

Spring 5-2010

RNA-dependent selenocysteine biosynthesis in eukaryotes and Archaea

Sotiria Palioura
Yale University.

Follow this and additional works at: <http://elischolar.library.yale.edu/ymtdl>



Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Palioura, Sotiria, "RNA-dependent selenocysteine biosynthesis in eukaryotes and Archaea" (2010). *Yale Medicine Thesis Digital Library*. 2253.

<http://elischolar.library.yale.edu/ymtdl/2253>

This Open Access Dissertation is brought to you for free and open access by the School of Medicine at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Yale Medicine Thesis Digital Library by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

RNA-dependent Selenocysteine Biosynthesis in Eukaryotes and Archaea

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Sotiria Palioura

Dissertation Director: Dieter G. Söll

May 2010

Abstract

RNA-dependent Selenocysteine Biosynthesis in Eukaryotes and Archaea

Sotiria Palioura

Yale University

2010

Selenocysteine (Sec), the 21st genetically encoded amino acid, is the major metabolite of the micronutrient selenium. Sec is inserted into nascent proteins in response to a UGA codon. The substrate for ribosomal protein synthesis is selenocysteinyl-tRNA^{Sec}. While the formation of Sec-tRNA^{Sec} from seryl-tRNA^{Sec} by a single bacterial enzyme selenocysteine synthase (SelA) has been well described, the mechanism of Sec-tRNA^{Sec} formation in archaea and eukaryotes remained poorly understood. Herein, biochemical and genetic data provide evidence that, in contrast to bacteria, eukaryotes and archaea utilize a different route to Sec-tRNA^{Sec} that requires the tRNA^{Sec}-dependent conversion of O-phosphoserine (Sep) to Sec. In this two-step pathway, O-phosphoseryl-tRNA kinase (PSTK) first converts Ser-tRNA^{Sec} to Sep-tRNA^{Sec}. This misacylated tRNA is the obligatory precursor for a Sep-tRNA: Sec-tRNA synthase (SepSecS); this protein was previously annotated as Soluble Liver Antigen/Liver Pancreas (SLA/LP). SepSecS genes from *Homo sapiens*, the lower eukaryote *Trypanosoma brucei* and the archaea *Methanocaldococcus jannaschii* and *Methanococcus maripaludis* complement an *Escherichia coli* $\Delta selA$ deletion strain *in vivo*. Furthermore, genetic analysis of selenoprotein biosynthesis in *T. brucei in vivo* demonstrated that eukaryotes have a single pathway to

Sec-tRNA^{Sec} that requires Sep-tRNA^{Sec} as an intermediate. Finally, purified recombinant SepSecS converts Sep-tRNA^{Sec} into Sec-tRNA^{Sec} *in vitro* in the presence of sodium selenite and purified *E. coli* selenophosphate synthetase.

The final step in Sec biosynthesis was further investigated by a structure-based mutational analysis of the *M. maripaludis* SepSecS and by determining the crystal structure of human SepSecS complexed with tRNA^{Sec}, phosphoserine and thiophosphate at 2.8 Å resolution. *In vivo* and *in vitro* enzyme assays support a mechanism of Sec-tRNA^{Sec} formation based on pyridoxal phosphate, while the lack of active site cysteines demonstrates that a perselenide intermediate is not involved in SepSecS-catalyzed Sec formation. Two tRNA^{Sec} molecules, with a fold distinct from other canonical tRNAs, bind to each human SepSecS tetramer through their unique 13 base-pair acceptor-TΨC arm. The tRNA binding induces a conformational change in the enzyme's active site that allows a Sep covalently attached to tRNA^{Sec}, but not free Sep, to be oriented properly for the reaction to occur.

ACKNOWLEDGEMENTS

This dissertation completes an 11-year cycle at Yale University since I first joined Yale College in August 1999. A completed dissertation, though, cannot be and is not the work of one person alone. It is rather, a collective effort of a group of individuals that includes advisors, co-authors, family, friends, relatives, and colleagues amongst others. To all the people that supported me at various points during this process, I am and will always remain grateful.

First and foremost, I would like to thank my PhD advisor, Dieter Söll, for his never-ending encouragement and support since my first experience in his laboratory as an undergraduate in 2001. I feel very privileged to have trained in science under Dieter's guidance and inspiration. Dieter has always been extremely generous in sharing his insight and knowledge with me not only about science but also about life. Being always humble, he has taught me the elegance of research, the joy of discovering a new insight, and the true ethos of a scholar. Although my road to this first «Ithaca » has not been very long, it was certainly « full of adventure, full of knowledge ». I primarily owe this to Dieter and I aspire to be an equally good mentor in the future as the example he has set for me.

I would like to thank my research committee, Susan Baserga and Gary Brudvig, for their patience, support, insightful critiques and thoughtful guidance. I would also like to thank the director of the MD/PhD program at Yale University School of Medicine, James Jamieson, for his time, advice and shrewd guidance.

The first part of this work was done in collaboration with Jing Yuan, whom I thank for her patience and help all along. The trypanosomal work was done in collaboration with Eric Aeby in the laboratory of André Schneider (University of Bern, Bern, Switzerland) and Elisabetta Ullu (Yale University, New Haven, CT), whom I thank for introducing me to trypanosomal biochemistry. The archaeal SepSecS structure was determined by Yuhei Araiso in the laboratory of Osamu Nureki (University of Tokyo, Tokyo, Japan). I thank Miljan Simonović for introducing me to X-ray crystallography and for his guidance and support during the last part of my thesis work. I would also like to thank Tom Steitz for his insight into the human SepSecS-tRNA^{Sec} complex structure. I am indebted to my fellow colleagues in the Söll lab, Anselm Sauerwald, Juan

Salazar, Alex Ambrogelly, Lynn Sherrer, Markus Englert, Michael Hohn, Patrick O'Donoghue, Theodoros Rampias, Dan Su and Kelly Sheppard, and to Gregor Blaha, Axel Innis and Bill Eliason in the Steitz lab for making this process an educational and enjoyable one.

Outside the lab, I especially would like to thank the circle of my closest friends – Ioannis Ioannou, Roula Tavoulari, Amanda Psyrris, Nasos Roussias, Zoe Cournia, Antonios Finitis, Tasos Gogakos, Dimitrios Zattas, Haralampos Stratigopoulos and Michael Gousgounis – who have been by my side every single time I needed them without thinking twice. They have been my closest “family” in the United States. I also sincerely thank all the great friends that I made here, in New Haven, for their help and support: Dario Englot, Eric Huebner, Mike Martinez, Yanni Hatzaras, George Galanis, Spyros Alogoskoufis, Olia Galenianou, Marios Panayides, Daniela Donno-Panayides, Andira Hernandez, Eleni Gousgounis, Irene Hatzidimitriou, Nikos Angelopoulos, Nicola Santoro, Ebe D’Adamo, Flavius Schackert, Mihalis Maniatakos and Antonis Stamboulis.

Lastly and most importantly, I would like to thank my family, and in particular, my father Dimitris, my mother Elisavet, and my sister Vassiliki, as well as my uncle Manthos and my aunt Zoe, for their patience throughout the years, for believing in me since day one, and without whom I could have never made it this far.

I am thankful and humbled to be surrounded by all these amazing people and I can only hope that each and every one of them will continue to support and challenge me in the years ahead.

To my beloved sister, Vassiliki

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
PUBLICATIONS	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xiii
INTRODUCTION.....	1
A. Selenocysteine, a versatile amino acid.....	1
B. Selenocysteine, the 21st genetically encoded amino acid.....	4
C. Selenocysteine is the universal exception to Crick’s paradigm.....	6
D. Biosynthesis of selenocysteine and UGA recoding in bacteria.....	9
E. Selenocysteine biosynthesis in eukaryotes and archaea: the last riddle in aminoacyl-tRNA synthesis.....	10
F. Scope of Dissertation.....	15
1. Solving the riddle of selenocysteine formation in eukaryotes and archaea.....	15
2. The PSTK/SepSecS pathway is the only route to Sec in Trypanosomes, and likely in all eukaryotes.....	16
3. Insights into archaeal Sec synthesis from the structure of <i>M. maripaludis</i> SepSecS....	17
4. The mechanism of tRNA ^{Sec} recognition and Sec formation by the human SepSecS..	19
MATERIALS AND METHODS.....	21
A. General.....	21
B. Bacterial strains and plasmids.....	21
C. <i>In vivo</i> SepSecS assay.....	23
D. Metabolic labeling with [⁷⁵ Se]selenite.....	24
E. Protein expression and purification.....	24
F. Preparation and purification of tRNA gene transcripts.....	25
G. Purification and crystallization of the SepSecS-tRNA ^{Sec} complex.....	27
H. Ligand soaks and data collection.....	27
I. Preparation of ³² P-labeled Sep-tRNA ^{Sec}	28
J. <i>In vitro</i> conversion of Sep-tRNA ^{Sec} to Cys-tRNA ^{Sec}	29
RESULTS.....	32
A. The missing step of selenocysteine biosynthesis in archaea and eukaryotes.....	32

1. SepSecS rescues selenoprotein biosynthesis in an <i>E. coli</i> Δ selA deletion strain.....	32
2. SepSecS converts Sep-tRNA ^{Sec} to Sec-tRNA ^{Sec} <i>in vitro</i>	34
B. <i>In vivo</i> evidence for a single pathway of Sec-tRNA ^{Sec} formation in <i>T. brucei</i>	36
C. Structural insights into archaeal RNA-dependent selenocysteine biosynthesis.....	40
1. Experimental approach.....	40
2. Overall structure of SepSecS from <i>Methanococcus maripaludis</i>	40
3. Archaeal SepSecS is active only as a tetramer.....	41
4. Active site mutants define PLP recognition.....	45
5. The active site.....	52
D. The human SepSecS-tRNA ^{Sec} complex.....	55
1. SepSecS can bind tRNA ^{Sec} only as a tetramer.....	55
2. The antigenic region.....	61
3. The unique 9/4 fold of tRNA ^{Sec} is recognized by the Sec-specific proteins.....	63
4. Sep binds to the active site only in the presence of tRNA ^{Sec}	67
5. SepSecS utilizes a PLP-dependent mechanism for Sec-tRNA ^{Sec} formation.....	73
DISCUSSION.....	76
A. SepSecS: the missing step in Sec biosynthesis in eukaryotes and archaea.....	76
B. The indirect RNA-dependent PSTK/SepSecS pathway is the only route to Sec.....	78
C. <i>M. maripaludis</i> SepSecS structure: insights into the reaction mechanism.....	79
D. The human SepSecS-tRNA ^{Sec} complex reveals the mechanism of Sec formation.....	79
REFERENCES.....	84
APPENDIX.....	94

PUBLICATIONS (* indicates co-first author)

1. **Palioura, S.**, Herkel, J., Simonović, M., Lohse, A., and Söll, D. (2010). Human SepSecS or SLA/LP: selenocysteine formation and autoimmune hepatitis. *Biol Chem (in press)*.
2. Yuan, J., O'Donoghue, P., Ambrogelly, A., Gundllapalli, S., Sherrer, R.L., **Palioura, S.**, Simonović, M., and Söll, D. (2010). Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett* 584, 342-349.
3. **Palioura, S.**, Sherrer, R.L., Steitz, T.A., Söll, D., and Simonović, M. (2009). The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. *Science* 325, 321-325.
4. Aeby, E., **Palioura, S.**, Pusnik, M., Marazzi, J., Lieberman, A., Ullu, E., Söll, D., and Schneider, A. (2009). The canonical pathway for selenocysteine insertion is dispensable in *Trypanosomes*. *Proc Natl Acad Sci U S A*. 106, 5088-92.
5. Su, D., Hohn, M.J., **Palioura, S.**, Sherrer, R.L., Yuan, J., Söll, D., and O'Donoghue, P. (2008). How an obscure archaeal gene inspired the discovery of selenocysteine biosynthesis in humans. *IUBMB Life* 61, 35-9.
6. Araiso, Y.*, **Palioura, S.***, Ishitani, R., Sherrer, R. L., Yuan, J., Oshikane, H., Domae, N., DeFranco, J., Söll, D., and Nureki, O. (2008). Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic Acids Res* 36, 1187-99.
7. Ambrogelly, A., **Palioura, S.**, and Söll, D. (2007). Natural expansion of the genetic code. *Nat Chem Biol* 3, 29-35.
8. Yuan, J.*, **Palioura, S.***, Salazar, J.C., Su, D., O'Donoghue, P., Hohn, M.J., Cardoso, A.M., Whitman, W.B., and Söll, D. (2006). RNA-dependent conversion of phosphoserine forms seleno-cysteine in eukaryotes and archaea. *Proc Natl Acad Sci USA* 103, 18923-18927.

LIST OF FIGURES

FIGURES		Page
Figure 1	Chemical structures of cysteine and selenocysteine	2
Figure 2	The genetic code	5
Figure 3	Aminoacyl-tRNA synthesis	8
Figure 4	The bacterial pathway to selenoprotein synthesis	11
Figure 5	The archaeal and eukaryal pathway to selenocysteine and the methanogenic pathway to cysteinyl-tRNA	14
Figure 6	Reaction schemes for SepSecS and related enzymes	18
Figure 7	<i>In vivo</i> complementation assay for archaeal and human SepSecS	33
Figure 8	<i>In vivo</i> complementation assay for <i>Trypanosoma brucei</i> SepSecS	35
Figure 9	<i>In vitro</i> assay for archaeal and human SepSecS	37
Figure 10	Selenoprotein expression in <i>T. brucei</i>	39
Figure 11	Overall architectures of archaeal SepSecS and related enzymes	42
Figure 12	Sequence alignment of SepSecS and related enzymes	43
Figure 13	The oligomeric states of archaeal SepSecS and related enzymes	44
Figure 14	<i>In vivo</i> assays of archaeal SepSecS mutants	46
Figure 15	PLP recognition in the archaeal SepSecS structure	47
Figure 16	<i>In vitro</i> assays for archaeal SepSecS mutants	49
Figure 17	Complementation assays of archaeal SepSecS mutants	50
Figure 18	The active site in archaeal SepSecS and related enzymes	53
Figure 19	Structure of human SepSecS in complex with unacylated tRNA ^{Sec}	56
Figure 20	Interactions between human SepSecS and tRNA ^{Sec}	58
Figure 21	<i>In vivo</i> assays of human SepSecS mutants	60
Figure 22	The proposed antigenic region in SepSecS or SLA/LP	62
Figure 23	Structure of human tRNA ^{Sec} compared to the canonical tRNA ^{Ser}	64
Figure 24	Selectivity of SepSecS towards tRNA ^{Sec}	66
Figure 25	Modeling of the SerRS-tRNA ^{Sec} complex	68
Figure 26	Ligand binding to the active sites in the SepSecS-tRNA ^{Sec} complex	70

Figure 27	Phosphoserine binding modes in the SepSecS-tRNA ^{Sec} complex	71
Figure 28	<i>In vitro</i> assays of wild type and PLP-cross-linked human SepSecS	74
Figure 29	The PLP-dependent mechanism of Sep to Sec conversion	75
Figure 30	Selenocysteine biosynthesis in eukaryotes	77

LIST OF ABBREVIATIONS

Aminoacyl-tRNA	aa-tRNA
Aminoacyl-tRNA synthetase	aaRS
<i>Archaeoglobus fulgidus</i> SepCysS	AFSepCysS
Benzyl viologen	BV
Cysteine	Cys
Cysteine desulfurase	IscS
<i>Escherichia coli</i> CsdB	ECCsdB
<i>Escherichia coli</i> IscS	ECIscS
Elongation factor Tu	EF-Tu
Formate dehydrogenase H	FDH _H
Lysine	Lys
<i>Methanococcus maripaludis</i> SepSecS	MMPSepSecS
O-phosphoserine	Sep
O-phosphoseryl-tRNA ^{Sec} kinase	PSTK
Pyridoxal phosphate	PLP
Pyrrolysine	Pyl
Selenocysteine	Sec
Selenocysteine Insertion Sequence	SECIS
Selenocysteine lyase	CsdB
Selenocysteine synthase	SelA
Selenophosphate synthase	SelD
Selenophosphate synthase 2	SPS2
Selenoprotein K	SelK
Selenoprotein T	SelT
Sep-tRNA:Cys-tRNA synthase	SepCysS
Sep-tRNA:Sec-tRNA synthase	SepSecS
Serine	Ser
Soluble Liver Antigen/Liver Pancreas	SLA/LP

Thin Layer Chromatography

Trypanosoma brucei SepSecS

Transfer ribonucleic acid

TLC

Tb-SepSecS

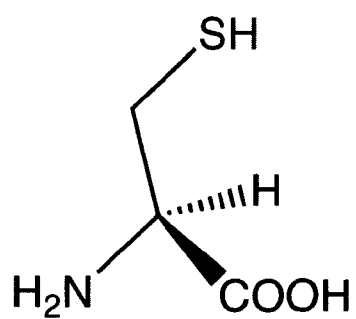
tRNA

INTRODUCTION

A. Selenocysteine, a versatile amino acid

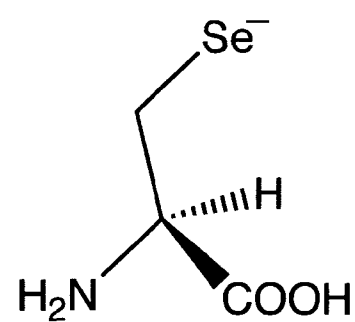
Selenium is a dietary trace element required for animal and human health. The antioxidant properties of selenium are thought to offer protection from diverse human ailments including adverse mood states, cardiovascular disease, viral infections and cancer (Rayman, 2000). Selenium carries out its functions in proteins in the form of selenocysteine (Sec), in which the sulfur-containing thiol group of cysteine has been replaced by the selenium-containing selenol moiety (Figure 1) (Cone et al., 1976). Although selenium and sulfur belong to the same group of elements in the periodic table and hence share some chemical properties (for example, electronegativity and major oxidation state), it is the different electronic structure of these two elements that makes the selenolate anion more stable than the thiolate one and gives selenoproteins their unique selenium-derived catalytic efficiencies. Indeed, the low pKa of selenol relative to a thiol (5.2 versus 8.0, respectively) and selenol's lower redox potential (- 488 mV, versus - 233 mV for the thiol) render it fully ionized at the physiological pH range and causes selenocysteine to be substantially more reactive than cysteine (Figure 1) (Ambrogelly et al., 2007). Thus, it is not surprising that most known selenoproteins are oxidoreductases with an essential (for efficient catalysis) selenocysteine active site residue. A selenocysteine-to-cysteine (Cys) mutation in such cases results in a >100-fold reduction in catalytic turnover (Axley et al., 1991; Johansson et al., 2005).

Biochemical studies and genomic analyses have established that selenocysteine and selenoproteins are found in organisms from all three domains of life – bacteria,



Cysteine

pK_a = 8.0



Selenocysteine

pK_a = 5.2

Figure 1: The chemical structures and ionization states of cysteine and selenocysteine at physiological pH.

archaea and eukaryotes — though not in all species. In fact, most organisms do not use this amino acid. Except for its occurrence in *Chlamydomonas spp.* (Fu et al., 2002) and green algae of the *Ostreococcus* and the *Volvox spp.* (Lobanov et al., 2009), selenocysteine is notably absent from fungi and higher plants. Although the mammalian selenoproteome (Kryukov et al., 2003) consists of more than 20 selenoproteins, only a limited number of prokaryotes contain selenocysteine (Kryukov and Gladyshev, 2004). For instance, within the archaea with known genomes (currently representing 43 genera), selenoproteins are restricted to two genera, *Methanococcus* and *Methanopyrus* (Stock and Rother, 2009).

Selenocysteine is the principal metabolite of selenium in the human body and the means for exerting its various health benefits (Rayman, 2000). Like the rest of the 20 canonical amino acids selenocysteine is delivered to the ribosome for protein synthesis by its cognate transfer RNA (tRNA) molecule, tRNA^{Sec}. Knockout of tRNA^{Sec} in the mouse is embryonic lethal, which underscores the importance of Sec-containing proteins (or selenoproteins) in mammalian development (Bösl et al., 1997). Mutations and polymorphisms in several of the currently known 25 human selenoproteins have been implicated in cancer and diseases of the muscular, nervous, immune and endocrine systems (Bellinger et al., 2009). Most human selenoenzymes (i.e. thioredoxin reductase, glutathione peroxidase, thioredoxin/glutathione reductase and methionine sulfoxide reductase) safeguard the cell from the detrimental effects of reactive oxygen species and regenerate important antioxidants such as vitamin C, vitamin E and coenzyme Q (Lu and Holmgren, 2009). Selenocysteine is also the catalytic residue in iodothyronine

deiodinase, the enzyme that enables circulating thyroid hormone to exert its pro-metabolic actions in peripheral tissues. The erroneous replacement of the active site Sec with the chemically similar serine (Ser) or Cys diminishes the activity of several selenoenzymes (Zhong and Holmgren, 2000; Kuiper et al., 2003).

B. Selenocysteine, the 21st genetically encoded amino acid

Ribosomal protein synthesis is the pathway to most natural proteins. The genetic code serves as the 'alphabet' used by the translation machinery to ensure faithful decoding of the genetic message into the specified amino acid sequence each cellular protein is comprised of. The code is usually represented as the assortment of mRNA codons that specify each amino acid (Figure 2). With the exception of tryptophan and methionine, the code is redundant such that more than one mRNA triplet specifies each canonical amino acid. At the same time it is not ambiguous; each codon has only one meaning since it specifies only one particular amino acid or a translational stop (codons UGA, UAA and UAG). Since the discovery of the code in the mid 1960s (Nirenberg et al., 1965; Söll et al., 1965) the canonical amino acids that participate in protein synthesis across all species were known to be twenty in number. Given the unambiguous nature of the code, the prospect for additional genetically encoded amino acids being discovered seemed very faint.

Nevertheless, in the late 1980s and early 1990s selenocysteine was established as the 21st genetically encoded amino acid and UGA as the only codon with an ambiguous meaning (Zinoni et al., 1986 ; Leinfelder et al., 1988; Söll, 1988). Although the occurrence

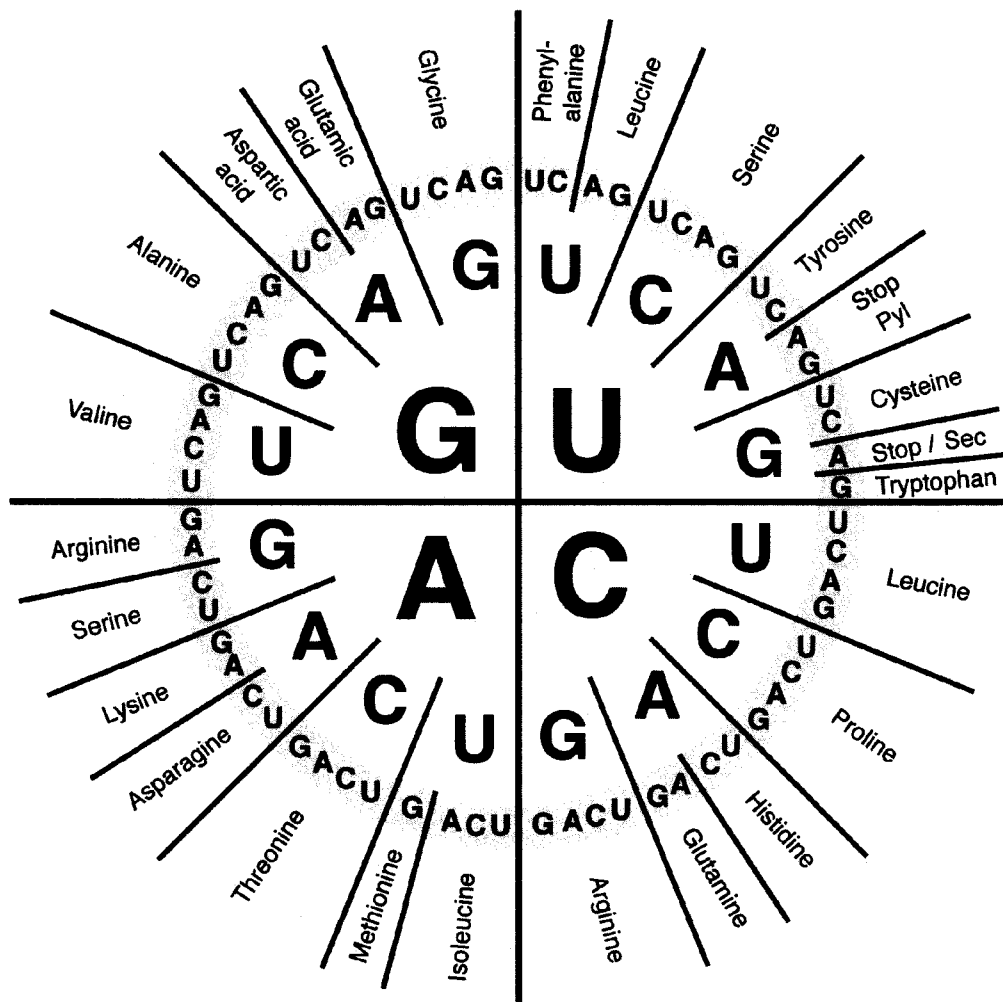


Figure 2: Circular representation of the genetic code (Adapted from Ambrogelly et al., 2007). The genetic code maps each trinucleotide sequence (or codon) found in mRNA to a specific amino acid. The first nucleotide of each codon lies in the inner circle and moving outward are the second and the third ones. The code is composed of 64 codons; 61 specify one of the 20 canonical amino acids and three are stop codons (UGA, UAG, UAA) that indicate translational termination. In selenoprotein-containing organisms UGA also codes for Sec (red). Similarly, in the *Methanosarcinacea* and in *Desulfitobacterium hafniense* UAG also codes for pyrrolysine (Pyl).

of Sec as the selenium moiety in naturally occurring proteins was discovered in 1976 (Cone et al., 1976), Sec was thought to be the result of a post-translational modification event (Forstrom et al., 1978). Apart from its most common function to signal translational termination, an in-frame UGA codon within the coding region of certain proteins also codes for selenocysteine in organisms from all three domains of life (Figure 2). An elegant recoding mechanism allows the translation machinery to accurately discriminate between a Sec UGA and a Stop UGA codon (Böck et al., 2005). The presence of a stem-loop structure in the mRNA of selenoproteins signals the insertion of selenocysteine to the translation apparatus (reviewed in Introduction, section D).

Currently, the repertoire of genetically encoded amino acids has expanded to include 22 members. UAG is also ambiguous in the *Methanosarcinaceae* and in the bacterium *Desulfitobacterium hafriense*, where in addition to serving as a translational stop it also encodes pyrrolysine (Pyl) (Figure 2), the 22nd cotranslationally inserted amino acid (Hao et al., 2002; Srinivasan et al., 2002).

C. Selenocysteine is the universal exception to Crick's adapter hypothesis

Faithful protein synthesis relies on the presence of an aminoacyl-tRNA (aa-tRNA) with an anticodon that matches the mRNA codon by the rules of the wobble hypothesis at the ribosome (Crick, 1966; Söll et al., 1966), and the cognate amino acid esterified to the tRNA's 3'-terminal adenosine (Ibba and Söll, 2000). Thus, synthesis of the correct aminoacyl-tRNA is critical in enabling the genetic code. Aminoacyl-tRNA synthesis is the task of a family of enzymes called aminoacyl-tRNA synthetases (aaRSs); each enzyme

esterifies its cognate amino acid to the corresponding tRNA species (Ibba and Söll, 2000). AARSs ultimately “read” the genetic code since they are responsible for selecting the right amino acid out of the cellular pool of amino acids and pairing it with its cognate tRNA (Figure 3).

According to Crick’s adapter hypothesis postulated 40 years ago, there should be one aaRS for each amino acid (Crick, 1958). Indeed, this is the case for the twenty canonical amino acids and for the most recent addition to the set of genetically encoded amino acids, pyrrolysine. Pyrrolysyl-tRNA synthetase is a newly discovered aaRS, responsible for the synthesis of pyrrolysyl-tRNA (Blight et al., 2004; Polycarpo et al., 2004). A prerequisite for the acylation of a tRNA with its cognate amino acid by the corresponding aaRS is the presence of a cellular pool of free amino acids synthesized prior to this step. Indeed, biochemical experimental efforts of the last century have identified biosynthetic pathways for all 20 canonical amino acids and recently for pyrrolysine as well (Longstaff et al., 2007).

Selenocysteine is the only exception; there is no biosynthetic pathway for free Sec formation within the cell and a selenocysteinyI-tRNA synthetase (SecRS) does not exist (Figure 3). Instead, selenocysteine synthesis occurs on its cognate tRNA in a route that is based on misacylation and subsequent tRNA-dependent amino acid transformation. The selenocysteine-specific tRNA has a UCA anticodon (thus, it pairs with UGA at the ribosome according to the wobble rules) and is denoted as tRNA^{Sec} (Leinfelder et al., 1988). According to common convention in the tRNA field, the superscript (in this

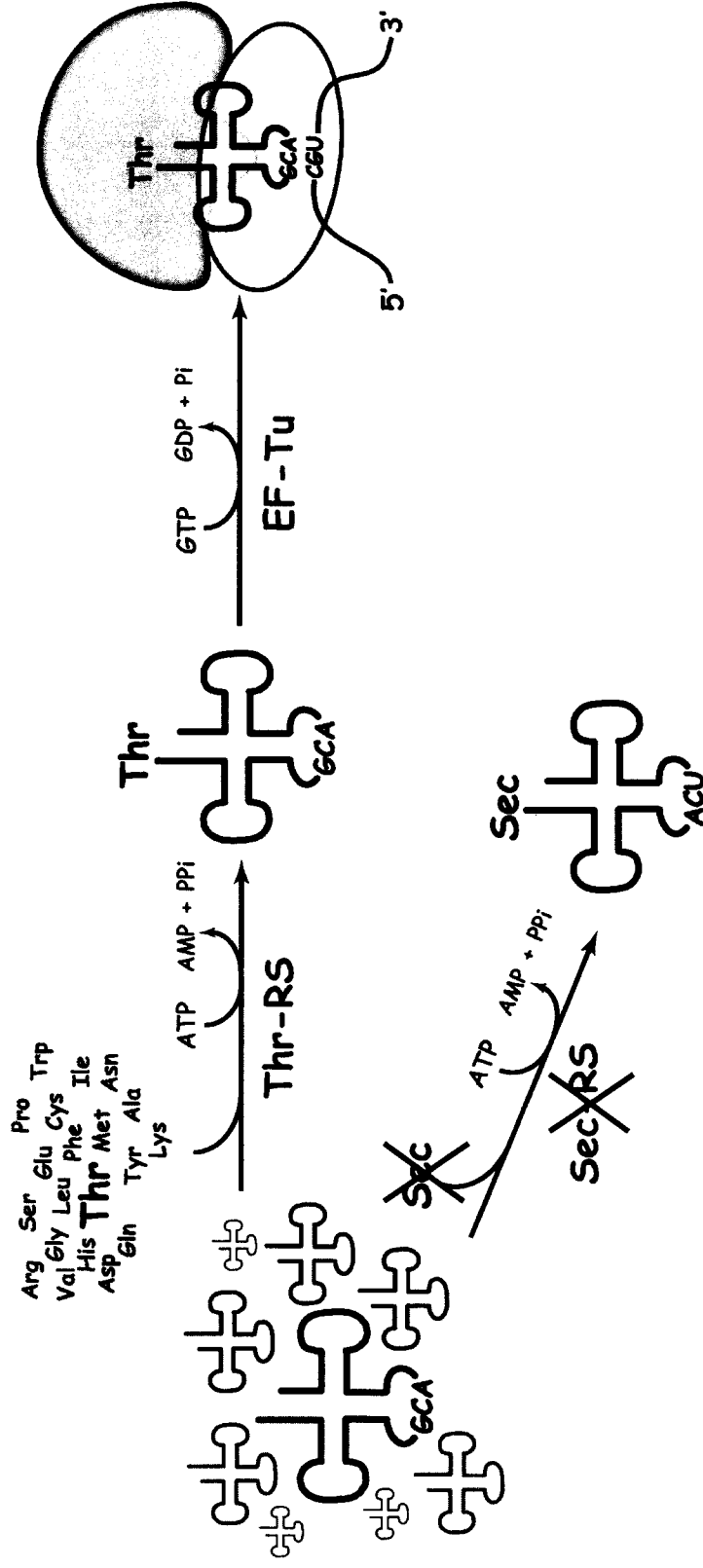


Figure 3: Synthesis of an aminoacyl-tRNA and its role in ribosomal translation. Each aminoacyl-tRNA synthetase (aaRS) is responsible for selecting its cognate tRNA and the corresponding amino acid from the available cellular pools of these molecules. The aminoacyl-tRNA is delivered to the ribosome by elongation factor Tu (EF-Tu) in bacteria and its anticodon can then interact with the corresponding codon in mRNA. The example shown illustrates how this leads to the translation of the codon GCA as threonine during the elongation phase of protein synthesis (Adapted from Ibba and Söll, 2000). According to Crick's adapter hypothesis there should be twenty aaRSs, one for each of the twenty canonical amino acids. Sec is the only exception; there is no cellular pool of free Sec and a selenocysteine-tRNA synthetase (SecRS) has never evolved.

case ^{Sec}) denotes the anticodon that is specific for the cognate amino acid (in this case Sec). The prefix before the tRNA denotes the amino acid attached to the 3' end of the tRNA; for example, Sec-tRNA^{Sec} denotes that selenocysteine is acylated to the 3' end of tRNA^{Sec}, while Ser-tRNA^{Sec} denotes that serine is attached (or misacylated) to the 3' end of tRNA^{Sec}.

D. Biosynthesis of selenocysteine and UGA recoding in bacteria

Pretranslational amino acid modification is the path to Sec-tRNA^{Sec} in bacteria (Figure 4). Selenocysteine biosynthesis begins with SerRS acylating tRNA^{Sec} with serine, but because of its unusual structure SerRS does so with only 1% efficiency as compared with the other five tRNA^{Ser} isoacceptors (Baron et al., 1990). Based on the requirement of *Escherichia coli* formate dehydrogenase for the presence of selenocysteine for activity, a genetic system was developed that led to the identification of four *E. coli* genes (*selA*, *selB*, *selC* and *selD*), all of which are essential for selenocysteine insertion (reviewed in Yoshizawa and Böck, 2009). In addition to tRNA^{Sec} (*SelC*), selenocysteine synthase (*SelA*) and selenophosphate synthase (*SelD*) are needed for conversion of Ser-tRNA^{Sec} to Sec-tRNA^{Sec}. *SelA* is a pyridoxal phosphate (PLP)-dependent enzyme with a conserved lysine (Lys) residue that catalyzes the formation of an enzyme-bound dehydroalanyl intermediate with elimination of a water molecule (Forchhammer et al., 1991). *SelD* synthesizes (from selenide and ATP) the selenium donor selenophosphate (Tormay et al., 1998). This compound reacts with the *SelA*-bound dehydroalanyl-tRNA^{Sec} to form

Sec-tRNA^{Sec}. Thus, selenocysteine is made by a tRNA-dependent amino acid modification process (Figure 4).

Recoding of UGA as selenocysteine requires the presence of an mRNA stem-loop structure known as the selenocysteine insertion sequence (SECIS) element (reviewed in Yoshizawa and Böck , 2009). In bacteria, the SECIS element is located immediately downstream of the UGA codon within the selenoprotein coding region. The general translation elongation factor Tu (EF-Tu) does not recognize Sec-tRNA^{Sec} effectively neither *in vivo* (Leinfelder et al., 1989) nor *in vitro* (Förster et al., 1990). Instead, the specialized elongation factor SelB performs two functions that are crucial for proper UGA recoding as selenocysteine: (i) its N-terminal domain, which is homologous to EF-Tu, binds to Sec-tRNA^{Sec} with high specificity but not to any other aa-tRNAs including Ser-tRNA^{Sec} (Forchhammer et al., 1989); and (ii) the binding of SelB to the SECIS element through its C-terminal domain is required for delivery of Sec-tRNA^{Sec} to the ribosomal A site, by outcompeting release-factor binding to the UGA codon. Thus, the dual properties of SelB guard the fidelity of protein translation by ensuring that only UGA codons in selenoprotein mRNAs are recoded (Allmang et al., 2009).

E. Selenocysteine biosynthesis in eukaryotes and archaea: the last riddle in aminoacyl-tRNA synthesis

Because a selenocysteinyl-tRNA synthetase (an enzyme that would directly esterify selenocysteine to its cognate tRNA^{Sec}) has never evolved, tRNA-dependent conversion of serine has been assumed to be the route to selenocysteine in eukaryotes

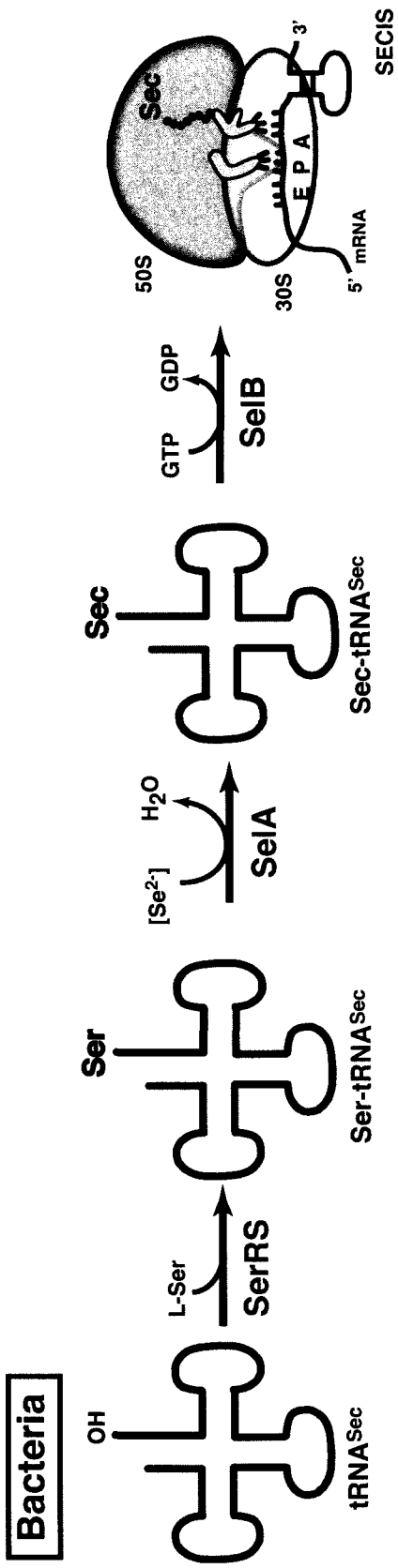


Figure 4: The bacterial pathway to selenoprotein synthesis. Serine is initially attached to tRNA^{Sec} by seryl-tRNA synthetase (SerRS). The resulting Ser-tRNA^{Sec} is then converted to Sec-tRNA^{Sec} by selenocysteine synthase (SelA) in the presence of the selenium donor selenophosphate. The resulting Sec-tRNA^{Sec} is delivered to the ribosome bound to the specialized elongation factor SelB. A Selenocysteine Insertion Sequence (SECIS) element immediately downstream of the mRNA UGA codon to be recoded interacts with SelB and directs the release of Sec-tRNA^{Sec} during selenoprotein translation.

and archaea. Indeed, serylation of tRNA^{Sec} by SerRS has been shown *in vitro* in both eukaryotes (Wu and Gross, 1993; Ohama et al., 1994; Stürchler-Pierrat et al., 1995; Geslain et al., 2006) and archaea (Rother et al., 2000; Bilokapic et al., 2004; Kaiser et al., 2005) and is likely to occur *in vivo* as well (Allmang and Krol, 2006; Geslain et al., 2006). However, the search for a selenocysteine synthase activity that would convert Ser-tRNA^{Sec} to Sec-tRNA^{Sec} proved futile (Kaiser et al., 2005), suggesting an alternative synthetic pathway in eukaryotes and archaea.

An elegant bioinformatic analysis revealed the enzyme *O*-phosphoseryl-tRNA^{Sec} (Sep-tRNA^{Sec}) kinase (PSTK) to be present only in genomes of organisms that have a selenoproteome. The activity was characterized *in vitro* both for the mouse (Carlson et al., 2004) and the *Methanocaldococcus jannaschii* (Kaiser et al., 2005) enzymes; this kinase phosphorylates Ser-tRNA^{Sec} to form Sep-tRNA^{Sec} in the presence of ATP and magnesium. PSTK exhibits a remarkable specificity for Ser-tRNA^{Sec}. It does not phosphorylate free Ser or Ser attached to its cognate tRNA^{Ser} (Carlson et al., 2004). Thus, only tRNA^{Sec} binding can appropriately position Ser in the active site of PSTK for phosphorylation to occur. Moreover, in contrast to most aaRSs that bind their cognate tRNAs with micromolar affinities, PSTK binds both unacylated and serylated tRNA^{Sec} with nanomolar affinity (Sherrer et al., 2008b). Such tight binding of tRNA^{Sec} to PSTK may compensate for the significantly lower abundance of tRNA^{Sec} than tRNA^{Ser} even in selenocysteine-rich tissues such as liver, kidney and testis (Diamond et al., 1993).

While the discovery of PSTK lent credence to the old finding of Sep-tRNA^{Sec} in mammalian cell extracts (Mäenpää and Bernfield, 1970), the concomitant discovery of Sep-tRNA:Cys-tRNA synthase (SepCysS) in some methanogenic archaea (Sauerwald et al., 2005) paved the way for the elucidation of the route to selenocysteine synthesis in archaea and eukarya (Figure 5). SepCysS uses PLP as a cofactor and a sulfur donor to convert Sep attached to tRNA^{Cys} to Cys. Given the similar chemistries of Cys and Sec, the presence of Sep-tRNA^{Cys} as an intermediate of Cys biosynthesis in certain methanogenic archaea suggested Sep-tRNA^{Sec} as an intermediate in the anabolic cycle of Sec (Su et al., 2009) and further supported the quest for an archaeal and eukaryal enzyme that would perform the Sep to Sec conversion (Figure 5). From a chemical standpoint, *O*-phosphoserine would provide a better leaving group (phosphate) than serine (water) for replacement with selenium.

Regarding UGA recoding on the ribosome, several distinguishing features account for a more complex picture and a still-hazy understanding of the process in archaea and eukaryotes. In particular, the established presence of the eukaryotic and archaeal SECIS elements in the 3' untranslated region of the mRNA, the inability of the eukaryotic and archaeal homolog of SelB to bind to the SECIS element, and the identification of other SECIS-binding proteins have led to the proposition of at least two unconfirmed models (Allmang et al., 2009) that may explain how the ribosome distinguishes a selenocysteine UGA from a termination codon and how the eukaryotic and archaeal selenoprotein mRNAs elude nonsense-mediated decay.

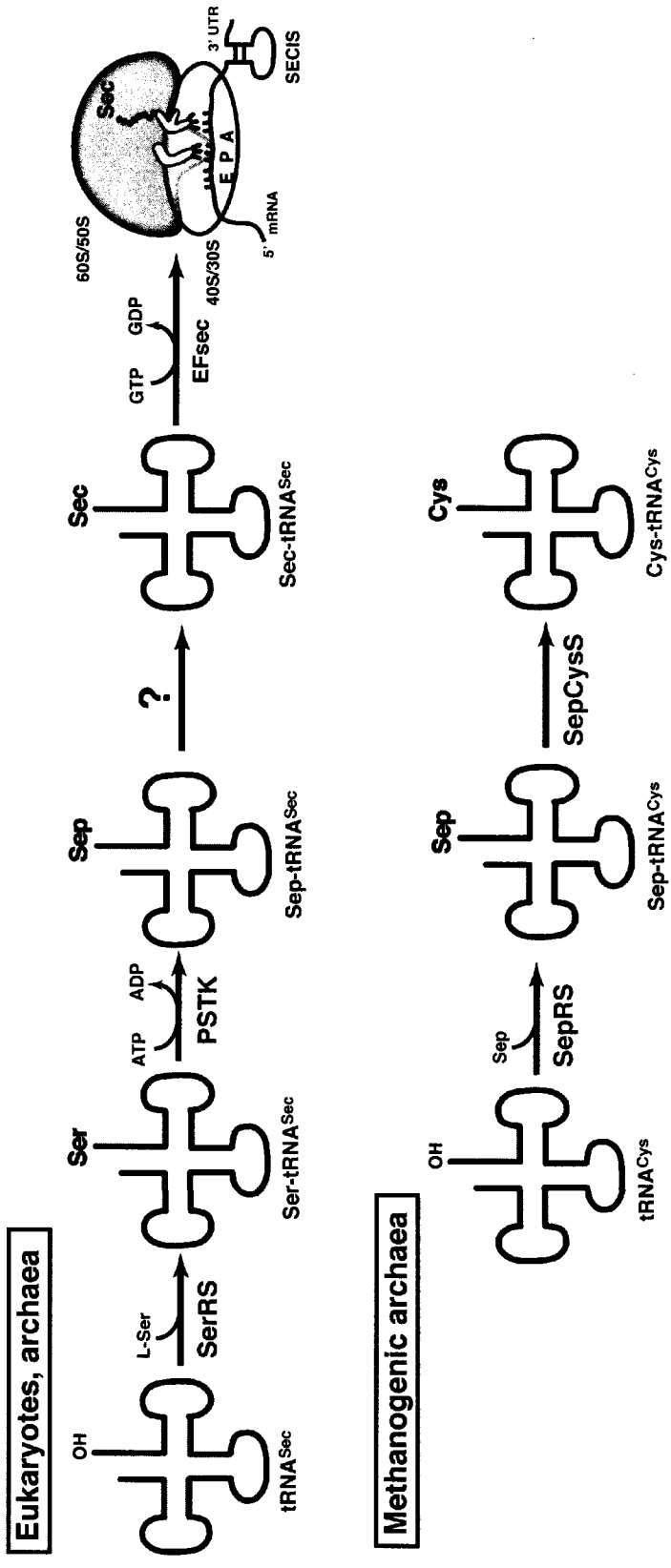


Figure 5: The eukaryotic pathway to selenoprotein synthesis shares a similar phosphorylated intermediate with the tRNA-dependent cysteine biosynthesis pathway in methanogenic archaea. In archaea and eukaryotes Ser-tRNA^{Sec} is converted to O-phosphoserlyl-tRNA^{Sec} (Sep-tRNA^{Sec}) through the action of O-phosphoserlyl-tRNA kinase (PSTK). The activity that converts Sep-tRNA^{Sec} to Sec-tRNA^{Sec} has been elusive for the past twenty years. In methanogenic archaea, Cys-tRNA^{Cys} synthesis proceeds in two steps. First, O-phosphoserlyl-tRNA synthetase (SepRS) acylates tRNA^{Cys} with Sep and then Sep-tRNA^{Cys}: Cys-tRNA synthase (SepCysS) converts Sep-tRNA^{Cys} to Cys in a PLP-dependent reaction. An enzyme analogous to SepCysS is likely to be performing the Sep to Sec conversion in eukaryotes and archaea.

F. Scope of Dissertation

1. Solving the riddle of selenocysteine formation in eukaryotes and archaea

Sec-encoding organisms have 21 genetically encoded amino acids, but have acquired only 20 aminoacyl-tRNA synthetases during their evolution. The biosynthetic route of aa-tRNA formation is understood for all genetically encoded amino acids except for selenocysteine in archaea and eukarya (Allmang and Krol, 2006; Hatfield et al., 2006). Given that a PLP-dependent selenocysteine synthase activity (as manifested by *E. coli* Sela (Tormay et al., 1998) for the conversion of Ser-tRNA^{Sec} to Sec-tRNA^{Sec}) has been elusive, we considered other proteins. The most promising enzyme candidate to exhibit such an activity is the human protein soluble liver antigen/liver pancreas (SLA/LP). This protein was first identified in the early 1990s as it co-precipitated with tRNA^{Sec} when mammalian cell extracts were treated with serum from patients with autoimmune hepatitis (Gelpi et al., 1992). Through a computational approach SLA/LP was classified as a PLP-dependent serine hydroxymethyltransferase (Kernebeck et al., 2001) and its archaeal orthologs are merely found in known Sec-containing archaea. Interestingly, the closest homologue of SLA/LP is SepCysS, the only known tRNA-dependent enzyme that converts Sep to Cys (Yuan et al., 2006; Su et al., 2009).

We used biochemical and molecular genetic studies to establish that human SLA/LP, as well as its *Methanococcus maripaludis* orthologue, convert Sep-tRNA^{Sec} to Sec-tRNA^{Sec} *in vivo* and *in vitro*. Thus, this enzyme is a Sep-tRNA: Sec-tRNA synthase (SepSecS), filling in the question mark in Figure 5 (Yuan et al., 2006).

2. The PSTK/SepSecS pathway is the only route to Sec in *Trypanosomes*, and likely in all eukaryotes

The identification of the human SLA/LP and the *M. maripaludis* MMP0595 proteins as SepSecS enzymes that convert Sep-tRNA^{Sec} → Sec-tRNA^{Sec} was done by *in vivo* complementation assays in a heterologous bacterial system and by *in vitro* activity assays using purified components. However, earlier RNA interference results in mammalian cells could not exclude the existence of an alternative SepSecS-independent pathway of Sec-tRNA^{Sec} formation (Xu et al., 2005). In that study, RNAi-mediated ablation of SepSecS had only a marginal effect on mammalian selenoprotein synthesis.

We investigated the potential presence of an alternative pathway to Sec-tRNA^{Sec} in eukaryotes in the insect form of the parasitic protozoon *Trypanosoma brucei*, which is a highly tractable genetic system. Double allelic knockout cell lines can easily be produced in *T. brucei* by homologous recombination-directed gene replacements (Beverley and Clayton, 1993), while RNAi-based methods for highly efficient inducible ablation of proteins are also available (Ullu et al., 2004). Moreover, *in silico* screens by several groups identified tRNA^{Sec}, SerRS, PSTK, SepSecS, SPS2 (the eukaryotic orthologue of the *E. coli* selenophosphate synthase SelD) and EFSec (the eukaryotic orthologue of the *E. coli* elongation factor SelB) in the *T. brucei* genome (Cassago et al., 2006; Lobanov et al., 2006). Genetic knockouts and *in vivo* selenoprotein analysis in *T. brucei* revealed the presence of a single pathway for Sec-tRNA^{Sec} synthesis involving the SerRS, PSTK and SepSecS sequence and strengthened the notion that all eukaryotes have a single route to Sec.

3. Insights into archaeal Sec synthesis from the structure of *M. maripaludis* SepSecS

The initial characterization of SepSecS revealed that this protein is a PLP-dependent enzyme (Yuan et al., 2006; Xu et al., 2007). In nature such enzymes are abundant; in some microbial genomes, they represent as much as 1.5% of all genes (Percudani and Peracchi, 2003). They have many diverse functions and are often involved in amino acid biosynthesis (Eliot and Kirsch, 2004). The structures of three other PLP-dependent enzymes that use substrates (selenocysteine, cysteine) chemically similar to those of SepSecS have been solved. One is the *Archaeoglobus fulgidus* SepCysS (Fukunaga and Yokoyama, 2007), while another one is the *E. coli* selenocysteine lyase (CsdB) (Lima, 2002), an enzyme that converts selenocysteine to alanine and elemental selenium (Mihara et al., 1999). The last enzyme is the *E. coli* cysteine desulfurase (IscS) that catalyzes the desulfuration of cysteine (Cupp-Vickery et al., 2003). The reactions catalyzed by these enzymes are illustrated in Figure 6; SepSecS and SepCysS carry out β -replacements on tRNA-bound Sep, while selenocysteine lyase removes the β -substituent of Sec to form elemental selenium, and cysteine desulfurase catalyzes the fragmentation of cysteine to alanine and elemental sulfur. The lack of structural and biochemical data on SepSecS prompted us to further characterize this enzyme. The crystal structure of *M. maripaludis* SepSecS (MMPSepSecS) at 2.5 Å resolution was solved and a structural comparison with *E. coli* CsdB (ECCsdB) and *A. fulgidus* SepCysS (AFSepCysS) was performed. I identified active site residues important for the enzymatic function of MMPSepSecS by employing a combination of mutational *in vivo* and *in vitro* activity analyses.

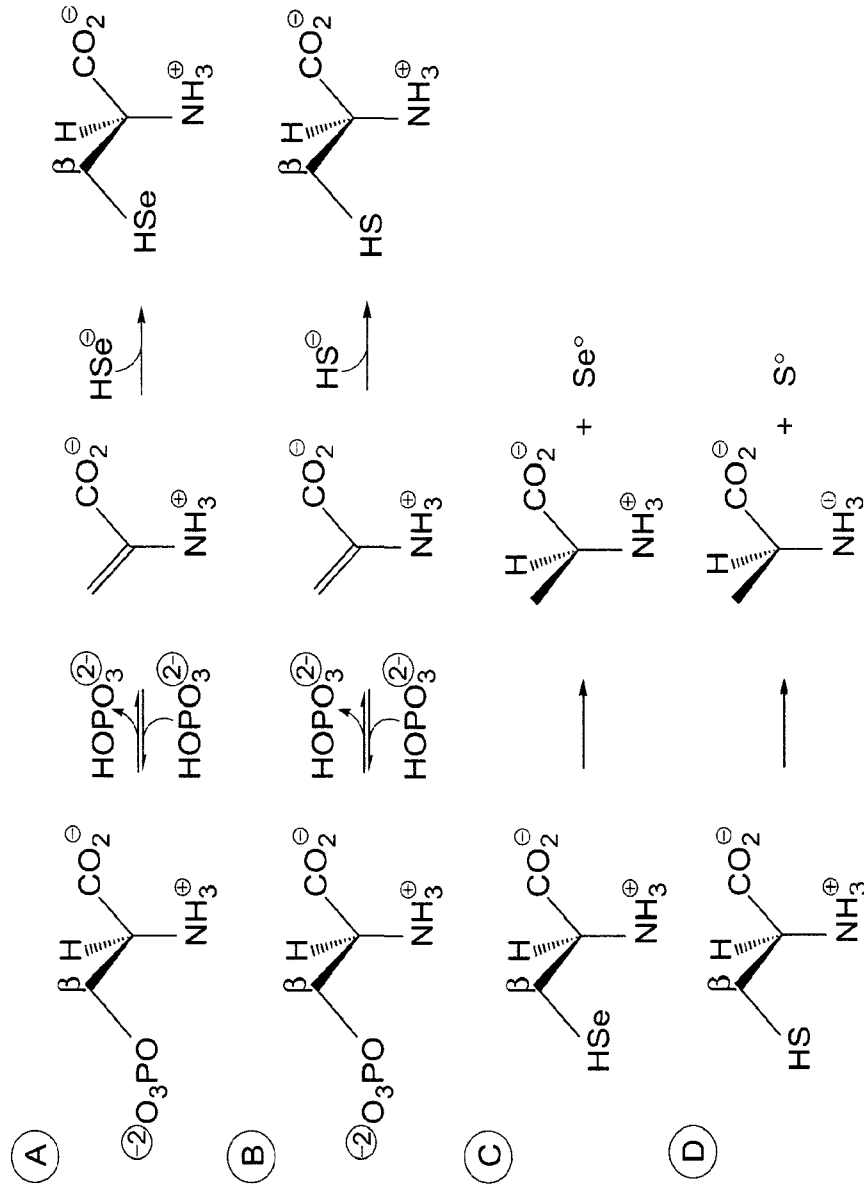


Figure 6: Graphic representation of the reaction schemes (Taken from Araiso et al., 2008). (A) *M. maripaludis* SepSecS catalyzes the conversion of tRNA^{Sec}-bound Sep to Sec. (B) *A. fulgidus* SepCysS mediates the tRNA^{Cys}-dependent transformation of Sep to Cys. (C) *E. coli* CsdB converts selenocysteine to alanine and elemental selenium. (D) *E. coli* IscS converts cysteine to alanine and elemental sulfur.

4. The mechanism of tRNA^{Sec} recognition and Sec formation by the human SepSecS

Despite the crystal structures of the archaeal and murine SepSecS apo-enzymes several questions remained unanswered. SepSecS acts on phosphoserine that is linked to tRNA^{Sec} and not on free phosphoserine or Ser-tRNA^{Sec} (Yuan et al., 2006; Xu et al., 2007; Ganichkin et al., 2008). However, the molecular basis for substrate discrimination and the roles of PLP and tRNA^{Sec} in the mechanism of Sep to Sec conversion were still poorly understood. Moreover, in contrast to its closest homologue SepCysS, which functions as a dimer to perform a similar reaction on Sep-tRNA^{Cys} (Sauerwald et al., 2005; Fukunaga and Yokoyama, 2007), SepSecS forms a stable tetramer (Araiso et al., 2008; Ganichkin et al., 2008). The tertiary structure of tRNA^{Sec} itself and the mode of binding to the tetrameric SepSecS had yet to be discovered as well. In fact, the unique structure of tRNA^{Sec} is thought to be critical for binding to all the enzymes of the Sec biosynthesis and translational incorporation machinery, i.e. SerRS, PSTK, SepSecS, and EFSec (Leinfelder et al., 1988; Baron and Böck, 1991; Wu and Gross, 1993; Wu and Gross, 1994; Ohama et al., 1994; Stürchler-Pierrat et al., 1995; Rudinger et al., 1996; Rother et al., 2000; Bilokapic et al., 2004; Geslain et al., 2006; Sherrer et al., 2008a).

To provide answers to these questions, we determined the crystal structure of the quaternary complex between human SepSecS, unacylated tRNA^{Sec}, thiophosphate and free Sep to 2.8 Å resolution. Unlike the physiologic selenium donor of SepSecS, selenophosphate, which is stable only under anaerobic conditions, thiophosphate is not oxygen-sensitive. Thiophosphate is a surrogate substrate for SepSecS *in vitro* (Araiso et al., 2008) and was, thus, used in the co-crystallization experiment. The structure

demonstrated the basis of SepSecS substrate specificity for Sep-tRNA^{Sec} versus Ser-tRNA^{Sec} and free Sep, while additional structure-guided genetic and biochemical experiments suggested a detailed PLP-dependent mechanism for the Sep-tRNA^{Sec} to Sec-tRNA^{Sec} conversion.

MATERIALS AND METHODS

A. General

Oligonucleotide synthesis and DNA sequencing were performed by the Keck Foundation Biotechnology Resource Laboratory at Yale University. The pET15b, pET20b and pACYCDuet-1 vectors were from Novagen (San Diego, CA, USA). The QuikChange site-directed mutagenesis kit was from Stratagene (LaJolla, CA, USA). [⁷⁵Se]selenite was purchased from the University of Missouri Research Reactor Facility (Columbia, MO). [¹⁴C]Serine (163 mCi/mmol) and [α -³²P]ATP (10 mmol/ μ Ci) were obtained from Amersham Pharmacia Biosciences (Piscataway, NJ, USA). The polyethyleneimine (PEI) cellulose 20 cm \times 20 cm thin layer chromatography (TLC) plates were from Merck (Darmstadt, Germany). Nickel-nitriloacetic acid (Ni-NTA) agarose and the MaxiPrep plasmid purification kit were from Qiagen (Valencia, CA, USA). The Ni- iminodiacetic acid (Ni-IDA) pre-packed column, the Resource-Q pre-packed column and the Sephadex G25 Microspin columns were from GE-Healthcare (Piscataway, NJ, USA). The protease inhibitor cocktail (Complete EDTA-free) was from Roche (Indianapolis, IN).

B. Bacterial strains and plasmids

The *selA* deletion strain was constructed by replacing the gene with the kanamycin cassette in BW25113 *E. coli* strain as described (Datsenko and Wanner, 2000). The T7 polymerase gene was inserted in the genome by P1 transduction. The *E. coli* BL21-CodonPlus (DE3)-RIL strain and the pUC18 vector were from Stratagene (LaJolla, CA, USA). SepSecS genes from *Methanococcus maripaludis* (MMP0595) and

Methanocaldococcus jannaschii (MJ0610) were amplified from genomic DNA, and the human SepSecS gene (GenBank accession no. BX648976) was amplified from an EST clone [clone ID: DKFZp686J1361, obtained from the RZPD German Resource Center for Genome Research, Berlin, Germany (Wiemann et al., 2001)] followed by cloning into the pET15b vector. *E. coli* selD was cloned into pET15b vector. The *M. jannaschii* PSTK gene was cloned into the pACYCDuet-1 vector. The *T. brucei* PSTK and SepSecS genes were cloned into the pACYCDuet-1 and the pET15b vectors, respectively.

The *M. maripaludis* SepSecS mutants R72A, R72Q, R72K, R94A, R94Q, H166A, H166F, H166Q, R307A, R307Q, R307K, Q102A, K278A, N247A, D277A, and K278A were generated using the QuikChange site-directed mutagenesis kit (Stratagene) and cloned into the pET15b vector with an N-terminal His-tag. An N-terminal Δ_{1-34} *M. maripaludis* SepSecS deletion mutant was constructed by PCR using the primers 5'-CCGCTCGAGC ATCGGAAAATTCCTGAAAACGGAATTGATGACG-3' and 5'- GCTAGTTATTGCTCA GCGGTGGCAGC-3' and the pET15b-*sepsecS* plasmid DNA as the template. The resulting DNA fragment was digested with *Bam*HI and *Xho*I and re-inserted into the pET15b vector.

The human SepSecS mutants (R26A, R26E, K38A, K38E, R75A, R97A, S98A, Q105A, K173A, K173M, K284T, N252A, N252D, R271A, R271E, R313A, T397A, T397E, T397V, R398A and R398E) were generated using the QuikChange site-directed mutagenesis kit and cloned into the pET15b vector with an N-terminal His-tag. An N-terminal Δ_{1-37} human SepSecS deletion mutant was constructed by PCR with the pET15b-*sepsecS* plasmid as template and the following primers: 5'-GGAATTCCATATGAAGG

GCAAGTGTCCAGAGAATGGCTGGGATGAAAGTACAC-3', and 5'-CGCGGATCCTC
ATGAAGAAGCATCCTGGTATGTGTCAAGAAGTACATTATC-3'. The resulting
DNA fragment was subcloned into the pET20b vector (Novagen) with a C-terminal His₆-
tag.

C. *In vivo* SepSecS assay

The *M. jannaschii* SepSecS gene, the *M. maripaludis* wild-type and mutant SepSecS genes, the *T. brucei* SepSecS gene and the human wild type and mutant SepSecS genes were transformed into the *E. coli* $\Delta selA$ strain with or without the *M. jannaschii* PSTK gene. The *T. brucei* PSTK gene was transformed into the *E. coli* $\Delta selA$ strain in the presence of *T. brucei* SepSecS, *M. jannaschii* SepSecS or human SepSecS. In strains expressing only one gene, the corresponding empty plasmid was cotransformed as a control.

Aerobic overnight cultures were streaked in aerobic conditions on LB-agar plates and/or M9 minimal medium-agar plates supplemented with 0.01 mM IPTG, 1 μ M Na₂MoO₄, 1 μ M Na₂SeO₃ and 50 mM sodium formate. The plates were placed in an anaerobic incubation jar that was flushed with an N₂: CO₂: H₂ (90 : 5 : 5) gas mix three times to give an anaerobic atmosphere and then grown for 5 h at 37 °C and 20 h at 30 °C. The plates were then overlaid with agar containing 1 mg/ml benzyl viologen (BV), 0.25 M sodium formate and 25 mM KH₂PO₄ adjusted to pH 7.0. The appearance of a blue/purple color is the indication of active formate dehydrogenase H.

D. Metabolic labeling with [⁷⁵Se]selenite

A total of 5×10^7 *T. brucei* cells from each transgenic strain were resuspended in 0.5 mL of FCS-supplemented SDM-79 medium (Brun and Schönenberger, 1979). The cultures were labeled with 9.6 μ Ci of Hepes-neutralized [⁷⁵Se]selenite in the presence of 100 μ g/mL cysteine at 27 °C for 3 h (for the Tb-SepSecS and Tb-PSTK knock-out strains) or 1 h (for the RNAi cell lines). Cells were harvested, washed with phosphate buffered saline pH 7.4, and lysed by heating to 100 °C for 10 min in buffer containing 100 mM Tris-Cl pH 6.8, 8% (w/v) SDS, 24% glycerol, and 0.02% Coomassie blue G-250. The cell lysates were subjected to Tris-tricine SDS/PAGE gels and analyzed by autoradiography.

E. Protein expression and purification

The *M. maripaludis* wild-type and mutant SepSecS proteins and the *E. coli* SelD were purified using Ni-NTA column chromatography. The respective constructs were transformed into the *E. coli* BL21-CodonPlus (DE3)-RIL strain. Cells were grown at 37 °C to A600 nm = ~0.6 at which point the culture was induced with 0.5 mM IPTG for an overnight expression at 15 °C. The cells were harvested by centrifugation at 5,000 rpm for 15 minutes and resuspended in 50 mM Hepes (pH 8.0), 300 mM NaCl, 10 mM imidazole, 3 mM DTT, 10 μ M PLP in the presence of a protease inhibitor cocktail (Complete EDTA-free, Roche). Following sonication and centrifugation at 14 000g for 1 h, the supernatant was applied to Ni-NTA resin (Qiagen). Any non-specifically bound protein was washed away with 50 mM imidazole, while the His₆-tagged proteins were eluted with 300 mM imidazole. Proteins were then thoroughly dialyzed against 50

mM Hepes (pH 8.0), 300 mM NaCl, 10 mM DTT, 10 μ M PLP, and 50% glycerol. PLP was omitted from the lysis, wash, elution and dialysis buffers during the purification of *E.coli* SeID.

Purification of the human SepSecS used for crystallization involved two chromatographic steps. Briefly, pET15b-*sepsecS* was transformed into the *E. coli* BL21-CodonPlus (DE3)-RIL strain, cells were grown to A₆₀₀ nm = 0.6 and gene expression was induced with 0.5 mM IPTG. After growth for 17 h at 15 °C, cells were harvested and resuspended in 20 mM Tris, pH 8.0, 50 mM NaCl, 10% (v/v) glycerol, 1 mM 2-mercaptoethanol, 10 μ M PLP in the presence of the protease inhibitor cocktail (Complete EDTA-free, Roche). The cell paste was lysed using a microfluidizer, the insoluble material was separated by centrifugation at 32,000 rpm for 40 minutes, and the cleared supernatant was loaded onto a Ni-IDA (GE-Healthcare) column pre-equilibrated with 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM 2-mercaptoethanol and 10 μ M PLP. The column was then washed with 50 mM imidazole and the His₆-SepSecS was eluted with 500 mM imidazole. The protein sample was diluted three-fold with 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DTT, 10 μ M PLP, concentrated to 3 mL and loaded onto a S-200pg size-exclusion column. The protein eluted as a tetramer, it was concentrated to ~2 mg/mL, flash frozen in liquid nitrogen and stored at -80 °C.

F. Preparation and purification of tRNA gene transcripts

The *M. maripaludis* tRNA^{Sec} and the human tRNA^{Sec} used as substrates in the *in vitro* assays were synthesized by *in vitro* T7 RNA polymerase run-off transcription as

described (Milligan et al., 1987). In particular, the *M. maripaludis* tRNA^{Sec} and the human tRNA^{Sec} genes together with the T7 promoter were constructed from overlapping chemically synthesized oligonucleotides, cloned into the pUC18 and pUC19 plasmids, respectively, and purified from *E. coli* DH5 α transformants using a MaxiPrep plasmid purification kit (Qiagen). The purified plasmids were digested with BstNI at 55 °C for 16 h. The *in vitro* transcription reactions were performed at 37 °C for 5 h in buffer containing 40 mM Tris-HCl (pH 8.0), 22 mM MgCl₂, 25 mM DTT, 2 mM spermidine, 50 μ g/ml BSA, 0.1 mg/ml pyrophosphatase, 4 mM of each nucleoside triphosphate, BstNI-digested vector containing the tRNA^{Sec} gene (60 μ g/ml) and 1 mM T7 RNA polymerase. The tRNA^{Sec} transcripts were purified by electrophoresis on a 12% denaturing polyacrylamide gel. Full-length tRNA^{Sec} was eluted and desalted on Sephadex G25 Microspin columns (Amersham). The tRNA^{Sec} transcripts were refolded by heating for 5 min at 70 °C in buffer containing 10 mM Tris-HCl (pH 7.0), followed by addition of 5 mM MgCl₂ and immediate cooling on ice.

