa. Representative diagram from SELENOP mRNAs predicted in each species we found with a conserved motif in the SBP2 N-Terminal Domain. Lophotrocozoans share high N-terminal motif conservation in their SBPs. UGAs that have become a UGC are noted in blue.

frame UGAs with two SECIS elements. In the Gastropoda subphylum all the SELENOP were found had a UGC in the N-terminal position of the thioredoxin fold domain and one SECIS element. Due to the lack of full genomic coverage of some of the organisms in the Gastropod phylum, we found more SELENOPs sequences than SBP sequences. In cephalopods we were able to find a well conserved N-terminal SBP2, however without PFVQ motif, but no SELENOP. In the annelids we were able to find a possible two UGA SELENOP with a single SECIS element in *Capitella telta*, however the context of the first ATG site is unclear. The summary of the novel SELENOP sequences discovered in Lophotrocozoa are found Figure 3-15.

The other major clade of Protostomia is Ecdysozoa (arthropods, tardigrades, and nematodes), which is characterized by at least a three-layered cuticle which is periodically molted. The support for this clade of organisms was identified by 16S rRNA genes and through broad phylogenic gene analysis (Aguinaldo et al., 1997; Dunn et al., 2008). As expected, based on prior work nematodes and hexapods (insects) do not have a high conservation of N-terminal SBP2, which is likely linked to their reduced usage of selenoproteins in comparison to other invertebrates (Chapple and Guigó, 2008).

We found a highly conserved SBP2L sequence in phylum Priapludia with the only sequenced species *Priapulus caudatus.* This organism was of interest to us due to a hybrid development of both deuterostomic and protostomic development and it is considered a living fossil and has demonstrated very slow genomic changes (Webster,

2006). We were unable to locate a conserved SELENOP gene however we found it is predicted to produce a two UGA SELENON.

In Arthropods we found the conservation of SBP2 N-terminal domain and SELENOP primarily in Chelicerates (Chelicerata) clade of Ecdysozoa, which notably includes spiders, horseshoe crabs, scorpions, mites, horseshoe crabs, sea spiders, and ticks. We found sequence conservation of two LSAD N-terminal motifs in only a single species of Crustacea (*Hyalella azteca*). The Myriapod (centipede) *Strigamia maritima*, which contains a short Sec containing N-terminal SELENOP2 gene with two SECIS elements and had only a C-terminal SBP2. In Chelicerates, we began with the order Xiphosura (horseshoe crabs), which is considered the living fossil of the clade and is generally considered a keystone taxon for the understanding of Chelicerate evolution. Of the horseshoe crabs only one species, *Limulus polyphemus* (Atlantic horseshoe crab), has been fully sequenced with EST data. *Limulus polyphemus* makes SBP2 with two LSAD motifs and a QEPP/LYFED motif in the N-terminus of its SBP. When we looked for SELENOP in *L. Polyphemus* we found a 3 UGA SELENOP with a single SECIS element. In the other orders of this clade we found similar retention of SBP2 motifs in spiders, ticks, and scorpions. *Parasteatoda tepidariorum* (house spider) had 10 in frame UGAs with and two SECIS elements in the 3' UTR. *Ixodes scapularis* (deer tick) retained 2 LSAD motifs and the QEPP motif and had a 3 UGA SELENOP with 1 SECIS element like the *Limulus polyphemus*. The golf coast tick, *Amblyomma maculatum*, is already been reported to have a SBP2 and single UGA SELENOP (Budachetri et al., 2017). SEICS data is unavailable for the golf coast tick as 3' UTR data for this gene is not currently available.

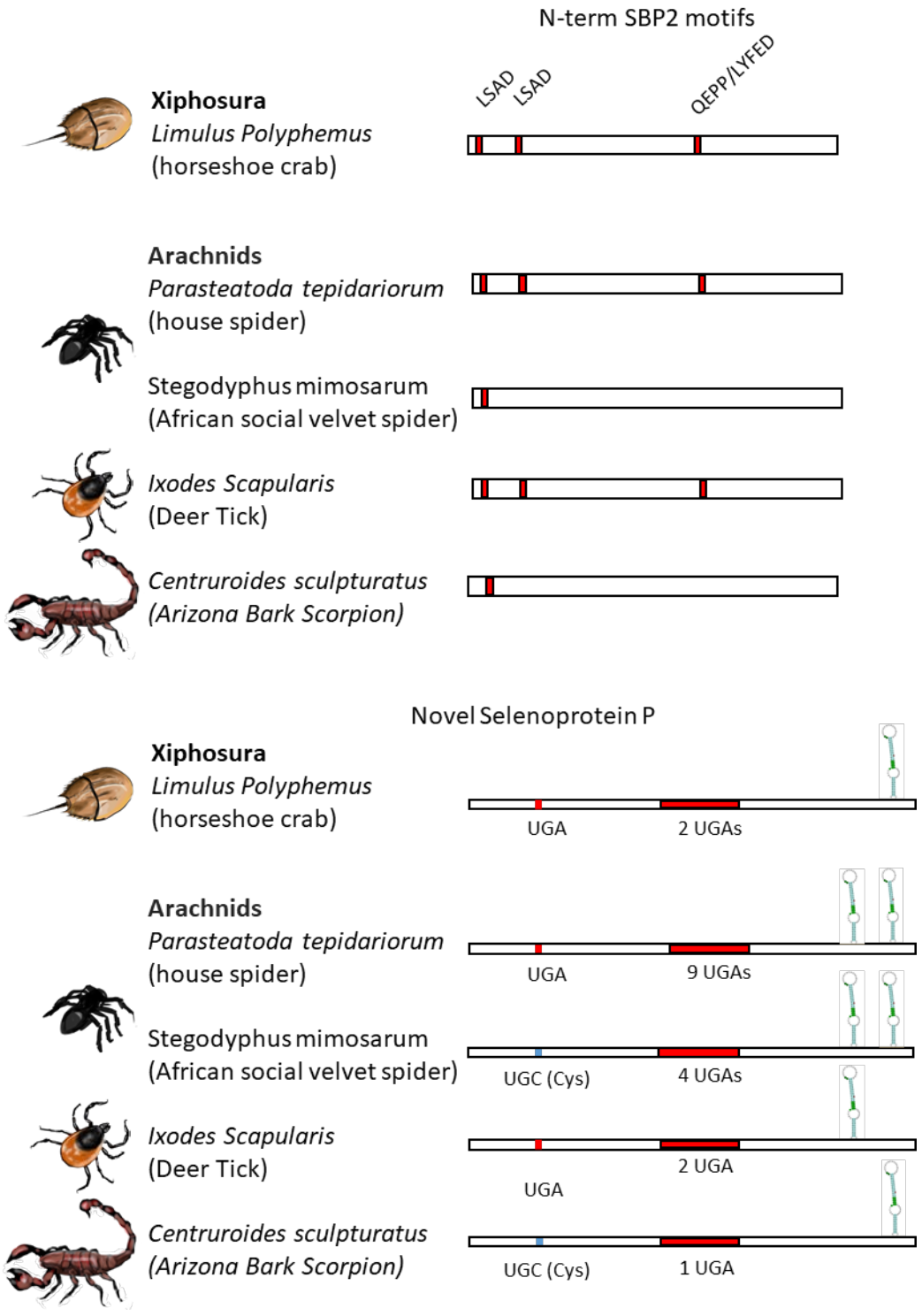


Figure 3-16. Representative diagram of N-terminal SBP2 motifs and Selenoprotein P found in Ecdyzoa. A representative diagram was created to summarize SBP N-terminal conservation (top) and novel SELENOP mRNASs (bottom). Only organisms from the Chelicerata clade (Xiphosura and Arachnids) contained SBP2 and SELENOP and are depicted. No N-terminal SBP2 motifs were found in Nematodes and Hexapods.

In *Stegodyphus mimosarum* (African social velvet spider) and in *Centruroides sculpturatus* (Arizona Bark Scorpion) we see relatively poor conservation of N-terminal motifs in SBP2 and found they have a UGC (Cys) in the N-terminal thioredoxin reductase fold. The summary of the retained N-terminal motifs of SBP2 and novel SELENOP sequences discovered in Chelicerates are found Figure 3-16.

Of the remaining phyla, Cnidarians and Placazoans were the last metazoans with conserved N-terminal motifs from SBP2 and SBP2L. The sole species of fully sequenced Porifera *Amphimedon queenslandica*, although possessing an N-terminal extension to its SBP2, shared no conserved motifs with any other organisms. We found only one species of Placozoa, with the LSAD N-terminal motif, but no SELENOP. While the typical LSAD motif was found in several species of the Cnidarian phyla, the order Scleractinia (stony corals) seemed to be the only one to have a second LSAD motif in the relative position of the PFVQ motif (Summarized in Figure 3-17). The stony corals also appear to retain part of the QEPP/LYFED motif as well, which was not present in the sister order Actiniaria (anemones) nor in the class Hydrozoa (sea hydras). In the stony corals there seemed to be two copies of SELENOP2 both with two SECIS elements which appeared to be localized near a predicted melanopsin gene. There are also multiple predicted initiation sites in the Cnidarian SELENOP2 genes, which are predicted to code for multiple UGAs (7 or more). More RNA based data would need to be published to confirm what the true start site is, as there was no consensus from the exon prediction programs. In the anemones, *Exaiptasia pallida* was the only other species we found with a Sec containing SELENOP2 with one full SECIS element, and one partial SECIS

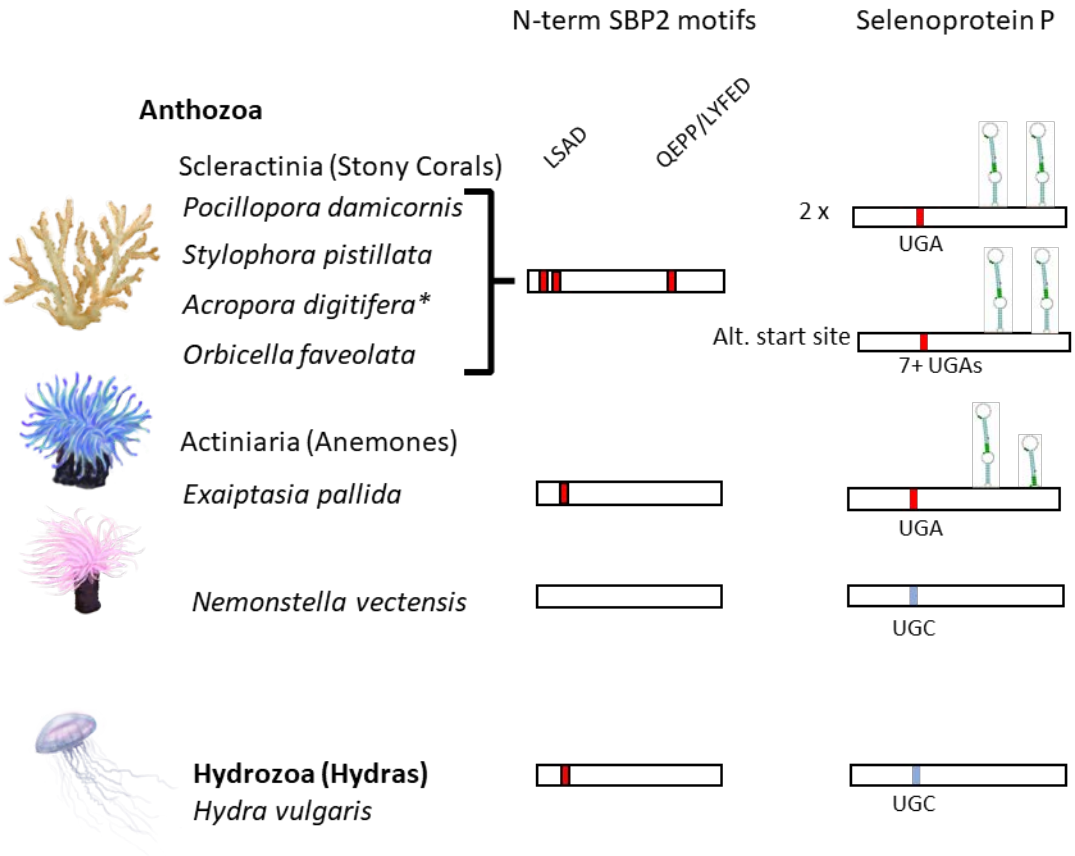


Figure 3-17. Representative diagram of N-terminal SBP2 motifs and Selenoprotein P found in Cnidarians. A representative diagram was created from SBP2 sequences analyzed for SBP2 n-terminal motifs and the presence of SELENOP. *Acropora digitifera* contained a large assembly gap of SBP2 in its genome, but still retained 3 of the N-terminal motifs.

element just downstream missing the first helix in the 3' UTR. The other SELENOP sequences we found in Hydrazoa and Actiniaria contained UGC (Cys) instead or Sec and contained no SECIS elements and both groups do not have a well conserved N-terminal domain of SBP2. The summary of the retained N-terminal motifs of SBP2 and novel SELENOP sequences discovered in Cnidarians are found in Figure 3-17.

While most organisms seemed to follow a pattern of conservation with the N-terminus of SBP2 and the presence of a SELENOP gene, there were some exceptions. Notably some organisms with very high conservation of the N-terminus to multiple UGA included the closest relative to chordates, Tunicata, did have any SELENOP gene. Evidence is emerging for genome consolidation and high rates of evolution, which could explain the lack of SELENOP in that phylum (Bern and Alvarez-Valin, 2014). Another exception was the Priapulid, *Priapulus caudatus,* but which also had every motif retained in it SBP2L. As a sister group to Ecdysozoa it may be possible that there was a loss of selenoprotein utilization over time, however more research would need to be conducted to make that conclusion.

We wondered why the N-terminal motifs of SBP2 would remain conserved in animals without a SELENOP gene and began to investigate if it could also be possible that the motifs could be involved with interactions of *cis* acting factors. For example, the Sec redefinition element (SRE), a hairpin loop found downstream of the UGA, such as the one found in SELENON and SELENOP (Howard et al., 2005; Mariotti et al., 2017). We then searched Tunicates and Priapulids along with other invertebrates for SELENON

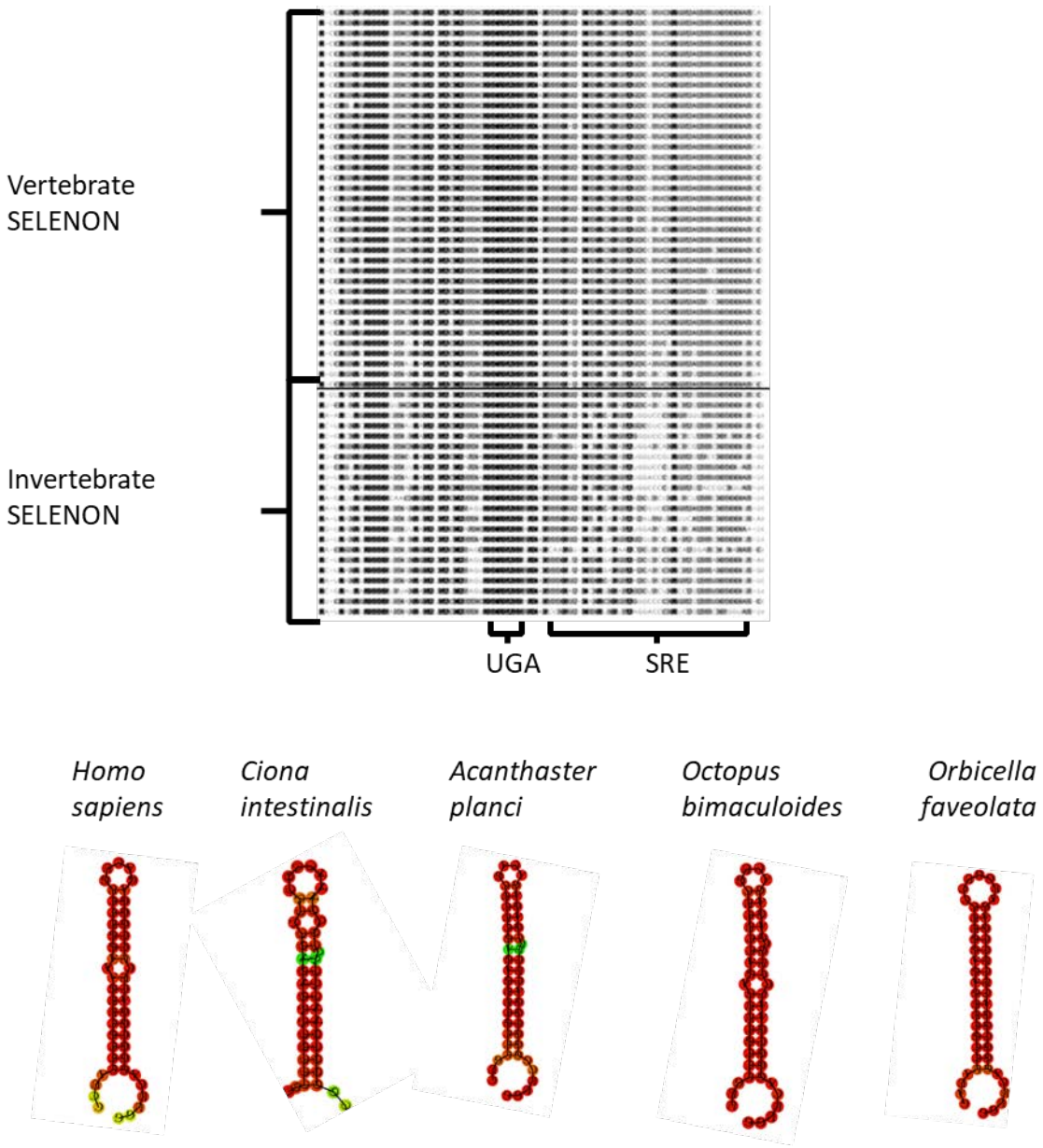


Figure 3-18. SELENON SREs are conserved in invertebrates with conserved N-terminal SBP2 motifs. Top: A nucleotide alignment of SELENON of vertebrates and invertebrates using the MUSCLE algorithm in the conserved sequence surrounding the C-terminal UGA and the SRE in brackets. Bottom: Structural prediction of SREs in humans and other invertebrates by the RNAfold webserver.

via a TBLASTN of SELENON. We then aligned the sequences and found a high conservation of the hairpin loop of the SRE, with only variation in base pairs in the apical loop. The location was well conserved in invertebrates after in the conserved position after UGA position (Figure 3-18, top). Inputting the SRE sequences into the RNAfold webserver, we found a hairpin loop to be predicted even in invertebrates (Gruber et al., 2008; Figure 3-18, bottom). Previously published results on the intrinsically disordered properties of the N-terminus of SBP2 led us to consider what possible function the conserved motifs LSAD, PFVQ, and QEPP/LYFED might be. To gain a better idea of the function, we used intrinsic disorder prediction to determine if there was any significance of these regions using DisEMBL 1.5 (Linding, et al., 2003). We found that the LSAD motif and the QEPP/LYFED motif is predicted to be a hot-loop via intrinsic disorder prediction in humans. Hot-loops are a specific subset of intrinsic disorder, where the loops have a high degree of mobility (Linding et al., 2003).

Discussion

Revisiting SBP2 and SBP2L evolution

The primary aim of our survey of SBPs was to determine the conservation of the N-terminal domain of SBP2 and SBP2L in relation the presence of a multiple UGA C-terminal SELENOP. We also wanted to see which motifs are conserved in the SBP2 and SBP2L vertebrates in order to determine which motifs to look for in invertebrates. We were then wanted to see if novel SELENOPs could be discovered in organisms with conserved these SBP2 and SBP2L N-terminal motifs. The secondary aim of looking at the

evolutionary relationship between SBP2 and SBP2L was to infer how the two proteins with similar functional domains function together without competing.

In order to determine what were the conserved N-terminal motifs the could potentially be important for processive Sec incorporation in SELENOP we decided to do multiple sequence alignments in vertebrates first. This is because we know vertebrates make a multiple UGA SELENOP, and in the case of vertebrates have a conserved C-terminal Sec rich domain. Also, by aligning more closely related phyla would be able to identify motifs more accurately due to the high variability of the N-terminus of SBPs across different phyla. By increasing the number of sequences and focusing on SBPs in vertebrates to search for unannotated genes we were able to discover motifs that originally did not align: DFPE/QTDF and QEPP/LYFED. Through multiple alignments of previously annotated genes when aligned and sorted emerged 3 major groups of SBPs: SBP2L of all chordates, non-mammalian SBP2, and mammalian SBP2. As of now, the story of SBP2 in vertebrate began with the first duplication of SBP2L, likely from a whole genome duplication based on data of 2 rounds of whole genome duplication events occurring in the vertebrate lineage (Ohno, 1970; Tang and Thomas, 2018; discussed in Kasahara, 2007). This duplication in early vertebrates lead to two different paralogous genes SBP2 and SBP2L. The second SBP2 duplication later in vertebrate evolution did not result in three SECIS binding proteins, but lead to the divergence of mammalian SBP2 from the rest of the vertebrates. It would be interesting to compare the difference of vertebrate SBP2s *in vitro*, to see if these differences were related to efficiency or

processivity. The differences in mammalian SBP2 could also be involved in other functions *in vivo*.

Our search for N-terminal motifs in SBP2L showed high conservation of the SEIS motif and the surrounding sequences connecting it to the SID. This led us to wonder if the motif may be part of the SID region, and through protein prediction software we were able to determine some of the SEIS sequence was part of a helix motif. This seemed odd because invertebrate CTSBP2L from *Capitella* was capable of Sec incorporation *in vitro* but not human CTSBP2L (Copeland et al., 2001; Donovan and Copeland, 2012). Upon further inspection we found this construct did started further upstream just after the after the LYFED motif, while the human construct started after the SEIS motif. This could be the reason recombinant hCTSBP2L has not functional *in vitro* due to the SID position being in a different location than the SID in SBP2. It does not explain, however, why we have been unable to make a functional assay using *in vitro* translated full length SBP2L in the past (Donovan, 2011).

When *de novo* predictions of SBP2 genes included *via* the Selenoprofiles pipeline lead to the discovery of a set of SBP2 pseudogenes present in mammals, which suggested a second gene duplication event for SBP2. The duplication also seems to be the reason for divergence in the conserved motifs of SBP2 in mammals. The second SBP2 duplication in early mammalian evolution seems to have led to a small sequence divergence in SBP2 compared to the other chordates.

The story and fate of SBPs changes with new evidence of a separate gene duplication of SBP2 in mammals. To date, the only phylogenetic analysis of SBPs across eukaryotic taxa was conducted by one group and provided evidence of SBP2L and SBP2 being paralogous genes and a high likelihood of SBP2L being the progenitor of SBP2 (Donovan and Copeland, 2009). The retention of SBP2L in mammals and a second duplication of SBP2 becoming a pseudogene provides more evidence of a specific role for SBP2L. The answer to why this occurred is not clear, but SBP2L could be providing some other unknown function *in vivo*. In terms of the evolutionary scale SBP2L is considered an ancient gene (greater than 500 million years ago). Mammalian ancient genes are typically express in more abundance their younger counterpart SBP2. It has been shown that younger genes in humans, which are defined to have duplicated at the emergence of mammals, tend to be express significantly more in the testis and liver than their paralog (Guschanski et al., 2017). On the other hand, ancient paralog in humans are expressed significantly more in the neural tissue (Guschanski et al., 2017). When looking at the global transcriptome data this turns out to be true for SBP2 and SBP2L, which supports our finding of a second SBP2 duplication in mammals (Fagerberg et al., 2014). SBP2 expression is regulated by 3' UTR binding proteins governing turnover and translation (Bubenik et al., 2009). The same RNA binding motifs are found in the 3' UTR of SBP2L as well, suggesting both SBPs transcriptional expression are linked.

Retention of paralogous genes from whole genome duplications, typically change their function, expression, or location to avoid conflicts when both proteins are

expressed. Prime examples of this phenomenon are Hox and opsin genes, which have diverged into different forms to specialize in different functions across many eons of evolution (Feuda et al., 2012; Reviewed in Dong et al., 2012; Feuda et al., 2016; Barucca et al., 2014; Osoegawa et al., 2014). One key difference In SBP2 and SBP2L in vertebrates was the loss of the known CRM1 dependent NLS domain in SBP2L in the jump from lancelet to fish, which could have caused a change the general localization of each protein in the cell. Localization differences would allow for the co expression of both SBPs while reducing the competition for the same factors for both genes. The change in localization between the nucleus and cytoplasm would allow both proteins to be expressed without competition between each other.