The human tRNA^{Sec} used for crystallization was cloned into pUC19, expressed in *E. coli* DH5 α and synthesized by *in vitro* T7 RNA polymerase run-off transcription as described above. The transcription reaction was performed at 37 °C for 2 h in buffer containing 40 mM Tris-HCl pH 8.1, 22 mM MgCl₂, 10 mM DTT, 1 mM spermidine, 0.01% Triton X-100, 50 μ g/ml BSA, 4 mM of nucleoside triphosphates, 16 mM GMP, BstNI-digested vector containing the tRNA^{Sec} gene (70 μ g/ml) and 3 mM T7 RNA polymerase. The reaction was briefly centrifuged to pellet the formed pyrophosphate, the supernatant reaction mixture was loaded onto a Resource-Q column (GE-Healthcare)

and tRNA^{Sec} was purified using a linear gradient of NaCl (0.2-1 M) in 20 mM Tris-HCl pH 8.0. Human tRNA^{Sec} eluted at ~0.6 M NaCl. The purity of the tRNA^{Sec}-containing fractions was checked by electrophoresis on a 12% denaturing polyacrylamide gel. Pure tRNA^{Sec} fractions were ethanol precipitated and kept at -80 °C until use.

G. Purification and crystallization of the SepSecS-tRNA^{Sec} complex

The complex was prepared by mixing human SepSecS with a 1.5-fold molar excess of tRNA^{Sec}. Thus, for each SepSecS tetramer there were 6 molecules of tRNA^{Sec} available. The mixture was concentrated to ~1 mL, loaded onto a S-200pg column (XK 16/60) and eluted with 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DTT, 10 μM PLP. The first peak contained the SepSecS-tRNA^{Sec} complex, whereas the second peak contained the unbound tRNA^{Sec}. The complex was concentrated to ~7.5 mg/mL. The complex was crystallized by the sitting-drop vapor-diffusion method at 12 °C by mixing equal volumes of the complex sample with the well buffer containing 0.3 M tri-lithium citrate and 18% (w/v) PEG 3,350. Addition of various additives to the mother liquor such as sorbitol, D-galactose, NDSB-201, Cymal-7, MPD and phenol improved both the crystal size and the quality of diffraction.

H. Ligand soaks and data collection

A mixture of thiophosphate and *O*-phosphoserine was added to the binary complex crystals to the final concentration of 1 mM and incubated for 1 h prior to freezing.

Crystals were cryoprotected for 5 minutes in the presence of 20% (v/v) ethylene-glycol

and then frozen in liquid propane. Data were collected at the NE-CAT 24IDE beamline (APS, Argonne National Laboratory, Chicago) and processed using HKL2000.

I. Preparation of ^{32}P -labeled Sep-tRNA^{Sec}

Refolded tRNA^{Sec} transcript was ^{32}P -labeled on the 3' terminus by using the *E. coli* CCA-adding enzyme and [α - ^{32}P]ATP (Amersham) as previously described with some modifications (Oshikane et al., 2006). Briefly, 6 μg of tRNA^{Sec} transcript was incubated with the CCA-adding enzyme and [α - ^{32}P]ATP (50 μCi) for 1 h at room temperature in buffer containing 50 mM Tris-HCl pH 8.0, 20 mM MgCl₂, 5 mM DTT and 50 μM sodium pyrophosphate. After phenol/chloroform extraction the sample was passed over a Sephadex G25 Microspin column (GE Healthcare) to remove excess ATP (Bullock et al., 2003).

The recovered [^{32}P]-labeled tRNA^{Sec} was serylated and phosphorylated by *M. maripaludis* SerRS (5 μM) and *M. jannaschii* PSTK (1 μM) for 75 min at 37 °C in buffer containing 50 mM HEPES (pH 7.5), 10 mM MgCl₂, 20 mM KCl, 1 mM DTT, 1 mM serine and 10 mM ATP. After phenol/chloroform extraction aminoacylated Sep-tRNA^{Sec} was ethanol precipitated at -20 °C for 45 min and collected as a pellet by centrifugation at 10,000 g at 4 °C for 30 min. After washing the pellet with 70% ethanol, it was allowed to dry on ice in order to avoid deacylation.

To check levels of serylation and Ser→Sep conversion, 1 μl aliquots at the start and end of the reaction were quenched on ice with 3 μl of 100 mM sodium citrate (pH 4.75) and 0.66 mg/ml of nuclease P1 (Sigma). Following nuclease P1 digestion at room

temperature for 1 h, 1.5 μ l of the sample was spotted onto PEI cellulose 20 cm \times 20 cm TLC plates (Merck). To separate the Sep-[³²P]AMP spot from [³²P]AMP and any remaining Ser-[³²P]AMP the plates were developed for 75 min in buffer containing 100 mM ammonium acetate, 5% acetic acid. The plates were exposed on an imaging plate (FujiFilms) for 14 h, scanned using a Molecular Dynamics Storm 860 scanner and quantified using the ImageQuant densitometry software. The amount of Sep-tRNA^{Sec} formed can be calculated by dividing the intensity of the Sep-[³²P]AMP spot by the sum of the intensities of all spots (Sep-[³²P]AMP, [³²P]AMP and Ser-[³²P]AMP).

J. *In vitro* conversion of Sep-tRNA^{Sec} to Cys-tRNA^{Sec}

Purified wild-type or mutant *M. maripaludis* SepSecS (1 μ M) was incubated with 1 μ M ³²P-labeled Sep-tRNA^{Sec} in buffer containing 50 mM HEPES (pH 7.0), 20 mM KCl, 10 mM MgCl₂, 5 mM DTT, 2 μ M PLP and 500 μ M sodium thiophosphate. Reactions were carried out anaerobically at 37 °C over 40 min. At each time point taken, 1 μ l reaction aliquots were quenched on ice with 3 μ l of 100 mM sodium citrate (pH 4.75) and 0.66 mg/ml of nuclease P1 (Sigma). Following nuclease P1 digestion, 1.5 μ l of the sample was spotted onto PEI cellulose TLC plates that were developed, scanned and quantified as described above. To separate the Cys-[³²P]AMP spot from the Sep-[³²P]AMP, the [³²P]AMP and any remaining Ser-[³²P]AMP spots the plates were developed for 75 min in buffer 100 mM ammonium acetate, 5% acetic acid. The plates were exposed on an imaging plate (FujiFilms) for 14 h, scanned using a Molecular Dynamics Storm 860 scanner and quantified using the ImageQuant densitometry software. The amount of

Cys-[³²P]AMP formed was calculated by dividing the intensity of the Cys-[³²P]AMP spot by the sum of the intensities of all spots (Cys-[³²P]AMP, Sep-[³²P]AMP, [³²P]AMP and Ser-[³²P]AMP). The added elevated concentration of PLP was used to assure that the mutant enzymes were saturated with the cofactor.

For the human *in vitro* SepSecS assay, refolded tRNA^{Sec} (10 μM) was serylated and phosphorylated by *M. maripaludis* SerRS (5 μM) and *M. jannaschii* PSTK (3 μM) for 75 min at 37 °C in buffer containing 50 mM HEPES pH 7.2, 10 mM MgCl₂, 20 mM KCl, 1 mM DTT, 10 mM ATP and 300 μM [¹⁴C]Ser. After phenol/chloroform extraction Sep-tRNA^{Sec} was purified by application on a Sephadex G25 Microspin column and ethanol precipitation. The conversion of Sep to Cys was done under anaerobic conditions using wild-type human SepSecS and human SepSecS treated with either hydroxylamine or NaBH₄. For the hydroxylamine treatment, the SepSecS sample was dialyzed for 5 hours against buffer containing 50 mM HEPES pH 7.2, 250 mM NaCl, 1 mM DTT, 7.5 mM hydroxylamine, whereas for the borohydride treatment the SepSecS sample was incubated in the dark for 1 hour in the presence of 100 mM NaBH₄. The excess of hydroxylamine and borohydride was removed from the samples by dialysis against 50 mM HEPES pH 7.2, 250 mM NaCl, 1 mM DTT. SepSecS (20 μg) was incubated with 10 μM Sep-tRNA^{Sec} in buffer containing 50 mM HEPES pH 7.2, 20 mM KCl, 10 mM MgCl₂, 1 mM DTT and 250 μM sodium thiophosphate. All buffers were prepared anaerobically and the reaction was carried out at 37°C inside an anaerobic chamber over 40 min. The reaction was stopped by phenol/chloroform extraction, the aqueous phase was applied to a Sephadex G25 Microspin column and the eluted aminoacylated-tRNAs were

ethanol precipitated. The purified aa-tRNAs were deacylated in 20 mM NaOH for 10 min at 22 °C. The released amino acids were oxidized with performic acid as described previously (Sauerwald et al., 2005). The final samples were spotted onto silica gel 60 TLC aluminium sheets (Merck) that were developed in 85% ethanol. To detect the labelled amino acids, the TLC plate was dried and then exposed on an imaging plate (FujiFilms) for 60 hours. The imaging plate was scanned using a Molecular Dynamics Storm 860 scanner.

RESULTS

A. The missing step of selenocysteine biosynthesis in archaea and eukaryotes

This part of my research was done in collaboration with Jing Yuan, Juan Salazar and Dan Su.

1. SepSecS rescues selenoprotein biosynthesis in an *E. coli* $\Delta selA$ deletion strain

When grown anaerobically, *E. coli* produces the selenium-dependent formate dehydrogenase H (FDH_H). Its activity enables the cells to reduce benzyl viologen (BV) in the presence of formate; this is usually observed by a blue/purple color in agar overlay plates under anaerobic conditions (Lacourciere et al., 2002). An *E. coli* $\Delta selA$ deletion strain was constructed in which selenoprotein production is abolished.

Complementation of this strain with archaeal SepSecS genes from *M. maripaludis* (MMP0595) and *M. jannaschii* (MJ0610) and with the human homologue allowed testing of the ability of these genes to restore selenoprotein biosynthesis as detected by FDH_H activity (Yuan et al., 2006). The *E. coli* cells were grown under anaerobic conditions for 24 h and overlaid with top agar containing BV and formate. Blue-colored cells indicated that BV was reduced and, therefore, FDH_H activity was present, whereas cells without FDH_H activity remained colorless.

Complementation was achieved when the SepSecS genes were cotransformed with the gene encoding *M. jannaschii* PSTK (Figure 7). Neither PSTK alone nor the archaeal and human SepSecS genes alone were able to restore Sec synthesis in the *E. coli* $\Delta selA$ deletion strain (Yuan et al., 2006). These data imply that Sec formation was achieved by a two-step process: PSTK phosphorylates the endogenous Ser-tRNA^{Sec} to Sep-tRNA^{Sec}, and then this misacylated aa-tRNA species is converted to Sec-tRNA^{Sec} by

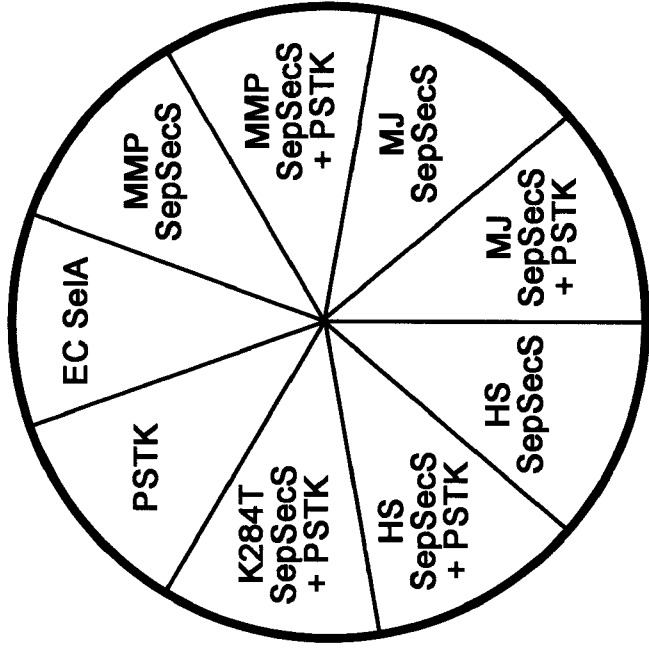


Figure 7: SepSecS genes restore FDH_H activity in an *E. coli* $\Delta selA$ deletion strain (Taken from Yuan et al., 2006). The *E. coli* $\Delta selA$ deletion strains complemented with the indicated SepSecS genes were grown anaerobically on glucose minimal medium plates supplemented with 0.01 mM IPTG at 30 °C for 2 days. FDH_H activity was observed by overlaying top agar containing formate and BV. Colonies with active FDH_H reduced BV to a blue color. The SepSecS genes were from *Methanocaldococcus jannaschii* (MJ), *Methanococcus maripaludis* (MMP), human (HS), and *E. coli selA* (EC). K284T denotes a mutant human SepSecS where the Lys residue critical to PLP binding is changed to Thr.

SepSecS. Moreover, Ser-tRNA^{Sec}, present in the *E. coli* $\Delta selA$ deletion strain, cannot be converted to Sec-tRNA^{Sec} in the absence of PSTK and, therefore, is not a substrate for SepSecS (Yuan et al., 2006). Finally, PSTK from the parasitic protozoon *T. brucei* in combination with either *T. brucei* SepSecS (Tb-SepSecS) or the corresponding archaeal or human enzymes reconstituted selenoprotein synthesis in the *E. coli* $\Delta selA$ deletion strain (Figure 8) (Aeby et al., 2009). Likewise, archaeal PSTK and Tb-SepSecS also restored FDH_H activity in the *E. coli* $\Delta selA$ deletion strain (Figure 8) (Aeby et al., 2009). Taken together, these results underscore the conserved nature of the PSTK/SepSecS pathway of Sec-tRNA^{Sec} formation in archaea and eukaryotes.

The bacterial Sela is a PLP-dependent enzyme with a critical Lys residue engaged in PLP binding (Tormay et al., 1998). As the archaeal and eukaryal SepSecS proteins also have PLP binding domains (Kernebeck et al., 2001), I constructed Lys to Ala or Thr mutations that replaced the critical lysine position in these enzymes (*M. maripaludis* SepSecS K278A, and human SepSecS K284T). Even in the presence of PSTK, the mutant SepSecS genes were no longer able to complement the bacterial *selA* deletion strain (Figure 7) (Yuan et al., 2006). This finding strengthens the notion that the SepSecS proteins need PLP to carry out the Sep \rightarrow Sec conversion.

2. SepSecS converts Sep-tRNA^{Sec} to Sec-tRNA^{Sec} *in vitro*

To prove that SepSecS is a Sela that catalyzes a tRNA-dependent Sep \rightarrow Sec synthesis we analyzed the conversion of [¹⁴C]Sep-tRNA to [¹⁴C]Sec-tRNA *in vitro*. To this end,

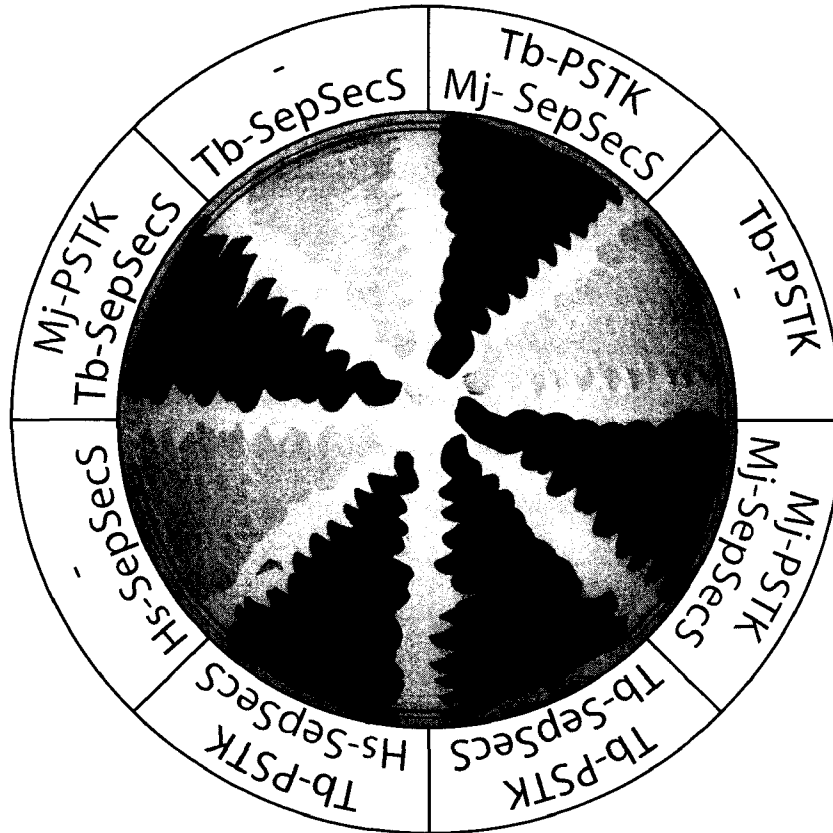


Figure 8: Expression of SepSecS together with PSTK restores FDH_H activity in an anaerobically-grown *E. coli* $\Delta selA$ deletion strain (Taken from Aeby et al., 2009). The indicated genes of the following organisms were tested: *T. brucei* (Tb-PSTK, Tb-SepSecS), *M. jannaschii* (Mj-PSTK, Mj-SepSecS), and human (Hs-SepSecS).

M. maripaludis tRNA^{Sec} was acylated with [¹⁴C]Ser by using pure *M. maripaludis* SerRS (Kim et al., 1998). Phosphorylation was performed with pure *M. jannaschii* PSTK. We then incubated the isolated [¹⁴C]Sep-tRNA under anaerobic conditions with purified recombinant SepSecS and *E. coli* SelD in the presence of selenite. After the reaction, aminoacyl-tRNA was deacylated, and the amino acid products were identified by thin layer chromatography (TLC) (Figure 9) (Yuan et al., 2006). The data show that SepSecS from both *M. maripaludis* (Figure 9, lane 3) and human (Figure 9, lane 4) were able to form Sec-tRNA^{Sec} from Sep-tRNA^{Sec}, but not from Ser-tRNA^{Sec} (Figure 9, lane 5) (Yuan et al., 2006). Thus, Sep-tRNA^{Sec} is the crucial precursor for Sec-tRNA^{Sec} formation in archaea and eukarya.

B. *In vivo* evidence for a single pathway of Sec-tRNA^{Sec} formation in *T. brucei*

This part of my thesis research was done in collaboration with Eric Aeby, who constructed the T. brucei RNAi and knock-out strains.

A bioinformatic analysis of the *T. brucei* genome predicted 3 selenoproteins. They include distant homologs of mammalian selenoprotein K (SelK) and selenoprotein T (SelT) and a selenoprotein, termed SelTryp, that is specific for the kinetoplastid line (Lobanov et al., 2006). Moreover, all major components of the eukaryotic Sec-inserting system have trypanosomal orthologues. Thus, Tb-tRNA^{Sec}, Tb-SerRS, Tb-PSTK, Tb-SepSecS, Tb-SPS2, and Tb-EFSec have been identified *in silico* (Cassago et al., 2006). However, of these only Tb-SerRS, Tb-tRNA^{Sec} (Cassago et al., 2006; Bouzaidi-Tiali et al.,

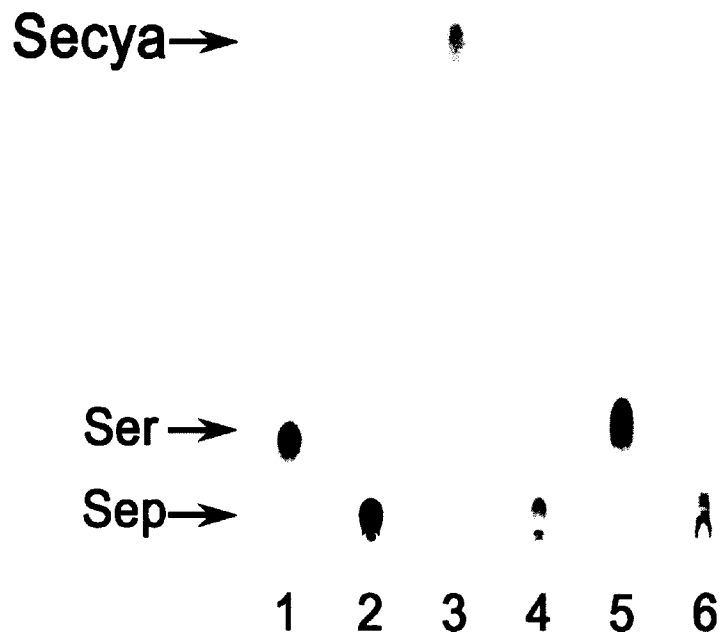


Figure 9: Conversion of *in vitro*-synthesized Sep-tRNA^{Sec} to Sec-tRNA^{Sec} (Taken from Yuan et al., 2006). Phosphorimages of TLC separation of [¹⁴C]Sep and [¹⁴C]Sec recovered from the aa-tRNAs of the SepSecS activity assay. Sec was analyzed in its oxidized form as selenocysteic acid (Secya). Lane 1, Ser marker; lane 2, Sep marker; lane 3, Sep-tRNA^{Sec} with *M. maripaludis* SepSecS; lane 4, Sep-tRNA^{Sec} with human SepSecS; lane 5, Ser-tRNA^{Sec} with *M. maripaludis* SepSecS; lane 6, reaction of lane 3 without selenite.

2007; Geslain et al., 2006), and Tb-SPS2 (Sculaccio et al., 2008) have been subject to preliminary experimental analyses. To analyze the Sec-tRNA^{Sec} formation pathway *in vivo* and establish its physiological importance for *T. brucei*, RNAi cell lines that allow inducible ablation of Tb-SerRS, Tb-SPS2, and Tb-EFSec were constructed. Moreover, we prepared knock-out cell lines that lack either Tb-PSTK or Tb-SepSecS, the two core components of the eukaryotic Sec-tRNA^{Sec} formation pathway. I analyzed all cell lines for selenoprotein synthesis by labeling with radioactive ⁷⁵Se (Figure 10) (Aeby et al., 2009).

Labeling of uninduced Tb-SerRS-RNAi cells with ⁷⁵Se and subsequent analysis by Tris-Tricine polyacrylamide gels revealed three bands whose molecular masses are consistent with the three predicted trypanosomal selenoproteins SelK, SelT, and SelTryp (Figure 10A) (Lobanov et al., 2006). Induction of Tb-SerRS RNAi causes, in line with the role of SerRS in serylation of tRNA^{Sec}, a significant reduction in selenoprotein labeling (Figure 10A) (Aeby et al., 2009). Similarly, RNAi-mediated ablation of Tb-SPS2 (Figure 10C) and of Tb-EFSec (Figure 10D) severely impairs selenoprotein synthesis (Aeby et al., 2009). Most importantly, though, labeling of the three selenoproteins was abolished in the Tb-PSTK and the Tb-SepSecS knock-out cell lines (Figure 10B). Moreover, tetracycline-inducible ectopic expression of Tb-SepSecS in the Tb-SepSecS knock-out cell line restored ⁷⁵Se labeling of all three proteins (Figure 10B) (Aeby et al., 2009). Taken together, these results show that both Tb-PSTK and Tb-SepSecS are indispensable for selenoprotein synthesis and imply that eukaryotes have a single pathway of Sec-tRNA^{Sec} synthesis that requires Sep-tRNA^{Sec} as an intermediate

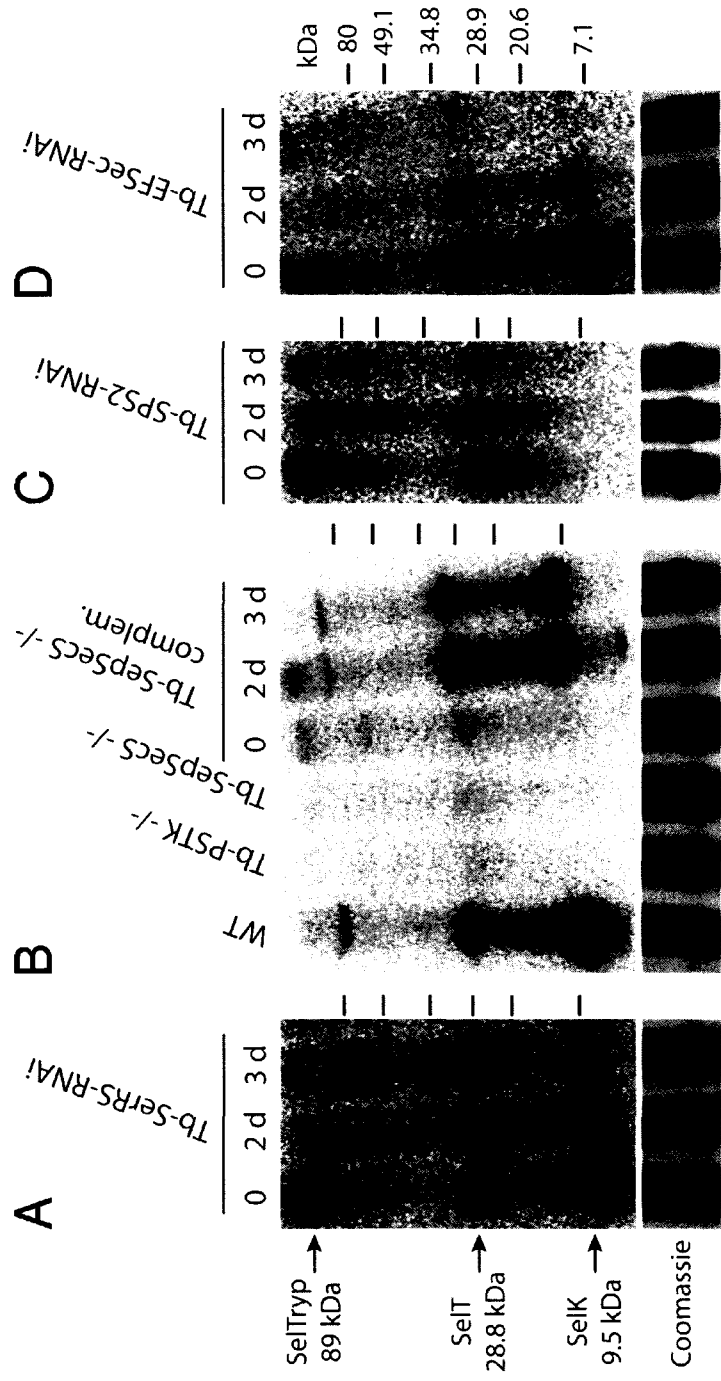


Figure 10: Selenoprotein expression in *T. brucei* cell lines (Taken from Aeby et al., 2009). Expression was analyzed by ^{75}Se labeling of living cells and subsequent analysis of the labeled proteins by 10–20% polyacrylamide Tris-Tricine gels. The following cell lines were analyzed. (A) Uninduced and induced Tb-SerRS-RNAi cells. (B) WT 427 cells, Tb-PSTK KO cells (Tb-PSTK $-/-$), Tb-SepSecS KO cells (Tb-SepSecS $-/-$), and a Tb-SepSecS KO cell line allowing inducible ectopic expression of Tb-SepSecS (Tb-SepSecS $-/-$ complm.). (C) Uninduced and induced Tb-SPS2-RNAi cells. (D) Uninduced and induced Tb-EFSec-RNAi cells. For the RNAi and the complemented Tb-SepSecS $-/-$ cell lines days of induction (d) by tetracycline are indicated. The putative identity of the three labeled selenoproteins and their molecular mass as predicted *in silico* are indicated on the left. Molecular mass markers are indicated on the right. Segments of the tubulin region (50–80 kDa) of the corresponding Coomassie-stained gels are shown as loading controls.

C. Structural insights into archaeal RNA-dependent selenocysteine biosynthesis

This part of my thesis research was done in collaboration with Yuhei Arais, who crystallized and solved the structure of SepSecS from Methanococcus maripaludis.

1. Experimental Approach

In order to investigate the mechanism of selenocysteine formation in archaea, I performed a structure-based biochemical characterization of SepSecS from *M. maripaludis*. The crystal structure of MMPSepSecS with covalently-bound PLP was determined at 2.5 Å resolution (Arais et al., 2008). To assess the activity of wild-type and mutant MMPSepSecS proteins, I used genetic complementation of the *E. coli* $\Delta selA$ deletion strain as an *in vivo* test (described in Results, section A.1) as well as two *in vitro* tests. The first detects the enzyme's final reaction product, Sec-tRNA^{Sec}, as determined by TLC of Sec released from tRNA^{Sec} (Yuan et al., 2006). Given the difficulty in working with selenophosphate (availability and oxygen sensitivity), I used thiophosphate as a surrogate substrate to measure the time course of Cys-tRNA^{Sec} formation by SepSecS. In this assay, I incubated ³²P-labeled Sep-tRNA^{Sec} anaerobically with wild-type or mutant SepSecS proteins and thiophosphate. After nuclease P1 digestion of the tRNA in the reaction mixture, the product Cys-[³²P]AMP was separated from Sep-[³²P]AMP by TLC and quantified.

2. Overall structure of SepSecS from *Methanococcus maripaludis*

The MMPSepSecS protein adopts an L-shaped structure consisting of the N-terminal extension domain (1–130), a catalytic domain (131–309) and a C-terminal domain (353–

436) (Figure 11A). Long, kinked helices (310–352) connect the catalytic and C-terminal domains. The PLP molecule is covalently bound to the conserved Lys residue K278 (Figure 12) at the active site. As can be seen in Figure 11 (Araiso et al., 2008), the overall architecture of MMPSepSecS is similar to the Fold Type I (Eliot and Kirsch, 2004) PLP enzymes, *A. fulgidus* SepCysS (Fukunaga and Yokoyama, 2007) and *E. coli* selenocysteine lyase (Lima, 2002). These enzymes consist of a catalytic domain similar to that of MMPSepSecS, and a 'small domain', formed by the N-terminal and C-terminal polypeptides, which also resembles the MMPSepSecS C-terminal domain (Figure 11B and C) (Fukunaga and Yokoyama, 2007; Lima, 2002). We used a structure-based sequence alignment (Figure 12) to compare MMPSepSecS to three PLP enzymes, AFSepCysS, ECCsdB and ECIsCS, that act upon chemically similar substrates (Araiso et al., 2008). SepCysS, CsdB and IcsS share 24 identical residues while these proteins have only 8 residues in common with SepSecS (Figure 12). The alignment also shows that the archaeal SepSecS proteins lack the C-terminal extension that has been identified as the major antigenic region for the human SLA/LP autoantibodies (Herkel et al., 2002).

3. Archaeal SepSecS is active only as a tetramer

The crystal of the apo-MMPSepSecS contains four SepSecS molecules in the asymmetric unit that are related by non-crystallographic symmetry (Figure 13A). This is in agreement with our gel-filtration results that have shown that SepSecS forms a homotetramer (dimer of homodimers) in solution (Figure 13E). The N-terminal extension domain of each subunit plays a pivotal role in the tetrameric organization of

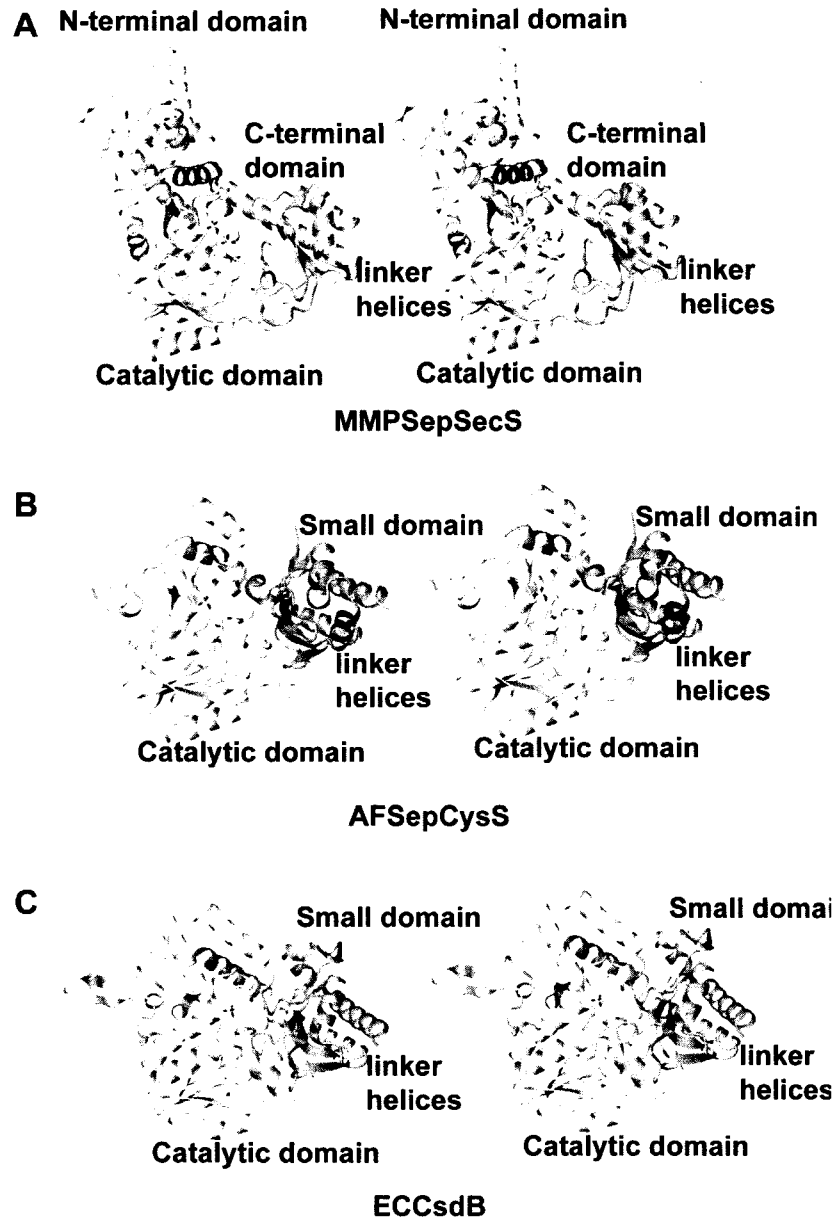


Figure 11: The overall architectures of SepSecS and related enzymes (Taken from Araiso et al., 2008). (A) Stereo view of the *M. maripaludis* SepSecS (MMPSepSecS) structure. The N-terminal domain (residues 1–130), the catalytic domain (residues 131–309), the linker helix (310–352) and the C-terminal domain (residues 353–434) are colored yellow, green, blue and pink, respectively. The carboxy-terminal two residues are disordered. The PLP molecule bound to the active site is shown as a ball and stick model. (B) Stereo view of the *Archaeoglobus fulgidus* SepCysS (AFSepCysS) structure in the same orientation as (A) (Fukunaga and Yokoyama, 2007). The small domain, the linker helices and the catalytic domain are colored pink, blue and green, respectively. (C) Stereo view of the *E. coli* selenocysteine lyase (ECCsdB) structure in the same orientation as (A) (Lima, 2002).

		α1																						
<i>M. maripaludis</i> (2z67)	1	---MLDPNIE LIPENMEK ELVMEYLFIEIDVFNHRRKIPEN I-----																				43		
<i>M. kandleri</i>	1	-----MR LIPDHMLEN RTVLDSYEPVEPVERLLSERMPPEE W-----																				38		
<i>H. sapiens</i>	1	MNRRESFAA ERLVSPAYVVRQ CEARRSHEHLISLLLEN KCPEN W-----																				46		
<i>T. thermophila</i>	1	MNSQNLQLASQITISKYIDLSSCALTRQNLI*VLLSQCLPDD W-----																				46		
SepCysS (2z7j)	8	-----D-----FHIND																				13		
CsdB (1jf9)	2	-----IFSVDRVADFPVLSREVN LP-LAYLDS																				29		
IscS (1p3w)	2	-----RLPIYLDY																				9		
		α2					β1					α3												
<i>M. maripaludis</i> (2z67)	44	-----D-DEKI-KLF--L-K--FLSMMDTDKDPHSVRI EREATYSKIHEELSS FCH I RS NLVDP																				101		
<i>M. kandleri</i>	39	-----P-DDVI-AIF--L-W--ELSRMDTDKDPNAARI EREAVASFLAEEVF FCH V RS TLVDP																				96		
<i>H. sapiens</i>	47	-----D-ESIL-ELF--L-H--ELAIMDSNNFL NC V ERE RVASALVARRHYRPHI I RS DISAV																				104		
<i>T. thermophila</i>	47	-----D-DLSI-EHF--L-Q--ELAMMDTNFFNM V ERE RVFSELVNRKHYMAH I RS DISAE																				104		
SepCysS (2z7j)	14	PLQT ELTEEARQ-ALLE-WP-----																				53		
CsdB (1jf9)	30	AAS--AQPSSVID-AEAE--FYR-H-Y-----A--A-V-----																				54		
IscS (1p3w)	10	SAT--TPVDRVAE-KMGC--FMMD TF-----N-P-----																				36		
		α4				β2				α5				β3										
<i>M. maripaludis</i> (2z67)	102	Q-PK----AS ASIMYALIN-KILESFFKOL LN-V-H--AIAPPIST MSISLCLSAARKKY S--N---VVI																				160		
<i>M. kandleri</i>	97	Q-PK----AP ASIMYALIN-RLVDFLRLR FR-I-E--AFVVP AT LSIALCLSALE E-----E---EVI																				153		
<i>H. sapiens</i>	105	Q-PK----AA SELLNKIIN-SLVLDITKLA VHTV-A--NCFVVPMAT MSITLCPFLTRHKRPA-K---YII																				166		
<i>T. thermophila</i>	105	Q-PK----AA SELLKLELE-YLVKDTLKIC YQSI-K--SVLIVPLAT MALSLVLSLSPNNPKR-R---YVI																				166		
SepCysS (2z7j)	54	-----PIHDFIHMPLKPL-----C-D--VARVE-N AREAFVAMHSLA---K---PD-AWV																				97		
CsdB (1jf9)	55	H-R IHTLSAQATERMENVR-FRASLFI-----N-ARSAEELVSV-R TTE INLVANSV NSN-V--RA DNII																				117		
IscS (1p3w)	37	ASR-SHRF WQAEAVDIAR-NQIADLV-----A-DPREIVFT-S ATESDNLAIN AANFY-QKK---KHII																				98		
		α6			β4			β5			β6			α7			β7							
<i>M. maripaludis</i> (2z67)	161	YPYASHKSPIKAVSFV----MMRLVETV-LD DRVYVPEVDIENAKKEIEL NRP-CVLSTLTF-FPPRNSD																				228		
<i>M. kandleri</i>	154	YPPAAHKSPIKAVRLA----FMRVVDTE-IE DRIVVDP DVEEALERSE---SPA-AVLSTLTF-FPPRNSD																				218		
<i>H. sapiens</i>	167	WPRIDQKSCPKSMITIA----PEPVVIENV-LE DELPDTLKAWEAKVQEL--PDCIL-CLHSTSC-FAPRVPD																				233		
<i>T. thermophila</i>	167	WPRIDQKICLKCIYIS----LEPLVIENI-QPEDRLTINIEEIKVIEK--QDEIL-CVHSTSC-FAPRVPD																				232		
SepCysS (2z7j)	98	MDENCHYSSVVAERA----LNIALVPRDY--PDYAITPENFAQIIEETKR--EVLALITPYDD NY--NLP																				164		
CsdB (1jf9)	118	ISQMEHHANIVPQMLCASV AELGVPLK-P--D ILQLETPLTF--DE---KRLIAITHVSNVLD- TEN																				181		
IscS (1p3w)	99	TSKTEHKAVALDTCQLERE--FEVYIAPQ-R---N IIDLKELEAM--RD---DTILVSIMHVNMEI- VVQ																				161		
		α8			β8			α9			β9			α10			β10			α11				
<i>M. maripaludis</i> (2z67)	229	DIVEIASICEYDIPHIIE AVAIQNNYYLEK-LKRAFKY-RVDAVSSSDRNLLTPI*GLVYSFDA--EFIVE																				299		
<i>M. kandleri</i>	219	PLPETAELCEEY VPHVVAAY IQHEQYRDL-LNRAIKR-RVDVVSSTDRNLLTPV*GLIVAPDE--EITRE																				290		
<i>H. sapiens</i>	234	RLEELAVICANYDIPHVNAY VOSSYCMHL-IQQ APV RIDAFVQSLDRKFMVVPV*AIIAAFND--SPIOE																				305		
<i>T. thermophila</i>	233	DIEFIAQLCAKAYNI HVVNNAY LQCTKIANS-VNM IKR-RVDCILSSTDRNLMVVPV*AFIYSHSE--KVIQE																				304		
SepCysS (2z7j)	165	DVEKIAFVCSYDVPILVH AYAI RM--P-VSLKEI--ADFI V S HKSMAAS PIIVM MK-E--EWAIE																				229		
CsdB (1jf9)	182	PLAEMITLALHGH AVVLVD AQAVMHH--P-VDVQAL--D--CDFYVFS HK-LY PPI I ILVVK-E--ALLOE																				245		
IscS (1p3w)	162	DIAAI EMCRAR ITYHVDAQTSV KL--P-IDLSOL--K--VDLMSFS*HK-TY PK I ALYVR-RKPRV--R																				225		
		α12																						
<i>M. maripaludis</i> (2z67)	300	IS-----LSYP RASA-TPVVNVLVLSLMS*SKN-YLELVKQKNSKRLDDELL																				346		
<i>M. kandleri</i>	291	VS-----RAYP RASA-APVAHALISLLSL MKQ-YRPLMRQKQEKKALLDELL																				337		
<i>H. sapiens</i>	306	IS-----KMYP RASA-SPSLDVLITLLSL*SNQ-YKKLLSKKEMFYSLSNQI																				352		
<i>T. thermophila</i>	305	IN-----QLYP RASA-PIILDVFI*LLSMQKK-LKDLLSRKENLOYLRDCL																				351		
SepCysS (2z7j)	230	VLRK-----SER--Y-KNK--EVELL CTAR ATITILMASFPHV-RER-IR-RNDEVEKARFFAEM																				285		
CsdB (1jf9)	246	M-PPWE*NSMIAIVSLSE*FIWTRKAPWFEA*TPNI*GEIIL*AALEYV-SAL*LNNAIAYQNLHMHYALSQ																				317		
IscS (1p3w)	226	I-EAQMHE-----H--ER MRS*TLPV-HQIV M EAYRIA-KEE-MATEMERLR LRRNLWN I																				281		
		β11			α13			β12			α14													
<i>M. maripaludis</i> (2z67)	347	NDLSFKT--KFLD-V--ES-P-IASCISVN-----SD-----PVEIAAKLYNLRVT*PR I																				392		
<i>M. kandleri</i>	338	EDLEARRDDVRVLD-V--DN-P-IASAVAVE-----HD-----PVDLAARLYVRVTP*PR V																				385		
<i>H. sapiens</i>	353	FKLGEAYN-EFLRHIP--HN-P-IESLAMTKTL-----DEHRDFA-----VIQL*EMLETRQVSRV																				406		
<i>T. thermophila</i>	352	OKFAERHS-ERLLDIP--EN-T-IESLALITTMKEQIILEKEQENKQD-----ITSLQ*ILYSKVM*SRIL																				413		
SepCysS (2z7j)	286	EKL-----LYQL--DN-PHNN-DMFFHAEVLY-----EISKKAK*RFPLRELSRKH I--IKP																				338		
CsdB (1jf9)	318	ESV--PD-LTLY--PO-NR-L-VIAFNL-----KHH-----AYDV*SFLDNY IAV-RT--																				360		
IscS (1p3w)	282	KDI--EE-VYLN--DLEH--APNILNVSNF-----YVE-----ESLIMAL-KD-LAV-SS--																				324		
		insertion																						
<i>M. maripaludis</i> (2z67)	393	K-----KTD-----HP*NCYL*TYT--HD--YIVMNAAI VRTEDIVNSVSKLEKILL																				436		
<i>M. kandleri</i>	386	R-----ADD-----PF*TSRLR*YH--SN--YITINAAT VREEDVKTAVRLEKELE E---																				431		
<i>H. sapiens</i>	407	HL*SMQTVS*YTER-----PMSHTNN-YF--CA--YLNAASAI MKMODVDFIRRLDRCLKAVRKE																				464		
<i>T. thermophila</i>	414	LNQTK*EVC*IEFK-----NY*SHSDS*FTY--LP--YMTIACAI MEKKEIDQVLLKLEDS*PQELRKE																				471		
SepCysS (2z7j)	339	-----LTRYFKLSTY--LSDEEVDVYVNAF*EIIIEKYS--																				371		
CsdB (1jf9)	361	-----HHCAM-----PLMAY--YN--VPA-MCRASLAMYNTHEEVDRLVT*LORIHRL																				406		
IscS (1p3w)	325	-----SACTISASLEPSYVLRAL--LNDELAHS-SIRFSL RFTTEEIIDYTIELVRSI*RLRDL																				383		
		antigenic region																						
<i>M. maripaludis</i> (2z67)		-----																				X	hydrophobic	
<i>M. kandleri</i>		-----																				X	polar	
<i>H. sapiens</i>	465	RSEFESDDNYDFTEDVDIEEMALKLDNVLLD*YODASS																				501	X	basic
<i>T. thermophila</i>	472	KVDKQ																				476	X	acidic
SepCysS (2z7j)		-----																				X	aromatic	
CsdB (1jf9)		-----																				X	glycine	
IscS (1p3w)	384	SPLWEMY*Q																				392	X	glycine

Figure 12: Structure-based sequence alignment (Taken from Araiso et al., 2008). The alignment includes SepSecS sequences from two archaea (*Methanococcus maripaludis* and *Methanopyrus kandleri*) and two eukaryotes (*Homo sapiens* and *Tetrahymena thermophila*) as well as *Archaeoglobus fulgidus* SepCysS, *E. coli* cysteine desulfurase (IscS) and *E. coli* selenocysteine lyase (CsdB). Amino acids are colored according to residue types, and strictly conserved residues are marked (*). The secondary structure of MMPSepSecS is shown above the alignment, and the active site arginines are highlighted in blue. The autoimmune antigenic region of human SepSecS (Herkelel et al, 2002) and a eukaryotic specific insertion are shown in gray boxes and green, respectively.

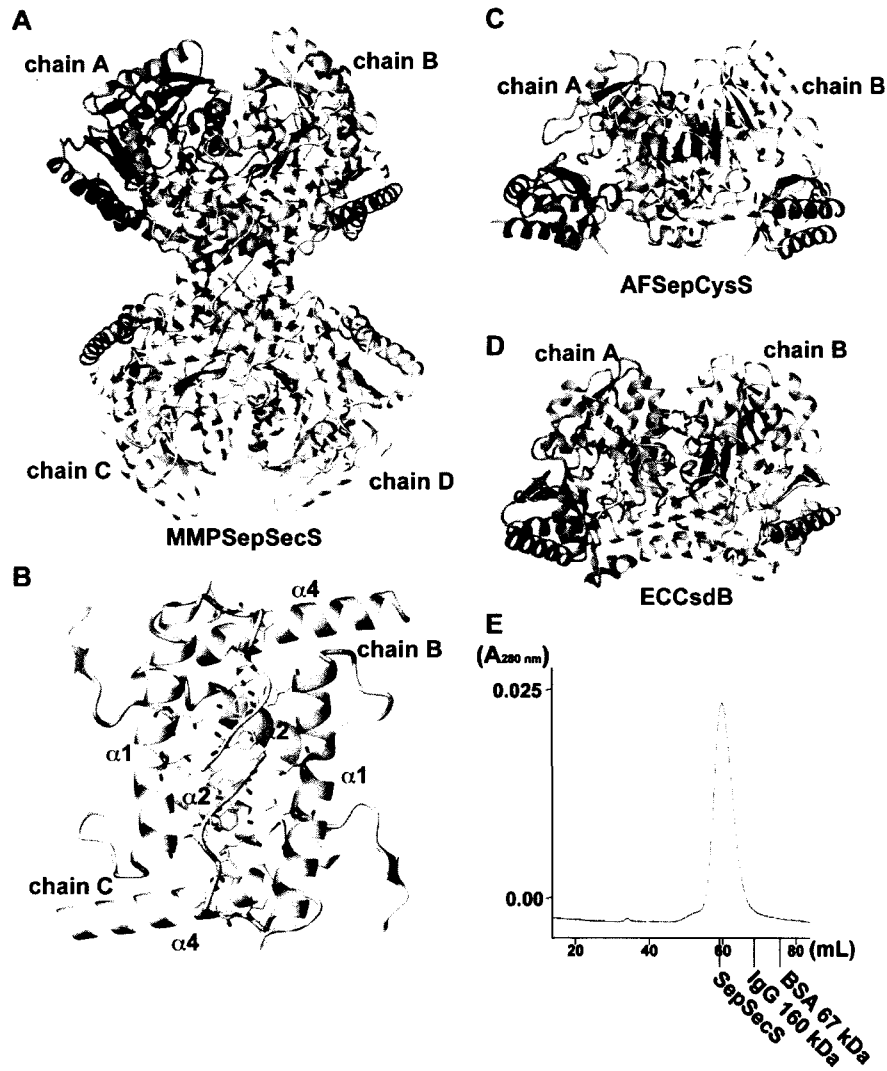


Figure 13: The oligomeric states of SepSecS and related enzymes (Taken from Araiso et al., 2008). (A) The overall architecture of the MMPSepSecS tetramer. Chains A and B, and chains C and D form dimers, respectively. Chain A, chain B, chain C and chain D are colored pink, blue, green and yellow, respectively. (B) The N-terminal extension domains of chains B and C form a hydrophobic core that stabilizes the tetrameric state of MMPSepSecS. Three α -helices (1, 2, 4) are labeled. Chains B and C are colored blue and green, respectively. (C) Stereo view of the AFSepCysS dimer in the same orientation as (A) (Fukunaga and Yokoyama, 2007). Chains A and B are colored pink and blue, respectively. The PLP molecules are shown as ball and stick models. (D) Stereo view of the ECCsdB dimer in the same orientation as (A) (Lima et al, 2002). The coloring scheme and the PLP representation are the same as in (C). (E) Gel filtration of MMPSepSecS on Sephacryl S-300. Absorbance at 280 nm is shown as a blue line. The elution volumes of other oligomeric proteins are indicated in the chromatogram. The molecular weight of the MMPSepSecS monomer is 50 kDa. MMPSepSecS eluted at the size expected for a tetrameric species.

SepSecS (Araiso et al., 2008). These domains interact with each other to form a hydrophobic core (chain A with chain D, and chain B with chain C), which involves inward-facing hydrophobic residues protruding from the helices, $\alpha 1$, $\alpha 2$ and $\alpha 4$ (Figure 13B). Thus, the N-terminal extension facilitates tetramer formation and its deletion is predicted to produce a dimeric SepSecS. Interestingly, AFSepCysS (Fukunaga and Yokoyama, 2007) and ECCsdB (Lima, 2002) lack the N-terminal extension domain and form dimers (Figure 13C and D).

In order to investigate whether SepSecS would be active in a homodimeric form I assayed the enzymatic activity of an MMPSepSecS protein that lacks the N-terminal extension residues 1–34. The Δ_{1-34} SepSecS mutant did not form Sec-tRNA^{Sec} *in vivo* (Figure 14) since it did not rescue selenoprotein biosynthesis in the complementation test of the *E. coli* Δ_{selA} deletion strain (Araiso et al., 2008). Since the active site of SepSecS, like other PLP enzymes, is formed at the dimer interface it is clear that SepSecS would not function as a monomer. Why the tetrameric organization of SepSecS is critical for function remained unclear until the determination of the human SepSecS-tRNA^{Sec} co-crystal structure (reviewed in Results, section D) (Palioura et al., 2009).

4. Active site mutants define PLP recognition

As in other Fold Type I PLP enzymes, the active sites of MMPSepSecS lie on the dimer interface with each monomer contributing essential residues (Eliot and Kirsch, 2004). The active site of chain A is formed by chains A and B that both recognize the PLP molecule (Figure 15) (Araiso et al., 2008). The catalytic domain harbors a seven-stranded

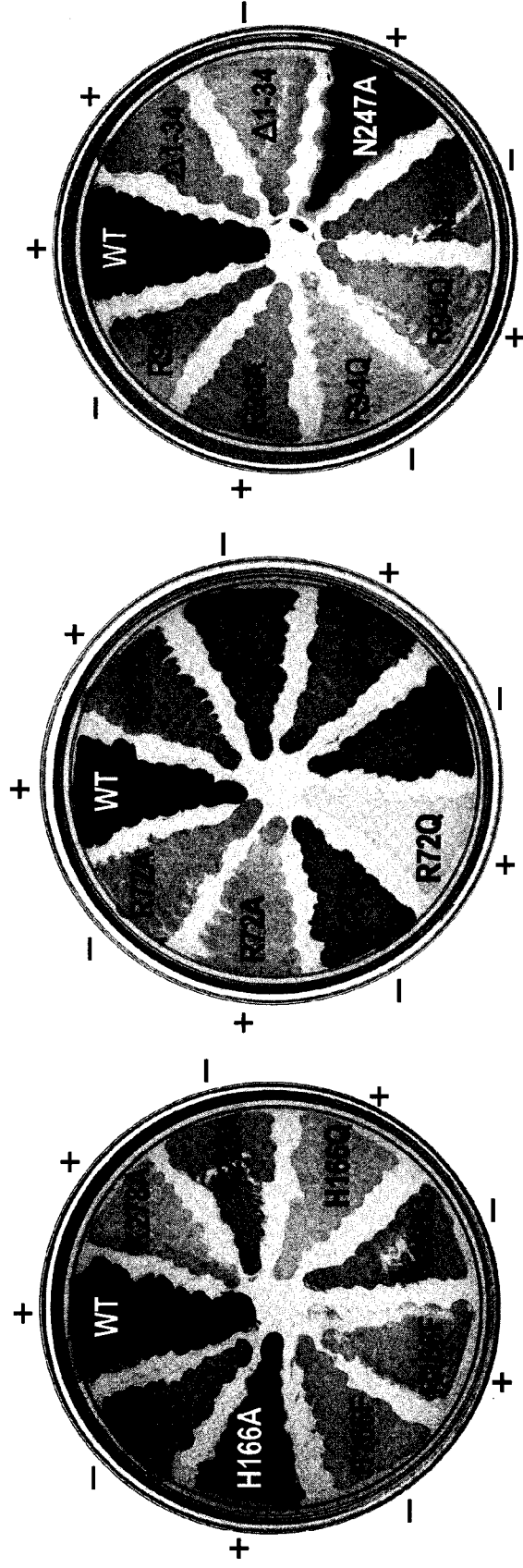


Figure 14: *In vivo* assays of MMPSepSecS mutants (Taken from Araiso et al., 2008). Formation of Sec-tRNA^{Sec} *in vivo* is assayed by the ability of the wild-type MMPSepSecS and its mutant variants (N-terminal deletion Δ1-34, R72A, R72Q, R72K, R94A, R94Q, H166A, H166F, H166Q, K278A, N247A, D277A) to restore the BV reducing activity of the selenoprotein FDH_H in the *E.coli* Δ*selA* deletion strain. Cotransformation of the PSTK gene (indicated with +) from *M. jannaschii* is required for the formation of the Sep-tRNA^{Sec} intermediate.

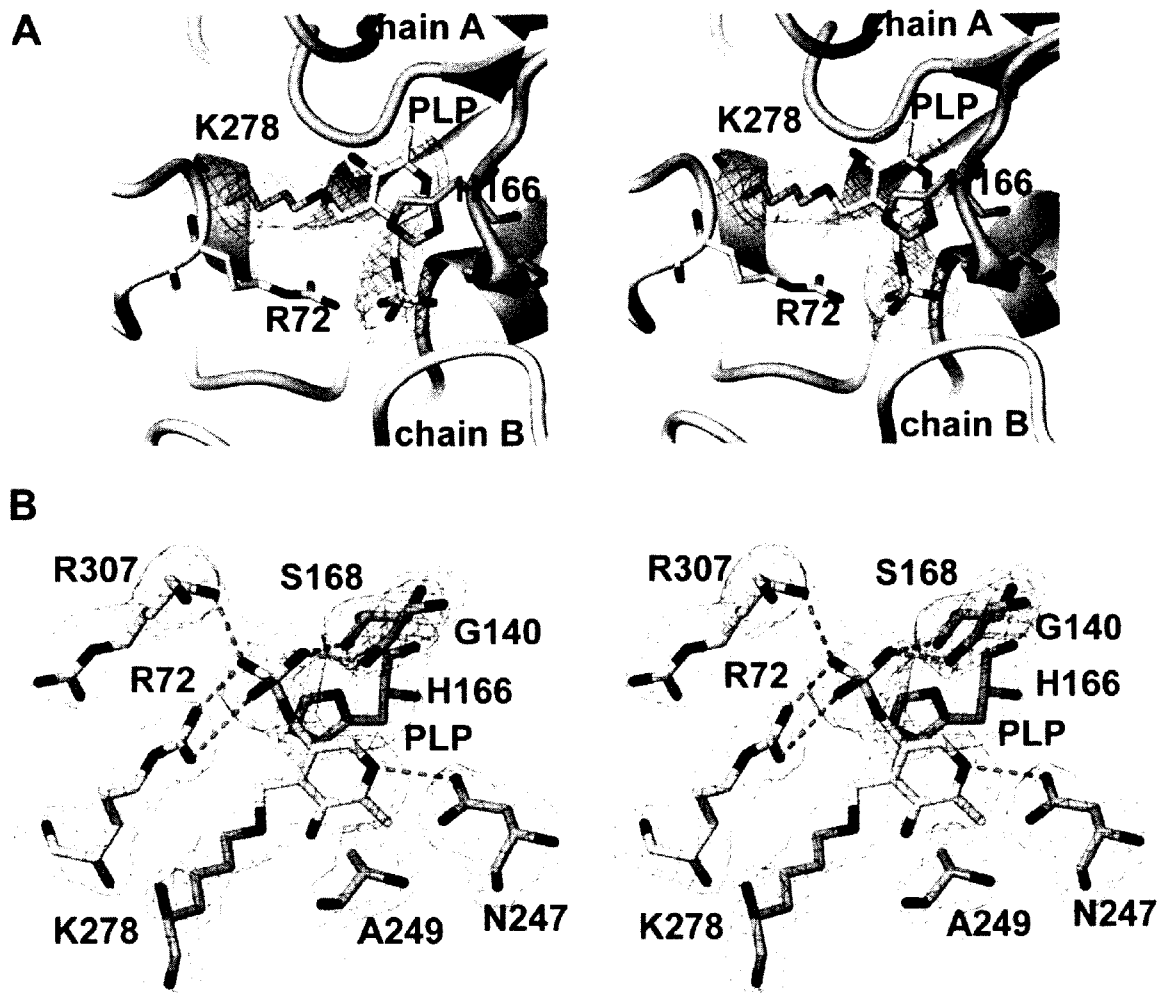


Figure 15: PLP recognition (Taken from Araiso et al., 2008). Stereo views of (A) Ribbon representation of the active site of MMP-SepSecS in the dimer interface between chains A and B that are colored pink and blue, respectively. The PLP molecule is covalently bound to Lys278 of chain A. The F_o-F_c omit map of the PLP molecule, contoured at 3.5σ , is shown. (B) The amino acid residues that recognize the PLP molecule. The residues of chains A and B are colored pink and blue, respectively. The F_o-F_c omit map (contoured at 3.5σ) of all of the residues and PLP is shown.

β -sheet, with only the sixth β -strand being antiparallel; this is a common feature of Fold Type I PLP enzymes (Eliot and Kirsch, 2004; Jansonius, 1998) including AFSepCysS (Fukunaga and Yokoyama, 2007) and ECCsdB (Lima, 2002). In the MMPSepSecS structure, PLP is covalently bound via a Schiff base to the strictly conserved Lys278 (chain A) (Figure 12), which is located between the sixth and seventh β strand in the active site (Figure 15A) (Araiso et al., 2008). Indeed, a Lys278Ala mutation abolishes MMPSepSecS catalytic activity as shown both *in vivo* by the lack of BV reduction by FDH_H (Figure 14), and *in vitro* by the inability of the Lys278Ala mutant to form Cys-tRNA^{Sec} (Figure 16) (Araiso et al., 2008).

All known Fold Type I PLP enzymes possess a critical aspartate that forms a hydrogen bond with the N1 atom of the pyridine ring (Schneider et al., 2000), which is thought to increase the electron sink character of the PLP cofactor. SepSecS is an exception to this paradigm; Asn247 is found in the corresponding position, and it forms a hydrogen bond with the pyridinium nitrogen (Figure 15). The other 'nonconforming' enzyme is SepCysS where the structure also reveals a similar Asn contact with PLP (Fukunaga and Yokoyama, 2007). To test whether Asn247 is responsible for the discrimination that MMPSepSecS exhibits towards Ser-tRNA^{Sec}, we mutated it to Asp and Ala. Though still fully active towards Sep-tRNA^{Sec}, the Asn247Asp MMPSepSecS mutant is inactive towards Ser-tRNA^{Sec} (Figure 17) (Araiso et al., 2008). Finally, the replacement of Asn247 with the uncharged Ala yields an enzyme that is only partially active (Figure 14). Asn247, thus, plays the role of the highly conserved Asp247 in the

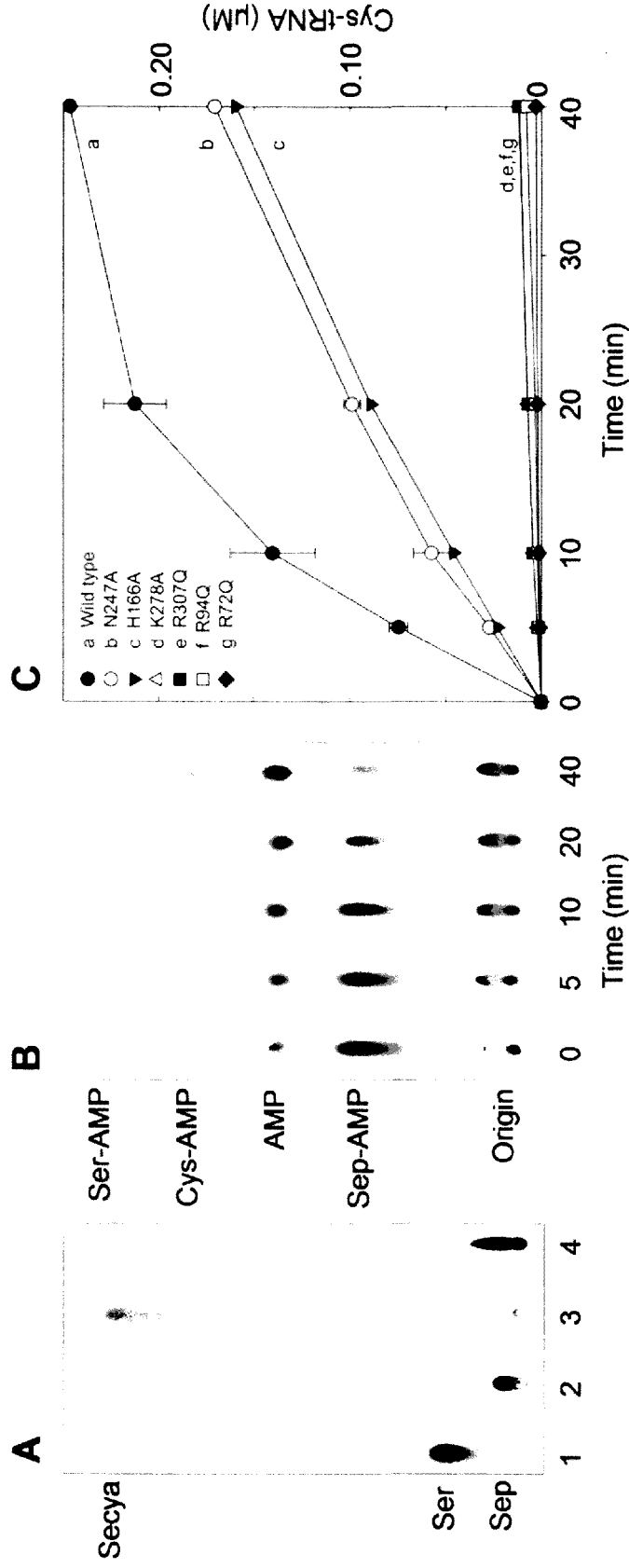


Figure 16: *In vitro* conversion of Sep-tRNA^{Sec} to Sec-tRNA^{Sec} or Cys-tRNA^{Sec} (Taken from Araiso et al., 2008). (A) Phosphorimages of TLC separation of [¹⁴C]Sep and [¹⁴C]Sec recovered from the aa-tRNAs of the SepSecS activity assay. Sec was analyzed in its oxidized form as selenocysteic acid (Secya). Lane 1, Ser marker; lane 2, Sep marker; lane 3, Sep-tRNA^{Sec} with wild-type MMPSepSecS; lane 4, Sep-tRNA^{Sec} with the R72Q MMPSepSecS mutant. (B) Representative phosphorimage for the H166A SepSecS mutant of the separation of Ser-[³²P]AMP, Cys-[³²P]AMP and Sep-[³²P]AMP. At the indicated time points aliquots of the SepSecS reaction were quenched, digested with nuclease P1 and spotted onto PEI-cellulose TLC plates. (C) Plot of Cys-tRNA^{Sec} formed versus time with 1 μM of wild-type and mutant SepSecS enzymes using Sep-tRNA^{Sec} (1 μM) and thiophosphate (500 μM) as substrates. Following quantification of the intensities of Ser-[³²P]AMP, Cys-[³²P]AMP, and Sep-[³²P]AMP using ImageQuant, the concentration of Cys-tRNA^{Sec} formed at each time point was calculated by dividing the intensity of the Cys-[³²P]AMP spot by the total intensity.

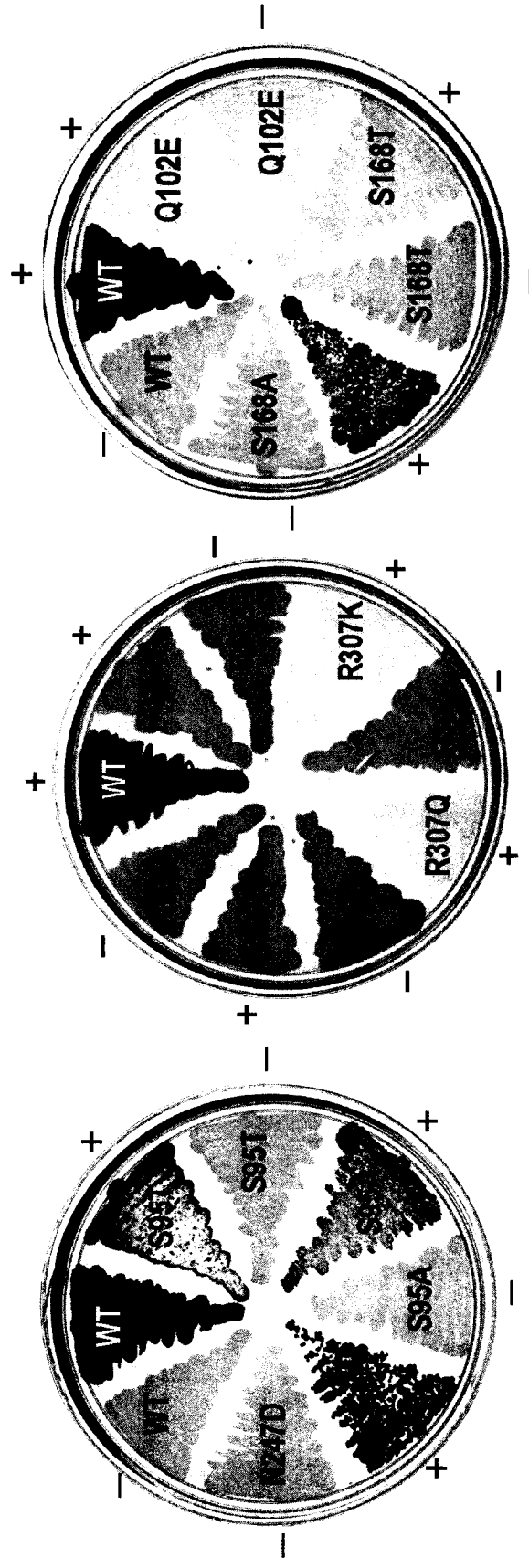


Figure 17: *In vivo* assays of MMPSepSecS mutants (Adapted in part from Araiso et al., 2008). Formation of Sec-tRNA^{Sec} *in vivo* is assayed by the ability of the wild-type MMPSepSecS and its mutant variants (N247D, Q102A, Q102E, R307A, R307K, R307Q, S95A, S95T, S168A, S168T) to restore the BV reducing activity of the selenoprotein FDH_H in the *E. coli selA* deletion strain. Cotransformation of the PSTK gene (indicated with +) from *M. jannaschii* is required for the formation of the Sep-tRNA^{Sec} intermediate.

Fold Type I PLP enzymes. The protonated PLP is critical for the stabilization of the covalently attached carbanion intermediate in the reaction pathway.

The aromatic pyridine ring of PLP is sandwiched between His166 (chain A) and Ala249 (chain A) through hydrophobic interactions at the floor of the catalytic site (Figure 15). This recognition mode is commonly observed in the structures of Fold Type I PLP enzymes where the His residue increases the electron sink character of PLP's pyridine ring through the stacking interactions with it and may also play a role in substrate activation and acid base catalysis. In the latter case, the His residue makes a direct hydrogen bond with the critical aspartate that contacts the N1 atom of the coenzyme pyridine ring in all Fold Type I PLP enzymes and is, thus, thought to assist in the dissipation of the negative charge generated around PLP's ring during catalysis (Yano et al., 1991). Furthermore, the phosphate moiety hydrogen bonds with the main-chain amide and carbonyl groups of Gly140 and Ser168, respectively (Figure 15B). The MMPSepSecS mutant His166Ala was partially active in forming Sec-tRNA^{Sec} *in vivo* (Figure 14). *In vitro*, the His166Ala MMPSepSecS mutant was partially active in forming Cys-tRNA^{Sec} as can be seen by the increase in intensity of the Cys-[³²P]AMP spot and the concomitant decrease of the Sep-[³²P]AMP spot during the course of the reaction (Figure 16B) (Araiso et al., 2008). Obviously, the mutant enzyme with Ala166 has PLP still in a partially functional position. A mutation of the corresponding His residue (position 143) in *E. coli* aspartate aminotransferase also led to a functional enzyme (Yano et al., 1991). In contrast, both the His166Gln and His166Phe mutants were inactive *in vivo* (Figure 14).

Since the catalytic pocket is formed at the dimerization interface, chain B also contributes residues that participate in PLP recognition. In particular, the guanidinium group of Arg72 and the main-chain amide group of Arg307 (from chain B) hydrogen bond to the phosphate moiety of PLP (Figure 15B). Mutations of Arg72 and Arg307 to Ala, Gln or Lys resulted in MMPSepSecS mutants that were significantly less active in Sec-tRNA^{Sec} formation *in vivo* (Figures 14 and 17) and Cys-tRNA^{Sec} formation *in vitro* (Figure 16C). Such PLP recognition differs from that of AFSepCysS (Fukunaga and Yokoyama, 2007) and ECCsdB (Lima, 2002). Unlike in the MMPSepSecS structure, the adjacent subunit of ECCsdB does not come close to the active site (Figure 18A, B) (Araiso et al., 2008). In ECCsdB, an additional $\beta\alpha\beta$ structural motif covers the active site, presumably to stabilize the PLP and Sec substrates inside the pocket. This motif also precludes interaction between residues from the neighboring subunit and the active site (Figure 18B). Although the active site of AFSepCysS is formed by residues from chains A and B, the chain B amino acids are located too distant for PLP recognition (Figure 18C). The active sites of MMPSepSecS and AFSepCysS are spacious enough to accommodate the Sep-CCA end of the tRNA^{Sec} or tRNA^{Cys} species, respectively.

5. The active site

The crystallization solution contained 10 mM magnesium sulfate. In the present structure, we observed a strong electron density peak (4.5σ), presumably corresponding to a sulfate ion, adjacent to the PLP molecule of chains A and B (Figure 18A). The sulfate ion may be mimicking the position of the phosphate moiety of Sep attached to tRNA^{Sec}.

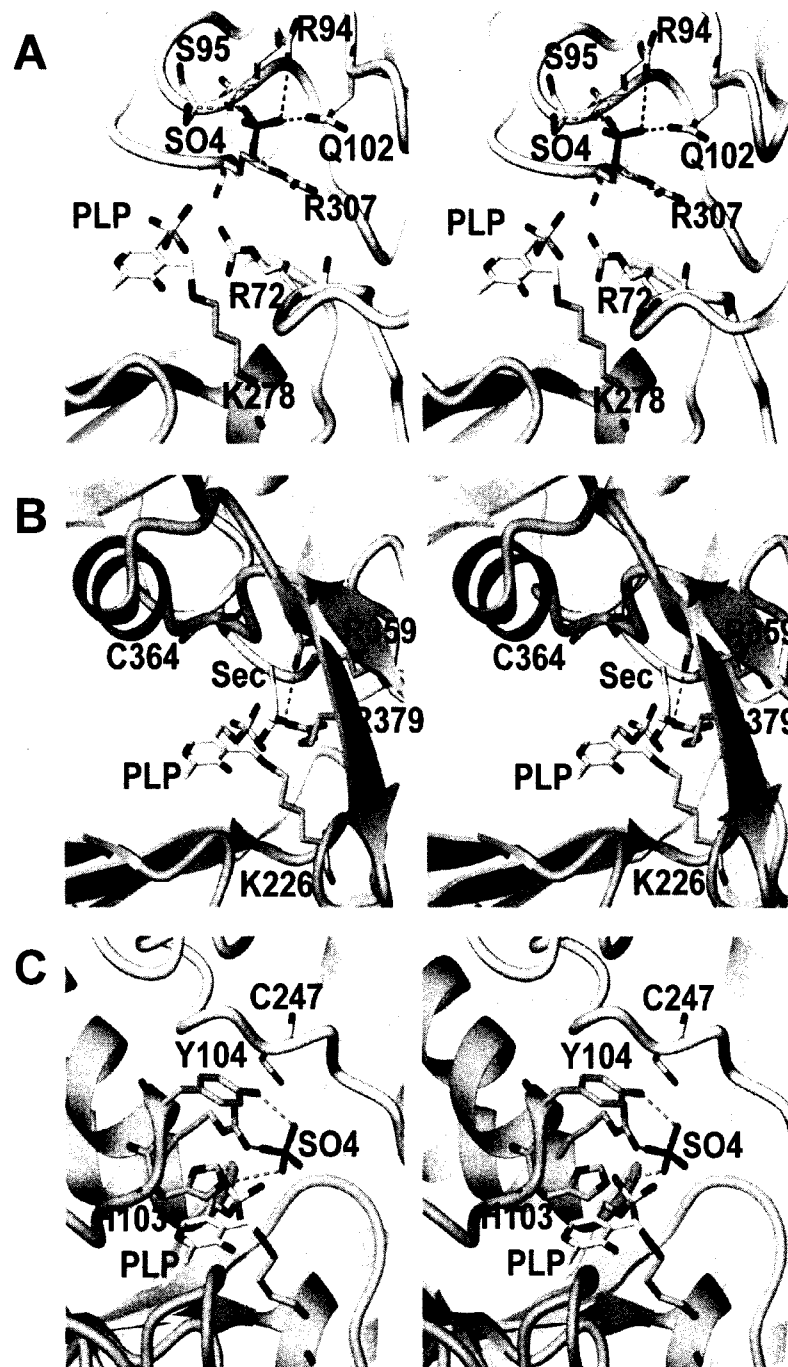


Figure 18: The active site in MMPSepSecS, ECCsdB, and AFSepCysS (Adapted from Araiso et al., 2008). Chains A and B are colored pink and blue, respectively. PLP is covalently bound to chain A. (A) The sulfate ion is shown as a ball-and-stick model in the active site of MMPSepSecS. (B) In the active site of ECCsdB PLP is covalently bound to Lys226 of chain A and is covered by the $\beta\alpha\beta$ structural motif. (C) In the active site of AFSepCysS, PLP is covalently bound to Lys209 of chain A. The orientation of PLP is the same as that in (A). The chain A residues that recognize PLP and sulfate are shown.

Indeed, the quaternary structure of the human SepSecS complexed with tRNA^{Sec}, Sep and thiophosphate revealed that the phosphate moiety of Sep occupies the same position in the active site of the human enzyme as the sulfate ion in the MMPSepSecS structure (reviewed in Results, section D) (Palioura et al., 2009). Mutations of Arg94, Gln102 and Arg307 of chain B that recognize the sulfate ion (Figure 18A) significantly decreased the catalytic activity of MMPSepSecS both *in vivo* (Figures 14 and 17) and *in vitro* (Figure 16) (Araiso et al., 2008).

The MMPSepSecS structure reveals three conserved arginines (Arg72, Arg94 and Arg307) that are located in proximity of each other and close to the active site; they recognize the phosphate groups of PLP and presumably of Sep acylated to tRNA^{Sec} (Figures 12 and 18A). Mutations of Arg72, Arg94 and Arg307 to Ala, Gln or Lys yielded MMPSepSecS enzymes that were unable to form Sec-tRNA^{Sec} *in vivo* (Figures 14 and 17) (Araiso et al., 2008). Asp277 interacts electrostatically with Arg72, and the Asp277Ala enzyme was also inactive *in vivo* (Figure 14). The Arg72Gln, Arg94Gln and Arg307Gln MMPSepSecS mutants were inactive in forming Cys-tRNA^{Sec} *in vitro* (Figure 16) (Araiso et al., 2008). In addition to the PLP-conjugated Lys278, the side chains of the Sep-binding arginines may also facilitate general acid/base catalysis as was shown for the tRNA modification enzyme TrmH where an arginine activated by a phosphate group acts as a general base (Nureki et al., 2004).

Finally, to investigate whether the reaction proceeds through an enzyme-bound phosphorylated intermediate I mutated the two Ser residues located close to the active site, Ser95 and Ser168. Ser95 interacts with the sulfate ion, while Ser168 recognizes the

PLP phosphate moiety. Since the Ser95Ala, Ser95Thr and Ser168Ala mutants are active in forming Sec-tRNA^{Sec} *in vivo* (Figure 17), a phosphorylated state of the enzyme is unlikely to be part of the reaction mechanism.

D. The human SepSecS-tRNA^{Sec} complex

This part of my thesis research was done in collaboration with Miljan Simonović. I crystallized the human SepSecS-tRNA^{Sec} complex and performed in vivo and in vitro activity assays and Dr. Simonović solved the structure.

1. SepSecS can bind tRNA^{Sec} only as a tetramer

In order to investigate the mode of binding of tRNA^{Sec} to SepSecS and the role of PLP in the catalytic mechanism, we determined the crystal structure of the quaternary complex between human SepSecS, unacylated tRNA^{Sec}, Sep and Thiophosphate to 2.8 Å resolution. In this complex, SepSecS forms a tetramer that is bound to two tRNA^{Sec} molecules (Figure 19) (Palioura et al., 2009). The physiological tetramer in this crystal is the same as that found in studies of the archaeal and murine apo-enzyme (Figure 19) (Araiso et al., 2008; Ganichkin et al., 2008). Each monomer has a PLP cofactor covalently linked to the N ϵ -amino group of the conserved Lys284 via formation of a Schiff base (internal aldimine). Two SepSecS monomers form a homodimer, and two active sites are formed at the dimer interface. The two homodimers associate into a tetramer through interactions between the N-terminal α 1-loop- α 2 motifs (Figure 19C, D). Given that SepCysS, the closest PLP Fold Type I homologue of SepSecS that also acts on tRNA-bound Sep (Sep-tRNA^{Cys}), is a dimer and that the active sites of one homodimer do not

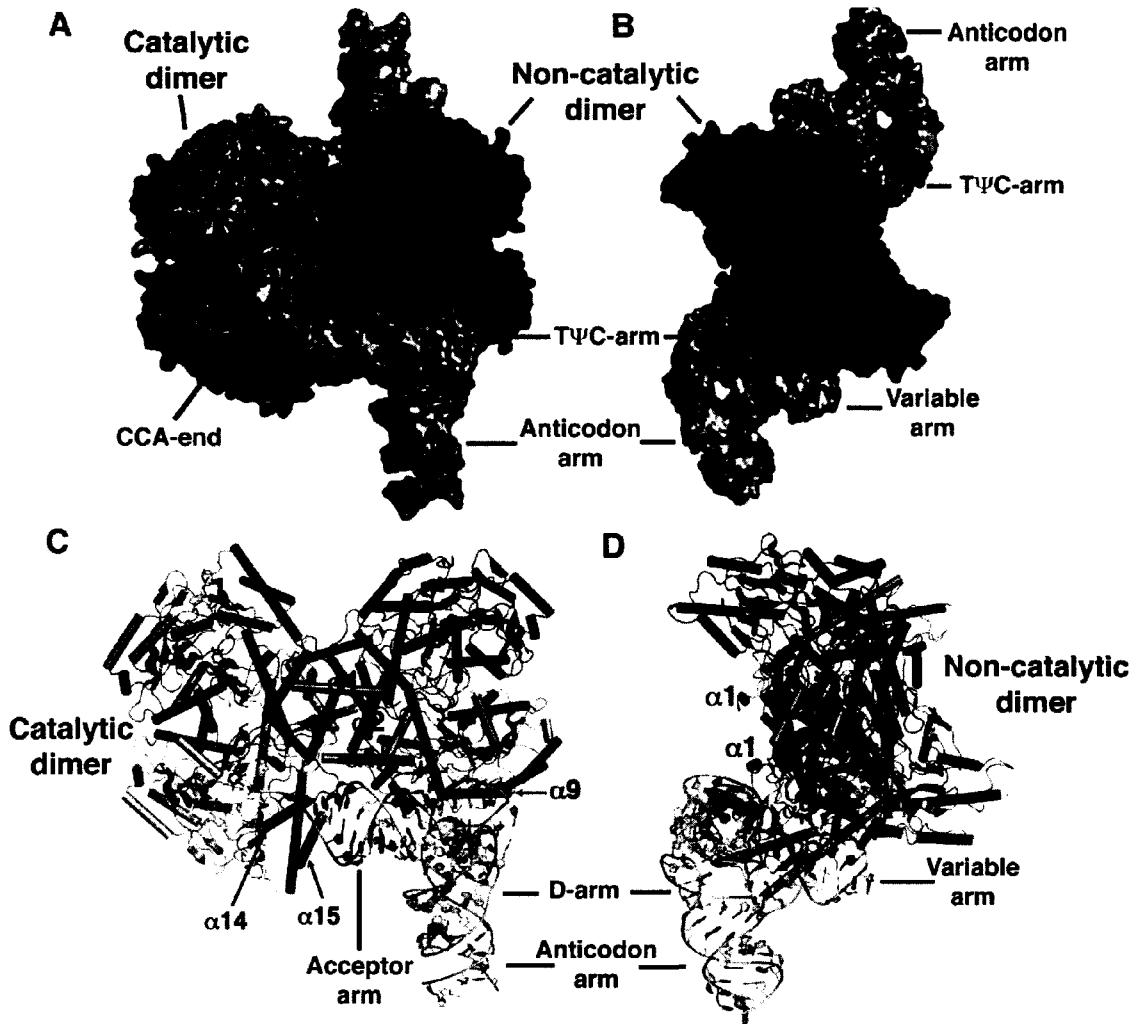


Figure 19: Structure of human SepSecS in complex with unacylated tRNA^{Sec} (Adapted from Palioura et al., 2009). (A, B) Surface representation of the complex of SepSecS with tRNA^{Sec}. The subunits of the catalytic dimer are dark and light blue, those of the noncatalytic dimer are dark and light red; the backbone and bases of tRNA^{Sec} are green and gray, respectively. (C, D) Ribbon diagram of the SepSecS-tRNA^{Sec} complex. The catalytic dimer interacts with the acceptor arm of tRNA^{Sec} through helices $\alpha 14$ and $\alpha 15$ (blue). The $\alpha 1$ helix (red) of the noncatalytic dimer interacts with the rest of the acceptor-T Ψ C arm. The variable arm of tRNA^{Sec} interacts with helix $\alpha 9$ (red) of the noncatalytic dimer. The $\alpha 1$ -loop- $\alpha 2$ motif is important for the tetramerization of the enzyme and is annotated in both dimers. The regions of tRNA^{Sec} that interact with SepSecS are shown in orange; the rest is green. One tRNA^{Sec} molecule is shown for clarity. The view in B and D is rotated $\sim 90^\circ$ clock-wise around the vertical axis relative to that in A and C.

communicate with the active sites of the other homodimer in the apo-SepSecS tetramer (Araiso et al., 2008; Ganichkin et al., 2008), the reason for the tetrameric organization of SepSecS was not known. The mode of tRNA^{Sec} binding to SepSecS provides an answer to this question.

The CCA-ends of both tRNA^{Sec} molecules point to the active sites of the same homodimer, which will be referred to as the catalytic dimer (Figure 19). The other homodimer, which will be referred to as the noncatalytic dimer, serves as a binding platform that orients tRNA^{Sec} for catalysis (Figure 19) (Palioura et al., 2009). The structure reveals that SepSecS binds only to the acceptor-, TΨC- and variable arms of tRNA^{Sec} (Figure 19C, D). The tip of the acceptor arm interacts with the C-terminus of the catalytic dimer, whereas the rest of the acceptor-, TΨC- and variable arms wrap around a monomer from the non-catalytic dimer (Figure 19). The most important binding element is an interaction between the discriminator base G73 of tRNA^{Sec} and the conserved Arg398 of the catalytic dimer (Figure 20A) (Palioura et al., 2009). The guanidinium group of Arg398 forms hydrogen bonds with the Hoogsteen face of G73. The discriminator base G73 of tRNA^{Sec} is universally conserved in archaea and eukaryotes. Neither adenine nor cytidine in position 73 could form hydrogen bonds with Arg398 because they have amino-groups instead of the keto-group. Also, if a cytidine or uridine was in position 73, the C5 and C6 atoms of the pyrimidine ring would clash with the side-chain of Thr397, thus, preventing the interaction of these bases with Arg398.

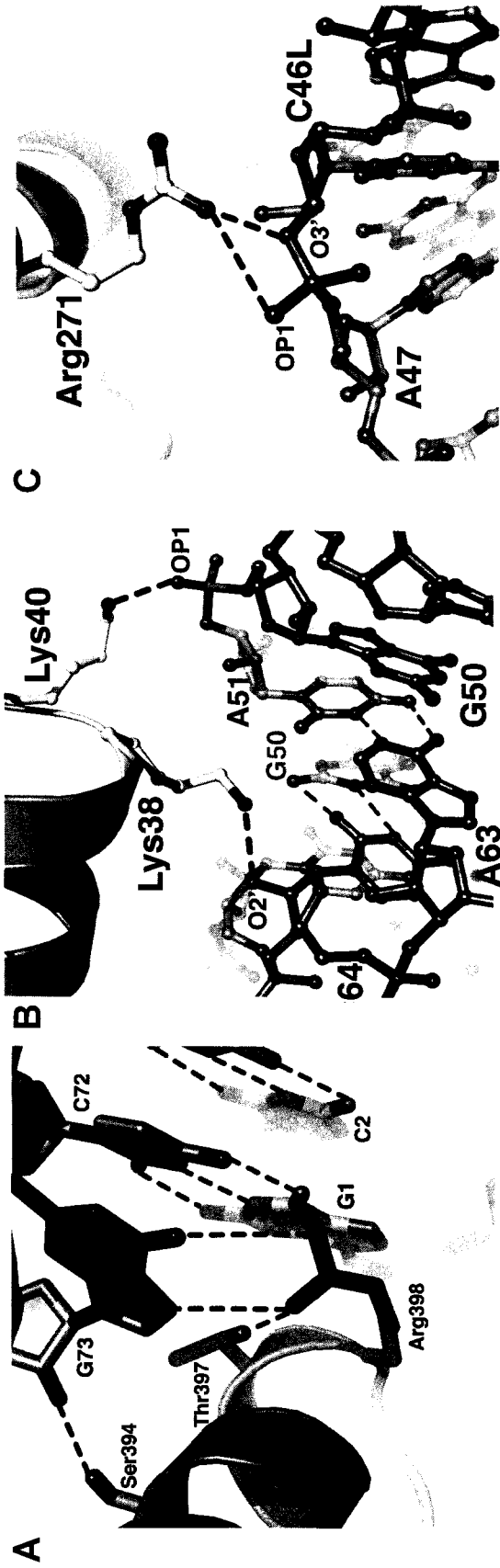


Figure 20: Close-up view of interactions between human SepSecS and tRNA^{Sec} (Adapted from Palioura et al., 2009) (A) Interactions between the discriminator base G73 and the conserved Arg398 in the α 14- β 11 loop. In A, B and C, tRNA^{Sec} is green and the protein side-chains are gold. (B) The side-chains of Lys38 and Lys40 (α 1- α 2 loop) interact with the T Ψ C arm. Lys38 and Lys40 interact with C64 and G50, respectively. (C) Arg271 (α 9) in the noncatalytic dimer interacts with the variable arm (nucleotides C46L-A47).

The majority of interactions between the enzyme and tRNA^{Sec}, however, occur between the highly conserved basic residues in the N-terminal α 1 helix of the noncatalytic dimer and the phosphate-sugar backbone of the acceptor-T Ψ C arm (Figure 19C, D). For instance, the side-chain of Arg26 forms hydrogen bonds with the phosphate backbone of C2-C3 of the acceptor arm, the N ϵ -amino group of Lys38 forms a hydrogen bond with the 2'-hydroxyl group of C64 (Figure 20B), whereas the side-chain of Lys40 interacts with the phosphate oxygen of G50 from the T Ψ C arm (Figure 20B). Moreover, the 13-base pair (bp) length of the acceptor-T Ψ C arm is important for positioning the variable arm to interact with the neighboring noncatalytic dimer. The nucleotides C46A and C46L interact with Ser260 and Arg271 from the noncatalytic dimer, respectively (Figure 20C) (Palioura et al., 2009).

The interactions between SepSecS and tRNA^{Sec} observed in the crystal are consistent with *in vivo* activity assays of SepSecS mutants (Figure 21) (Palioura et al., 2009). Replacing Arg398 with either alanine or glutamate renders the enzyme completely inactive suggesting that the interaction between the discriminator base and the highly conserved Arg398 of the catalytic dimer is critical for tRNA^{Sec} recognition (Figure 21). Furthermore, Thr397Ala is completely active, Thr397Val is partially active and Thr397Glu is inactive *in vivo* (Figure 21). Thus, Thr397 strengthens the interaction of the acceptor arm with the α 14 helix. In addition, the Arg271Glu mutant is partially active suggesting that the interaction between the variable arm and the noncatalytic dimer is important for complex formation as well (Figure 21). When Lys38 is replaced by

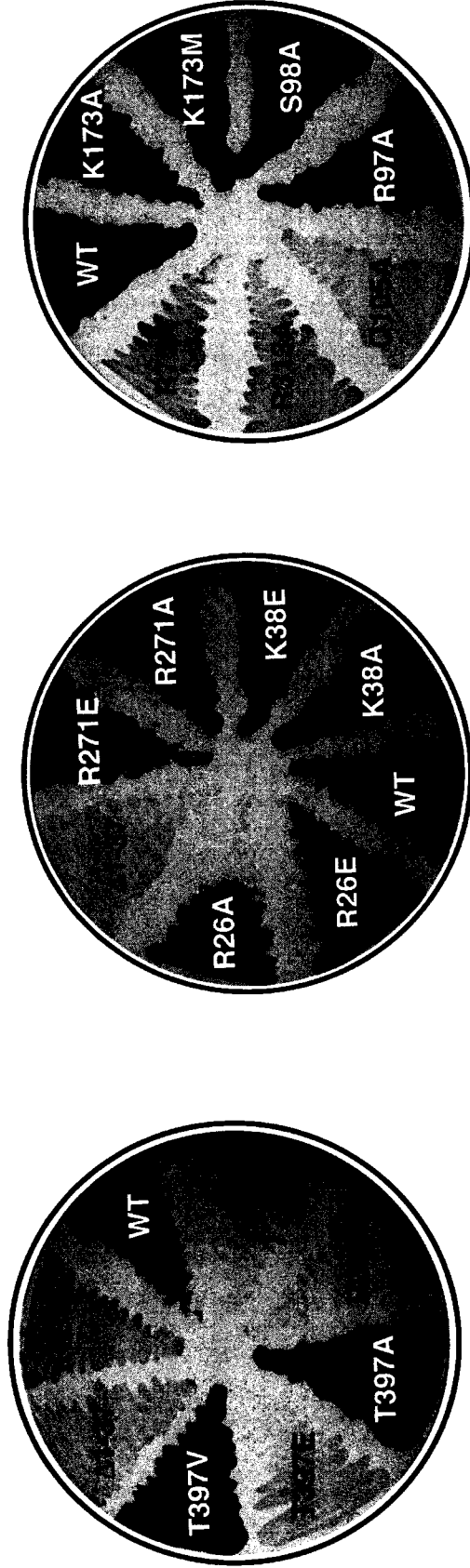


Figure 21: *In vivo* assays of SepSecS mutants (Taken from Palioura et al., 2009). Formation of Sec-tRNA^{Sec} *in vivo* is assayed by the ability of the wild-type human SepSecS and its mutant variants (N-terminal deletion Δ 1-37, R398A, R398E, R26A, R26E, K38A, K38E, T397A, T397E, T397V, R271A, R271E, K173A, K173M, S98A, R97A, Q105A, R313A and R75A) to restore the benzyl viologen reducing activity of the selenoprotein FDH_H in the *E. coli* Δ selA deletion strain. The *M. jannaschii* PSTK is co-transformed, except where indicated with a “-”, to provide the necessary Sep-tRNA^{Sec} intermediate.

Glu, the enzyme is partially active *in vivo*, whereas both the Arg26Ala and Arg26Glu mutants are completely active (Figure 21), suggesting that the Arg26-C2 interaction alone is not critical for complex formation. Finally, both the human and archaeal SepSecS enzymes are completely inactive *in vivo* when the α 1 helix is deleted (Figure 21) (Araiso et al., 2008; Palioura et al., 2009).

2. The antigenic region

In the structure of the SepSecS-tRNA^{Sec} binary complex, the first twelve residues of the proposed antigenic region (residues 452-463) form the C-terminal helix α 15. The remaining residues of the antigenic region (residues 463-501) form the extreme C-terminal tail of SepSecS and have been disordered in all crystal structures of the enzyme determined to date (Araiso et al., 2008; Ganichkin et al., 2008). Interestingly, helix α 15 is spatially located near the entrance to the active-site cleft and it is proximal to helix α 14 (Figures 19B and 22A) (Palioura et al., 2009; Palioura et al., 2010). Both α 14 and α 15 interact with distinct parts of the acceptor arm (Figure 22) (Palioura et al., 2010). The side chains of Thr397 and Arg398 from helix α 14 interact with the discriminator base G73 and, thus, establish the identity of the bound tRNA molecule (Figure 22B). The importance of Thr397 and Arg398 for tRNA^{Sec} binding has been confirmed in the *in vivo* assays (Figure 21) (Palioura et al., 2009). On the other hand, the residues Arg453, Arg456 and Lys463 in α 15 form the 5'-phosphate binding groove and they interact with the tRNA backbone atoms (Figure 22B) (Palioura et al., 2010).

Since autoantibodies from patients with autoimmune hepatitis can precipitat

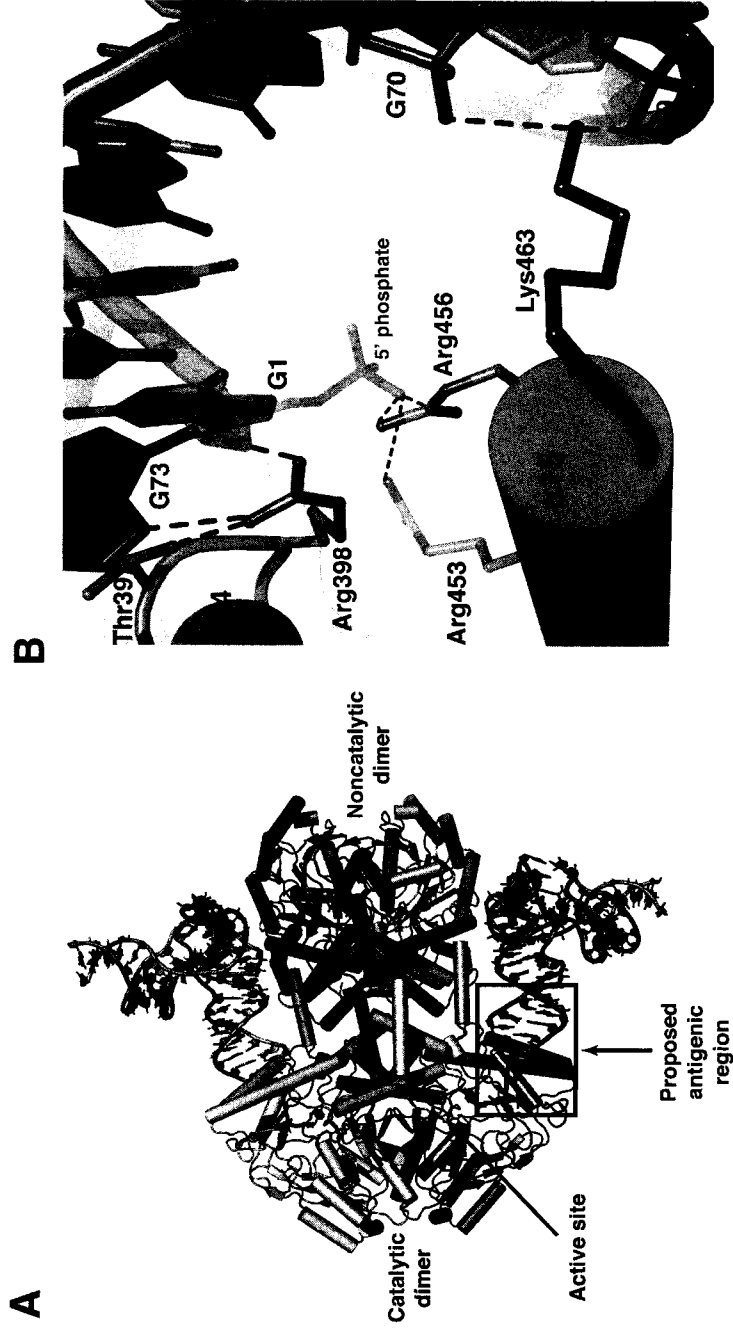


Figure 22: The proposed antigenic region in SepSecS or SLA/LP (Taken from Palioura et al., 2010). (A) The putative antigenic region of SepSecS is located near the active site of the enzyme (arrow). This region interacts with the tip of the acceptor arm of tRNA^{Sec} and is critical for tRNA recognition. The catalytic dimer of SepSecS is in shades of blue, the non-catalytic dimer is colored in shades of pink, two molecules of tRNA^{Sec} are green, and the antigenic region is orange and demarcated with a box. (B) A close-up view of the interactions at the enzyme-tRNA interface. Residues from helices $\alpha 14$ and $\alpha 15$ interact with the tip of tRNA^{Sec}. The helix $\alpha 14$ residues interact with the discriminator base G73: Arg398 forms hydrogen bonds with the Hoogsteen face of G73, while Thr397 stabilizes this interaction. The C-terminal helix $\alpha 15$ interacts with the tRNA backbone and with the 5' phosphate. Arg453 and Arg456 are within hydrogen bonding distance from the 5' phosphate, whereas Lys463 interacts with the non-bridging oxygens of C69 and G70.

the ribonucleoprotein SepSecS-tRNA^{Sec} complex, such autoantibodies would bind to an interface that lies between the $\alpha 14$ and $\alpha 15$ helices of SepSecS and the tip of the acceptor arm of tRNA^{Sec} (Figure 22B). Binding of the antibody near the active site of SepSecS may, in fact, inhibit the enzyme's function directly or indirectly by stabilizing the entire protein-RNA complex and preventing release of tRNA^{Sec}. The inhibitory effect of antibodies would resemble the mechanism by which LKM-1 autoantibodies inhibit cytochrome P450 isoenzyme 2D6 (CYP2D6) in a subgroup of autoimmune hepatitis patients (Manns et al., 1989).

3. The unique 9/4 fold of tRNA^{Sec} is recognized by the Sec-specific proteins

Human tRNA^{Sec} contains 90 nucleotides rather than the conventional 75 nucleotides of canonical tRNA molecules. The structure shows that human tRNA^{Sec} adopts a unique 9/4 fold with a 13 bp acceptor-T Ψ C arm (where 9 and 4 reflect the number of base pairs in the acceptor and T Ψ C arms, respectively) and a long variable arm (Figure 23) (Palioura et al., 2009). This resolves a controversy between conflicting models in which the 7/5 model suggested 12 bp in the acceptor-T Ψ C arm as found in all known tRNA structures (Ioudovitch and Steinberg, 1998), whereas the 9/4 model suggested a unique 13 bp acceptor-T Ψ C arm (Stürchler et al., 1993).

In the 9/4 fold of the human tRNA^{Sec}, the linker between the acceptor- and the D-stem is two residues long (A8-U9) as was proposed earlier (Stürchler et al., 1993). The base of A8 does not participate in any tertiary interactions, while the base of U9 stacks with A49 from the variable arm-T Ψ C linker. The first pair of the D-arm is a C10-G2

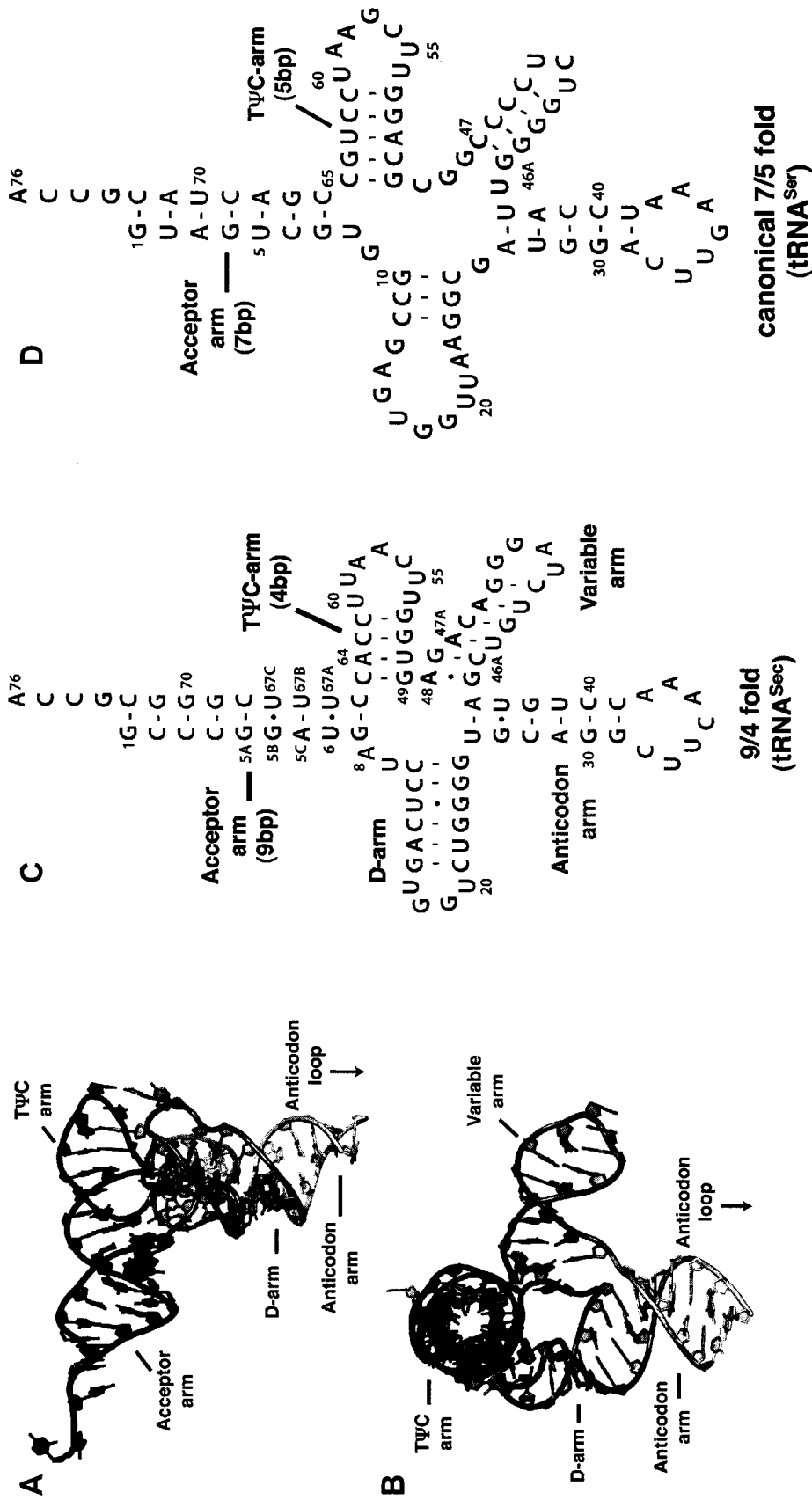


Figure 23: Structure of human tRNA^{Sec} compared to the canonical tRNA^{Ser} (Adapted from Palioura et al., 2009). (A) Ribbon diagram of the human tRNA^{Sec} molecule observed in complex with SepSecS. The major structural elements are colored as follows: the acceptor arm is red, the D-arm is blue, the anticodon arm is light blue, the variable arm is orange and the TΨC arm is dark red. (B) The view is rotated by ~90° clockwise around the vertical axis. Secondary structure diagram of (C) human tRNA^{Sec} and (D) tRNA^{Ser}.

couple followed by a C11-G24 pair (Figure 23C) (Palioura et al., 2009). Four nucleotides U16, G18, G19 and U20 form the D-loop and interact with U55, C56, A57 and A58 from the TΨC loop closing the cloverleaf structure as proposed (Stürchler et al., 1993). The rest of the tRNA^{Sec} structure agrees well with the proposed model (Stürchler et al., 1993).

The 13 bp acceptor-TΨC arm and a long variable arm are distinct structural features that serve as major recognition motifs for binding to SepSecS (Figure 19). The structure of the human SepSecS in complex with tRNA^{Sec} along with the *in vivo* mutagenesis data reveal that SepSecS can bind tRNA^{Sec} only as a tetramer. The interaction between Arg398 and G73 along with the longer 13 bp acceptor-TΨC arm and the variable arm of tRNA^{Sec} secure binding of the tRNA to the enzyme. The only tRNA with a G73 discriminator base and a long variable arm yet with a 12 bp acceptor-TΨC arm is tRNA^{Ser}. However, neither unacylated tRNA^{Ser} nor Ser-tRNA^{Ser} is able to bind to SepSecS (Xu et al., 2007). To examine the importance of the length of the acceptor-TΨC arm for SepSecS binding, we superimposed tRNA^{Asp} (Eiler et al., 1999) onto tRNA^{Sec} (Figure 24) (Palioura et al., 2009). The tRNA^{Asp} was chosen because it is a canonical tRNA molecule with G73 as the discriminator base and because the complete structure of tRNA^{Ser} is not available. Two approaches were used for superimposing the two tRNAs. In the first approach, the atoms of the sugar-phosphate backbone of the acceptor-, TΨC- and anticodon arms were superimposed. Although the secondary structure elements of tRNA^{Asp} were oriented as observed in tRNA^{Sec}, G73 of tRNA^{Asp} could not reach Arg398 due to its shorter acceptor-TΨC arm (Figure 24B). In the second approach, the discriminator base of tRNA^{Asp} was first modeled to form hydrogen bonds

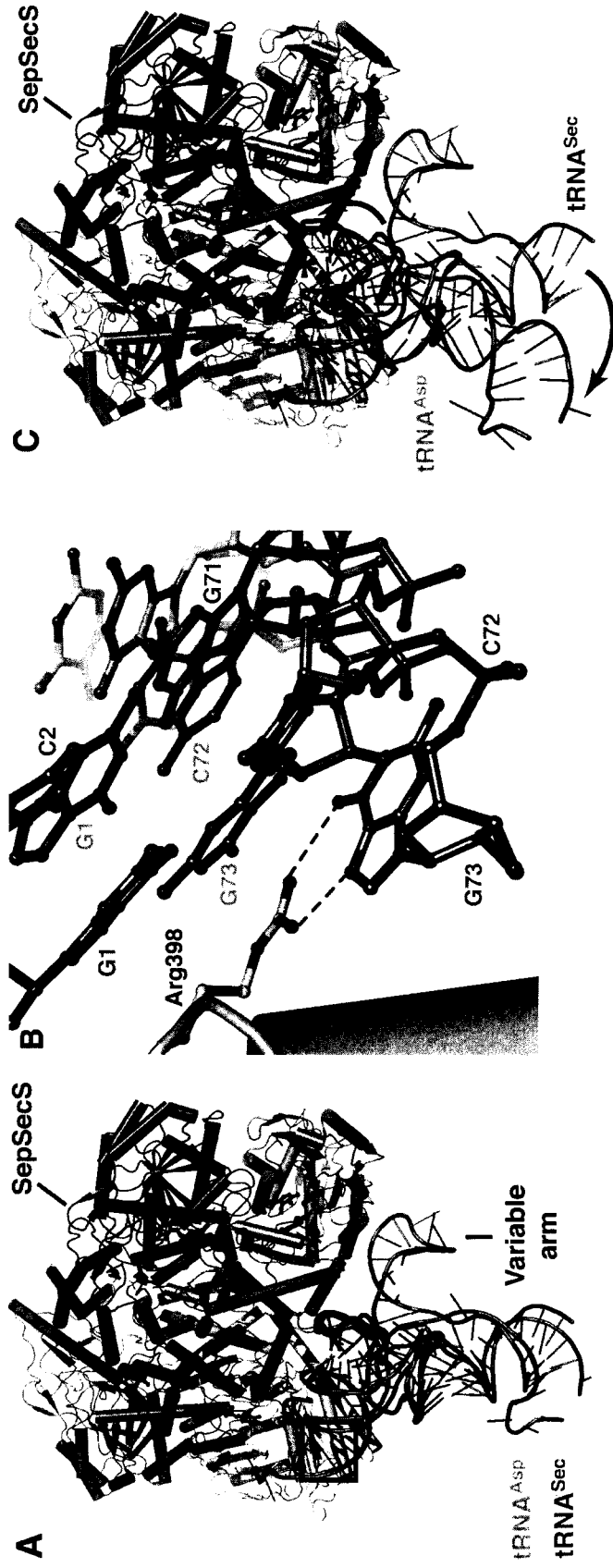


Figure 24: Selectivity of SepSecS towards tRNA^{sec} (Taken from Palioura et al, 2009). (A) Modeling of the complex between SepSecS and canonical tRNAs. tRNA^{Asp} (orange) and tRNA^{Sec} (green) were superimposed onto tRNA^{Sec} (green) using the sugar-phosphate backbone atoms of the acceptor, TΨC and anticodon arms. The long variable arm of tRNA^{Sec} interacts with the noncatalytic dimer of SepSecS (shades of red). The box delineates the close-up view shown in B. (B) The short acceptor-TΨC arm of canonical tRNAs prevents binding to SepSecS. G73 of tRNA^{Asp} (orange) cannot reach Arg398. (C) Steric clashes (asterisks) between the α1 helix of the noncatalytic dimer of SepSecS (shades of red) and the TΨC arm of tRNA^{Asp} (orange) prevent the SepSecS-tRNA^{Asp} complex formation. G73 from tRNA^{Asp} was modeled to form hydrogen bonds with Arg398 using its Hoogsteen face. Only the tip of the acceptor arm (nucleotides 1-3 and 70-73) was used in the superposition. Arrows show the movement of both the anticodon and TΨC arms in tRNA^{Asp} relative to its orientation shown in A.

with Arg398 using its Hoogsteen face, and then the tips of the acceptor arms (nucleotides 1-3 and 70-73) were superimposed. In this case, the anticodon and variable arms of tRNA^{Asp} are positioned away from the enzyme, whereas its TΨC arm clashes with the α 1 helix of the noncatalytic dimer (Figure 24C), thus preventing the formation of the SepSecS-tRNA^{Asp} complex. In addition, the absence of a long and properly oriented variable arm in tRNA^{Asp} diminishes further its capacity for interaction with SepSecS (Figure 24A, C) (Palioura et al., 2009). Thus, modeling of the SepSecS-tRNA^{Asp} complex suggests that tRNA^{Ser} is not able to bind to SepSecS due to its shorter acceptor-TΨC arm.

Further, it has been proposed that the variable arm of tRNA^{Sec} serves as the major recognition motif for its binding to SerRS, which allows Ser-tRNA^{Sec} formation to occur (Bilokapic et al., 2004; Geslain et al., 2006; Leinfelder et al., 1988; Ohama et al., 1994; Rother et al., 2000; Sherrer et al., 2008a; Stürchler-Pierrat et al., 1995; Wu and Gross, 1993, 1994). Modeling of the SerRS-tRNA^{Sec} complex based on the crystal structure of the bacterial SerRS complexed with tRNA^{Ser} (Biou et al., 1994) suggests that SerRS can recognize the variable arm of tRNA^{Sec} similarly to tRNA^{Ser} providing an explanation for the inability of SerRS to discriminate between tRNA^{Ser} and tRNA^{Sec} (Figure 25).

4. Sep binds to the active site only in the presence of tRNA^{Sec}

To explore the mechanism by which the phosphoryl group is converted to the selenocysteinylyl moiety, I soaked crystals of the binary SepSecS - tRNA^{Sec} complex in a solution containing a mixture of *O*-phosphoserine and thiophosphate. I used Sep and

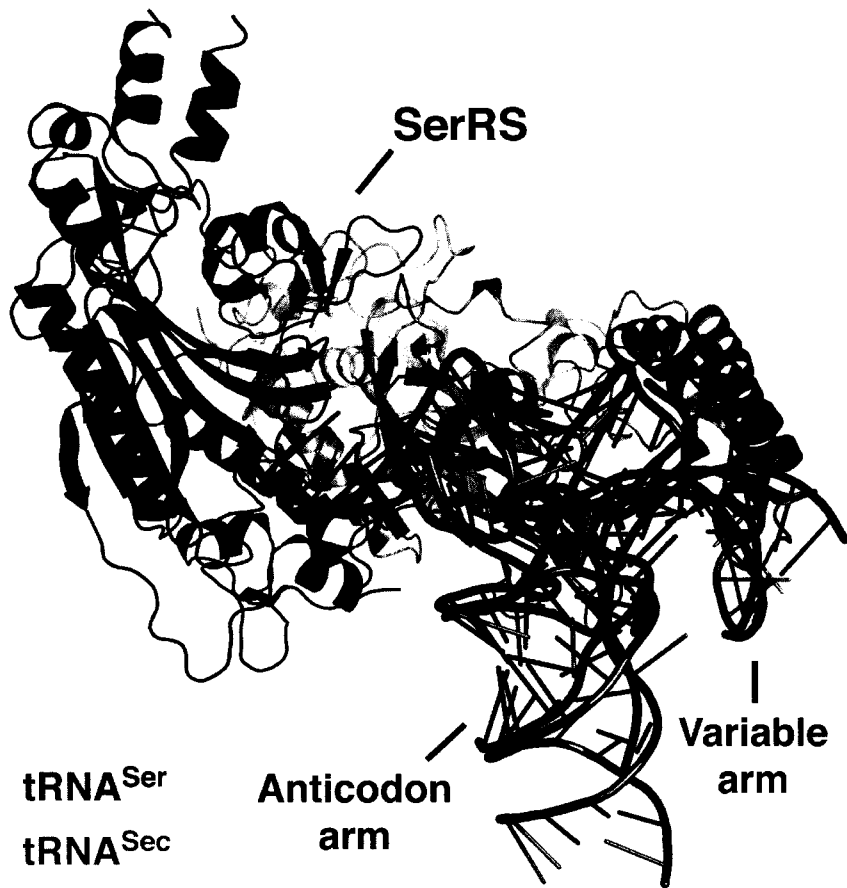


Figure 25: Modeling of the SerRS-tRNA^{Sec} complex (Taken from Palioura et al., 2009). tRNA^{Sec} was superimposed onto the tRNA^{Ser} that is in complex with SerRS. SerRS is a non-discriminating aminoacyl-tRNA synthetase because it recognizes variable arms of both tRNA^{Ser} (red) and tRNA^{Sec} (green)

and thiophosphate as mimics of the phosphoseryl group attached to tRNA^{Sec} and selenophosphate, respectively. Although both ligands were used in the soaking experiments, Sep bound only to the active sites of the catalytic dimer (Figure 26A) (Palioura et al., 2009), whereas thiophosphate bound only to the active sites of the non-catalytic dimer (Figure 26B), i.e. Sep can bind to an active site only in the presence of tRNA^{Sec}. Both thiophosphate and the phosphoryl group of Sep bound to the same binding pocket (Figure 26) suggesting that a specific active site can accommodate only one ligand at a time (Palioura et al., 2009).

Sep binds to the catalytic active site in either of two different orientations (Figure 26A). The phosphoryl group occupies a similar location in the two orientations, whereas the seryl moieties are rotated by $\sim 90^\circ$ around the phosphate group. In what appears to be a non-productive orientation of Sep (Sep^N), its seryl group is sandwiched between the side-chains of Arg97 and Lys173 at the catalytic dimer interface, while its amino group lies ~ 12 Å away from the PLP Schiff base (Figure 27A) (Palioura et al., 2009). Thus, it is unlikely that SepSecS would act on Sep^N, unless a PLP-independent mechanism is utilized. In the productive-like orientation of Sep (Sep^P), the amino-group of Sep is ~ 3.5 Å away from the Schiff base yet it is not positioned for attack due to a hydrogen bond with Gln172 (Figure 27B). The carboxyl group of Sep^P also forms a hydrogen bond with the side-chain of Gln172. Electron density for both the amino and the carboxyl groups of Sep^P is weak suggesting higher mobility in this part of Sep^P. This is presumably due to free rotation around the C α -C β bond. The interactions of the phosphoryl group of Sep^P

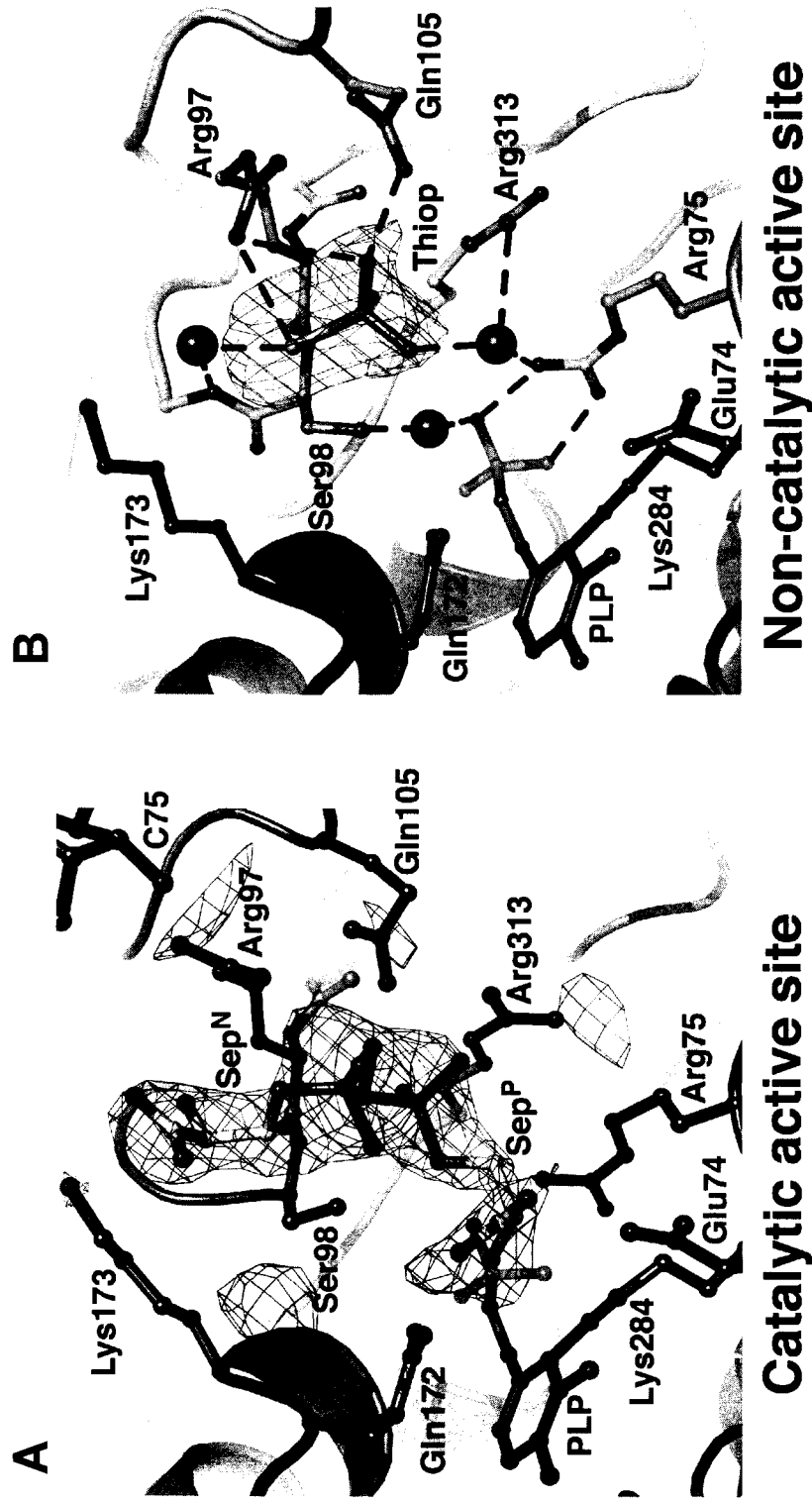


Figure 26: Ligand binding to the active sites in the SepSecS-tRNA^{Sec} complex (Taken from Palioura et al., 2009). (A) Phosphoserine binds to the catalytic sites in two orientations (Sep^P and Sep^N). The catalytic PLP-monomer is purple, the P-loop monomer is light blue, Sep is gold, tRNA is green and PLP is magenta. Simulated annealing omit electron density map (green mesh) is contoured at 3.5 σ . (B) Thiophosphate (Thiop) binds only to the non-catalytic site. The PLP-monomer is brown and the P-loop monomer is pink.

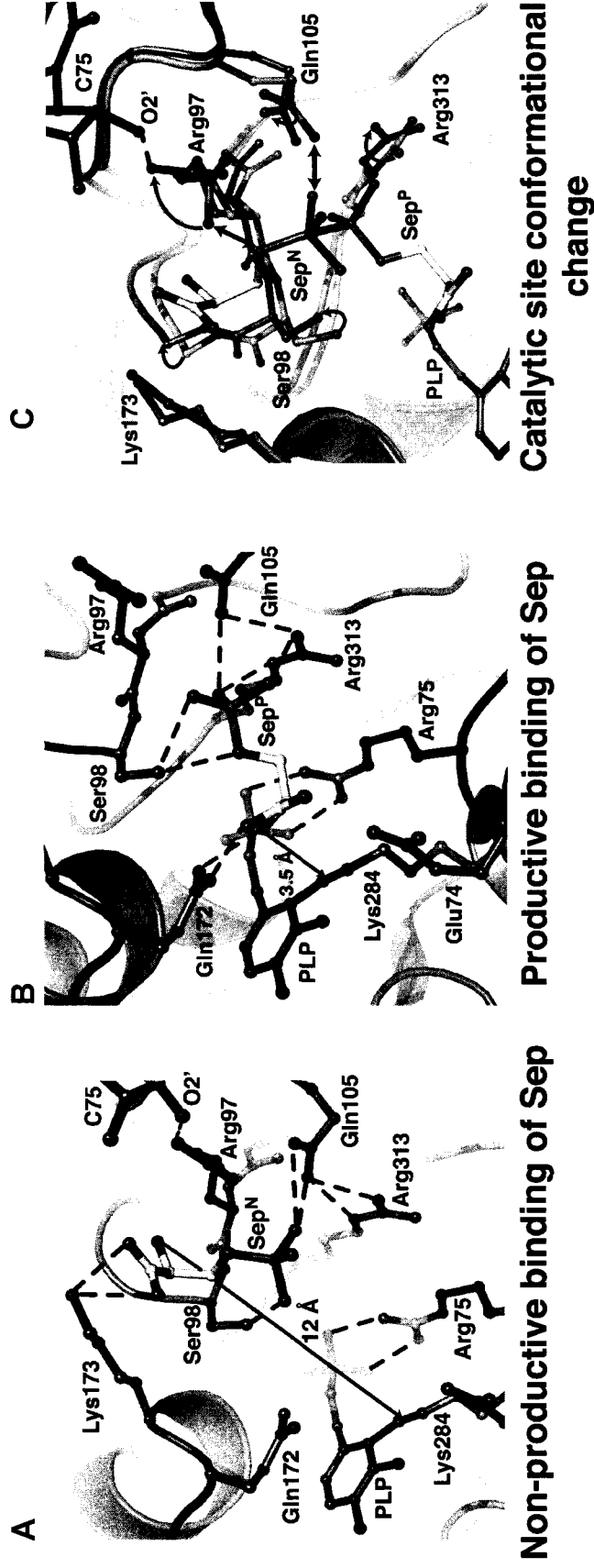


Figure 27: Phosphoserine binds to the active site of the SepSecS-tRNA^{Sec} complex in two orientations (Taken from Palioura et al., 2009). The catalytic PLP-monomer is purple, the P-loop monomer is light blue, Sep is gold, tRNA is green and PLP is magenta. (A) The Sep^N amino-group is ~12 Å away from the Schiff base. Arg97, Gln105 and Arg313 coordinate phosphate and the carboxyl group interacts with Lys173. (B) The amino group of Sep^P interacts with Gln172 and is 3.5 Å away from the Schiff base. The carboxyl group interacts with Lys172, while Ser98, Gln105, and Arg313 coordinate phosphate. (C) The P-loop adopts a different conformation upon tRNA^{Sec} binding. The noncatalytic dimer is pink, the catalytic dimer is light blue. Steric clashes between the noncatalytic P loop and Sep^N are shown (double arrow).

with the side-chains of Ser98, Gln105 and Arg313 anchor Sep in the active site. That the phosphoryl group is required to properly position the ligand for catalysis explains why the obligate substrate for SepSecS is Sep-tRNA^{Sec} and not Ser-tRNA^{Sec}, which is the physiological substrate for the bacterial selenocysteine synthase.

A comparison of the catalytic and noncatalytic sites reveals that Sep binds to the active site only in the presence of tRNA^{Sec} because the conformation of the P-loop (residues Gly96-Lys107) differs between the two active sites (Figure 27C). In the noncatalytic dimer the guanidinium group of Arg97 and the side-chain of Gln105 are rotated towards the phosphate-binding groove where they coordinate thiophosphate (Figures 26B and 27C) (Palioura et al., 2009). In the catalytic dimer, the side-chain of Arg97 rotates away from the phosphate-binding groove, and forms a hydrogen bond with the 2'-OH group of C75. Gln105 also rotates away from the phosphate and this concerted movement of Arg97 and Gln105 in the P-loop upon tRNA^{Sec} binding allows Sep to bind to the active site (Figure 27C). Free Sep enters the active site through the gate formed by the side-chains of Arg97, Gln105 and Lys173. The affinity for the Sep^N is high enough to trap the free ligand in this orientation as well. Moreover, the free amino group of Sep^P cannot be positioned appropriately for attack onto the Schiff base probably due to free rotation around the C α -C β bond explaining why the reaction does not occur in the crystal. Thus, the covalent attachment of Sep to tRNA^{Sec} is necessary for the proper placement of the Sep moiety into the active site and for orienting the amino-group of Sep for attack onto the Schiff base of PLP.

5. SepSecS utilizes a PLP-dependent mechanism for Sec-tRNA^{Sec} formation

I used both *in vivo* and *in vitro* activity assays to investigate the mechanism of Sep-tRNA^{Sec} to Sec-tRNA^{Sec} conversion by human SepSecS. Several lines of evidence suggest that SepSecS catalyzes the Sep-tRNA^{Sec} to Sec-tRNA^{Sec} conversion in a PLP-dependent reaction. First, the reduction of the Schiff base by sodium borohydride to form a chemically stable secondary amine and, thus, cross-link PLP to Lys284 renders SepSecS completely inactive *in vitro* (Figure 28) (Palioura et al., 2009). The catalytic activity of SepSecS is also quenched upon removal of PLP by treatment with hydroxylamine (Figure 28) (Palioura et al., 2009). Some residual activity that is observed upon hydroxylamine treatment is probably due to an incomplete removal of PLP (Figure 28). Second, we show that the Arg75Ala, Gln105Ala and Arg313Ala mutants are inactive *in vivo* (Figure 21) (Palioura et al., 2009). These residues are involved in coordinating either the phosphate group of PLP or that of Sep. Finally, the *in vivo* activities of the Arg97Ala, Arg97Gln, Lys173Ala and Lys173Met mutants are indistinguishable from that of the wild-type enzyme confirming that Arg97 and Lys173 are involved only in the non-productive binding of free Sep (Figure 21). Therefore, a PLP-dependent mechanism for the Sep to Sec conversion is proposed (Figure 29) (reviewed in Discussion, section D) (Palioura et al., 2009).

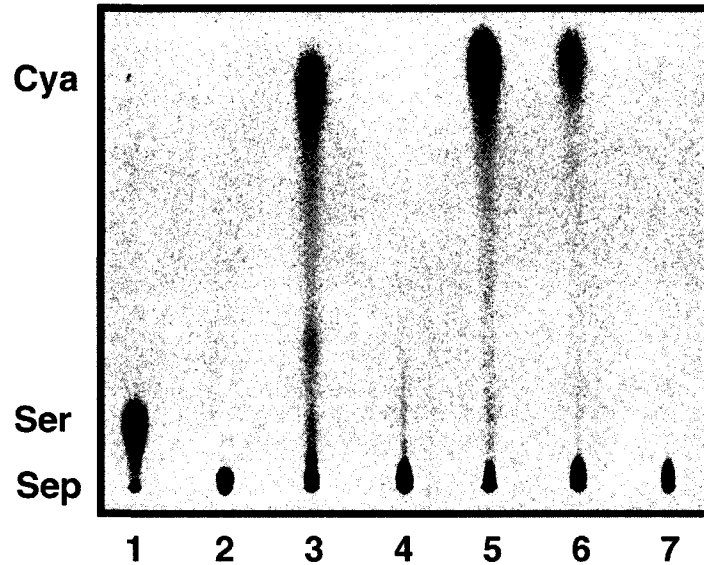


Figure 28: *In vitro* conversion of Sep-tRNA^{Sec} to Cys-tRNA^{Sec} by wild type and PLP-cross-linked SepSecS (Taken from Palioura et al., 2009). Phosphorimages of TLC separation of [¹⁴C]Sep and [¹⁴C]Cys recovered from the aminoacylated-tRNAs of the SepSecS activity assays. Cysteine was analyzed in its oxidized form as cysteic acid (Cya). Lane 1, Ser marker; lane 2, Sep marker; lane 3, Cys marker; lane 4, Sep-tRNA^{Sec} with human SepSecS in the absence of thiophosphate; lane 5, Sep-tRNA^{Sec} with wild-type human SepSecS; lane 6, Sep-tRNA^{Sec} with human SepSecS treated with 7.5 mM hydroxylamine; lane 7, Sep-tRNA^{Sec} with human SepSecS treated with 100 mM sodium borohydride.

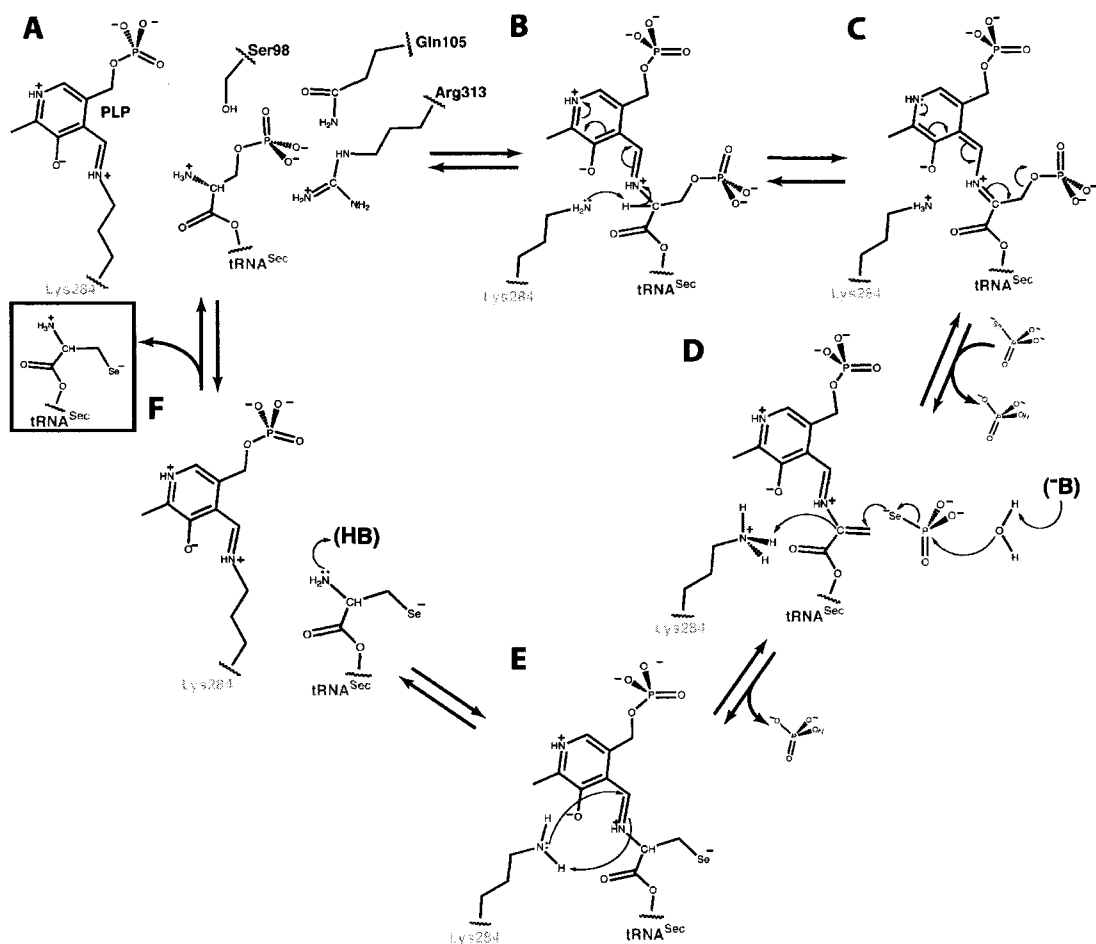


Figure 29: The PLP-dependent mechanism of Sep to Sec conversion (Taken from Palioura et al., 2009). (A) The phosphoseryl moiety of Sep-tRNA^{Sec} is bound to the active site similarly to Sep^P. The amino group is oriented for attack onto the Schiff base (internal aldimine), while the phosphoryl group is stabilized by the side-chains of Ser98, Gln105 and Arg313. Hydrogen bonds are shown in dashed lines. (B) After the formation of the external aldimine, the side-chain of Lys284 rearranges and abstracts the C α proton from Sep. The protonated pyridine ring of PLP stabilizes the carbanion. (C) Electron delocalization leads to a rapid β -elimination of phosphate and to the formation of dehydroalanyl-tRNA^{Sec}. Free phosphate dissociates and selenophosphate binds to the active site. (D) An unidentified base (-B) activates water that hydrolyzes selenophosphate. Free phosphate dissociates again, and selenium attacks the dehydroalanyl-tRNA^{Sec}. Lys284 returns the proton to the C α carbon and the selenocysteiny moiety is formed. (E) The reaction of reverse transaldimination is shown. Lys284 forms the Schiff base with PLP leading to a release of the oxidized form of Sec-tRNA^{Sec} (red box). (F) The free amino-group of Sec-tRNA^{Sec} is protonated, and the active site of SepSecS is regenerated.