The role of SBP2L still remains unclear, but one hypothesis for function could be for a ROS mediated stress response. There is no *in vitro* data to support a role for SBP2L directly in Sec incorporation and the only *in vivo* data is indirectly based on SBP2 deletion in cells still express selenoproteins (Seeher and Schweizer, 2014; Fu et al., 2017; Dubey and Copeland, 2016). SBP2 is localized to the nucleus under high ROS conditions, but it would not make much sense to downregulate ROS mitigating proteins under high ROS conditions. SBP2 appears to also play a role in selenoprotein mRNA regulation, as deletions of the gene reveal a global reduction in mRNA expression possibly by prevention of NMD (Mariotti et al., 2017). Under these conditions SBP2L could temporarily take over the role of SBP2 Sec incorporation in the cytoplasm, while SBP2 increases mRNA expression by prevention of NMD when shuttle into the nucleus. SBP2L is likely to be as sensitive to oxidation as well but appears to be expressed in all

tissues at a higher level than SBP2 according to global transcriptome data (Fagerberg et al., 2014; Duff et al., 2015). The higher expression could allow for a higher turnover rate for SBP2L under ROS stress conditions and could potentially compensate for this. Another reason for the apparent lack of SBP2L Sec incorporation activity *in vitro* could be that it requires post translational modification or chaperone mediate folding. This is because it is an intrinsically disordered protein (Oliéric et al., 2009). Disorder proteins can also become ordered by RNA binding or post translational modifications, which could be the fate of SBP2L (Tompa et al., 2005; Peng et al., 2014). Work is currently being conducted in our lab in Zebrafish and mammalian cell lines with CRISPR deleted SBP2L to get a better understanding the of its role *in vivo*.

Conservation of N-terminus of SBP2L and SELENOP by Invertebrates

By determining which parts of the N-terminal domain of SBPs were conserved in SELENOP producing animals we were able to find several specific motifs which were retained throughout chordate evolution. With a vertebrate only alignment of SBP2 and SBP2L we were able to gain a consensus on N-terminal motifs to search for in invertebrates. The LSAD motif was a highly conserved motif among organisms with an N-terminal extension on their SECIS binding protein with SELENOP. Trichoplax was only organism we found with a conserved N-terminal motif (LSAD) without a SELENOP and SELENON. One unique characteristic about Trichoplax however, is it also has an abnormally large number of iodothyronine deiodinase (DIO) like proteins, 11 in all (Jiang et al., 2012). Even in the early metazoan Cnidaria there were several SELENOP2 genes with two SECIS elements predicted, and some with a possible multiple UGA depending

on the start site. An interesting finding was with most cases of two SELENOP2 genes in the Scleractinia (stony corals), was their proximity to a predicted melanopsin gene. Opsins have been well studied for their retention after gene duplication events and eventual differentiation for different kinds of photoreceptors (Dong et al., 2012). This could be insight into the origin of the first duplication event of SELENOP, as the proximity of both SELENOP2 genes in the corals are relatively close. If the melanopsin gene was translocated the two genes could have merged over time and though separate gene duplication events create the Sec rich C-terminus in an early progenitor to bilaterians.

The function of the N-terminal domain of SBP2 in processive Sec incorporation is still speculative, however the retention of the N-terminus in the metazoans is correlated with SELENOP and SRE containing selenoproteins. The simplest reason for the conservation of the SRE of SELENON in invertebrates suggests that the SRE was an evolutionary advantage by increasing UGA recoding efficiency. Since the availability of selenium is variable on land and generally lower than the ocean, more efficient Sec incorporation could have been a selection pressure especially in the case of processive Sec incorporation. It is plausible that the N-terminus interacts with the SRE during the process of Sec incorporation and allows more efficient delivery of EFSec-GTP and Sec-tRNA$^{[Ser]Sec}$ to the ribosome. The interaction might help also help with the stability and UGA recognition as many selenoproteins can have long 3’ UTRs in which the SECIS element is far from an in frame UGA. The LSAD motif retention with a basic motif just

before the SID is evolutionarily conserved with SELENOP or SELNEON across metazoans implicates its importance as a requirement for processive Sec incorporation.

Chapter 4: Evaluation of Processive Sec incorporation by Full length SBP2 *in vitro*.

Introduction

Although unnamed at that time, SBP2 was first discovered through UV-crosslinking of a 120Kd protein from testis extracts to the glutathione peroxidase 4 (GPX4) 3' UTR (Lesoon, 1997). From the same paper, this 120 kDa protein was also found to require the AUGA in the SECIS element for binding and correlated to selenoprotein synthesis (Lesoon, 1997). Around the same time there were reports of other SECIS binding proteins, but none of these reports demonstrated any involvement for selenoprotein synthesis. The 120kDa protein was later purified and characterized as a *bone fide* selenoprotein synthesis protein after demonstration as a requirement *in vitro* and was also named SECIS binding protein 2 (SBP2) (Copeland and Driscoll, 1999; Copeland, et al., 2000).

Since the discovery of SBP2 as an essential part of eukaryotic Sec incorporation, SBP2 has been well studied in the C-terminal RBD and SID. In comparison, the N-terminus of SBP2 has not been well characterized. What is known is the N-terminus is it not essential for single Sec Incorporation in WGL and RRL nor is it required for SELENOP synthesis in RRL (Copeland et al., 2001; Gupta et al., 2013). Furthermore, the N-terminal domain does not enhance single Sec incorporation activity *in vitro* (Mehta et al., 2004). The only published evidence for any kind of function comes from partial truncation mutants causing selenium deficiency and low serum SELENOP, which is not changed by

Se supplementation (Donovan and Copeland 2009; Di Cosmo et al., 2009). The rest of the evidence for N-terminal function in SELENOP synthesis is the previous work in chapters 2 and 3.

Based on the bioinformatics in chapters 2 and 3 we began to wonder if processive Sec incorporation could be mediated by the N-terminal domain of SBP2. This is also supported based on the addition of retic RSW, which was able to synthesize full length SELENOP without the addition of SBP2. Implicating that either endogenous full length SBP2 or SBP2L was capable of processive SELENOP synthesis. Since in our assays recombinant SBP2L is not functional we focused on evaluating the processive Sec incorporation activity of full length SBP2. The goal of this chapter was to determine if the N-terminus of SBP2 is a requirement for processive Sec incorporation *in vitro*.

Materials and Methods

Constructs and mRNA synthesis: All constructs used for mRNA synthesis for use in vitro was linearized with NotI (New England Biolabs) and used as a template for T7 RNA polymerase in the presence of $m^7G(5')ppp(5')G$ using the mMessage mMachine kit from Ambion. The Rat SELENOP construct containing the coding region, full 3' UTR, C-terminal FLAG tag, was made by isolating from McArdle 7777 cells and cloned by Topo-TA ligation into PCDNA 3.1 (Invitrogen) as described in Shetty et al., 2014. The wildtype Zebrafish SELENOP construct was a gift from Dr. Vadim Gladyshev (Harvard Medical School, Boston, MA). The coding region and 3' UTR was ligated into PCDNA3.1 and a C-terminal Flag tag was added to the C-terminus and SECIS mutants of the Zebrafish

construct using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies). The resulting Zebrafish SELENOP construct containing the zebrafish SELENOP coding region, C-terminal FLAG tag, and the full wild type zebrafish 3' UTR was cut and ligated to zebrafish SECIS I alone or human SECIS I alone. Luciferase mRNAs used in this chapter were made as described before (Shetty et al., 2014; Gupta and Copeland, 2007).

In vitro Luciferase activity assay: The Sec incorporation was determined using an *in vitro* RRL translation system with a luciferase reporter mRNA, containing an in-frame UGA codon at position 258 followed by the wild type rat SELENOP SECIS element in the 3' UTR WT or with SECIS 1 or SECIS 2 deleted). Each 12.5 µl reaction contained 6ul (48%) of Flexi Rabbit reticulocyte lysate, 0.5 pmol recombinant FLSBP2 or CTSBP2, 20 µM amino acid mix, 125 ng of mRNA, a final Mg^{2+} concentration of 2mM, and 400nM final concentration of either cold or ^{75}Se sodium selenite with a specific activity 6.29 µCi/µl (depending if SDS-PAGE analysis was being conducted). The reactions were incubated at 30°C for 1 hour then brought up the final volume 100µl with 1xPBS before measurement on a 96-well plate luminometer (Berthold Tristar).

Purification of CT and Full length human SBP2: Recombinant Xpress/ His-tagged human C-terminal SBP2 (CTSBP2) and Full length SBP2 (FLSBP2) and was purified as previously described in (Kinzy et al., 2005) and (Pinkerton and Copeland, 2017). After initial purification, the proteins were buffer exchanged by dialysis. To increase yield and specificity and to reduce background contaminants from the His column in Figure 4-7, SBP2 was buffer exchanged immediately following elution from a nickel column onto a GE Sephadex-200 column in 1x PBS, 2 mM DTT and 20% glycerol. 0.5mL fractions were

collected and a western blot was used to determine peak fractions. To quantify the proteins the peak fractions of both CTSBP2 and FLSBP2 were pooled and ran on the same SDS-PAGE gel with a BSA standard. The gel was then coomassie stained and bands were quantified by densitometry by GE Imagequant software.

Invitro translation of SBP2 and SBP2L for retic fractionation: a 125 µl standard *in vitro* translation reaction for RRL with full length SBP2 or SBP2L mRNA was conducted with and without 35S Met labeling for each mRNA. After 45min at 30°C the reactions were fractionated as described in Chapter 2: Fractionation of Mammalian Lysate for addition to Wheat Germ Lysate to be used in a WGL assay.

In vitro translation of Rat Selenoprotein P in Wheat Germ Lysate: Using a rat SELENOP construct containing the coding region and the full 3' UTR. Each 12.5 µl reaction contained 6.25 µl WGL (Promega), 320 nM recombinant SBP2, 320 nM wild-type His-tagged eEFSec, 20 µM amino acid mix, 200 ng of SELENOP mRNA and 75ng/µl (6000cpm) RPC-5 pool of total ^{75}Se Sec-tRNA$^{[Ser]Sec}$, methylated and unmethylated. The reactions were incubated at 25 °C for 2 h. 12.5µl Rabbit reticulocyte lysate of a control reaction contained 320 nM recombinant CTSBP2, 320 nM wild-type His-tagged eEFSec, 20 µM amino acid mix, 200 ng of SELENOP mRNA and 75ng/µl (6000cpm) RPC-5 pool of total 75Se Sec-tRNA$^{[Ser]Sec}$ (methylated and unmethylated). Samples were analyzed using SDS PAGE electrophoresis, the entire reaction was loaded for WGL, while 4µl was used for RRL. The gels were dried and exposed to a phosphoimage screen and scanned on a Typhoon Phosphoimager (GE Healthcare). Bands were quantified by densitometry with GE Imagequant software.

In vitro translation of Zebrafish SELENOP in Flexi Rabbit Reticulocyte lysate: Each 12.5 µl reaction contained 6µl (48%) of Flexi Rabbit reticulocyte lysate (Promega), 0.5 pmol recombinant FL or CT SBP2, 20 µM amino acid mix, 125 ng of mRNA, a final Mg^{2+} concentration of 2mM, and 400nM final concentration of ^{75}Se sodium selenite with a specific activity 6.29 µCi/µl. The reactions were incubated at 30°C for 1 hour. Samples were analyzed using SDS PAGE electrophoresis, where 4µl of the reaction was loaded. The gels were dried and exposed to a phosphoimage screen and scanned on a Typhoon Phosphoimager (GE Healthcare). Bands were quantified by densitometry with GE Imagequant software.

Results

Based on our results we hypothesized that SBP2, SBP2L or both could be a requirement for processive Sec incorporation for SELENOP synthesis. First, we wanted to determine if commercial RRL used in our previous assays to promote processive Sec incorporation contained detectable amounts of SBP2 and SBP2L. We analyzed 3 different batches of RRL and probed with either an SBP2 or SBP2L primary antibody. We were able to detect SBP2 and SBP2L. SBP2L had a low signal to noise ratio in our blots and the primary detected product in our SBP2 blots were smaller than expected, we typically observe SBP2 and SBP2L to run approximately 20 kDa higher than predicted (Figure 4-1, top and bottom respectively). We are not able to rule out if the resulting lower bands were from degradation or initiation from a downstream AUG. We also wanted to confirm SBP2 and SBP2L localizes and concentrates to the RSW of RRL using

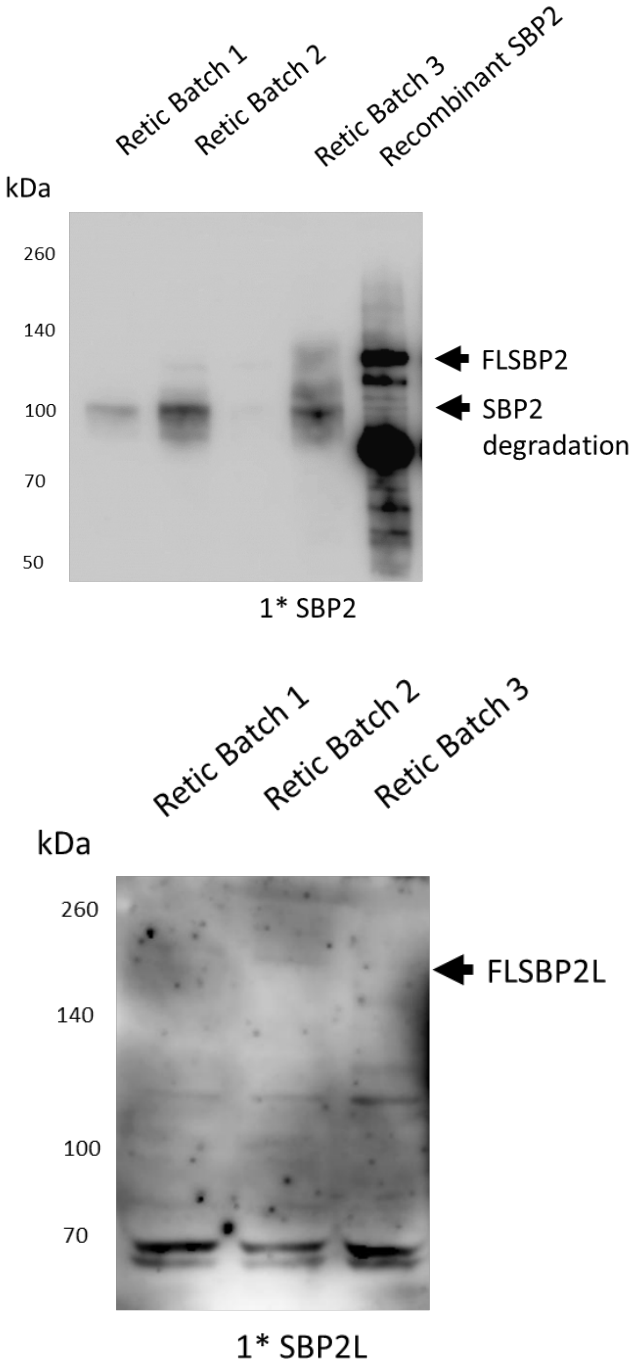


Figure 4-1. Detection of endogenous SBP2 and SBP2L in RRL by immunoblotting. Immunoblot analysis RRL lysates from multiple batches with either SBP2 (top) or SBP2L (bottom). Lysates were run on a 10% SDS-PAGE gel for immunoblot detection. Expected size of full length protein is indicated by arrows.

our procedures in Chapter 2. To confirm if SBP2 and SBP2L localized to the RSW in our experiments, we *in vitro* translated full length SBP2 and SBP2L mRNAs in RRL using ^{35}S Met labeling to trace SBP2 and SBP2L and fractionated via two 300,000 x g spins to separate crude ribosomes from retic. The crude ribosomes were washed with a high salt buffer (0.5M KCl) and spun again at 300,000 x g to separate the ribosomal salt washed proteins from the ribosomes. We found that SBP2 and SBP2L was sparse in the supernatant, but concentrated in the RSW and the P300 ribosomal pellet (Figure 4-2, lanes 1, 3, 5, and 7). Salt washed ribosomes were completely removed of SBP2 and SBP2L binding as well (Figure 4-2, top, lanes 4 and 8). A western blot also revealed detectable levels of SBP2 in the ribosomal salt wash at the expected molecular weight and S300 of pretranslated SBP2 RRL as well (Figure 4-2, bottom).

After confirmation of the localization of both SBP2 and SBP2L to the ribosomal salt wash, we wanted to see if either of these proteins could increase processivity. A technical issue that arises with synthesis of full length SBP2 (FLSBP2) and full length SBP2L (FLSBP2L) is low yield and degradation of the final products. The main cause of this issue with the SBP2 and SBP2L is likely due to the disordered N-terminal domain, which makes it vulnerable to proteases (Olieric et al., 2009). The caveat of using recombinant protein meant it was challenging to purify high concentrations of SBP2, which reduced the maximum amount of protein we were able to use in the following assays.

Due to difficulties in producing high concentrations of FLSBP2 we wondered if the N-terminal domain of SBP2 or SBP2L could increase processivity separate from Sec

incorporation. Previous reports using the SBP2 SID and RBD as separate recombinant proteins have reported successful Sec incorporation *in vitro* which led us to test this hypothesis (Donovan et al., 2008). We added 2 pmol CTSBP2 in a WGL reaction with increasing amounts of either FLSBP2 or FLSBP2L to see if the addition of these proteins could increase processivity. FLSBP2 alone appeared to produce a faint full-length band and a band corresponding to the 2nd UGA, while CTSBP2 had multiple bands as observed before (Figure 4-3 lanes 1 and 2). The addition of increasing amounts of FLSBP2 in combination of CTSBP2 saw a slight dose dependent shift in full length product signal compared to early termination products (Figure 4-3 compare lane 1 to lanes 4-6). The addition of FLSBP2L, however, seemed to ablate Sec incorporation in a dose dependent manner (Figure 4-3, lanes 7-9). While the changes were minor, it could have been due to the small quantities of FLSBP2 added compared to CTSBP2. Also, based on this result it seems unlikely that the observed increase in processivity with RSW in WGL was due to CTSBP2 and SBP2 or 2L N-terminal domain interacting separately to increase processivity.

Due to the function of processivity not likely to be function as separate domains. We wanted to determine if FLSBP2 was able to increase processivity when comparing equal amounts of CTSBP2 independently. FLSBP2 production was optimized to increase the highest yield with the least possible degradation products. We added a range of equal amounts of rat CTSBP2 (used in previous experiments) or human FLSBP2 at concentrations of 0.4, 0.8, 1.2, 1.6 pmol into a WGL reaction. In a comparison of FLSBP2

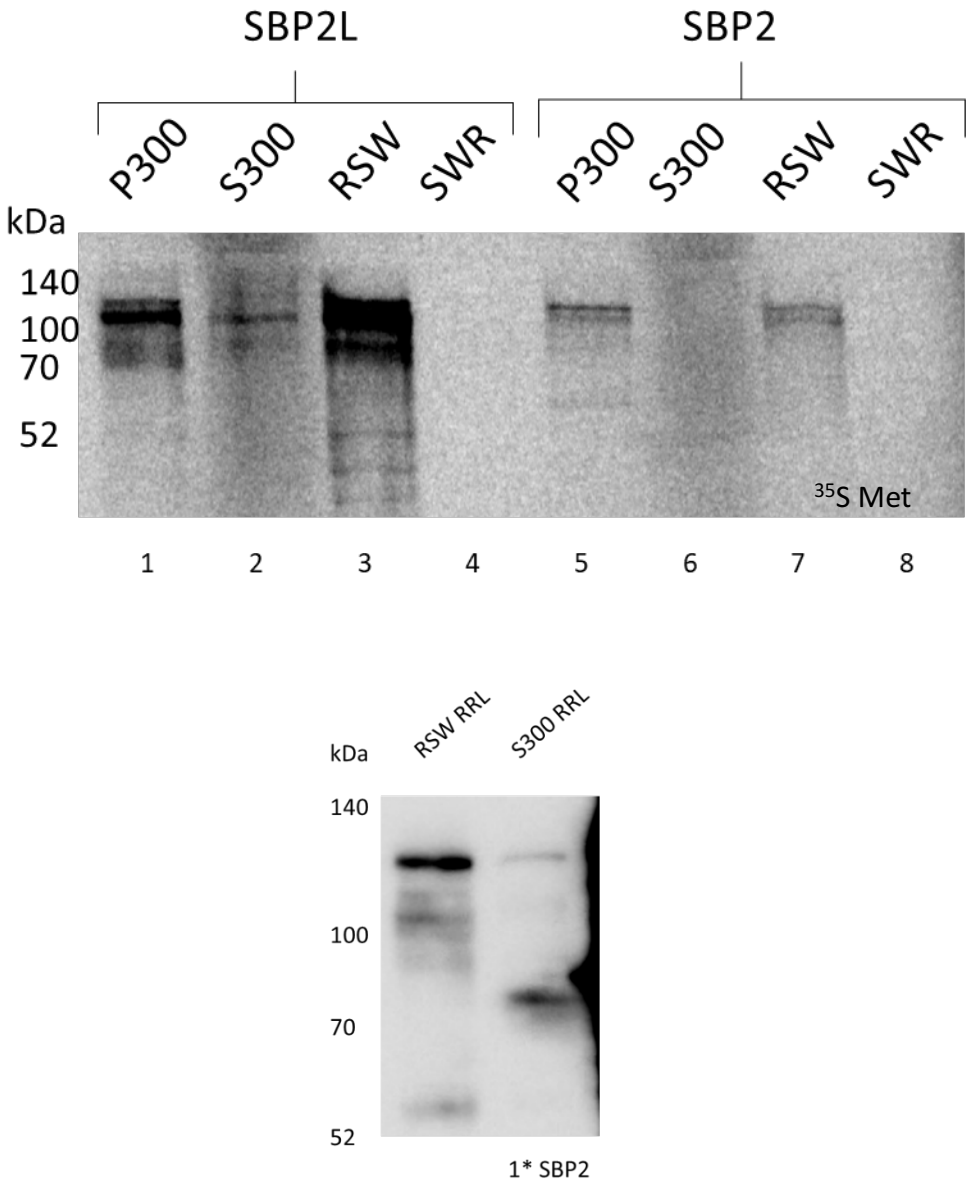


Figure 4-2. Localization of pretranslated SBP2 and SBP2L in RRL after fractionation. SBP2 or SBP2L mRNA was pretranslated in RRL in a standard *in vitro* reaction with and without ^{35}S Met (top and bottom respectively). After the reaction was complete RRL was centrifuged at 300,000 x g to separate crude ribosomes (P300) from the remaining RRL (S300) (prepared as described in Figure 2-6). The P300 was washed with 0.5M KCL to obtained and spun again to obtain a salt wash ribosomes (SWR) in the pellet and a ribosomal salt wash (RSW) in the supernatant. Fractions from the ^{35}S Met reactions were ran on a 12% SDS-PAGE gel for autoradiography analysis (top). Unlabeled SBP2 translated reactions of the RSW and S300 fractions were ran on a separate gel for immunoblot analysis with SBP2 primary antibody.

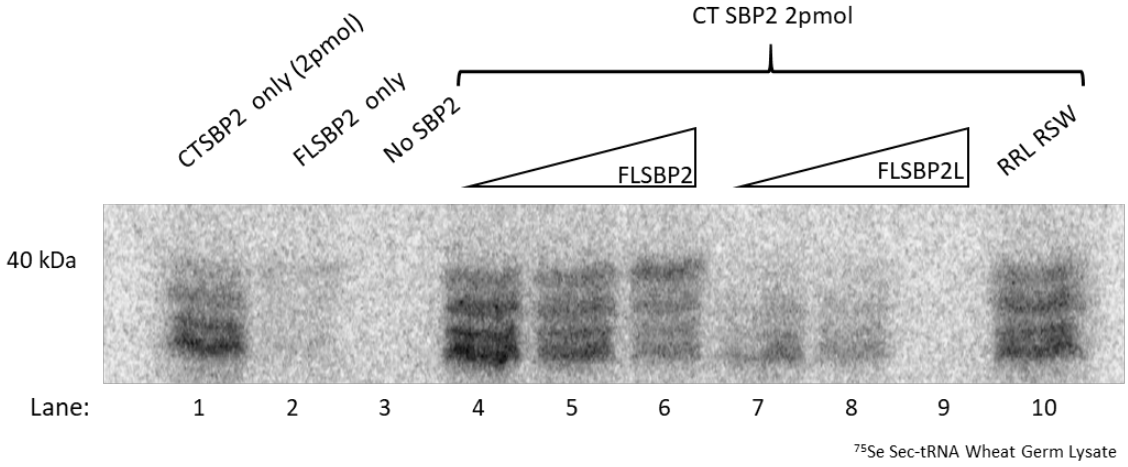
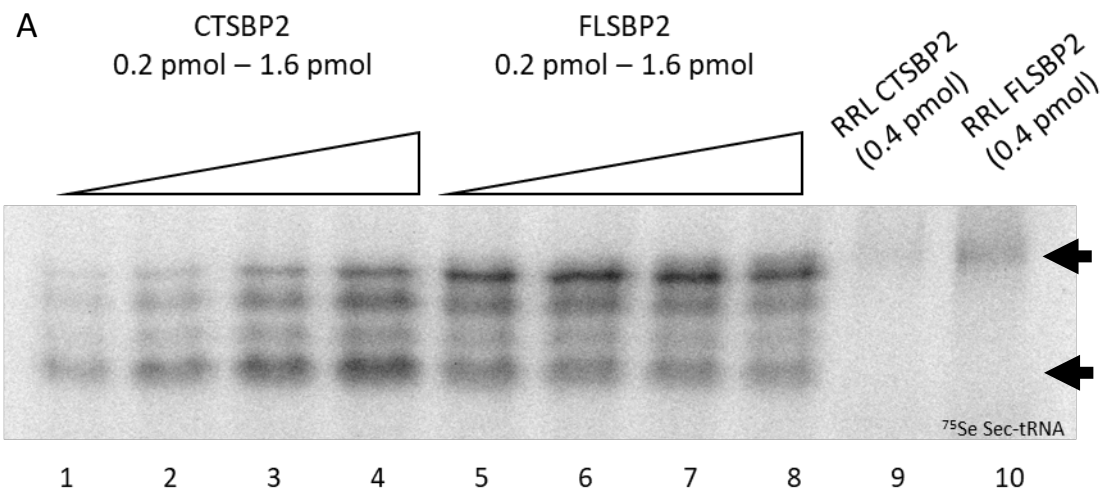


Figure 4-3. FLSBP2, but not FLSBP2L can increase full length SELENOP synthesis in wheat germ lysate. SELENOP mRNA was translated in WGL supplemented with RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$, eEFSec, 2pmol CTSBP2. The standard reaction was then supplemented with additional FLSBP2 (0.2-0.6pmol), FL SBP2L (0.2-0.6pmol), or RRL RSW. Reactions were loaded onto a 15% SDS-PAGE gel for autoradiography analysis.

to CTSBP2 there is a notable increase in the high molecular weight products compared to total SELENOP products (Figure 4-4, top, compare lanes 1-4 to lanes 5-8). There also appears to be a fifth band that becomes visible in the highest concentrations of FLSBP2 that is not present in the higher concentrations of CTSBP2. In RRL, there was also more full length SELENOP product present with the addition of full length SBP2. We wonder if the full-length protein was more efficient for single Sec incorporation too but found the full-length protein to be approximately 2.5 times less efficient for single Sec incorporation in this assay (Figure 4-4, bottom). This this result highlights the fundamental difference of efficient single Sec incorporation being separate from processive Sec incorporation.