DISCUSSION

A. SepSecS: the missing step in Sec biosynthesis in eukaryotes and archaea

We used *in vivo* complementation assays and *in vitro* activity assays to establish the human SLA/LP and its archaeal and trypanosomal orthologs as Sela analogs able to catalyze Sec-tRNA^{Sec} formation. However, in contrast to bacterial Sela that converts Ser-tRNA^{Sec} directly to Sec-tRNA^{Sec}, SepSecS acts on the phosphorylated intermediate, Sep-tRNA^{Sec} (Figure 30) (Palioura et al., 2009). Therefore, the bacterial Sela is replaced by PSTK and SepSecS in archaea and eukaryotes. Thus, archaeal and eukaryal Sec biosynthesis requires three enzymes, SerRS, PSTK, and SepSecS, whereas bacteria accomplish the same task without PSTK. Interestingly, both Sela and SepSecS are PLP-dependent enzymes. From a chemical standpoint, a Sep → Sec conversion is desirable, as Sep would provide a better leaving group (phosphate) than Ser (water) for replacement with selenium. Sep-tRNA is reported to be more stable than Ser-tRNA (Carlson et al., 2004), which is known to deacylate faster than any other aa-tRNA species (Matthaei et al., 1966). This would presumably improve the overall efficiency of the selenocysteinylation reaction and the rate of production of Sec-tRNA in all organisms that possess an extended selenoproteome. Most selenoprotein-containing bacteria only have one to three selenoproteins that are expressed under specific environmental conditions and a nonessential tRNA^{Sec} (Kryukov and Gladyshev, 2004). In contrast, humans have an essential tRNA^{Sec} and an extended selenoproteome whose 25 members have been variously implicated in health and disease (Lobanov et al., 2009).

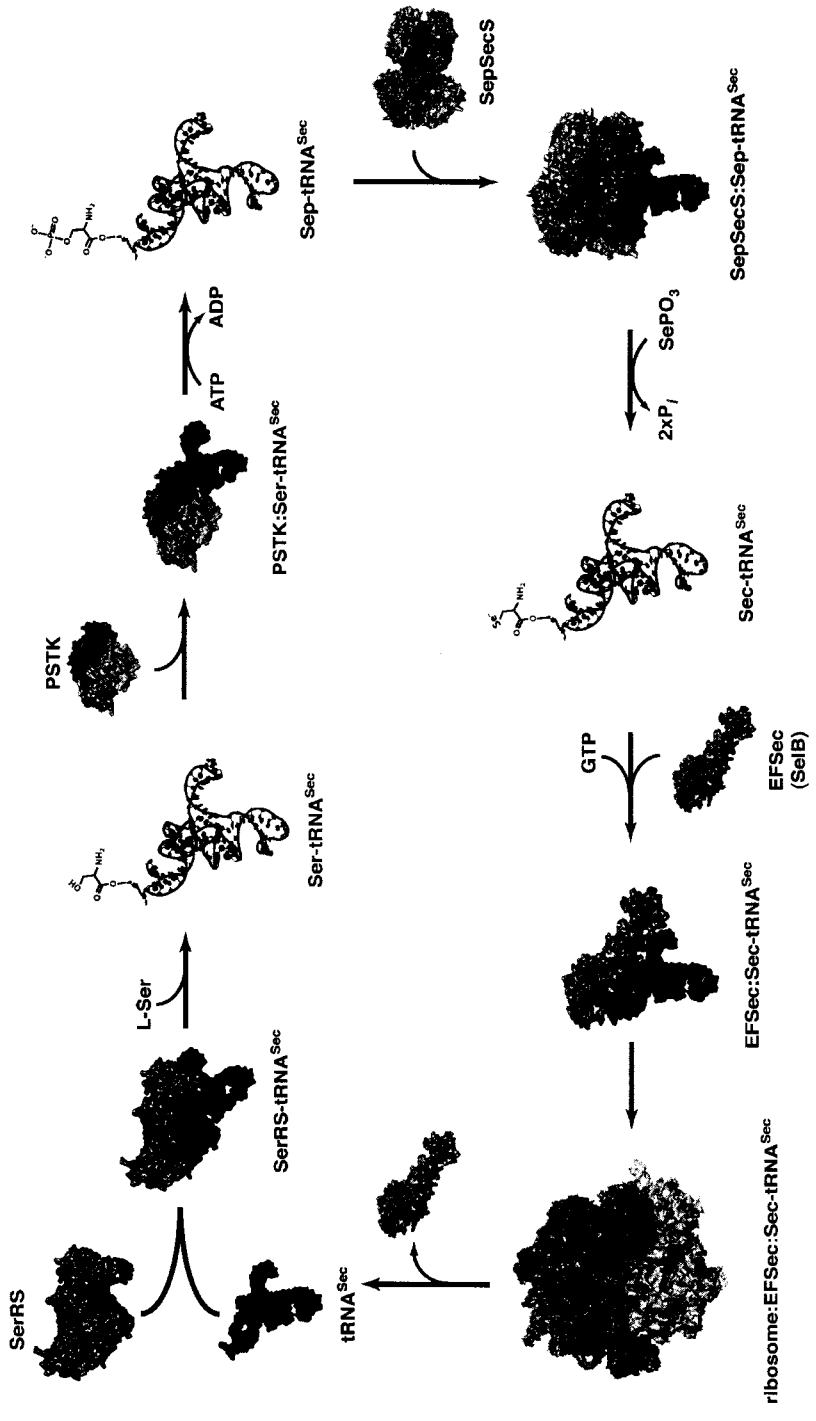


Figure 30: Selenocysteine biosynthesis in eukaryotes (Taken from Palioura et al., 2010). Following serylation of tRNA^{Sec} (red) by SerRS (light and dark blue), PSTK (light and dark grey) phosphorylates Ser-tRNA^{Sec} into Sep-tRNA^{Sec}. SepSecS (gold and olive) then binds Sep-tRNA^{Sec} and catalyzes the conversion of Sep into Sec using selenophosphate. Sec-tRNA^{Sec} is delivered to the 80S ribosome (orange and beige) by the specialized elongation factor EFSec (green). Once the Sec residue is inserted into the nascent polypeptide chain, free tRNA^{Sec} is released. Crystal structures of the bacterial SerRS (Biou et al., 1994), the archaeal PSTK (Araiso et al., 2009), the human SepSecS-tRNA^{Sec} complex (Palioura et al., 2009), the human tRNA^{Sec} (Itoh et al., 2009), the archaeal SeIB (Leibundgut et al., 2005) and that of the bacterial 70S ribosome in complex with EF-Tu (Schmeing et al., 2009) were used for modeling.

B. The indirect RNA-dependent PSTK/SepSecS pathway is the only route to Sec

Despite the *in vitro* experiments that indicated that PSTK and SepSecS are the core components for selenocysteine biosynthesis in eukaryotes (Xu et al., 2007; Yuan et al., 2006), the *in vivo* analysis of Sec-tRNA^{Sec} formation in eukaryotes has been hampered by the lack of a suitable system. Using *T. brucei* to produce knock-out cell lines for the trypanosomal orthologues of PSTK and SepSecS and RNAi strains of Tb-SerRS, Tb-PSTK and Tb-EFSec, I carried out analysis of the eukaryotic selenoprotein synthesis machinery *in vivo*. Formation of Sec-tRNA^{Sec} in *T. brucei* requires the sequential action of Tb-PSTK and Tb-SepSecS along with Tb-SerRS, Tb-SPS2 and Tb-EFSec.

These results also reveal the existence of a single pathway to Sec-tRNA^{Sec} that requires Sep-tRNA^{Sec} as an intermediate. While RNAi knock-down of SepSecS in mammalian cells had only a minor effect on selenoprotein synthesis (Xu et al., 2005), in *T. brucei* there is no alternative route to Sec-tRNA^{Sec} independent of SepSecS. However, the results obtained in *T. brucei* might not be readily comparable with the study in mammalian cells. In *T. brucei*, selenoprotein synthesis was assayed in a double knock-out cell line completely devoid of Tb-SepSecS, whereas in mammalian cells the RNAi-mediated knock-down may have not completely depleted the human SepSecS protein; western blot analysis of SepSecS protein levels in the mammalian knock-down strains would confirm this hypothesis. Nevertheless, it is likely that even in mammalian cells, just as in *T. brucei*, only a single pathway for Sec-tRNA^{Sec} formation exists.

C. *M. maripaludis* SepSecS structure: insights into the reaction mechanism

SepSecS is a PLP enzyme catalyzing a β -replacement (Figure 6), leading to the exchange of a phosphate group for a selenol moiety. The active site of MMPSepSecS lacks a Cys residue. In contrast, the crystal structures of ECCsdB (Lima, 2002) and of AFSepCysS (Fukunaga and Yokoyama, 2007), which are close homologs of SepSecS and catalyze similar reactions, reveal Cys364 and Cys247 in the active sites of ECCsdB and AFSepCysS, respectively (Figure 18) (Araiso et al., 2008). The Cys364 residue of ECCsdB recognizes Sec and withdraws its selenium to form perselenide as an intermediate (Lima, 2002). Cys364 of ECCsdB resides in the $\beta\alpha\beta$ structural motif which does not exist in MMPSepSecS. On the other hand, Cys247 of AFSepCysS is the candidate residue that forms a persulfide, which is the sulfur source for the enzyme-catalyzed Cys-tRNA^{Cys} formation (Fukunaga and Yokoyama, 2007). The absence of such a Cys active site residue that could be involved in the formation of a catalytically important perselenide intermediate indicates that the reaction mechanisms and chemistries of SepSecS are fundamentally different from those of CsdB and SepCysS. Thus, SepSecS and SepCysS, two related PLP enzymes that perform chemically analogous tRNA-dependent transformations of Sep to Sec or Cys, respectively, proceed with different selenium and sulfur transfer mechanisms.

D. The human SepSecS-tRNA^{Sec} complex reveals the mechanism of Sec formation

The structure of the quaternary complex of SepSecS-tRNA^{Sec} with Sep and thiophosphate along with the *in vivo* and *in vitro* activity assays suggest the following

PLP-based mechanism of Sep-tRNA^{Sec} to Sec-tRNA^{Sec} conversion. The reaction begins by the covalently attached Sep being brought into the proximity of the Schiff base when Sep-tRNA^{Sec} binds to SepSecS. The amino group of Sep can then attack the Schiff base formed between Lys284 and PLP, which yields an external aldimine (Figure 29A, B) (Palioura et al., 2009). The reoriented side chain of Lys284 abstracts the C α proton from Sep (Figure 29C), and the electron delocalization by the pyridine ring assists in rapid β -elimination of the phosphate group, which produces an intermediate dehydroalanyl-tRNA^{Sec} (Figure 29C, D). After phosphate dissociation and binding of selenophosphate, the concomitant attack of water on the selenophosphate group and of the nucleophilic selenium onto the highly reactive dehydroalanyl moiety yield an oxidized form of Sec-tRNA^{Sec} (Figure 29D). The protonated Lys284 returns the proton to the C α carbon and then attacks PLP to form an internal aldimine (Figure 29E). Finally, Sec-tRNA^{Sec} is released from the active site (Figure 29F) (Palioura et al., 2009).

This mechanism is clearly distinct from the persulfide-intermediate mechanism in the Sep-tRNA^{Cys} to Cys-tRNA^{Cys} reaction (Hauenstein and Perona, 2008) and explains why SepSecS does not group together with its closest homologue, SepCysS, in the family tree of fold-type I PLP enzymes (Araiso et al., 2008). Moreover, the proposed mechanism for SepSecS is similar to the one used by the bacterial Sela that also proceeds through a dehydroalanyl-tRNA^{Sec} intermediate (Forchhammer and Böck, 1991). SepSecS, therefore, uses a primordial tRNA-dependent catalytic mechanism in which the PLP cofactor is directly involved, while using a tetrameric Fold Type I architecture as the scaffold for binding the distinct structure of tRNA^{Sec}.

Comprised of 90 nucleotides, tRNA^{Sec} is the largest eukaryotic tRNA. Its acceptor-TΨC 'helix' contains an additional base pair resulting in a 9/4 fold, in contrast to the 7/5 fold adopted by all known canonical tRNAs (Figure 23) (Palioura et al., 2009; Ithoh et al., 2009). Except for tRNA^{Sec}, all tRNAs are transferred to the ribosome bound to either EF-Tu in bacteria or eEF1A in eukaryotes. The atypical 9/4 fold of tRNA^{Sec} accounts for the evolution of the specialized elongation factor SelB in bacteria (Forchhammer et al., 1989) and EFSec in eukaryotes (Fagegaltier et al., 2000), which binds only Sec-tRNA^{Sec} and not other aminoacyl-tRNAs. Both the variable and D arms of tRNA^{Sec} are longer than the corresponding elements in canonical tRNAs, while the eighth position in tRNA^{Sec} is occupied by adenine instead of the highly conserved uridine that is found in all canonical tRNAs. In a striking contrast to U8, the base of A8 does not form any tertiary interactions with the D- and TΨC arms leaving a hole in the core of the tRNA molecule (Palioura et al., 2009; Itoh et al., 2009; Yuan et al., 2010). All these features result in a distinct three-dimensional structure of tRNA^{Sec}, which is likely to be recognized by all Sec-specific enzymes.

In fact, while SerRS recognizes common structural features of tRNA^{Ser} and tRNA^{Sec} (Figure 25), it is thought that PSTK, SepSecS and EFSec recognize the distinct structural elements of tRNA^{Sec} instead. The structure of the human SepSecS in complex with tRNA^{Sec} and the *in vivo* mutagenesis data reveal that SepSecS does exactly that. It binds the longer acceptor-TΨC 'helix', the long variable arm, the 5' phosphate and the acceptor-TΨC-variable elbow (Figure 19). The tRNA^{Sec} molecule is anchored to the SepSecS tetramer on one end by interactions between the tip of its acceptor arm (G73)

and the catalytic dimer (Arg398), and on the other end by interactions between its variable arm (C46L) and the non-catalytic dimer (Arg271) (Figure 20). The rest of the interactions between the $\alpha 1$ helix of the non-catalytic dimer of SepSecS (Arg26, Lys38 and Lys40) and the acceptor-T Ψ C arm (C2, G50, and C64) further stabilize the complex (Palioura et al., 2009). Practically, SepSecS measures the length of the acceptor-T Ψ C 'helix' as the distance between the variable arm and the acceptor tip of tRNA^{Sec}. Modeling of canonical tRNAs onto the human SepSecS showed that the length of their acceptor-T Ψ C 'helix' is too short to reach the active site of the enzyme (Figure 24A, B). Even if a productive interaction between the tip of the acceptor arm and the enzyme were forced to form, multiple steric clashes between the T Ψ C arm and the N-terminal $\alpha 1$ helix would prevent binding of the canonical tRNA to SepSecS (Figure 24C).

The mode of tRNA^{Sec} recognition by human SepSecS is reminiscent of the tRNA-dependent amidotransferases GatDE and GatCAB that also recognize only the acceptor stem of the cognate tRNA (tRNA^{Gln}) (Nakamura et al., 2006; Oshikane et al., 2006). Given that SepSecS only binds to one side of the acceptor-T Ψ C arm of tRNA^{Sec}, it is plausible that PSTK would bind to a different side of the acceptor-T Ψ C arm or to the D-stem as it is proposed for the archaeal and human enzymes, respectively (Araiso et al., 2009; Sherrer et al., 2008a). This would then allow PSTK, SepSecS and tRNA^{Sec} to form a ternary complex *in vivo* similar to that of the transamidosome involved in prokaryotic tRNA-dependent asparagine biosynthesis (Bailly et al., 2007).

Finally, the observations that the archaeal PSTK and SepSecS are able to act on *E. coli* tRNA^{Sec} *in vivo*, that human SepSecS can use *E. coli* tRNA^{Sec} *in vivo* as well as the

archaeal tRNA^{Sec} *in vitro* (Yuan et al., 2006), and that the bacterial selenocysteine synthase can use both archaeal and murine tRNA^{Sec} *in vitro* (Xu et al., 2007) suggest that the length of the acceptor-TΨC arm of tRNA^{Sec}, the position of the variable arm and the mode of tRNA^{Sec} recognition are likely to be conserved in all domains of life.

Taken together, the structure of the SepSecS-tRNA^{Sec} complex in the presence of Sep and thiophosphate provided key insights into the mechanism of Sec formation in humans. (1) SepSecS can interact with tRNA^{Sec} only as a tetramer. (2) The crystal structure of human tRNA^{Sec} revealed a novel fold distinct from other canonical tRNA molecules. (3) The discriminator base G73, a 13 bp long acceptor-TΨC arm and a long variable arm, are the major recognition motifs for binding to SepSecS and may be critical for binding to other Sec-specific proteins such as SerRS and the Sec-specific elongation factors SelB or EFsec. (4) The epitope of human SepSecS autoantibodies is part of the interface of the enzyme's active site and the acceptor arm of tRNA^{Sec}. (5) SepSecS can act only on Sep linked to tRNA^{Sec} and not on free Sep or on Ser-tRNA^{Sec}. (6) Both structural and biochemical data support a PLP-dependent mechanism of Sep-tRNA^{Sec} to Sec-tRNA^{Sec} conversion. The mechanism of selenocysteine formation and the general mode of tRNA^{Sec} recognition presented here is likely to be conserved in all other organisms that encode selenocysteine.

REFERENCES

- Aeby, E., Palioura, S., Pusnik, M., Marazzi, J., Lieberman, A., Ullu, E., Söll, D., and Schneider, A.** (2009). The canonical pathway for selenocysteine insertion is dispensable in *Trypanosomes*. *Proc Natl Acad Sci U S A.* 106, 5088-5092.
- Allmang, C., and Krol, A.** (2006). Selenoprotein synthesis: UGA does not end the story. *Biochimie* 88, 1561-1571.
- Allmang, C., Wurth, L., and Krol, A.** (2009). The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated. *Biochim Biophys Acta* 1790, 1415-1423.
- Ambrogelly, A., Palioura, S., and Söll, D.** (2007) Natural expansion of the genetic code. *Nat. Chem. Biol.* 3, 29-35.
- Araiso, Y., Palioura, S., Ishitani, R., Sherrer, R.L., O'Donoghue, P., Yuan, J., Oshikane, H., Domae, N., Defranco, J., Söll, D., and Nureki, O.** (2008). Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic Acids Res* 36, 1187-1199.
- Araiso, Y., Sherrer, R.L., Ishitani, R., Ho, J.M., Söll, D., and Nureki, O.** (2009). Structure of a tRNA-dependent kinase essential for selenocysteine decoding. *Proc Natl Acad Sci USA* 106, 16215-16220.
- Axley, M.J., Böck, A., Stadtman, T.C.** (1991). Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium. *Proc Natl Acad Sci USA* 88, 8450-8454.
- Bailly, M., Blaise, M., Lorber, B., Becker, H.D., and Kern, D.** (2007). The transamidosome: a dynamic ribonucleoprotein particle dedicated to prokaryotic tRNA-dependent asparagine biosynthesis. *Mol Cell* 28, 228-239.
- Baron, C., and Böck, A.** (1991). The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA^{Sec} of *Escherichia coli* is the determinant for binding to elongation factors SELB or Tu. *J Biol Chem* 266, 20375-20379.
- Baron, C., Heider, J., and Böck, A.** (1990). Mutagenesis of selC, the gene for the selenocysteine-inserting tRNA-species in *E. coli*: effects on *in vivo* function. *Nucleic Acids Res* 18, 6761-6766.

- Bellinger, F.P., Raman, A.V., Reeves, M.A., and Berry, M.J.** (2009). Regulation and function of selenoproteins in human disease. *Biochem J* 422, 11-22.
- Beverley, S.M., and Clayton, C.E.** (1993). Transfection of *Leishmania* and *Trypanosoma brucei* by electroporation. *Methods Mol Biol* 21, 333-348.
- Bilokapic, S., Korencic, D., Söll, D., and Weygand-Durasevic, I.** (2004). The unusual methanogenic seryl-tRNA synthetase recognizes tRNA^{Ser} species from all three kingdoms of life. *Eur J Biochem* 271, 694-702.
- Biou, V., Yaremchuk, A., Tukalo, M., and Cusack, S.** (1994). The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA(Ser). *Science* 263, 1404-1410.
- Blight, S.K., Larue, R.C., Mahapatra, A., Longstaff, D.G., Chang, E., Zhao, G., Kang, P.T., Green-Church, K.B., Chan, M.K., and Krzycki, J.A.** (2004). Direct charging of tRNA_{CUA} with pyrrolysine *in vitro* and *in vivo*. *Nature* 431, 333-335.
- Böck, A., Thanbichler, M., Rother, M., and Resch, A.** (2005). Selenocysteine. In *Aminoacyl-tRNA Synthetases*, M. Ibba, C.S. Francklyn, and S. Cusack, eds. (Georgetown, TX, Landes Bioscience), pp. 320-327.
- Bösl, M.R., Takaku, K., Oshima, M., Nishimura, S., and Taketo, M.M.** (1997). Early embryonic lethality caused by targeted disruption of the mouse selenocysteine tRNA gene (*Trsp*). *Proc Natl Acad Sci USA* 94, 5531-5534.
- Bouzaidi-Tiali, N., Aeby, E., Charriere, F., Pusnik, M., and Schneider, A.** (2007). Elongation factor 1a mediates the specificity of mitochondrial tRNA import in *T. brucei*. *EMBO J* 26, 4302-4312.
- Brun, R., and Schönenberger** (1979). Cultivation and *in vitro* cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. *Acta tropica* 36, 289-292.
- Bullock, T.L., Uter, N., Nissan, T.A., and Perona, J.J.** (2003). Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J Mol Biol* 328, 395-408.
- Carlson, B.A., Xu, X.M., Kryukov, G.V., Rao, M., Berry, M.J., Gladyshev, V.N., and Hatfield, D.L.** (2004). Identification and characterization of phosphoseryl-tRNA^{[Ser]Sec} kinase. *Proc Natl Acad Sci USA* 101, 12848-12853.

- Cassago, A., Rodrigues, E.M., Prieto, E.L., Gaston, K.W., Alfonzo, J.D., Iribar, M.P., Berry, M.J., Cruz, A.K., and Thiemann, O.H.** (2006). Identification of *Leishmania* selenoproteins and SECIS element. *Mol Biochem Parasitol* 149, 128-134.
- Cone, J.E., Del Rio, 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 USA* 73, 2659-2663.
- Crick, F.H.C.** (1958). On protein synthesis. *Symp Soc Exp Biol* 12, 138-163.
- Crick, F.H.C.** (1966). Codon-anticodon pairing: the wobble hypothesis. *J Mol Biol* 19, 548-555.
- Cupp-Vickery, J.R., Urbina, H., and Vickery, L.E.** (2003). Crystal structure of IscS, a cysteine desulfurase from *Escherichia coli*. *J Mol Biol* 330, 1049-1059.
- Datsenko, K.A., and Wanner, B.L.** (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97, 6640-6645.
- Diamond, A.M., Choi, I.S., Crain, P.F., Hashizume, T., Pomerantz, S.C., Cruz, R., Steer, C.J., Hill, K.E., Burk, R.F., McCloskey, J.A., and Hatfield, D.L.** (1993). Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA([Ser]Sec). *J Biol Chem* 268, 14215-14223.
- Eiler, S., Dock-Bregeon, A., Moulinier, L., Thierry, J.C., and Moras, D.** (1999). Synthesis of aspartyl-tRNA(Asp) in *Escherichia coli* –a snapshot of the second step. *EMBO J* 18, 6532-6541.
- Eliot, A.C., and Kirsch, J.F.** (2004). Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem* 73, 383-415.
- Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P., and Krol, A.** (2000). Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J* 19, 4796-4805.
- Forchhammer, K., and Böck, A.** (1991). Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. *J Biol Chem* 266, 6324-6328.
- Forchhammer, K., Boesmiller, K., and Böck, A.** (1991). The function of selenocysteine synthase and SELB in the synthesis and incorporation of selenocysteine. *Biochimie* 73, 1481-1486.

- Forchhammer, K., Leinfelder, W., and Böck, A.** (1989). Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. *Nature* 342, 453-456.
- Förster, C., Ott, G., Forchhammer, K., and Sprinzl, M.** (1990). Interaction of a selenocysteine-incorporating tRNA with elongation factor Tu from *E.coli*. *Nucleic Acids Res* 18, 487-491.
- Forstrom, J.W., Zakowski, J.J., and Tappel, A.L.** (1978) Identification of the catalytic site of rat liver glutathione peroxidase as selenocysteine. *Biochemistry* 17, 2639-2644.
- Fu, L.H., Wang, X.F., Eyal, Y., She, Y.M., Donald, L.J., Standing, K.G., and Ben-Hayyim, G.** (2002). A selenoprotein in the plant kingdom. Mass spectrometry confirms that an opal codon (UGA) encodes selenocysteine in *Chlamydomonas reinhardtii* glutathione peroxidase. *J Biol Chem* 277, 25983-25991.
- Fukunaga, R., and Yokoyama, S.** (2007). Structural insights into the second step of RNA-dependent cysteine biosynthesis in archaea: crystal structure of Sep-tRNA: Cys-tRNA synthase from *Archaeoglobus fulgidus*. *J Mol Biol* 370, 128-141.
- Ganichkin, O.M., Xu, X.M., Carlson, B.A., Mix, H., Hatfield, D.L., Gladyshev, V.N., and Wahl, M.C.** (2008). Structure and catalytic mechanism of eukaryotic selenocysteine synthase. *J Biol Chem* 283, 5849-5865.
- Gelpi, C., Sontheimer, E.J., and Rodriguez-Sanchez, J.L.** (1992). Autoantibodies against a serine tRNA-protein complex implicated in cotranslational selenocysteine insertion. *Proc Natl Acad Sci USA* 89, 9739-9743.
- Geslain, R., Aeby, E., Guitart, T., Jones, T.E., Castro de Moura, M., Charriere, F., Schneider, A., and Ribas de Pouplana, L.** (2006). *Trypanosoma* seryl-tRNA synthetase is a metazoan-like enzyme with high affinity for tRNA^{Sec}. *J Biol Chem* 281, 38217-38225.
- Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A., and Chan, M.K.** (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 296, 1462-1466.
- Hatfield, D.L., Carlson, B.A., Xu, X.M., Mix, H., and Gladyshev, V.N.** (2006). Selenocysteine incorporation machinery and the role of selenoproteins in development and health. *Prog Nucleic Acid Res Mol Biol* 81, 97-142.

- Hauenstein, S.I., and Perona, J.J.** (2008). Redundant synthesis of cysteinyl-tRNA^{Cys} in *Methanosarcina mazei*. *J Biol Chem* 283, 22007-22017.
- Herkel, J., Heidrich, B., Nieraad, N., Wies, I., Rother, M., and Lohse, A.W.** (2002). Fine specificity of autoantibodies to soluble liver antigen and liver/pancreas. *Hepatology* 35, 403-408.
- Ibba, M., and Söll, D.** (2000). Aminoacyl-tRNA synthesis. *Annu Rev Biochem* 69, 617-650.
- Ioudovitch, A., and Steinberg, S.V.** (1998). Modeling the tertiary interactions in the eukaryotic selenocysteine tRNA. *RNA* 4, 365-373.
- Itoh, Y., Chiba, S., Sekine, S., and Yokoyama, S.** (2009). Crystal structure of human selenocysteine tRNA. *Nucleic Acids Res* 37, 6259-6268.
- Jansonius, J.N.** (1998). Structure, evolution and action of vitamin B6-dependent enzymes. *Curr Opin Struct Biol* 8, 759-769.
- Johansson, L., Gafvelin, G., and Arner, E.S.** (2005). Selenocysteine in proteins-properties and biotechnological use. *Biochim Biophys Acta* 1726, 1-13.
- Kaiser, J.T., Gromadski, K., Rother, M., Engelhardt, H., Rodnina, M.V., and Wahl, M.C.** (2005). Structural and functional investigation of a putative archaeal selenocysteine synthase. *Biochemistry* 44, 13315-13327.
- Kernebeck, T., Lohse, A.W., and Grötzinger, J.** (2001). A bioinformatical approach suggests the function of the autoimmune hepatitis target antigen soluble liver antigen/liver pancreas. *Hepatology* 34, 230-233.
- Kim, H.S., Vothknecht, U.C., Hedderich, R., Celic, I., and Söll, D.** (1998). Sequence divergence of seryl-tRNA synthetases in archaea. *J Bacteriol* 180, 6446-6449.
- Kryukov, G.V., Castellano, S., Novoselov, S.V., Lobanov, A.V., Zehtab, O., Guigo, R., and Gladyshev, V.N.** (2003). Characterization of mammalian selenoproteomes. *Science* 300, 1439-1443.
- Kryukov, G.V., and Gladyshev, V.N.** (2004). The prokaryotic selenoproteome. *EMBO Rep* 5, 538-543.
- Kuiper, G.G., Klootwijk, W., and Visser, T.J.** (2003). Substitution of cysteine for selenocysteine in the catalytic center of type III iodothyronine deiodinase reduces catalytic efficiency and alters substrate preference. *Endocrinology* 144, 2505-2513.

- Lacourciere, G.M., Levine, R.L., and Stadtman, T.C.** (2002). Direct detection of potential selenium delivery proteins by using an *Escherichia coli* strain unable to incorporate selenium from selenite into proteins. *Proc Natl Acad Sci USA* 99, 9150-9153.
- Leinfelder, W., Stadtman, T.C., and Böck, A.** (1989). Occurrence *in vivo* of selenocysteyl-tRNA^{Ser}_{UCA} in *Escherichia coli*. Effect of *selC* mutations. *J Biol Chem* 264, 9720-9723.
- Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M.A., and Böck, A.** (1988). Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* 331, 723-725.
- Lima, C.D.** (2002). Analysis of the *E. coli* NifS CsdB protein at 2.0 Å reveals the structural basis for perselenide and persulfide intermediate formation. *J Mol Biol* 315, 1199-1208.
- Lobanov, A.V., Gromer, S., Salinas, G., and Gladyshev, V.N.** (2006). Selenium metabolism in *Trypanosoma*: characterization of selenoproteomes and identification of a Kinetoplastida-specific selenoprotein. *Nucleic Acids Res* 34, 4012-4024.
- Lobanov, A.V., Hatfield, D.L., and Gladyshev, V.N.** (2009). Eukaryotic selenoproteins and selenoproteomes. *Biochim Biophys Acta* 1790, 1424-1428.
- Longstaff, D.G., Larue, R.C., Faust, J.E., Mahapatra, A., Zhang, L., Green-Church, K.B., and Krzycki, J.A.** (2007). A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc Natl Acad Sci USA* 104, 1021-1026.
- Lu, J., and Holmgren, A.** (2009). Selenoproteins. *J Biol Chem* 284, 723-727.
- Mäenpää, P.H., and Bernfield, M.R.** (1970). A specific hepatic transfer RNA for phosphoserine. *Proc Natl Acad Sci USA* 67, 688-695.
- Manns, M.P., Johnson, E.F., Griffin, K.J., Tan, E.M., and Sullivan, K.F.** (1989). Major antigen of liver kidney microsomal autoantibodies in idiopathic autoimmune hepatitis is cytochrome P450db1. *J Clin Invest* 83, 1066-1072.
- Matthaei, J.H., Voigt, H.P., Heller, G., Neth, R., Schöch, G., Kübler, H., Amelunxen, F., Sander, G., and Parmeggiani, A.** (1966). Specific interactions of ribosomes in decoding. *Cold Spring Harb Symp Quant Biol* 31, 25-38.

- Mihara, H., Maeda, M., Fujii, T., Kurihara, T., Hata, Y., and Esaki, N.** (1999). A nifS-like gene, *csdB*, encodes an *Escherichia coli* counterpart of mammalian selenocysteine lyase. Gene cloning, purification, characterization and preliminary x-ray crystallographic studies. *J Biol Chem* *274*, 14768-14772.
- Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C.** (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* *15*, 8783-8798.
- Nakamura, A., Yao, M., Chimnarong, S., Sakai, N., and Tanaka, I.** (2006). Ammonia channel couples glutaminase with transamidase reactions in GatCAB. *Science* *312*, 1954-1958.
- Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., and O'Neal, C.** (1965). RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc Natl Acad Sci USA* *53*, 1161-1168.
- Nureki, O., Watanabe, K., Fukai, S., Ishii, R., Endo, Y., Hori, H., and Yokoyama, S.** (2004). Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme. *Structure* *12*, 593-602.
- Ohama, T., Yang, D.C., and Hatfield, D.L.** (1994). Selenocysteine tRNA and serine tRNA are aminoacylated by the same synthetase, but may manifest different identities with respect to the long extra arm. *Arch Biochem Biophys* *315*, 293-301.
- Oshikane, H., Sheppard, K., Fukai, S., Nakamura, Y., Ishitani, R., Numata, T., Sherrer, R.L., Feng, L., Schmitt, E., Panvert, M., Söll, D., and Nureki, O.** (2006). Structural basis of RNA-dependent recruitment of glutamine to the genetic code. *Science* *312*, 1950-1954.
- Palioura, S., Sherrer, R.L., Steitz, T.A., Söll, D., and Simonović, M.** (2009). The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. *Science* *325*, 321-325.
- Palioura, S., Herkel, J., Simonović, M., Lohse, A., and Söll, D.** (2010). Human SepSecS or SLA/LP: selenocysteine formation and autoimmune hepatitis. *Biol Chem (in press)*.
- Percudani, R., and Peracchi, A.** (2003). A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep* *4*, 850-854.

- Polycarpo, C., Ambrogelly, A., Berube, A., Winbush, S.M., McCloskey, J.A., Crain, P.F., Wood, J.L., and Söll, D.** (2004). An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc Natl Acad Sci USA* *101*, 12450-12454.
- Rayman, M.P.** (2000). The importance of selenium to human health. *Lancet* *356*, 233-241.
- Rother, M., Wilting, R., Commans, S., and Böck, A.** (2000). Identification and characterisation of the selenocysteine-specific translation factor SelB from the archaeon *Methanococcus jannaschii*. *J Mol Biol* *299*, 351-358.
- Rudinger, J., Hillenbrandt, R., Sprinzl, M., and Giegé, R.** (1996). Antideterminants present in minihelix(Sec) hinder its recognition by prokaryotic elongation factor Tu. *EMBO J* *15*, 650-657.
- Sauerwald, A., Zhu, W., Major, T.A., Roy, H., Palioura, S., Jahn, D., Whitman, W.B., Yates, J.R., 3rd, Ibba, M., and Söll, D.** (2005). RNA-dependent cysteine biosynthesis in archaea. *Science* *307*, 1969-1972.
- Schneider, G., Kack, H., and Lindqvist, Y.** (2000). The manifold of vitamin B6 dependent enzymes. *Structure* *8*, R1-6.
- Sculaccio, S.A., Rodrigues, E.M., Cordeiro, A.T., Magalhaes, A., Braga, A.L., Alberto, E.E., and Thiemann, O.H.** (2008). Selenocysteine incorporation in Kinetoplastid: selenophosphate synthetase (SELD) from *Leishmania major* and *Trypanosoma brucei*. *Mol Biochem Parasitol* *162*, 165-171.
- Sherrer, R.L., Ho, J.M., and Söll, D.** (2008a). Divergence of selenocysteine tRNA recognition by archaeal and eukaryotic *O*-phosphoseryl-tRNA^{Sec} kinase. *Nucleic Acids Res* *36*, 1871-1880.
- Sherrer, R.L., O'Donoghue, P., and Söll, D.** (2008b). Characterization and evolutionary history of an archaeal kinase involved in selenocysteinyl-tRNA formation. *Nucleic Acids Res* *36*, 1247-1259.
- Söll, D.** (1988). Genetic code: enter a new amino acid. *Nature* *331*, 662-663.
- Söll, D., Jones, D.S., Ohtsuka, E., Faulkner, R.D., Lohrmann, R., Hayatsu, H., and Khorana, H.G.** (1966). Specificity of sRNA for recognition of codons as studied by the ribosomal binding technique. *J Mol Biol* *19*, 556-573.

- Söll, D., Ohtsuka, E., Jones, D.S., Lohrmann, R., Hayatsu, H., Nishimura, S., and Khorana, H.G.** (1965). Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *Proc Natl Acad Sci USA* *54*, 1378-1385.
- Srinivasan, G., James, C.M., and Krzycki, J.A.** (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* *296*, 1459-1462.
- Stock, T., and Rother, M.** (2009). Selenoproteins in Archaea and Gram-positive bacteria. *Biochim Biophys Acta* *1790*, 1520-1532.
- Stürchler-Pierrat, C., Hubert, N., Totsuka, T., Mizutani, T., Carbon, P., and Krol, A.** (1995). Selenocystenylation in eukaryotes necessitates the uniquely long aminoacyl acceptor stem of selenocysteine tRNA^{Sec}. *J Biol Chem* *270*, 18570-18574.
- Stürchler, C., Westhof, E., Carbon, P., and Krol, A.** (1993). Unique secondary and tertiary structural features of the eukaryotic selenocysteine tRNA(Sec). *Nucleic Acids Res* *21*, 1073-1079.
- Su, D., Hohn, M.J., Palioura, S., Sherrer, R.L., Yuan, J., Söll, D., and O'Donoghue, P.** (2009). How an obscure archaeal gene inspired the discovery of selenocysteine biosynthesis in humans. *IUBMB Life* *61*, 35-39.
- Tormay, P., Wilting, R., Lottspeich, F., Mehta, P.K., Christen, P., and Böck, A.** (1998). Bacterial selenocysteine synthase—structural and functional properties. *Eur J Biochem* *254*, 655-661.
- Ullu, E., Tschudi, C., and Chakraborty, T.** (2004). RNA interference in protozoan parasites. *Cell Microbiol* *6*, 509-519.
- Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glassl, S., Ansorge, W., Bocher, M., Blocker, H., Bauersachs, S., Blum, H., et al.** (2001). Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Res* *11*, 422-435.
- Wu, X.Q., and 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*, 5589-5594.

Wu, X.Q., and Gross, H.J. (1994). The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosphorylation. *EMBO J* 13, 241-248.

Xu, X.M., Carlson, B.A., Mix, H., Zhang, Y., Saira, K., Glass, R.S., Berry, M.J., Gladyshev, V.N., and Hatfield, D.L. (2007). Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol* 5, e4.

Xu, X.M., Mix, H., Carlson, B.A., Grabowski, P.J., Gladyshev, V.N., Berry, M.J., and Hatfield, D.L. (2005). Evidence for direct roles of two additional factors, SECp43 and soluble liver antigen, in the selenoprotein synthesis machinery. *J Biol Chem* 280, 41568-41575.

Yano, T., Kuramitsu, S., Tanase, S., Morino, Y., Hiromi, K., and Kagamiyama, H. (1991). The role of His143 in the catalytic mechanism of *Escherichia coli* aspartate aminotransferase. *J Biol Chem* 266, 6079-6085.

Yoshizawa, S., and Böck, A. (2009). The many levels of control on bacterial selenoprotein synthesis. *Biochim Biophys Acta* 1790, 1404-1414.

Yuan, J., Palioura, S., Salazar, J.C., Su, D., O'Donoghue, P., Hohn, M.J., Cardoso, A.M., Whitman, W.B., and Söll, D. (2006). RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc Natl Acad Sci USA* 103, 18923-18927.

Yuan, J., O'Donoghue, P., Ambrogelly, A., Gundllapalli, S., Lynn Sherrer, R., Palioura, S., Simonović, M., and Söll, D. (2010). Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett* 584, 342-349.

Zhong, L., and Holmgren, A. (2000). Essential role of selenium in the catalytic activities of mammalian thioredoxin reductase revealed by characterization of recombinant enzymes with selenocysteine mutations. *J Biol Chem* 275, 18121-18128.

Zinoni, F., Birkmann, A., Stadtman, T.C., and Böck, A. (1986). Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proc Natl Acad Sci USA* 83, 4650-4654.

APPENDIX

- Yuan, J.*, Palioura, S.*, Salazar, J.C., Su, D., O'Donoghue, P., Hohn, M.J., Cardoso, A.M., Whitman, W.B., and Söll, D.** (2006). RNA-dependent conversion of phosphoserine forms seleno-cysteine in eukaryotes and archaea. *Proc Natl Acad Sci USA* 103, 18923-18927.
- Ambrogelly, A., Palioura, S., and Söll, D.** (2007). Natural expansion of the genetic code. *Nat Chem Biol* 3, 29-35.
- Araiso, Y.*, Palioura, S.*, Ishitani, R., Sherrer, R. L., Yuan, J., Oshikane, H., Domae, N., DeFranco, J., Söll, D., and Nureki, O.** (2008). Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic Acids Res* 36, 1187-99.
- Su, D., Hohn, M.J., Palioura, S., Sherrer, R.L., Yuan, J., Söll, D., and O'Donoghue, P.** (2008). How an obscure archaeal gene inspired the discovery of selenocysteine biosynthesis in humans. *IUBMB Life* 61, 35-9.
- Aeby, E., Palioura, S., Pusnik, M., Marazzi, J., Lieberman, A., Ullu, E., Söll, D., and Schneider, A.** (2009). The canonical pathway for selenocysteine insertion is dispensable in *Trypanosomes*. *Proc Natl Acad Sci U S A*. 106, 5088-92.
- Palioura, S., Sherrer, R.L., Steitz, T.A., Söll, D., and Simonović, M.** (2009). The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. *Science* 325, 321-325.
- Yuan, J., O'Donoghue, P., Ambrogelly, A., Gundllapalli, S., Sherrer, R.L., Palioura, S., Simonović, M., and Söll, D.** (2010). Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett* 584, 342-349.

RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea

Jing Yuan*, Sotiria Palioura*, Juan Carlos Salazar*, Dan Su*, Patrick O'Donoghue*, Michael J. Hohn*, Alexander Machado Cardoso*, William B. Whitman[†], and Dieter Söll*^{‡§}

Departments of *Molecular Biophysics and Biochemistry and [‡]Chemistry, Yale University, New Haven, CT 06520-8114; and [†]Department of Microbiology, University of Georgia, Athens, GA 30602-2605

Contributed by Dieter Söll, November 1, 2006 (sent for review October 30, 2006)

The trace element selenium is found in proteins as selenocysteine (Sec), the 21st amino acid to participate in ribosome-mediated translation. The substrate for ribosomal protein synthesis is selenocysteinyl-tRNA^{Sec}. Its biosynthesis from seryl-tRNA^{Sec} has been established for bacteria, but the mechanism of conversion from Ser-tRNA^{Sec} remained unresolved for archaea and eukarya. Here, we provide evidence for a different route present in these domains of life that requires the tRNA^{Sec}-dependent conversion of O-phosphoserine (Sep) to Sec. In this two-step pathway, O-phosphoseryl-tRNA^{Sec} kinase (PSTK) converts Ser-tRNA^{Sec} to Sep-tRNA^{Sec}. This misacylated tRNA is the obligatory precursor for a Sep-tRNA:Sec-tRNA synthase (SepSecS); this protein was previously annotated as SLA/LP. The human and archaeal SepSecS genes complement *in vivo* an *Escherichia coli* Sec synthase (SelA) deletion strain. Furthermore, purified recombinant SepSecS converts Sep-tRNA^{Sec} into Sec-tRNA^{Sec} *in vitro* in the presence of sodium selenite and purified recombinant *E. coli* selenophosphate synthetase (SelD). Phylogenetic arguments suggest that Sec decoding was present in the last universal common ancestor. SepSecS and PSTK coevolved with the archaeal and eukaryotic lineages, but the history of PSTK is marked by several horizontal gene transfer events, including transfer to non-Sec-decoding Cyanobacteria and fungi.

aminoacyl-tRNA | evolution | formate dehydrogenase | pyridoxal phosphate

Amino acids enter protein synthesis as aminoacyl-tRNAs (aa-tRNAs). The identity of an amino acid inserted into the nascent polypeptide is determined by two factors: the interaction of the aa-tRNA anticodon with an appropriate mRNA codon, and the correct pairing of amino acid and tRNA anticodon in the aa-tRNA. Thus, faithful protein synthesis requires the presence in the cell of a full set of correctly aminoacylated tRNAs (1). Two routes to aa-tRNA synthesis exist (2): (i) direct acylation of the amino acid onto its cognate tRNA catalyzed by an aa-tRNA synthetase, and (ii) tRNA-dependent modification of a noncognate amino acid attached to tRNA. Although the majority of aa-tRNAs are made by direct aminoacylation, the indirect pathway is also widely used, e.g., for Gln-tRNA^{Gln} and Asn-tRNA^{Asn} formation (3).

The biosynthetic route of aa-tRNA formation is understood for all of the currently known cotranslationally inserted amino acids. The sole exception is the case of selenocysteine (Sec) in eukarya and archaea (e.g., refs. 4 and 5). The bacterial case was solved in the 1990s when genetic and biochemical studies with *Escherichia coli* revealed that bacterial Sec-tRNA^{Sec} was synthesized by an indirect tRNA-dependent amino acid transformation mechanism (6). Sec is cotranslationally inserted into proteins in response to the codon UGA; thus the tRNA^{Sec} species has the corresponding anticodon UCA (7). This tRNA is misacylated with serine by *E. coli* seryl-tRNA synthetase (SerRS) to form Ser-tRNA^{Sec} (8). Then a pyridoxal phosphate (PLP)-dependent Sec synthase (SelA) catalyzes the Ser-tRNA^{Sec} → Sec-tRNA^{Sec} conversion using selenophosphate as selenium donor (Fig. 1

Upper). For archaea and eukarya our understanding is still fragmentary. Serylation of tRNA^{Sec} by eukaryal (9–12) and archaeal (13–15) SerRS enzymes has been well established. However, the demonstration of an analogous reaction to the bacterial pathway converting Ser-tRNA^{Sec} → Sec-tRNA^{Sec} has remained elusive. In particular, the archaeal protein MJ0158, annotated as the ortholog to bacterial SelA, did not support *in vitro* Sec formation (15).

Recently, a pathway for cysteine formation was discovered in methanogens based on a tRNA-dependent conversion of O-phosphoserine (Sep) to cysteine (16). This route involves the attachment of Sep to tRNA^{Cys} by SepRS and the subsequent Sep-tRNA^{Cys} → Cys-tRNA^{Cys} transformation by Sep-tRNA:Cys-tRNA synthase, a PLP-dependent enzyme using a still unknown sulfur donor (Fig. 1 Lower). The implications of this pathway for Sec formation were obvious (16), especially because a phosphoseryl-tRNA^{Sec} kinase (PSTK) had just been identified (17). Its activity was characterized *in vitro* for the murine (17) and the *Methanocaldococcus jannaschii* (15) enzyme; this kinase phosphorylates specifically Ser-tRNA^{Sec} to Sep-tRNA^{Sec}. This finding would lend credence to early reports of the existence of mammalian Sep-tRNA formed from Ser-tRNA (18, 19). Because the initial search for a PLP-dependent SelA activity as manifested by *E. coli* SelA was unsuccessful (15), we considered other proteins. The most promising candidate was the human protein SLA/LP (soluble liver antigen/liver pancreas), which forms a ribonucleoprotein antigenic complex with tRNA^{Sec} in patients suffering from an autoimmune chronic hepatitis (20). This protein was classified computationally as a PLP-dependent serine hydroxymethyltransferase that might have SelA function (21). To lend further credibility, this protein (22, 23) has very well conserved archaeal orthologs but only in *Methanocaldococcus jannaschii* (MJ0610), *Methanococcus maripaludis* (MMP0595), *Methanopyrus kandleri* (MK0672), *Methanococcus voltae*, *Methanothermococcus thermolithotrophicus*, and *Methanococcus vannielii*; these are the known archaea that use Sec. Considering all of these data, it appeared probable that this protein was the missing Sep-tRNA:Sec-tRNA synthase (SepSecS) (Fig. 1 Upper).

Here, we report the identification of the human SLA/LP and *Methanococcus maripaludis* MMP0595 proteins as SepSecS en-

Author contributions: J.Y. and S.P. contributed equally to this work; J.Y., S.P., J.C.S., D. Su, P.O., W.B.W., and D. Söll designed research; J.Y., S.P., J.C.S., D. Su, and P.O. performed research; M.J.H., A.M.C., and W.B.W. contributed new reagents/analytic tools; J.Y., S.P., D. Su, P.O., and D. Söll analyzed data; and J.Y., S.P., P.O., and D. Söll wrote the paper.

The authors declare no conflict of interest.

Abbreviations: aa-tRNA, aminoacyl-tRNA; BV, benzyl viologen; FDH_H, formate dehydrogenase H; IPTG, isopropyl β-D-thiogalactoside; PLP, pyridoxal phosphate; PSTK, phosphoseryl-tRNA^{Sec} kinase; Sec, selenocysteine; SelA, selenocysteine synthase; SelB, elongation factor SelB; SelD, selenophosphate synthetase; Sep, O-phosphoserine; SepSecS, Sep-tRNA:Sec-tRNA synthase; SerRS, seryl-tRNA synthetase.

[§]To whom correspondence should be addressed. E-mail: soll@trna.chem.yale.edu.

© 2006 by The National Academy of Sciences of the USA

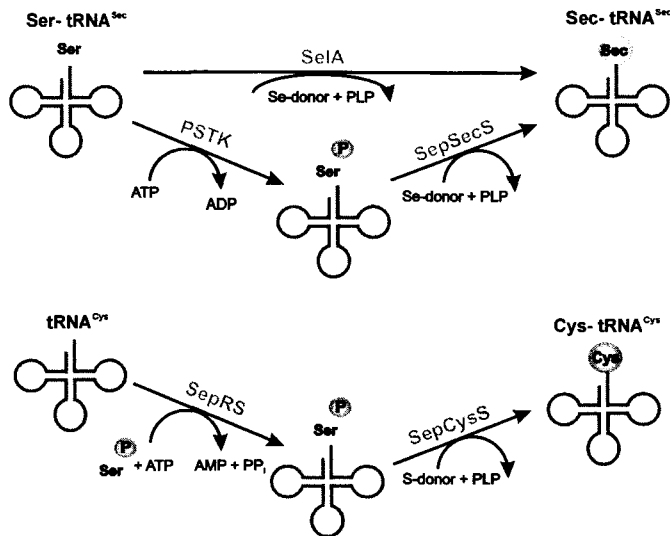


Fig. 1. tRNA-dependent amino acid transformations leading to Sec and cysteine. (Upper) The SelA route is the bacterial pathway (6). The PSTK/SepSecS is the archaeal/eukaryal route. (Lower) The SepRS/Sep-tRNA:Cys-tRNA synthase (SepCysS) pathway operates in methanogens to synthesize cysteine (16).

zymes (encoded by *spcS*) that convert Sep-tRNA^{Sec} → Sec-tRNA^{Sec}.

Results

SepSecS Rescues Selenoprotein Biosynthesis in an *E. coli selA* Deletion Strain. When grown anaerobically, *E. coli* produces the selenium-dependent formate dehydrogenase H (FDH_H). Its activity enables the cells to reduce benzyl viologen (BV) in the presence of formate, which is usually observed in agar overlay plates under anaerobic conditions (24). We therefore constructed an *E. coli selA* deletion strain (JS1) in which selenoprotein production is abolished. Complementation of the JS1 strain with archaeal SepSecS genes from *Methanococcus maripaludis* (MMP0595) and *Methanocaldococcus jannaschii* (MJ0610) and with the human homolog (Q9HD40) allowed us to test the ability of these genes to restore selenoprotein biosynthesis as detected by FDH_H activity. The *E. coli* cells were grown under anaerobic conditions for 24 h and overlaid with top agar containing BV and formate. Blue-colored cells indicated that BV was reduced and therefore FDH_H activity was present, whereas cells without FDH_H activity remained colorless. Neither the archaeal nor the human SepSecS genes were able to complement the *E. coli* JS1 strain (Fig. 2). However, complementation was achieved when the SepSecS genes were cotransformed with the gene encoding *Methanocaldococcus jannaschii* PSTK; the latter gene alone was not able to restore Sec synthesis in the *E. coli* JS1 strain. Confirmation of these results was with MacConkey nitrate agar plates where cells with inactive formate dehydrogenase N form dark red colonies (data not shown). These data imply that Sec formation was achieved by a two-step process: PSTK phosphorylates the endogenous Ser-tRNA^{Sec} to Sep-tRNA^{Sec}, and then this misacylated aa-tRNA species is converted to Sec-tRNA^{Sec} by SepSecS. Thus, the human and archaeal SepSecS enzymes are active in *E. coli*. The data also demonstrate that Ser-tRNA^{Sec}, present in *E. coli* JS1, cannot be converted to Sec-tRNA^{Sec}, and therefore is not a substrate for SepSecS.

The bacterial SelA is a PLP-dependent enzyme with a critical lysine residue engaged in PLP binding (25). As the archaeal and eukaryal SepSecS proteins also have PLP binding domains (21), we introduced Lys → Ala mutations to replace the critical lysine position in these enzymes (*Methanocaldococcus jannaschii* SepSecS

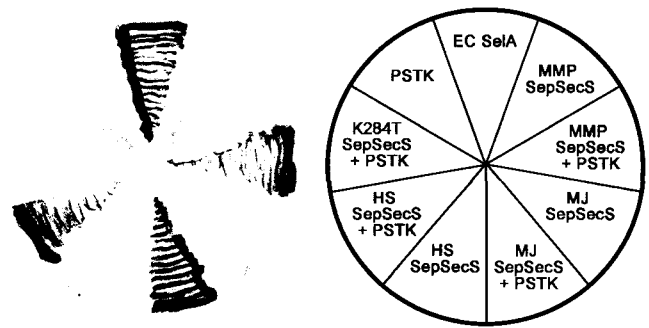


Fig. 2. SepSecS genes restore FDH_H activity in an *E. coli selA* deletion strain. The *E. coli* JS1 strains complemented with the indicated SepSecS genes were grown anaerobically on glucose minimal medium plates supplemented with 0.01 mM IPTG at 30°C for 2 days. FDH_H activity was observed by overlaying top agar containing formate and BV. Colonies with active FDH_H reduced BV to a blue color. The SepSecS genes were from *Methanocaldococcus jannaschii* (MJ), *Methanococcus maripaludis* (MMP), human (HS), and *E. coli selA* (EC). K284T denotes a mutant human SepSecS where the Lys residue critical to PLP binding is changed to Thr.

K277A, *Methanococcus maripaludis* SepSecS K278A, and human SepSecS K284A). Even in the presence of PSTK, the mutant SepSecS genes were no longer able to complement the bacterial *selA* deletion (the result for the human SepSecS mutant is shown in Fig. 2). This finding strengthens the notion that the SepSecS proteins need PLP to carry out the Sep → Sec conversion.

⁷⁵Selenium Incorporation into *E. coli* Proteins. To directly follow selenoprotein production in *E. coli*, we labeled with ⁷⁵Se the *E. coli* JS1 transformants (with SepSecS genes). These JS1 transformant strains were grown anaerobically in the presence of [⁷⁵Se]selenite and formate. Subsequently, the cell extracts were separated on a polyacrylamide gel and the radioactively labeled FDH_H was visualized by autoradiography. In agreement with the BV assay results, ⁷⁵Se-labeled FDH_H was detectable in the presence of archaeal SepSecS only in the strains also transformed with PSTK, whereas no labeling occurred when PSTK was absent (Fig. 3). The relatively lower labeling efficiency compared with the positive control, which was the transformant with *E. coli* WT *selA*, may be caused by the inefficient expression of the archaeal enzyme and the heterologous Sec biosynthesis

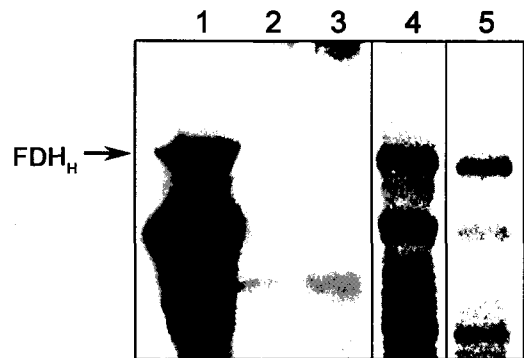


Fig. 3. ⁷⁵Se incorporation into the *E. coli* selenoprotein FDH_H. Cells were grown in the presence of [⁷⁵Se]selenite in TGYEP medium (see Materials and Methods) supplemented with 0.05 mM IPTG anaerobically at 37°C for 24 h. Cell extracts were separated by 10% SDS/PAGE, and the formation of ⁷⁵Se-containing FDH_H was followed by autoradiography. The *E. coli* strain JS1 was complemented with *E. coli* SelA (lane 1), empty plasmids (lane 2), *Methanocaldococcus jannaschii* SepSecS (lane 3), *Methanocaldococcus jannaschii* SepSecS with PSTK (lane 4), and *Methanocaldococcus jannaschii* SepSecS, PSTK with *Methanococcus maripaludis* tRNA^{Sec} (lane 5).

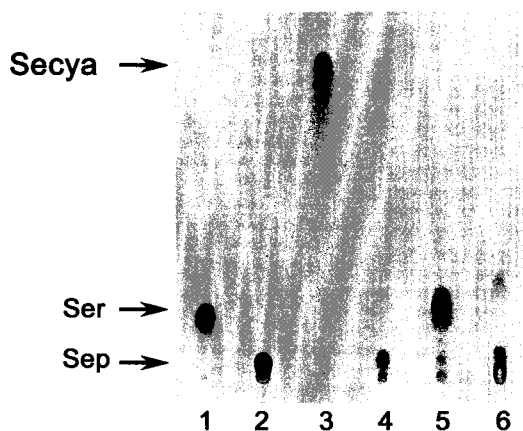


Fig. 4. Conversion of *in vitro*-synthesized Sep-tRNA^{Sec} to Sec-tRNA^{Sec}. Phosphorimages of TLC separation of [¹⁴C]Sep and [¹⁴C]Sec recovered from the aa-tRNAs of the SepSecS activity assay (see *Materials and Methods*). Sec was analyzed in its oxidized form as selenocysteic acid (Secya). Lane 1, Ser marker; lane 2, Sep marker; lane 3, Sep-tRNA^{Sec} with *Methanococcus maripaludis* SepSecS; lane 4, Sep-tRNA^{Sec} with human SepSecS; lane 5, Ser-tRNA^{Sec} with *Methanococcus maripaludis* SepSecS; lane 6, reaction of lane 3 with selenite omitted.

system. The second labeled band with lower molecular weight was likely caused by the degradation of FDH_{II} considering the cells were harvested in stationary phase.

In a different experiment we investigated the effect of a homologous SepSecS:tRNA^{Sec} system on selenium incorpora-

tion by coexpressing archaeal tRNA^{Sec} in the *E. coli* JS1 strain. However, no significant increase in FDH_{II} production was observed, which suggests that the endogenous *E. coli* tRNA^{Sec} is a good substrate for archaeal SepSecS. As the secondary structure of tRNA^{Sec} differs considerably among tRNA species from the different domains, SepSecS might display a relaxed recognition of the tRNA moiety of the Sep-tRNA substrate.

SepSecS Converts Sep-tRNA^{Sec} to Sec-tRNA^{Sec} *In Vitro*. To prove that SepSecS is a SelA that catalyzes a tRNA-dependent Sep → Sec synthesis we analyzed the conversion of [¹⁴C]Sep-tRNA to [¹⁴C]Sec-tRNA. To this aim *Methanococcus maripaludis* tRNA^{Sec} was acylated with [¹⁴C]serine by using pure *Methanococcus maripaludis* SerRS (26). Phosphorylation was performed with pure *Methanocaldococcus jannaschii* PSTK. The isolated tRNA was then incubated under anaerobic conditions with purified recombinant SepSecS and *E. coli* selenophosphate synthetase (SelD) in the presence of selenite. After the reaction aa-tRNA was deacylated, and the amino acids were identified by TLC in two systems (one is shown in Fig. 4). The data show that SepSecS from both *Methanococcus maripaludis* (Fig. 4, lane 3) and human (Fig. 4, lane 4) were able to form Sec-tRNA^{Sec} from Sep-tRNA^{Sec}, but not from Ser-tRNA^{Sec} (Fig. 4, lane 5). Thus, Sep-tRNA is the crucial precursor for Sec-tRNA formation in archaea and eukarya.

Evolution of Two Sec Biosynthesis Pathways. Although genetic coding of Sec is limited in organismal distribution, selenoproteins are found in all domains of life. Phylogenetic evidence demonstrates that the core Sec-decoding genes (tRNA^{Sec}, SelB, and SelD), and thus translation of genetically encoded Sec, is an ancient process

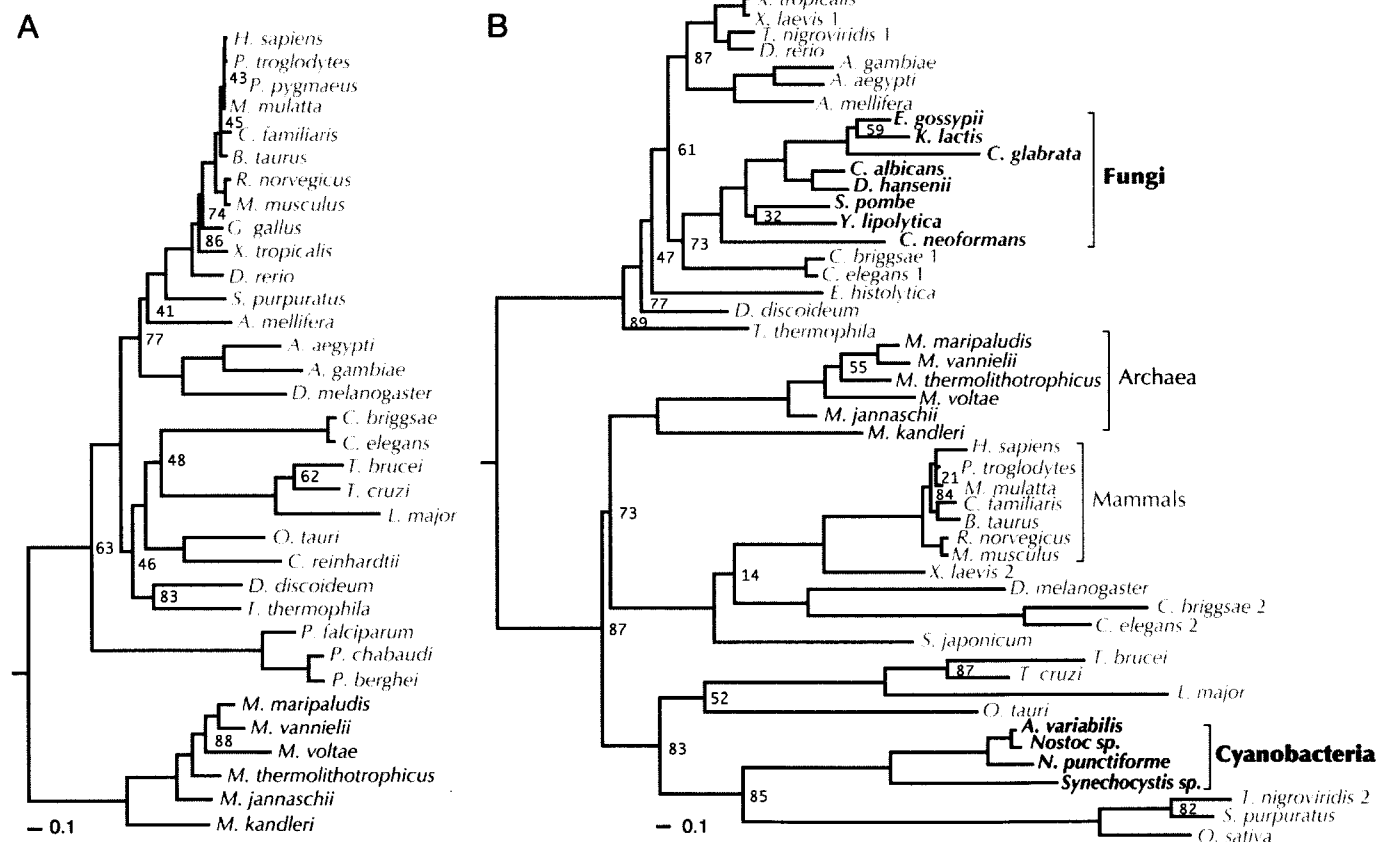


Fig. 5. Phylogenies of SepSecS (A) and PSTK (B). Organism names are color-coded according to the domain of life: Eukarya (green), Archaea (blue), and Bacteria (red). Non-Sec-decoding organisms are labeled in bold. Scale bar shows 0.1 changes per site. Only bootstrap values <90 are shown.

that was already in existence at the time of the last universal common ancestor (27–29). SelA is of bacterial origin and was vertically inherited in that domain (30). Phylogenetic analysis of SepSecS and PSTK reveals distinct archaeal and eukaryal versions of these enzymes. This pathway can be traced back (at least) to the evolutionary split between the archaeal and eukaryal sister lineages (Fig. 5). SepSecS is confined to Sec decoding archaea and eukarya. Its evolutionary history is consistent with vertical inheritance and gene loss in non-Sec-decoding lineages (Fig. 5A).

More interesting is the complex picture presented in the PSTK phylogeny (Fig. 5B). Like SepSecS, PSTK shows an initial deep divide between a strictly eukaryal group on the one hand and an archaeal group on the other. The archaeal PSTK appears to have been horizontally transferred to several of the Sec-decoding eukarya. The unexpected association between *Oryza sativa* and *Strongylocentrotus purpuratus* implies multiple horizontal gene transfer events, whereas a single transfer to mammals early in their evolution accounts for the clustering of these organisms within the archaeal genre. An additional horizontal gene transfer of the archaeal version to non-Sec-decoding Cyanobacteria is also suggested. The eukaryal PSTK is clearly discernable in a distinct group of non-Sec-decoding fungi that lack all other Sec biosynthesis components. These fungi form a coherent cluster in the eukaryal group, implying vertical inheritance. Retention of PSTK in this group suggests that it was recruited to perform a different function. Some eukarya (e.g., *Caenorhabditis* and *Xenopus*) retained both their native eukaryal PSTK and the acquired archaeal version, whereas in other lineages, including ancestors of mammals and trypanosomes, the archaeal version displaced its eukaryal counterpart. Most surprisingly, *Plasmodium falciparum* is Sec-decoding (30, 31), yet neither the PSTK nor SelA gene can be found, which contributes to the idea that a third pathway for Sec formation may exist in some organisms.

In summary, the last universal common ancestor genetically encoded Sec, and whether one or both pathways for converting Ser-tRNA^{Sec} → Sec-tRNA^{Sec} were present in the common ancestor remains to be seen.

Discussion

SepSecS has been suggested to be involved in Sec synthesis in eukarya since the early 1990s (20). In the current work, *in vivo* and *in vitro* results revealed that SepSecS is indeed a SelA analog able to catalyze the Sec-tRNA^{Sec} formation. However, in contrast to bacterial SelA that converts Ser-tRNA^{Sec} directly to Sec-tRNA^{Sec}, SepSecS acts on the phosphorylated intermediate, Sep-tRNA^{Sec} (Fig. 1). Therefore archaeal and eukaryal Sec biosynthesis requires three enzymes, SerRS, PSTK, and SepSecS, whereas bacteria accomplish the same task without PSTK. Both SelA and SepSecS are PLP-dependent enzymes. From a chemical standpoint a Sep → Sec conversion is desirable, as Sep would provide a better leaving group (phosphate) than Ser (water) for replacement with selenium. The detailed reaction mechanism of SepSecS should be subject to future structural and biochemical investigations. Sep-tRNA is reported to be more stable than Ser-tRNA (17), which is known to deacylate faster than any other aa-tRNA species (32). This fact would greatly improve the overall efficiency of the selenocysteinylation reaction and the rate of production of Sec-tRNA in all organisms that possess an extended selenoproteome. Most selenoprotein-containing bacteria only have one to three selenoproteins that are expressed under specific environmental conditions and a nonessential tRNA^{Sec} (33). In contrast, humans have an essential tRNA^{Sec} and an extended selenoproteome whose 25 members have been variously implicated in health and disease (4, 34).