We then repeated this experiment across the optimal range of SBP2 for processive Sec incorporation for quantitation to see if this phenomenon was repeatable. Gel Band quantitation was performed with using GE Imagequant on SDS-PAGE gels (n=4) comparing the 0.5 pmol FLSBP2 and CTSBP2 reactions to quantitate the 42.6 kDa band in proportion to total detectable products, a student's t test reported a p value of 0.0019. Indicating that the N-terminal domain of SBP2 significantly increases processive Sec incorporation in SELENOP over CTSBP2 only (Figure 4-5).

Even though we saw an increase in processivity with FLSBP2 we did not observe the same shift in processivity as we did with the addition of ribosomal salt wash from rabbit retic lysate. Due to this observation, we wanted to test if the endogenous protein is more "active" and if the ribosomal salt wash was able to help with folding or the



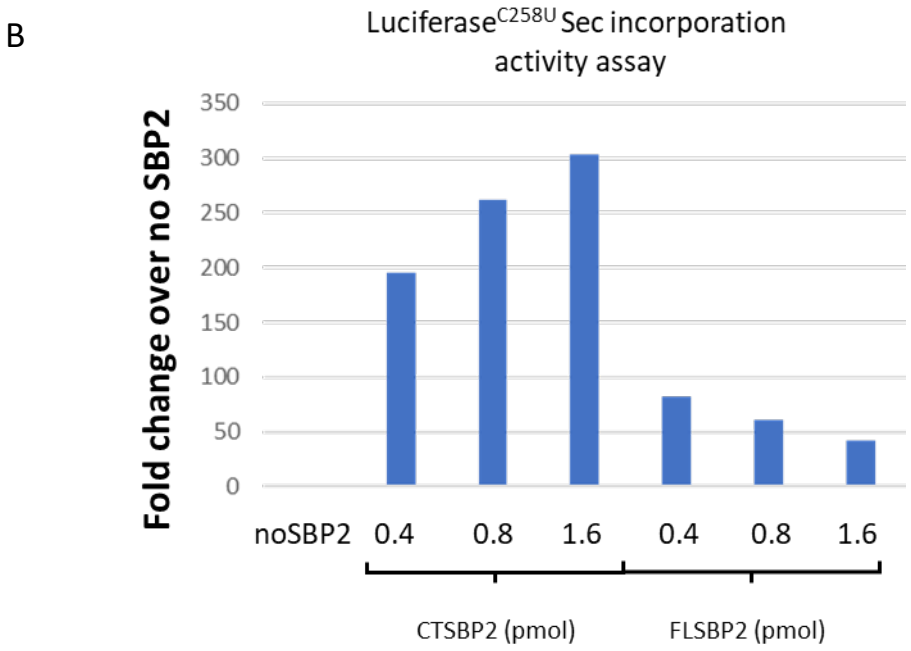
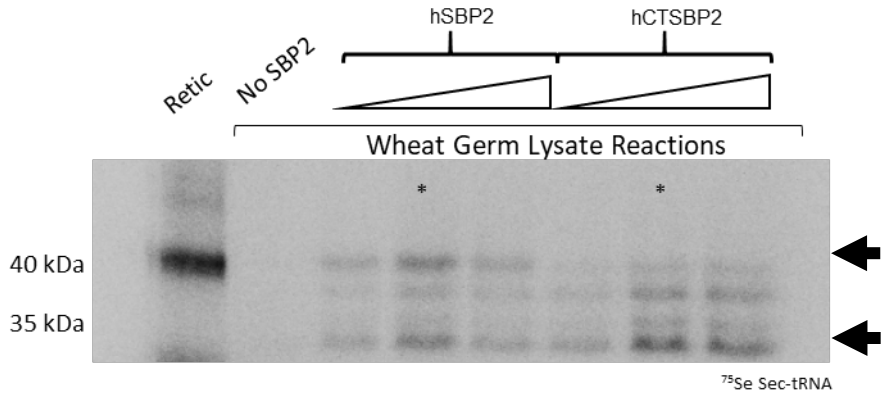


Figure 4-4. FLSBP2 can increase processive Sec incorporation in SELENOP but does not increase single Sec incorporation efficiency in WGL. A. SELENOP mRNA was translated in WGL supplemented with RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$, eEFSec, and the addition of either recombinant FL SBP2 at 0.2 pmol to 1.6 pmol by doubling the concentration in each lane. SELENOP mRNA was also translated in a control RRL reaction supplemented with 0.4pmol of CTSBP2 or FLSBP2 and RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$. Reactions were loaded onto a 15% SDS-PAGE gel for autoradiography analysis. **B.** Luciferase^{C258U} mRNA was translated in WGL supplemented with testis total tRNA, eEFSec and a range of CTSBP2 or FL SBP2



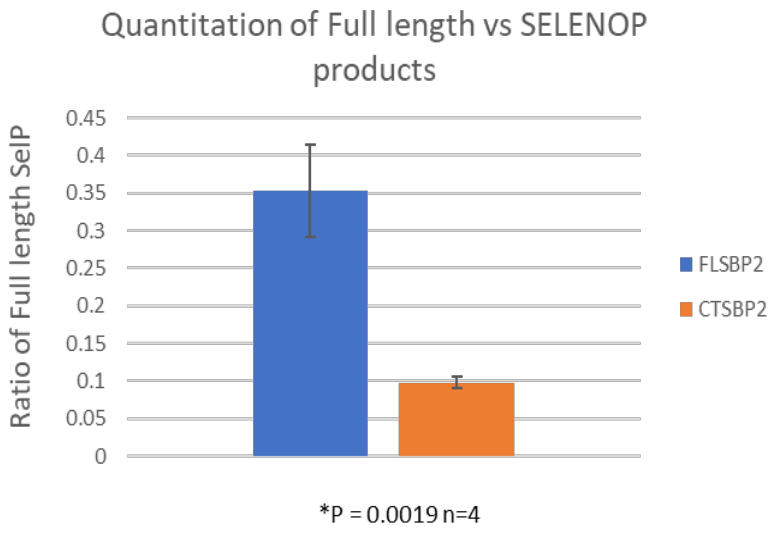


Figure 4-5. Full length SBP2 Increases full Length SELENOP synthesis compared to C-terminal SBP2 only. The top gel image was a reaction of SELENOP mRNA translated in WGL supplemented with RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$, eEFSec, and the addition of either recombinant FL SBP2 in the following amounts 0.25, 0.5, or 1 pmol. SELENOP mRNA was also translated in a control RRL reaction supplemented with CTSBP2 and RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$. Reactions were loaded onto a 15% SDS-PAGE gel for autoradiography analysis. Arrows indicate SELENOP products from early termination at the 2nd UGA based and full length SELENOP. Gel band quantitation of lanes marked with * (0.5 pmol of CTSBP2 and FLSBP2) were performed in quadruplicate and quantified by densitometry using a ratio of full length product over total Selenoprotein P (bottom). Error bars and P value were calculated from standard deviation and a student's t-test calculated from four replicates.

stability of SBP2 for processive Sec incorporation. This hypothesis is supported by previously published work which saw a 50 to 100-fold increase in Sec incorporation from *in vitro* translated CT-SBP2 (Bubenik et al., 2014). The findings by Bubenik et al., led us to wonder if there was set of factors involved in SBP2 folding or a separate unknown factor involved in processive Sec incorporation contained in the RSW. In this case, we wanted to see if FLSBP2 processive Sec incorporation activity could be enhanced by the addition of the ribosomal salt wash in WGL. We therefore wanted to see the difference between CTSBP2 and FLSBP2 in WGL with and without added RSW from RRL for SELENOP synthesis. As expected, we observed a dose dependent increase in processivity with the addition of FLSBP2 instead of CTSBP2 (Figure 4-6, compare lanes 1-3 with lanes 6-8). With the further addition of RSW from RRL we see less early termination products with FLSBP2 than CTSBP2 (Figure 4-6, lanes 5 and 9). Indicating that FLSBP2 has its own role in processive Sec incorporation, but mammalian RSW can further enhance this effect.

In order to further independently test the role of the N-terminus of FL-SBP2 on processivity we wanted to use reticulocyte lysate using a zebrafish SELENOP mRNA that were previously not fully processive *in vitro* from Shetty, et al., 2018. This allowed us to evaluate SBP2 CT vs FL *in vitro* with all other Sec incorporation factors controlled for in the context of processivity. Zebrafish SELENOP adds another seven in frame UGAs for a total of seventeen, and in order to control for variables such as magnesium concentration we used a defined reticulocyte lysate to keep magnesium levels at 2mM. Our lab has previously shown both magnesium and SECIS element are two important

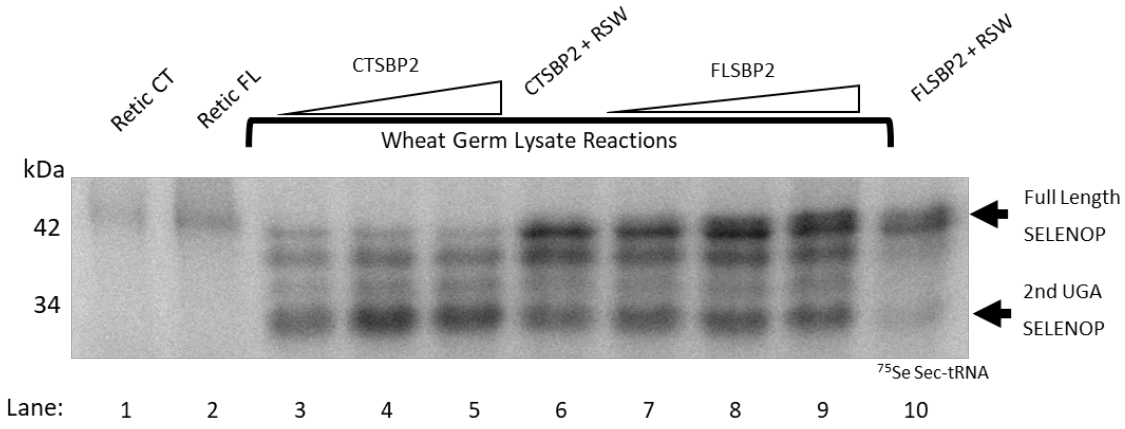


Figure 4-6. Full length SBP2 processive Sec incorporation is enhanced with the addition of RRL RSW. SELENOP mRNA translated in WGL supplemented with RCP-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$, eEFSec, and the addition of either FLSBP2 or CTSBP2 in the following amounts 0.25, 0.5, or 1 pmol. SELENOP mRNA was also translated in a control RRL reaction supplemented with CTSBP2 or FLSBP2 and RCP-5 75Se Sec-tRNA$^{[Ser]Sec}$. Additionally, another reaction of 1pmol of CTSBP2 or FLSBP2 was supplemented with RRL RSW. Reactions were loaded onto a 15% SDS-PAGE gel for autoradiography analysis. Arrows indicate SELENOP products from early termination at the 2nd UGA based and full length SELENOP.

factors for processive Sec incorporation (Shetty, et al., 2018). We compared three mRNAs, which used the wild type Zebrafish SELENOP (ZFSELENOP) coding region with different 3' UTRS: a wild type zebrafish full 3' UTR, Zebrafish SELENOP SECIS 1 alone, and Human SELENOP SECIS 1 (Figure 4-7, top). The wild type ZFSELENOP showed no significant difference in processivity between CTSBP2 and FLSBP2 both of which showed primarily an early termination product corresponding to 2nd UGA termination and the full-length product (Figure 4-7, bottom, lanes 2 and 3). Changing the construct to a single SECIS element caused a significant defect in processive Sec incorporation with CTSBP2 (lanes 5, 8), but not FLSBP2 (lanes 6 and 8).

We were also interested to see if there were any differences in single Sec incorporation between CTSBP2 vs FLSBP2. We used a luciferase assay using an mRNA containing the luciferase protein with an in frame UGA at UGC at position 258 and the 3' UTR of rat SELENOP as described before (Shetty et al., 2014; Gupta and Copeland, 2007) to evaluate any changes in single Sec incorporation (Figure 4-8). Interestingly, we find a difference between CT-SBP2 and FL-SBP2 but only in terms of the difference of read through activity upon halving the SECIS binding protein concentrations. In our preliminary results we find a differential in UGA readthrough depending on SECIS 1 and SECIS 2 deletion in the 3' UTR of our Luciferase^{C258U} construct (Figure 4-9). CTSBP2 appears to have more activity with SECIS 1 deleted, while with SECIS 2 FLSBP2 appeared to have higher Sec incorporation activity in the luciferase construct.

Discussion

The purpose of this chapter was to test the necessity of the N-terminal domain of full length SBP2 and/or SBP2L being a necessary factor of processive Sec incorporation. The reason we wanted to look at these two factors was based on our previous results with RSW from retic lysate increasing processivity in WGL and the synthesis of full length SELENOP in RRL without supplemented CTSBP2. This meant that SBP2 or SBP2L were the only possible endogenous factors for selenocysteine incorporation that could be in retic RSW. HSP90 seems to be able to interact with some recombinant proteins from mammals, but other reports have shown proteins unable to interact with wheat HSP90 (Pongratz et al., 1991; Reddy et al., 1998). It seems plausible that the HSP90 system in WGL would not interact with SBP2 or SPB2L as they are not native proteins to the kingdom. The "activation" by proper folding of SBP2 is further reiterated when CTSBP2 is in vitro translated in RRL is found to have several fold higher activities compared to recombinant E. coli sources (Bubenik et al., 2014). In the aforementioned study in vitro translated CTSBP2 was used, but it is possible that the increase would be applicable to the full-length protein as well. The increase in processivity with the addition of RSW to FLSBP2 could be due to HSP90 chaperone activity increasing recombinant FLSBP2 activity. Alternatively, the increase processivity with FLSBP2 could be caused by endogenous FLSBP2 from retic RSW outcompetes the recombinant protein. This seems unlikely as CTSBP2 and RSW does not see the same loss of truncation products in SELENOP compared to FLSBP2 and RSW. We suspected that the N-terminus would likely be the cause of processivity in Wheat Germ Lysate. Initially we tested for

processive Sec incorporation for both FLSBP2 and with CTSBP2 in a WGL assay and observed a slight increase in full length synthesis of SELENOP. Meanwhile, the addition of FLSBP2L seemed to interfere with Sec incorporation. Competitive binding of CTSBP2 and SBP2L to the SECIS element and eEFSec could be the cause of low amounts of translation products. This also highlights the importance of keeping both proteins separate in the cell to avoid competition. The role of SBP2 in processive Sec incorporation is solidified after it shows it can increase processive Sec incorporation alone, albeit at lower than optimized levels due to limitations of SBP2 expression in bacteria. Overall, these experiments indicated that the N-terminal domain of SBP2 significantly increases processive Sec incorporation in SELENOP vs only the C-terminal domain.

The addition of recombinant FLSBP2 did not appear to increase in processivity to the same extent as ribosomal salt wash in RRL. The reason for this could be either the possibility of another factor involved or the lack of proper folding of recombinant SBP2 to provide the same activity as the native protein. SBP2 and other L7ae proteins are known to interact with the HSP90 complex, and inhibition of HSP90 in cells are reported to reduce the amount of transfected SBP2 protein expression (Boulon et al., 2008). The HSP90 system is present in RRL and a similar system is also found in WGL (Schumacher et al., 1994; Owens-Grillo et al., 1996; Reddy et al., 1998; Boulon et al., 2008). We switched our assay system to SBP2 and RRL in order to evaluate what effect CTSBP2 and FLSBP2 could have in a lysate native for Sec incorporation. This was necessary in order to see what effect the N-terminal domain would have with all the mammalian factors

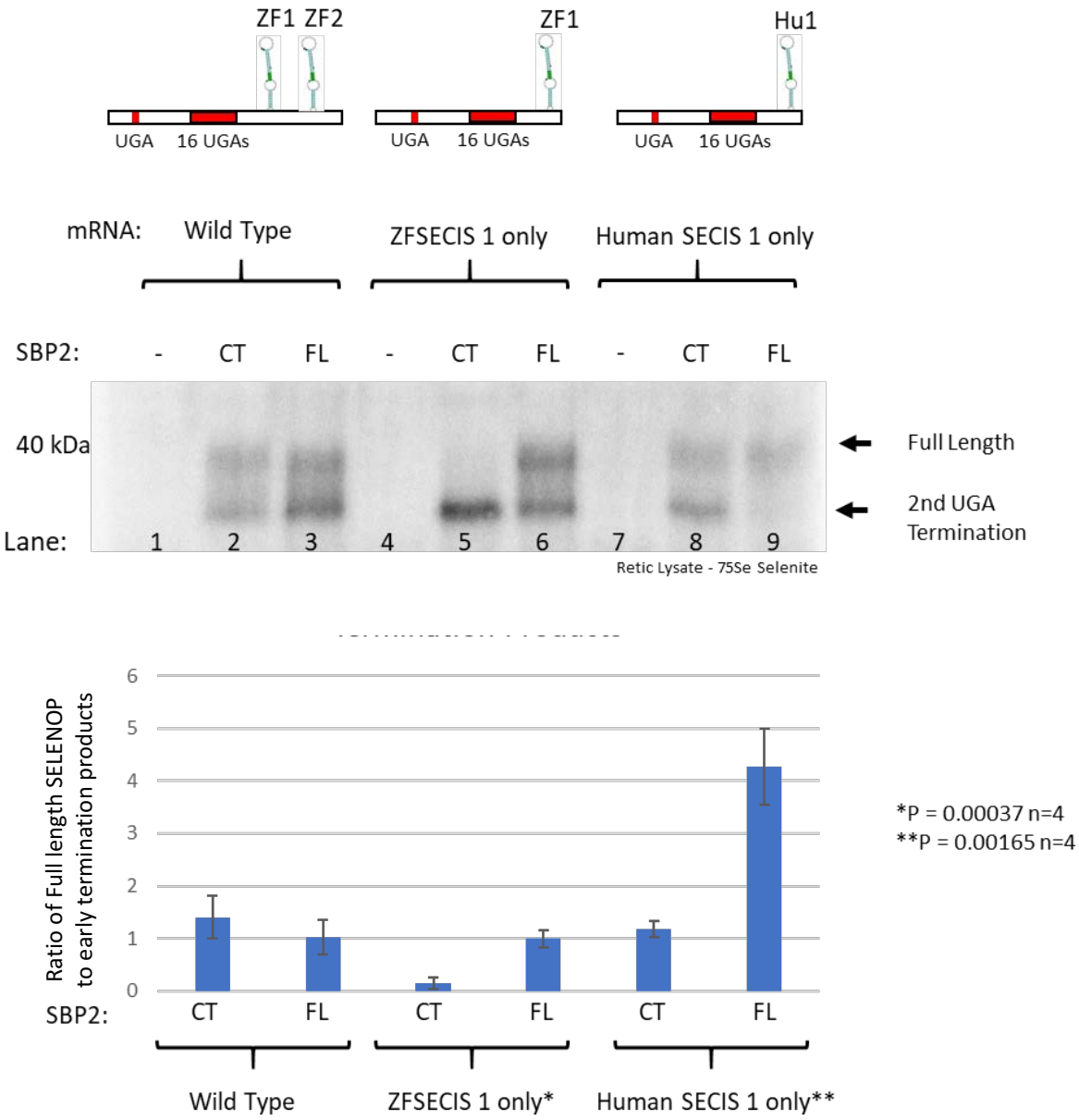


Figure 4-7. Difference between FLSBP2 and CTSBP2 processivity depends on *cis* factors in the 3' UTR. Zebrafish SELENOP mRNA with the wild type 3' UTR or mutant 3' UTR with a spacer and Zebrafish SECIS 1 (ZF1) or Human SEICS 1 (Hu1) were translated in RRL. A diagram of each mRNA is depicted (top). RRL was supplemented with 400nM ^{75}Se and FLSBP2 or CTSBP2 (1 pmol). Reactions were loaded onto a 15% SDS-PAGE gel for autoradiography analysis (middle). Arrows indicate SELENOP products from early termination at the 2nd UGA based and full length SELENOP. The gel bands were quantitated by densitometry using a ratio of full length product over early termination products. P value represents standard deviation calculated from four replicates, * and ** represent a significant difference.

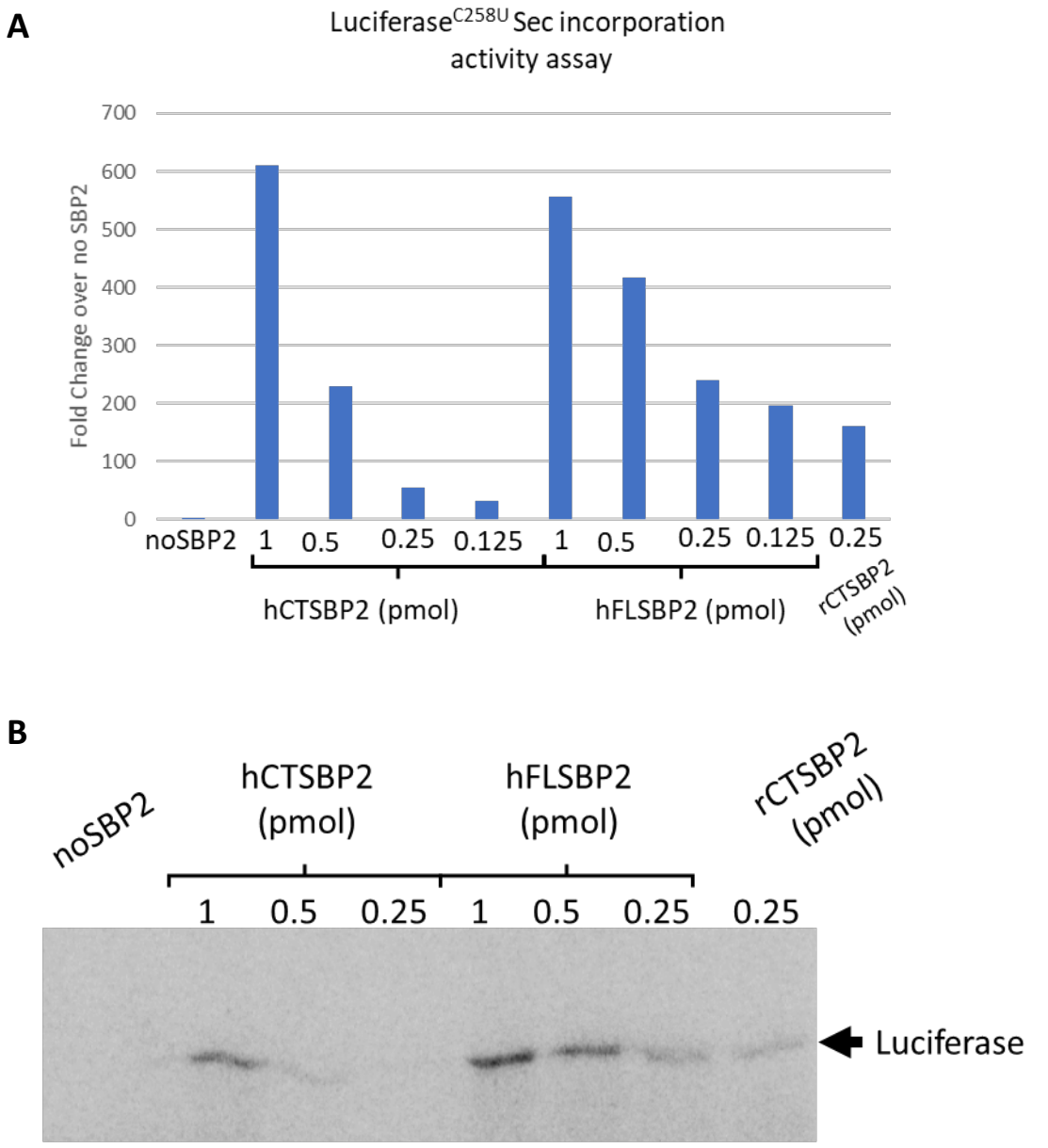


Figure 4-8. CTSBP2 and FLSBP2 translate Luciferase mRNA at different efficiencies at low concentrations in RRL. A. Luciferase^{C258U} mRNA with a wild type 3' UTR of SELENOP was translated in RRL supplemented with 400nM Se, and a range of CTSBP2 or FLSBP2 from human or CTSBP2 from rat. The reaction was then read on a luminometer and fold change was normalized to the noSBP2 reaction. **B.** Luciferase^{C258U} mRNA with a wild type 3' UTR of SELENOP was translated in RRL supplemented with 400nM ^{75}Se, and a range of CTSBP2 or FLSBP2 from human or CTSBP2 from rat. After the reactions were complete the reactions were ran on a 12% SDS-PAGE gel for autoradiography analysis.

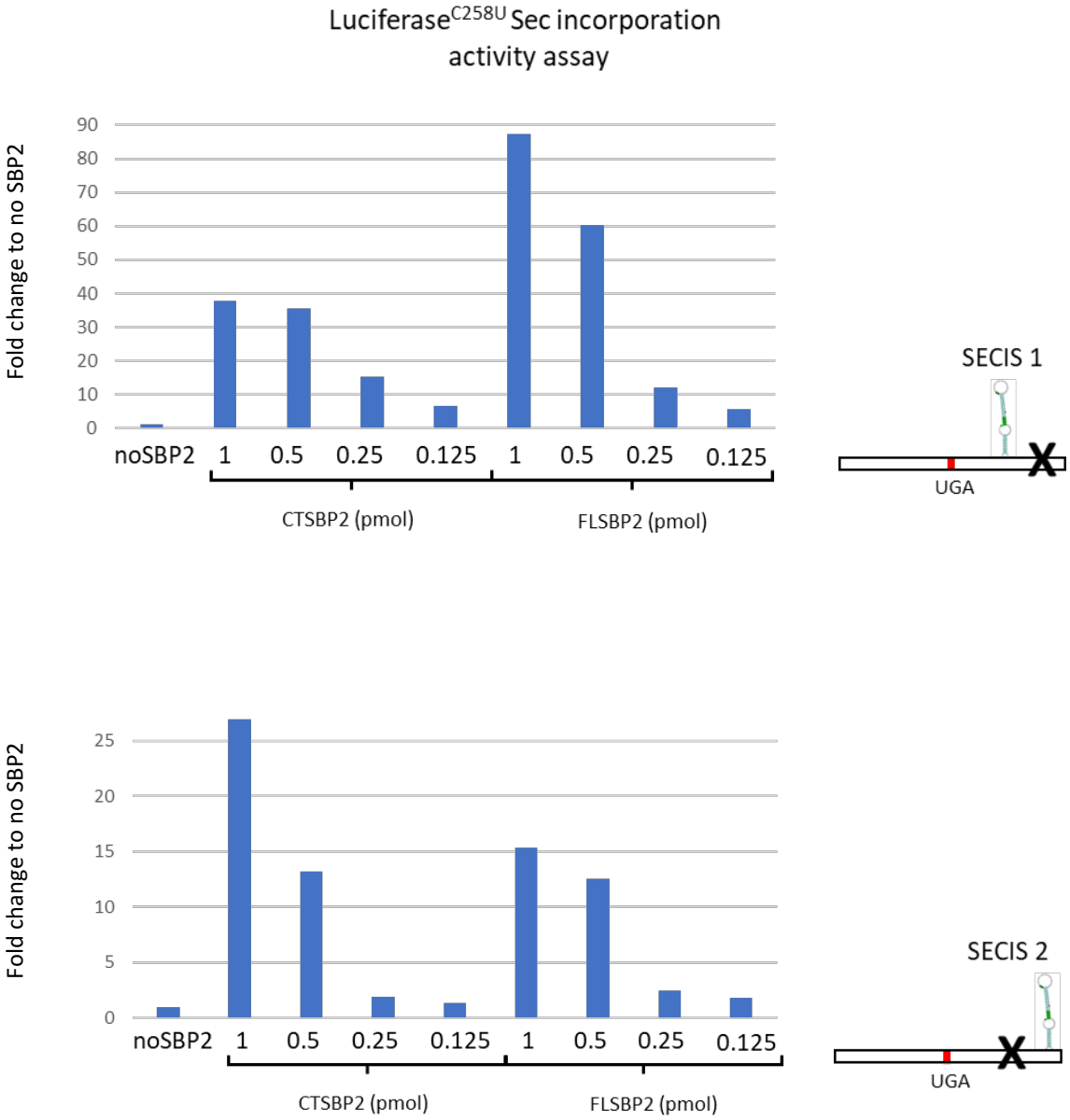


Figure 4-9. C-terminal and full length SBP2 have different efficiencies depending on the SECIS element. Luciferase^{C258U} mRNA with the 3' UTR of SELENOP with either SECIS 2 deleted (top) or SECIS 1 deleted (bottom), which is depicted to the right each graph. The RRL supplemented with 400nM Se, and a range of CTSBP2 or FLSBP2 from human. The reaction was then read on a luminometer and fold change was normalized to the no SBP2 reaction.

involved in Sec incorporation. We found we do not see a significant difference in processive Sec incorporation with the wild type construct of ZFSELENOP, however, but we see slight differences in overall efficiency. We saw a significant difference in CT and FLSBP2 in the context of a single SECIS element without the full downstream 3' UTR. Previous *in vitro* experiments using the same constructs with larger quantities (8pmol) of rat-CTSBP2 observed similar defects in processivity (Shetty et al., 2018). Interestingly, the addition of FLSBP2 with human SECIS 1 restores nearly all full-length synthesis of SELENOP, which was not observed with CTSBP2 nor zebrafish SECIS 1. The luciferase assays would need more replicates to solidify the effect we see. Currently, it seems like FLSBP2 has a higher Sec incorporation activity with SECIS 1, while CTSBP2 has a higher Sec incorporation activity in SECIS 2. This could have been one of the reasons we see more full length SELENOP when we had only SECIS 1 in the 3' UTRs. Again, this highlights the importance of the SECIS element driving efficiency and processivity in a SBP2 driven manner. The SECIS type driving different efficiencies based on CTSBP2 or FLSBP2 would be an interesting finding, and supports an idea of the SBP2 RBD diverged in evolution along with the SECIS elements.

Chapter 5: Other Data and Final Conclusions

Some of the data used for the Luciferase assay of eEFSec mutants appears in Nature Communications:

Dobosz-Bartoszek, M., Pinkerton, M. H., Otwinowski, Z., Chakravarthy, S., Soll, D., Copeland, P. R., & Simonovic, M. (2016). Crystal structures of the human elongation factor eEFSec suggest a non-canonical mechanism for selenocysteine incorporation. Nature Communications, 7, 12941.

Analysis of Sec Incorporation Activity from the Perspective of eEFSec Mutant GTP, Sec-tRNA$^{[Ser]Sec}$ Binding Pocket, and SerRS.

Introduction

Selenocysteine incorporation can be considered a specialized elongation cycle to accommodate Sec-tRNA$^{[Ser]Sec}$ in the presences of a UGA codon utilizing a specialized elongation factor eEFSec. The general translation elongation factors eEF1A (in eukaryotes) and EF-Tu (in bacteria), supply the translating ribosome during the elongation stage of protein synthesis with cognate aminoacyl-tRNAs (aa-tRNAs) to the ribosomal A site when there is a match of a codon anticodon. Both eEF1A and EF-Tu share many similarities to their Sec specific counterparts eEFSec and SelB respectively. Much of what was known about eEFSec was based on SelB research, the bacterial equivalent of eEFSec, because of their shared sequence homology. It was because of this similarity that archaeal SelB from *Methanococcus jannaschii* was used to find

eEFSec in invertebrates up to mammals by two different groups (Fagegaliter et al., 2000; Tujebajeva et al., 2000). The primary difference between prokaryotic and eukaryotic Sec incorporation is SelB performs the role of SBP2 and eEFSec of SECIS binding and Sec-tRNA$^{[Ser]Sec}$ delivery to the ribosome A site. eEFSec, as discussed earlier, is a translational GTP binding protein, which shares similarity to the other primary eukaryotic elongation factor eEF1a that specifically binds Sec-tRNA$^{[Ser]Sec}$ with high affinity and plays a pivotal role during decoding of UGA by taking the place of eIF1a for Sec incorporation. eEFSec was found to have similar binding affinities to GTP and GDP, implicating that eEFSec has its own GEF activity (Fagegaltier et al., 2000; Tujebajeva et al., 2000). It is also shown that Domain IV regulates nucleotide binding and regulates the GTPase activity of the mammalian eEFSec (Gonzalez-Flores et al., 2012). Utilizing a hybrid of animal and plant lysates were able to identify potential functional motifs of eEFSec domain IV via bioinformatics. To test the functionality of the potential motifs a hybrid system of SF21 cell free lysates to supplement WGL for Sec incorporation was able to test via penta-alanine mutants using a critical was first identified by (Gonzalez-Flores et al., 2012). To further our understanding of the protein, our collaborators from the Simonović lab were able to determine the crystal structures of the full human eEFSec in complex with GDP and non-hydrolyzable GTP analogues. They were also able to show that the four domains of human eEFSec fold into a chalice-like structure resembling the archaeal SelB, IF2, and eIF5B (Dobosz-Bartoszek et al., 2016). Surprisingly, they found that unlike in eEF1A and EF-Tu, guanine nucleotide exchange does not cause a major conformational change in Domain I of eEFSec, but instead

induces a swing of Domain IV. Domain IV is found to be unique to eEFSec, while Domains I-III share sequence similarity to eEF1A based on bioinformatics. The primary goal of the following experiments was to determine the critical amino acids necessary for eEFSec binding to GTP and Sec-tRNA$^{[Ser]Sec}$ based on structural data. We also wanted to see if single amino acid changes to the Sec-tRNA$^{[Ser]Sec}$ binding pocket would reduce stringency of Sec-tRNA$^{[Ser]Sec}$ specificity.

Methods

Constructs and Recombinant proteins: Human eEFSec and mutant constructs, SerRS, and *in vitro* tRNA was generously provided by the Simonović lab and was purified as described in Dobosz-Bartoszek, 2016. The proteins were quantified and normalized based on coomassie gel band quantitation to an ovalbumin standard. All eEFSec proteins were diluted to 8µM in 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 5 mM $MgCl_2$.

In vitro Luciferase activity assay: The Sec incorporation activity of recombinant WT and eEFSec mutants was determined using an *in vitro* wheat germ lysate translation system with a luciferase reporter mRNA, containing an in-frame UGA codon at position 258 followed by rat SELENOP SECIS element in the 3' UTR. Each 12.5 µl reaction contained 6.25 µl wheat germ extract, 320 nM recombinant SBP2, 320 nM wild-type or mutant His-tagged eEFSec, 20 mM amino acid mix, 20mM additional potassium acetate, 125 ng of luciferase mRNA and 1.25 µg total aminoacyl-tRNA pool from rat testes (a rich source of selenium). The reactions were incubated at 25°C for 2 h and then brought up the

final volume 100µl with 1xPBS before measurement on a 96-well plate luminometer (Berthold Tristar).

Testing of SerRS and *in vitro* tRNA: Accommodation of SerRS and *in vitro* tRNA in our luciferase assay was conducted in the same way as in our eEFSec mutant assay except no added testis tRNA was added (except for the positive control). 320 nM recombinant SerRS (calculated by gel quantitation) was added to each indicated reaction along with 75ng of total *in vitro* transcribed tRNA.

Results and Discussion

Our collaborators in Simonović lab managed to successfully crystalize eEFSec, which revealed several important amino acids belonging to the GTP and the Sec-tRNA$^{[Ser]Sec}$ binding pockets (Dobosz-Bartoszek et al., 2016). We evaluated the effects of point mutations to the key amino acids of each of these binding pockets. In the GTPase site the key identified amino acids were T48, D92, and H96 (Dobosz-Bartoszek et al., 2016). Mutations to each of these amino acids caused a significant reduction in UGA readthrough in our WGL luciferase Sec incorporation assay (Figure 5-1 A). Similarly, the amino acids found to be in important positions of the Sec binding pocket of eEFSec were: D229, H230, and R285 (Dobosz-Bartoszek et al., 2016). In order to test the results of the crystal structure data, mutants were tested for each amino acid in our WGL luciferase assay. We found that mutations to these amino acids also caused significant defects in Sec incorporation. Even the R285N mutation, which corresponds to the equivalent to EF-Tu in *T. aquaticus* was unable to read through UGA in our assay (Figure

5-1 A). As a control we tested eEFSec preincubated and stored in 1mM GTP vs the APO form for Sec incorporation activity in our luciferase assay. We found no significant

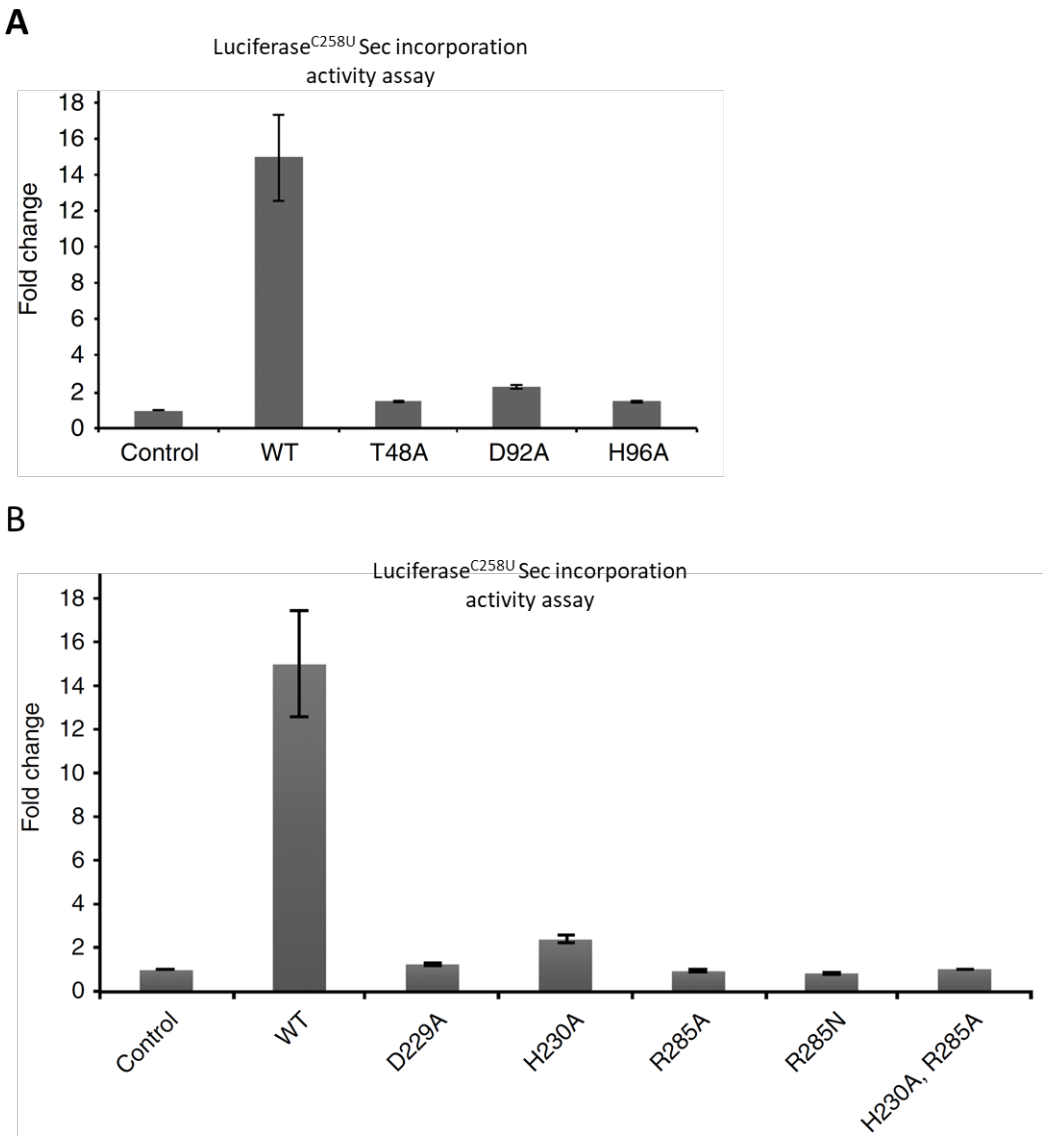


Figure 5-1. Evaluation of critical amino acids of the GTPase and putative Sec-tRNA binding pocket in human eEFSec based on structural data in a wheat germ lysate assay. A. Luciferase^{C258U} mRNA was translated in WGL supplemented with testis total tRNA, CTSBP2 and wild type eEFSec, a no eEFSec control, or an eEFSec mutant to the GTPase site (Thr48, Asp92 and His96 to Ala) **B.** The same WGL reaction was conducted as in A, but eEFSec mutants were in the Sec-tRNA$^{[Ser]Sec}$ binding pocket instead (Asp229, His230, and Arg285 to Ala). Luciferase activity was evaluated on a luciferase activity on a luminometer in both **A.** and **B.** All mutants demonstrated a significant difference in Sec incorporation activity and error bars for both **A.** and **B.** represent standard deviation calculated from three replicates. Fold change was calculated to a no eEFSec control in each replicate.

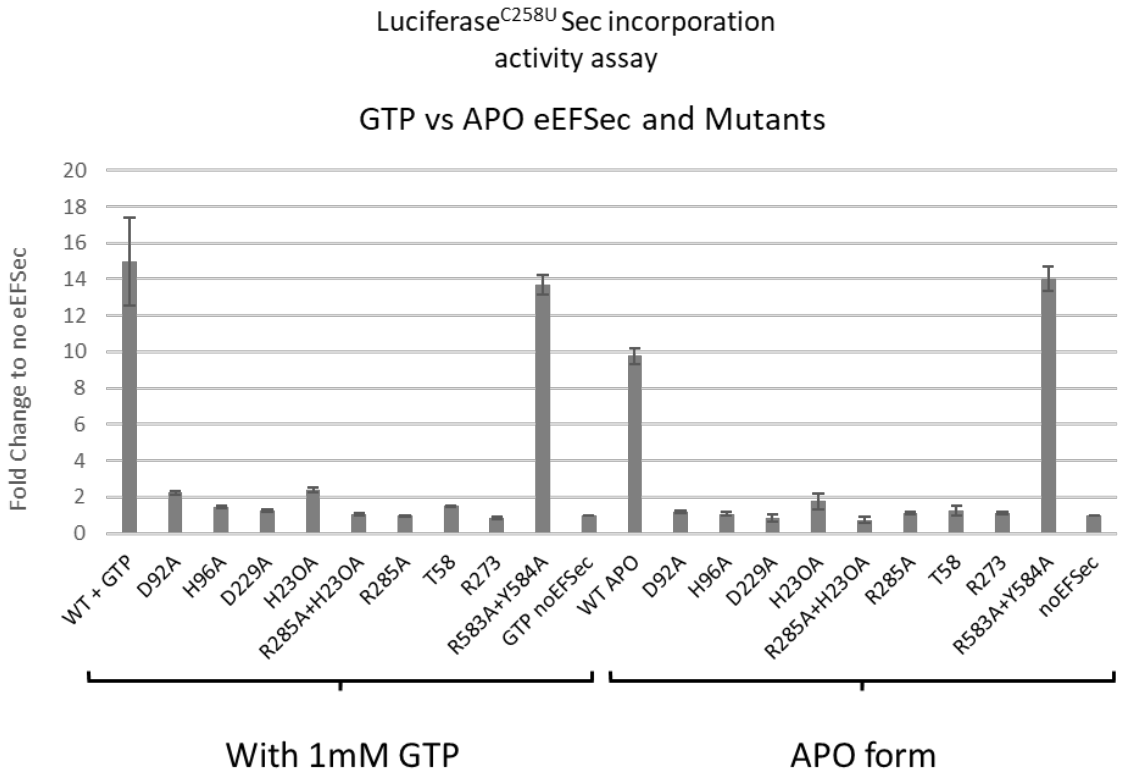


Figure 5-2. Preincubation and storage of eEFSec and mutants with GTP does not significantly impact Sec incorporation. Luciferase^{C258U} mRNA was translated in WGL supplemented with testis total tRNA, and CTSBP2. WT eEFSec or eEFSec mutants listed above with and without preincubation and storage of eEFSec in 1mM of GTP or in APO form. All mutants demonstrated a significant difference in Sec incorporation activity and error bars represent standard deviation calculated from three replicates. Luciferase activity was evaluated on a luciferase activity on a luminometer. Fold change was calculated to a no eEFSec control in each replicate.

difference between the two, likely due to WGL system providing enough GTP for the added eEFSec (Figure 5-2).

The next part of eEFSec we wanted to test was the ^{582}KRYVF^{586} motif, a motif in Domain IV which was previously identified to be important for Sec incorporation and Sec-tRNA$^{[Ser]Sec}$ binding (Gonzalez-Flores et al., 2012). Structural data provided by Dobosz-Bartoszek et al., 2016 indicated this is the hinge region for eEFSec and that R583 and Y584 could be important to this region of the motif. We evaluated a double mutant of these amino acids in eEFSec for Sec incorporation in a luciferase assay (R583A/Y584A). We only observed a slight decrease in activity in the R583A/Y584A (Figure 5-2 and 5-3). Previous assays of KRYVF motif mutations showed Sec incorporation activity similar to a no eEFSec control but used a full penta-alanine mutant of that motif (Gonzalez-Flores et al., 2012). The retention Sec incorporation activity in the mutant eEFSec$^{R583A/Y584A}$ is likely due to structural data indicating the structure has not changed(Dobosz-Bartoszek et al., 2016). Furthermore, the crystal structure suggest Lys582 can form a salt bridge to Glu372, which is also conserved archaeal SelB (Dobosz-Bartoszek et al., 2016). The next step would be to make a point mutation of Lys582 for evaluation of structure, Sec-tRNA$^{[Ser]Sec}$ binding, and Sec incorporation activity.

We also wanted to look at tRNA binding specificity in eEFSec and if mutants to the tRNA binding domain could accommodate serine charged Sec-tRNA$^{[Ser]Sec}$. We hypothesized that the Sec-tRNA$^{[Ser]Sec}$ would be accommodated into the tRNA binding pocket of eEFSec in the H230A mutant. This is because the crystal structural data

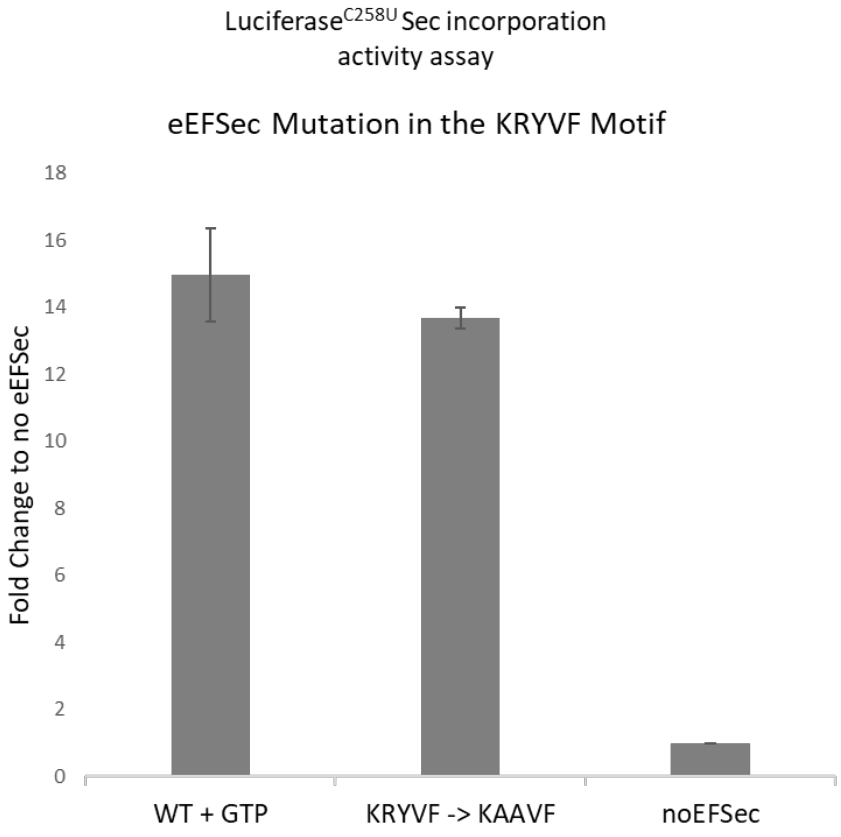


Figure 5-3. Mutations of the R583A/Y584A of the of the KRYVF hinge motif does not significantly affect Sec incorporation *in vitro*. Luciferase^{C258U} mRNA was translated in WGL supplemented with testis total tRNA, and CTSBP2. A WT eEFSec or the mutant eEFSec$^{R583A/Y584A}$ was evaluated for luciferase activity and compared to a no eEFSec control. Luciferase activity was measured on a luminometer. Error bars represent standard deviation of 3 replicates.