Phylogenetic analyses showed that SelA is of bacterial origin, whereas SepSecS is confined to archaea and eukarya, and PSTK, principally an archaeal and eukaryal enzyme, has been horizontally transferred to the Cyanobacteria. tRNA^{Sec} has distinct secondary structure features: the long extra arm for recognition

by SerRS, the elongated acceptor stem for SelB binding (35), and still unknown recognition sites for PSTK and SepSecS. Considering the number of proteins that tRNA^{Sec} interacts with, the clear division of bacterial and archaeal/eukaryal protein lineages and the discernable differences in tRNA^{Sec} structures in the three domains of life, it is interesting to see that *E. coli* tRNA^{Sec} can be recognized by archaeal and human SepSecS at least *in vivo* (35). Moreover, although the biological archaeal and mammalian selenium donors have not been characterized (6), our results indicate that selenophosphate provided by *E. coli* SelD can be used by SepSecS both *in vitro* and *in vivo*.

Are there other pathways? The absence of both PSTK and SelA genes in *P. falciparum* brings out interesting questions: Is there a third pathway or enzyme to make Sec? And why is Sec the only amino acid among the 22 naturally cotranslationally inserted protein building blocks that has not developed its own aa-tRNA synthetase?

Materials and Methods

General. Oligonucleotide synthesis and DNA sequencing was performed by the Keck Foundation Biotechnology Resource Laboratory at Yale University. [¹⁴C]serine (163 mCi/mmol) was obtained from Amersham Pharmacia Biosciences (Piscataway, NJ). [⁷⁵Se]selenite was purchased from the University of Missouri Research Reactor Facility (Columbia, MO).

Bacterial Strains and Plasmids. The *selA* deletion strain JS1 was constructed by replacing the gene with kanamycin cassette in BW25113 *E. coli* strain as described (36). The T7 polymerase gene was inserted in the genome by P1 transduction. The *selC* deletion strain FM460 (37) was obtained from the Yale *E. coli* Stock Center. SepSecS genes from *Methanococcus maripaludis* (MMP0595) and *Methanocaldococcus jannaschii* (MJ0610) were amplified from genomic DNA, and the human SepSecS gene (GenBank accession no. BX648976) was amplified from an EST clone [clone ID: DKFZp686J1361, obtained from the RZPD German Resource Center for Genome Research, Berlin, Germany (38)] followed by cloning into the pET15b vector. *E. coli selD* was cloned into pET15b vector. *Methanococcus maripaludis* tRNA^{Sec} and *Methanocaldococcus jannaschii* PSTK gene were cloned into the pACYC vector individually or together. Human SepSecS K284T was generated by using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's directions.

Protein Expression and Purification. Human SepSecS, *Methanococcus maripaludis* SepSecS, and *E. coli* SelD were transformed into BL21(DE3) cd+ strain. Cells were grown to A₆₀₀ = 1.0, and protein production was induced by the addition of 0.5 mM isopropyl β-D-thiogalactoside (IPTG). The cells were grown for 20 h at 15°C and then spun down and resuspended in 50 mM Hepes (pH 8.0), 300 mM NaCl, 10 mM imidazole, 3 mM DTT, 10 μM PLP, and protease inhibitor mix (Roche, Indianapolis, IN). After sonication, cell lysates were applied to Ni-NTA resin (Qiagen, Valencia, CA). The resin was washed, and the His₆-tagged proteins were eluted according to the manufacturer's manual. Proteins were then thoroughly dialyzed against 50 mM Hepes (pH 8.0), 300 mM NaCl, 10 mM DTT, 10 μM PLP, and 50% glycerol.

Complementation of an *E. coli* ΔselA Strain. *Methanocaldococcus jannaschii*, *Methanococcus maripaludis*, or human SepSecS were transformed into a ΔselA *E. coli* JS1 strain, with or without the *Methanocaldococcus jannaschii* PSTK gene. The deletion strain complemented with its own *selA* gene served as positive control. *E. coli* ΔselA strain plus PSTK, SepSecS, or both genes was plated on glucose-minimum medium agar plates with 0.01 mM IPTG and grown anaerobically for 2 days at 30°C. The plates then were

overlaid with agar containing 1 mg/ml BV, 0.25 M sodium formate, and 25 mM KH_2PO_4 adjusted to pH 7.0. The appearance of blue/purple color is the indication of active FDH_H (24). The same experiments were also carried out for SepSecS mutants. Formate dehydrogenase N activity was tested with MacConkey nitrate medium plates.

Metabolic Labeling with [^{75}Se]selenite. *E. coli* cells were grown aerobically overnight and diluted (1:50) under anaerobic conditions in 5 ml of TGYEP medium (0.5% glucose, 1% tryptone, 0.5% yeast extract, 1.2% K_2HPO_4 , 0.3% KH_2PO_4 , 0.1% formate, 1 μM Na_2MoO_4 , pH adjusted to 6.5) (39) with 1 μCi [^{75}Se]selenite and 0.05 mM IPTG added. The cells were further grown anaerobically at 37°C for another 24 h and harvested. Cell lysates were prepared and subjected to SDS/PAGE, followed by autoradiography.

Preparation of tRNA Substrates. The *Methanococcus maripaludis* tRNA^{Sec} gene was cloned into pUC18 with T7 promoter and terminator. The plasmid was transformed together with a pA-CYC vector encoding T7 RNA polymerase gene to the *E. coli* *selC* deletion strain (FM460). tRNA^{Sec} was overexpressed and purified with biotinylated DNA oligonucleotides as described (40). Purified tRNA^{Sec} (10 μM) was serylated in the presence of [^{14}C]Ser (200 μM) and *Methanococcus maripaludis* SerRS (6 μM) in reaction buffer (100 mM Hepes, pH 7.0/10 mM KCl/10 mM magnesium acetate/1 mM DTT/0.1 mg/ml BSA) at 37°C for 1 h. To prepare Sep-tRNA^{Sec}, 3 μM PSTK was added to the reaction. Aminoacylated tRNA products were purified by phenol extraction and ethanol precipitation followed by G-25 desalting column application.

In Vitro Sep \rightarrow Sec Conversion. Purified recombinant SepSecS was incubated in reaction buffer (100 mM Hepes, pH 7.0/300 mM KCl/10 mM MgCl_2) with Sep-tRNA^{Sec} or Ser-tRNA^{Sec} (10 μM),

DTT (1 mM), Na_2SeO_3 (250 μM), and purified recombinant *E. coli* SelD (100 μM) at 37°C for 30 min. The oxygen in the buffer was removed, and the reaction was carried out in the anaerobic chamber. After incubation, proteins were removed by phenol extraction. The tRNAs were purified with G-25 column and ethanol precipitation to remove small molecules and salts. Purified tRNA products were deacylated in 20 mM NaOH at room temperature for 10 min. The released amino acids were oxidized with performic acid (16) and subjected to TLC analysis [system 1: 85% ethanol; system 2: butanol/acetic acid/water (4:1:1)].

Phylogenetic Analysis. Sequences, taken from the National Center for Biotechnology Information nonredundant database or the ERGO database, were aligned by using Multiseq in VMD 1.8.5 (41). The most parsimonious trees were generated with PAUP4b10 (42). Gaps were counted as missing data and a parsimony cost matrix, based on the Blossum45 substitution matrix (43), was used. Of the 1,000 most parsimonious trees, the topology with the maximum likelihood was chosen and branch lengths were optimized with PHYML v.2.4.4 (44). The JTT+ Γ model with eight rate categories was used for amino acid substitution, and adjustable parameters were derived from maximum-likelihood estimates. Bootstrap values were computed by using the re-estimation of log likelihoods (RELL) method in PROTML from the Molphy 3.2 package (45).

We thank R. Lynn Sherrer and Kelly Sheppard (Yale University) for gifts of enzymes and experimental advice and Carmen Gelpi and Lennart Randau for helpful discussions. M.J.H. held a Feodor Lynen Postdoctoral Fellowship of the Alexander von Humboldt Stiftung, and P.O. holds a National Science Foundation postdoctoral fellowship in biological informatics. This work was supported by National Institute of General Medical Sciences Grant GM22854 and Department of Energy Grant DE-FG02-98ER20311.

- Ibba M, Söll D (1999) *Science* 286:1893–1897.
- Ibba M, Söll D (2000) *Annu Rev Biochem* 69:617–650.
- Ibba M, Söll D (2004) *Genes Dev* 18:731–738.
- Allmang C, Krol A (2006) *Biochimie* 10.1016/j.biochi.2006.04.015.
- Hatfield DL, Carlson BA, Xu XM, Mix H, Gladyshev VN (2006) *Prog Nucleic Acid Res Mol Biol* 81:97–142.
- Böck A, Thanbichler M, Rother M, Resch A (2005) in *Aminoacyl-tRNA Synthetases*, eds Ibba M, Francklyn CS, Cusack S (Landes Bioscience, Georgetown, TX), pp 320–327.
- Schön A, Böck A, Ott G, Sprinzl M, Söll D (1989) *Nucleic Acids Res* 17:7159–7165.
- Leinfelder W, Zehelein E, Mandrand-Berthelot MA, Böck A (1988) *Nature* 331:723–725.
- Wu XQ, Gross HJ (1993) *Nucleic Acids Res* 21:5589–5594.
- Ohama T, Yang DC, Hatfield DL (1994) *Arch Biochem Biophys* 315:293–301.
- Sturchler-Pierrat C, Hubert N, Totsuka T, Mizutani T, Carbon P, Krol A (1995) *J Biol Chem* 270:18570–18574.
- Geslain R, Aeby E, Guitart T, Jones TE, Castro de Moura M, Charriere F, Schneider A, Ribas de Pouplana L (2006) *J Biol Chem* 10.1074/jbc.M607862200.
- Rother M, Wiltling R, Commans S, Böck A (2000) *J Mol Biol* 299:351–358.
- Bilokapic S, Korencic D, Söll D, Weygand-Durasevic I (2004) *Eur J Biochem* 271:694–702.
- Kaiser JT, Gromadski K, Rother M, Engelhardt H, Rodnina MV, Wahl MC (2005) *Biochemistry* 44:13315–13327.
- Sauerwald A, Zhu W, Major TA, Roy H, Palioura S, Jahn D, Whitman WB, Yates JR, 3rd, Ibba M, Söll D (2005) *Science* 307:1969–1972.
- Carlson BA, Xu XM, Kryukov GV, Rao M, Berry MJ, Gladyshev VN, Hatfield DL (2004) *Proc Natl Acad Sci USA* 101:12848–12853.
- Mäenpää PH, Bernfield MR (1970) *Proc Natl Acad Sci USA* 67:688–695.
- Sharp SJ, Stewart TS (1977) *Nucleic Acids Res* 4:2123–2136.
- Gelpi C, Sontheimer EJ, Rodriguez-Sanchez JL (1992) *Proc Natl Acad Sci USA* 89:9739–9743.
- Kernebeck T, Lohse AW, Grötzinger J (2001) *Hepatology* 34:230–233.
- Costa M, Rodriguez-Sanchez JL, Czaja AJ, Gelpi C (2000) *Clin Exp Immunol* 121:364–374.
- Wies I, Brunner S, Henninger J, Herkel J, Kanzler S, Meyer zum Büschenfelde KH, Lohse AW (2000) *Lancet* 355:1510–1515.
- Lacourciere GM, Levine RL, Stadtman TC (2002) *Proc Natl Acad Sci USA* 99:9150–9153.
- Tormay P, Wiltling R, Lottspeich F, Mehta PK, Christen P, Böck A (1998) *Eur J Biochem* 254:655–661.
- Kim HS, Vothknecht UC, Hedderich R, Celic I, Söll D (1998) *J Bacteriol* 180:6446–6449.
- Keeling PJ, Fast NM, McFadden GI (1998) *J Mol Evol* 47:649–655.
- Foster CB (2005) *Mol Biol Evol* 22:383–386.
- Zhang Y, Romero H, Salinas G, Gladyshev VN (2006) *Genome Biol* 7:R94.
- Lobanov AV, Delgado C, Rahlfs S, Novoselov SV, Kryukov GV, Gromer S, Hatfield DL, Becker K, Gladyshev VN (2006) *Nucleic Acids Res* 34:496–505.
- Mourier T, Pain A, Barrell B, Griffiths-Jones S (2005) *RNA* 11:119–122.
- Matthaei JH, Voigt HP, Heller G, Neth R, Schöch G, Kübler H, Amelunxen F, Sander G, Parmeggiani A (1966) *Cold Spring Harb Symp Quant Biol* 31:25–38.
- Kryukov GV, Gladyshev VN (2004) *EMBO Rep* 5:538–543.
- Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigo R, Gladyshev VN (2003) *Science* 300:1439–1443.
- Baron C, Böck A (1991) *J Biol Chem* 266:20375–20379.
- Datsenko KA, Wanner BL (2000) *Proc Natl Acad Sci USA* 97:6640–6645.
- Sawers G, Heider J, Zehelein E, Böck A (1991) *J Bacteriol* 173:4983–4993.
- Wiemann S, Weil B, Wellenreuther R, Gassenhuber J, Glassl S, Ansong W, Bocher M, Blocker H, Bauersachs S, Blum H, et al. (2001) *Genome Res* 11:422–435.
- Begg YA, Whyte JN, Haddock BA (1977) *FEMS Microbiol Lett* 2:47–50.
- Rinehart J, Krett B, Rubio MA, Alfonso JD, Söll D (2005) *Genes Dev* 19:583–592.
- Roberts E, Eargle J, Wright D, Luthy-Schulten Z (2006) *BMC Bioinformatics* 7:382.
- Swofford D (2003) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA), version 4.
- Marsh TL, Reich CI, Whitlock RB, Olsen GJ (1994) *Proc Natl Acad Sci USA* 91:4180–4184.
- Guindon S, Gascuel O (2003) *Syst Biol* 52:696–704.
- Adachi M, Hasegawa M (1996) *Comp Sci Monogr* 28:1–150.

Natural expansion of the genetic code

Alexandre Ambrogelly, Sotiria Palioura & Dieter Söll

At the time of its discovery four decades ago, the genetic code was viewed as the result of a “frozen accident.” Our current knowledge of the translation process and of the detailed structure of its components highlights the roles of RNA structure (in mRNA and tRNA), RNA modification (in tRNA), and aminoacyl-tRNA synthetase diversity in the evolution of the genetic code. The diverse assortment of codon reassignments present in subcellular organelles and organisms of distinct lineages has ‘thawed’ the concept of a universal immutable code; it may not be accidental that out of more than 140 amino acids found in natural proteins, only two (selenocysteine and pyrrolysine) are known to have been added to the standard 20-member amino acid alphabet. The existence of phosphoseryl-tRNA (in the form of tRNA^{Cys} and tRNA^{Sec}) may presage the discovery of other cotranslationally inserted modified amino acids.

Ribosomal protein synthesis is the pathway to most natural proteins. Faithful decoding of the genetic message relies on the presence of an aminoacyl-tRNA whose anticodon matches the mRNA codon, with the cognate amino acid esterified to the tRNA's 3'-terminal adenosine. Thus, synthesis of the correct aminoacyl-tRNA is critical in enabling the genetic code. Aminoacyl-tRNA (aa-tRNA) synthesis is the task of the aminoacyl-tRNA synthetases (aaRSs); each enzyme esterifies its cognate amino acid to the corresponding tRNA species¹. In certain cases, however, aa-tRNA synthesis is carried out by an evolutionarily older route based on misacylation and subsequent tRNA-dependent amino acid modification². Although over 140 different amino acids are found in natural proteins, most of them arise by post-translational modification of the 20 canonical amino acids³. The biosynthesis of proteins containing modified amino acids, which has long been a subject of great interest, relies on the existence of a designated codon and of enzymes that are able to acylate a specific tRNA with the desired amino acid. Though great strides have been made in introducing man-made amino acids into proteins based on the development of such orthogonal pairs⁴ of aaRSs and tRNA species^{5,6}, there is also a natural expansion of the genetic code beyond the 20-amino-acid repertoire that leads to the coding of modified amino acids. Here we review our current knowledge of the makeup of the genetic code and the cotranslational insertion of noncanonical amino acids into proteins.

Diversity of genetic code includes modified amino acids

The apparent universality of the genetic code at the time of its elucidation was thought to attest to the principle of common descent of all life from a single organism and provided support for Crick's “frozen accident” hypothesis regarding the evolution of the code itself⁷. According to this theory, ambiguous decoding that inevitably leads to the generation of a pool of “statistical proteins”⁸ (that is, by translation of any given gene into a group of related proteins) must have been tolerated by early life forms and may have even been advantageous for shaping the cellular machinery to its current state. However, in order to account for its universality, the code was thought to be frozen to its existing form once a certain level of cellular complexity was reached. The already improved accuracy of protein synthesis at that stage, along with any further structural and functional refinement of the translation apparatus from there on, would preclude additional codon reassignments because they would inevitably lead to disruption of an organism's whole proteome; the vast production of misfolded and aberrant proteins would greatly challenge survival of any such organism.

It was only 15 years after its deciphering that the discovery of deviations from the universal code in vertebrate mitochondria (where AUA was found to code for methionine and UGA for tryptophan) challenged the concept of a frozen, nonevolving genetic code. Further genome analyses have revealed ten codon reassignments in prokaryotic and eukaryotic nuclear codes and 16 changes in mitochondrial codes (reviewed in refs. 9–12) (Fig. 1). The correlation of mRNA codons with amino acids is the product of the interpretation of the code by the translational machinery, and therefore it is only static as long as the components of this machinery do not change and evolve. It is not surprising then that the documented codon reassignments can always be traced back to alterations in the components of the translational apparatus that are primarily involved in the decoding process: the aaRSs, which ensure correct acylation of each tRNA species with its cognate amino acid¹; the tRNA molecules, whose anticodon base pairs with the correct mRNA codon by the rules of the wobble hypothesis at the ribosome^{13,14}; and the peptide chain termination factors that recognize the termination codons¹⁵.

Interestingly, the same codon reassignments have been observed in organisms from different lineages, and a predisposition for frequent reassignment of certain codons to several alternatives has also been documented. This is particularly true for the termination (stop) codons UAG, UAA and UGA, whose single occurrence per gene renders them particularly good candidates for reassignment given that such a change would cause minimal damage to the proteome. Examination of *Escherichia coli* laboratory strains in the 1960s led to the discovery of nonsense suppression caused by tRNA species able to recognize the codons UAG (amber suppressor), UAA (ochre suppressor) or UGA (opal suppressor) (reviewed in ref. 16). UGA is the only codon

Alexandre Ambrogelly and Sotiria Palioura are in the Department of Molecular Biophysics and Biochemistry, and Dieter Söll is in the Department of Molecular Biophysics and Biochemistry and the Department of Chemistry, Yale University, New Haven, Connecticut 06520-8114, USA. e-mail: dieter.soll@yale.edu

Published online 15 December 2006; doi:10.1038/nchembio847

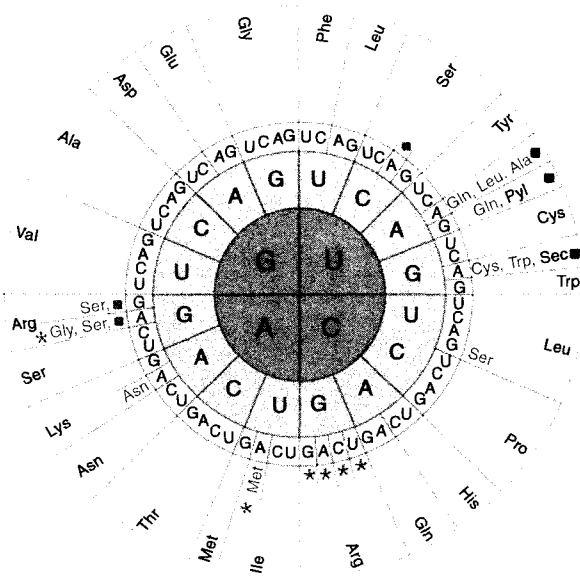


Figure 1 Summary of genetic code changes. Codons not affected by reassignments are in black. Shown in red are codons that also have changed assignments to the amino acids shown in the outer circle in blue (standard amino acids) or red (noncanonical amino acids). Filled squares denote stop codons. The blue asterisk indicates codons that may be unassigned in some organisms or organelles.

with an ambiguous meaning in organisms from all three domains of life; apart from functioning as a stop codon, an in-frame UGA also encodes selenocysteine (Sec), the 21st cotranslationally inserted amino acid¹⁷, through a recoding mechanism that requires a tRNA with a UCA anticodon (tRNA^{Sec})¹⁸, a specialized translation elongation factor (SelB) and an mRNA stem-loop structure known as the selenocysteine insertion sequence element (SECIS)¹⁹. UAG is also ambiguous in the Methanosarcinaceae, where in addition to serving as a translational stop it also encodes pyrrolysine (Pyl), the 22nd cotranslationally inserted amino acid^{20,21}; in this case, a new tRNA synthetase, pyrrolysyl-tRNA synthetase (PylRS)^{22,23}, is essential for this recoding event.

The aforementioned genetic and biochemical studies of the translation machinery over the past four decades and the recent genome analyses have made it clear that the genetic code is still evolving, even though fixation of mutations that lead to codon reassignment is not favored¹⁰. We should note that, curiously enough, the ten nuclear code changes are a subset of the 16 mitochondrial ones; that is, every single nuclear codon alteration has been observed in mitochondria as well. Thus, the same type of changes seem to occur through a conserved codon reassignment process for both nuclear and mitochondrial codes, though mitochondrial idiosyncrasies, such as the small genome (for example, 16 kilobases in humans) and the compartmentalized nature of the organelle, may account for the different rate of such changes. The mitochondrial genome is thought to be under significant pressure for genome minimization, and as such loss of a tRNA gene would reduce genome size by about 1% and bestow any such organism with a significant replication advantage. In fact, mammalian mitochondria have only 22 tRNA species. This is explained by the wobble rules: if a single tRNA can recognize up to three codons, then translation of the 61 sense codons requires a minimum of 32 tRNA species. Furthermore, the unmodified uracil present in the first anticodon position of many mitochondrial tRNAs can read all four codons in a box, thereby reducing the required number to 24 (ref. 24); sequence analysis of the *Neurospora*

crassa mitochondrial genome revealed 27 tRNA genes. Lastly, mitochondria have only one tRNA^{Met} (serving as initiator and elongator tRNA) and decode AUA as methionine (with no need for a second tRNA^{Met}). In addition, the codon-anticodon pairing rules have been stretched by the realization that anticodon conformation is shaped by the anticodon loop and stem structure²⁵, which is critically dependent on nucleotide modification (reviewed in ref. 26). Moreover, targeting of tRNA modification enzymes or edited tRNAs to mitochondria allows for codon reassignments specifically within this organelle without any effect in the nuclear codes. This is indeed the case for the 7-methylguanosine modification at the first anticodon position of tRNA^{Ser}_{GCU} that expands its decoding from AGU to AGN and thus reassigns the arginine codons AGA and AGG to serine in squid mitochondria²⁷. A similar situation has been documented in echinoderm mitochondria for the pseudouridine modification at position 35 of the tRNA^{Asn} anticodon that reassigns the lysine AAA codon to asparagine²⁸.

The above observations underscore the importance of tRNA modifications for the generation of alternate codon assignments (Table 1). RNA versatility, through post-transcriptional tRNA base modifications, thus accounts for the generation of most of the nonstandard codes. Codon reassignment in such a case would have to occur through an ambiguous decoding step in which mischarged tRNAs (if modification alters the specificity of the aminoacylation reaction) or misreading tRNAs (if modification alters codon recognition) compete with wild-type tRNAs for mRNA codon binding at the ribosomal A site. This, in turn, implies that reassignment also depends on the capacity of the organism to tolerate a somewhat destabilized proteome, which has traditionally been thought to decrease survival fitness. In fact, ambiguous decoding may not be as detrimental as it was once thought. The proteome of *Candida albicans* is unstable because the tRNA_{CAG} decodes the CUG codon as both serine and leucine; this tRNA can be charged by both seryl-tRNA synthetase (SerRS) and leucyl-tRNA synthetase²⁹. Under environmental stress, a destabilized proteome seems to be beneficial for such species²⁹. *E. coli* is able to cope with some mis-made proteins via an upregulated heat shock response (B. Ruan and S.P., unpublished data) when subjected to the survival stress of correcting the missense mutation in an essential enzyme³⁰. This indicates that ambiguous decoding is not necessarily lethal and may even prove advantageous under certain conditions. The flexibility that the genetic code can attain through tRNA modifications and the cellular tolerance of an altered proteome has made possible the engineering of microorganisms that can incorporate non-natural amino acids into their protein repertoire³¹.

As a final point, we should note that a recently proposed theory of collective evolution³² can account for the genetic code's universality and optimality (with respect to codon assignments and amino acid relatedness) without prohibiting the propensity for additional evolvability. In this theory, an innovation-sharing principle is proposed to govern survival in the primordial communal world through horizontal gene-transfer events involving translational components that would provide a selective evolutionary advantage for the community as a whole. The universality of the code established at this stage provides the *lingua franca* for efficient genetic information exchange. Darwinian vertical evolution and its "survival of the fittest" dictum would in this case stem not from a single organism but from a dynamically interacting, gene-sharing community that comprises the last universal common ancestor.

Selenocysteine: a versatile amino acid

Selenium is a dietary trace element required for animal and human health. Its antioxidant properties account for its numerous health benefits. Selenium offers protection from diverse human ailments including adverse mood states, cardiovascular disease, viral infections and cancer³³.

Selenium carries out its functions in proteins in the form of selenocysteine, in which the sulfur-containing thiol group of cysteine has been replaced by the selenium-containing selenol moiety (**Scheme 1**)³⁴.

Biochemical studies and genomic analyses have established that selenocysteine and selenoproteins are found in organisms from all three domains of life, though not in all species. In fact, most organisms do not use this amino acid. Except for its occurrence in *Chlamydomonas* spp.³⁵ and the alga *Emiliana huxleyi*³⁶, and the presence of tRNA^{Sec} in some plants³⁷, selenocysteine is notably absent from fungi and higher plants. Although the mammalian selenoproteome³⁸ contains more than 20 members that are implicated in health (for example, the essential antioxidant enzyme glutathione peroxidase and the thyroid hormone deiodinase) and disease, only a limited number of prokaryotes contain selenocysteine³⁹. For instance, within the 32 archaea with known genomes, only three species (*Methanocaldococcus jannaschii*, *Methanococcus maripaludis* and *Methanopyrus kandleri*) encode tRNA^{Sec}.

Although selenium and sulfur belong to the same group of elements in the periodic table and hence share some chemical properties (for example, electronegativity and major oxidation state), it is the different electronic structure of these two elements that makes the selenolate anion more stable than the thiolate one and gives selenoproteins their unique selenium-derived catalytic efficiencies. Indeed, the low pK_a of selenol relative to a thiol (5.2 versus 8.5, respectively) and selenol's lower redox potential (−488 mV, versus −233 mV for the thiol) renders it fully ionized at the physiological pH range and causes selenocysteine to be substantially more reactive than cysteine. Thus, it is not surprising that most known selenoproteins are oxidoreductases with an essential (for efficient catalysis) selenocysteine active site residue. A selenocysteine-to-cysteine mutation in such cases results in a >100-fold reduction in catalytic turnover (reviewed in ref. 40). A recent bioinformatic approach took into account both the catalytic advantage provided by selenocysteine and its reactivity in the presence of oxygen to explain its dynamic evolution⁴¹.



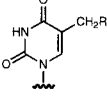
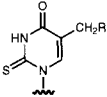
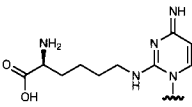
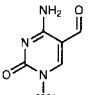
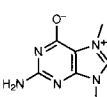
Biosynthesis of selenocysteine and UGA recoding in bacteria. Pretranslational amino acid modification is the path to Sec-tRNA^{Sec} (**Scheme 1**, top). Selenocysteine biosynthesis commences with SerRS acylating tRNA^{Sec} with serine, but because of its unusual structure SerRS does so with only 1% efficiency as compared with the other five tRNA^{Ser} isoacceptors⁴². Based on the requirement of *E. coli* formate dehydrogenase for the presence of selenocysteine for activity, a genetic system was developed that led to the identification of four *E. coli* genes (*selA*, *selB*, *selC* and *selD*), all of which are essential for selenocysteine insertion (reviewed in ref. 19). In addition to tRNA^{Sec} (*selC*), selenocysteine synthase (*selA*) and selenophosphate synthetase (*selD*) are needed for conversion of Ser-tRNA^{Sec} to Sec-tRNA^{Sec}. *selA* is a pyridoxal phosphate (PLP)-dependent enzyme with a conserved lysine residue that catalyzes the formation of an enzyme-bound dehydroalanyl intermediate with elimination of a water molecule⁴³. *selD* synthesizes (from

selenide and ATP) the selenium donor monoselenophosphate⁴⁴. This compound reacts with the *selA*-bound dehydroalanyl-tRNA^{Sec} to form Sec-tRNA. Thus, selenocysteine is made by a tRNA-dependent amino acid modification process.

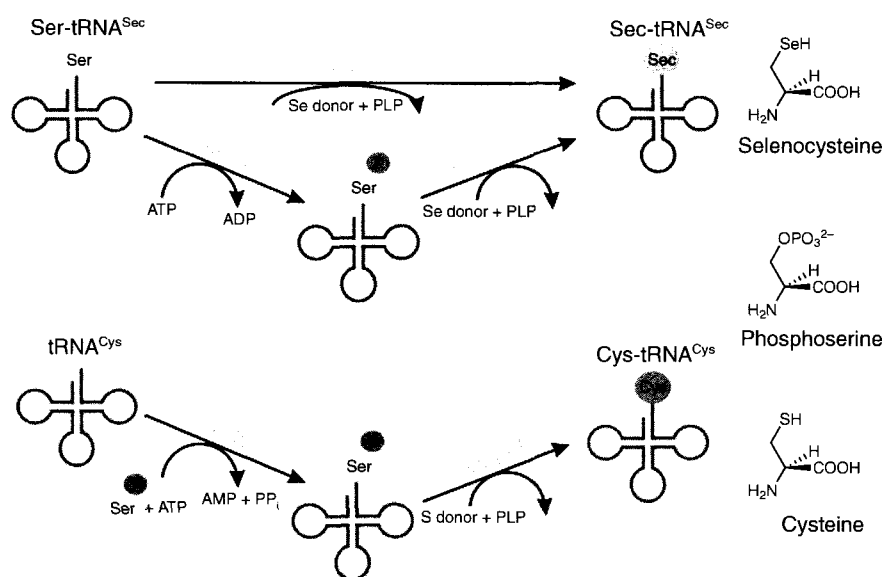
Recoding of UGA as selenocysteine requires the presence of the SECIS stem-loop mRNA structure and the specialized translation elongation factor SelB. In bacteria, the SECIS element is located immediately downstream of the UGA codon within the selenoprotein coding region. The general translation elongation factor Tu (EF-Tu) does not recognize Sec-tRNA^{Sec} effectively either *in vivo*⁴⁵ or *in vitro*⁴⁶. SelB performs two functions that are crucial for proper UGA recoding as selenocysteine: (i) its N-terminal domain, which is homologous to EF-Tu, binds to Sec-tRNA^{Sec} with high specificity—unlike EF-Tu it does not bind to any other aa-tRNA, and it is even able to discriminate against Ser-tRNA^{Sec} (ref. 47); and (ii) the binding of SelB to the SECIS element through its C-terminal domain is required for delivery of Sec-tRNA^{Sec} to the ribosomal A site, by outweighing release-factor binding to the UGA codon. Thus, the dual properties of SelB guard the fidelity of protein translation by ensuring that only UGA codons in selenoprotein mRNAs are recoded.

Selenocysteine biosynthesis and incorporation in eukaryotes and archaea. Because a selenocysteinyl-tRNA synthetase (an enzyme that would directly esterify selenocysteine to its cognate tRNA^{Sec}) has never been found, tRNA-dependent conversion of serine is assumed to be the route to selenocysteine in eukaryotes and archaea (**Scheme 1**, top). Indeed, serylation of tRNA^{Sec} by SerRS has been shown *in vitro* in both eukaryotes^{48–51} and archaea^{52–54} and is likely to occur *in vivo*

Table 1 Some tRNA modifications and their effects on codon recognition

1st anticodon base	3rd codon position recognized	Species or organelle	Structure
U	U, C, A or G	Mitochondria, chloroplasts, <i>Mycoplasma</i> spp.	
xo ⁵ U	U, A or G	Bacteria, eukaryotes	
xm ⁵ U	A > G	Bacteria, eukaryotes, mitochondria	
xm ⁵ s ² U	A > G	Bacteria, eukaryotes	
k ² C (L)	A only	Bacteria	
f ⁵ C	A or G	Mitochondria of <i>D. melanogaster</i> , nematode, <i>Bos taurus</i>	
m ⁷ G	U, C, A or G	Echinoderm, squid mitochondria	

xo⁵U, 5-hydroxymethyluridine derivative; xm⁵U, 5-methyluridine derivative; xm⁵s²U, 2-thio-5-methyluridine derivative; k²C (L), lysidine; f⁵C, 5-formylcytidine; m⁷G, 7-methylguanosine. Only the base of the nucleotides is represented.



Scheme 1 tRNA-dependent amino acid transformations leading to selenocysteine and cysteine. Top: the SelA route is the bacterial pathway¹⁹; the PSTK/SepSecS pathway is found in archaea and eukaryotes⁵². Bottom: Cys-tRNA^{Cys} synthesis in methanogenic archaea⁵⁹. The chemical structures of selenocysteine, *O*-phosphoserine and cysteine are shown on the side.

as well^{51,55}. A recent bioinformatic analysis revealed the enzyme phosphoseryl-tRNA^{Sec} (Sep-tRNA^{Sec}) kinase (PSTK) to be present only in genomes of organisms that have a selenoproteome. The activity was characterized *in vitro* for both the mouse⁵⁶ and the *M. jamareschii*⁵² enzymes; this kinase phosphorylates Ser-tRNA^{Sec} specifically to form Sep-tRNA^{Sec}. This lends credence to the old finding of Sep-tRNA^{Sec} in mammalian cell extracts^{57,58}. From a chemical standpoint, *O*-phosphoserine would provide a better leaving group (phosphate) than serine (water) for replacement with selenium. Given that a PLP-dependent selenocysteine synthase activity (as manifested by *E. coli* SelA⁴⁴ for the conversion of Ser-tRNA^{Sec} to Sec-tRNA^{Sec}) has been elusive, the mechanistically similar SepCysS enzyme⁵⁹ involved in Cys-tRNA^{Cys} formation in methanogens (see below) might engage in a tRNA-dependent synthesis of selenocysteine from *O*-phosphoserine. Another possible enzyme candidate is the human protein soluble liver antigen/liver pancreas (SLA/LP), which forms a ribonucleoprotein antigenic complex with tRNA^{Sec} in people suffering from an autoimmune chronic hepatitis⁶⁰. Bioinformatic support for this idea suggested that SLA/LP, a member of the family of PLP-dependent transferases, may have selenocysteine synthase function⁶¹. Recently, biochemical and molecular genetic studies have established that human SLA/LP, as well as its *M. maripaludis* ortholog, convert Sep-tRNA^{Sec} to Sec-tRNA^{Sec} *in vivo* and *in vitro*. Thus, the enzyme is a Sep-tRNA:Sec-tRNA synthase (SepSecS)⁶². The natural selenium donor substrate for SepSecS is unknown, but SPS1 and SPS2 are two mammalian candidate selenophosphate synthetase enzymes^{63,64}.

Regarding UGA recoding on the ribosome, several distinguishing features account for a more complex picture and a still-hazy understanding of the process in archaea and eukaryotes. In particular, the established presence of the eukaryotic and archaeal SECIS elements in the 3' untranslated region of the mRNA, the

inability of the eukaryotic and archaeal homolog of SelB to bind to the SECIS element, and the identification of other SECIS-binding proteins have led to the proposition of at least two unconfirmed models (reviewed in ref. 55) that may explain how the ribosome distinguishes a selenocysteine UGA from a termination codon and how the eukaryotic and archaeal selenoprotein mRNAs elude nonsense-mediated decay.

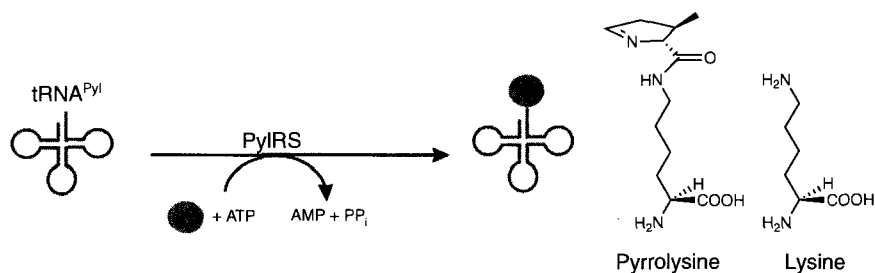
Pyrrolysine: the 22nd amino acid

Methanogens are anaerobic archaea that produce methane commonly from carbon dioxide and dihydrogen. The Methanosarcinaceae are unique among methanogenic archaea in that they are capable of methanogenesis using methylamines⁶⁵. Methane production from monomethylamine, for example, requires the enzyme monomethylamine methyltransferase. The *mtmB1* gene encoding this enzyme has an in-frame UAG codon that is translated in the Methanosarcinaceae, thereby producing the full-length methyltransferase protein (MtmB)⁶⁶. The crystal structure of this protein from *Methanosarcina barkeri* revealed that

the UAG-encoded residue is a lysine with a pyrroline-ring derivative attached to its ϵ -amino group^{20,67}; this amino acid was named pyrrolysine (**Scheme 2**). An amino acid with the expected mass was found at this position in the active site of the native MtmB enzyme⁶⁸; pyrrolysine is a crucial residue for the enzymatic reaction to proceed⁶⁶. Based on the currently known genomes, pyrrolysine is found in the Methanosarcinaceae and in the bacterium *Desulfitobacterium hafniense*.

Pyrrolysine insertion requires an orthogonal aaRS-tRNA pair. The discovery of a new *M. barkeri* amber suppressor tRNA gene suggested that pyrrolysine might be cotranslationally inserted into MtmB1 during protein synthesis and therefore genetically encoded by UAG²¹. The tRNA^{Pyl} gene (*pylT*) is located next to *pylS*, which encodes a protein similar to a class II aaRS. These two genes precede the *pylB*, *pylC* and *pylD* genes²¹.

Initial experiments²¹ suggested that Pyl-tRNA^{Pyl} synthesis proceeds via Lys-tRNA^{Pyl} (formed by PylS) and a subsequent tRNA-dependent amino acid modification. Support for a possible indirect route was also provided by the finding that Lys-tRNA^{Pyl} can be made by the simultaneous action of class I and class II lysyl-tRNA synthetases (LysRSs) that exist in the *M. barkeri* genome⁶⁹ (**Scheme 2**). However, direct

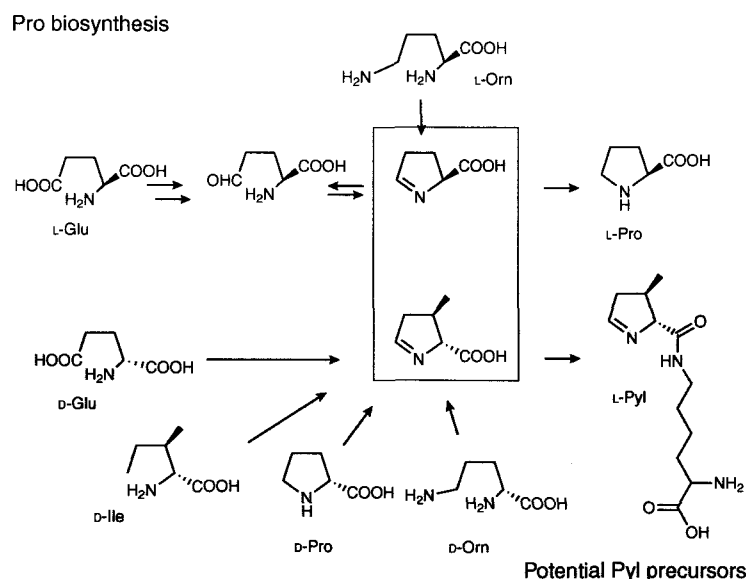


Scheme 2 Schematic representation of Pyl-tRNA^{Pyl} formation. Chemical structures of lysine and pyrrolysine are shown on the side.

charging of pyrrolysine onto tRNA^{Pyl} could not be overlooked, in particular because the *pylS* gene product, which has all the features of a *bona fide* class II aaRS, was a plausible candidate for this task. Indeed, *in vitro* studies using synthetic pyrrolysine subsequently showed that the PylS protein can charge pyrrolysine directly to tRNA^{Pyl}; lysine was not a substrate for this enzyme^{22,23} (Scheme 2). The reaction also occurs *in vivo*²³. Because of this enzymatic activity, the product of the *pylS* gene was named pyrrolysyl-tRNA synthetase (PylRS)^{22,23}. This enzyme is the first example of a tRNA synthetase that is specific for a modified amino acid, and PylRS and tRNA^{Pyl} are a naturally occurring aaRS-tRNA pair that is effectively orthogonal to the canonical genetic code.

Pyrrolysine biosynthesis. Attempts at predicting the genes involved in pyrrolysine biosynthesis rested on the idea that, like *pylS* and *pylT*, their occurrence should be restricted to pyrrolysine-incorporating organisms; only *pylD* has these characteristics⁷⁰. Although *pylB* and *pylC* were found to have homology to biotin synthase and D-alanine-D-alanine ligase, respectively, their conserved presence in the same transcriptional unit as *pylT* and *pylS* suggests some level of functional relevance to pyrrolysine metabolism or biosynthesis. Given the D stereoconformation of the pyrrolysine pyrroline ring, it is tempting to speculate that PylC may be responsible for the ligation of the pyrroline ring to the ε-amino group of lysine. Although the pyrroline ring shares strong structural homology with the direct proline precursor Δ¹-pyrroline-5-carboxylate (Scheme 3), the absence of homology of *pylB* and *pylC* to the proline biosynthetic genes (γ-glutamylphosphate reductase and γ-glutamyl kinase on one hand and ornithine aminotransferase on the other) might indicate that proline and pyrrolysine pyrroline ring biosynthesis do not share the same glutamate or ornithine precursors. It is interesting to speculate whether the pyrroline ring is instead derived from the metabolite D-isoleucine (Scheme 3), which would free the organism of the need to conduct a difficult methylation step. The 3-methyl-Δ¹-pyrroline-5-carboxylate species, a regioisomer of the pyrrolysine precursor, has been previously obtained from L-leucine by a chemically analogous route⁷¹. The pyrrolysine precursor might also potentially be obtained from the catabolism of D-proline (Scheme 3). This reaction may be carried out by proline dehydrogenase⁷², which performs this function for L-proline and whose gene is present in the Methanosarcinaceae genomes.

A stem-loop structure in mRNA is not essential for UAG recoding. A currently unanswered question is whether the *mtmB1* UAG codon is actively recoded to pyrrolysine (as is the case with UGA recoding to selenocysteine¹⁹) or whether this is a standard nonsense suppression¹⁶ event in which the presence of charged suppressor tRNA is sufficient to ensure a certain degree of UAG read-through⁷⁰. An active recoding mechanism requires additional components (specific proteins or RNA elements) that a normal suppression event does not require. The cotranslational insertion of pyrrolysine into *M. barkeri* MtmB observed in *E. coli*²³ implies that bacterial elongation factors deliver Pyl-tRNA^{Pyl} to the ribosome. This is also supported by the fact that *Thermus thermophilus* EF-Tu can bind *in vitro* a mutant tRNA^{Pyl} (with a UUU anticodon) aminoacylated with lysine⁷³.



Scheme 3 Comparison of proline biosynthesis and potential pyrrolysine precursors. Intermediate metabolites in proline biosynthesis ((5S)-Δ¹-pyrroline-5-carboxylate) and pyrrolysine biosynthesis ((4S,5R)-4-methyl-Δ¹-pyrroline-5-carboxylate) that share structural homology are boxed.

PYLIS elements (RNA structures resembling SECIS elements) were identified in *M. barkeri*⁷⁴ and *D. hafniense*² methylamine methyltransferase genes immediately downstream of the in-frame UAG codon. A synthetic oligonucleotide with sequence corresponding to the *M. barkeri* *mtmB1* PYLIS element did fold in the predicted secondary structure⁷⁵. However, a bioinformatic analysis of the regions surrounding putative SECIS and PYLIS elements suggested that selenocysteine and pyrrolysine have different decoding strategies⁷⁰ and questioned the functional significance of PYLIS elements. A recent study demonstrated successful suppression of *lacZ* and *trpA* amber mutants in *E. coli* strains transformed with *M. barkeri* *pylS* and *pylT* and grown in a medium supplemented with *N*-ε-cyclopentylloxycarbonyl-lysine, an analog that substitutes for the commercially unavailable pyrrolysine⁷⁶. The production of active β-galactosidase and tryptophan synthetase enzymes in the absence of PYLIS elements suggests that this RNA motif is not essential (at least in an *E. coli* context) to achieve sufficient UAG suppression. Future genetic studies in *Methanosarcina* spp. will provide a definitive answer to the question of the physiological significance of PYLIS elements.

Sep-tRNA: precursor to Cys-tRNA in methanogenic archaea

Cysteine is part of the 20-canonical-amino-acid repertoire ubiquitously used for protein synthesis. Though cysteine is present in proteins of *M. jannaschii*, *Methanothermobacter thermautotrophicus* and *M. kandleri* to an extent similar to that in other organisms, the genomes of these methanogens do not encode a canonical cysteinyl-tRNA synthetase (CysRS). This puzzle was solved when a new pathway (Scheme 1, bottom) for pre-translational amino acid modification was discovered for Cys-tRNA^{Cys} synthesis in these organisms⁵⁹. The first enzyme, an unusual aaRS that is homologous to the phenylalanyl-tRNA synthetase α subunit, accurately and selectively acylates tRNA^{Cys} with *O*-phosphoserine. A second enzyme then converts Sep-tRNA^{Cys} to Cys-tRNA^{Cys} in the presence of PLP and a still-unidentified sulfur donor. The two enzymes were named phosphoseryl-tRNA synthetase (SepRS, encoded by *sepS*) and Sep-tRNA:Cys-tRNA synthase (SepCysS), respectively. Searches in genomic

databases revealed the presence of homologs of these proteins not only in *M. jannaschii*, *M. kandleri* and *M. thermautotrophicus* but also in genomes of other archaea such as *M. maripaludis*, *Methanococcoides burtonii*, *Methanospirillum hungatei*, *Archaeoglobus fulgidus* and the Methanosarcinaceae, organisms that also have the canonical CysRS⁵⁹. Biochemical examination showed that SepRS is a typical class II aaRS specific solely for its substrates *O*-phosphoserine and tRNA^{Cys}. After PylRS, this is now the second natural aaRS that evolved for activating a modified amino acid.

An indication of the physiological and evolutionary significance of the SepRS/SepCysS pathway in these organisms has been provided by genetic experiments in *M. maripaludis*. This organism contains CysRS, horizontally transferred from bacteria, in addition to the indirect pathway of Cys-tRNA formation⁷⁷. Although the inactivation of the *cysS* gene encoding the canonical CysRS had no effect on cell growth in different media⁷⁸, an *M. maripaludis* strain containing a *sepS* deletion showed cysteine auxotrophy⁵⁹. The dispensability of the canonical CysRS together with the auxotrophic nature of the *M. maripaludis sepS* knockout strain demonstrates that (i) in the presence of exogenous cysteine the SepRS/SepCysS pathway and CysRS are functionally equivalent in Cys-tRNA synthesis, and (ii) the indirect route to Cys-tRNA is the sole source of free cellular cysteine in this organism⁵⁹. Thus, in *M. maripaludis* the SepRS/SepCysS pathway provides the cell with both Cys-tRNA^{Cys} and free cysteine.

Because *M. maripaludis* contains both CysRS and SepRS, the single *M. maripaludis* tRNA^{Cys} species is aminoacylated by the two unrelated synthetases—the class II enzyme SepRS (with *O*-phosphoserine) and the class I enzyme CysRS (with cysteine). The existence of two unrelated decoding systems allows a reexamination of the role of aaRSs in the evolution of the genetic code. Biochemical analysis has established that the major tRNA^{Cys} identity elements are shared by both synthetases, even though they belong to different classes⁷⁹. In three other similar cases distinct aaRSs recognize common major identity elements in tRNA (tRNA^{Lys} (ref. 80), tRNA^{Ser} (ref. 81) and tRNA^{Gly} (refs. 82, 83)). The evidence indicates convergence of the aaRSs to recognize a preestablished tRNA identity. This RNA record supports the idea that the universal genetic code evolved in the RNA world⁷⁹.

Discovery of the tRNA-dependent cysteine biosynthetic route in *M. jannaschii* may have implications that reach far beyond the problem of how Cys-tRNA^{Cys} is formed in three methanogenic archaea. Indeed, the fact that cysteine biosynthesis in *M. jannaschii* proceeds via Sep-tRNA^{Cys} and a subsequent conversion of *O*-phosphoserine to cysteine catalyzed by SepCysS suggests that the same (or a similar) enzyme may also carry out the transformation of *O*-phosphoserine to selenocysteine⁵⁹ in archaea and eukaryotes, as discussed above.

Outlook

It is exciting to know that RNA modifications and specialized RNA structures are essential in defining the coding response by fine-tuning codon-anticodon interactions and by recoding stop codons, and it is fascinating to visualize some of these structures in detail⁸⁴. Indeed, tRNA modifications seem to coevolve with the pattern of codon usage seen in different organisms, thus facilitating codon identity change. Imbalance of the proteome was once thought to be the prohibitive factor in the introduction of new amino acids to the standard alphabet. The absence of a selenocysteinyl-tRNA synthetase and the use of intricate machinery for selenocysteine incorporation was regarded as further proof that the current code and amino acid repertoire are hard to change. The ability of *Candida* spp. and *E. coli* to tolerate and even benefit from a destabilized proteome and the unexpected existence of aaRSs specific for the noncanonical amino acids pyrrolysine

and *O*-phosphoserine have revealed that the code can be adaptive to evolutionary innovations. This, in turn, raises the possibility that additional encoded amino acids with particular functions might exist in still-uncharacterized genomes.

ACKNOWLEDGMENTS

We thank P. Agris for critical comments on the paper. S.P. holds a fellowship of the Yale University School of Medicine MD/PhD Program. Work in the authors' laboratory was supported by grants from the US National Institute of General Medical Sciences (GM22854), the US Department of Energy (DE-FG02-98ER20311) and the US National Science Foundation (DBI-0535566).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/naturechemicalbiology>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Ibba, M. & Söll, D. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69**, 617–650 (2000).
2. Ibba, M. & Söll, D. Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes Dev.* **18**, 731–738 (2004).
3. Uy, R. & Wold, F. Posttranslational covalent modification of proteins. *Science* **198**, 890–896 (1977).
4. Furter, R. Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci.* **7**, 419–426 (1998).
5. Wang, L., Xie, J. & Schultz, P.G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249 (2006).
6. Köhrer, C. & RajBhandary, U.L. in *The Aminoacyl-tRNA Synthetases* (eds. Ibba, M., Francklyn, C.S. & Cusack, S.) 353–363 (Landes Bioscience, Georgetown, Texas, USA, 2005).
7. Crick, F.H.C. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
8. Woese, C.R. On the evolution of the genetic code. *Proc. Natl. Acad. Sci. USA* **54**, 1546–1552 (1965).
9. Osawa, S., Jukes, T.H., Watanabe, K. & Muto, A. Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **56**, 229–264 (1992).
10. Knight, R.D., Freeland, S.J. & Landweber, L.F. Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev. Genet.* **2**, 49–58 (2001).
11. Santos, M.A., Moura, G., Massey, S.E. & Tuite, M.F. Driving change: the evolution of alternative genetic codes. *Trends Genet.* **20**, 95–102 (2004).
12. Miranda, I., Silva, R. & Santos, M.A. Evolution of the genetic code in yeasts. *Yeast* **23**, 203–213 (2006).
13. Crick, F.H.C. Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548–555 (1966).
14. Söll, D. *et al.* Specificity of sRNA for recognition of codons as studied by the ribosomal binding technique. *J. Mol. Biol.* **19**, 556–573 (1966).
15. Kisselev, L., Ehrenberg, M. & Frolova, L. Termination of translation: interplay of mRNA, tRNAs and release factors? *EMBO J.* **22**, 175–182 (2003).
16. Eggertsson, G. & Söll, D. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiol. Rev.* **52**, 354–374 (1988).
17. Söll, D. Genetic code: enter a new amino acid. *Nature* **331**, 662–663 (1988).
18. Schön, A., Böck, A., Ott, G., Sprinzl, M. & Söll, D. The selenocysteine-inserting opal suppressor serine tRNA from *E. coli* is highly unusual in structure and modification. *Nucleic Acids Res.* **17**, 7159–7165 (1989).
19. Böck, A., Thanbichler, M., Rother, M. & Resch, A. in *The Aminoacyl-tRNA Synthetases* (eds. Ibba, M., Francklyn, C.S. & Cusack, S.) 320–327 (Landes Bioscience, Georgetown, Texas, USA, 2005).
20. Hao, B. *et al.* A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296**, 1462–1466 (2002).
21. Srinivasan, G., James, C.M. & Krzycki, J.A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462 (2002).
22. Polycarpo, C. *et al.* An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. USA* **101**, 12450–12454 (2004).
23. Blight, S.K. *et al.* Direct charging of tRNA_{CUA} with pyrrolysine *in vitro* and *in vivo*. *Nature* **431**, 333–335 (2004).
24. Heckman, J.E., Sarnoff, J., Alzner-DeWeerd, B., Yin, S. & RajBhandary, U.L. Novel features in the genetic code and codon reading patterns in *Neurospora crassa* mitochondria based on sequences of six mitochondrial tRNAs. *Proc. Natl. Acad. Sci. USA* **77**, 3159–3163 (1980).
25. Yarus, M. Translational efficiency of transfer RNAs: uses of an extended anticodon. *Science* **218**, 646–652 (1982).
26. Agris, P.F. Decoding the genome: a modified view. *Nucleic Acids Res.* **32**, 223–238 (2004).
27. Tomita, K., Ueda, T. & Watanabe, K. 7-Methylguanosine at the anticodon wobble position of squid mitochondrial tRNA^{Ser}_{GCU}: molecular basis for assignment of AGA/AGG codons as serine in invertebrate mitochondria. *Biochim. Biophys. Acta* **1399**, 78–82 (1998).
28. Tomita, K., Ueda, T. & Watanabe, K. The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine

- codon as asparagine in echinoderm mitochondria. *Nucleic Acids Res.* **27**, 1683–1689 (1999).
29. Suzuki, T., Ueda, T. & Watanabe, K. The 'polysemous' codon—a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *EMBO J.* **16**, 1122–1134 (1997).
 30. Min, B. *et al.* Protein synthesis in *Escherichia coli* with mischarged tRNA. *J. Bacteriol.* **185**, 3524–3526 (2003).
 31. Mehl, R.A. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).
 32. Vetsigian, K., Woese, C. & Goldenfeld, N. Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. USA* **103**, 10696–10701 (2006).
 33. Rayman, M.P. The importance of selenium to human health. *Lancet* **356**, 233–241 (2000).
 34. Cone, J.E., Del Rio, R.M., Davis, J.N. & Stadtman, T.C. Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. *Proc. Natl. Acad. Sci. USA* **73**, 2659–2663 (1976).
 35. Fu, L.H. *et al.* A selenoprotein in the plant kingdom. Mass spectrometry confirms that an opal codon (UGA) encodes selenocysteine in *Chlamydomonas reinhardtii* glutathione peroxidase. *J. Biol. Chem.* **277**, 25983–25991 (2002).
 36. Obata, T. & Shiraiwa, Y. A novel eukaryotic selenoprotein in the haptophyte alga *Emiliania huxleyi*. *J. Biol. Chem.* **280**, 18462–18468 (2005).
 37. Hatfield, D., Choi, I.S., Mischke, S. & Owens, L.D. Selenocysteyl-tRNAs recognize UGA in *Beta vulgaris*, a higher plant, and in *Glucadium virens*, a filamentous fungus. *Biochem. Biophys. Res. Commun.* **184**, 254–259 (1992).
 38. Kryukov, G.V. *et al.* Characterization of mammalian selenoproteomes. *Science* **300**, 1439–1443 (2003).
 39. Kryukov, G.V. & Gladyshev, V.N. The prokaryotic selenoproteome. *EMBO Rep.* **5**, 538–543 (2004).
 40. Johansson, L., Gafvelin, G. & Arner, E.S. Selenocysteine in proteins—properties and biotechnological use. *Biochim. Biophys. Acta* **1726**, 1–13 (2005).
 41. Zhang, Y., Romero, H., Salinas, G. & Gladyshev, V.N. Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox-active cysteine residues. *Genome Biol.* **7**, R94 (2006).
 42. Baron, C., Heider, J. & Böck, A. Mutagenesis of *selC*, the gene for the selenocysteine-inserting tRNA-species in *E. coli*: effects on *in vivo* function. *Nucleic Acids Res.* **18**, 6761–6766 (1990).
 43. Forchhammer, K., Boesmillier, K. & Böck, A. The function of selenocysteine synthase and SELB in the synthesis and incorporation of selenocysteine. *Biochimie* **73**, 1481–1486 (1991).
 44. Tormay, P. *et al.* Bacterial selenocysteine synthase—structural and functional properties. *Eur. J. Biochem.* **254**, 655–661 (1998).
 45. Leinfelder, W., Stadtman, T.C. & Böck, A. Occurrence *in vivo* of selenocysteyl-tRNA^{Sec} in *Escherichia coli*. Effect of *sel* mutations. *J. Biol. Chem.* **264**, 9720–9723 (1989).
 46. Forster, C., Ott, G., Forchhammer, K. & Sprinzl, M. Interaction of a selenocysteine-incorporating tRNA with elongation factor Tu from *E. coli*. *Nucleic Acids Res.* **18**, 487–491 (1990).
 47. Forchhammer, K., Leinfelder, W. & Böck, A. Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. *Nature* **342**, 453–456 (1989).
 48. Wu, X.Q. & Gross, H.J. The long extra arms of human tRNA^{(Ser)Sec} and tRNA^{Ser} function as major identify elements for serylation in an orientation-dependent, but not sequence-specific manner. *Nucleic Acids Res.* **21**, 5589–5594 (1993).
 49. Sturchler-Pierrat, C. *et al.* Selenocysteylation in eukaryotes necessitates the uniquely long aminoacyl acceptor stem of selenocysteine tRNA^{Sec}. *J. Biol. Chem.* **270**, 18570–18574 (1995).
 50. Ohama, T., Yang, D.C. & Hatfield, D.L. Selenocysteine tRNA and serine tRNA are aminoacylated by the same synthetase, but may manifest different identities with respect to the long extra arm. *Arch. Biochem. Biophys.* **315**, 293–301 (1994).
 51. Geslain, R. *et al.* Trypanosoma seryl-tRNA synthetase is a metazoan-like enzyme with high affinity for tRNA^{Sec}. *J. Biol. Chem.*, published online 13 October 2006 (doi:10.1074/jbc.M607862200).
 52. Kaiser, J.T. *et al.* Structural and functional investigation of a putative archaeal selenocysteine synthase. *Biochemistry* **44**, 13315–13327 (2005).
 53. Rother, M., Wilting, R., Commans, S. & Böck, A. Identification and characterization of the selenocysteine-specific translation factor SelB from the archaeon *Methanococcus jannaschii*. *J. Mol. Biol.* **299**, 351–358 (2000).
 54. Bilokapic, S., Korencic, D., Söll, D. & Weygand-Durasevic, I. The unusual methanogenic seryl-tRNA synthetase recognizes tRNA^{Ser} species from all three kingdoms of life. *Eur. J. Biochem.* **271**, 694–702 (2004).
 55. Allmang, C. & Krol, A. Selenoprotein synthesis: UGA does not end the story. *Biochimie* **88**, 1561–1571 (2006).
 56. Carlson, B.A. *et al.* Identification and characterization of phosphoseryl-tRNA^{(Ser)Sec} kinase. *Proc. Natl. Acad. Sci. USA* **101**, 12848–12853 (2004).
 57. Mäenpää, P.H. & Bernfield, M.R. A specific hepatic transfer RNA for phosphoserine. *Proc. Natl. Acad. Sci. USA* **67**, 688–695 (1970).
 58. Sharp, S.J. & Stewart, T.S. The characterization of phosphoseryl tRNA from lactating bovine mammary gland. *Nucleic Acids Res.* **4**, 2123–2136 (1977).
 59. Sauerwald, A. *et al.* RNA-dependent cysteine biosynthesis in archaea. *Science* **307**, 1969–1972 (2005).
 60. Gelpi, C., Sontheimer, E.J. & Rodriguez-Sanchez, J.L. Autoantibodies against a serine tRNA-protein complex implicated in cotranslational selenocysteine insertion. *Proc. Natl. Acad. Sci. USA* **89**, 9739–9743 (1992).
 61. Kernebeck, T., Lohse, A.W. & Grötzinger, J. A bioinformatical approach suggests the function of the autoimmune hepatitis target antigen soluble liver antigen/liver pancreas. *Hepatology* **34**, 230–233 (2001).
 62. Yuan, J. *et al.* RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc. Natl. Acad. Sci. USA* **103**, 18923–18927 (2006).
 63. Small-Howard, A. *et al.* Supramolecular complexes mediate selenocysteine incorporation *in vivo*. *Mol. Cell. Biol.* **26**, 2337–2346 (2006).
 64. Guimaraes, M.J. *et al.* Identification of a novel *selD* homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism? *Proc. Natl. Acad. Sci. USA* **93**, 15086–15091 (1996).
 65. Boone, D.R., Whitman, W.B. & Rouvière, P. in *Methanogenesis* (ed. Ferry, J.G.) 35–80 (Chapman & Hall, New York, 1993).
 66. Krzycki, J.A. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr. Opin. Chem. Biol.* **8**, 484–491 (2004).
 67. Hao, B. *et al.* Reactivity and chemical synthesis of L-pyrrolysine—the 22nd genetically encoded amino acid. *Chem. Biol.* **11**, 1317–1324 (2004).
 68. Soares, J.A. *et al.* The residue mass of L-pyrrolysine in three distinct methylamine methyltransferases. *J. Biol. Chem.* **280**, 36962–36969 (2005).
 69. Polycarpo, C. *et al.* Activation of the pyrrolysine suppressor tRNA requires formation of a ternary complex with class I and class II lysyl-tRNA synthetases. *Mol. Cell* **12**, 287–294 (2003).
 70. Zhang, Y., Baranov, P.V., Atkins, J.F. & Gladyshev, V.N. Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* **280**, 20740–20751 (2005).
 71. Fu, S.L. & Dean, R.T. Structural characterization of the products of hydroxyl-radical damage to leucine and their detection on proteins. *Biochem. J.* **324**, 41–48 (1997).
 72. Zhang, M. *et al.* Structures of the *Escherichia coli* PutA proline dehydrogenase domain in complex with competitive inhibitors. *Biochemistry* **43**, 12539–12548 (2004).
 73. Théobald-Dietrich, A., Frugier, M., Giegé, R. & Rudinger-Thirion, J. Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA. *Nucleic Acids Res.* **32**, 1091–1096 (2004).
 74. Namy, O., Rousset, J.P., Naphtine, S. & Brierley, I. Reprogrammed genetic decoding in cellular gene expression. *Mol. Cell* **13**, 157–168 (2004).
 75. Théobald-Dietrich, A., Giegé, R. & Rudinger-Thirion, J. Evidence for the existence in mRNAs of a hairpin element responsible for ribosome dependent pyrrolysine insertion into proteins. *Biochimie* **87**, 813–817 (2005).
 76. Polycarpo, C.R. *et al.* Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.*, published online 20 November 2006 (doi:10.1016/j.febslet.2006.11.028).
 77. Li, T. *et al.* Cysteinyl-tRNA formation: the last puzzle of aminoacyl-tRNA synthesis. *FEBS Lett.* **462**, 302–306 (1999).
 78. Stathopoulos, C. *et al.* Cysteinyl-tRNA synthetase is not essential for viability of the archaeon *Methanococcus maripaludis*. *Proc. Natl. Acad. Sci. USA* **98**, 14292–14297 (2001).
 79. Hohn, M.J., Park, H.-S., O'Donoghue, P., Schnitzbauer, M. & Söll, D. Emergence of the universal genetic code imprinted in an RNA record. *Proc. Natl. Acad. Sci. USA* **103**, 18095–18100 (2006).
 80. Ibba, M., Bono, J.L., Rosa, P.A. & Söll, D. Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci. USA* **94**, 14383–14388 (1997).
 81. Korencic, D., Polycarpo, C., Weygand-Durasevic, I. & Söll, D. Differential modes of transfer RNA^{Ser} recognition in *Methanosarcina barkeri*. *J. Biol. Chem.* **279**, 48780–48786 (2004).
 82. Mazauric, M.H. *et al.* Glycyl-tRNA synthetase from *Thermus thermophilus*—wide structural divergence with other prokaryotic glycyl-tRNA synthetases and functional inter-relation with prokaryotic and eukaryotic glycylation systems. *Eur. J. Biochem.* **251**, 744–757 (1998).
 83. Mazauric, M.H., Roy, H. & Kern, D. tRNA glycylation system from *Thermus thermophilus*. tRNA^{Gly} identity and functional interrelation with the glycylation systems from other phylae. *Biochemistry* **38**, 13094–13105 (1999).
 84. Murphy, F.V. IV, Ramakrishnan, V., Malkiewicz, A. & Agris, P.F. The role of modifications in codon discrimination by tRNA^{Ala}_{CUU}. *Nat. Struct. Mol. Biol.* **11**, 1186–1191 (2004).

Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation

Yuhei Araiso¹, Sotiria Palioura², Ryuichiro Ishitani¹, R. Lynn Sherrer²,
Patrick O'Donoghue², Jing Yuan², Hiroyuki Oshikane¹, Naoshi Domae³,
Julian DeFranco², Dieter Söll^{2,*} and Osamu Nureki^{1,4,*}

¹Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa 226-8501, Japan, ²Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114, USA, ³Biomolecular Characterization, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198 and ⁴SORST, JST, Honcho, Kawaguchi-shi, Saitama 332-0012, Japan

Received October 25, 2007; Revised November 29, 2007; Accepted November 30, 2007

ABSTRACT

The micronutrient selenium is present in proteins as selenocysteine (Sec). In eukaryotes and archaea, Sec is formed in a tRNA-dependent conversion of O-phosphoserine (Sep) by O-phosphoseryl-tRNA:selenocysteinyl-tRNA synthase (SepSecS). Here, we present the crystal structure of *Methanococcus maripaludis* SepSecS complexed with PLP at 2.5 Å resolution. SepSecS, a member of the Fold Type I PLP enzyme family, forms an (α_2)₂ homotetramer through its N-terminal extension. The active site lies on the dimer interface with each monomer contributing essential residues. In contrast to other Fold Type I PLP enzymes, Asn247 in SepSecS replaces the conserved Asp in binding the pyridinium nitrogen of PLP. A structural comparison with *Escherichia coli* selenocysteine lyase allowed construction of a model of Sep binding to the SepSecS catalytic site. Mutations of three conserved active site arginines (Arg72, Arg94, Arg307), protruding from the neighboring subunit, led to loss of *in vivo* and *in vitro* activity. The lack of active site cysteines demonstrates that a perselenide is not involved in SepSecS-catalyzed Sec formation; instead, the conserved arginines may facilitate the selenation reaction. Structural phylogeny shows that SepSecS evolved early in the history of PLP enzymes, and indicates that tRNA-dependent Sec formation is a primordial process.