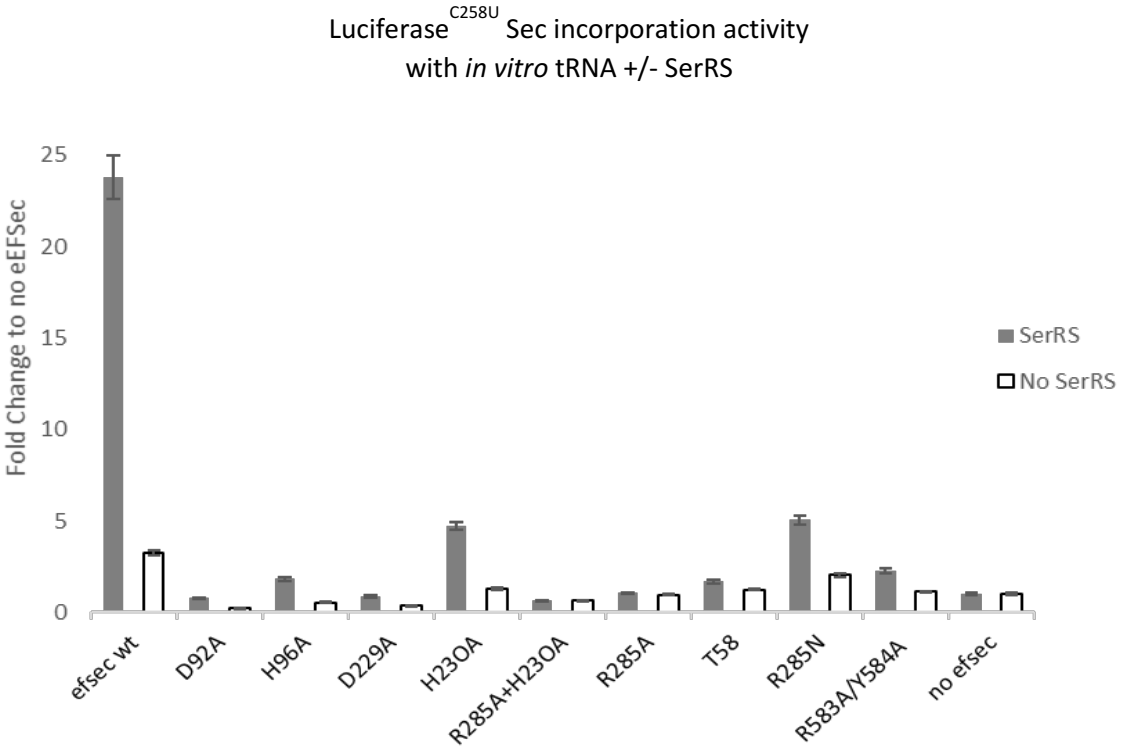


Figure 5-4. Serylation of *in vitro* transcribed Sec-tRNA Sec incorporation in wheat germ lysate. Luciferase^{C258U} mRNA was translated in WGL supplemented with *in vitro* transcribed tRNA$^{[Ser]Sec}$, and CTSBP2 with or without SerRS. A WT eEFSec or a mutant eEFSec was also added to the WGL reactions. Luciferase activity was measured on a luminometer and compared to a no eEFSec control. Error bars represent standard deviation of three replicates.

Predicts the positive charge of histidine to provide specificity in attracting selenocysteine. In order to test this hypothesis, we added purified human SerRS and *in vitro* transcribed tRNA, which generously provided by the Simonivić lab as well, and added it into our WGL luciferase assay using different eEFSec mutants. No aminoacylated Sec-tRNA$^{[Ser]Sec}$ was added to the reaction and because WGL lacks any endogenous Sec aminoacylation components so the *in vitro* tRNA$^{[Ser]Sec}$ added must be via serylation by SerRS to give Ser-tRNA$^{[Ser]Sec}$. We were surprised to find that wild type eEFSec was the most accommodating for Sec incorporation for *in vitro* tRNA, while the H230A and R285N (EF-Tu equivalent) were distant runners up in activity. The explanation for this result is still being investigated, however it could be possible that eEFSec is not as specific for Sec aminoacylated Sec-tRNA$^{[Ser]Sec}$ as originally thought. Alternatively, it could be possible another factor is involved in specificity for Sec-tRNA$^{[Ser]Sec}$ eEFSec binding.

Further investigation into these results could be utilized to assay Sec incorporation without the use ^{75}Se. This would be a great service to the field of study as ^{75}Se is expensive and can be difficult to obtain. Currently, efforts have been focused on attempting to reconstitute the Sec aminoacylation pathway for Sec-tRNA$^{[Ser]Sec}$ synthesis for *in vitro* assays as well, which would greatly help further *in vitro* research in the field.

Regulation of SBP2 under Stress Conditions

Introduction

Enterovirus 71 (EV71) is global public health issue as the primary cause of hand, foot, and mouth disease, with high prevalence in the Asia-Pacific region. A typical infection resolves itself and includes fever, sore throat, and sores in the hands, feet, and mouth. Currently there are no effective therapies nor vaccines to treat the virus, however vaccines are in development with one in Phase IV clinical trials (Tung et al., 2007; Zhu et al., 2014; Guan, 2019). Depending on the viral strain, infection can be serious, and outbreaks have resulted in CNS related cardiopulmonary failure and delayed neurodevelopment in young children (Reviewed in Chang et al., 2016). EV71 is a member of genus *Enterovirus*, family Picornaviridae, which is a diverse group of small RNA viruses. Enteroviruses contain a positive-strand RNA genome of ~7,400 nt that encodes a large polyprotein. The polyprotein becomes processed by 2A and 3C proteases into 4 structural proteins VP1-4 and 7 non-structural proteins, 2A-2C and 3A-3D (Wimmer et al., 1993; Chase et al., 2012; Bedard et al., 2004). Many Picornavirus proteins cause potent biological effects on host cells to maximize viral propagation and limit host immunity (Chase et al., 2012). In EV71, 3C protease can induce apoptosis of human glioblastoma cells via caspase activation (Li et al., 2002). In addition, 3C cleaves CstF-64, a pre-mRNA processing and polyadenylation factor. An analysis of 3C cleavage revealed that target proteins can be classified into at least nine different functional groups (Zhang et al., 2014).

Another enterovirus included in this family is the coxsackie virus, which has established itself as a risk factor for selenium deficiency associated cardiomyopathies (reviewed in Levander, 1999). HIV-1 infections also observe a strong correlation with selenium deficiency and disease progression and mortality (reviewed in Stone et al., 2010). Oxidative stress is well documented to be associated with viral infection, it is likely that the antioxidant function of selenoproteins is important in preventing ROS-induced damage that accompanies viral infection (Reshi et al., 2014). While there is no reported association between EV71 and selenium deficiencies, more recent work has shown a direct correlation with increased ROS and EV71 replication efficiency (Ho et al., 2008; Cheng et al., 2014). The effect of other viruses causing oxidative stress led us to explore if EV71 could affect the regulation of selenoproteins and the selenoprotein synthesis machinery. We initially focused around SBP2 expression and binding activity as a result of viral infection based on previous reports on ROS governing sensitivity and localization of SBP2. We were also interested in exploring a potential mechanism involved viral replication and SBP2 expression in cells and how it could affect selenoprotein expression.

Materials and Methods

Cell Culture and Infection with EV71, Transfection EV71 3C protease, and Cell Lysis: All cell lines were maintained at 37°C under a humidified 5% CO2 atmosphere. HEPG2 cells were cultured in DMEM (Gibco) media containing 10% FBS and supplemented with 50 nM of sodium selenite. HEK293 cells were cultured in EMEM media (Gibco) containing 10% FBS and supplemented with 50 nM of sodium selenite. HAP1 cells (Horizon

Genomics) were grown in Iscove's Modified Eagle Medium (IMEM; Gibco). SF268 cells and EV71 virus Strain TW2231 at a concentration of 10^8 cfu were a gift from the Brewer Lab. SF268 cells were grown in RPMI medium supplemented with 10% fetal bovine serum. Cells were infected at approximately 90% confluence with EV71 virus, which was diluted 5-fold or 10-fold in RPMI media with 2.5% FBS to infect SF268 cells for 1 to 2 hours at 37°C before washing unbound virus (described in Lin et al., 2015). 3C protease transfected cells were transfected with lipofectamine either with wild type or C147S, a catalytically inactive mutant EV71 3C. Cells were harvested by lysis with 1xPBS and 0.1% tween and passing through a 21g needle, then spun at 17,000 x g for 15 min at 4°C. The pellet was discarded, and the supernatants were quantified by BCA assay.

Etoposide treatment of Cells: SF268 cells were treated with 25, 50, or 100 ug/mL of Etoposide for up to 72 hours, control cells were treated with the equivalent volumes of DMSO. Cells were harvested by standard lysis with 1xPBS and 0.1% tween.

UV Cross linking to SELENOV SECIS RNA for SBP2 Detection: After cell lysates were quantitated via BCA assay and equal amounts of total lysate were incubated with 250 µg yeast tRNA (Sigma), 10 mM DTT, and 5 µg soybean trypsin inhibitor (Sigma) and 20 fmol [32P]-a-UTP labeled wild type or mutant SELENOV SECIS element that harbors a mutation in the conserved SECIS Core sequence required for SBP2 binding (Lesoon et al., 1997; Copeland and Driscoll, 1999). The wild type and mutant SELENOV SECIS elements were *in vitro* transcribed using the T7 RiboMAX™ system (Promega). The reaction was brought to a final volume of 20µl by 1x PBS and then incubated for 30min at 37°C. Following incubation, complexes were UV irradiated at 254 nm for 10 min and

subsequently treated with 20 µg RNase A for 15 min at 37°C. Samples were resolved by 10% SDS-PAGE and visualized by phosphorimaging on a GE Typhoon Phosphoimager.

Western Blot of Cells: Cell lysates were quantified by BCA assay (Pierce) and equal amounts of protein were loaded for each sample. Antibodies and dilutions used for analysis were: rabbit anti-SBP2 antibody 1:1000 dilution (Proteintech, 12798-1-AP); mouse anti β-actin (Sigma, A5316).

MG132 treatment of Cell Lines: HepG2 cells were seeded in 6 well plates at 2.5 x 10^6 cells and harvested every 24 hours after MG132 or DMSO treatment. Cells were treated with 10 µM of MG132 (Sigma) or DMSO (Sigma) for four hours prior to harvesting in NP40 lysis buffer (150 mM NaCl, 1% NP-40, 50 mM Tris pH 8.0) and EDTA-free protease inhibitor (Roche).

Results and Discussion

We were initially interested in SBP2 levels in response to an EV71 infection because it has been reported that EV71 increases cellular ROS levels (Cheng et al., 2014). Since high ROS levels cause SBP2 to be shuttled in the nucleus we wanted to use UV crosslinking to a wild type or mutant SELENOV SECIS to measure levels of SBP2 binding and check SBP2 protein levels by western blot. We made lysates from SF268 glioblastoma cells after 24 hours of treatment with either wild type EV71 virus infection (Strain TW2231), a mock infection, or transfection with the EV71 3C protease or a non-functional 3C mutant (C147S). Cell lysates were then UV crosslinked to SELENOV in our first analysis. We found infected cells and cells transfected with the 3C protease

showed a higher level of SECIS binding activity by SBP2 compared to the negative controls (Figure 5-5, compare Lanes 3,5 with Lanes 7, 9). We wanted to see at what time peak SBP2 SECIS binding activity occurred. The experiment was repeated as a time course of EV71 infection and lysates were collected 4, 20, 24, 28, and 48 hours post infection. We saw a relatively narrow timepoint of a significant change in SECIS binding to SELENOV at 20 and 28 hours post infection, with 24 giving the strongest signal (Figure 5-6, lanes 4,6,8). By 48 hours SBP2 seemed to be reduced to almost control levels of signal (Figure 5-6, lane 10). We then wanted to repeat the experiment by collecting at only 24 and 48 hour time points with a western blot to look at SBP2 and actin protein levels as well. We thought that the reduction of SBP2 signal at the 48-hour time point could be from cell death, so we wanted to make sure SBP2 was still being expressed. Once again in the crosslinking assay we saw little difference in SBP2 SECIS binding between control and infected cells by 48 hours (Figure 5-6 A., lane 8 and 10). We did a difference in total SBP2 levels between infected and uninfected cells even at the 24 hour timepoint (Figure 5-7 B., lanes 3 and 4). This meant the increase of SBP2 SECIS binding by EV71 and 3C is independent of the actual protein levels of SBP2. We also noticed in our western blot SBP2 appeared to be highly degraded at the 48 hour time point in both infected and control cells, which we investigated later (Figure 5-7 B., lane 5 and 6).

Further work needs to be conducted to understand this contradictory observation of increased SECIS binding activity with either no change or decreasing protein expression. The 3C protease seems like a main point of focus to

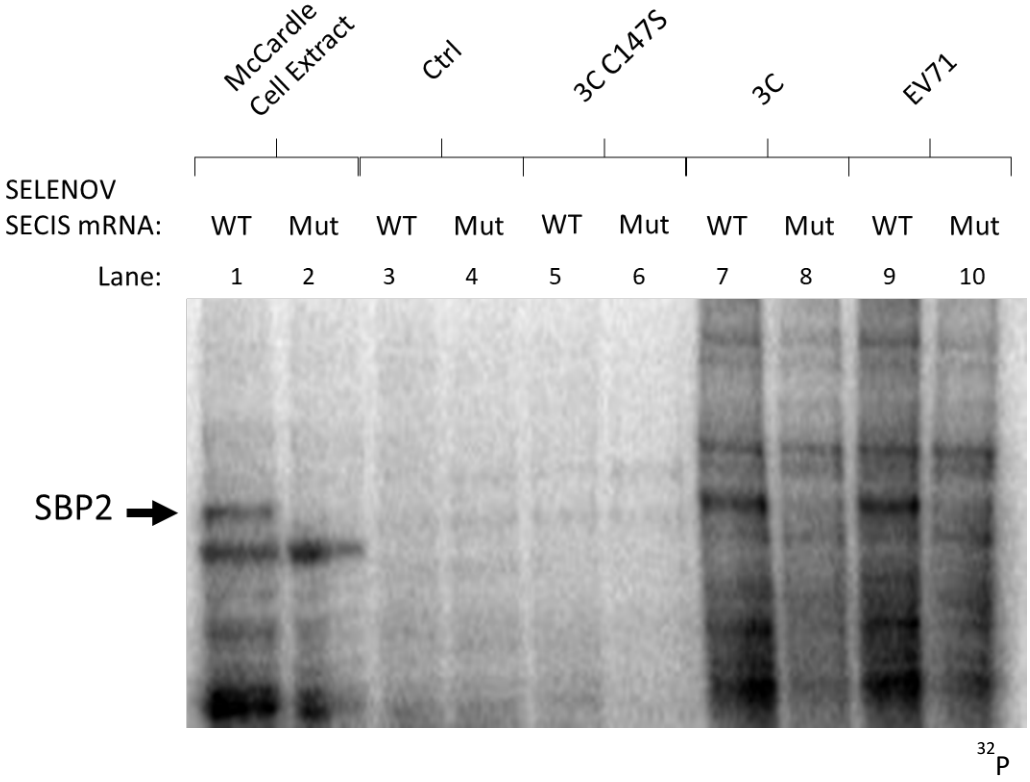


Figure 5-5. SBP2 binding is increased in SF268 cells infected with EV71 and 3C protease transfection. SF268 glioblastoma cells were infected with EV71 strain TW2231 or transfected with 2 µg of plasmids containing the 3C protease cDNA or a catalytically inactive C147S mutant for 24 hours. Cell lysates were incubated with radiolabeled with ^{32}P-UTP labeled wild type or mutant SelV SECIS RNA and subjected to UV crosslinking. Lanes 1 and 2 contain extracts from liver hepatoma cells, which serve as a positive control for SBP2 binding. SBP2 is marked and determined based on its relative size and the missing band in our mutant SECIS mRNA in our control extract.

understanding the mechanism and to gain insight into a possible hierarchy for selenoprotein expression. As ROS products are a toxic byproduct of viral parthenogenesis, it would be important to investigate if enhancing the anti-oxidant cellular properties could resist cell apoptosis. The critical piece of future research would 3C protease could be cleaving other SECIS binding proteins, such as L30, EIF4A3, or SBP2L which can potentially compete with SBP2 for SECIS binding (Budiman et al., 2009). 3C could also be regulating another factor involved in post translational modification of SBP2 increasing its binding activity.

Degradation of SBP2 in cell culture is Proteasome mediated

To see if the degradation products of SBP2 were created because of increase in confluence we grew HAP1 cells at a high confluence over 96 hours, with daily changes to their media. We observed was a significant loss of full length SBP2 by 96 hours with degradation products detectable by 48 hours as the cells grew to confluence (Figure 5-8). After observing cells that had grown to confluence having significant SBP2 degredation we wanted to test if other cells see the same phenomenon. We then grew HepG2 and HEK293 cells for 96 hours and ran a western blot probing for SBP2. Similar SBP2 degradation patterns were observed in these cells as well, meaning it was not a cell type specific phenomenon (Figure 5-9 A and B). The primary truncated SBP2 products observed were approximately at 70 kDa, 50 kDa, and 30 kDa but many more bands were observed as well. While SBP2 is reported to have multiple initiation and alternative splice sites, it would still not explain all the various sizes of SBP2 we observe (Papp et al., 2008). The molecular weights of the four isoforms at downstream

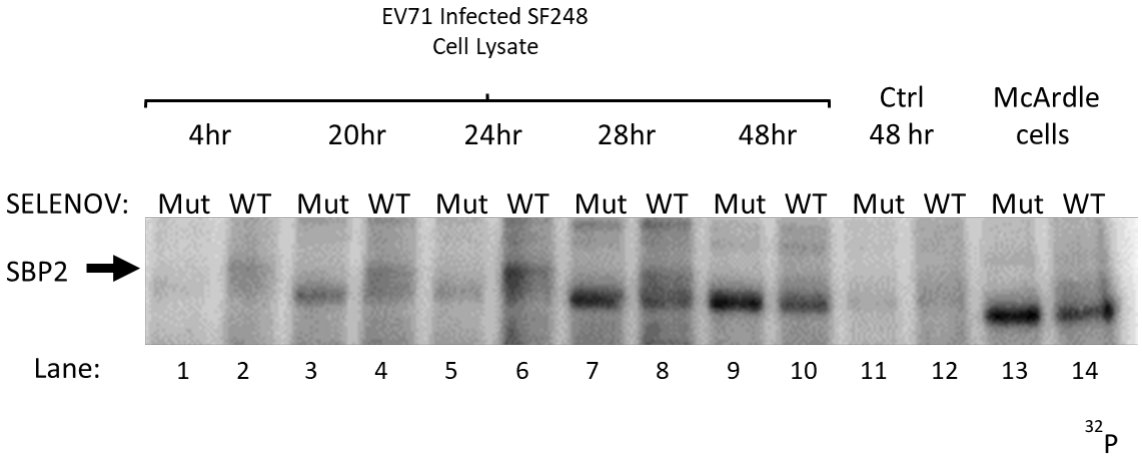


Figure 5-6. Time Course Evaluation of EV71 infection of SF248 cell SBP2 binding to SELENOV during infection. SF268 glioblastoma cells were infected with EV71 strain TW2231 or mock treated. Cell lysates were incubated with radiolabeled with ^{32}P-UTP labeled wild type or mutant SELENOV SECIS RNA and subjected to UV crosslinking. Cells were collected at the indicated timepoints. Lanes 13 and 14 contain extracts from liver hepatoma cells, which serve as a positive control for SBP2 binding. McArdle cells were used as a control for labeling.

methionine are predicted to be: 85.9 kDa (mtSBP2), 86 kDa, 69.2 kDa, and 62.4 kDa, which does not account for all the bands of SBP2 in our western blots. We therefore explored the possibility of SBP2 degradation being mediated by the proteasome. The reason for our hypothesis stemmed from SBP2 being lysine rich and predicted by UbiSite and UPRED web servers to contain multiple ubiquitination sites (Summarized in Figure 5-10, A). In order to test the hypothesis of proteasome mediated degradation, we used a 26S proteasomal inhibitor MG132 and added it to HepG2 cells after allowing them to grow to a high confluence. We saw a clear difference in MG132 treated cells in the expression of full length SBP2 protein vs untreated cells especially at the 96 hour time point (Figure 5-11, B). The MG132 experiment established a link of the proteasome pathway to SBP2 degradation. Further experiments would need to be conducted to determine if SBP2 is polyubiquitinated. In the case that SBP2 is not ubiquitinated, it leaves open the possibility that another protein that regulates SBP2 degradation, is sent to the proteasome for degradation instead.

We found that cell to cell contact as cells approached confluence was the most reproduceable way to see high levels of SBP2 degradation products. Work is currently being done to clarify the specific triggers of high confluence causing this degradation, ie. cell senescence, oxidative stress, cell to cell signaling. We do not believe that media starvation of selenium or amino acids could be the cause, due to the changing of media every 24 hours during our experimental time courses. Another possibility could be

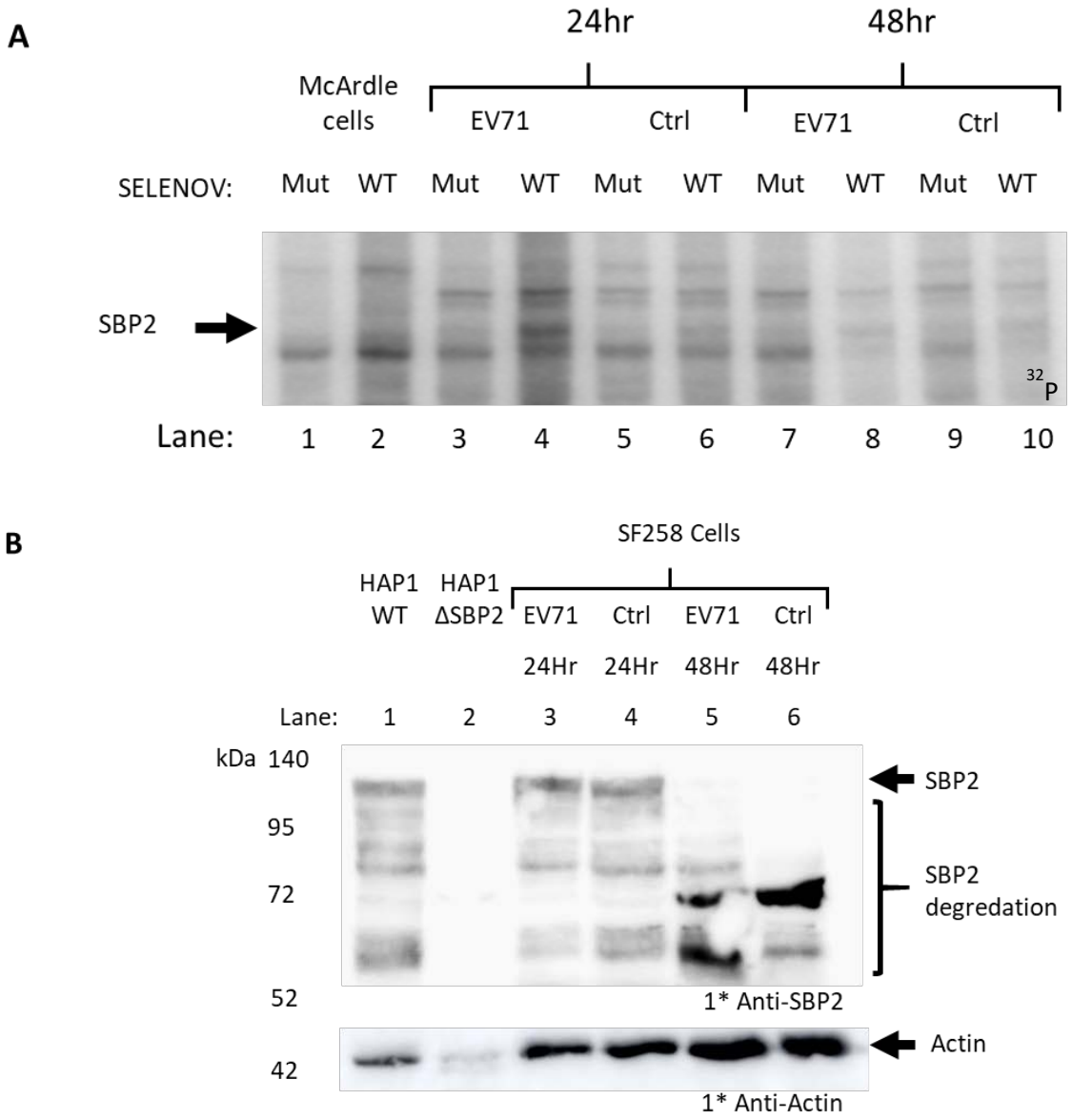


Figure 5-7. Time Course Evaluation of EV71 infection of SF248 cells in SBP2 expression. A. SF268 glioblastoma cells were infected with EV71 strain TW2231 or ctrl treated and collected at the 24 hour and 48 hour time points. Cell lysates were incubated with radiolabeled with ^{32}P-UTP labeled wild type or mutant SELENOV SECIS RNA and subjected to UV crosslinking. As a control lanes 1 and 2 contain extracts from McArdle cells hepatoma cells. B. The same SF268 cell lysate extracts probed for SBP2 and actin by western blot. Lane 1 and 2 were run with HAP1 WT cells and HAP1 cells CRISPR deleted for SBP2 as a control for background bands.