INTRODUCTION

The indirect tRNA-dependent pathways of aminoacyl-tRNA formation, in which a noncognate amino acid bound to tRNA is converted to the cognate one, are widely distributed in nature. In fact, the tRNA-dependent pathways for Gln and Asn formation are evolutionarily older than the corresponding direct aminoacylation route catalyzed by the aminoacyl-tRNA synthetases (1). Although selenocysteine occurs in organisms from all three domains of life (2,3), Sec-tRNA is synthesized solely by the indirect route; actually it is the only natural amino acid found in proteins for which a cognate aminoacyl-tRNA synthetase did not evolve. Seryl-tRNA synthetase (SerRS) forms Ser-tRNA^{Sec} in bacteria (4), archaea (5,6) and eukaryotes (7). Using the selenium donor selenophosphate bacteria convert this misacylated aminoacyl-tRNA species to Sec-tRNA^{Sec} by the action of the Sela protein, a PLP-dependent selenocysteine synthase (3). Some methanogenic archaea harbor a gene that was thought to encode a Sela homolog (e.g. MJ0158), but its product is unable to synthesize Sec-tRNA^{Sec} *in vitro* (6).

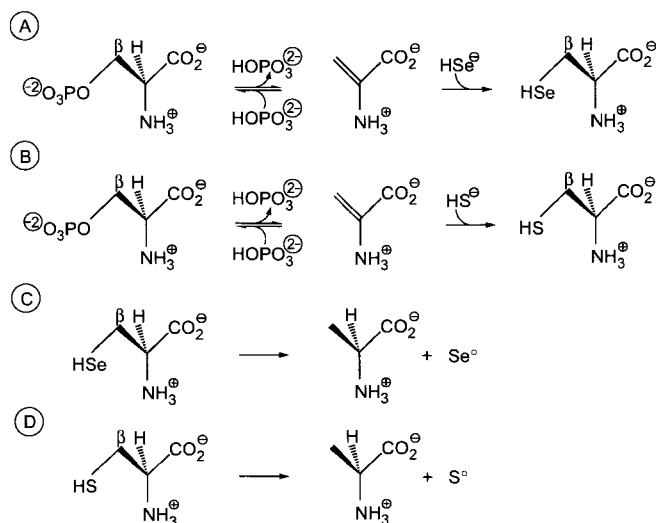
On the other hand, eukaryotes and archaea require an additional phosphorylation step catalyzed by O-phosphoseryl-tRNA^{Sec} kinase (PSTK) (8–10) and convert the resulting Sep-tRNA^{Sec} to Sec-tRNA^{Sec} by Sep-tRNA:Sec-tRNA synthase (SepSecS) (11,12). An unexpected and important property of the human SepSecS protein is the fact that it is the target antigen for soluble liver antigen/liver-pancreas (SLA/LP) autoantibodies (13–15) that are found in about a quarter of the patients with autoimmune hepatitis (16). The reactions catalyzed by PSTK and SepSecS are reminiscent of the indirect pathway of

*To whom correspondence should be addressed. Tel: +1 203 432 6200; Fax: +1 203 432 6202; Email: dieter.soll@yale.edu
Correspondence may also be addressed to Osamu Nureki. Tel: +81 45 924 5711; Fax: +81 45 924 5831; Email: nureki@bio.titech.ac.jp

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.



Scheme 1. Graphic representation of the reaction schemes. (A) MMPSepSecS catalyzes the conversion of tRNA^{Sec}-bound Sep to Sec. (B) AFSepCysS mediates the tRNA^{Cys}-dependent transformation of Sep to Cys. (C) ECCsdB converts selenocysteine to alanine and elemental selenium. (D) ECIsC converts cysteine to alanine and elemental sulfur.

Cys-tRNA^{Cys} synthesis in archaeal methanogens (17) where Sep-tRNA^{Cys} is converted to Cys-tRNA^{Cys} by Sep-tRNA:Cys-tRNA synthase (SepCysS), a PLP-dependent enzyme carrying out a β -replacement on tRNA-bound Sep. The crystal structure of *Archaeoglobus fulgidus* SepCysS has been reported (18).

The initial characterization of SepSecS revealed that this protein is a PLP-dependent enzyme (11,12). In nature such enzymes are abundant; in some microbial genomes they represent as much as 1.5% of all genes (19). They have many diverse functions and are often involved in amino acid biosynthesis (20). The structures of three other PLP-dependent enzymes that use substrates (selenocysteine, cysteine) chemically similar to those of SepSecS have been solved. One is the *A. fulgidus* SepCysS (AFSepCysS) (18), while another one is *Escherichia coli* selenocysteine lyase (ECCsdB) (21), an enzyme that converts selenocysteine to alanine and elemental selenium (22). The last enzyme is the *E. coli* cysteine desulfurase (ECIsC) that catalyzes the desulfuration of cysteine (23). The reactions catalyzed by these enzymes are illustrated in Scheme 1; SepSecS and SepCysS carry out β -replacements on tRNA-bound Sep, while selenocysteine lyase removes the β -substituent of Sec to form elemental selenium, and cysteine desulfurase catalyzes the fragmentation of cysteine to alanine and elemental sulfur. Biochemical data on SepSecS and SepCysS currently do not exist.

Here, we report the crystal structure of *Methanococcus maripaludis* SepSecS (MMPSepSecS) at 2.5 Å resolution and perform a structural comparison with ECCsdB and AFSepCysS. We propose active site residues important for the enzymatic function of MMPSepSecS by employing a combination of mutational *in vivo* and *in vitro* activity analyses. Finally, we present a structural phylogeny of the

Fold Type I family of PLP-dependent enzymes that documents the evolutionary history of SepSecS.

MATERIALS AND METHODS

General

Oligonucleotide synthesis and DNA sequencing was performed by the Keck Foundation Biotechnology Resource Laboratory at Yale University. [¹⁴C]Serine (163 mCi/mmol) and [α -³²P]ATP (10 mmol/ μ Ci) were obtained from Amersham Pharmacia Biosciences (Piscataway, NJ, USA). The *E. coli* BL21-CodonPlus (DE3)-RIL strain and the pUC18 vector were from Stratagene (LaJolla, CA, USA). The pET15b and the pACYC184 vectors were from Novagen (San Diego, CA, USA). Nickel-nitriloacetic acid agarose was from Qiagen (Valencia, CA, USA). Nickel-sepharose and Resource PHE were from GE Healthcare Bio-Sciences KK (Tokyo, Japan).

Bacterial strains and plasmids

Construction of the *E. coli* *ΔselA* deletion strain JS1 (DE3), cloning of the *M. maripaludis* SepSecS gene (MMP0595) and the *E. coli* SelD gene into the pET15b vector, of the *M. maripaludis* tRNA^{Sec} gene into the pUC18 vector, and of the *M. jannaschii* PSTK gene into the pACYC184 vector were described previously (11). The *M. maripaludis* SepSecS mutants R72A, R72Q, R72K, R94A, R94Q, H166A, H166F, H166Q, R307A, R307Q, R307K, Q102A, K278A, N247A, D277A, and K278A were generated using the QuikChange site-directed mutagenesis kit (Stratagene) and cloned into the pET15b vector with an N-terminal His-tag. An N-terminal Δ_{1-34} SepSecS deletion mutant was constructed by PCR using the primers 5'-CCGCTCGAGCATCGGAAAATTCCTGA AAACGGAATTGATGACG-3' and 5'-GCTAGTTAT TGCTCAGCGGTGGCAGC-3' and the pET15b-*sepsecS* plasmid DNA as the template. The resulting DNA fragment was digested with *Bam*HI and *Xho*I and re-inserted into the pET15b vector.

Protein expression and purification

Expression and purification of the *M. maripaludis* wild-type and mutant SepSecS proteins and the *E. coli* SelD (used for biochemical experiments) was done as described previously using Ni-NTA column chromatography (11). Purification of the wild-type SepSecS used for crystallization involved two chromatographic steps. Briefly, pET15b-*sepsecS* was transformed into the *E. coli* BL21-CodonPlus (DE3)-RIL strain, cells were grown to A₆₀₀ = 0.6 and gene expression was induced with 0.4 mM IPTG. After growth for 17 h at 20°C, cells were harvested, resuspended in 50 mM Tris-HCl (pH 7.0), 300 mM NaCl, 10 μ M PLP, 5 mM 2-mercaptoethanol, 10% glycerol, 1 mM PMSF and gently sonicated. After centrifugation at 14 000g for 30 min, the supernatant was collected and purified by sequential passage through a Ni-Sepharose and a Resource PHE chromatography column.

Table 1. Data collection and phasing statistics

X-ray source	SPRing-8 BL41XU			
	Peak	Edge	Reml	Remh
Data collection statistics	SeMet			
Wavelength (Å)	0.9791	0.9793	0.9820	0.9770
Resolution (Å)	50–2.5 (2.54–2.5)	50–2.5 (2.54–2.5)	50–2.5 (2.54–2.5)	50–2.5 (2.54–2.5)
Unique reflections	59 938	59 274	59 463	59 123
Redundancy	5.0 (2.4)	4.9 (2.3)	4.8 (2.2)	4.8 (2.0)
Completeness (%)	97.9 (86.1)	97.6 (86.1)	97.0 (81.2)	96.6 (79.3)
<i>I</i> / <i>s</i> (<i>I</i>)	17.0 (2.9)	16.4 (2.7)	16.3 (2.6)	15.3 (2.4)
<i>R</i> _{sym}	0.135 (0.258)	0.127 (0.272)	0.125 (0.261)	0.127 (0.272)
Phasing statistics				
No. of Se sites	31	31	31	31
Phasing power				
Iso (cen./acen.)	0.633/0.587	–	1.006/0.879	0.706/0.653
Ano	1.399	0.79	0.107	0.811
<i>R</i> _{culis}				
Iso (cen./acen.)	0.896/0.833	–	0.713/0.767	0.749/0.794
Ano	0.754	0.89	0.987	0.89
Mean FOM				
Cen./Acen.	0.406/0.461			

The numbers in parentheses are for the last shell.

$$R_{\text{sym}} = \frac{\sum_i |\bar{I} - I_i|}{\sum_i I_i}, R_{\text{culis}} = \frac{\sum_i |F_{\text{PH}} + F_{\text{P}} - F_{\text{H}}^{\text{calc}}|}{\sum_i |F_{\text{PH}}|}$$

Gel filtration of MMPSepSecS

A 0.5 ml sample of a 1.5 mg/ml purified solution of selenomethionine-labeled MMPSepSecS was loaded onto a HiPrep 16/60 Sephacryl S-300 HR column (GE Healthcare). The column was run at 0.5 ml/min in the same buffer as crystallization, containing 20 mM HEPES (pH 7.0), 300 mM NaCl, 10 μM PLP and 5 mM DTT. The elution volume of MMPSepSecS was compared to the elution volumes of other oligomeric proteins according to the GE healthcare web site (http://www.gelifesciences.co.jp/catalog/pdf_attach/18106088AC.pdf).

Crystallization, structure determination and refinement

The purified SepSecS was dialyzed against crystallization buffer, containing 20 mM HEPES–NaOH (pH 7.0), 10 μM PLP and 5 mM DTT, 300 mM NaCl and was concentrated. Crystals of *M. maripaludis* SepSecS were grown within a day at 20°C by the sitting-drop vapor diffusion method. Drops were prepared by mixing equal volumes of the 6 mg/ml SepSecS solution and the reservoir solution, containing 45 mM HEPES–NaOH (pH 7.0), 10 mM MES–HCl (pH 6.5), 90 mM KCl, 9 mM CaCl₂, 160 mM MgSO₄ and 10% PEG550MME. Selenomethionine-labeled SepSecS was prepared by the conventional method and was purified in the same manner as the wild-type.

The SepSecS crystals were flash-cooled in a nitrogen stream at 100 K. All diffraction data sets were collected at the BL41XU at SPRing-8 (Harima, Japan), and were processed with the HKL2000 suite. The crystals belong to the primitive monoclinic space group *P*2₁, with unit-cell parameters *a* = 75.7, *b* = 108.1, *c* = 110.4 Å, β = 97°. There are four SepSecS molecules in the asymmetric unit. A multiwavelength anomalous dispersion (MAD) data set of the selenomethionine-substituted crystals was collected, and was used to search for the locations of the selenium atoms by using the program SnB (24).

Subsequent phase refinements were performed with the program SHARP (25), and the model was manually built into the electron density maps by using the program O (26). The model was refined against reflections up to 2.5 Å resolution by using the program CNS (27). The backbones of all residues were clearly defined in the final 2*F*_o–*F*_c electron density maps. Graphic representations were prepared with CueMol (<http://www.cuemol.org>).

Statistics on data collection, phasing and refinement are shown in Tables 1 and 2.

In vivo SepSecS assay

The *M. maripaludis* wild-type and mutant SepSecS genes were transformed into the *ΔselA E. coli* JS1 strain with or without the *M. jannaschii* PSTK gene. Aerobic overnight cultures were streaked in aerobic conditions on LB-agar plates supplemented with 0.01 mM IPTG, 1 μM Na₂MoO₄, 1 μM Na₂SeO₃ and 50 mM sodium formate. The plates were placed in an anaerobic incubation jar that was flushed with an N₂: CO₂: H₂ (90 : 5 : 5) gas mix three times to give an anaerobic atmosphere and then grown for 16 h at 37°C and 36 h at 30°C. The plates were then overlaid with agar containing 1 mg/ml benzyl viologen (BV), 0.25 M sodium formate and 25 mM KH₂PO₄ adjusted to pH 7.0. The appearance of a blue/purple color is the indication of active formate dehydrogenase H (FDH_H).

Preparation and purification of tRNA gene transcripts

The *M. maripaludis* tRNA^{Sec} used as substrate in the *in vitro* assays was synthesized by *in vitro* T7 RNA polymerase run-off transcription as described (28). The tRNA^{Sec} gene together with the T7 promoter was constructed from overlapping chemically synthesized oligonucleotides, cloned into the pUC18 plasmid and purified from *E. coli* DH5α transformants using a MaxiPrep plasmid purification kit (Qiagen). The purified

Table 2. Structure refinement statistics

Refinement statistics	Se-Met
Resolution (Å)	50–2.5
No. of atoms	
Protein	13 552
Water	146
PLP	160
SO ₄	10
Luzzati coordinate error (Å)	0.3
Cross-validated Luzzati coordinate error (Å)	0.39
RMSD of	
Bond length (Å)	0.007
Bond angle (°)	1.39
Dihedral angle (°)	22.2
Improper angle (°)	0.89
Average <i>B</i> factor (Å ²)	38.5
Ramachandran plot	
Core region (%)	88.2
Additionally allowed region (%)	11.3
Generously allowed region (%)	0.3
Disallowed region (%)	0.2
<i>R</i> _{work} / <i>R</i> _{free}	0.208/0.269

$R_{\text{work}} = \sum |F_o - F_c| / \sum F_o$ for reflections of work set.

$R_{\text{free}} = \sum |F_o - F_c| / \sum F_o$ for reflections of test set (10% of total reflections).

plasmid was digested with *Bst*NI at 55°C for 16 h. The *in vitro* transcription reaction was performed at 37°C for 5 h in buffer containing 40 mM Tris–HCl (pH 8), 22 mM MgCl₂, 25 mM DTT, 2 mM spermidine, 50 µg/ml BSA, 0.1 mg/ml pyrophosphatase, 4 mM of each nucleoside triphosphate, *Bst*NI-digested vector containing the tRNA^{Sec} gene (60 µg/ml) and 1 mM T7 RNA polymerase. The tRNA^{Sec} transcript was purified by electrophoresis on a 12% denaturing polyacrylamide gel. Full-length tRNA was eluted and desalted on Sephadex G25 Microspin columns (Amersham). The tRNA transcripts were refolded by heating for 5 min at 70°C in buffer containing 10 mM Tris–HCl (pH 7.0), followed by addition of 5 mM MgCl₂ and immediate cooling on ice (5).

Preparation of ³²P-labeled Sep-tRNA^{Sec}

Refolded tRNA^{Sec} transcript was ³²P-labeled on the 3' terminus by using the *E. coli* CCA-adding enzyme and [α-³²P]AMP (Amersham) as previously described with some modifications (29). Briefly, 6 µg of tRNA^{Sec} transcript was incubated with the CCA-adding enzyme and [α-³²P]ATP (50 µCi) for 1 h at room temperature in buffer containing 50 mM Tris–HCl (pH 8.0), 20 mM MgCl₂, 5 mM DTT and 50 µM sodium pyrophosphate. After phenol/chloroform extraction the sample was passed over a Sephadex G25 Microspin column (Amersham) to remove excess ATP (30).

The recovered [³²P]-labeled tRNA^{Sec} was serylated and phosphorylated by *M. maripaludis* SerRS (5 µM) and *M. jannaschii* PSTK (1 µM) for 75 min at 37°C in buffer containing 50 mM HEPES (pH 7.5), 10 mM MgCl₂, 20 mM KCl, 1 mM DTT, 1 mM serine and 10 mM ATP. After phenol/chloroform extraction aminoacylated

Sep-tRNA^{Sec} was ethanol precipitated at –20°C for 45 min and collected as a pellet by centrifugation at 10 000 g at 4°C for 30 min. After washing the pellet with 70% ethanol, it was allowed to dry on ice in order to avoid deacylation.

To check levels of serylation and Ser→Sep conversion 1 µl aliquots at the start and end of the reaction were quenched on ice with 3 µl of 100 mM sodium citrate (pH 4.75) and 0.66 mg/ml of nuclease P1 (Sigma). Following nuclease P1 digestion at room temperature for 1 h, 1.5 µl of the sample was spotted onto polyethyleneimine (PEI) cellulose 20 cm × 20 cm thin layer chromatography (TLC) plates (Merck). To separate the Sep-[³²P]AMP spot from [³²P]AMP and any remaining Ser-[³²P]AMP the plates were developed for 75 min in buffer containing 100 mM ammonium acetate, 5% acetic acid. The plates were exposed on an imaging plate (FujiFilms) for 14 h, scanned using a Molecular Dynamics Storm 860 scanner and quantified using the ImageQuant densitometry software. The amount of Sep-tRNA^{Sec} formed can be calculated by dividing the intensity of the Sep-[³²P]AMP spot by the sum of the intensities of all spots (Sep-[³²P]AMP, [³²P]AMP and Ser-[³²P]AMP).

In vitro conversion of Sep-tRNA^{Sec} to Cys-tRNA^{Sec}

Purified wild-type or mutant MMPSepSecS (1 µM) was incubated with 1 µM ³²P-labeled Sep-tRNA^{Sec} in buffer containing 50 mM HEPES (pH 7.0), 20 mM KCl, 10 mM MgCl₂, 5 mM DTT, 2 µM PLP and 500 µM sodium thiophosphate. Reactions were carried out anaerobically at 37°C over 40 min. At each time point taken, 1 µl reaction aliquots were quenched on ice with 3 µl of 100 mM sodium citrate (pH 4.75) and 0.66 mg/ml of nuclease P1 (Sigma). Following nuclease P1 digestion, 1.5 µl of the sample was spotted onto PEI cellulose TLC plates that were developed, scanned and quantified as described above. To separate the Cys-[³²P]AMP spot from the Sep-[³²P]AMP, the [³²P]AMP and any remaining Ser-[³²P]AMP spots the plates were developed for 75 min in buffer 100 mM ammonium acetate, 5% acetic acid. The plates were exposed on an imaging plate (FujiFilms) for 14 h, scanned using a Molecular Dynamics Storm 860 scanner and quantified using the ImageQuant densitometry software. The amount of Cys-[³²P]AMP formed was calculated by dividing the intensity of the Cys-[³²P]AMP spot by the sum of the intensities of all spots (Cys-[³²P]AMP, Sep-[³²P]AMP, [³²P]AMP and Ser-[³²P]AMP). The added elevated concentration of PLP was used to assure that the mutant enzymes were saturated with the cofactor.

In vitro conversion of Sep-tRNA^{Sec} to Sec-tRNA^{Sec}

The Sep-to-Sec conversion reaction was carried out as described before (11). Briefly, purified tRNA^{Sec} (10 µM) was incubated with *M. maripaludis* SerRS (6 µM) and *M. jannaschii* PSTK (3 µM) in reaction buffer containing 100 µM [¹⁴C]Ser, 100 mM HEPES (pH 7.0), 10 mM KCl, 10 mM magnesium acetate, 1 mM DTT and 0.1 mg/ml BSA at 37°C for 1 h. The aminoacylated Sep-tRNA^{Sec}

products were purified by phenol extraction followed by passage over a Sephadex G25 Microspin column (Amersham) and ethanol precipitation. Purified wild-type SepSecS or the R72Q mutant was incubated with 10 μ M Sep-tRNA^{Sec} and 100 μ M purified *E. coli* SelD in reaction buffer containing 100 mM HEPES pH 7.0, 300 mM KCl, 10 mM MgCl₂, 1 mM DTT and 250 μ M Na₂SeO₃. All buffers were prepared anaerobically and the reaction was carried out in an anaerobic chamber at 37°C. After 30 min incubation the reaction was stopped by phenol extraction and the tRNAs were purified by application on a Sephadex G25 Microspin column (Amersham) and ethanol precipitation. Purified tRNA products were deacylated in 20 mM NaOH at room temperature for 10 min. The released amino acids were oxidized with performic acid and spotted onto silica gel 60 TLC aluminium sheets (Merck) that were subsequently developed in 85% ethanol.

Alignment and phylogeny

The STAMP (31) structural superposition algorithm in the Multiseq 2.0 module of VMD 1.8.6 (32) was used to establish a structure-based alignment between SepSecS and the other members of the fold type I PLP-dependent family. The structural similarity measure Q_H (33) was used to determine evolutionary distances between members of the fold type I group for the structural phylogeny shown in Figure 8. The tree was drawn using the programs NEIGHBOR and DRAWTREE in the Phylip 3.66 package (34). A similar structure-based alignment was used for the structure-based sequence alignment shown in Figure 2. First, SepSecS was structurally aligned to SepCysS, IscS and CsdB. This structure-based alignment was then supplemented with three additional SepSecS sequences, which had been previously aligned to MMPSepSecS with CLUSTAL (35). Some alignment ambiguities were corrected by manual adjustment to the structure-based sequence alignment.

Accession numbers

The Protein Data Bank (<http://www.rcsb.org/pdb>) accession number for the coordinates of MMPSepSecS conjugated with PLP is 2Z67.

RESULTS

Experimental outline

The MMPSepSecS protein was overproduced in *E. coli*, then purified by two column chromatographic steps, and crystallized. Using selenomethionine-labeled protein the structure was solved by the MAD method. The complex of SepSecS with covalently-bound PLP was refined to an R_{free} of 26.9% at 2.5 Å resolution.

In this first characterization of *M. maripaludis* SepSecS (MMPSepSecS) activity, we used the genetic complementation of an *E. coli* *AselA* deletion strain as an *in vivo* test. When grown anaerobically, *E. coli* produces the selenium-dependent formate dehydrogenase FDH_H. Its activity

enables the cells to reduce benzyl viologen (BV) in the presence of formate; this is usually observed by a blue/purple color in agar overlay plates under anaerobic conditions (36). Furthermore, we employed two *in vitro* tests. The first detects the enzyme's final reaction product, Sec-tRNA^{Sec}, as determined by TLC of Sec released from tRNA^{Sec} (11). Given the difficulty in working with selenophosphate (availability and oxygen sensitivity) we used thiophosphate as a surrogate substrate to measure the time course of Cys-tRNA^{Sec} formation by SepSecS. In this assay, ³²P-labeled Sep-tRNA^{Sec} was incubated anaerobically with wild-type or mutant SepSecS proteins and thiophosphate. After nuclease P1 digestion of the tRNA in the reaction mixture, the product Cys-[³²P]AMP was separated from Sep-[³²P]AMP by TLC and quantitated.

Overall structure of *M. maripaludis* SepSecS

The MMPSepSecS protein adopts an L-shaped structure consisting of the N-terminal extension domain (1–130), a catalytic domain (131–309) and a C-terminal domain (353–436) (Figure 1A). Long, kinked helices (310–352) connect the catalytic and C-terminal domains. The PLP molecule is covalently bound to the conserved Lys residue K278 (Figure 2) at the active site. As can be seen in Figure 1, the overall architecture of MMPSepSecS is similar to the Fold Type I (20) PLP enzymes AFSepCysS (18) and ECCsdB (21) (Figure 1). These enzymes consist of a catalytic domain similar to that of MMPSepSecS, and a 'small domain', formed by the N-terminal polypeptide and C-terminal polypeptide, which also resembles the MMPSepSecS C-terminal domain (Figure 1B and C) (18,21). We constructed a structure-based sequence alignment (Figure 2) in order to accurately compare MMPSepSecS to three PLP enzymes, AFSepCysS, ECCsdB and ECIScS, that act upon chemically similar substrates. According to structural similarity measures SepCysS, CsdB and IscS are significantly more closely related to each other than they are to SepSecS (Figure 3). Sequence relationships show a similar trend with SepCysS, CsdB and IscS sharing 24 identical residues while these proteins have only 8 residues in common with SepSecS (Figure 2). The alignment also shows that the archaeal SepSecS proteins lack the C-terminal extension that has been identified as the major antigenic region for the SLA/LP autoantibodies (37).

In solution, SepSecS is a tetramer as revealed by gel filtration (Figure 4E). The asymmetric unit of the crystal contains four SepSecS molecules (chains A, B, C and D) related by non-crystallographic symmetry, suggesting that SepSecS forms a homotetramer (dimer of homodimers) (Figure 4A). The four molecules are mostly identical, with root mean square (RMS) deviations of about 0.5 Å between the subunits. The N-terminal extension domain of each subunit plays a pivotal role in the tetrameric organization of SepSecS. These domains interact with each other to form a hydrophobic core (chain A with chain D, and chain B with chain C), which involves inward-facing hydrophobic residues protruding from the helices, $\alpha 1$, $\alpha 2$ and $\alpha 4$ (Figure 4B). Thus, the N-terminal

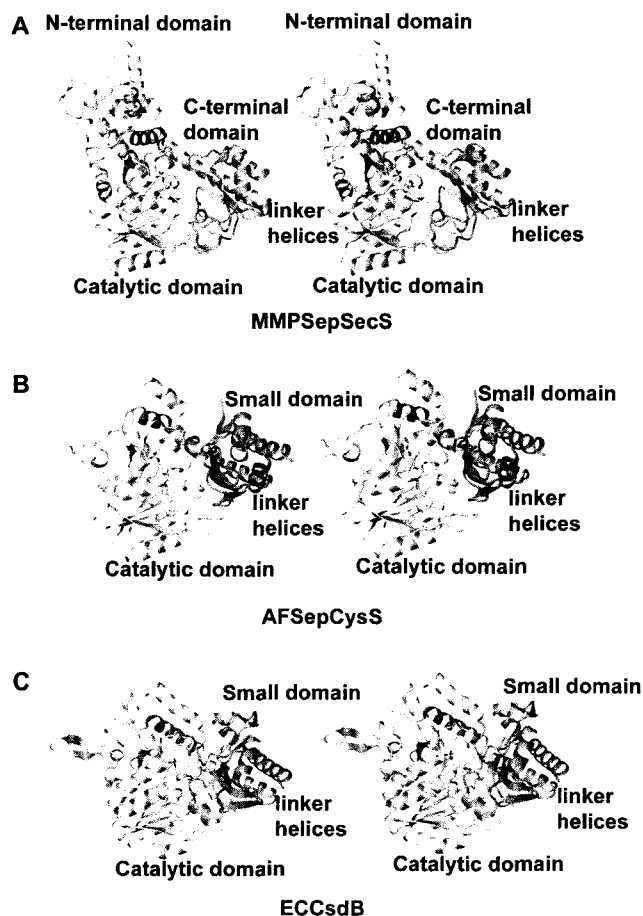


Figure 1. The overall architectures of SepSecS and related enzymes. (A) Stereo view of the MMPSepSecS structure. The N-terminal domain (residues 1–130), the catalytic domain (residues 131–309), the linker helix (310–352) and the C-terminal domain (residues 353–434) are colored yellow, green, blue and pink, respectively. The carboxy-terminal two residues are disordered. The PLP molecule bound to the active site is shown as a ball and stick model. (B) Stereo view of the AFSepCysS structure in the same orientation as (A) (18). The small domain, the linker helices and the catalytic domain are colored pink, blue and green, respectively. (C) Stereo view of the ECCsdB structure in the same orientation as (A) (21). The coloring scheme is the same as in (B).

extension facilitates tetramer formation and its deletion is predicted to produce a dimeric SepSecS. Interestingly, AFSepCysS (18) and ECCsdB (21) lack the N-terminal extension domain and form dimers (Figure 4C and D). We investigated whether SepSecS would be active in a homodimeric form by assaying the enzymatic activity of an MMPSepSecS protein that lacks the N-terminal extension residues 1–34. The Δ_{1-34} SepSecS enzyme did not form Sec-tRNA^{Sec} *in vivo* (Figure 5). It was not able to rescue selenoprotein biosynthesis in our *in vivo* complementation test of the *E. coli* *ΔselA* deletion strain. Since the active site of SepSecS, like other PLP enzymes, is formed at the dimer interface it is clear that SepSecS would not function as a monomer. Why the tetrameric organization of SepSecS is critical for function remains unclear, but it is possible that the quaternary structure of SepSecS is important for tRNA recognition. In such a scenario, when a tRNA acceptor stem is bound to the active site of

one dimer, the other dimer in the tetramer could be interacting with other regions of the tRNA.

Recognition of PLP

As in other Fold Type I PLP enzymes the active sites of MMPSepSecS lie on the dimer interface with each monomer contributing essential residues (20). The active site of chain A is formed by chains A and B that both recognize the PLP molecule (Figure 6). The catalytic domain harbors a seven-stranded β -sheet, with only the sixth β -strand being antiparallel; this is a common feature of Fold Type I PLP enzymes (20,38) including AFSepCysS (18) and ECCsdB (21). In the MMPSepSecS structure, PLP is covalently bound via a Schiff base to the strictly conserved Lys278 (chain A) (Figure 2), which is located between the sixth and seventh β strand in the active site (Figure 6A). Indeed, a Lys278Ala mutation abolishes MMPSepSecS catalytic activity as shown both *in vivo* by the lack of BV reduction by FDH_H (Figure 5), and *in vitro* by the inability of the Lys278Ala mutant to form Cys-tRNA^{Sec} (Figure 7).

All known Fold Type I PLP enzymes possess a critical aspartate that contacts the N1 atom of the pyridine ring (39), which is thought to further increase the electron sink character of the PLP cofactor. SepSecS is an exception to this paradigm, Asn247 is found in the corresponding position, and it forms a hydrogen bond with the pyridinium nitrogen (Figure 6). The other ‘nonconforming’ enzyme is SepCysS where the structure also reveals a similar Asn contact with PLP (18). The replacement of Asn247 with the uncharged Ala247 yields a partially active enzyme (Figures 5 and 7). The phosphate moiety of Sep appears to be such a good leaving group that the presence of an Asp or Asn residue is not required for the catalytic activity of SepSecS.

The aromatic pyridine ring of PLP is sandwiched between His166 (chain A) and Ala249 (chain A) through hydrophobic interactions at the bottom of the catalytic site (Figure 6). This recognition mode is commonly observed in the structures of Fold Type I PLP enzymes where the His residue increases the electron sink character of PLP’s pyridine ring through the stacking interactions with it and may also play a role in substrate activation and acid base catalysis. In the latter case, the His residue makes a direct hydrogen bond with the critical aspartate that contacts the N1 atom of the coenzyme pyridine ring in all Fold Type I PLP enzymes (see above discussion) and is thus thought to assist in the dissipation of the negative charge generated around PLP’s ring during catalysis (40). Furthermore, the phosphate moiety hydrogen bonds with the main-chain amide and carbonyl groups of Gly140 and Ser168, respectively (Figure 6B). The MMPSepSecS mutant His166Ala was partially active in forming Sec-tRNA^{Sec} *in vivo* (Figure 5). *In vitro*, the His166Ala MMPSepSecS mutant was partially active in forming Cys-tRNA^{Sec} as can be seen by the increase in intensity of the Cys-[³²P]AMP spot and the concomitant decrease of the Sep-[³²P]AMP spot during the course of the reaction (Figure 7B). Obviously, the mutant enzyme with Ala166 has PLP still in a partially functional position. A mutation

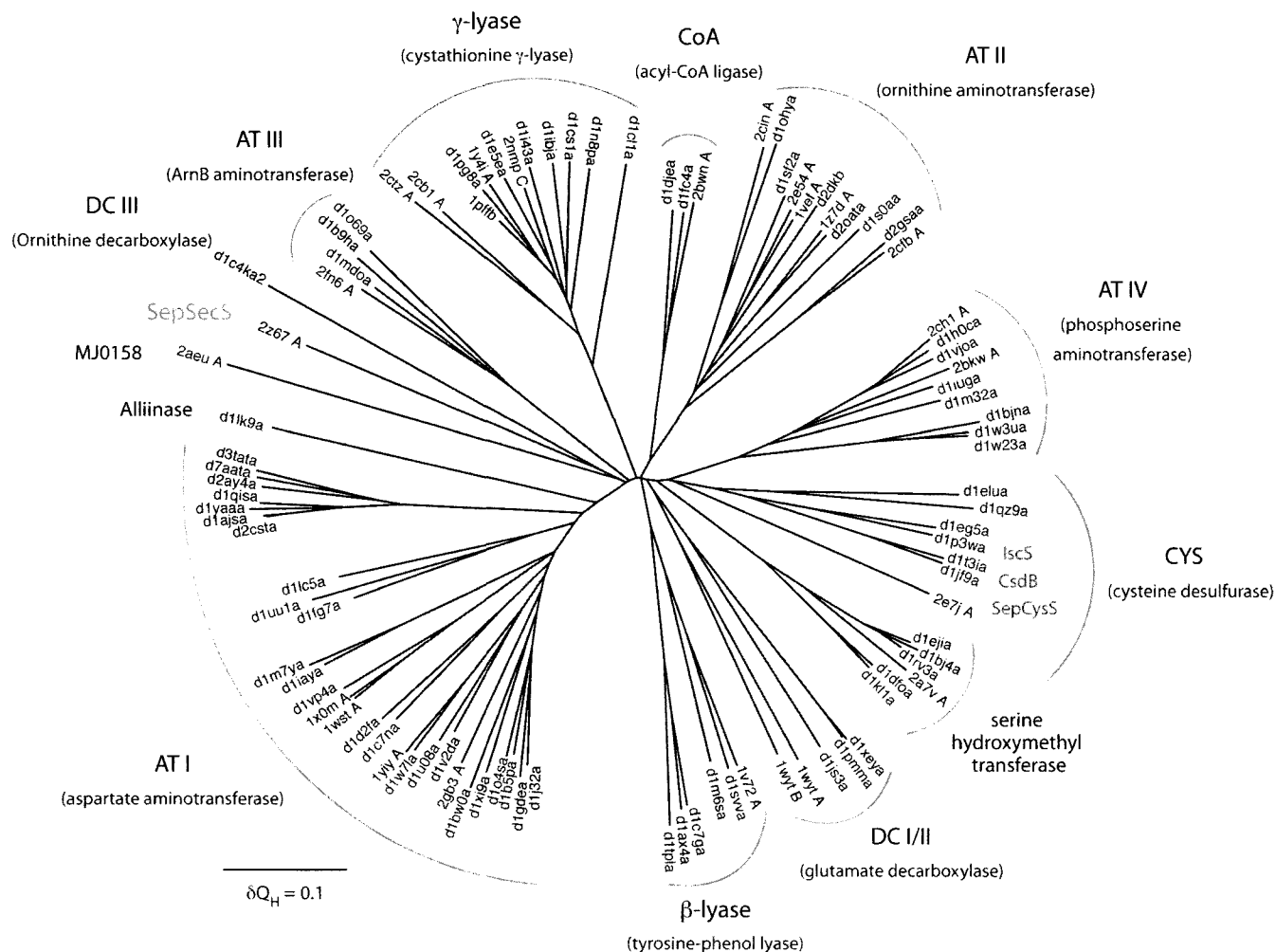


Figure 3. Structural phylogeny of the Fold Type I PLP-dependent enzyme family. The phylogenetic tree depicts the evolutionary history of the Fold Type I family. Branch lengths in the tree are proportional to structural differences between the members of the family according to the structural similarity measure Q_H (see 'Materials and methods' section). The subfamily clusters are denoted by arced lines and under each subfamily name (e.g. AT I) a representative enzyme name (e.g. aspartate aminotransferase) is given in parentheses. Isofunctional subfamilies are labeled only by their corresponding enzyme names. Protein structures used to build the tree are noted according to PDB (46) or SCOP (47) codes.

of the corresponding His residue (position 143) in *E. coli* aspartate aminotransferase also led to a functional enzyme (40). In contrast, both the His166Gln and His166Phe mutants were inactive *in vivo* (Figure 5). The lack of enzymatic activity of the His166Phe mutant implicates His166 in the catalytic mechanism and not only in structural ring stacking interactions with the pyridine ring of PLP. It is interesting to note that with the exception of *Plasmodium* all known eukaryotic SepSecS proteins contain Gln at a homologous position to 166 in their sequences (Figure 2 and data not shown).

Since the catalytic pocket is formed in the dimerization interface, chain B also contributes residues that participate in PLP recognition. In particular, the guanidinium group of Arg72 and the main-chain amide group of Arg307 (from chain B) hydrogen bond to the phosphate moiety of PLP (Figure 6B). Mutations of Arg72 and Arg307 to Ala, Gln or Lys resulted in MMPSepSecS mutants that were significantly less active in Sec-tRNA^{Sec} formation *in vivo*

(Figure 5) and Cys-tRNA^{Sec} formation *in vitro* (Figure 7C). We also show that the Arg72Gln mutant enzyme is unable to form Sec-tRNA^{Sec} *in vitro* (Figure 7A).

Such PLP recognition differs from that of AFSepCysS (18) and ECCsdB (21) (Figure 8). Unlike in the MMPSepSecS structure, the adjacent subunit of ECCsdB does not come close to the active site. In ECCsdB, an additional $\beta\alpha\beta$ structural motif covers the active site, presumably to stabilize the PLP and Sec substrates inside the pocket. This motif also precludes interaction between residues from the neighboring subunit and the active site (Figure 8D). Although the active site of AFSepCysS is formed by residues from chains A and B, the chain B amino acids are located too distant for PLP recognition (Figure 8E). The active sites of MMPSepSecS and AFSepCysS are spacious enough to accommodate the *O*-phosphoserine-CCA end of the tRNA^{Sec} or tRNA^{Cys} species, respectively.

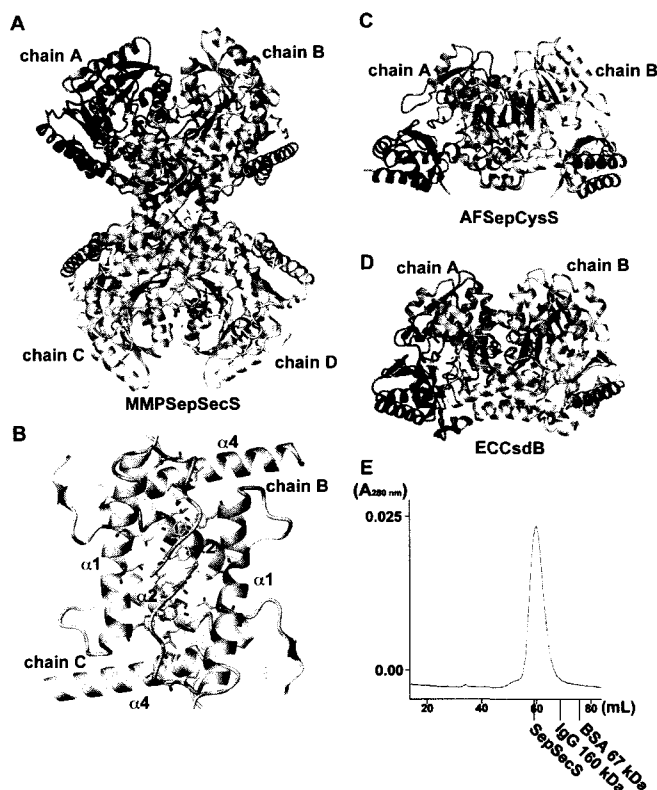


Figure 4. The oligomeric states of SepSecS and related enzymes. (A) The overall architecture of the MMPSepSecS tetramer. Chains A and B, and chains C and D form dimers, respectively. The PLP molecules bound to each subunit are shown as ball and stick models. Chain A, chain B, chain C and chain D are colored pink, blue, green and yellow, respectively. (B) The N-terminal extension domains of chains B and C form a hydrophobic core that stabilizes the tetrameric state of MMPSepSecS. Three α -helices ($\alpha 1, \alpha 2, \alpha 4$) are labeled. Chains B and C are colored blue and green, respectively. (C) Stereo view of the AFSepCysS dimer in the same orientation as (A) (18). Chains A and B are colored pink and blue, respectively. The PLP molecules are shown as ball and stick models. (D) Stereo view of the ECCsdB dimer in the same orientation as (A) (21). The coloring scheme and the PLP representation are the same as in (C). (E) Gel filtration of selenomethionine-labeled SepSecS on Sephacryl S-300. Absorbance at 280 nm is shown as a blue line. The elution volumes of other oligomeric proteins are indicated in the chromatogram. The molecular weight of the MMPSepSecS monomer is 50 kDa. MMPSepSecS eluted at the size expected for a tetrameric species.

Phosphoserine binding model

Our crystallization solution contained 10 mM magnesium sulfate. In the present structure, a strong electron density (4.5σ), presumably corresponding to a sulfate ion, was observed adjacent to the PLP molecule of chains A and B (Figure 8A). The sulfate ion is recognized by Arg94, Ser95, Gln102 and Arg307 of chain B (Figure 8A). The structure of ECCsdB with Sec bound to the active site (Figure 8D) was reported (21). Using the CE program (41) we superposed the active site structure of MMPSepSecS onto that of ECCsdB. This allowed modeling of Sec from CsdB into the MMPSepSecS active site (Figure 8B). The distance between the selenium atom of the modeled Sec and the sulfur atom of the bound sulfate is 3.74 Å. We overlaid a Sep molecule on the position of the modeled Sec in the active site. The phosphate moiety of this modeled Sep overlapped with the bound sulfate, suggesting that this sulfate mimics the phosphate moiety of Sep attached to tRNA^{Sec}. This allowed us to construct the Sep-binding model of MMPSepSecS (Figure 8C). Mutations of Arg94, Gln102 and Arg307 residues that according to our Sep-binding model recognize the phosphate moiety of Sep significantly decreased the catalytic activity of MMPSepSecS both *in vivo* (Figure 5) and *in vitro* (Figure 7). While the distance of the Schiff base linkage between the C4 α atom of PLP and the N ζ atom of Lys278 is 1.74 Å, the PLP C4 α atom is 2.19 Å away from the amino group of the modeled Sep; thus, the amino group of Sep forms a Schiff base with PLP as a reaction intermediate.

The active site

We should first note that the active site of MMPSepSecS lacks a Cys residue. In contrast, the crystal structures of ECCsdB (21) and of AFSepCysS (18) each possess a cysteine-active site residue; Cys364 in ECCsdB and Cys247 in AFSepCysS (Figure 8). As revealed by a structural study (21), Cys364 of ECCsdB recognizes Sec and withdraws its selenium to form perselenide as an intermediate. Cys364 of ECCsdB resides in the $\beta\alpha$ structural motif which does not exist in MMPSepSecS. On the other hand, Cys247 (from the neighboring subunit) of AFSepCysS is the candidate residue that forms a

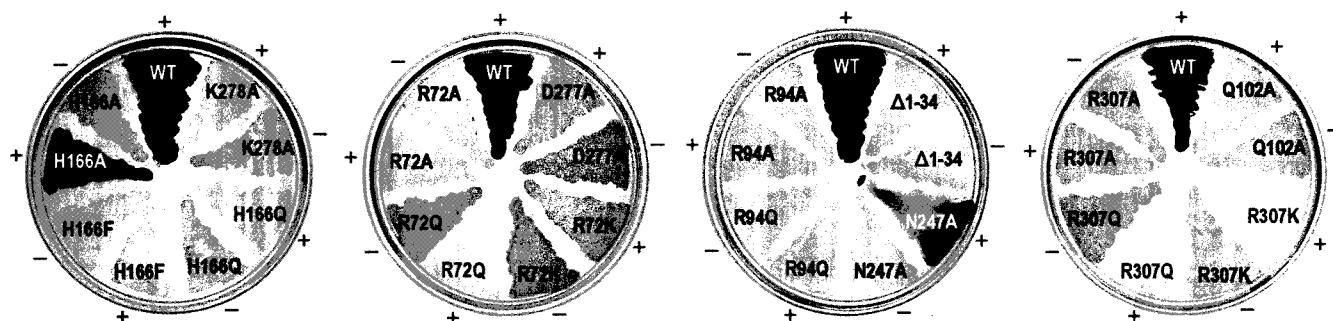


Figure 5. *In vivo* assays of SepSecS mutants. Formation of Sec-tRNA^{Sec} *in vivo* is assayed by the ability of the wild-type MMPSepSecS and its mutant variants (N-terminal deletion $\Delta 1-34$, R72A, R72Q, R72K, R94A, R94Q, H166A, H166F, H166Q, R307A, R307Q, R307K, Q102A, N247A, D277A, K278A) to restore the BV reducing activity of the selenoprotein FDH_H in the *E. coli* *selA* deletion strain JS1. Cotransformation of the PSTK gene (indicated with +) from *M. jannaschii* is required for the formation of the Sep-tRNA^{Sec} intermediate.

persulfide that is the sulfur source for enzyme-catalyzed Cys-tRNA^{Cys} formation (18). The absence of such a Cys active site residue indicates that the reaction mechanisms and chemistries of SepSecS are fundamentally different from those of CsdB and SepCysS. Interestingly, SepSecS

has an active site arginine (Arg307) in a homologous position to Cys247 in SepCysS (Figure 2).

The MMPSepSecS structure reveals three conserved arginines (Arg72, Arg94 and Arg307) that are located in proximity of each other and close to the active site; they recognize the phosphate groups of PLP and presumably of Sep acylated to tRNA^{Sec} (Figures 2 and 8A–C). Mutations of Arg72, Arg94 and Arg307 to Ala, Gln or Lys yielded MMPSepSecS enzymes that were unable to form Sec-tRNA^{Sec} *in vivo* (Figure 5). Asp277 interacts electrostatically with Arg72, and the Asp277Ala enzyme was also inactive *in vivo* (Figure 5). The Arg72Gln, Arg94Gln and Arg307Gln MMPSepSecS mutants were inactive in forming Cys-tRNA^{Sec} *in vitro* (Figure 7). In addition to the PLP-conjugated Lys278, Sep-binding arginines may also facilitate general acid/base catalysis as was shown for the tRNA modification enzyme TrmH where an arginine activated by a phosphate group acts as a general base (42).

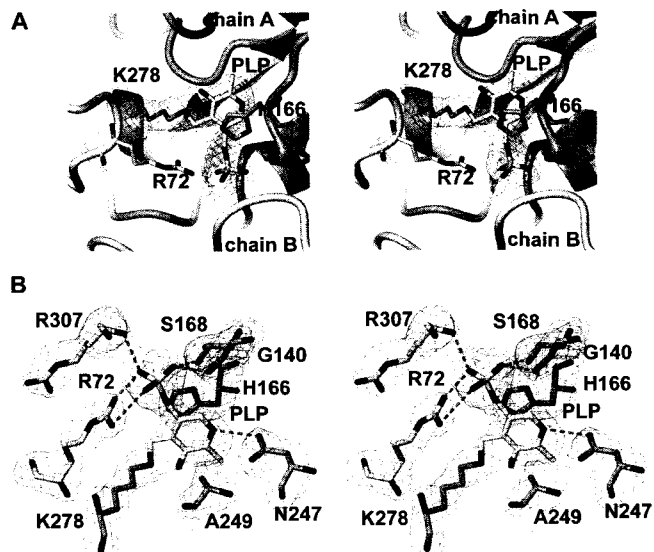


Figure 6. PLP recognition. Stereo views of (A) Ribbon representation of the active site of MMPSepSecS in the dimer interface between chains A and B that are colored pink and blue, respectively. The PLP molecule is covalently bound to Lys278 of chain A. The $F_o - F_c$ omit map of the PLP molecule, contoured at 3.5σ , is shown. (B) The amino acid residues that recognize the PLP molecule. The residues of chains A and B are colored pink and blue, respectively. The $F_o - F_c$ omit map (contoured at 3.5σ) of all of the residues and PLP is shown.

DISCUSSION

The mechanism

SepSecS is a PLP enzyme catalyzing a β -replacement (Scheme 1), leading to the exchange of a phosphate group for a selenol moiety. We were interested to see if SepSecS might employ a perselenide intermediate during catalysis, which would require the presence of an active site cysteine. There are four moderately-to-highly conserved cysteine residues in the SepSecS sequences, which are found in 35–97% of the SepSecSs. Three of these cysteines (Cys146, Cys214 and Cys237 in MMPSepSecS) are distant from the active site and have no chance to contribute to catalysis.

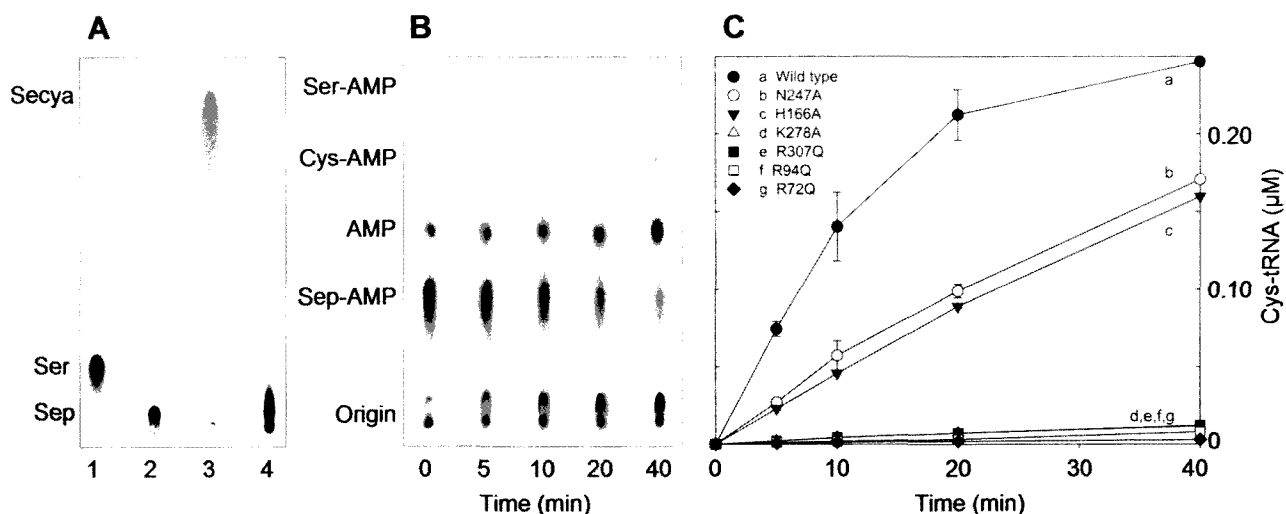


Figure 7. *In vitro* conversion of Sep-tRNA^{Sec} to Sec-tRNA^{Sec} or Cys-tRNA^{Sec}. (A) Phosphorimages of TLC separation of [¹⁴C]Sep and [¹⁴C]Sec recovered from the aa-tRNAs of the SepSecS activity assay (see 'Materials and methods' section). Sec was analyzed in its oxidized form as selenocysteic acid (Secya). Lane 1, Ser marker; lane 2, Sep marker; lane 3, Sep-tRNA^{Sec} with wild-type MMPSepSecS; lane 4, Sep-tRNA^{Sec} with the R72Q MMPSepSecS mutant. (B) Representative phosphorimage for the H166A SepSecS mutant of the separation of Ser-[³²P]AMP, Cys-[³²P]AMP, [³²P]AMP and Sep-[³²P]AMP. At the indicated time points aliquots of the SepSecS reaction were quenched, digested with nuclease P1 and spotted onto PEI-cellulose TLC plates as described in the 'Materials and methods' section. (C) Plot of Cys-tRNA^{Sec} formed versus time with 1 μM of wild-type and mutant SepSecS enzymes using Sep-tRNA^{Sec} (1 μM) and thiophosphate (500 μM) as substrates. Following quantification of the intensities of Ser-[³²P]AMP, Cys-[³²P]AMP, [³²P]AMP and Sep-[³²P]AMP using ImageQuant, the concentration of Cys-tRNA^{Sec} formed at each time point was calculated by dividing the intensity of the Cys-[³²P]AMP spot by the total intensity. The experiment was carried out in duplicate.

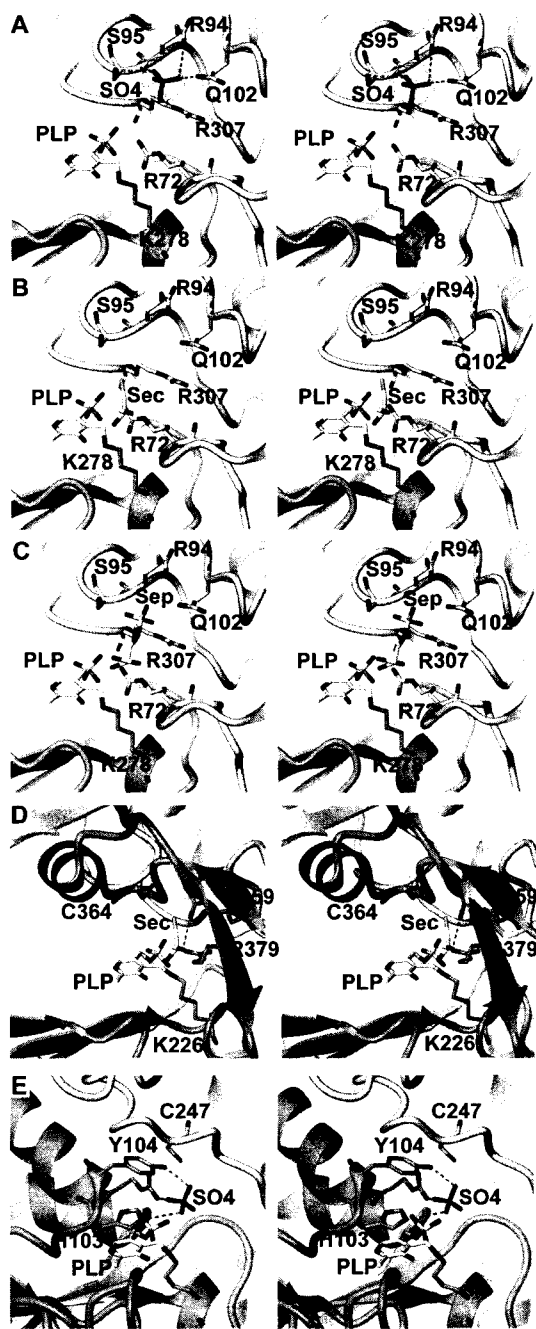


Figure 8. The active site. Close-up stereo views of the active site regions in MMPSepSecS, ECCsdB, and AFSepCysS. Chains A and B are colored pink and blue, respectively. The PLP molecule (shown as a ball and stick model) is covalently bound to chain A. (A) The recognition manner of the sulfate ion in the active site of MMPSepSecS. The sulfate ion is shown as a ball-and-stick model. The $F_o - F_c$ omit map of the sulfate ion, contoured at 4.5 σ , is shown. (B) Sec docking model of MMPSepSecS, based on the structural comparison with ECCsdB. (C) Sep binding model of MMPSepSecS, based on the locations of the sulfate ion and Sep. (D) Close-up view of the active site region in ECCsdB. The PLP molecule is covalently bound to Lys226 of chain A. The orientation of the PLP molecule is the same as that in (A). The chain A residues that recognize PLP are shown. The PLP molecule is covered by the $\beta\alpha$ structural motif. (E) Close-up view of the active site region in AFSepCysS. The PLP molecule is covalently bound to Lys209 of chain A. The orientation of PLP is the same as that in (A). The chain A residues that recognize PLP and sulfate are shown.

The least conserved of these residues is a strictly conserved proline (Pro169) in the archaeal sequences, while the residue is conserved as a cysteine in 78% of the eukaryotic sequences. Interestingly, Pro169 in MMPSepSecS is in the active site and in close contact with the PLP moiety (3.77 Å at closest approach), which would then mean that most eukaryotic SepSecSs have a cysteine adjacent to the PLP. It is unclear whether the eukaryotic sequences make use of this cysteine in catalysis; nevertheless, this position is an intriguing candidate for mutagenesis in the eukaryotic context.

Concerning the archaeal SepSecSs, there are no cysteine residues that could be involved in the formation of a catalytically important perselenide intermediate. In contrast, SepCysS is proposed to use a persulfide mechanism (18). Thus, SepSecS and SepCysS, two related PLP enzymes that perform chemically analogous tRNA-dependent transformations of Sep to Sec or Cys, respectively, proceed with different selenium and sulfur transfer mechanisms. Based on the observed active site residues we propose that the arginine residues, that bind the phosphate groups of PLP and of Sep, also recruit the selenium donor selenophosphate. In this context, Arg307 in the present structure resides at a similar position to that of Cys364 in ECCsdB and of Cys247 in AFSepCysS.

Structural phylogeny of Fold Type I PLP enzymes

SepSecS belongs to the largest and most diverse family of PLP-dependent enzymes found in nature. The evolutionary history of the Fold Type I PLP-dependent enzymes, also referred to as the α -family, has been studied in detail with sequence-based phylogenetic methods (43). By applying the family profile analysis (FPA) technique (44), the authors were able to use sequence-based phylogenetics to partially capture the distant evolutionary events recorded in the sequences of the members of this family, some of whom share only 5% sequence identity. However, the low level of sequence similarity may lead to artifacts in the tree reconstruction process. Since crystal structures for most members of this protein family exist, we applied the technique of structural phylogeny (45) to explore the history of the Fold Type I family. This technique derives phylogenetic information directly from three-dimensional structures and thus allows an accurate reconstruction of the most distantly detectable evolutionary events. Our structural phylogeny (Figure 3) is largely in agreement with the previous sequence-based work, but some significant rearrangements in the tree can be seen. For example, the DC I and DC II subfamilies coalesce into a single subfamily according to structural similarity, and the γ -lyase and CYS groups are not one but two evolutionary distinct subfamilies.

There are two kinds of subfamilies in the Fold Type I group. The first type includes several proteins with distinct but related functions, which may have descended from a progenitor with a promiscuous enzymatic function. Examples of these progenitor enzymes include the four distinct aminotransferases (AT I–AT IV), an amino acid decarboxylase (DC I/II), a persulfide forming cysteine desulfurase/sulphydrylase and finally separate β - and

γ -amino acid lyases. Subsequent evolution in each of these subfamilies ultimately produced the specific enzymatic functions observed in modern PLP-dependent enzymes.

The second kind of subfamilies are those that include a set of isofunctional enzymes. The modern enzymes in these subfamilies trace back to ancestors that had already evolved their modern enzymatic specificity in the initial evolutionary radiation of the Fold Type I family. These isofunctional subfamilies include serine hydroxymethyl transferase, alliinase and an archaeal SelA-like protein (MJ0158), which is of unknown function.

Importantly, the structural phylogeny allows the accurate placement of SepSecS within its family tree and reveals how this enzyme came into being. SepSecS is also a founding member of the Fold Type I family. It shows no specific relationship to any of the other subfamilies and emerges near the root of its family tree. While SepCysS, CsdB and IscS are part of a multifunctional subfamily (CYS in Figure 8) and these proteins have chemically similar substrates to SepSecS, there is no special relationship between SepSecS and the CYS subfamily. This result indicates that SepSecS is truly an ancient enzyme, and thus tRNA-dependent selenocysteine biosynthesis is a primordial process.

ACKNOWLEDGEMENTS

We thank the beam-line staff at BL41XU of SPring-8 (Harima, Japan) for technical help in data collection. Dan Su, Markus Englert, Kelly Sheppard, Michael Hohn, Hee-Sung Park, Juan Salazar, Motoyuki Hattori, Kotaro Nakanishi and Tomoyuki Numata participated in many discussions on this topic.

This work was supported by a SORST Program grant from Japan Science and Technology (to O.N.), by grants from the Ministry of Education, Culture, Sports, Science and Technology (to R.I. and O.N.), by a National Project on Protein Structural and Functional Analyses grant from the Ministry of Education, Culture, Sports, Science and Technology (to O.N.) and by grants from the Department of Energy, the National Institute of General Medical Sciences and the National Science Foundation (to D.S.). S.P. holds a fellowship of the Yale University School of Medicine MD/PhD Program. R.L.S. is the recipient of a Ruth L. Kirschstein National Research Service Award from the National Institute of General Medical Sciences. P.O'D. holds a National Science Foundation Postdoctoral Fellowship in Biological Informatics. Funding to pay the Open Access publication charges for this article was provided by GM22854 (to D.S.).

Conflict of interest statement. None declared.

REFERENCES

- Ibba, M. and Söll, D. (2004) Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes Dev.*, **18**, 731–738.
- Hatfield, D.L. and Gladyshev, V.N. (2002) How selenium has altered our understanding of the genetic code. *Mol. Cell Biol.*, **22**, 3565–3576.
- Böck, A., Thanbichler, M., Rother, M. and Resch, A. (2005) Selenocysteine. In Ibba, M., Francklyn, C.S. and Cusack, S. (eds),

- Aminoacyl-tRNA Synthetases*. Landes Bioscience, Georgetown, TX, pp. 320–327.
- Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M.A. and Böck, A. (1988) Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature*, **331**, 723–725.
- Bilokapic, S., Korencic, D., Söll, D. and Weygand-Durasevic, I. (2004) The unusual methanogenic seryl-tRNA synthetase recognizes tRNA^{Ser} species from all three kingdoms of life. *Eur. J. Biochem.*, **271**, 694–702.
- Kaiser, J.T., Gromadski, K., Rother, M., Engelhardt, H., Rodnina, M.V. and Wahl, M.C. (2005) Structural and functional investigation of a putative archaeal selenocysteine synthase. *Biochemistry*, **44**, 13315–13327.
- Ohama, T., Yang, D.C.H. and Hatfield, D.L. (1994) Selenocysteine tRNA and serine tRNA are aminoacylated by the same synthetase, but may manifest different identities with respect to the long extra arm. *Arch. Biochem. Biophys.*, **315**, 293–301.
- Mäenpää, P.H. and Bernfield, M.R. (1970) A specific hepatic transfer RNA for phosphoserine. *Proc. Natl Acad. Sci. USA*, **67**, 688–695.
- Sharp, S.J. and Stewart, T.S. (1977) The characterization of phosphoserine tRNA from lactating bovine mammary gland. *Nucleic Acids Res.*, **4**, 2123–2136.
- Carlson, B.A., Xu, X.M., Kryukov, G.V., Rao, M., Berry, M.J., Gladyshev, V.N. and Hatfield, D.L. (2004) Identification and characterization of phosphoserine tRNA^{(Ser)Sec} kinase. *Proc. Natl Acad. Sci. USA*, **101**, 12848–12853.
- Yuan, J., Palioura, S., Salazar, J.C., Su, D., O'Donoghue, P., Hohn, M.J., Cardoso, A.M., Whitman, W.B. and Söll, D. (2006) RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc. Natl Acad. Sci. USA*, **103**, 18923–18927.
- Xu, X.M., Carlson, B.A., Mix, H., Zhang, Y., Saira, K., Glass, R.S., Berry, M.J., Gladyshev, V.N. and Hatfield, D.L. (2007) Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol.*, **5**, e4. doi:10.1371/journal.pbio.0050004.
- Gelpi, C., Sontheimer, E.J. and Rodriguez-Sanchez, J.L. (1992) Autoantibodies against a serine tRNA-protein complex implicated in cotranslational selenocysteine insertion. *Proc. Natl Acad. Sci. USA*, **89**, 9739–9743.
- Costa, M., Rodriguez-Sanchez, J.L., Czaja, A.J. and Gelpi, C. (2000) Isolation and characterization of cDNA encoding the antigenic protein of the human tRNP^{(Ser)Sec} complex recognized by autoantibodies from patients with type-I autoimmune hepatitis. *Clin. Exp. Immunol.*, **121**, 364–374.
- Wies, I., Brunner, S., Henninger, J., Herkel, J., Kanzler, S., Meyer zum Büschenfelde, K.H. and Lohse, A.W. (2000) Identification of target antigen for SLA/LP autoantibodies in autoimmune hepatitis. *Lancet*, **355**, 1510–1515.
- Herkel, J., Manns, M.P. and Lohse, A.W. (2007) Selenocysteine, soluble liver antigen/liver-pancreas, and autoimmune hepatitis. *Hepatology*, **46**, 275–277.
- Sauerwald, A., Zhu, W., Major, T.A., Roy, H., Palioura, S., Jahn, D., Whitman, W.B., Yates, J.R.III, Ibbá, M. et al. (2005) RNA-dependent cysteine biosynthesis in archaea. *Science*, **307**, 1969–1972.
- Fukunaga, R. and Yokoyama, S. (2007) Structural insights into the second step of RNA-dependent cysteine biosynthesis in archaea: crystal structure of Sep-tRNA:Cys-tRNA synthase from *Archaeoglobus fulgidus*. *J. Mol. Biol.*, **370**, 128–141.
- Percudani, R. and Peracchi, A. (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep.*, **4**, 850–854.
- Eliot, A.C. and Kirsch, J.F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.*, **73**, 383–415.
- Lima, C.D. (2002) Analysis of the *E. coli* NifS CsdB protein at 2.0 Å reveals the structural basis for perselenide and persulfide intermediate formation. *J. Mol. Biol.*, **315**, 1199–1208.
- Mihara, H., Maeda, M., Fujii, T., Kurihara, T., Hata, Y. and Esaki, N. (1999) A nifS-like gene, csdB, encodes an *Escherichia coli* counterpart of mammalian selenocysteine lyase. Gene cloning, purification, characterization and preliminary x-ray crystallographic studies. *J. Biol. Chem.*, **274**, 14768–14772.