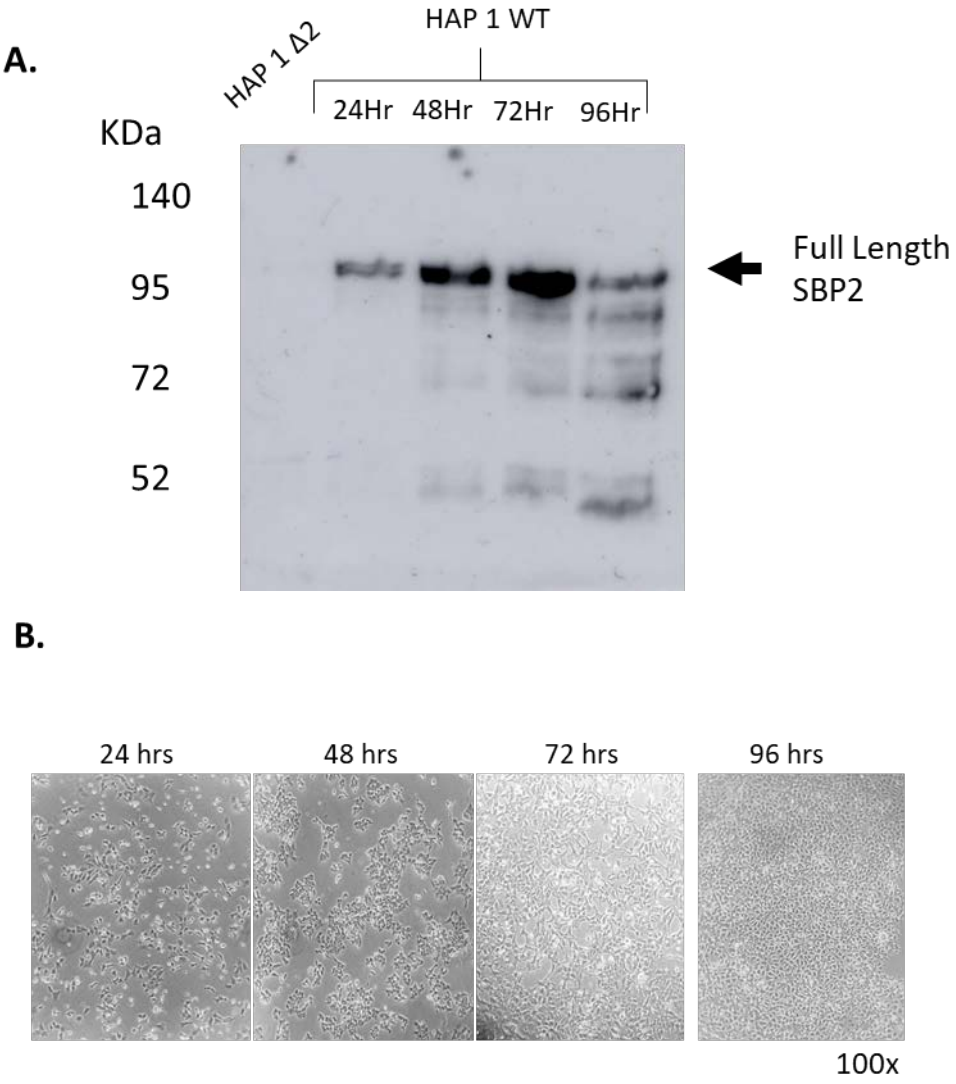


Figure 5-7. Time Course Evaluation of SBP2 degradation in HAP1 cells. A. 250,000 HAP1 cells were seeded in a 6 well plate and allowed to grow with daily media changes. Lysates were collected at the indicated timepoints and probed for SBP2 by immunoblotting. **B.** 100x magnification of HAP1 cells before each harvesting timepoint.

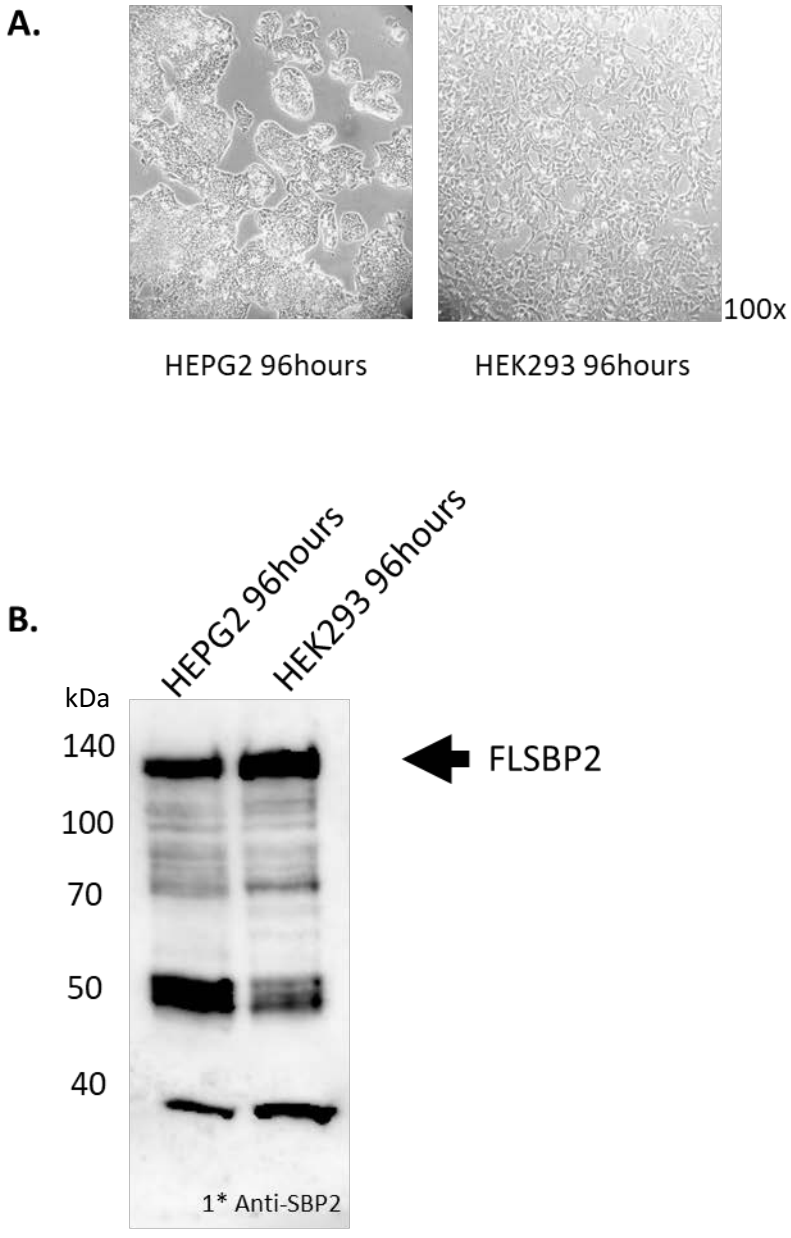


Figure 5-8. Evaluation of SBP2 degradation in HepG2 and HEK293 cells gown in culture for 96 hours. A. 250,000 HEPG2 and HEK293 cells were seeded in a 6 well plate and allowed to grow with daily media changes for 96 hours. Image was taken at 100x before cells were harvested. **B.** Lysates were made from HepG2 and HEK293 cells after the 96 hour time point and probed for SBP2 by immunoblotting.

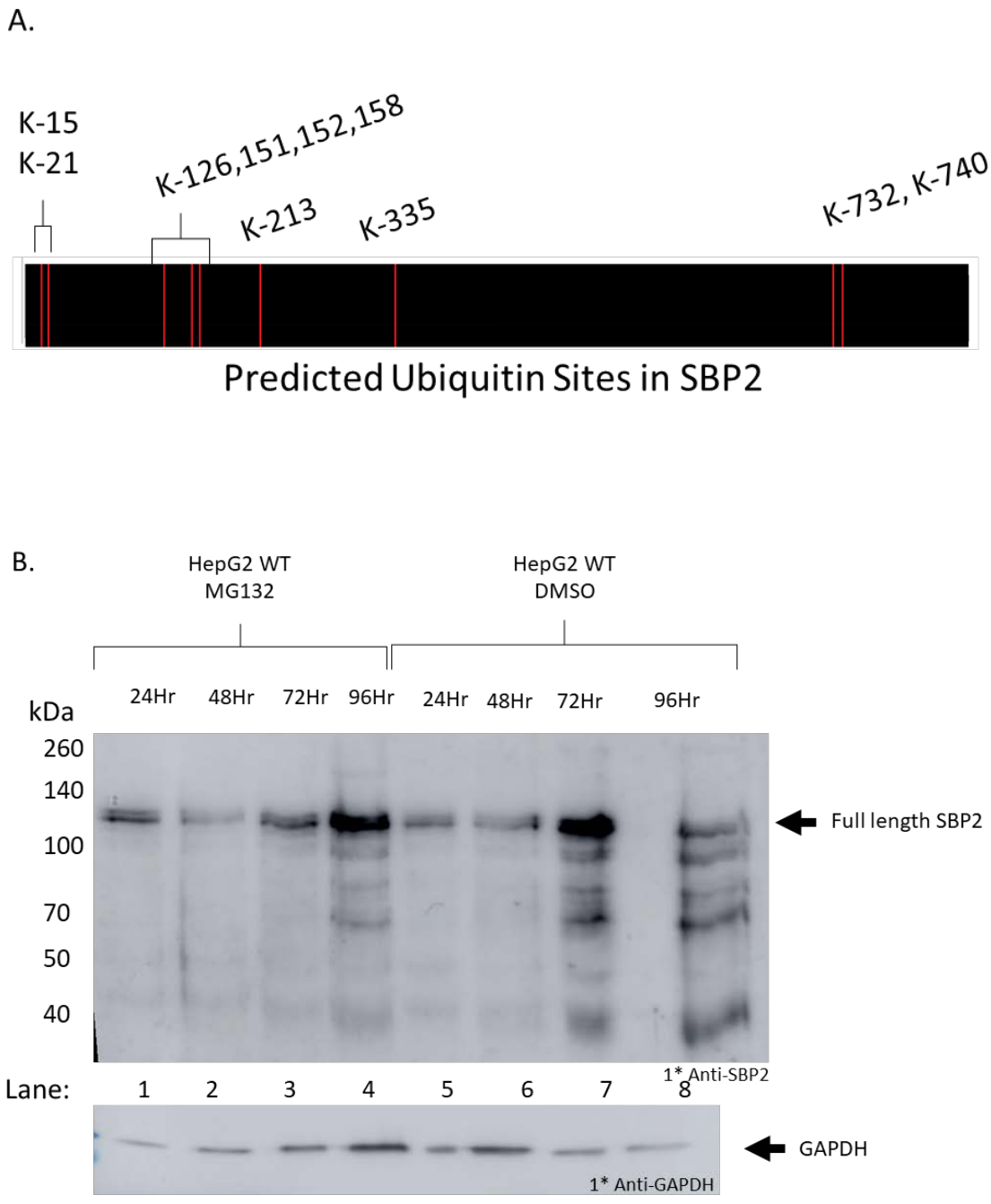


Figure 5-9. MG132 inhibition of the proteasome changes SBP2 degradation patterns. A. Representative diagram of predicted Ubiquitin Sites in SBP2. Ubiquitin sites were predicted by UbiSite and UPRED web servers. **B.** Immunoblot analysis of SBP2 and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a loading control. 250,000 HepG2 cells were seeded and grown in a 24-96 hour time course. Lanes 1-4 were treated with 10 μM of MG132. Lanes 5-8 were treated with 10 μM of DMSO.

based on cell to cell contact causing a signal transduction cascade such as integrin signaling with the Wnt/β-catenin (Crampton et al., 2009). The Wnt signaling pathway has demonstrated involvement with glutathionine peroxidase 2 (GPx2) as well as thioredoxin reductases 2 and 3 (Kipp et al., 2012; Brigelius-Flohe and Kipp, 2013). β-catenin, is also reported to undergo ubiquitin mediated degradation at levels of high confluence, causing the inhibition of the Wnt pathway. It is possible this cascade could cause subsequent downregulation of the aforementioned selenoproteins though SBP2 degradation, but more data is needed to make that conclusion.

Final Summary and Conclusions

The main goal of the data presented in this thesis was to expand our knowledge of the mechanism in the recoding of ten in frame UGA stop codons into Sec in SELENOP. There were two main *in vitro* lysates used in this thesis to study this topic, a mammalian RRL, which contains mammalian factors involved in Sec incorporation. The other *in vitro* lysate is WGL, which contains no Sec incorporation factors. Previous research demonstrated that WGL could read through single UGAs selenoprotein mRNAs with the addition of all the known Sec incorporation factors: eEFSec, Sec-tRNA$^{[Ser]Sec}$, and CTSBP2. Yet, these known factors could not read through past the 2nd UGA in a SELENOP mRNA, even though these factors were sufficient for full length synthesis in RRL (Shetty et al., 2014). The results suggested another unknow factor was involved, which had to be contained in RRL.

The beginning of this work attempted to reconstitute processive Sec incorporation in Wheat Germ Lysate. Attempts to add high levels of high purity RPC-5 ^{75}Se Sec-

tRNA$^{[Ser]Sec}$ increased the visualization of termination products past the 2nd UGA, however it was far from achieving the primarily full length SELENOP seen in other RRL experiments. It was unknown if it was even possible for WGL to be compatible with processive Sec incorporation up to this point. It took fractionation of RRL to ribosomal pellets and 0.5M KCl salt washes from ribosomal pellets where we were able to reconstitute highly processive SELENOP synthesis in the WGL system. This led to speculation of the involvement of the N-terminal domain of SBP2 and SBP2L, especially when RSW demonstrates its own SBP2 independent Sec incorporation activity in RRL. *In vitro* comparison of full length SBP2 and C-terminal SBP2 had significant difference processive Sec incorporation. The combination of the 0.5M KCl salt washes also saw a difference between full length SBP2 and C-terminal SBP2. This suggested SBP2 might need to be activated by proper folding or by post translational modifications. Although we learned WGL is a very powerful system, we were unfortunately unable to manage to obtain a robust processive Sec incorporation assay without the use of highly purified RPC-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$. The main issue with this technique for Sec-tRNA$^{[Ser]Sec}$ purification is that it is not a practical due to high cost, the amount of radioactivity required, and the low availability of high specific activity ^{75}Se selenite. We have exhausted all attempts to find an assay close to the robustness of that of RPC-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$. Some of these multiple alternative methods were not entirely practical due to background, such as ^{35}S Cys for use in WGL due to the lower efficiency of the first UGA readthrough and each subsequent UGAs compared to RRL. We found ^{35}S Cys was not sensitive enough to visualize higher molecular weight products

of SELENOP without the deletion of the first UGA, and background was still a persistent issue with faint bands. We found that ^{75}Se labeled HepG2 tRNA to be the closest labeling technique to RPC-5 ^{75}Se Sec-tRNA$^{[Ser]Sec}$ labeling but found that WGL was sensitive to contamination of rRNAs and mRNAs. We found column cleanup using benzoylated DEAE cellulose, while effective of removing nonspecific RNAs, also reduced the maximum specific activity we could obtain. We had the same challenge with using eEFSec ternary complex capture of ^{75}Se radiolabeled HepG2 total tRNA which was also unable to provide high enough Sec-tRNA$^{[Ser]Sec}$ concentration and specific activity, despite having a high purity. This technique also ended up being time consuming and cost prohibitive. The most effective method at retaining the highest specific activity, while being relatively quick and cost-effective was using ultra centrifugation to remove mRNAs and rRNAs. The use of this tRNA in a high enough concentration can detect termination products past the 2nd UGA with ease. We hope this technique can be further utilized for in vitro analysis of processive Sec incorporation and other selenoproteins in the future. Perhaps future work in modification of the Sec-tRNA$^{[Ser]Sec}$ binding pocket of eEFSec or *in vitro* reconstitution of the full Sec aminoacylation pathway could help advance our capabilities with experiments using RRL and WGL to study Sec incorporation.

We established the N-terminal domain of SBP2 being at least one of the requirements for full processive Sec incorporation in our WGL experiments comparing CTSBP2 vs FL SBP2. What showed the dynamic nature of SBP2 N-terminal domain involvement in Sec incorporation was when we used RRL to test the differences

between CTSBP2 and FLSBP2. Using a ZFSELENOP construct with a wild type 3'UTR two 3'UTR mutants with only a single type I SECIS element from zebrafish or human SELENOP, which have previously reported some early truncation products. We do not see a significant difference between CT and FLSBP in the full 3'UTR construct but see a significant difference in processivity when only a single SEICS element from SELENOP from either zebrafish or human is added instead. Additional testing would need to be performed in the future with only SECIS deletions.

Our study of processive Sec incorporation led us to take another look at the SECIS binding proteins in the context of bioinformatics and evolution. In this case, we shaped our analysis of SBP2 and SBP2L based on our hypothesis that the N-terminal domain was an important component for processive Sec incorporation. Based on previous work by Donovan and Copeland, 2009 we looked for invertebrates with conserved N-terminal motifs specifically those of known to produce SELENOP and SELENOP2. This proved to be quite difficult due to the variable distance between conserved motifs and inaccurate predicted annotations due to multiple initiation sites, so we first aligned sequences of vertebrate SBP2 and SBP2L. We also utilized the help of the bioinformatic tool Selenoprofiles, by Marco Mariotti, to help find SBPs *de novo* as well using a profile from a wide range of vertebrate and invertebrate SBPs. While these *de novo* predictions were not always 100% accurate, it showed us were to look for SBP2 sequences in genomic and EST data. To improve the profile, there would likely need to be adjustment for C-terminal only SBPs and SBPs that contain a full N-terminus for accuracy.

What was unexpected, were the findings of SBP2s in chordates where we discovered there were 3 distinct branches of SBP2s in vertebrates and a pseudo gene of SBP2 found to be retained in the genetic sequence of a small subset of mammals based on phylogenetic analysis. It seems that somewhere in the evolutionary timeline the SBP2L duplication from a common ancestor of vertebrates and lancelets caused a divergence in to SBP2L and the emergence of SBP2. This is suspected to be from one of the whole genome duplication events in the history of vertebrates. It seems that sometime during the emergence of early mammals, SBP2 also duplicated again. The source of this seems to be a chromosomal duplication predicted of "late" genes in mammalian evolution. It also provides an explanation for the apparent faster evolution rate of SBP2 according to phylogenetic trees.

After we were able to fully align both SBP2 and SBP2L in vertebrates, they revealed a number of N-terminal motifs, which were not found to align previously due to the inclusion of less related invertebrate animals. When aligning SBP2 and SBP2L alone we found the SID of SBP2L may actually be longer than originally thought based on our alignment data, and in silico protein structural prediction. This could explain why we were able to see Capitella CTSBP2L Sec incorporation activity, but not in human CTSBP2L, which was missing a piece of the predicted SID region. In our *in vitro* test with a predicted SELENOP producing organism we observed Sec incorporation activity using oyster SBP2L, however we did not see a level of processivity we see with other mammalian SBP2s. A closer look into the SID of the oyster SBP2L revealed that it is missing a number of conserved motifs in mammalian SBP2, which would affect eEFSec

interaction in RRL. We contemplate if activity would have increased if we had used a recombinant eEFSec from the oyster as well. Taken together we hope this will further future research into SBP2L interactions in vitro, however we have not been able to make a functional full length protein to date. It could be possible that in humans SBP2L and SBP2 functions have diverged so much that SBP2L no longer participates in Sec incorporation directly. The support for SBP2L function directly in Sec incorporation is supported by CRISPR deletion of SBP2 in cultured cells and knockout experiments in mice. In these cases, we still find the production of selenoproteins, albeit at lower levels. This may be the basis for future investigational efforts by using an organism from an earlier evolutionary time point of vertebrates, *in vitro* and in cell lines.

Finally, after discovering all of the N-terminal motifs of vertebrates we also searched for invertebrates with conserved motifs in their N-terminal domains as well. Several clades of invertebrates were revealed to have highly conserved motifs of the N-terminus, but we noticed a large number of them were not predicted to contain SELENOP. Through closer examination into sequence data with Selenoprofiles and alignments with closely related organisms, we were able to find that virtually all but a few of those organisms had a predicted SELENOP gene. Most of these organisms contained multiple in frame UGAs. We found SELENOP all the way to cnidarians, which had two SECIS elements and in many cases two separate genes. For the few exceptions that did not produce SELENOP, we found multiple UGA containing SELENON and DIO-like proteins. Organisms that lacked motifs in the N-terminus also lacked a multiple UGA SELENOP.

We found the most highly conserved motifs in these multiple UGA selenoprotein producing animals was the LSAD motif in duplicate or with the PFVQ motif in later metazoans, and the QEPP/LYFED motif. We utilized disorder prediction to see what kind of structure the LSAD and PFVQ motif could be and found they were predicted to form flexible loops. Due to their prediction to form loops and their location on the very end of the protein, it seemed possible that these motifs could be used to bind to the SRE of proteins. It could be possible that while bound to the SECIS element the N-terminal domain is free to interact with the SRE of selenoprotein mRNAs. We have found that in the case of SELENON the SRE appears to be highly conserved. Further investigation would need to be conducted using mutants of these regions. A CRISPR cell line for SBP2 would be valuable for this research, as mutant SBPs could be transfected back and global selenoprotein expression could be observed via labeling. Obtaining CryoEM structural data would also be very informative to the field, as attempts to use X-ray crystallography has been unsuccessful to date likely due to the large sections of disorder within the protein. Knowledge of SBP2 structure would be groundbreaking in the understanding of the mechanism of the many functions of SBP2.

Appendix:

Profile for SBP2 search was added to the newest version of Selenoprofiles (3.5b)

https://github.com/marco-mariotti/selenoprofiles

De novo SBPs annotated by Selenoprofiles 3.5b, annotated SBPs, and *de novo* SELENOP sequences can be found online at: https://pastebin.com/CSnCMMcp

Backup link: https://docs.google.com/document/d/1IOGSQcsaebTGgO6VLjVcW1GpxgisWYPzaOpZzEswreo/edit?usp=sharing

List of Abbreviations

3' UTR	3' untranslated region
aa	amino acid
aa-tRNA	amino-acyl tRNA
CRM-1	chromosome region maintenance 1
CT	C-terminal
Cys	cysteine
DDC	duplication-degeneration-complementation
DIO	iodothyronine deiodinase
eEF1A	eukaryotic elongation factor 1A
eEFSec	Sec specific translation elongation factor
EF-Tu	elongation factor Tu
eIF4A3	eukaryotic initiation factor 4AIII
eIF4A3	eukaryotic initiation factor 4AIII
eRF	eukaryotic release factor
fmol	femtomole
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GDP	guanine diphosphate
GEF	guanine exchange factor
GPX	glutathione peroxidase
GTP	guanine triphosphate
IRES	internal ribosome entry site
Kd	dissociation constant
kDa	kilodalton
mcm^5U	methylcarboxymethyl-2'-O-uridine
mcm^5Um	methylcarboxymethyl-2'-O-methyluridine
mRNA	messenger RNA
NES	nuclear export signal
NLS	nuclear localization signal
NMD	nonsense mediated decay
NT	N-terminal
RBD	RNA binding domain
RNA	ribonucleic acid
ROS	reactive oxygen species
RP	ribosomal protein
RRL	rabbit reticulocyte lysate
S	Sulfur
SBP2	SECIS binding protein 2
SBP2L	SECIS binding protein 2-like
SDS	sodium dodecyl sulfate
Se	Selenium
Sec	selenocysteine
SECIS	Sec insertion sequence

Ser	Serine
SHAPE	selective 2’ -hydroxyl acylation analyzed by primer extension
SID	Sec incorporation domain
SRE	Sec redefinition element
tRNA	transfer RNA
Trp	Tryptophan
TXNRD	thioredoxin reductase
UTP	uridine triphosphate
WGL	wheat germ lysate

References

Achsel, T., & Gross, H. J. (1993). Identity determinants of human tRNA(Ser): sequence elements necessary for serylation and maturation of a tRNA with a long extra arm. The EMBO Journal, 12(8), 3333–3338. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8344269%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC413601

Aguinaldo, A. M. A., Turbeville, J. M., Linford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A., & Lake, J. A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature, 387(6632), 489–493. https://doi.org/10.1038/387489a0

Allamand, V., Richard, P., Lescure, A., Ledeuil, C., Desjardin, D., Petit, N., … Guicheney, P. (2006). A single homozygous point mutation in a 3'untranslated region motif of selenoprotein N mRNA causes SEPN1-related myopathy. EMBO Reports, 7(4), 450–454. https://doi.org/10.1038/sj.embor.7400648

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 (New York, N.Y.), 8(10), 1308–1318.

Anonymous (1979) Chin Med J Engl 92:471–476

Asahara, H., Himeno, H., Tamura, K., Nameki, N., Hasegawa, T., & Shimizu, M. (1994). Escherichia coli seryl-tRNA synthetase recognizes tRNA(Ser) by its characteristic tertiary structure. Journal of Molecular Biology, 236(3), 738–748.

Barucca, M., Canapa, A., & Biscotti, M. (2016). An Overview of Hox Genes in Lophotrochozoa: Evolution and Functionality. Journal of Developmental Biology, 4(1), 12. https://doi.org/10.3390/jdb4010012

Bedard, K. M., & Semler, B. L. (2004). Regulation of picornavirus gene expression. Microbes and Infection, 6(7), 702–713. https://doi.org/10.1016/j.micinf.2004.03.001

Bentley, M. A., Holland, P. W. H., Feuda, R., & Marle, F. (2016). Conservation, Duplication, and Divergence of Five Opsin Genes in Insect Evolution. Genome Biology and Evolution, 8(3), 579–587. https://doi.org/10.5287/bod-leian

Berná, L., & Alvarez-Valin, F. (2014). Evolutionary genomics of fast evolving tunicates. Genome Biology and Evolution, 6(7), 1724–1738. https://doi.org/10.1093/gbe/evu122

Berry, M. J., & Larsen, P. R. (1993b). Molecular cloning of the selenocysteine-containing enzyme type I iodothyronine deiodinase. The American Journal of Clinical Nutrition, 57(2 Suppl), 249S-255S. https://doi.org/10.1093/ajcn/57.2.249S

Berry, M. J., Banu, L., Chen, Y. Y., Mandel, S. J., Kieffer, J. D., Harney, J. W., & Larsen, P. R. (1991). Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. Nature, 353(6341), 273–276. https://doi.org/10.1038/353273a0

Berry, M. J., Banu, L., Harney, J. W., & Larsen, P. R. (1993a). Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. The EMBO Journal, 12(8), 3315–3322.

Bertram, G., Innes, S., Minella, O., Richardson, J., & Stansfield, I. (2001). Endless possibilities: translation termination and stop codon recognition. Microbiology (Reading, England), 147(Pt 2), 255–269. https://doi.org/10.1099/00221287-147-2-255

Bifano, A. L., Atassi, T., Ferrara, T., & Driscoll, D. M. (2013). Identification of nucleotides and amino acids that mediate the interaction between ribosomal protein L30 and the SECIS element. BMC Molecular Biology, 14, 12. https://doi.org/10.1186/1471-2199-14-12

Boulon, S., Marmier-Gourrier, N., Pradet-Balade, B., Wurth, L., Verheggen, C., Jády, B. E., … Charpentier, B. (2008). The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. Journal of Cell Biology, 180(3), 579–595. https://doi.org/10.1083/jcb.200708110

Brigelius-Flohe, R., & Kipp, A. P. (2013). Selenium in the redox regulation of the Nrf2 and the Wnt pathway. Methods in Enzymology, 527, 65–86. https://doi.org/10.1016/B978-0-12-405882-8.00004-0

Bubenik, J. L., & Driscoll, D. M. (2007). Altered RNA binding activity underlies abnormal thyroid hormone metabolism linked to a mutation in selenocysteine insertion sequence-binding protein 2. The Journal of Biological Chemistry, 282(48), 34653–34662. https://doi.org/10.1074/jbc.M707059200

Bubenik, J. L., Ladd, A. N., Gerber, C. A., Budiman, M. E., & Driscoll, D. M. (2009). Known turnover and translation regulatory RNA-binding proteins interact with the 3' UTR of SECIS-binding protein 2. RNA Biology, 6(1), 73–83.

Budachetri, K., Crispell, G., & Karim, S. (2017). Amblyomma maculatum SECIS binding protein 2 and putative selenoprotein P are indispensable for pathogen replication and tick fecundity. Insect Biochemistry and Molecular Biology, 88, 37–47. https://doi.org/10.1016/j.ibmb.2017.07.006

Budiman, M. E., Bubenik, J. L., & Driscoll, D. M. (2011). Identification of a signature motif for the eIF4a3-SECIS interaction. Nucleic Acids Research, 39(17), 7730–7739. https://doi.org/10.1093/nar/gkr446

Budiman, M. E., Bubenik, J. L., Miniard, A. C., Middleton, L. M., Gerber, C. A., Cash, A., & Driscoll, D. M. (2009). Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation. Molecular Cell, 35(4), 479–489. https://doi.org/10.1016/j.molcel.2009.06.026

Bulygin, K. N., Bartuli, Y. S., Malygin, A. A., Graifer, D. M., Frolova, L. Y., & Karpova, G. G. (2016). Chemical footprinting reveals conformational changes of 18S and 28S rRNAs at different steps of translation termination on the human ribosome. Rna, 22(2), 278–289. https://doi.org/10.1261/rna.053801.115

Burge, C. B., & Karlin, S. (1998). Finding the genes in genomic DNA. Current Opinion in Structural Biology, 8(3), 346–354.

Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology, 268(1), 78–94. https://doi.org/10.1006/jmbi.1997.0951

Caban, K., & Copeland, P. R. (2006). Size matters: a view of selenocysteine incorporation from the ribosome. Cellular and Molecular Life Sciences : CMLS, 63(1), 73–81. https://doi.org/10.1007/s00018-005-5402-y

Caban, K., & Copeland, P. R. (2012). Selenocysteine insertion sequence (SECIS)-binding protein 2 alters conformational dynamics of residues involved in tRNA accommodation in 80 S ribosomes. The Journal of Biological Chemistry, 287(13), 10664–10673. https://doi.org/10.1074/jbc.M111.320929

Caban, K., Kinzy, S. A., & 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. Molecular and Cellular Biology, 27(18), 6350–6360. https://doi.org/10.1128/MCB.00632-07

Carlson, B. A., Gupta, N., Pinkerton, M. H., Hatfield, D. L., & Copeland, P. R. (2017). The utilization of selenocysteine-tRNA([Ser]Sec) isoforms is regulated in part at the level of translation in vitro. Translation (Austin, Tex.), 5(1), e1314240. https://doi.org/10.1080/21690731.2017.1314240

Carlson, B. A., Lee, B. J., Tsuji, P. A., Copeland, P. R., Schweizer, U., Gladyshev, V. N., & Hatfield, D. L. (2018). Selenocysteine tRNA([Ser]Sec), the Central Component of Selenoprotein Biosynthesis: Isolation, Identification, Modification, and Sequencing. Methods in Molecular Biology (Clifton, N.J.), 1661, 43–60. https://doi.org/10.1007/978-1-4939-7258-6_4

Carlson, B. A., Lee, B. J., Tsuji, P. A., Tobe, R., Park, J. M., Schweizer, U., … Hatfield, D. L. (2016). Selenocysteine tRNA[Ser]Sec: From Nonsense Suppressor tRNA to the Quintessential Constituent in Selenoprotein Biosynthesis BT - Selenium: Its Molecular Biology and Role in Human Health. In D. L. Hatfield, U. Schweizer, P. A. Tsuji, & V. N. Gladyshev (Eds.) (pp. 3–12). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41283-2_1

Carlson, B. A., Xu, X.-M., Gladyshev, V. N., & Hatfield, D. L. (2005). Selective rescue of selenoprotein expression in mice lacking a highly specialized methyl group in selenocysteine tRNA. The Journal of Biological Chemistry, 280(7), 5542–5548. https://doi.org/10.1074/jbc.M411725200

Castellano, S., Gladyshev, V. N., Guigo, R., & Berry, M. J. (2008). SelenoDB 1.0 : a database of selenoprotein genes, proteins and SECIS elements. Nucleic Acids Research, 36(Database issue), D332-8. https://doi.org/10.1093/nar/gkm731

Castellano, S., Novoselov, S. V, Kryukov, G. V, Lescure, A., Blanco, E., Krol, A., … Guigo, R. (2004). Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution. EMBO Reports, 5(1), 71–77. https://doi.org/10.1038/sj.embor.7400036

Chang, P.-C., Chen, S.-C., & Chen, K.-T. (2016). The Current Status of the Disease Caused by Enterovirus 71 Infections: Epidemiology, Pathogenesis, Molecular Epidemiology, and Vaccine Development. International Journal of Environmental Research and Public Health, 13(9), 890. https://doi.org/10.3390/ijerph13090890

Chapple, C. E., & Guigo, R. (2008). Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. PloS One, 3(8), e2968. https://doi.org/10.1371/journal.pone.0002968

Chase, A. J., & Semler, B. L. (2012). Viral subversion of host functions for picornavirus translation and RNA replication. Future Virology, 7(2), 179–191.

Chavatte, L., Brown, B. A., & Driscoll, D. M. (2005). Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. Nature Structural & Molecular Biology, 12(5), 408–416. https://doi.org/10.1038/nsmb922

Cheng, M. L., Weng, S. F., Kuo, C. H., & Ho, H. Y. (2014). Enterovirus 71 induces mitochondrial reactive oxygen species generation that is required for efficient replication. PLoS ONE, 9(11). https://doi.org/10.1371/journal.pone.0113234

Chipman, A. D., Ferrier, D. E. K., Brena, C., Qu, J., Hughes, D. S. T., Schröder, R., … Richards, S. (2014). The First Myriapod Genome Sequence Reveals Conservative Arthropod Gene Content and Genome Organisation in the Centipede Strigamia maritima. PLoS Biology, 12(11). https://doi.org/10.1371/journal.pbio.1002005

Cone, J. E., Del Río, R. M., Davis, J. N., and Stadtman, T. C. (1976) Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. Proc. Natl. Acad. Sci. U. S. A. 73, 2659−2663.

Copeland, P. R., Fletcher, J. E., Carlson, B. A., Hatfield, D. L., & Driscoll, D. M. (2000). A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. The EMBO Journal, 19(2), 306–314. https://doi.org/10.1093/emboj/19.2.306

Copeland, P. R., Stepanik, V. A., & Driscoll, D. M. (2001). Insight into mammalian selenocysteine insertion: domain structure and ribosome binding properties of Sec insertion sequence binding protein 2. Molecular and Cellular Biology, 21(5), 1491–1498. https://doi.org/10.1128/MCB.21.5.1491-1498.2001

Crampton, S. P., Wu, B., Park, E. J., Kim, J. H., Solomon, C., Waterman, M. L., & Hughes, C. C. W. (2009). Integration of the β-catenin-dependent wnt pathway with integrin signaling through the adaptor molecule Grb2. PLoS ONE, 4(11). https://doi.org/10.1371/journal.pone.0007841

Di Cosmo, C., McLellan, N., Liao, X.-H., Khanna, K. K., Weiss, R. E., Papp, L., & Refetoff, S. (2009). Clinical and molecular characterization of a novel selenocysteine insertion sequence-binding protein 2 (SBP2) gene mutation (R128X). The Journal of Clinical Endocrinology and Metabolism, 94(10), 4003–4009. https://doi.org/10.1210/jc.2009-0686

Diamond, A. M., Choi, I. S., Crain, P. F., Hashizume, T., Pomerantz, S. C., Cruz, R., … Hatfield, D. L. (1993). Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA([Ser]Sec). The Journal of Biological Chemistry, 268(19), 14215–14223.

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. https://doi.org/10.1016/0092-8674(81)90068-4

Dobosz-Bartoszek, M., Pinkerton, M. H., Otwinowski, Z., Chakravarthy, S., Soll, D., Copeland, P. R., & Simonovic, M. (2016). Crystal structures of the human elongation factor eEFSec suggest a non-canonical mechanism for selenocysteine incorporation. Nature Communications, 7, 12941. https://doi.org/10.1038/ncomms12941

Dong, C., Zhang, J., Qiao, J., & He, G. (2012). Positive selection and functional divergence after melanopsin gene duplication. Biochemical Genetics, 50(3–4), 235–248. https://doi.org/10.1007/s10528-011-9466-0

Donovan, J. (2011). Post transcriptional regulation of selenoprotine expression by SECIS binding proteins. (PhD dissertation). Rutgers University. Picataway, NJ.

Donovan, J., & Copeland, P. R. (2009). Evolutionary history of selenocysteine incorporation from the perspective of SECIS binding proteins. BMC Evolutionary Biology, 9, 229. https://doi.org/10.1186/1471-2148-9-229

Donovan, J., & Copeland, P. R. (2010). The efficiency of selenocysteine incorporation is regulated by translation initiation factors. Journal of Molecular Biology, 400(4), 659–664. https://doi.org/10.1016/j.jmb.2010.05.026

Donovan, J., & Copeland, P. R. (2010). Threading the needle: getting selenocysteine into proteins. Antioxidants & Redox Signaling, 12(7), 881–892. https://doi.org/10.1089/ars.2009.2878

Donovan, J., Caban, K., Ranaweera, R., Gonzalez-Flores, J. N., & Copeland, P. R. (2008). A novel protein domain induces high affinity selenocysteine insertion sequence binding and elongation factor recruitment. The Journal of Biological Chemistry, 283(50), 35129–35139. https://doi.org/10.1074/jbc.M806008200

Dubey, A., & Copeland, P. R. (2016). The Selenocysteine-Specific Elongation Factor Contains Unique Sequences That Are Required for Both Nuclear Export and Selenocysteine Incorporation. PloS One, 11(11), e0165642. https://doi.org/10.1371/journal.pone.0165642

Duff, M. O., Olson, S., Wei, X., Garrett, S. C., Osman, A., Bolisetty, M., … Graveley, B. R. (2015). Genome-wide identification of zero nucleotide recursive splicing in Drosophila. Nature, 521(7552), 376–379. https://doi.org/10.1038/nature14475

Dumitrescu, A. M., & Refetoff, S. (2013). The syndromes of reduced sensitivity to thyroid hormone. Biochimica et Biophysica Acta, 1830(7), 3987–4003. https://doi.org/10.1016/j.bbagen.2012.08.005

Dumitrescu, A. M., & Refetoff, S. (2013). The syndromes of reduced sensitivity to thyroid hormone. Biochimica et Biophysica Acta, 1830(7), 3987–4003. https://doi.org/10.1016/j.bbagen.2012.08.005

Dumitrescu, A. M., Cosmo, C. Di, Liao, X.-H., Weiss, R. E., & Refetoff, S. (2009). The Syndrome of Inherited Partial SBP2 Deficiency in Humans. Antioxidants & Redox Signaling, 12(7), 905–920. https://doi.org/10.1089/ars.2009.2892

Dumitrescu, A. M., Di Cosmo, C., Liao, X.-H., Weiss, R. E., & Refetoff, S. (2010). The syndrome of inherited partial SBP2 deficiency in humans. Antioxidants & Redox Signaling, 12(7), 905–920. https://doi.org/10.1089/ars.2009.2892

Dumitrescu, A. M., Liao, X.-H., Abdullah, M. S. Y., Lado-Abeal, J., Majed, F. A., Moeller, L. C., … Refetoff, S. (2005). Mutations in SECISBP2 result in abnormal thyroid hormone metabolism. Nature Genetics, 37(11), 1247–1252. https://doi.org/10.1038/ng1654

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., … Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. Nature, 452, 745. Retrieved from https://doi.org/10.1038/nature06614

Eggert, R. O., E. Patterson, W. J. Akers, and E. L. R. Stokstad. 1957. The role of vitamin E and selenium in the nutrition of the pig. J. Anim. Sci. 16:1037.

Engelberg-Kulka, H. (1981). UGA suppression by normal tRNATrP in Escherichia coli: codon context effects. Nucleic Acids Research, 9(4), 983–991. https://doi.org/10.1093/nar/9.4.983

Fagegaltier, D., Carbon, P., & Krol, A. (2001). Distinctive features in the SelB family of elongation factors for selenoprotein synthesis. A glimpse of an evolutionary complexified translation apparatus. BioFactors (Oxford, England), 14(1–4), 5–10.

Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P., & Krol, A. (2000). Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. The EMBO Journal, 19(17), 4796–4805. https://doi.org/10.1093/emboj/19.17.4796

Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P., & Krol, A. (2000). Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. The EMBO Journal, 19(17), 4796–4805. https://doi.org/10.1093/emboj/19.17.4796

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., … Uhlén, M. (2013). Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. Molecular & Cellular Proteomics, 13(2), 397–406. https://doi.org/10.1074/mcp.m113.035600

Feuda, R., Hamilton, S. C., McInerney, J. O., & Pisani, D. (2012). Metazoan opsin evolution reveals a simple route to animal vision. Proceedings of the National Academy of Sciences, 109(46), 18868–18872. https://doi.org/10.1073/pnas.1204609109

Fixsen, S. M., & Howard, M. T. (2010). Processive selenocysteine incorporation during synthesis of eukaryotic selenoproteins. Journal of Molecular Biology, 399(3), 385–396. https://doi.org/10.1016/j.jmb.2010.04.033

Fletcher, J. E., Copeland, P. R., Driscoll, D. M., & Krol, A. (2001). The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. RNA (New York, N.Y.), 7(10), 1442–1453.

Forstrom, J. W., Zakowski, J. J., & Tappel, A. L. (1978). Identification of the catalytic site of rat liver glutathione peroxidase as selenocysteine. Biochemistry, 17(13), 2639–2644.

Fradejas-Villar, N., Seeher, S., Anderson, C. B., Doengi, M., Carlson, B. A., Hatfield, D. L., … Howard, M. T. (2017). The RNA-binding protein Secisbp2 differentially modulates UGA codon reassignment and RNA decay. Nucleic Acids Research, 45(7), 4094–4107. https://doi.org/10.1093/nar/gkw1255

Franke, K. W., and Painter, E. P. (1936) Selenium in proteins from toxic foodstuffs I: Remarks on the occurrence and nature of the selenium present in a number of foodstuffs or their derived products. Cereal Chem. 13, 67−70.

Fu, J., Fujisawa, H., Follman, B., Liao, X.-H., & Dumitrescu, A. M. (2017). Thyroid Hormone Metabolism Defects in a Mouse Model of SBP2 Deficiency. Endocrinology, 158(12), 4317–4330. https://doi.org/10.1210/en.2017-00618

Fu, J., Fujisawa, H., Follman, B., Liao, X.-H., & Dumitrescu, A. M. (2017). Thyroid Hormone Metabolism Defects in a Mouse Model of SBP2 Deficiency. Endocrinology, 158(12), 4317–4330. https://doi.org/10.1210/en.2017-00618

Gehring, N. H., Kunz, J. B., Neu-Yilik, G., Breit, S., Viegas, M. H., Hentze, M. W., & Kulozik, A. E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Molecular Cell, 20(1), 65–75. https://doi.org/10.1016/j.molcel.2005.08.012

Gladyshev, V. (2016). Eukaryotic Selenoproteomes. In Selenium: Its Molecular Biology and Role in Human Health: Fourth Edition (pp. 127–139). https://doi.org/10.1007/978-3-319-41283-2_11

Gonzalez-Flores, J. N., Gupta, N., DeMong, L. W., & Copeland, P. R. (2012). The selenocysteine-specific elongation factor contains a novel and multi-functional domain. The Journal of Biological Chemistry, 287(46), 38936–38945. https://doi.org/10.1074/jbc.M112.415463

Grundner-Culemann, E., Martin, G. W. 3rd, Harney, J. W., & Berry, M. J. (1999). Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. RNA (New York, N.Y.), 5(5), 625–635.

Guan X. (2019). Efficacy Trial of a Commercial EV71 Vaccine. ClinicalTrials.gov Identifier: NCT03903926

Gupta, M., & Copeland, P. R. (2007). Functional analysis of the interplay between translation termination, selenocysteine codon context, and selenocysteine insertion sequence-binding protein 2. The Journal of Biological Chemistry, 282(51), 36797–36807. https://doi.org/10.1074/jbc.M707061200

Gupta, M., & Gupta, S. (2017). An Overview of Selenium Uptake, Metabolism, and Toxicity in Plants. Frontiers in Plant Science, 7(January), 1–14. https://doi.org/10.3389/fpls.2016.02074

Gupta, N., DeMong, L. W., Banda, S., & Copeland, P. R. (2013). Reconstitution of selenocysteine incorporation reveals intrinsic regulation by SECIS elements. Journal of Molecular Biology, 425(14), 2415–2422. https://doi.org/10.1016/j.jmb.2013.04.016

Guschanski, K., Warnefors, M., & Kaessmann, H. (2017). The evolution of duplicate gene expression in mammalian organs, 1461–1474. https://doi.org/10.1101/gr.215566.116.3

Hatfield, D., Lee, B. J., Hampton, L., & Diamond, A. M. (1991). Selenium induces changes in the selenocysteine tRNA[Ser]Sec population in mammalian cells. Nucleic Acids Research, 19(4), 939–943.

Hill, K. E., Lloyd, R. S., & Burk, R. F. (1993). Conserved features of selenoprotein P cDNA. Biochemical Society Transactions, 21(4), 832–835.

Hill, K. E., Lloyd, R. S., & Burk, R. F. (1993). Conserved nucleotide sequences in the open reading frame and 3' untranslated region of selenoprotein P mRNA. Proceedings of the National Academy of Sciences of the United States of America, 90(2), 537–541.

Hirosawa-Takamori, M., Chung, H.-R., & Jäckle, H. (2004). Conserved selenoprotein synthesis is not critical for oxidative stress defence and the lifespan of Drosophila. EMBO Reports, 5(3), 317–322. https://doi.org/10.1038/sj.embor.7400097

Ho, H.-Y., Cheng, M.-L., Weng, S.-F., Chang, L., Yeh, T.-T., Shih, S.-R., & Chiu, D. T.-Y. (2008). Glucose-6-phosphate dehydrogenase deficiency enhances enterovirus 71 infection. The Journal of General Virology, 89(Pt 9), 2080–2089. https://doi.org/10.1099/vir.0.2008/001404-0

Howard, M. T., Aggarwal, G., Anderson, C. B., Khatri, S., Flanigan, K. M., & Atkins, J. F. (2005). Recoding elements located adjacent to a subset of eukaryal selenocysteine-specifying UGA codons. The EMBO Journal, 24(8), 1596–1607. https://doi.org/10.1038/sj.emboj.7600642

Howard, M. T., Carlson, B. A., Anderson, C. B., & Hatfield, D. L. (2013). Translational redefinition of UGA codons is regulated by selenium availability. The Journal of Biological Chemistry, 288(27), 19401–19413. https://doi.org/10.1074/jbc.M113.481051

Huber, R. E. J., & Criddle, R. S. (1967). Comparison and of the Chemical Selenocystine Properties Their Sulfur of Selenocysteine Analogs ' with their chemical properties , and differ only slightly in properties such as ionic radius , several attempts have been made to replace the sulfur needed fo. Archives of Biochemistry and Biophysics, 122, 164–173.

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Molecular Biology and Evolution, 33(6), 1635–1638. https://doi.org/10.1093/molbev/msw046

Hüttenhofer, A., & Böck, A. (1998). Selenocysteine Inserting RNA Elements Modulate GTP Hydrolysis of Elongation Factor SelB. Biochemistry, 37(3), 885–890. https://doi.org/10.1021/bi972298k

Huttenhofer, A., Westhof, E., & Bock, A. (1996). Solution structure of mRNA hairpins promoting selenocysteine incorporation in Escherichia coli and their base-specific interaction with special elongation factor SELB. RNA (New York, N.Y.), 2(4), 354–366.

Inagaki, Y., Ehara, M., Watanabe, K. I., Hayashi-Ishimaru, Y., & Ohama, T. (1998). Directionally Evolving Genetic Code: The UGA Codon from Stop to Tryptophan in Mitochondria. Journal of Molecular Evolution, 47(4), 378–384. https://doi.org/10.1007/PL00006395

Inamine, J. M., Ho, K. C., Loechel, S., & Hu, P. C. (1990). Evidence that UGA is read as a tryptophan codon rather than as a stop codon by Mycoplasma pneumoniae, Mycoplasma genitalium, and Mycoplasma gallisepticum. Journal of Bacteriology, 172(1), 504–506. https://doi.org/10.1128/jb.172.1.504-506.1990

Ivanova, N. N., Schwientek, P., Tripp, H. J., Rinke, C., Pati, A., Huntemann, M., … Rubin, E. M. (2014). Stop codon reassignments in the wild. Science (New York, N.Y.), 344(6186), 909–913. https://doi.org/10.1126/science.1250691

Jiang, L., & Liu, Q. (2018). SelGenAmic: An Algorithm for Selenoprotein Gene Assembly. Methods in Molecular Biology (Clifton, N.J.), 1661, 29–39. https://doi.org/10.1007/978-1-4939-7258-6_3

Jiang, L., Liu, Q., & Ni, J. (2010). In silico identification of the sea squirt selenoproteome. BMC Genomics, 11, 289. https://doi.org/10.1186/1471-2164-11-289

Jiang, L., Ni, J., & Liu, Q. (2012). Evolution of selenoproteins in the metazoan. BMC Genomics, 13, 446. https://doi.org/10.1186/1471-2164-13-446

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences : CABIOS, 8(3), 275–282.