23. Cupp-Vickery, J.R., Urbina, H. and Vickery, L.E. (2003) Crystal structure of IscS, a cysteine desulfurase from *Escherichia coli*. *J. Mol. Biol.*, **330**, 1049–1059.
24. Weeks, C.M. and Miller, R. (1999) The design and implementation of SnB version 2.0. *J. Appl. Cryst.*, **32**, 120–124.
25. de La Fortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.
26. Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A.*, **47**(Pt 2), 110–119.
27. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.*, **54**, 905–921.
28. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
29. Oshikane, H., Sheppard, K., Fukai, S., Nakamura, Y., Ishitani, R., Numata, T., Sherrer, R.L., Feng, L., Schmitt, E. *et al.* (2006) Structural basis of RNA-dependent recruitment of glutamine to the genetic code. *Science*, **312**, 1950–1954.
30. Bullock, T.L., Uter, N., Nissan, T.A. and Perona, J.J. (2003) Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J. Mol. Biol.*, **328**, 395–408.
31. Russell, R.B. and Barton, G.J. (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, **14**, 309–323.
32. Roberts, E., Eargle, J., Wright, D. and Luthey-Schulten, Z. (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*, **7**, 382.
33. O'Donoghue, P. and Luthey-Schulten, Z. (2005) Evolutionary profiles derived from the QR factorization of multiple structural alignments gives an economy of information. *J. Mol. Biol.*, **346**, 875–894.
34. Felsenstein, J. (1989) PHYLIP - Phylogeny Interference Package (Version 3.2). *Cladistics*, **5**, 164–166.
35. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.*, **31**, 3497–3500.
36. Lacourciere, G.M., Levine, R.L. and Stadtman, T.C. (2002) Direct detection of potential selenium delivery proteins by using an *Escherichia coli* strain unable to incorporate selenium from selenite into proteins. *Proc. Natl Acad. Sci. USA*, **99**, 9150–9153.
37. Herkel, J., Heidrich, B., Nieraad, N., Wies, I., Rother, M. and Lohse, A.W. (2002) Fine specificity of autoantibodies to soluble liver antigen and liver/pancreas. *Hepatology*, **35**, 403–408.
38. Jansonius, J.N. (1998) Structure, evolution and action of vitamin B6-dependent enzymes. *Curr. Opin. Struct. Biol.*, **8**, 759–769.
39. Schneider, G., Kack, H. and Lindqvist, Y. (2000) The manifold of vitamin B6 dependent enzymes. *Structure*, **8**, R1–R6.
40. Yano, T., Kuramitsu, S., Tanase, S., Morino, Y., Hiromi, K. and Kagamiyama, H. (1991) The role of His143 in the catalytic mechanism of *Escherichia coli* aspartate aminotransferase. *J. Biol. Chem.*, **266**, 6079–6085.
41. Shindyalov, I.N. and Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, **11**, 739–747.
42. Nureki, O., Watanabe, K., Fukai, S., Ishii, R., Endo, Y., Hori, H. and Yokoyama, S. (2004) Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme. *Structure*, **12**, 593–602.
43. Christen, P. and Mehta, P.K. (2001) From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Chem. Rev.*, **1**, 436–447.
44. Mehta, P.K., Argos, P., Barbour, A.D. and Christen, P. (1999) Recognizing very distant sequence relationships among proteins by family profile analysis. *Proteins*, **35**, 387–400.
45. O'Donoghue, P. and Luthey-Schulten, Z. (2003) On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol. Mol. Biol. Rev.*, **67**, 550–573.
46. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
47. Andreeva, A., Howorth, D., Brenner, S.E., Hubbard, T.J., Chothia, C. and Murzin, A.G. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.*, **32**, D226–D229.

Critical Review

How an Obscure Archaeal Gene Inspired the Discovery of Selenocysteine Biosynthesis in Humans

Dan Su¹, Michael J. Hohn¹, Sotiria Palioura¹, R. Lynn Sherrer¹, Jing Yuan¹, Dieter Söll^{1,2} and Patrick O'Donoghue¹

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

²Department of Chemistry, Yale University, New Haven, CT, USA

Summary

Selenocysteine (Sec) is the 21st genetically encoded amino acid found in organisms from all three domains of life. Sec biosynthesis is unique in that it always proceeds from an aminoacyl-tRNA precursor. Even though Sec biosynthesis in bacteria was established almost two decades ago, only recently the pathway was elucidated in archaea and eukaryotes. While other aspects of Sec biology have been reviewed previously (Allmann and Krol, *Biochimie* 2006;88:1561–1571, Hatfield et al., *Prog Nucleic Acid Res Mol Biol* 2006;81:97–142, Squires and Berry, *IUBMB Life* 2008;60:232–235), here we review the biochemistry and evolution of Sec biosynthesis and coding and show how the knowledge of an archaeal cysteine biosynthesis pathway helped to uncover the route to Sec formation in archaea and eukaryotes. © 2008 IUBMB

IUBMB Life, 61(1): 35–39, 2009

Keywords selenocysteine; tRNA^{Sec}; SepSecS; PSTK; genetic code.

INTRODUCTION

Since its identification as an element, selenium was only famous for its toxicity to animals and plants. Yet in 1954 selenium was found to be required for formate dehydrogenase activity in bacteria (1) and its presence in glutathione peroxidase was demonstrated in 1973 (2). The occurrence of selenocysteine (Sec) as the selenium moiety in naturally occurring proteins was first discovered in 1976 (3). Subsequent work docu-

mented the necessity of Sec in additional bacterial and eukaryotic proteins (4). Later Sec was identified as the catalytic residue in the active site of glutathione peroxidase in 1978 (5). Without the extensive genomic data on hand today the authors already concluded that “it is highly unlikely that Se-Cys is incorporated specifically via normal transcription and translation” (5).

Although later it was shown that Sec corresponds to certain UGA codons (6), the final proof that Sec is a cotranslationally incorporated amino acid came when the tRNA^{Sec} in *Escherichia coli* was identified in 1988 (7). Many of the basic questions in this area (8) were addressed in the following years, including elucidation of Sec biosynthesis and encoding in bacteria. There are several special requirements for Sec encoding to occur: a PLP-dependent selenocysteine synthase (SelA) that catalyzes the conversion of Ser-tRNA^{Sec} to Sec-tRNA^{Sec} (9, 10) (Fig. 1B), an mRNA secondary structure element (SECIS) (11), an elongation factor (SelB) specific for Sec-tRNA^{Sec} (12, 13), and an enzyme (SelD) that produces selenophosphate (14). Even though the counterparts of SelB, SelD, and SECIS have all been found in archaea and eukaryotes (15, 16), the mechanism of Sec biosynthesis, however, remained unclear until 2006.

RNA-DEPENDENT CYSTEINE BIOSYNTHESIS IN ARCHAEA

Cysteine (Cys) biosynthesis is one of the critical metabolic processes. The two well-known Cys biosynthesis pathways are the sulfur assimilation pathway, which is the sole route of making cysteine in most bacteria, all known plants, and at least one fungus; and the trans-sulfuration pathway, which is present in most fungi, nematodes, and vertebrates. In both cases Cys is synthesized through multiple enzymatic steps, one of which is catalyzed by a pyridoxal-phosphate (PLP)-dependent enzyme. Following its synthesis, Cys is charged to tRNA^{Cys} by cys-

Received 22 July 2008; accepted 7 August 2008

Address correspondence to: Patrick O'Donoghue, Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, New Haven, CT 06520-8114, USA. Tel: +1-203-432-6205. E-mail: patrick.odonoghue@yale.edu (or) Dan Su, Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, New Haven, CT 06520-8114, USA. Tel: +1-203-432-6205. E-mail: d.su@yale.edu.

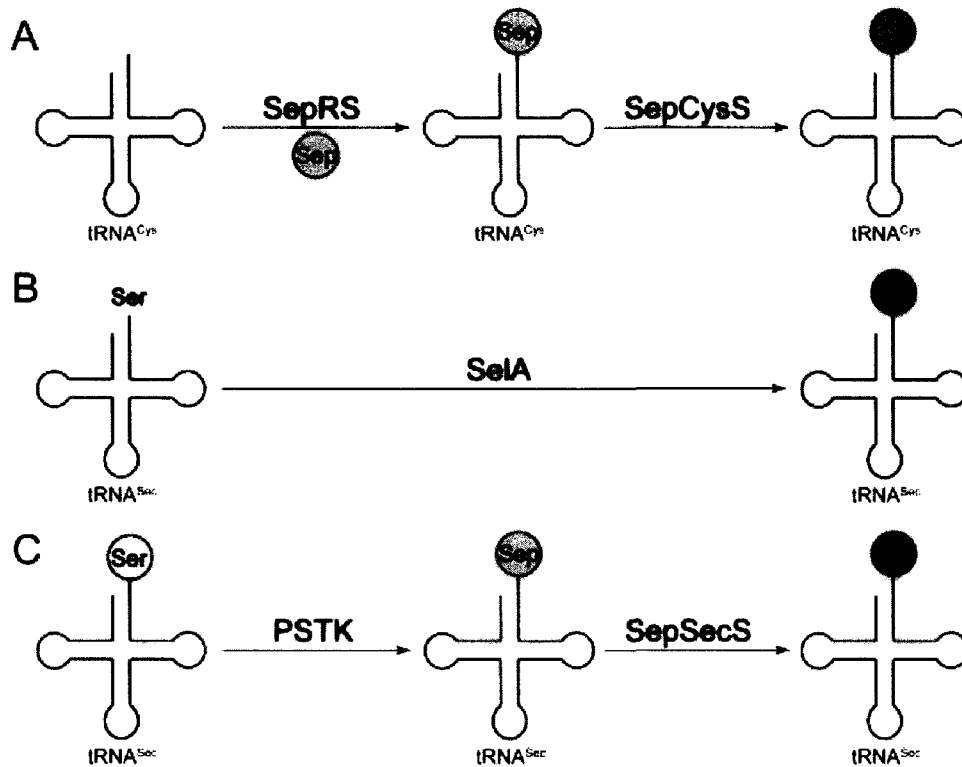


Figure 1. A: Cys-tRNA^{Cys} formation via the archaeal route, *O*-phosphoserine-tRNA synthetase (SepRS) forms the intermediate Sep-tRNA^{Cys} and the PLP-dependent enzyme Sep-tRNA:Cys-tRNA synthetase (SepCysS) catalyzes formation of the cognate pair Cys-tRNA^{Cys}. B: In bacterial selenocysteine biosynthesis, SslA catalyzes the direct formation of Sec-tRNA^{Sec} from Ser-tRNA^{Sec}, which is produced by seryl-tRNA synthetase (SerRS). C: Like Cys-tRNA formation in archaea, the archaeal and eukaryotic route for Sec-tRNA synthesis utilizes a Sep-tRNA^{Sec} intermediate produced by phosphoserine-tRNA^{Sec} kinase (PSTK) and the PLP-dependent enzyme Sep-tRNA:Sec-tRNA synthetase (SepSecS) gives Sec-tRNA^{Sec}. The initial substrate, Ser-tRNA^{Sec}, is generated by SerRS.

teinyI-tRNA synthetase (CysRS) to make Cys-tRNA^{Cys}, which then enters ribosomal translation.

When the genomes of some methanogenic archaea, including *Methanocaldococcus jannaschii* (17), *Methanothermobacter thermautotrophicus* (18), and *Methanopyrus kandleri* (19), were sequenced, *cysS* was not found. It was believed then that CysRS was present but not yet identified. Thus much effort was applied in finding this mysterious CysRS. Two proteins, ProRS (20, 21) and MJ1477 (22), were proposed to be the missing CysRS based on *in vitro* charging experiments and *in vivo* complementation of an *E. coli cysS* mutant strain. Later findings placed the complementation data into a different light because the strain retained low CysRS activity (23). In addition, some of these methanogens lacked recognizable cysteine biosynthesis genes (24).

In 2005, the mystery was finally solved. By purification coupled with enzymatic assays, our group identified two proteins in *Methanocaldococcus jannaschii* that work together to make Cys-tRNA^{Cys} (25). The first protein (a class II tRNA synthetase) ligates activated phosphoserine (Sep) onto tRNA^{Cys},

while the second protein is a PLP-dependent enzyme that converts Sep-tRNA^{Cys} to Cys-tRNA^{Cys} (Fig. 1A). These two proteins were thus named *O*-phosphoserine-tRNA synthetase (SepRS) and Sep-tRNA:Cys-tRNA synthetase (SepCysS). Deletion of SepRS in *Methanococcus maripaludis* yielded a cysteine auxotroph, indicating that the RNA-dependent Cys-tRNA^{Cys} synthesis pathway is the sole route for Cys biosynthesis in *Methanococcus maripaludis*.

SepCysS belongs to the fold Type I PLP-dependent enzyme family, more specifically, the cysteine desulfurase subfamily, based on sequence and structure comparisons (26, 27). With sodium sulfide present, SepCysS catalyzes the sulfuration of the phosphoserine moiety of Sep-tRNA^{Cys}. The natural sulfur donor in this reaction remains unknown. Cysteine desulfurase catalyzes the removal of S or Se from Cys or Sec, utilizing a conserved Cys residue to form a persulfide or perselenide, respectively. These S or Se relay mechanisms in related proteins may also operate in SepCysS, which contains several conserved Cys residues in the vicinity of the active site (26, 27), some which are required for its activity (unpublished data, (26, 27)).

SEP-tRNA IS AN INTERMEDIATE IN SEC BIOSYNTHESIS IN EUKARYOTES AND ARCHAEA

As early as 1964, Sep-tRNA was found in chicken liver (28). Later studies confirmed this finding (29) and suggested that Sep-tRNA was formed by phosphorylation of Ser-tRNA, not by direct acylation of phosphoserine onto tRNA (30). In 1981, RNA sequencing showed that the Sep moiety was carried by a putative suppressor tRNA^{Ser} (31), which was later identified as tRNA^{Sec} (32). Following the landmark discovery of *E. coli* tRNA^{Sec} in 1988 (7), it was speculated that Sep-tRNA might be an intermediate in Sec-tRNA^{Sec} formation (8). *In vitro* synthesis of glutathione peroxidase in mouse extracts suggested that this was the case (33). However, the successful identification and characterization of *E. coli* SelA, the selenocysteine synthase that catalyzes Sec-tRNA^{Sec} formation directly from Ser-tRNA^{Sec} (Fig. 1B), showed that the bacterial pathway does not involve Sep-tRNA (10). This result may have influenced the interpretation of another study of mammalian Sec-tRNA synthesis that concluded that the mammalian mechanism does not proceed through Sep-tRNA, and is similar to that in *E. coli* (34).

During the next decade little progress was made concerning Sec-tRNA^{Sec} formation in eukaryotes and archaea. A SelA-like protein from *M. jannaschii* was investigated and even crystallized, but then shown to lack the hypothesized Sec synthase activity (35). As more genomic data became available, an *in silico* search first identified a kinase that was only present in Sec-decoding archaea (36). *In vitro* characterization of the mouse enzyme, now called phosphoseryl-tRNA^{Sec} kinase (PSTK), showed that it converted Ser-tRNA^{Sec} to Sep-tRNA^{Sec} in an ATP and Mg²⁺-dependent reaction (36). Yet the description of this important enzyme did not resolve mammalian Sec formation, but deferred to future studies (36). However, the Sep moiety was again implicated as an intermediate in Sec synthesis, when the two-step tRNA-dependent Cys biosynthesis pathway (Fig. 1A) in archaea was discovered (25); the need for Sep-tRNA^{Cys}, the required precursor for Cys-tRNA^{Cys} formation, suggested that a similar mechanism may exist for Sec-tRNA^{Sec} formation in archaea and eukaryotes (25).

Given the chemical similarity between Cys and Sec, the reaction from Sep-tRNA to Sec-tRNA seemed plausible, and a single PLP-dependent enzyme could catalyze such a reaction. The soluble liver antigen/liver pancreas antigen (SLA/LP) protein had long been suspected to be the missing Sec synthase. SLA/LP coprecipitated with Sec-tRNA^{Sec} by autoantibodies from autoimmune chronic hepatitis patients (37). Although by the year 2000, the cDNA sequence of SLA/LP was available (38, 39), workers in the field could only conclude that SLA/LP might be the Sec-tRNA synthase, a human analog of the translation factor SelB, or perhaps a selenium-tRNA protecting factor (37–39). In the following year, using a fold recognition algorithm, a bioinformatics analysis was able to correctly classify SLA/LP as a PLP-dependent enzyme and the group predicted its involvement in Sec-tRNA synthesis (40).

Archaeal cysteine biosynthesis provided the conceptual template for Sec-tRNA formation in archaea and eukaryotes, and it quickly became clear that these two domains of life share a homologous mechanism for Sec-tRNA formation that was inherited from their distant common ancestor. Using an *in vivo* complementation assay, we showed that SLA/LP is responsible for Sec-tRNA^{Sec} formation (41). When grown under anaerobic atmosphere *E. coli* produces the selenoprotein formate dehydrogenase, which reduces benzyl viologen in the medium in the presence of formate. Reduced benzyl viologen is clearly visible due to its purple color and can be used to monitor selenoprotein biosynthesis inside *E. coli* cells. We complemented a *selA* deletion strain of *E. coli* with archaeal PSTK, with archaeal SLA/LP, or with both proteins, and monitored the formate dehydrogenase activity. Only when both PSTK and SLA/LP were present, the formate dehydrogenase activity was recovered from the $\Delta selA$ *E. coli* cells, indicating that PSTK and SLA/LP constituted a two-step pathway to make selenocysteine. Metabolic labeling of the same cells with ⁷⁵Se supported the result, showing ⁷⁵Se incorporation into proteins only in the presence of both PSTK and SLA/LP. These data established that the archaeal SLA/LP, which was appropriately renamed Sep-tRNA: Sec-tRNA synthase (SepSecS), converts Sep-tRNA^{Sec} to Sec-tRNA^{Sec} (41). Having established the archaeal system, we then tested the mammalian SLA/LP protein. The human SLA/LP displayed similar results in both the benzyl viologen and the ⁷⁵Se labeling assays, showing that mammals and archaea share a common mechanism for Sec biosynthesis. *In vitro* assays, both in our laboratory (41) and independently by another group (42), showed that SepSecS is responsible for the final step of Sec-tRNA^{Sec} formation.

Recently it was established that SepSecS is in the same enzyme family as SepCysS (27, 41, 42) through phylogenetic analyses based on both its sequence and structure from mouse (43) and archaea (41), revealing an evolutionary linkage between the tRNA-dependent Cys and Sec formation.

SEC ENCODING WAS AN ANCIENT BUT “LATE” ADDITION TO THE GENETIC CODE

The discovery of Sec argues against components of two of Crick’s famous hypotheses: the wobble hypothesis (44) and the “frozen accident” theory (45). Crick maintained that the wobble hypothesis “does not permit UGA to code for any amino acid other than cysteine or tryptophan.” While Cys and Sec are chemically similar, they are not universally interchangeable. The substitution of Cys for Sec at a critical active site location can have a severe effect on catalytic efficiency (46, 47). The “frozen accident” theory states that “the code is universal because at the present time any change would be lethal, or at least very strongly selected against.” If Sec encoding emerged before the code was “frozen,” then the code obviously has never been “frozen” because it was shown that Sec decoding

was later lost in certain organisms (48); if Sec encoding emerged after the “freezing” of the genetic code, then the code was not “frozen” *per se*. Our understanding of Sec evolution based on phylogenetic data (27, 41, 49–52), indicates that the genetic code expanded to include Sec at the time of the last universal common ancestor and was then independently lost in many lineages during evolution.

Among the 22 known ribosome-incorporated amino acids, including pyrrolysine, only Sec lacks its own aminoacyl-tRNA synthase. A SecRS has apparently never evolved (and perhaps never will) because typical protein substrate recognition, which depends on distinct differences between the shape and charge of cognate versus noncognate substrates, cannot differentiate Cys from Sec (53, 54). The indirect route for Sec-tRNA formation evolved, therefore, to ensure that critical Sec codons were read with an exceedingly high fidelity by eliminating or minimizing the possibility that Cys-tRNA^{Sec} would be formed. Similarly, since CysRS cannot discriminate Sec from Cys and can easily form Sec-tRNA^{Cys}, the presence of excessive free selenocysteine might be unfavorable for positions in the peptide chain that require Cys instead of Sec. The Sec codon has a dual meaning (also functioning as a stop codon), so the special translational components (SelB, SECIS) evolved to ensure that translation did not stop at Sec codons and that it terminated properly in response to UGA stop codons.

Although Sec encoding is ancient, the evolution of the genetic code is yet an older event; the biochemical and phylogenetic data indicate that Sec was a late addition to the code. For the reasons mentioned above, it seems clear that Sec was introduced only after UGA had acquired the termination meaning and possibly after Cys was added to the code. The additional coding machinery for Sec hints that Sec decoding represents a complication or addition to a pre-existing, simpler coding process—such as that applied to the 20 canonical amino acids.

ACKNOWLEDGEMENTS

The authors' laboratory is supported by grants (to D.S.) from the National Institute of General Medical Sciences, the Department of Energy, and the National Science Foundation.

REFERENCES

- Pinsent, J. (1954) The need for selenite and molybdate in the formation of formic dehydrogenase by members of the coli-aerogenes group of bacteria. *Biochem. J.* **57**, 10–16.
- Flohé, L., Günzler, W. A., and Schock, H. H. (1973) Glutathione peroxidase: a selenoenzyme. *FEBS Lett.* **32**, 132–134.
- Cone, J. E., Del Rio, 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. USA* **73**, 2659–2663.
- Stadtman, T. C. (1980) Selenium-dependent enzymes. *Annu. Rev. Biochem.* **49**, 93–110.
- Forstrom, J. W., Zakowski, J. J., and Tappel, A. L. (1978) Identification of the catalytic site of rat liver glutathione peroxidase as selenocysteine. *Biochemistry* **17**, 2639–2644.
- Zinoni, F., Birkmann, A., Stadtman, T. C., and Böck, A. (1986) Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **83**, 4650–4654.
- Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M. A., and Böck, A. (1988) Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* **331**, 723–725.
- Söll, D. (1988) Genetic code: enter a new amino acid. *Nature* **331**, 662–663.
- Forchhammer, K., Leinfelder, W., Boesmiller, K., Veprek, B., and Böck, A. (1991) Selenocysteine synthase from *Escherichia coli*. Nucleotide sequence of the gene (*selA*) and purification of the protein. *J. Biol. Chem.* **266**, 6318–6323.
- Forchhammer, K. and Böck, A. (1991) Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. *J. Biol. Chem.* **266**, 6324–6328.
- Berry, M. J., Banu, L., Chen, Y. Y., Mandel, S. J., Kieffer, J. D., Harney, J. W., and Larsen, P. R. (1991) Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. *Nature* **353**, 273–276.
- Forchhammer, K., Leinfelder, W., and Böck, A. (1989) Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. *Nature* **342**, 453–456.
- Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P., and Krol, A. (2000) Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J.* **19**, 4796–4805.
- Leinfelder, W., Forchhammer, K., Veprek, B., Zehelein, E., and Böck, A. (1990) *In vitro* synthesis of selenocysteinyl-tRNA^{UCA} from seryl-tRNA^{UCA}: involvement and characterization of the *selD* gene product. *Proc. Natl. Acad. Sci. USA* **87**, 543–547.
- Allmang, C. and Krol, A. (2006) Selenoprotein synthesis: UGA does not end the story. *Biochimie* **88**, 1561–1571.
- Hatfield, D. L., Carlson, B. A., Xu, X. M., Mix, H., and Gladyshev, V. N. (2006) Selenocysteine incorporation machinery and the role of selenoproteins in development and health. *Prog. Nucleic Acid Res. Mol. Biol.* **81**, 97–142.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S., and Venter, J. C. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073.
- Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrovski, S., Church, G. M., Daniels, C. J., Mao, J.-I., Rice, P., Nölling, J., and Reeve, J. N. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135–7155.
- Slesarev, A. I., Mezhevaya, K. V., Makarova, K. S., Polushin, N. N., Shcherbinina, O. V., Shakhova, V. V., Belova, G. I., Aravind, L., Natale, D. A., Rogozin, I. B., Tatusov, R. L., Wolf, Y. I., Stetter, K. O., Malykh, A. G., Koonin, E. V., and Kozyavkin, S. A. (2002) The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl. Acad. Sci. USA* **99**, 4644–4649.

20. Stathopoulos, C., Li, T., Longman, R., Vothknecht, U. C., Becker, H. D., Ibba, M., and Söll, D. (2000) One polypeptide with two aminoacyl-tRNA synthetase activities. *Science* **287**, 479–482.
21. Lipman, R. S., Sowers, K. R., and Hou, Y. M. (2000) Synthesis of cysteinyl-tRNA^{Cys} by a genome that lacks the normal cysteine-tRNA synthetase. *Biochemistry* **39**, 7792–7798.
22. Fabrega, C., Farrow, M. A., Mukhopadhyay, B., de Crecy-Lagard, V., Ortiz, A. R., and Schimmel, P. (2001) An aminoacyl tRNA synthetase whose sequence fits into neither of the two known classes. *Nature* **411**, 110–114.
23. Ruan, B., Nakano, H., Tanaka, M., Mills, J. A., DeVito, J. A., Min, B., Low, K. B., Battista, J. R., and Söll, D. (2004) Cysteinyl-tRNA^{Cys} formation in *Methanocaldococcus jannaschii*: the mechanism is still unknown. *J. Bacteriol.* **186**, 8–14.
24. Ambrogelly, A., Kamtekar, S., Sauerwald, A., Ruan, B., Tumbula-Hansen, D., Kennedy, D., Ahel, I., and Söll, D. (2004) Cys-tRNA^{Cys} formation and cysteine biosynthesis in methanogenic archaea: two faces of the same problem? *Cell Mol. Life Sci.* **61**, 2437–2445.
25. Sauerwald, A., Zhu, W., Major, T. A., Roy, H., Palioura, S., Jahn, D., Whitman, W. B., Yates, J. R. III, Ibba, M., and Söll, D. (2005) RNA-dependent cysteine biosynthesis in archaea. *Science* **307**, 1969–1972.
26. Fukunaga, R. and Yokoyama, S. (2007) Structural insights into the second step of RNA-dependent cysteine biosynthesis in archaea: crystal structure of Sep-tRNA:Cys-tRNA synthase from *Archaeoglobus fulgidus*. *J. Mol. Biol.* **370**, 128–141.
27. Araiso, Y., Palioura, S., Ishitani, R., Sherrer, R. L., O'Donoghue, P., Yuan, J., Oshikane, H., Domae, N., Defranco, J., Söll, D., and Nureki, O. (2008) Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic Acids Res.* **36**, 1187–1199.
28. Carlsen, E. N., Trelle, G. J., and Schjeide, O. A. (1964) Transfer Ribonucleic Acids. *Nature* **202**, 984–986.
29. Mäenpää, P. H. and Bernfield, M. R. (1970) A specific hepatic transfer RNA for phosphoserine. *Proc. Natl. Acad. Sci. USA* **67**, 688–695.
30. Sharp, S. J. and Stewart, T. S. (1977) The characterization of phosphoseryl tRNA from lactating bovine mammary gland. *Nucleic Acids Res.* **4**, 2123–2136.
31. Diamond, A., Dudock, B., and Hatfield, D. (1981) Structure and properties of a bovine liver UGA suppressor serine tRNA with a tryptophan anticodon. *Cell* **25**, 497–506.
32. Lee, B. J., Worland, P. J., Davis, J. N., Stadtman, T. C., and Hatfield, D. L. (1989) Identification of a selenocysteyl-tRNA^{Ser} in mammalian cells that recognizes the nonsense codon, UGA. *J. Biol. Chem.* **264**, 9724–9727.
33. Mizutani, T. and Hitaka, T. (1988) The conversion of phosphoserine residues to selenocysteine residues on an opal suppressor tRNA and casein. *FEBS Lett.* **232**, 243–248.
34. Mizutani, T., Kurata, H., and Yamada, K. (1991) Study of mammalian selenocysteyl-tRNA synthesis with [⁷⁵Se]HSe⁻. *FEBS Lett.* **289**, 59–63.
35. Kaiser, J. T., Gromadski, K., Rother, M., Engelhardt, H., Rodnina, M. V., and Wahl, M. C. (2005) Structural and functional investigation of a putative archaeal selenocysteine synthase. *Biochemistry* **44**, 13315–13327.
36. Carlson, B. A., Xu, X. M., Kryukov, G. V., Rao, M., Berry, M. J., Gladyshev, V. N., and Hatfield, D. L. (2004) Identification and characterization of phosphoseryl-tRNA^{Ser} kinase. *Proc. Natl. Acad. Sci. USA* **101**, 12848–12853.
37. Gelpi, C., Sontheimer, E. J., and Rodriguez-Sanchez, J. L. (1992) Autoantibodies against a serine tRNA-protein complex implicated in cotranslational selenocysteine insertion. *Proc. Natl. Acad. Sci. USA* **89**, 9739–9743.
38. Wies, I., Brunner, S., Henninger, J., Herkel, J., Kanzler, S., Meyer zum Buschenfelde, K. H., and Lohse, A. W. (2000) Identification of target antigen for SLA/LP autoantibodies in autoimmune hepatitis. *Lancet* **355**, 1510–1515.
39. Costa, M., Rodriguez-Sanchez, J. L., Czaja, A. J., and Gelpi, C. (2000) Isolation and characterization of cDNA encoding the antigenic protein of the human tRNP^{(Ser)Sec} complex recognized by autoantibodies from patients with type-1 autoimmune hepatitis. *Clin. Exp. Immunol.* **121**, 364–374.
40. Kernebeck, T., Lohse, A. W., and Grötzinger, J. (2001) A bioinformatical approach suggests the function of the autoimmune hepatitis target antigen soluble liver antigen/liver pancreas. *Hepatology* **34**, 230–233.
41. Yuan, J., Palioura, S., Salazar, J. C., Su, D., O'Donoghue, P., Hohn, M. J., Cardoso, A. M., Whitman, W. B., and Söll, D. (2006) RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc. Natl. Acad. Sci. USA* **103**, 18923–18927.
42. Xu, X. M., Carlson, B. A., Mix, H., Zhang, Y., Saira, K., Glass, R. S., Berry, M. J., Gladyshev, V. N., and Hatfield, D. L. (2007) Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol.* **5**, e4.
43. Ganichkin, O. M., Xu, X. M., Carlson, B. A., Mix, H., Hatfield, D. L., Gladyshev, V. N., and Wahl, M. C. (2008) Structure and catalytic mechanism of eukaryotic selenocysteine synthase. *J. Biol. Chem.* **283**, 5849–5865.
44. Crick, F. H. (1966) Codon—anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548–555.
45. Crick, F. H. (1968) The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379.
46. Su, D. and Gladyshev, V. N. (2004) Alternative splicing involving the thioredoxin reductase module in mammals: a glutaredoxin-containing thioredoxin reductase 1. *Biochemistry* **43**, 12177–12188.
47. Kim, H. Y. and Gladyshev, V. N. (2005) Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol.* **3**, e375.
48. Zhang, Y., Romero, H., Salinas, G., and Gladyshev, V. N. (2006) Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome Biol.* **7**, R94.
49. Böck, A., Forchhammer, K., Heider, J., and Baron, C. (1991) Selenoprotein synthesis: an expansion of the genetic code. *Trends Biochem. Sci.* **16**, 463–467.
50. Keeling, P. J., Fast, N. M., and McFadden, G. I. (1998) Evolutionary relationship between translation initiation factor eIF-2 γ and selenocysteine-specific elongation factor SELB: change of function in translation factors. *J. Mol. Evol.* **47**, 649–655.
51. Romero, H., Zhang, Y., Gladyshev, V. N., and Salinas, G. (2005) Evolution of selenium utilization traits. *Genome Biol.* **6**, R66.
52. Sherrer, R. L., O'Donoghue, P., and Söll, D. (2008) Characterization and evolutionary history of an archaeal kinase involved in selenocysteinyl-tRNA formation. *Nucleic Acids Res.* **36**, 1247–1259.
53. Young, P. A. and Kaiser, II. (1975) Aminoacylation of *Escherichia coli* cysteine tRNA by selenocysteine. *Arch. Biochem. Biophys.* **171**, 483–489.
54. Shrift, A., Bechard, D., and Harcup, C. (1976) Utilization of Selenocysteine by a Cysteinyl-tRNA Synthetase from *Phaseolus aureus*. *Plant Physiol.* **58**, 248–252.

The canonical pathway for selenocysteine insertion is dispensable in *Trypanosomes*

Eric Aeby^a, Sotiria Palioura^b, Mascha Pusnik^a, Janine Marazzi^a, Allyson Lieberman^b, Elisabetta Ullu^c, Dieter Söll^{b,d,1}, and André Schneider^{a,1}

^aDepartment of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, CH-3012 Bern, Switzerland; Departments of ^bMolecular Biophysics and Biochemistry and ^cChemistry, Yale University, New Haven, CT 06520-8114; and ^dDepartments of Internal Medicine and Cell Biology, Yale University School of Medicine, New Haven, CT 06536-0812

Contributed by Dieter Söll, February 11, 2009 (sent for review January 18, 2009)

The micronutrient selenium is found in proteins as selenocysteine (Sec), the 21st amino acid cotranslationally inserted in response to a UGA codon. In vitro studies in archaea and mouse showed that Sec-tRNA^{Sec} formation is a 3-step process starting with serylation of tRNA^{Sec} by seryl-tRNA synthetase (SerRS), phosphorylation of serine to form phosphoserine (Sep)-tRNA^{Sec} by phosphoseryl-tRNA^{Sec} kinase (PSTK), and conversion to Sec-tRNA^{Sec} by Sep-tRNA:Sec-tRNA synthase (SepSecS). However, a complete study of eukaryotic selenoprotein synthesis has been lacking. Here, we present an analysis of Sec-tRNA^{Sec} formation in the parasitic protozoan *Trypanosoma brucei* in vivo. Null mutants of either PSTK or SepSecS abolished selenoprotein synthesis, demonstrating the essentiality of both enzymes for Sec-tRNA^{Sec} formation. Growth of the 2 knockout strains was not impaired; thus, unlike mammals, trypanosomes do not require selenoproteins for viability. Analysis of conditional RNAi strains showed that SerRS, selenophosphate synthase, and the Sec-specific elongation factor, EFSec, are also essential for selenoprotein synthesis. These results with *T. brucei* imply that eukaryotes have a single pathway of Sec-tRNA^{Sec} synthesis that requires Sep-tRNA^{Sec} as an intermediate.

phosphoseryl-tRNA^{Sec} kinase | selenocysteine tRNA |
Sep-tRNA:Sec-tRNA synthase | *Trypanosoma brucei* | selenoprotein

Selenium is an essential dietary trace element. It is present in proteins as selenocysteine (Sec), a cotranslationally-inserted amino acid encoded by UGA. Sec is not attached directly to tRNA^{Sec}, but is formed by the tRNA-dependent conversion of serine (reviewed in refs. 1–4). In the first step tRNA^{Sec} is misacylated by seryl-tRNA synthetase (SerRS). The subsequent conversion to Sec proceeds by 2 different pathways in nature. In bacteria Ser-tRNA^{Sec} is directly transformed to Sec-tRNA^{Sec} in a pyridoxal-5'-phosphate-dependent reaction by Sec synthase, the *selA* gene product (5). Archaea and eukaryotes require an additional step, the formation of the intermediate phosphoserine (Sep) by phosphoseryl-tRNA^{Sec} kinase (PSTK). The resultant Sep-tRNA^{Sec} is then converted into the Sec-tRNA^{Sec} by the pyridoxal-5'-phosphate-dependent enzyme Sep-tRNA:Sec-tRNA synthase (SepSecS) (6, 7). The Se donor for this reaction, selenophosphate, is synthesized by selenophosphate synthase (SPS2) (8). Then EFSec, the tRNA^{Sec}-specific elongation factor, carries the Sec-tRNA^{Sec} to the ribosome (9) where a translational recoding process allows UGA to be read as Sec (Fig. 1). Selenoproteins are found in organisms from all 3 domains of life. Humans have 25 selenoproteins, many of them are essential for organismal viability (10). Some selenoproteins are predicted to be redox proteins containing catalytic Sec residues (e.g., glutathione peroxidase or thioredoxin reductase).

Most of our knowledge of eukaryotic Sec-tRNA^{Sec} formation comes from in vitro reconstitution experiments using components of mammalian cells (7). In vivo formation of eukaryotic Sec-tRNA^{Sec} has been addressed in a study that showed that RNAi-mediated ablation of SepSecS in mammalian cells did not completely abolish selenoprotein expression (11). Thus, the

existence of an alternative SepSecS-independent pathway for Sec-tRNA^{Sec} formation could not be excluded. Moreover, the in vivo role of PSTK has not yet been analyzed.

Here, we present a comprehensive analysis of the in vivo role of the 5 major proteins required for eukaryotic Sec-tRNA^{Sec} formation and function. The study was done in the insect form of the parasitic protozoan *Trypanosoma brucei* as double allelic KO cell lines can easily be produced in this system by homologous recombination-directed gene replacements (12). Because *T. brucei* is diploid, double KO cell lines are null mutants in both alleles. Furthermore, RNAi-based methods for highly efficient inducible ablation of proteins are also available (13). Moreover, *T. brucei* is an excellent model for exploring eukaryotic diversity. It represents a different branch of the eukaryotic evolutionary tree than the phylogenetically more closely related classical model organisms (e.g., mouse, *Drosophila*, *Caenorhabditis elegans*, and yeast) (14).

Results

***T. brucei* Components Involved in Selenoprotein Formation.** A bioinformatic analysis of the *T. brucei* genome predicts 3 selenoproteins. They include distant homologs of mammalian SelK and SelT and a selenoprotein, termed SelTryp, that is specific for the kinetoplastid line (15). Moreover, *in silico* screens by several groups have identified Tb-tRNA^{Sec}, Tb-SerRS, Tb-PSTK, Tb-SepSecS, Tb-SPS2, and Tb-EFSec, the trypanosomal orthologues of essentially all major components of the Sec-inserting system (16). However, of these only Tb-SerRS, Tb-tRNA^{Sec} (16–18), and Tb-SPS2 (19) have been subject to preliminary experimental analyses.

To analyze the Sec-tRNA^{Sec} formation pathway in vivo and establish its physiological importance for *T. brucei* we used RNAi cell lines allowing inducible ablation of Tb-SerRS, Tb-SPS2, and Tb-EFSec. Moreover, we prepared KO cell lines that lack either Tb-PSTK or Tb-SepSecS, the 2 core components of the eukaryotic Sec-tRNA^{Sec} formation pathway (Fig. S1). All cell lines were analyzed for selenoprotein synthesis by labeling with radioactive ⁷⁵Se (Fig. 2). The Tb-SerRS-RNAi cell line had been analyzed before by other methods, and it was shown that Tb-SerRS activity is required to charge the Tb-tRNA^{Sec} with serine (18). Labeling of uninduced Tb-SerRS-RNAi cells with ⁷⁵Se and subsequent analysis by Tris-Tricine polyacrylamide gels revealed 3 bands whose molecular mass are consistent with the 3 predicted trypanosomal selenoproteins SelK, SelT, and SelTryp (Fig. 2A). SelTryp migrated in some experiments slightly faster

Author contributions: E.A., S.P., D.S., and A.S. designed research; E.A., S.P., M.P., J.M., A.L., and E.U. performed research; E.A., S.P., M.P., D.S., and A.S. analyzed data; and E.A., S.P., D.S., and A.S. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence may be addressed. E-mail: dieter.soll@yale.edu or andre.schneider@bc.unibe.ch.

This article contains supporting information online at www.pnas.org/cgi/content/full/0901575106/DCSupplemental.

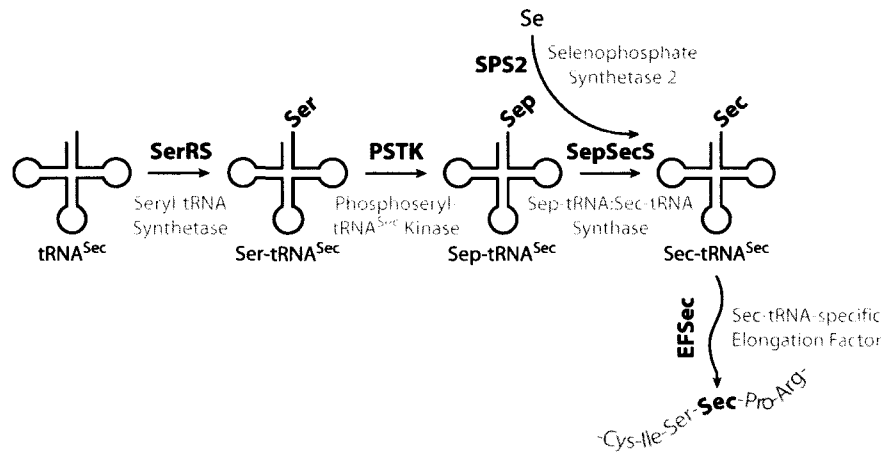


Fig. 1. tRNA^{Sec}-dependent amino acid transformations leading to Sec in eukaryotes as elucidated by using recombinantly produced mammalian components (6, 7).

than expected (Fig. 2B and C). We do not know why but the simplest explanation is proteolytic clipping during sample preparation, which could also account for the simultaneous presence of both bands as seen in Fig. 2B, lane 2d.

Induction of Tb-SerRS-RNAi causes, in line with the role of SerRS in serylation of tRNA^{Sec}, an efficient but incomplete reduction of the selenoprotein labeling (Fig. 2A). This finding is in contrast to the Tb-PSTK KO and Tb-SepSecS KO cell lines in which the labeling of the 3 selenoproteins was abolished (Fig. 2B). Moreover, tetracycline-inducible ectopic expression of Tb-

SepSecS in the Tb-SepSecS KO cell line restored ⁷⁵Se labeling of all 3 proteins. In the case of SelTryp and SelT the restoration was to wild-type level, whereas in the case of SelK the complementation was not quantitative. These experiments show that both Tb-PSTK and Tb-SepSecS are indispensable for selenoprotein synthesis.

Next, we isolated and analyzed aminoacylated tRNAs from the different cell lines by acid urea polyacrylamide electrophoresis (Fig. 2E) (20). This technique allows the separation of the faster-migrating Sep-tRNA^{Sec} from the slower-migrating Ser-

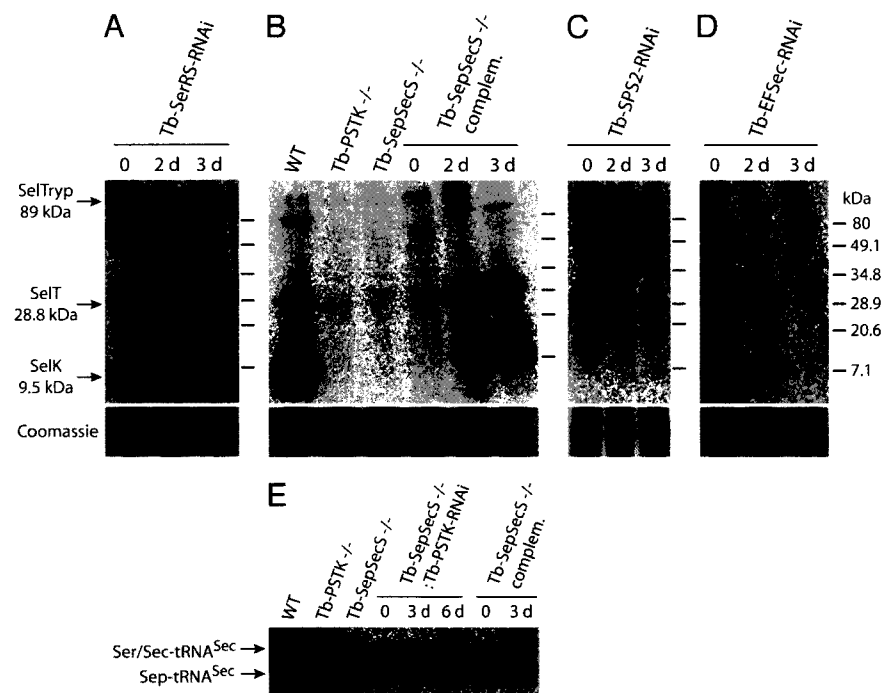


Fig. 2. Selenoprotein expression in various *T. brucei* cell lines. Expression was analyzed by ⁷⁵Se labeling of living cells and subsequent analysis of the labeled proteins by 10–20% polyacrylamide Tris-Tricine gels. The following cell lines were analyzed. (A) Uninduced and induced Tb-SerRS-RNAi cells. (B) WT 427 cells, Tb-PSTK KO cells (Tb-PSTK^{-/-}), Tb-SepSecS KO cells (Tb-SepSecS^{-/-}), and a Tb-SepSecS KO cell line allowing inducible ectopic expression of Tb-SepSecS (Tb-SepSecS^{-/-} complem.). (C) Uninduced and induced Tb-SPS2-RNAi cells. (D) Uninduced and induced Tb-EFSec-RNAi cells. For the RNAi and the complemented Tb-SepSecS^{-/-} cell lines days of induction (d) by tetracycline are indicated. The putative identity of the 3 labeled selenoproteins and their molecular mass as predicted *in silico* are indicated on the left. Molecular mass markers are indicated on the right. Segments of the tubulin region (~50–80 kDa) of the corresponding Coomassie-stained gels are shown as loading controls. (E) Total RNA isolated for the cell lines analyzed in B and a Tb-SepSecS KO line capable of RNAi-mediated ablation of Tb-PSTK (Tb-SepSecS^{-/-}:Tb-PSTK RNAi) was separated on a long acidic urea gel and analyzed for the presence of the different forms of tRNA^{Sec} by Northern analysis (18). The positions of the Sep-tRNA^{Sec} and the comigrating Ser/Sec-tRNA^{Sec} are indicated.

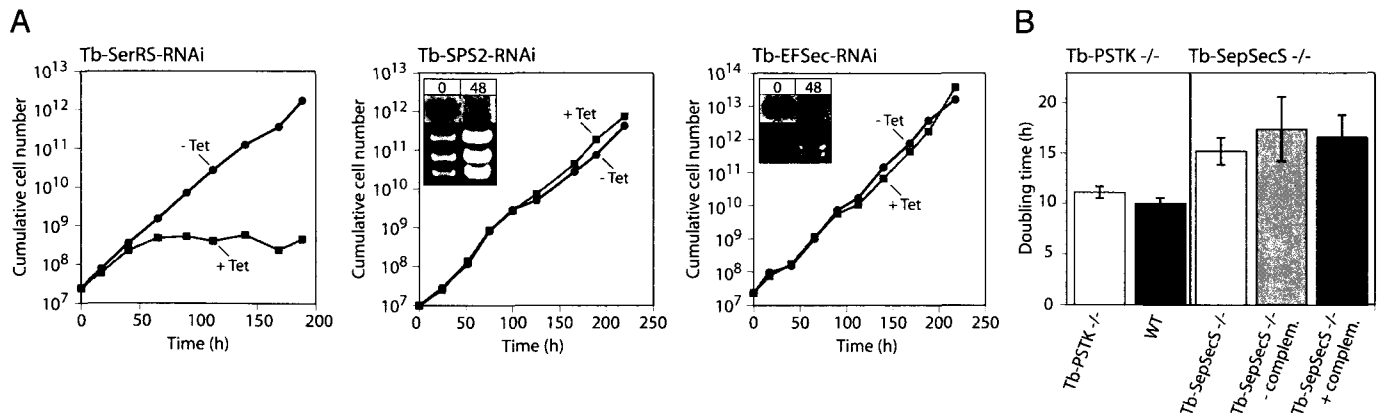


Fig. 3. Trypanosomal selenoproteins are not essential. (A) Representative growth curves in standard culture medium SDM-79 of uninduced and induced (–Tet, +Tet) clonal RNAi cell lines for Tb-SerRS (redrawn from ref. 18), Tb-SPS2, and Tb-EFSec. The Northern blots verifying mRNA ablation are indicated for the cell lines not showing a growth arrest. (B) (Left) Comparisons of doubling times of the Tb-PSTK KO cell line (Tb-PSTK^{-/-}) and the parent WT strain 427 (WT). (Right) Comparisons of doubling times of the Tb-SepSecS KO cell line (Tb-SepSecS^{-/-}) and the corresponding cell line allowing tetracycline-dependent ectopic Tb-SepSecS expression (SepSec^{-/-} plus or minus complem.). The parent cell line for the Tb-SepSecS KO strain is *T. brucei* 29-13 that has a longer doubling time than *T. brucei* 427. Standard errors ($n = 6$) are indicated.

tRNA^{Sec} and Sec-tRNA^{Sec} species. Distinction between the latter 2 species is not possible as they comigrate. The Northern blot analysis in Fig. 2E shows 2 tRNA^{Sec} bands in wild-type cells. The lower one corresponds to Sep-tRNA^{Sec} and the upper corresponds to Sec-tRNA^{Sec} (as inferred from the pattern observed in the PSTK and SepSecS KO cell lines). In the Tb-PSTK KO cell line only the upper band is detected, which most likely represents the Ser-tRNA^{Sec} form, because this species is expected to accumulate in the absence of PSTK activity (see Fig. 1). In the Tb-SepSecS-KO cell line, however, only the Sep-tRNA^{Sec} band is detected. Moreover, if Tb-PSTK is depleted by inducible RNAi in the Tb-SepSecS-KO cell line, a time-dependent partial recovery of the upper band is observed, which indicates that Ser-tRNA^{Sec} accumulates as would be expected in the absence of both PSTK and SepSecS. Moreover, the upper band, which is absent in the KO cells, reappears if the KO cell line is complemented by inducible ectopic expression of Tb-SepSecS. Thus, the accumulation of distinct intermediates of the Sec-tRNA^{Sec} synthesis pathway in the 2 KO cell lines together with the ⁷⁵Se-labeling experiments show that the Tb-PSTK and Tb-SepSecS act sequentially in the indicated order (Fig. 1).

SPS2 generates selenophosphate, the Se donor required by SepSecS. Mammalian SPS2, which itself is a selenoprotein, is essential for selenoprotein synthesis *in vivo* (8). Tb-SPS2 an SPS2 orthologue capable of complementing *SelA*-deficient *Escherichia coli* has been identified in trypanosomatids (19). Like the *E. coli* orthologue, but in contrast to its mammalian counterpart, Tb-SPS2 is not a selenoprotein. Fig. 2C shows that RNAi-mediated ablation of the Tb-SPS2 severely impairs selenoprotein synthesis, indicating that SPS2 (with a cysteine in place of the Sec) is catalytically active in trypanosomes. Surprisingly, even before the induction of RNAi we reproducibly see labeling of SelTryp and SelT but not of SelK. It is possible that the SPS2-RNAi cell line is leaky and that even in the absence of tetracycline a fraction of the SPS2 mRNA is partially down-regulated. If SelK labeling is more sensitive to SPS2 levels than labeling of the other 2 trypanosomal selenoproteins its absence could be explained. Indeed selenium labeling of SelK shows a larger variation in the different cell lines than that of the other 2 selenoproteins, indicating that it might be quite sensitive to small changes of the labeling conditions (compare Fig. 2A, lane 0 with Fig. 2B, lane WT). However, without further experiments this explanation must remain speculative at present.

Sec-tRNA^{Sec} is an elongator tRNA; however, it does not interact with elongation factor 1a but requires its own elongation factor, EFSec. Consequently, ablation of Tb-EFSec by RNAi essentially abolishes selenoprotein synthesis (Fig. 2D).

Selenoprotein Synthesis in *T. brucei* Is Not Essential. Selenoproteins are found in all 3 domains, indicating that they were acquired early in the evolution of life (4). However, in all domains there are many species that then lost the Sec-inserting system (e.g., fungi and plants), which raises the question of whether organisms that possess a Sec-inserting system require selenoproteins for viability. Clearly, this is the case in mammals, because a tRNA^{Sec} KO mouse has an embryonic lethal phenotype (21). In other systems, however, this question has not been rigorously addressed. With our *T. brucei* cell lines described above we now can ask whether selenoproteins are essential for this unicellular eukaryote. Therefore, we analyzed the effects RNAi-mediated ablation of Tb-SerRS, Tb-SPS2, and Tb-EFSec (Fig. 3A) and the effects the complete lack of Tb-PSTK and Tb-SepSecS (Fig. 3B) have on *T. brucei* growth. Ablation of Tb-SerRS caused a growth arrest, which can be explained by the down-regulation of Ser-tRNA^{Ser} levels with concomitant reduction and cessation of protein synthesis (18). Growth of all other RNAi and KO cell lines was not or only marginally impaired. For the RNAi cell lines this is not very informative because of the possible residual activity of the ablated enzymes. However, the normal growth of the Tb-PSTK and Tb-SepSecS double KO cell lines clearly demonstrates that *T. brucei* does not depend on selenoproteins.

Conserved Sec-tRNA^{Sec} Formation in Eukaryotes and Archaea. When grown anaerobically *E. coli* produces selenium-dependent formate dehydrogenase (FDH). Its activity can easily be visualized on plates overlaid with benzyl viologen, because in the presence of formate active FDH reduces benzyl viologen to a blue substance (22). Using this assay it was shown that an *E. coli selA* deletion strain could be rescued by coexpression of archaeal PSTK and human or archaeal SepSecS (6). We extended these studies and showed that Tb-PSTK in combination with either endogenous *T. brucei* SepSecS or the corresponding archaeal or human enzymes reconstitute selenoprotein synthesis in an *E. coli selA* deletion strain (Fig. 4). Likewise, archaeal PSTK and Tb-SepSecS also restored FDH activity in the special *E. coli* strain. Taken together, these results underscore the highly

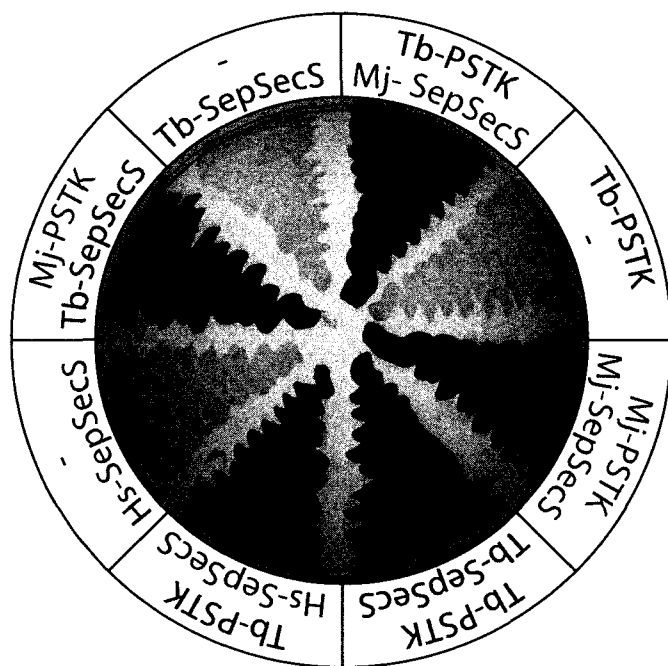


Fig. 4. Expression of SepSecS together with PSTK restores FDH activity in an anaerobically-grown *E. coli* *selA* deletion strain. The indicated genes of the following organisms were tested: *T. brucei* (Tb-PSTK, Tb-SepSecS), the Archaeon *Methanocaldococcus jannaschii* (Mj-PSTK, Mj-SepSecS), and humans (Hs-SepSecS).

conserved nature of the PSTK/SepSecS pathway of Sec-tRNA^{Sec} formation in Archaea and eukaryotes.

Discussion

In vitro experiments indicate that PSTK and SepSecS are the core components for Sec biosynthesis in eukaryotes (6, 7). However, the in vivo analysis of Sec-tRNA^{Sec} formation in eukaryotes has been hampered by the lack of a suitable system. Here, we have used *T. brucei* to produce KO cell lines for the trypanosomal orthologues of PSTK and SepSecS. Analysis of these cell lines has shown that formation of Sec-tRNA^{Sec} in living *T. brucei* requires the sequential action of Tb-PSTK and Tb-SepSecS. It remains unknown why eukaryotes use the additional phosphorylation step and thus require 2 enzymes, even though direct conversion of Ser-tRNA^{Sec} to Sec-tRNA^{Sec} is possible, as exemplified in bacteria (5).

Our results also show that no alternative pathway for Sec-tRNA^{Sec} formation exists that is independent of Tb-SepSecS and Tb-PSTK. Interestingly, RNAi-mediated ablation of SepSecS had only a marginal effect on mammalian selenoprotein synthesis (11). However, the results obtained in *T. brucei* might not be readily comparable with the study in mammalian cells for the following reason. In *T. brucei* selenoprotein synthesis was assayed in a double KO cell line completely devoid of Tb-SepSecS, whereas in mammalian cells the analysis might have been obscured by the fact that RNAi cannot be relied on to completely deplete protein levels. Thus, it is likely that even in mammalian cells, just as in *T. brucei*, only a single pathway for Sec-tRNA^{Sec} formation is operational.

Insect-stage *T. brucei* lacking selenoproteins grew as well as wild type in standard medium SDM-79, which shows that normal growth of *T. brucei* does not require selenoproteins and indicates that neither Tb-PSTK nor Tb-SepSecS have a second essential role that is unlinked to Sec-tRNA^{Sec} formation. The unimpaired growth of selenoprotein-lacking *T. brucei* was unexpected, because it had been reported that both insect- and bloodstream-

stage *T. brucei* cells were sensitive to nanomolar concentrations of auranofin, a compound suggested to inactivate selenoproteins (15). *T. brucei* is the causative agent of human sleeping sickness. Based on the reported auranofin experiments it was proposed that selenoprotein synthesis might be a novel target for development of an antiparasitic drug (15). Because auranofin interacts with selenol and thiol groups (23), we suggest that auranofin-induced cell death might have been caused by the drug's interaction with thiols. All of our experiments were carried out in insect-stage *T. brucei* cells; therefore it is still possible that Sec is essential in *T. brucei* bloodstream forms, but it seems unlikely because all 3 predicted selenoproteins are already expressed in the insect-stage form. Thus, the prospect of using the pathway of Sec or selenoprotein synthesis as drug targets against *T. brucei* is questionable. However, because *T. brucei* in its normal habitat adapts to situations and environments that are difficult to reproduce in the laboratory, selenoproteins may play a role under such conditions.

In summary, this work illustrates that *T. brucei* is an experimentally highly-tractable system for examining the in vivo formation of eukaryotic Sec-tRNA^{Sec}. Our analysis revealed the presence of a single pathway for Sec-tRNA^{Sec} synthesis involving SerRS, PSTK, and SepSecS and strengthens the notion that all eukaryotes have only a single route to Sec. Furthermore, we show that selenoproteins are dispensable for insect-stage *T. brucei* under standard growth conditions.

Materials and Methods

Cell Culture. Procytic *T. brucei*, strain 427 and strain 29-13, and the corresponding transgenic cell lines were grown at 27 °C in SDM-79 (24) supplemented with 5% and 15% FCS, respectively.

Transgenic Cell Lines. RNAi cell lines were produced by using pLew-100-based stem loop constructs containing the puromycin resistance gene (25–27). As inserts we used a 451-bp fragment (nucleotides 181–632) for the Tb-SP52 gene and a 528-bp fragment of the Tb-EF5ec gene (nucleotides 1–528). Production and initial characterization of Tb-SerRS RNAi cell line has been described (18). The double allelic replacements of the Tb-PSTK gene, in *T. brucei* 427, and of the Tb-SepSecS gene, in *T. brucei* 29–13, are described in Fig. S1. Inducible ectopic expression of Tb-SepSecS was done by using pLew-100 carrying the phleomycin resistance gene (25). Transfection, selection of transformants, and production of clonal cell lines were done by using standard procedures as described (12).

⁷⁵Selenium Labeling. A total of 5×10^7 cells was resuspended in 0.5 mL of FCS-supplemented SDM-79. The cultures were labeled with 9.6 μ Ci of Hepsen-neutralized [⁷⁵Se] selenite (University of Missouri Research Reactor, Columbia) in the presence of 100 μ g/mL cysteine at 27 °C for 3 h (for the KO strains) or 1 h (for the RNAi cell lines). After the labeling, the culture was washed with PBS, and the resulting pellet was resuspended in sample buffer and heated to 100 °C for 10 min. Finally, the labeled proteins ($\sim 10^7$ cell equivalents) were analyzed on 10–20% Tris-tricine gels (Ready gel; Bio-Rad) and visualized with a phosphorimager.

Acid Gel Analysis of tRNA^{Sec} Population. Total RNA was isolated as described (29) and resuspended in 10 mM Na-acetate, pH 4. Isolated RNA corresponding to 8×10^6 cell equivalents was run on a 50-cm-long acidic sequencing gel as described (20). The gel was run at 4 °C in 0.1 M Na-acetate, pH 5 until the xylene blue front migrated to ≈ 28 cm from the top. The section of the gel containing the RNA was blotted to a Genescreen plus membrane and analyzed by Northern blots using oligonucleotide hybridization as described (28).

Benzyl Viologen Assay for Active FDH. Benzyl viologen-dependent FDH assays were performed as described (6). The PSTK and SepSecS genes of the different species were cloned into the pACYCDuet-1 and pET15b expression vectors (Novagen), respectively. In strains expressing only 1 gene, the corresponding empty plasmid was cotransformed as a control.

ACKNOWLEDGMENTS. We thank G. Cross (The Rockefeller University, New York) for cell lines and plasmids and Isabel Roditi (University of Bern) for helpful discussions. This work was supported by Swiss National Foundation Grant 3100-067906 (to A.S.) and National Institutes of Health Grants AI028798 (to E.U.) and GM22854 (to D.S.).

1. Commins S, Böck A (1999) Selenocysteine inserting tRNAs: An overview. *FEMS Microbiol Rev* 23:335–351.
2. Hatfield DL, Gladyshev VN (2002) How selenium has altered our understanding of the genetic code. *Mol Cell Biol* 22:3565–3576.
3. Hatfield DL, et al. (2006) Selenocysteine incorporation machinery and the role of selenoproteins in development and health. *Prog Nucleic Acid Res Mol Biol* 81:97–142.
4. Su D, et al. (2009) How an obscure archaeal gene inspired the discovery of selenocysteine biosynthesis in humans. *IUBMB Life* 61:35–39.
5. Forchhammer K, Böck A (1991) Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. *J Biol Chem* 266:6324–6328.
6. Yuan J, et al. (2006) RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc Natl Acad Sci USA* 103:18923–18927.
7. Xu XM, et al. (2007) Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol* 5:e4.
8. Xu XM, et al. (2007) Selenophosphate synthetase 2 is essential for selenoprotein biosynthesis. *Biochem J* 404:115–120.
9. Forchhammer K, Leinfelder W, Böck A (1989) Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. *Nature* 342:453–456.
10. Kryukov GV, et al. (2003) Characterization of mammalian selenoproteomes. *Science* 300:1439–1443.
11. Xu XM, et al. (2005) Evidence for direct roles of two additional factors, SECp43 and soluble liver antigen, in the selenoprotein synthesis machinery. *J Biol Chem* 280:41568–41575.
12. Beverley SM, Clayton CE (1993) Transfection of *Leishmania* and *Trypanosoma brucei* by electroporation. *Methods Mol Biol* 21:333–348.
13. Ullu E, Tschudi C, Chakraborty T (2004) RNA interference in protozoan parasites. *Cell Microbiol* 6:509–519.
14. Dacks JB, Walker G, Field MC (2008) Implications of the new eukaryotic systematics for parasitologists. *Parasitol Int* 57:97–104.
15. Lobanov AV, Gromer S, Salinas G, Gladyshev VN (2006) Selenium metabolism in *Trypanosoma*: Characterization of selenoproteomes and identification of a Kinetoplastida-specific selenoprotein. *Nucleic Acids Res* 34:4012–4024.
16. Cassago A, et al. (2006) Identification of *Leishmania* selenoproteins and SECIS element. *Mol Biochem Parasitol* 149:128–134.
17. Bouzaidi-Tiali N, et al. (2007) Elongation factor 1a mediates the specificity of mitochondrial tRNA import in *T. brucei*. *EMBO J* 26:4302–4312.
18. Geslain R, et al. (2006) *Trypanosoma* seryl-tRNA synthetase is a metazoan-like enzyme with high affinity for tRNA^{Sec}. *J Biol Chem* 281:38217–38225.
19. Sculaccio SA, et al. (2008) Selenocysteine incorporation in Kinetoplastid: Selenophosphate synthetase (SELD) from *Leishmania major* and *Trypanosoma brucei*. *Mol Biochem Parasitol* 162:165–171.
20. Varshney U, Lee C-P, RajBhandary UL (1991) Direct analysis of aminoacylation levels of tRNAs in vivo. *J Biol Chem* 266:24712–24718.
21. Bösl MR, et al. (1997) Early embryonic lethality caused by targeted disruption of the mouse selenocysteine tRNA gene (*Trsp*). *Proc Natl Acad Sci USA* 94:5531–5534.
22. Lacourciere GM, Levine RL, Stadtman TC (2002) Direct detection of potential selenium delivery proteins by using an *Escherichia coli* strain unable to incorporate selenium from selenite into proteins. *Proc Natl Acad Sci USA* 99:9150–9153.
23. Rigobello MP, Scutari G, Boscolo R, Bindoli A (2002) Induction of mitochondrial permeability transition by auranofin, a Gold(I)-phosphine derivative. *Br J Pharmacol* 136:1162–1168.
24. Brun R, Schönenberger M (1979) Cultivation in vitro cloning of procyclic culture forms of *Trypanosoma brucei* in a semidefined medium. *Acta Tropica* 36:289–292.
25. Wirtz E, Leal S, Ochatt C, Cross GA (1999) A tightly regulated inducible expression system for conditional gene knockouts and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol* 99:89–101.
26. Morris JC, et al. (2001) Replication of kinetoplast DNA: An update for the new millennium. *Int J Parasitol* 31:453–458.
27. Bochud-Allemann N, Schneider A (2002) Mitochondrial substrate level phosphorylation is essential for growth of procyclic *Trypanosoma brucei*. *J Biol Chem* 277:32849–32854.
28. Tan THP, et al. (2002) tRNAs in *Trypanosoma brucei*: Genomic organization, expression, and mitochondrial import. *Mol Cell Biol* 22:3707–3717.
29. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156–159.

24. F. F. Jesus, J. F. Wilkins, V. N. Solferini, J. Wakeley, *Genet. Mol. Res.* **5**, 466 (2006).
25. T. Ohta, *Nature* **246**, 96 (1973).
26. S. L. Kosakovsky Pond, S. D. W. Frost, *Mol. Biol. Evol.* **22**, 478 (2004).
27. A. Gnirke *et al.*, *Nat. Biotechnol.* **27**, 182 (2009).
28. E. Hodges *et al.*, *Nat. Genet.* **39**, 1522 (2007).
29. J. P. Noonan *et al.*, *Science* **314**, 1113 (2006).
30. F. Ronquist, J. P. Huelsenbeck, *Bioinformatics* **19**, 1572 (2003).
31. We thank P. Johnson, M. Knapp, A.-S. Malaspinas, M. Meyer, J. Sullivan, M. Slatkin, and anonymous reviewers

for comments; the Croatian Academy of Sciences and Arts and the Berlin-Brandenburg Academy of Sciences for logistic and scientific support; and the Presidential Innovation Fund of the Max Planck Society for financial support. The government of the Principado de Asturias funded excavations at the El Sidron site. C.L.-F. was supported by the Spanish Ministry of Education and Science and J.M.G. by an NSF international postdoctoral fellowship (OISE-0754461). Sequences are deposited at the EBI (European Bioinformatics Institute) nucleotide database with the following accession numbers: Neandertal 1 (Feldhofer 1), FM865407; Neandertal 2

(Feldhofer 2), FM865408; Sidron 1253, FM865409; Vindija 33.25, FM865410; Mezmaiskaya 1, FM865411.

Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5938/318/DC1

Materials and methods

Figs. S1 to S14

Tables S1 to S8

References

Appendix 1

3 April 2009; accepted 3 June 2009

10.1126/science.1174462

The Human SepSecS-tRNA^{Sec} Complex Reveals the Mechanism of Selenocysteine Formation

Sotiria Palioura,¹ R. Lynn Sherrer,¹ Thomas A. Steitz,^{1,2,3} Dieter Söll,^{1,2*} Miljan Simonović^{4*}

Selenocysteine is the only genetically encoded amino acid in humans whose biosynthesis occurs on its cognate transfer RNA (tRNA). *O*-Phosphoseryl-tRNA:selenocysteinyl-tRNA synthase (SepSecS) catalyzes the final step of selenocysteine formation by a poorly understood tRNA-dependent mechanism. The crystal structure of human tRNA^{Sec} in complex with SepSecS, phosphoserine, and thiophosphate, together with *in vivo* and *in vitro* enzyme assays, supports a pyridoxal phosphate-dependent mechanism of Sec-tRNA^{Sec} formation. Two tRNA^{Sec} molecules, with a fold distinct from other canonical tRNAs, bind to each SepSecS tetramer through their 13-base pair acceptor-T Ψ C arm (where Ψ indicates pseudouridine). The tRNA binding is likely to induce a conformational change in the enzyme's active site that allows a phosphoserine covalently attached to tRNA^{Sec}, but not free phosphoserine, to be oriented properly for the reaction to occur.