Kasahara, M. (2007). The 2R hypothesis : an update, 547–552. https://doi.org/10.1016/j.coi.2007.07.009

Kasaikina, M. V, Lobanov, A. V, Malinouski, M. Y., Lee, B. C., Seravalli, J., Fomenko, D. E., … Gladyshev, V. N. (2011). Reduced utilization of selenium by naked mole rats due to a

specific defect in GPx1 expression. The Journal of Biological Chemistry, 286(19), 17005–17014. https://doi.org/10.1074/jbc.M110.216267

Kekez, M., Bauer, N., Saric, E. et al. J. Plant Biol. (2016) 59: 44. https://doi.org/10.1007/s12374-016-0370-3

Kekez, M., Zanki, V., Kekez, I., Baranasic, J., Hodnik, V., Duchene, A.-M., … Rokov-Plavec, J. (2019). Arabidopsis seryl-tRNA synthetase: the first crystal structure and novel protein interactor of plant aminoacyl-tRNA synthetase. The FEBS Journal, 286(3), 536–554. https://doi.org/10.1111/febs.14735

Kinzy, S. A., Caban, K., & Copeland, P. R. (2005). Characterization of the SECIS binding protein 2 complex required for the co-translational insertion of selenocysteine in mammals. Nucleic Acids Research, 33(16), 5172–5180. https://doi.org/10.1093/nar/gki826

Kipp, A. P., Muller, M. F., Goken, E. M., Deubel, S., & Brigelius-Flohe, R. (2012). The selenoproteins GPx2, TrxR2 and TrxR3 are regulated by Wnt signalling in the intestinal epithelium. Biochimica et Biophysica Acta, 1820(10), 1588–1596. https://doi.org/10.1016/j.bbagen.2012.05.016

Klein, D. J., Schmeing, T. M., Moore, P. B., & Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. The EMBO Journal, 20(15), 4214–4221. https://doi.org/10.1093/emboj/20.15.4214

Koonin, E. V, 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 Research, 22(11), 2166–2167. https://doi.org/10.1093/nar/22.11.2166

Korotkov, K. V, Novoselov, S. V, Hatfield, D. L., & Gladyshev, V. N. (2002). Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element. Molecular and Cellular Biology, 22(5), 1402–1411.

Kossinova, O., Malygin, A., Krol, A., & Karpova, G. (2013). A novel insight into the mechanism of mammalian selenoprotein synthesis. RNA (New York, N.Y.), 19(8), 1147–1158. https://doi.org/10.1261/rna.036871.112

Krol, A. (2002). Evolutionarily different RNA motifs and RNA-protein complexes to achieve selenoprotein synthesis. Biochimie, 84(8), 765–774.

Kryukov, G. V, & Gladyshev, V. N. (2000). Selenium metabolism in zebrafish: multiplicity of selenoprotein genes and expression of a protein containing 17 selenocysteine residues. Genes to Cells : Devoted to Molecular & Cellular Mechanisms, 5(12), 1049–1060.

Kryukov, G. V, Castellano, S., Novoselov, S. V, Lobanov, A. V, Zehtab, O., Guigo, R., & Gladyshev, V. N. (2003). Characterization of mammalian selenoproteomes. Science (New York, N.Y.), 300(5624), 1439–1443. https://doi.org/10.1126/science.1083516

Latreche, L., Duhieu, S., Touat-Hamici, Z., Jean-Jean, O., & Chavatte, L. (2012). The differential expression of glutathione peroxidase 1 and 4 depends on the nature of the SECIS element. RNA Biology, 9(5), 681–690. https://doi.org/10.4161/rna.20147

Latreche, L., Jean-Jean, O., Driscoll, D. M., & Chavatte, L. (2009). Novel structural determinants in human SECIS elements modulate the translational recoding of UGA as selenocysteine. Nucleic Acids Research, 37(17), 5868–5880. https://doi.org/10.1093/nar/gkp635

Lenhard, B., Orellana, O., Ibba, M., & Weygand-Durašević, I. (1999). tRNA recognition and evolution of determinants in seryl-tRNA synthesis. Nucleic Acids Research, 27(3), 721–729. https://doi.org/10.1093/nar/27.3.721

Lescure, A., Gautheret, D., Carbon, P., & Krol, A. (1999). Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif. The Journal of Biological Chemistry, 274(53), 38147–38154.

Lesoon, A., Mehta, A., Singh, R., Chisolm, G. M., & Driscoll, D. M. (1997). An RNA-binding protein recognizes a mammalian selenocysteine insertion sequence element required for cotranslational incorporation of selenocysteine. Molecular and Cellular Biology, 17(4), 1977–1985.

Levander, O. A., & Beck, M. A. (1999). Selenium and viral virulence. British Medical Bulletin, 55(3), 528–533. https://doi.org/10.1258/0007142991902592

Li, M.-L., Hsu, T.-A., Chen, T.-C., Chang, S.-C., Lee, J.-C., Chen, C.-C., … Shih, S.-R. (2002). The 3C protease activity of enterovirus 71 induces human neural cell apoptosis. Virology, 293(2), 386–395. https://doi.org/10.1006/viro.2001.1310

Li, Z., Tiley, G. P., Galuska, S. R., Reardon, C. R., Kidder, T. I., Rundell, R. J., & Barker, M. S. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. Proceedings of the National Academy of Sciences, 115(18), 4713–4718. https://doi.org/10.1073/pnas.1710791115

Lin, J.-Y., Brewer, G., & Li, M.-L. (2015). HuR and Ago2 Bind the Internal Ribosome Entry Site of Enterovirus 71 and Promote Virus Translation and Replication. PLOS ONE, 10(10), e0140291. Retrieved from https://doi.org/10.1371/journal.pone.0140291

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. Structure (London, England : 1993), 11(11), 1453–1459.

Lobanov, A. V., Hatfield, D. L., & Gladyshev, V. N. (2008). Reduced reliance on the trace element selenium during evolution of mammals. Genome Biology, 9(3). https://doi.org/10.1186/gb-2008-9-3-r62

Mariotti, M., Lobanov, A. V, Guigo, R., & Gladyshev, V. N. (2013). SECISearch3 and Seblastian: new tools for prediction of SECIS elements and selenoproteins. Nucleic Acids Research, 41(15), e149. https://doi.org/10.1093/nar/gkt550

Mariotti, M., Lobanov, A. V, Manta, B., Santesmasses, D., Bofill, A., Guigo, R., … Gladyshev, V. N. (2016). Lokiarchaeota Marks the Transition between the Archaeal and Eukaryotic Selenocysteine Encoding Systems. Molecular Biology and Evolution, 33(9), 2441–2453. https://doi.org/10.1093/molbev/msw122

Mariotti, M., Ridge, P. G., Zhang, Y., Lobanov, A. V, Pringle, T. H., Guigo, R., … Gladyshev, V. N. (2012). Composition and evolution of the vertebrate and mammalian selenoproteomes. PloS One, 7(3), e33066. https://doi.org/10.1371/journal.pone.0033066

Mariotti, M., Salinas, G., Gabaldón, T., & Gladyshev, V. N. (2019). Utilization of selenocysteine in early-branching fungal phyla. Nature Microbiology. https://doi.org/10.1038/s41564-018-0354-9

Mariotti, M., Santesmasses, D., Capella-Gutierrez, S., Mateo, A., Arnan, C., Johnson, R., … Guigo, R. (2015). Evolution of selenophosphate synthetases: emergence and relocation of function through independent duplications and recurrent subfunctionalization. Genome Research, 25(9), 1256–1267. https://doi.org/10.1101/gr.190538.115

Mariotti, M., Shetty, S., Baird, L., Sen, W., Loughran, G., Copeland, P. R., … Howard, M. T. (2017). Multiple RNA structures affect translation initiation and UGA redefinition efficiency during synthesis of selenoprotein P. Nucleic Acids Research, 45(22), 13004–13015. https://doi.org/10.1093/nar/gkx982

Mehta, A., Rebsch, C. M., Kinzy, S. A., Fletcher, J. E., & Copeland, P. R. (2004). Efficiency of mammalian selenocysteine incorporation. The Journal of Biological Chemistry, 279(36), 37852–37859. https://doi.org/10.1074/jbc.M404639200

Nauser, T., Steinmann, D., Grassi, G., and Koppenol, W. H. (2014) Why selenocysteine replaces cysteine in thioredoxin reductase: a radical hypothesis. Biochemistry 53, 5017−5022.

Negrutskii, B. S., & Deutscher, M. P. (1991). Channeling of aminoacyl-tRNA for protein synthesis in vivo. Proceedings of the National Academy of Sciences of the United States of America, 88(11), 4991–4995. https://doi.org/10.1073/pnas.88.11.4991

Noble, C.G. & Song, H. Cell. Mol. Life Sci. (2008) 65: 1335. https://doi.org/10.1007/s00018-008-7495-6

Ohno S. (1970). Evolution by gene duplication. Springer-Verlag. ISBN 0-04-575015-7.

Oldfield, J. E. (1987). The two faces of selenium. The Journal of Nutrition, 117(12), 2002–2008. https://doi.org/10.1093/jn/117.12.2002

Oldfield, J. E. (1995). SERENDIPITY. Chemtech, 25(3), 52–55.

Oldfield, J. E. (2002). A brief history of selenium research: From alkali disease to prostate cancer (from poison to prevention). American Society of Animal Science, 1–4. Retrieved from https://www.asas.org/docs/publications/oldfieldhist.pdf?sfvrsn=0

Oliéric, V., Wolff, P., Takeuchi, A., Bec, G., Birck, C., Vitorino, M., … Dumas, P. (2009). SECIS-binding protein 2, a key player in selenoprotein synthesis, is an intrinsically disordered protein. Biochimie, 91(8), 1003–1009. https://doi.org/https://doi.org/10.1016/j.biochi.2009.05.004

Osoegawa, K., Jong, P. De, Grimwood, J., Chapman, J. A., Grigoriev, V., Lindberg, D. R., … Nicholas, H. (2014). Insights into bilaterian evolution from three spiralian genomes. Nature, 493(7433), 526–531. https://doi.org/10.1038/nature11696.Insights

Paleskava, A., Konevega, A. L., & Rodnina, M. V. (2012). Thermodynamics of the GTP-GDP-operated Conformational Switch of Selenocysteine-specific Translation Factor SelB. Journal of Biological Chemistry , 287(33), 27906–27912. https://doi.org/10.1074/jbc.M112.366120

Papp, L. V, Wang, J., Kennedy, D., Boucher, D., Zhang, Y., Gladyshev, V. N., … Khanna, K. K. (2008). Functional characterization of alternatively spliced human SECISBP2 transcript variants. Nucleic Acids Research, 36(22), 7192–7206. https://doi.org/10.1093/nar/gkn829

Papp, L. V., Lu, J., Striebel, F., Kennedy, D., Holmgren, A., & Khanna, K. K. (2006). The Redox State of SECIS Binding Protein 2 Controls Its Localization and Selenocysteine Incorporation Function. Molecular and Cellular Biology, 26(13), 4895–4910. https://doi.org/10.1128/mcb.02284-05

Patterson, E. L., R. Milstrey, and E. L. R. Stokstad. 1957. Effect of selenium in preventing exudative diathesis in chicks. Proc. Soc. Exp. Exp. Biol. Med. 95:617–620.

Peng, Z., Oldfield, C. J., Xue, B., Mizianty, M. J., Dunker, A. K., Kurgan, L., & Uversky, V. N. (2014). A creature with a hundred waggly tails: Intrinsically disordered proteins in the ribosome. Cellular and Molecular Life Sciences, 71(8), 1477–1504. https://doi.org/10.1007/s00018-013-1446-6

Pilon-smits, E. A. H., & Quinn, C. F. (2010). Cell Biology of Metals and Nutrients, 17. https://doi.org/10.1007/978-3-642-10613-2

Pinsent J. (1954). The need for selenite and molybdate in the formation of formic dehydrogenase by members of the coli-aerogenes group of bacteria. The Biochemical journal, 57(1), 10–16. doi:10.1042/bj0570010

Piovesan, D., Tabaro, F., Paladin, L., Necci, M., Mieti, I., Camilloni, C., … Tosatto, S. C. E. (2018). MobiDB 3.0: More annotations for intrinsic disorder, conformational diversity and interactions in proteins. Nucleic Acids Research, 46(D1), D471–D476. https://doi.org/10.1093/nar/gkx1071

Pongratz, I., Mason, G. G. F., & Poellinger, L. (1992). Dual roles of the 90-kDa heat shock protein hsp90 in modulating functional activities of the dioxin receptor. Evidence that the dioxin receptor functionally belongs to a subclass of nuclear receptors which require hsp90 both for ligand binding activity an. Journal of Biological Chemistry, 267(19), 13728–13734.

Razga, F., Koca, J., Sponer, J., & Leontis, N. B. (2005). Hinge-like motions in RNA kink-turns: the role of the second a-minor motif and nominally unpaired bases. Biophysical Journal, 88(5), 3466–3485. https://doi.org/10.1529/biophysj.104.054916

Reddy, R. K., Kurek, I., Silverstein, A. M., Chinkers, M., Breiman, A., & Krishna, P. (2002). High-Molecular-Weight FK506-Binding Proteins Are Components of Heat-Shock Protein 90 Heterocomplexes in Wheat Germ Lysate. Plant Physiology, 118(4), 1395–1401. https://doi.org/10.1104/pp.118.4.1395

Reddy, R. K., Kurek, I., Silverstein, A. M., Chinkers, M., Breiman, A., & Krishna, P. (2002). High-Molecular-Weight FK506-Binding Proteins Are Components of Heat-Shock Protein 90 Heterocomplexes in Wheat Germ Lysate. Plant Physiology, 118(4), 1395–1401. https://doi.org/10.1104/pp.118.4.1395

Reich, H. J., & Hondal, R. J. (2016). Why Nature Chose Selenium. ACS Chemical Biology, 11(4), 821–841. https://doi.org/10.1021/acschembio.6b00031

Reshi, M. L., Su, Y.-C., & Hong, J.-R. (2014). RNA Viruses: ROS-Mediated Cell Death. International Journal of Cell Biology, 2014, 467452. https://doi.org/10.1155/2014/467452

Rokov-Plavec, J., Bilokapić, S., Gruic-Sovulj, I., Močibob, M., Glavan, F., Brgles, M., & Weygand-Durašević, I. (2004). Unilateral flexibility in tRNASer recognition by heterologous seryl-tRNA synthetases. Periodicum Biologorum (Vol. 106).

Roll-Mecak, A., Cao, C., Dever, T. E., & Burley, S. K. (2000). X-Ray Structures of the Universal Translation Initiation Factor IF2/eIF5B: Conformational Changes on GDP and

GTP Binding. Cell, 103(5), 781–792. https://doi.org/https://doi.org/10.1016/S0092-8674(00)00181-1

Romagne, F., Santesmasses, D., White, L., Sarangi, G. K., Mariotti, M., Hubler, R., … Castellano, S. (2014). SelenoDB 2.0: annotation of selenoprotein genes in animals and their genetic diversity in humans. Nucleic Acids Research, 42(Database issue), D437-43. https://doi.org/10.1093/nar/gkt1045

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Santesmasses D, Mariotti M, Guigó R (2017) Computational identification of the selenocysteine tRNA (tRNASec) in genomes. PLoS Comput Biol 13(2): e1005383. https://doi.org/10.1371/journal.pcbi.1005383

Santesmasses, D., Mariotti, M., & Guigo, R. (2018). Selenoprofiles: A Computational Pipeline for Annotation of Selenoproteins. Methods in Molecular Biology (Clifton, N.J.), 1661, 17–28. https://doi.org/10.1007/978-1-4939-7258-6_2

Schoenmakers, E., Agostini, M., Mitchell, C., Schoenmakers, N., Papp, L., Rajanayagam, O., … Chatterjee, K. (2010). Mutations in the selenocysteine insertion sequence-binding protein 2 gene lead to a multisystem selenoprotein deficiency disorder in humans. The Journal of Clinical Investigation, 120(12), 4220–4235. https://doi.org/10.1172/JCI43653

Schoenmakers, E., Schoenmakers, N., & Chatterjee, K. (2016). Mutations in Humans That Adversely Affect the Selenoprotein Synthesis Pathway. In Selenium: Its Molecular Biology and Role in Human Health: Fourth Edition. https://doi.org/10.1007/978-3-319-41283-2_44

Schumacher, R. J., Hurst, R., Sullivan, W. P., McMahon, N. J., Toft, D. O., & Matts, R. L. (1994). ATP-dependent chaperoning activity of reticulocyte lysate. Journal of Biological Chemistry, 269(13), 9493–9499.

Schwarz, K., & Foltz, C. M. (1957). SELENIUM AS AN INTEGRAL PART OF FACTOR 3 AGAINST DIETARY NECROTIC LIVER DEGENERATION. Journal of the American Chemical Society, 79(12), 3292–3293. https://doi.org/10.1021/ja01569a087

Seeher, S., & Schweizer, U. (2014). Targeted deletion of Secisbp2 reduces, but does not abrogate, selenoprotein expression and leads to striatal interneuron loss. Free Radical Biology & Medicine, 75 Suppl 1, S9. https://doi.org/10.1016/j.freeradbiomed.2014.10.849

Seeher, S., Atassi, T., Mahdi, Y., Carlson, B. A., Braun, D., Wirth, E. K., … Schweizer, U. (2014). Secisbp2 is essential for embryonic development and enhances selenoprotein

expression. Antioxidants & Redox Signaling, 21(6), 835–849. https://doi.org/10.1089/ars.2013.5358

Shepherdley, C. A., Klootwijk, W., Makabe, K. W., Visser, T. J., & Kuiper, G. G. J. M. (2004). An Ascidian Homolog of Vertebrate Iodothyronine Deiodinases. Endocrinology, 145(3), 1255–1268. https://doi.org/10.1210/en.2003-1248

Shetty, S. P., Shah, R., & Copeland, P. R. (2014). Regulation of selenocysteine incorporation into the selenium transport protein, selenoprotein P. The Journal of Biological Chemistry, 289(36), 25317–25326. https://doi.org/10.1074/jbc.M114.590430

Shetty, S. P., Sturts, R., Vetick, M., & Copeland, P. R. (2018). Processive incorporation of multiple selenocysteine residues is driven by a novel feature of the selenocysteine insertion sequence. Journal of Biological Chemistry, 293(50), 19377–19386. https://doi.org/10.1074/jbc.RA118.005211

Sivaram, P., & Deutscher, M. P. (1990). Existence of two forms of rat liver arginyl-tRNA synthetase suggests channeling of aminoacyl-tRNA for protein synthesis. Proceedings of the National Academy of Sciences of the United States of America, 87(10), 3665–3669. https://doi.org/10.1073/pnas.87.10.3665

Snider, G. W., Ruggles, E. L., Khan, N., and Hondal, R. J. (2013) Selenocysteine confers resistance to inactivation by oxidation in thioredoxin reductase: Comparison of selenium and sulfur enzymes. Biochemistry

Souvorov, A., Kapustin, Y., Kiryutin, B., Chetvernin, V., Tatusova, T., & Lipman, D. (2012). Gnomon – NCBI eukaryotic gene prediction.

Stapulionis, R., & Deutscher, M. P. (1995). A channeled tRNA cycle during mammalian protein synthesis. Proceedings of the National Academy of Sciences of the United States of America, 92(16), 7158–7161. https://doi.org/10.1073/pnas.92.16.7158

Steinmann, D., Nauser, T., Beld, J., Tanner, M., Gunther, D., Bounds, P. L., & Koppenol, W. H. (2008). Kinetics of tyrosyl radical reduction by selenocysteine. Biochemistry, 47(36), 9602–9607. https://doi.org/10.1021/bi801029f

Stone, C. A., Kawai, K., Kupka, R., & Fawzi, W. W. (2010). Role of selenium in HIV infection. Nutrition Reviews, 68(11), 671–681. https://doi.org/10.1111/j.1753-4887.2010.00337.x

Stoytcheva, Z., Tujebajeva, R. M., Harney, J. W., & Berry, M. J. (2006). Efficient incorporation of multiple selenocysteines involves an inefficient decoding step serving as a potential translational checkpoint and ribosome bottleneck. Molecular and Cellular Biology, 26(24), 9177–9184. https://doi.org/10.1128/MCB.00856-06

Sweeney, T. R., Abaeva, I. S., Pestova, T. V., & Hellen, C. U. T. (2014). The mechanism of translation initiation on type 1 picornavirus IRESs. EMBO Journal, 33(1), 76–92. https://doi.org/10.1002/embj.201386124

Tang, H., & Thomas, P. D. (2018). Inference of gene loss rates after whole genome duplications at early vertebrates through ancient genome reconstructions. Retrieved from http://arxiv.org/abs/1802.07201

Tompa, P., Szasz, C., & Buday, L. (2005). Structural disorder throws new light on moonlighting. Trends in Biochemical Sciences, 30(9), 484–489. https://doi.org/10.1016/j.tibs.2005.07.008

Trofast, J. (2011). Berzelius' Discovery of Selenium. Chemistry International, 33, 16–19.

Tsuji, P. A., Carlson, B. A., Anderson, C. B., Seifried, H. E., Hatfield, D. L., & Howard, M. T. (2015). Dietary Selenium Levels Affect Selenoprotein Expression and Support the Interferon-gamma and IL-6 Immune Response Pathways in Mice. Nutrients, 7(8), 6529–6549. https://doi.org/10.3390/nu7085297

Tujebajeva, R. M., Copeland, P. R., Xu, X. M., Carlson, B. A., Harney, J. W., Driscoll, D. M., … Berry, M. J. (2000). Decoding apparatus for eukaryotic selenocysteine insertion. EMBO Reports, 1(2), 158–163. https://doi.org/10.1038/sj.embor.embor604

Tung, W. S., Bakar, S. A., Sekawi, Z., & Rosli, R. (2007). DNA vaccine constructs against enterovirus 71 elicit immune response in mice. Genetic Vaccines and Therapy, 5, 6. https://doi.org/10.1186/1479-0556-5-6

Turanov, A. A., Lobanov, A. V, Hatfield, D. L., & Gladyshev, V. N. (2013). UGA codon position-dependent incorporation of selenocysteine into mammalian selenoproteins. Nucleic Acids Research, 41(14), 6952–6959. https://doi.org/10.1093/nar/gkt409

Turner DC, Stadtman TC. Purification of protein components of the clostridial glycine reductase system and characterization of protein A as a selenoprotein. Arch Biochem Biophys 1973;154:366–381.

Varadi, M., Zsolyomi, F., Guharoy, M., Tompa, P., & Levy, Y. K. (2015). Functional advantages of conserved intrinsic disorder in RNA-binding proteins. PLoS ONE, 10(10), 1–16. https://doi.org/10.1371/journal.pone.0139731

Vilardell, J., Yu, S. J., & Warner, J. R. (2000). Multiple functions of an evolutionarily conserved RNA binding domain. Molecular Cell, 5(4), 761–766.

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

Webster, B. L., Copley, R. R., Jenner, R. A., Mackenzie-Dodds, J. A., Bourlat, S. J., Rota-Stabelli, O., … Telford, M. J. (2006). Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral Ecdysozoan. Evolution and Development, 8(6), 502–510. https://doi.org/10.1111/j.1525-142X.2006.00123.x

Weeks, M. E. (1932). The discovery of the elements. VI. Tellurium and selenium. Journal of Chemical Education, 9(3), 474. https://doi.org/10.1021/ed009p474

Weiss, S. L., & Sunde, R. A. (1998). Cis-acting elements are required for selenium regulation of glutathione peroxidase-1 mRNA levels. RNA (New York, N.Y.), 4(7), 816–827.

Wimmer, E., Hellen, C. U., & Cao, X. (1993). Genetics of poliovirus. Annual Review of Genetics, 27, 353–436. https://doi.org/10.1146/annurev.ge.27.120193.002033

Wu, S., Mariotti, M., Santesmasses, D., Hill, K. E., Baclaocos, J., Aparicio-prat, E., … Atkins, J. F. (2016). Human selenoprotein P and S variant mRNAs with different numbers of SECIS elements and inferences from mutant mice of the roles of multiple SECIS elements. Open Biology.

Wu, X. Q., & Gross, H. J. (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 Research, 21(24), 5589–5594. https://doi.org/10.1093/nar/21.24.5589

Xu, X.-M., Carlson, B. A., Mix, H., Zhang, Y., Saira, K., Glass, R. S., … Hatfield, D. L. (2007). Biosynthesis of selenocysteine on its tRNA in eukaryotes. PLoS Biology, 5(1), e4. https://doi.org/10.1371/journal.pbio.0050004

Xu, X.-M., Turanov, A. A., Carlson, B. A., Yoo, M.-H., Everley, R. A., Nandakumar, R., … Hatfield, D. L. (2010). Targeted insertion of cysteine by decoding UGA codons with mammalian selenocysteine machinery. Proceedings of the National Academy of Sciences of the United States of America, 107(50), 21430–21434. https://doi.org/10.1073/pnas.1009947107

Yang, J.-G., Hill, K. E., & Burk, R. F. (1989). Dietary Selenium Intake Controls Rat Plasma Selenoprotein P Concentration. The Journal of Nutrition, 119(7), 1010–1012. https://doi.org/10.1093/jn/119.7.1010

Yeh, R. F., Lim, L. P., & Burge, C. B. (2001). Computational inference of homologous gene structures in the human genome. Genome Research, 11(5), 803–816. https://doi.org/10.1101/gr.175701

Zhang, B., Wu, X., Huang, K., Li, L., Zheng, L., Wan, C., … Zhao, W. (2014). The variations of VP1 protein might be associated with nervous system symptoms caused by

enterovirus 71 infection. BMC Infectious Diseases, 14, 243. https://doi.org/10.1186/1471-2334-14-243

Zhu, F., Xu, W., Xia, J., Liang, Z., Liu, Y., Zhang, X., … Wang, N. (2014). Efficacy, Safety, and Immunogenicity of an Enterovirus 71 Vaccine in China. New England Journal of Medicine, 370(9), 818–828. https://doi.org/10.1056/NEJMoa1304923