The 21st amino acid, selenocysteine (Sec), is distinct from other amino acids not only because it lacks its own tRNA synthetase, but also because it is the only one that is synthesized on the cognate tRNA in all domains of life [reviewed in (1–4)] in a process that is reminiscent of the tRNA-dependent synthesis of glu-

tamine, asparagine, and cysteine in prokaryotes (4). The importance of Sec is illustrated by the embryonic lethal phenotype of the tRNA^{Sec} knockout mouse (5) and by the presence of Sec in the active sites of enzymes involved in removing reactive oxidative species and in thyroid hormone activation (6, 7). It is intriguing that the codon for

Sec is UGA, which is normally a translational stop signal (1). During translation of selenoprotein mRNAs, UGA is recoded by the interaction of a specialized elongation factor, SelB in bacteria and EFsec in humans, with a downstream Sec-insertion sequence element that forms a stem loop (1, 3).

The first step in Sec formation involves the misacylation of tRNA^{Sec} by seryl-tRNA synthetase (SerRS) to give Ser-tRNA^{Sec} [reviewed in (3, 8)]. Although the tertiary structure of tRNA^{Sec} was unknown, it was proposed that the mischarging reaction is possible because of similarities between the tRNA^{Sec} and tRNA^{Ser} structures (3, 8). In archaea and eukaryotes, the γ -hydroxyl group of Ser-tRNA^{Sec} is subsequently phosphorylated by *O*-phosphoseryl-tRNA kinase (PSTK) (9) to give *O*-phosphoseryl-tRNA^{Sec} (Sep-tRNA^{Sec}), which is then used as a substrate for the last synthetic

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. ²Department of Chemistry, Yale University, New Haven, CT 06520, USA. ³Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA. ⁴Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, USA.

*To whom correspondence should be addressed. E-mail: dieter.soll@yale.edu (D.S.); msimon5@uic.edu (M.S.)

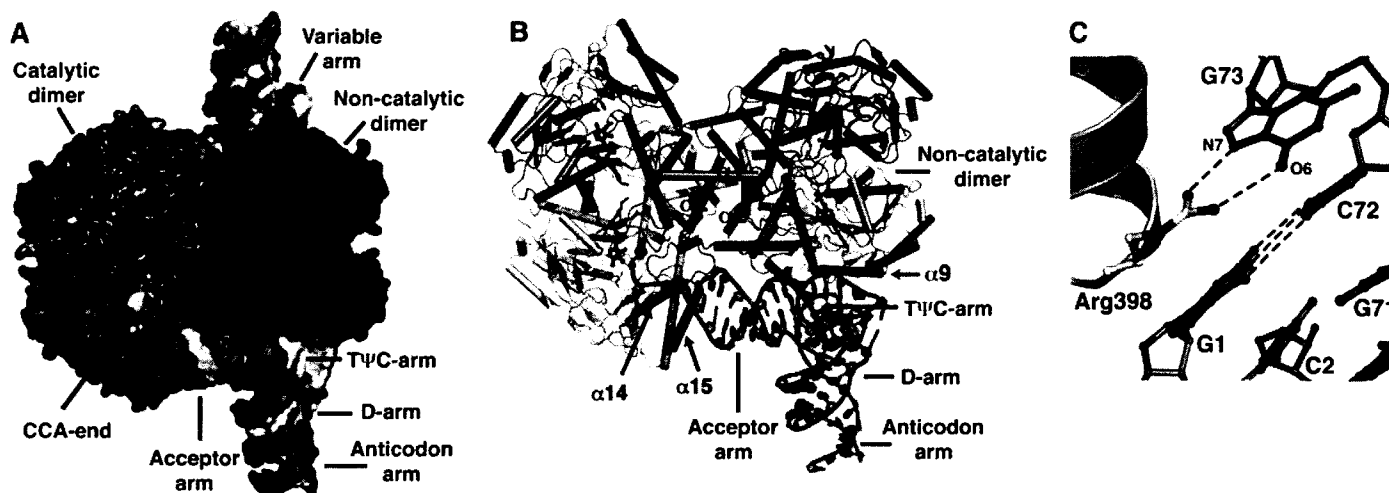


Fig. 1. Structure of human SepSecS in complex with unacylated tRNA^{Sec}. (A) Surface representation of the physiological complex of SepSecS with tRNA^{Sec}. The subunits of the catalytic dimer are dark and light blue, those of the noncatalytic dimer are dark and light red; the backbone and the bases of tRNA^{Sec} are green and gray, respectively. (B) The catalytic dimer interacts with the acceptor arm of tRNA^{Sec} through helices $\alpha 14$ and $\alpha 15$

(blue). The $\alpha 1$ helix (red) of the noncatalytic dimer interacts with the rest of the acceptor-T Ψ C arm. The regions of tRNA^{Sec} that interact with SepSecS are shown in orange; the rest is green. One tRNA^{Sec} molecule is shown for clarity. (C) Interactions between the discriminator base G73 and the conserved Arg³⁹⁸ in the $\alpha 14$ - $\beta 11$ loop. The protein side chains are gold, and tRNA^{Sec} is green.

enzyme, SepSecS (8, 10). SepSecS catalyzes the conversion of the phosphoserine moiety into the selenocysteine group by using selenophosphate as the selenium donor. An early observation that autoantibodies isolated from patients

with type I autoimmune hepatitis targeted a ribonucleoprotein complex containing tRNA^{Sec} led to the identification and characterization of the archaeal and the human SepSecS (2, 11). The crystal structures of the archaeal and murine

SepSecS apo-enzymes and phylogenetic analysis suggested that SepSecS forms its own branch in the family of fold-type I pyridoxal phosphate (PLP) enzymes that goes back to the last universal common ancestor (12, 13).

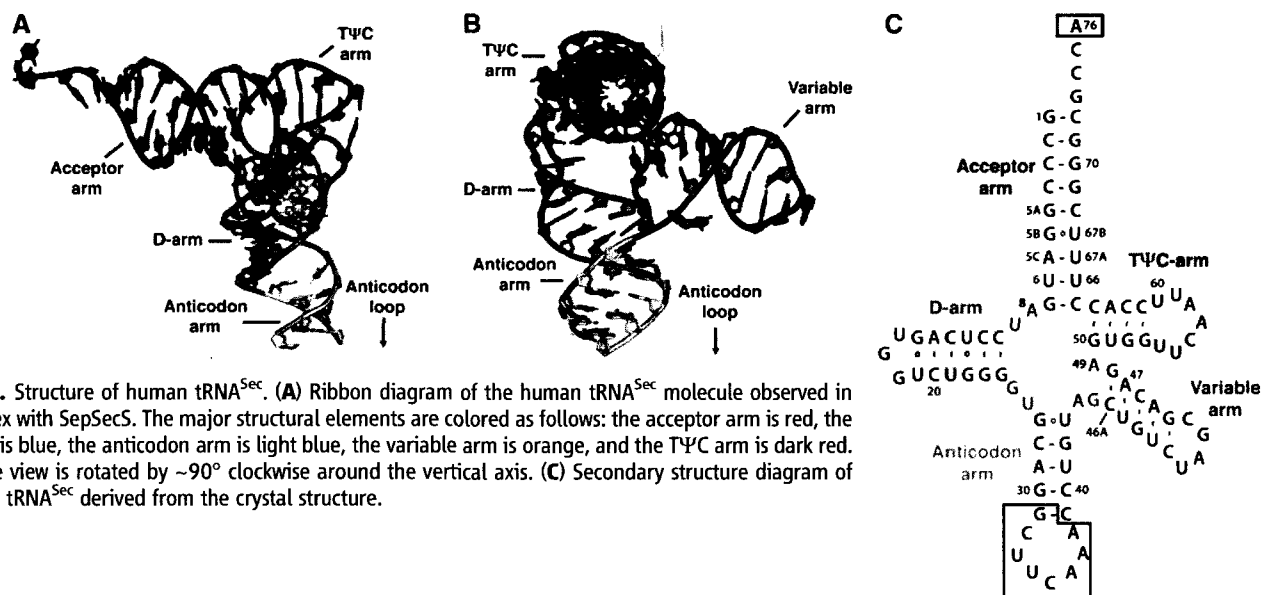
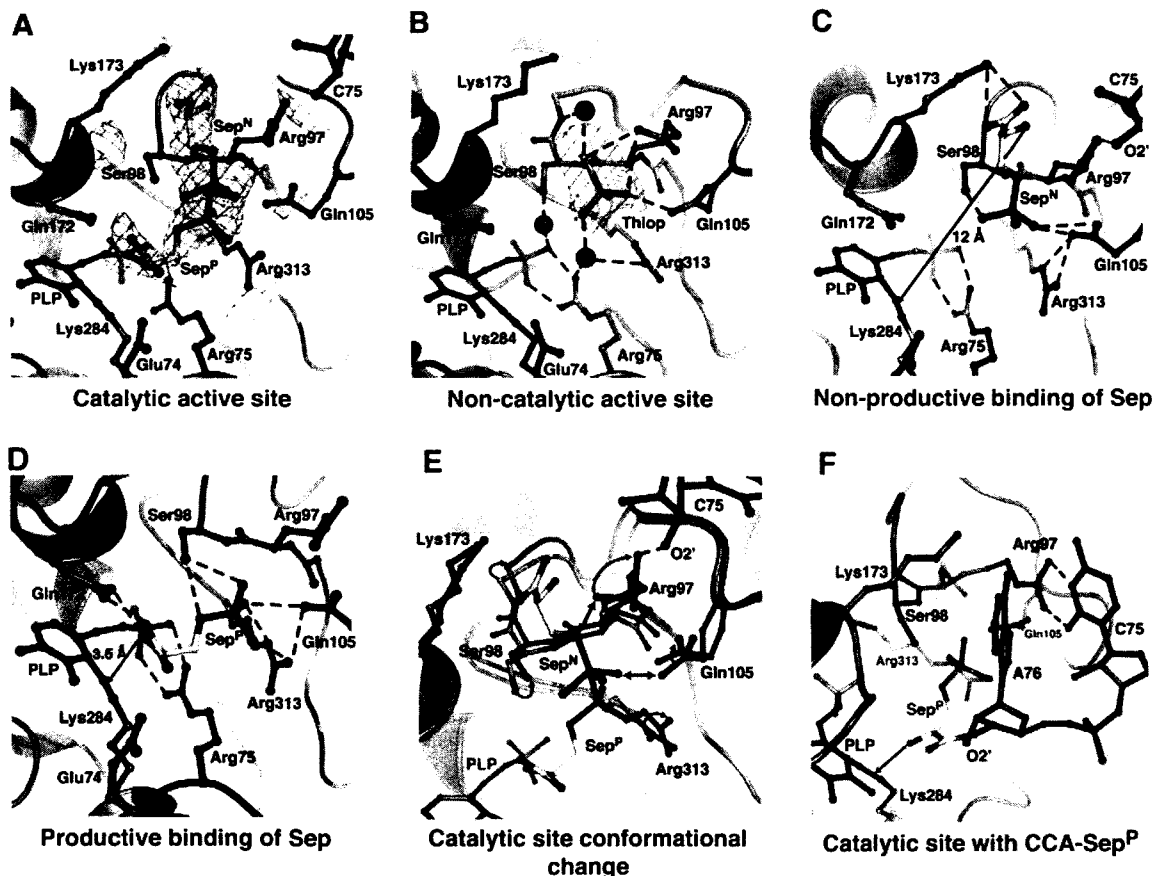


Fig. 3. Ligand binding to the active sites in the SepSecS-tRNA^{Sec} complex. In (A), (C), (D), (E), and (F), the catalytic PLP-monomer is purple, the P-loop monomer is light blue, Sep is gold, tRNA is green, and PLP is magenta. In (A) and (B), the unbiased omit electron density map (green mesh) is contoured at 3.5 σ. (A) Phosphoserine binds to the catalytic sites in two orientations (Sep^P and Sep^N). (B) Thiop (orange) binds only to the non-catalytic site. The PLP-monomer is brown, and the P-loop monomer is pink. (C) The Sep^N amino group is ~12 Å away from the Schiff base. Arg⁹⁷, Gln¹⁰⁵, and Arg³¹³ coordinate phosphate, and the carboxyl group interacts with Lys¹⁷³. (D) The amino group of Sep^P interacts with Gln¹⁷² and is 3.5 Å away from the Schiff base. The carboxyl group interacts with Gln¹⁷², whereas Ser⁹⁸, Gln¹⁰⁵, and Arg³¹³ coordinate phosphate. (E) The P-loop adopts different conformation after tRNA^{Sec} binding. The noncatalytic dimer is pink, the catalytic dimer is light blue. Steric clashes between the noncatalytic P-loop and Sep^N



are shown (double arrow). (F) A model of CCA-Sep^P (gold) in the catalytic active site. A76 binds between the side chains of Arg⁹⁷ and Lys¹⁷³, whereas C75 interacts with Arg⁹⁷.

In contrast to its closest homologue, SepCysS, which functions as a dimer to convert Sep-tRNA^{Cys} to Cys-tRNA^{Cys} in archaea (14, 15), SepSecS forms a stable tetramer (12, 13). SepSecS acts on phosphoserine that is linked to tRNA^{Sec} and not on free phosphoserine or Ser-tRNA^{Sec} (8, 13). However, the molecular basis for substrate discrimination and the roles of PLP and tRNA^{Sec} in the mechanism of Sep to Sec conversion are not clear. To explore these questions, we have determined the crystal structure of the quaternary complex between human SepSecS, unacylated tRNA^{Sec}, and a mixture of *O*-phosphoserine (Sep) and thio-

phosphate (Thiop) to 2.8 Å resolution. The observed intensity divided by its standard deviation [$I/\sigma(I)$] of the x-ray diffraction data are 2 at 3.0 Å resolution, but the data out to 2.8 Å resolution, where $I/\sigma(I)$ drops to 1, were included in the structure refinement.

Human SepSecS forms a tetramer that is bound to two tRNA^{Sec} molecules in the crystal (Fig. 1, A and B, and fig. S1). Computational modeling suggests that the tetrameric enzyme could potentially bind up to four tRNA^{Sec} molecules (fig. S2), but these additional tRNA-binding sites are blocked by crystal packing in

our crystal (16). Electron density for the 23 N-terminal and 15 C-terminal residues of SepSecS, as well as for the anticodon loop (nucleotides 31 to 38) and A76 of tRNA^{Sec}, was of poor quality, and these residues were not included in the final model. Each SepSecS monomer has a PLP co-factor covalently linked to the Ne-amino group of the conserved Lys²⁸⁴ by means of formation of a Schiff base (internal aldimine). Two SepSecS monomers form a homodimer, and two active sites are formed at the dimer interface. The two homodimers associate into a tetramer through interactions between the N-terminal α 1-loop- α 2

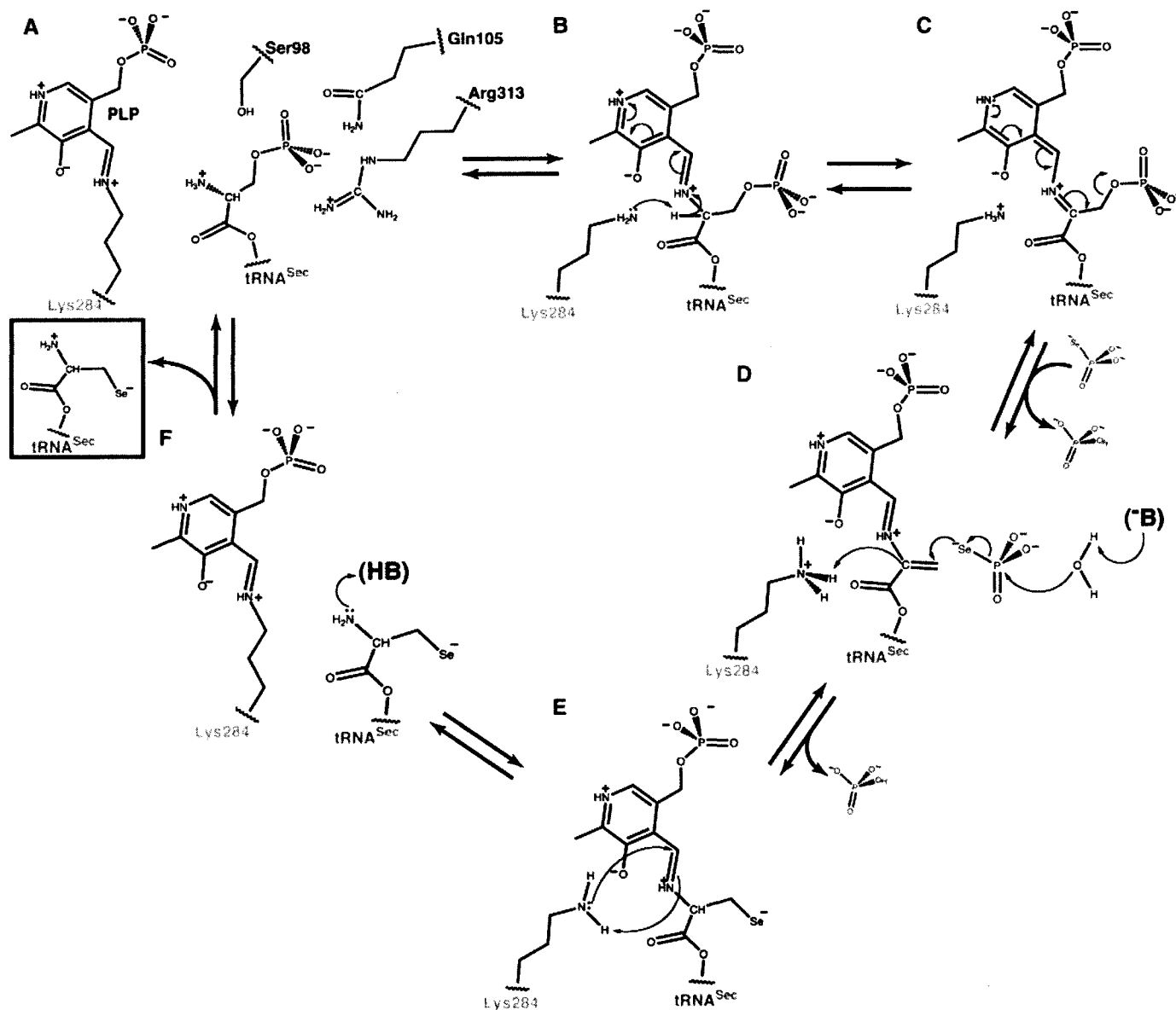


Fig. 4. The PLP-dependent mechanism of Sep to Sec conversion. **(A)** The phosphoserine moiety of Sep-tRNA^{Sec} is bound to the active site similar to Sep^P. The amino group is oriented for attack on the Schiff base, whereas the phosphoryl group is stabilized by the side chains of Ser⁹⁸, Gln¹⁰⁵, and Arg³¹³. Hydrogen bonds are shown in dashed lines. **(B)** After the formation of the external aldimine, the side chain of Lys²⁸⁴ rearranges and abstracts the C α proton from Sep. The protonated pyridine ring of PLP stabilizes the carbanion. **(C)** Electron delocalization leads to a rapid β -elimination of phosphate and to the formation

of dehydroalanyl-tRNA^{Sec}. Free phosphate dissociates, and selenophosphate binds to the active site. **(D)** An unidentified base (**B**) activates water that hydrolyzes selenophosphate. Free phosphate dissociates again, and selenium attacks the dehydroalanyl-tRNA^{Sec}. Lys²⁸⁴ returns the proton to the C α carbon, and the selenocysteine moiety is formed. **(E)** The reaction of reverse transaldimination is shown. Lys²⁸⁴ forms the Schiff base, with PLP leading to a release of the oxidized form of Sec-tRNA^{Sec} (red box). **(F)** The free amino group of Sec-tRNA^{Sec} is protonated, and the active site of SepSecS is regenerated.

motifs (Fig. 1B). Given that SepCysS is a dimer and that the active sites of one homodimer do not communicate with the active sites of the other homodimer in the apo-SepSecS tetramer (12, 13), the reason for the tetrameric organization of SepSecS was not known. The mode of tRNA^{Sec} binding to SepSecS provides an answer to this question.

The CCA ends of both tRNA^{Sec} molecules point to the active sites of the same homodimer, which we shall refer to as the catalytic dimer (Fig. 1A). The other homodimer, which we shall refer to as the noncatalytic dimer, serves as a binding platform that orients tRNA^{Sec} for catalysis (Fig. 1A). The structure reveals that SepSecS binds only to the acceptor-, T Ψ C-, and variable arms of tRNA^{Sec} (Fig. 1B and fig. S3A). (Ψ , pseudouridine.) The tip of the acceptor arm interacts with the C terminus of the catalytic dimer, whereas the rest of the acceptor-, T Ψ C-, and variable arms wrap around a monomer from the noncatalytic dimer (Fig. 1, A and B, and fig. S3A). The most important binding element is an interaction between the discriminator base G73 of tRNA^{Sec} and the conserved Arg³⁹⁸ of the catalytic dimer (Fig. 1C). The guanidinium group of Arg³⁹⁸ forms hydrogen bonds with the Hoogsteen face of G73. The discriminator base G73 of tRNA^{Sec} is universally conserved in archaea and eukaryotes. Neither adenine nor cytosine in position 73 could form hydrogen bonds with Arg³⁹⁸ because they have amino groups instead of the keto group. Also, if a cytosine or uridine were in position 73, the C5 and C6 atoms of the pyrimidine ring would clash with the side chain of Thr³⁹⁷ and thus prevent the interaction of these bases with Arg³⁹⁸.

Moreover, the side chain of Lys⁴⁶³ from the C-terminal helix α 15 forms a hydrogen bond with the backbone oxygen of G69 from the acceptor arm. We propose that autoantibodies bind to an interface that lies between the α 15 helix of SepSecS and the tip of the acceptor arm of tRNA^{Sec} (Fig. 1B), which enables them to precipitate the entire ribonucleoprotein complex (11). The interaction of the antibody with a region that lies close to the acceptor stem–active site interface may inhibit the function of SepSecS. This would be similar to the mechanism of autoimmune hepatitis type 2, where LKM-1 autoantibodies (anti-liver-kidney microsomal antibodies) inhibit cytochrome P450 isoenzyme 2D6 (CYP2D6) and contribute to the pathogenesis of the disease (17).

The tRNA^{Sec} molecule is anchored to the SepSecS tetramer, on one end by interactions between the tip of its acceptor arm (G73) and the catalytic dimer (Arg³⁹⁸) (Fig. 1C) and on the other end by interactions between its variable arm (C46L) and the noncatalytic dimer (Arg²⁷¹) (fig. S3B). The rest of the interactions between the α 1 helix of the noncatalytic dimer of SepSecS (Arg²⁶, Lys³⁸, and Lys⁴⁰) and the acceptor-T Ψ C arm (C2, G50, and C64) further stabilize the complex (fig. S3C). For instance, the side chain of Arg²⁶ forms hydrogen bonds with the phosphate backbone of C2-C3 of the acceptor arm

and the Ne-amino group of Lys³⁸ is within the hydrogen bonding distance from both the O2'-hydroxyl group and the O2 atom of C64 (fig. S3C), whereas the side chain of Lys⁴⁰ interacts with the phosphate oxygen of G50 from the T Ψ C arm (fig. S3C). The interactions between SepSecS and tRNA^{Sec} seen in the crystal are consistent with *in vivo* activity assays of SepSecS mutants (fig. S4A) (16). For example, replacing Arg³⁹⁸ with either alanine or glutamate renders the enzyme completely inactive, which suggests that the interaction between the discriminator base and the highly conserved Arg³⁹⁸ of the catalytic dimer is critical for tRNA^{Sec} recognition (fig. S4A).

Human tRNA^{Sec} contains 90 nucleotides rather than the conventional 75 nucleotides of canonical tRNA molecules. The structure shows that human tRNA^{Sec} adopts a unique 9/4 fold with a 13-base pair (bp) acceptor-T Ψ C arm (where 9 and 4 reflect the number of base pairs in the acceptor and T Ψ C arms, respectively) and a long variable arm (Fig. 2) (16). This resolves a controversy between conflicting models in which the 7/5 model suggested 12 bp in the acceptor-T Ψ C arm, as found in all known tRNA structures (18), whereas the 9/4 model suggested a unique 13-bp acceptor-T Ψ C arm (19). The 13-bp acceptor-T Ψ C arm and a long variable arm are distinct structural features that serve as major recognition motifs for binding to SepSecS (Fig. 1, A and B, and fig. S3). Indeed, based on our modeling analysis with tRNA^{Asp}, we suggest that tRNA^{Ser}, which contains both the G73 discriminator base and a long variable arm, is not able to bind to SepSecS because of its shorter acceptor-T Ψ C arm (fig. S5, A, B, and C). This is reminiscent of the rejection of tRNA^{Sec} by bacterial EF-Tu (20). However, modeling of the SerRS-tRNA^{Sec} complex based on the crystal structure of the bacterial SerRS in complex with tRNA^{Ser} (21) suggests that SerRS can recognize the variable arm of tRNA^{Sec} similarly to tRNA^{Ser}, which provides an explanation for the inability of SerRS to discriminate between tRNA^{Ser} and tRNA^{Sec} (fig. S5D). Finally, the observations that the archaeal PSTK and SepSecS can act on *E. coli* tRNA^{Sec} *in vivo*; that human SepSecS can use *E. coli* tRNA^{Sec} *in vivo*, as well as the archaeal tRNA^{Sec} *in vitro* (8); and that the bacterial selenocysteine synthase can use both archaeal and murine tRNA^{Sec} *in vitro* (10) suggest that the length of the acceptor-T Ψ C arm of tRNA^{Sec}, the position of the variable arm, and the mode of tRNA^{Sec} recognition are likely to be conserved in all domains of life.

To explore the mechanism by which the phosphoryl group is converted to the selenocysteinyl moiety, we soaked crystals of the binary SepSecS-tRNA^{Sec} complex in a solution containing a mixture of *O*-phosphoserine and thiophosphate. Sep and Thiop were used as mimics of the phosphoserine group attached to tRNA^{Sec} and selenophosphate, respectively. Although both ligands were used in the soaking experiments, Sep bound only to the active sites of the

catalytic dimer (Fig. 3A), whereas Thiop bound only to the active sites of the noncatalytic dimer (Fig. 3B), i.e., Sep can bind to an active site only in the presence of tRNA^{Sec}. Both Thiop and the phosphoryl group of Sep bind to the same binding pocket (Fig. 3, A and B), which suggests that a specific active site can accommodate only one ligand at a time. Sep binds to the catalytic active site in either of two different orientations (Fig. 3A). The phosphoryl group occupies a similar location in the two orientations, whereas the seryl moieties are rotated by $\sim 90^\circ$ around the phosphate group. In the nonproductive orientation of Sep (Sep^N), its seryl group is sandwiched between the side chains of Arg⁹⁷ and Lys¹⁷³ at the catalytic dimer interface, whereas its amino group lies ~ 12 Å away from the PLP Schiff base (Fig. 3C). Thus, it is unlikely that SepSecS would act on Sep^N, unless a PLP-independent mechanism is utilized. In the productive-like orientation of Sep (Sep^P), the amino group of Sep is ~ 3.5 Å away from the Schiff base, yet it is not positioned for attack because of a hydrogen bond with Gln¹⁷² (Fig. 3D). The carboxyl group of Sep^P also forms a hydrogen bond with the side chain of Gln¹⁷², whereas its phosphoryl group anchors the ligand into the active site through its interactions with the side chains of Ser⁹⁸, Gln¹⁰⁵, and Arg³¹³. That the phosphoryl group is required to properly position the ligand for catalysis explains why the obligate substrate for SepSecS is Sep-tRNA^{Sec} and not Ser-tRNA^{Sec}, which is the physiological substrate for the bacterial selenocysteine synthase.

A comparison of the catalytic and noncatalytic sites reveals that Sep binds to the active site only in the presence of tRNA^{Sec} because the conformation of the P-loop (residues Gly⁹⁶ to Lys¹⁰⁷) differs between the two active sites (Fig. 3E). In the noncatalytic dimer, the guanidinium group of Arg⁹⁷ and the side chain of Gln¹⁰⁵ are rotated toward the phosphate-binding groove, where they coordinate Thiop (Fig. 3, B and E). In the catalytic dimer, the side chain of Arg⁹⁷ rotates away from the phosphate-binding groove and forms a hydrogen bond with the 2'-OH group of C75. Gln¹⁰⁵ also rotates away from the phosphate, and this concerted movement of Arg⁹⁷ and Gln¹⁰⁵ in the P-loop on tRNA^{Sec} binding allows Sep to bind to the active site (Fig. 3E). Free Sep enters the active site through the gate formed by the side chains of Arg⁹⁷, Gln¹⁰⁵, and Lys¹⁷³ and gets trapped in the Sep^N orientation. Electron density for both the amino and the carboxyl groups of Sep^P is weak, which suggests higher mobility in this part of Sep^P, presumably because of free rotation around the C α -C β bond. This explains why the free amino group of Sep^P cannot be positioned appropriately for attack onto the Schiff base and why the reaction does not occur in the crystal. Thus, the covalent attachment of Sep to tRNA^{Sec} is necessary for the proper placement of the Sep moiety into the active site and for orienting the amino group of Sep for attack onto the Schiff base of PLP (Fig. 3F).

We used both in vivo and in vitro activity assays to investigate the mechanism of Sep-tRNA to Sec-tRNA conversion by human SepSecS. First, the reduction of the Schiff base by sodium borohydride to form a chemically stable secondary amine and thus to cross-link PLP to Lys²⁸⁴ renders SepSecS completely inactive in vitro (fig. S4B). The catalytic activity of SepSecS is also quenched on removal of PLP by treatment with hydroxylamine (fig. S4B). Some residual activity that is observed after hydroxylamine treatment is probably because of incomplete removal of PLP (fig. S4B). Second, we show that the Arg75Ala, Gln105Ala, and Arg313Ala mutants are inactive in vivo (fig. S4A). These residues are involved in coordinating either the phosphate group of PLP or that of Sep. Finally, the in vivo activities of the Arg97Ala, Arg97Gln, Lys173Ala, and Lys173Met mutants are indistinguishable from that of the wild-type enzyme, which confirms that Arg⁹⁷ and Lys¹⁷³ are involved only in the nonproductive binding of free Sep (fig. S4A).

On the basis of our findings, we propose the following PLP-based mechanism of Sep-tRNA to Sec-tRNA conversion. The reaction begins by the covalently attached Sep being brought into the proximity of the Schiff base when Sep-tRNA^{Sec} binds to SepSecS. The amino group of Sep can then attack the Schiff base formed between Lys²⁸⁴ and PLP, which yields an external aldimine (Fig. 4, A and B). The reoriented side chain of Lys²⁸⁴ abstracts the C α proton from Sep (Fig. 4C), and the electron delocalization by the pyridine ring assists in rapid β -elimination of the phosphate group, which produces an intermediate dehydroalanyl-tRNA^{Sec} (Fig. 4, C and D). After phosphate dissociation and binding of selenophosphate, the concomitant attack of water on the selenophosphate group and of the nucleophilic selenium onto

the highly reactive dehydroalanyl moiety yield an oxidized form of Sec-tRNA^{Sec} (Fig. 4D). The protonated Lys²⁸⁴, returns the proton to the C α carbon and then attacks PLP to form an internal aldimine (Fig. 4E). Finally, Sec-tRNA^{Sec} is released from the active site (Fig. 4F).

This mechanism is clearly distinct from the persulfide-intermediate mechanism in the Sep-tRNA^{Cys} to Cys-tRNA^{Cys} reaction (22) and explains why SepSecS does not group together with its closest homolog, SepCysS, in the family tree of fold-type I PLP enzymes (12). Moreover, the proposed mechanism for SepSecS is similar to the one used by the bacterial SelA that also proceeds through a dehydroalanyl-tRNA^{Sec} intermediate (23). SepSecS therefore uses a primordial tRNA-dependent catalytic mechanism in which the PLP cofactor is directly involved, while using a tetrameric fold-type I architecture as the scaffold for binding the distinct structure of tRNA^{Sec}.

References and Notes

1. A. Böck, M. Thanbichler, M. Rother, A. Resch, in *Aminoacyl-tRNA Synthetases*, M. Ibba, C. Francklyn, S. Cusack, Eds. (Landes Bioscience, Georgetown, TX, 2005), pp. 320–327.
2. D. Su *et al.*, *IUBMB Life* **61**, 35 (2009).
3. A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* **3**, 29 (2007).
4. K. Sheppard *et al.*, *Nucleic Acids Res.* **36**, 1813 (2008).
5. M. R. Bösl, K. Takaku, M. Oshima, S. Nishimura, M. M. Taketo, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5531 (1997).
6. M. P. Rayman, *Lancet* **356**, 233 (2000).
7. G. V. Kryukov *et al.*, *Science* **300**, 1439 (2003).
8. J. Yuan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18923 (2006).
9. B. A. Carlson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12848 (2004).
10. X. M. Xu *et al.*, *PLoS Biol.* **5**, e4 (2007).
11. C. Gelpi, E. J. Sontheimer, J. L. Rodriguez-Sanchez, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9739 (1992).
12. Y. Arais *et al.*, *Nucleic Acids Res.* **36**, 1187 (2008).
13. O. M. Ganichkin *et al.*, *J. Biol. Chem.* **283**, 5849 (2008).
14. A. Sauerwald *et al.*, *Science* **307**, 1969 (2005).
15. R. Fukunaga, S. Yokoyama, *J. Mol. Biol.* **370**, 128 (2007).
16. Materials and methods are available as supporting material on Science Online.
17. J. Herkel, M. P. Manns, A. W. Lohse, *Hepatology* **46**, 275 (2007).
18. A. Ioudovitch, S. V. Steinberg, *RNA* **4**, 365 (1998).
19. C. Sturchler, E. Westhof, P. Carbon, A. Krol, *Nucleic Acids Res.* **21**, 1073 (1993).
20. J. Rudinger, R. Hillenbrandt, M. Sprinzl, R. Gieger, *EMBO J.* **15**, 650 (1996).
21. V. Biou, A. Yaremchuk, M. Tukalo, S. Cusack, *Science* **263**, 1404 (1994).
22. S. I. Hauenstein, J. J. Perona, *J. Biol. Chem.* **283**, 22007 (2008).
23. K. Forchhammer, A. Böck, *J. Biol. Chem.* **266**, 6324 (1991).
24. We thank C. Axel Innis, Gregor Blaha, Robin Evans, Michael Strickler, Michael E. Johnson, Bernard D. Santarsiero, Jimin Wang, and the NE-CAT beamline staff (APS, ANL, Chicago) for their help during data collection and structure determination. We thank Dan Su, Theodoros Rampias, and Kelly Sheppard for helpful discussions. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (code 3HL2). Supported by grants from DOE (to D.S.) and NIGMS (to T.A.S. and D.S.). In the initial phase of this study M.S. was supported by HHMI at Yale University. S.P. holds a fellowship of the Yale University School of Medicine MD/PhD Program. R.L.S. was supported by a Ruth L. Kirschstein National Research Service Award. Authors' contributions: S.P. designed the research, collected and analyzed the data, and wrote the manuscript; R.L.S. did research and read the manuscript; T.A.S. designed the research and wrote the manuscript; D.S. designed the research and wrote the manuscript; M.S. designed the research, collected and analyzed the data, and wrote the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5938/321/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Table S1
References

18 March 2009; accepted 3 June 2009
10.1126/science.1173755

Tiger Moth Jams Bat Sonar

Aaron J. Corcoran,^{1*} Jesse R. Barber,² William E. Conner^{1*}

In response to sonar-guided attacking bats, some tiger moths make ultrasonic clicks of their own. The lepidopteran sounds have previously been shown to alert bats to some moths' toxic chemistry and also to startle bats unaccustomed to sonic prey. The moth sounds could also interfere with, or "jam," bat sonar, but evidence for such jamming has been inconclusive. Using ultrasonic recording and high-speed infrared videography of bat-moth interactions, we show that the palatable tiger moth *Bertholdia trigona* defends against attacking big brown bats (*Eptesicus fuscus*) using ultrasonic clicks that jam bat sonar. Sonar jamming extends the defensive repertoire available to prey in the long-standing evolutionary arms race between bats and insects.

The ability to pinpoint airborne insects in darkness, by echolocation (1), allowed bats to master nocturnal insectivory (2) and set the stage for the evolution of defensive countermeasures by insect prey (3). Some insects gained ears (4) and evasive maneuvering (5). Tiger moths (Lepidoptera: Arctiidae) developed the ability to click ultrasonically in response to attacking bats (6, 7). Decades of research on moth click de-

fenses have led to three, not mutually exclusive, hypotheses regarding their function—startle (8), acoustic aposematism ("warning") (8–14), and sonar interference ("jamming") (15–17). When ultrasonic clicks are paired with unpalatable prey, bats learn to perceive clicks as a warning of unprofitability (8, 10–14). Moth clicks also startle inexperienced bats or bats that have not heard clicks for multiple days (8). Because bats habit-

uate to startle quickly, its effectiveness as a defense requires clicking moths to be rare. This situation does not appear typical in nature (12).

Finally, moth clicks may disrupt the sonar of an attacking bat (15–17). Clicks might diminish a bat's acuity in determining target distance (17–19) or feign echoes from objects that do not exist (15). However, evidence that moth clicks can disrupt bat attacks by jamming sonar is lacking. One recent study found that moth clicks had no discernible effect on attacking bats unless clicks were paired with defensive chemistry (13). All previous studies, however, tested moths with relatively low duty cycles, or sound production per unit time (20). High-duty-cycle moth clicks, such as those of *Bertholdia trigona* (Fig. 1 and

¹Department of Biology, Wake Forest University, Winston-Salem, NC 27106, USA. ²Department of Fish, Wildlife and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA.

*To whom correspondence should be addressed. E-mail: corcaj8@wfu.edu (A.J.C.); conner@wfu.edu (W.E.C.)



Supporting Online Material for

The Human SepSecS-tRNA Complex Reveals the Mechanism of Selenocysteine Formation

Sotiria Palioura, R. Lynn Sherrer, Thomas A. Steitz, Dieter Söll,* Miljan Simonović*

*To whom correspondence should be addressed. E-mail: dieter.soll@yale.edu (D.S.); msimon5@uic.edu (M.S.)

Published 17 July 2009, *Science* **325**, 321 (2009)

DOI: 10.1126/science.1173755

This PDF file includes

Materials and Methods

SOM Text

Figs. S1 to S5

Table S1

References

SUPPORTING ONLINE MATERIAL

Results and Discussion

The 9/4 fold of human tRNA^{Sec}. The linker between the acceptor- and the D-stem is two residues long (A8-U9) as was proposed earlier (*1*). The base of A8 interacts with the sugar-phosphate backbone of both the variable and TΨC arms, while the base of U9 stacks with A49 from the variable arm-TΨC linker. The first pair of the D-arm is a loose C10-G24 couple followed by a C11-G23 pair (Fig. 2C). A U12-G22 wobble pair follows and here the only hydrogen bond is formed between O2-U12 and N2-G22. This observation agrees well with the earlier findings that N3-U12 is reactive toward carbodiimide (*1*). Further, the D-stem contains a C13-U21 pair, and A14-C20 and G15-U19 wobble pairs (Fig. 2C). These interactions explain why N1-A14, N1-G15 and N3-U21 were not reactive towards various chemical probes under native conditions (*1*). Three nucleotides U16, G17 and G18 form the D-loop and they interact with C56, A57, A58 and U59 from the TΨC loop closing the cloverleaf structure as proposed (*1*). The linker between the D-arm and the anticodon stem is one residue longer than in the theoretical 9/4 model and is composed of two unpaired nucleotides, G25 and U26. The anticodon arm begins with the G26-U43 base pair, whereas the anticodon loop is disordered in our crystal. A44 and G45 link the anticodon and variable arms and are modeled as single stranded. The rest of the tRNA^{Sec} structure agrees well with the proposed model (*1*).

Results of the *in vivo* activity assays are consistent with the SepSecS-tRNA^{Sec} interactions observed in the crystal. Analyzing the activity of the SepSecS mutants *in vivo* assessed the importance of interactions other than Arg398-G73, which is described in the main text. We find that the Thr397Ala mutant is completely active, Thr397Val is partially active, and Thr397Glu is inactive *in vivo* (Fig. S4A). Thus, Thr397 strengthens the interaction of the acceptor arm with the α 14 helix through its interactions with Arg398. In addition, the mutant Arg271Glu is partially active suggesting that the interaction between the variable arm and the non-catalytic dimer is important for complex formation as well (Fig. S4A). When Lys38 is replaced by Glu, the enzyme is partially active *in vivo*, whereas both the Arg26Ala and Arg26Glu mutants are completely active (Fig. S4A), suggesting that the Arg26-C2 interaction alone is not critical for complex formation. Finally, the human and archaeal SepSecS enzymes are completely inactive *in vivo* when the α 1 helix is deleted (Fig. S4A) (*2*).

SepSecS, PSTK and tRNA^{Sec} may form a ternary complex *in vivo*. The mode of tRNA^{Sec} recognition by human SepSecS is reminiscent of the tRNA-dependent amidotransferases GatDE and GatCAB that also recognize only the acceptor stem of the cognate tRNA (tRNA^{Gln}) (*3, 4*). Given that SepSecS only binds to one side of the acceptor-TΨC arm of tRNA^{Sec} it is plausible that PSTK would bind to a different side of the

acceptor-T Ψ C arm or to the D-stem as is proposed for the archaeal and human enzymes, respectively (5). This would then allow PSTK, SepSecS and tRNA^{Sec} to form a ternary complex *in vivo* similar to that of the transamidosome involved in prokaryotic tRNA-dependent asparagine biosynthesis (6) or the SepRS, SepCysS, tRNA^{Cys} components engaged in Cys-tRNA formation in methanogens (7).

Canonical tRNAs with the G73 discriminator base cannot bind to SepSecS due to the shorter acceptor-T Ψ C arm. The structure of the human SepSecS in complex with tRNA^{Sec} along with the *in vivo* mutagenesis data reveal that SepSecS can bind tRNA^{Sec} only as a tetramer. We propose that the interaction between Arg398 and G73 along with the longer 13 bp acceptor-T Ψ C arm and the variable arm of tRNA^{Sec} secure binding of the tRNA to the enzyme. The only tRNA with a G73 discriminator base and a long variable arm yet with a 12 bp acceptor-T Ψ C arm is tRNA^{Ser}. However, neither unacylated tRNA^{Ser} nor Ser-tRNA^{Ser} is able to bind to SepSecS (8). To examine the importance of the length of the acceptor-T Ψ C arm for SepSecS binding, we superimposed tRNA^{Asp} onto tRNA^{Sec}. The tRNA^{Asp} was chosen because it is a canonical tRNA molecule with G73 as discriminator base and because the complete structure of the tRNA^{Ser} is not available. We used two approaches for superimposing the two tRNAs. In the first approach the atoms of the sugar-phosphate backbone of the acceptor-, T Ψ C- and anticodon arms were superimposed. Although the secondary structure elements of tRNA^{Asp} were oriented as observed in tRNA^{Sec}, we find that G73 of the tRNA^{Asp} cannot reach Arg398 due to its shorter acceptor-T Ψ C arm (Fig. S5A, B). In the second approach, the discriminator base of tRNA^{Asp} was first modeled to form hydrogen bonds with Arg398 with its Hoogsteen face, and then the tips of the acceptor arms (nucleotides 1-3 and 70-73) were superimposed. In this case, we find that both the anticodon and variable arms of tRNA^{Asp} are positioned away from the enzyme, whereas its T Ψ C arm clashes with the α 1 helix of the non-catalytic dimer (Fig. S5C), thus preventing the formation of the SepSecS-tRNA^{Asp} complex. In addition, the absence of a long and properly oriented variable arm in tRNA^{Asp} diminishes further its capacity for interaction with SepSecS (Fig. S5A, C).

Asn252 in SepSecS plays a role of the conserved Asp252. From a mechanistic standpoint, SepSecS and SepCysS are the only Fold Type I PLP enzymes that possess an asparagine in contact with the N1 atom of the PLP pyridine ring (2). All other known Fold Type I PLP enzymes have an aspartate in the corresponding position that is critical for catalysis since it increases the electron sink properties of the PLP cofactor (9, 10). To test whether this Asn252 is also responsible for the discrimination that SepSecS exhibits towards Ser-tRNA^{Sec} we mutated it to an Asp. Finally, the replacement of Asn252 with Ala renders the enzyme inactive, whereas the Asn252Asp mutant is completely active (Fig. S4A). Asn252, thus, plays the role of the highly conserved Asp252 in the Fold Type I PLP enzymes, which protonates the pyridine ring of PLP.

The protonated PLP is critical for the stabilization of the covalently attached carbanion intermediate in the reaction pathway. Though still active towards Sep-tRNA^{Sec}, the Asn252Asp human SepSecS mutant is inactive towards Ser-tRNA^{Sec} (Fig. S4A).

Materials and Methods

Expression and purification of human SepSecS. The human SepSecS gene was cloned into the pET15b (Novagen) expression vector with an N-terminal His-tag as described previously (11). The construct was transformed into BL21(DE3) cd+ expression strain. Cells were grown in LB media at 37°C to A₆₀₀ = ~0.6 at which point the culture was induced with 0.5mM IPTG for an overnight expression at 15°C. The cells were harvested by centrifugation at 5,000 rpm for 15 minutes and resuspended in 20 mM Tris, pH 8.0, 50 mM NaCl, 10% (v/v) glycerol, 1 mM 2-mercaptoethanol, 10 μM PLP in the presence of the protease inhibitor cocktail (Complete EDTA-free, Roche). The cell paste was lysed using a microfluidizer, the insoluble material was separated by centrifugation at 32,000 rpm for 40 minutes, and the cleared supernatant was loaded onto a Ni-IDA (GE-Healthcare) column pre-equilibrated with 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM 2-mercaptoethanol, 10 μM PLP. Any non-specifically bound protein was washed away with 50 mM imidazole, while the His₆-SepSecS was eluted with 500 mM imidazole. The protein sample was diluted three-fold with 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DTT, 10 μM PLP, concentrated to 3 mL and loaded onto a S-200pg size-exclusion column. The protein eluted as a tetramer, it was concentrated to ~2 mg/mL, flash frozen in liquid nitrogen and stored at -80°C.

***In vitro* transcription and purification of human tRNA^{Sec}.** The human tRNA^{Sec} was cloned into pUC19, expressed in *E. coli* DH5α and synthesized by *in vitro* T7 RNA polymerase run-off transcription as described (12). The transcription reaction was performed at 37°C for 2h in buffer containing 40 mM Tris-HCl pH 8.1, 22 mM MgCl₂, 10 mM DTT, 1 mM spermidine, 0.01% Triton X-100, 50 μg/ml BSA, 4 mM of nucleoside triphosphate, 16 mM GMP, BstNI-digested vector containing the tRNA^{Sec} gene (70 μg/ml) and 3 mM T7 RNA Polymerase. The reaction was briefly centrifuged to pellet the formed pyrophosphate, the supernatant reaction mixture was loaded onto a Resource-Q column (GE-Healthcare) and tRNA^{Sec} was purified using a linear gradient of NaCl (0.2-1 M) in 20 mM Tris-HCl, pH 8.0. Human tRNA^{Sec} eluted at ~0.6 M NaCl. The purity of the tRNA^{Sec}-containing fractions was checked by electrophoresis on a 12% denaturing polyacrylamide gel. Pure tRNA^{Sec} fractions were ethanol precipitated and kept at -80°C.

Purification and crystallization of the SepSecS-tRNA^{Sec} complex. The complex was prepared by mixing SepSecS with a 1.5-fold molar excess of tRNA^{Sec}. Thus, for each SepSecS tetramer there were 6 molecules of tRNA^{Sec} available. The mixture was concentrated to ~1 mL, loaded onto a S-200pg column (XK 16/60)

and eluted with 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DTT, 10 μ M PLP. The first peak contained the SepSecS-tRNA^{Sec} complex, whereas the second peak contained the unbound tRNA^{Sec}. The complex was concentrated to \sim 7.5 mg/mL. The complex was crystallized by sitting-drop vapor-diffusion method at 12°C by mixing equal volumes of the complex sample with the well buffer containing 0.3 M tri-lithium citrate, 18% (w/v) PEG 3,350. Addition of various additives to the mother liquor such as sorbitol, D-galactose, NDSB-201, Cymal-7, MPD and phenol improved both the crystal size and the quality of diffraction.

Ligand soaks and data collection. A mixture of thiophosphate and phosphoserine was added to the binary complex crystals to the final concentration of 1 mM and incubated for 1h prior to freezing. Crystals were cryoprotected for 5 minutes in the presence of 20% (v/v) ethylene-glycol and then frozen in liquid propane. Data were collected at the NE-CAT 24IDE beamline (APS, Argonne National Laboratory, Chicago) and processed using HKL2000 (Supplementary Table 1) (13).

Structure determination and refinement. The crystal structure of the complex between human SepSecS and tRNA^{Sec} was solved by molecular replacement using Phaser (14) and by using a monomer of the murine SepSecS (PDBID: 3BC8) (15) as a search model. The asymmetric unit contained four SepSecS molecules, which we shall refer to as the crystallographic tetramer (Fig. S1A). After a round of rigid-body refinement strong positive electron density peaks for tRNA^{Sec}, Thiop and Sep appeared in the Fo-Fc electron density map. The tRNA^{Sec} molecule was bound to the crystallographic tetramer in two alternate orientations and both orientations were modeled and refined (Fig. S1A). The actual physiological tetramer contains a half of the crystallographic tetramer and two other symmetry related SepSecS molecules (Fig. S1B). Consequently, the actual physiological SepSecS tetramer has two tRNA^{Sec} molecules bound each with half occupancy. The physiological tetramer is discussed in the manuscript. Structure refinement was done in Phenix (16), while model and map inspection was done in Coot (17). The final model had an excellent geometry with 94% and 6% residues in preferred and allowed regions of the Ramachandran plot, respectively. The final R_{cryst} was 20.3% and R_{free} was 23.8% (Table 1). All figures were generated in Pymol (18).

Modeling of the SepSecS-tRNA^{Asp} and SerRS-tRNA^{Sec} complexes. A model of SepSecS-tRNA^{Asp} was generated using the crystal structure of AspRS-tRNA^{Asp} (PDBID: 1C0A) (19). tRNA^{Asp} was superimposed onto tRNA^{Sec} using two approaches. On one hand the atoms of the sugar-phosphate backbone of the acceptor-, T Ψ C- and anticodon arms were used for superpositioning, whereas in the other approach, G73 of tRNA^{Asp} was first modeled as found in tRNA^{Sec} and then only the nucleotides 1-3 and 70-73 were superimposed. The CCA-end of both tRNA^{Asp} and tRNA^{Sec} was excluded from the calculation. A model of SerRS-tRNA^{Sec} complex was generated using the crystal structure of SerRS-tRNA^{Ser} complex (PDBID: 1SER)

(20). The atoms of the sugar-phosphate backbone of tRNA^{Sec} were superimposed onto the corresponding atoms of tRNA^{Ser}. The tip of the acceptor arm (nucleotides 1-2, 71-76) was excluded from the calculation. All calculations were done in Lsqman (21).

Modeling of CCA-Sep^P in the catalytic site of SepSecS. Modeling of CCA-Sep^P was done in Coot (17). The O3'-hydroxyl group of A76 replaced the carboxyl oxygen of Sep^P, thus forming an ester bond. The base of A76 was positioned between the side-chains of Arg97 and Lys173, while the side-chain of Arg97 and the entire P-loop was kept in the same conformation as found in the catalytic active site.

***In vivo* SepSecS assay.** The human SepSecS mutants (R26A, R26E, K38A, K38E, R75A, R97A, S98A, Q105A, K173A, K173M, N252A, N252D, R271A, R271E, R313A, T397A, T397E, T397V, R398A and R398E) were generated using the QuikChange site-directed mutagenesis kit (Stratagene) and cloned into the pET15b vector with an N-terminal His-tag. An N-terminal Δ_{1-37} SepSecS deletion mutant was constructed by PCR with the pET15b-*sepsecS* plasmid as the template and with the following primers: 5'-GGAATTCATATGAAGGGCAAGTGTCCAGAGAATGGCTGGGATGAAAGTACAC-3', and 5'-CGCGGATCCTCATGAAGAAGCATCCTGGTATGTGTCAAGAAGTACATTATC-3'. The resulting DNA fragment was subcloned into the pET20b vector (Novagen) with a C-terminal His-tag. Construction of the *E. coli* Δ *selA* deletion strain JS1 (DE3) and cloning of the *M. jannaschii* PSTK gene into the pACYC-Duet vector were described previously (11). The human wild type and mutant SepSecS genes were transformed into the *E. coli* Δ *selA* deletion strain JS1 (DE3) with or without the *M. jannaschii* PSTK gene. Aerobic overnight cultures were streaked on LB-agar plates and/or M9 minimal medium-agar plates supplemented with 0.01 mM IPTG, 1 μ M Na₂MoO₄, 1 μ M Na₂SeO₃ as described previously (2, 11). The LB-agar plates also contained 50mM sodium formate. The cells were grown anaerobically for 5h at 37°C and 20h at 30°C. The plates were then overlaid with agar containing 1 mg/ml benzyl viologen, 0.25 M sodium formate, and 25 mM KH₂PO₄, pH 7.0. The appearance of a blue/purple color is the indication of active formate dehydrogenase H.

***In vitro* SepSecS assay.** For this assay, following *in vitro* transcription as described above, the human tRNA^{Sec} transcript was purified by electrophoresis on a 12% denaturing polyacrylamide gel. Full-length tRNA was eluted and desalted on Sephadex G25 Microspin columns (Amersham). The tRNA^{Sec} was refolded by heating for 5 min. at 70°C in buffer containing 10 mM Tris-HCl (pH 7.0), followed by addition of 5 mM MgCl₂ and immediate cooling on ice. Refolded tRNA^{Sec} (10 μ M) was serylated and phosphorylated by *M. maripaludis* SerRS (5 μ M) and *M. jannaschii* PSTK (3 μ M) for 75 min. at 37°C in buffer containing 50 mM HEPES (pH 7.2), 10 mM MgCl₂, 20 mM KCl, 1 mM DTT, 10 mM ATP and 300 μ M

[¹⁴C]Ser. After phenol/chloroform extraction Sep-tRNA^{Sec} was purified by application on a Sephadex G25 Microspin column (Amersham) and ethanol precipitation. The conversion of Sep to Cys was done under anaerobic conditions using wild-type human SepSecS and human SepSecS treated with either hydroxylamine or NaBH₄. For the hydroxylamine treatment the SepSecS sample was dialyzed for 5 hr against the buffer containing 50 mM HEPES pH 7.2, 250mM NaCl, 1mM DTT, 7.5 mM hydroxylamine, whereas for the borohydride treatment the SepSecS sample was incubated in the dark for 1 hr in the presence of 100 mM NaBH₄. The excess of hydroxylamine and borohydride was removed from the samples by dialysis against 50 mM HEPES pH 7.2, 250 mM NaCl, 1 mM DTT. SepSecS (20 μg) was incubated with 10 μM Sep-tRNA^{Sec} in buffer containing 50 mM HEPES pH 7.2, 20 mM KCl, 10 mM MgCl₂, 1 mM DTT and 250 μM sodium thiophosphate. All buffers were prepared anaerobically and the reaction was carried out at 37°C inside an anaerobic chamber over 40 min. The reaction was stopped by phenol/chloroform extraction, the aqueous phase was applied to a Sephadex G25 Microspin column (Amersham) and the eluted aminoacyl-tRNAs were ethanol precipitated. The purified aa-tRNAs were deacylated in 20 mM NaOH for 10 min at 22°C. The released amino acids were oxidized with performic acid as described previously (22). The final samples were spotted onto silica gel 60 TLC aluminium sheets (Merck) that were developed in 85% ethanol. To detect the labelled amino acids the TLC plate was dried and then exposed on an imaging plate (FujiFilms) for 60 hr. The imaging plate was scanned using a Molecular Dynamics Storm 860 scanner.

References

1. C. Stürchler, E. Westhof, P. Carbon, A. Krol, *Nucleic Acids Res.* **21**, 1073 (1993).
2. Y. Araiso *et al.*, *Nucleic Acids Res.* **36**, 1187 (2008).
3. H. Oshikane *et al.*, *Science* **312**, 1950 (2006).
4. A. Nakamura, M. Yao, S. Chimnarongk, N. Sakai, I. Tanaka, *Science* **312**, 1954 (2006).
5. R. L. Sherrer, J. M. Ho, D. Söll, *Nucleic Acids Res.* **36**, 1871 (2008).
6. M. Bailly, M. Blaise, B. Lorber, H. D. Becker, D. Kern, *Mol. Cell* **28**, 228 (2007).
7. C. M. Zhang, C. Liu, S. Slater, Y. M. Hou, *Nat. Struct. Mol. Biol.* **15**, 507 (2008).
8. X. M. Xu *et al.*, *PLoS Biol.* **5**, e4 (2007).
9. G. Schneider, H. Kack, Y. Lindqvist, *Structure* **8**, R1 (2000).
10. A. C. Eliot, J. F. Kirsch, *Annu. Rev. Biochem.* **73**, 383 (2004).
11. J. Yuan *et al.*, *Proc. Nat. Acad. Sci. USA* **103**, 18923 (2006).
12. R. L. Sherrer, P. O'Donoghue, D. Söll, *Nucleic Acids Res.* **36**, 1247 (2008).
13. Z. Otwinowski, W. Minor, Eds., *Processing of X-ray Diffraction Data Collected in Oscillation Mode*, (Academic Press, New York, 1997) Vol. **276**, 307.
14. A. J. McCoy *et al.*, *J. Appl. Cryst.* **40**, 658 (2007).
15. O. M. Ganichkin *et al.*, *J. Biol. Chem.* **283**, 5849 (2008).
16. P. D. Adams *et al.*, *Acta Cryst. D* **58**, 1948 (2002).
17. P. Emsley, K. Cowtan, *Acta Cryst. D* **60**, 2126 (2004).
18. W. L. DeLano, *The Pymol Molecular Graphics System*. (DeLano Scientific, LLC, San Carlos, CA).
19. S. Eiler, A. Dock-Bregeon, L. Moulinier, J. C. Thierry, D. Moras, *EMBO J.* **18**, 6532 (1999).
20. V. Biou, A. Yaremchuk, M. Tukalo, S. Cusack, *Science* **263**, 1404 (1994).
21. G. J. Kleywegt, T. A. Jones, *CCP4/ESF-EACBM Newsletter on Protein Crystallography* **31**, 9 (1994).
22. A. Sauerwald *et al.*, *Science* **307**, 1969 (2005).

Supporting Figure legends

Figure S1. Asymmetric unit of the SepSecS-tRNA^{Sec} complex crystal. (A) Ribbon diagram of the crystallographic tetramer composed of two halves of the physiological SepSecS tetramer with one tRNA^{Sec} molecule bound in two alternate orientations. Catalytic monomers are colored in blue, non-catalytic monomers are shown in red, and the two conformations of tRNA^{Sec} are colored in green and orange. The definition of the monomers is as in the text. The cofactor PLP is shown as sticks and colored in red and blue in the catalytic and non-catalytic active sites, respectively. (B) A half of the asymmetric unit and two symmetry related molecules (shown in light blue and light pink) make up the physiological SepSecS tetramer.

Figure S2. The SepSecS tetramer can accommodate four tRNA^{Sec} molecules. (A) Ribbon diagram of the physiological SepSecS tetramer in a putative complex with four tRNA^{Sec} molecules bound. The catalytic dimer (blue) with two tRNA^{Sec} molecules (green) bound was superimposed onto the non-catalytic dimer (red) and the complex was generated. The generated tRNA^{Sec} molecules are shown in orange. (B) The view in B is rotated ~90° clock-wise around the vertical axis relative to A. The SepSecS-tRNA^{Sec} complex is in the same orientation as in Figure 1 of the manuscript.

Figure S3. Interactions between SepSecS and the TYC- and variable arms of tRNA^{Sec}. (A) The ribbon diagram of the complex showing interactions between parts of the variable arm of tRNA^{Sec} (orange) and the $\alpha 9$ helix (red) of SepSecS. The rest of the tRNA^{Sec} is green. The view is rotated ~90° clockwise around the vertical axis relative to Fig. 1A. (B) Arg271 ($\alpha 9$) in the non-catalytic dimer interacts with the variable arm (nucleotides C46L-A47). (C) The side-chains of Lys38 and Lys40 ($\alpha 1$ - $\alpha 2$ loop) interact with the TYC arm. Lys38 and Lys40 interact with C64 and G50, respectively. In B and C, tRNA^{Sec} is green and the protein side-chains are gold.

Figure S4. *In vivo* formation of Sec-tRNA^{Sec} by SepSecS mutants and *in vitro* conversion of Sep-tRNA^{Sec} to Cys-tRNA^{Sec} by wild type and PLP-cross-linked SepSecS. (A) *In vivo* assays of SepSecS mutants. Formation of Sec-tRNA^{Sec} *in vivo* is assayed by the ability of the wild-type human SepSecS and its mutant variants (N-terminal deletion $\Delta 1$ -37, R398A, R398E, R26A, R26E, K38A, K38E, T397A, T397E, T397V, R271A, R271E, K173A, K173M, S98A, R97A, Q105A, R313A, R75A, N252 and N252D) to restore the benzyl viologen reducing activity of the selenoprotein FDH_H in the *E. coli selA* deletion strain JS1. The *M. jannaschii* PSTK is co-transformed, except where indicated with a “-”, to provide the necessary Sep-tRNA^{Sec} intermediate. (B) *In vitro* assays of SepSecS. Phosphorimages of TLC separation of [¹⁴C]Sep and [¹⁴C]Cys recovered from the aminoacylated-tRNAs of the SepSecS activity assays. Cysteine

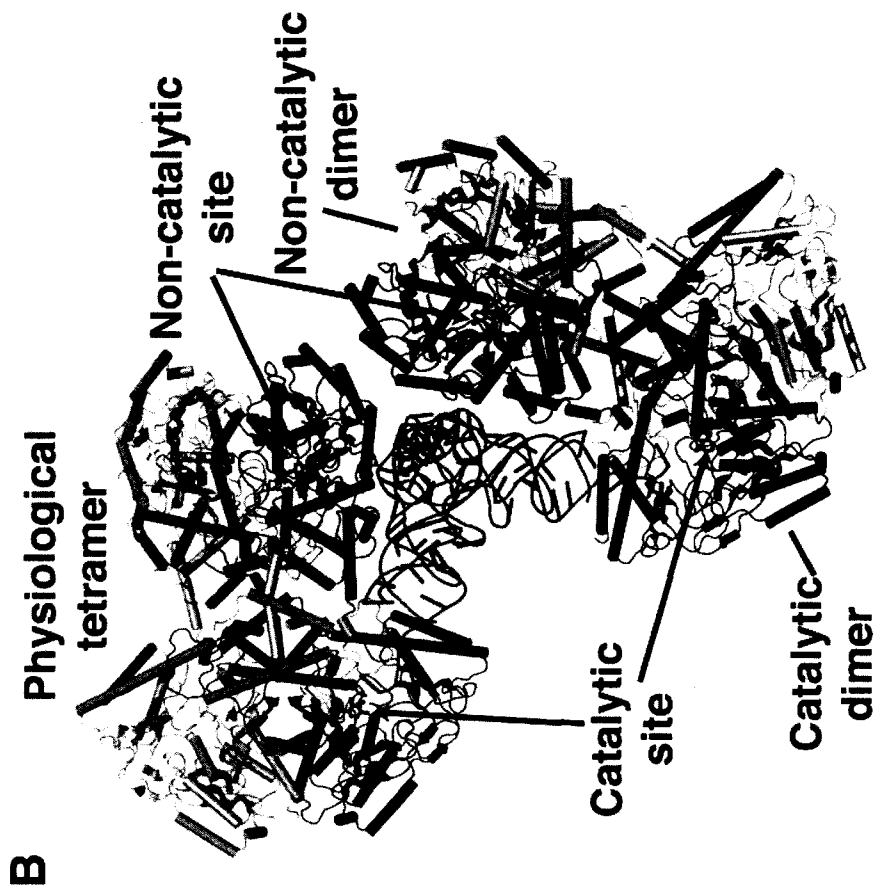
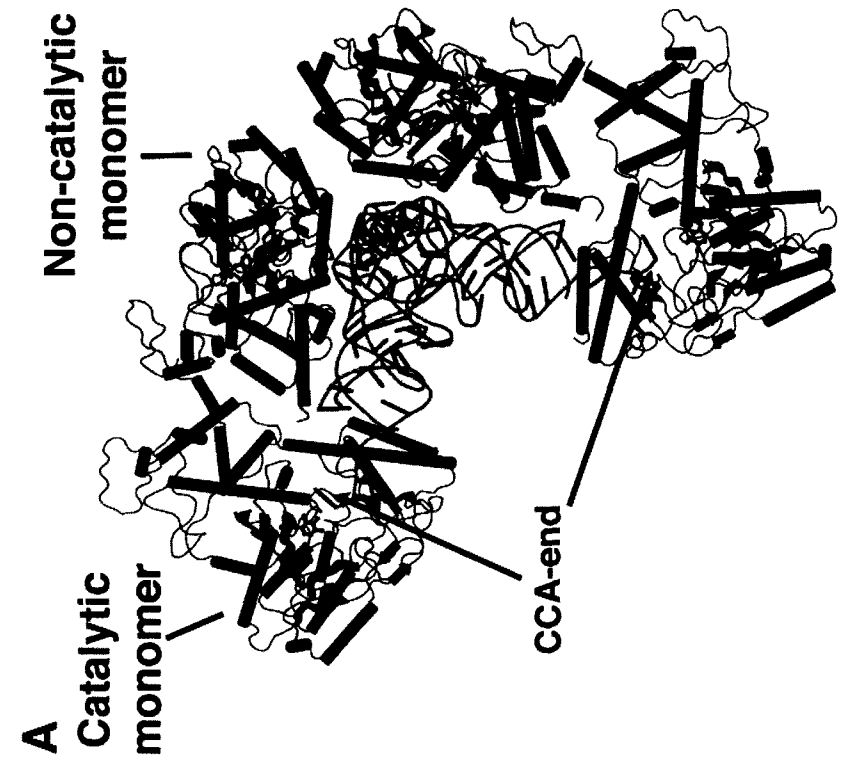
was analyzed in its oxidized form as cysteic acid (Cya). Lane 1, Ser marker; lane 2, Sep marker; lane 3, Cys marker; lane 4, Sep-tRNA^{Sec} with human SepSecS in the absence of thiophosphate; lane 5, Sep-tRNA^{Sec} with wild-type human SepSecS; lane 6, Sep-tRNA^{Sec} with human SepSecS treated with 7.5 mM hydroxylamine; lane 7, Sep-tRNA^{Sec} with human SepSecS treated with 100 mM sodium borohydride.

Figure S5. Selectivity of SepSecS and SerRS. (A) Modeling of the complex between SepSecS and canonical tRNAs. tRNA^{Asp} (orange) was superimposed onto tRNA^{Sec} (green) using the sugar-phosphate backbone atoms of the acceptor-, TΨC- and anticodon arms. The long variable arm of tRNA^{Sec} interacts with the non-catalytic dimer of SepSecS (shades of red). The box delineates the close-up view shown in B. **(B)** The short acceptor-TΨC arm of canonical tRNAs prevents binding to SepSecS. G73 of tRNA^{Asp} (orange) cannot reach Arg398. **(C)** Steric clashes (asterisk) between the α 1 helix of the non-catalytic-dimer of SepSecS (shades of red) and the TΨC arm of tRNA^{Asp} (orange) prevent the SepSecS-tRNA^{Asp} complex formation. G73 from tRNA^{Asp} was modeled to form hydrogen bonds with Arg398 using its Hoogsteen face. Only the tip of the acceptor arm (nucleotides 1-3 and 70-73) was used in the superposition. Arrows show the movement of both the anticodon and TΨC arms in tRNA^{Asp} relative to its orientation shown in A. **(D)** Modeling of the SerRS-tRNA^{Sec} complex. tRNA^{Sec} was superimposed onto the tRNA^{Ser} that is in complex with SerRS. SerRS is a non-discriminating aminoacyl-tRNA synthetase because it recognizes variable arms of both tRNA^{Ser} (red) and tRNA^{Sec} (green).

Table 1
Data collection and refinement statistics

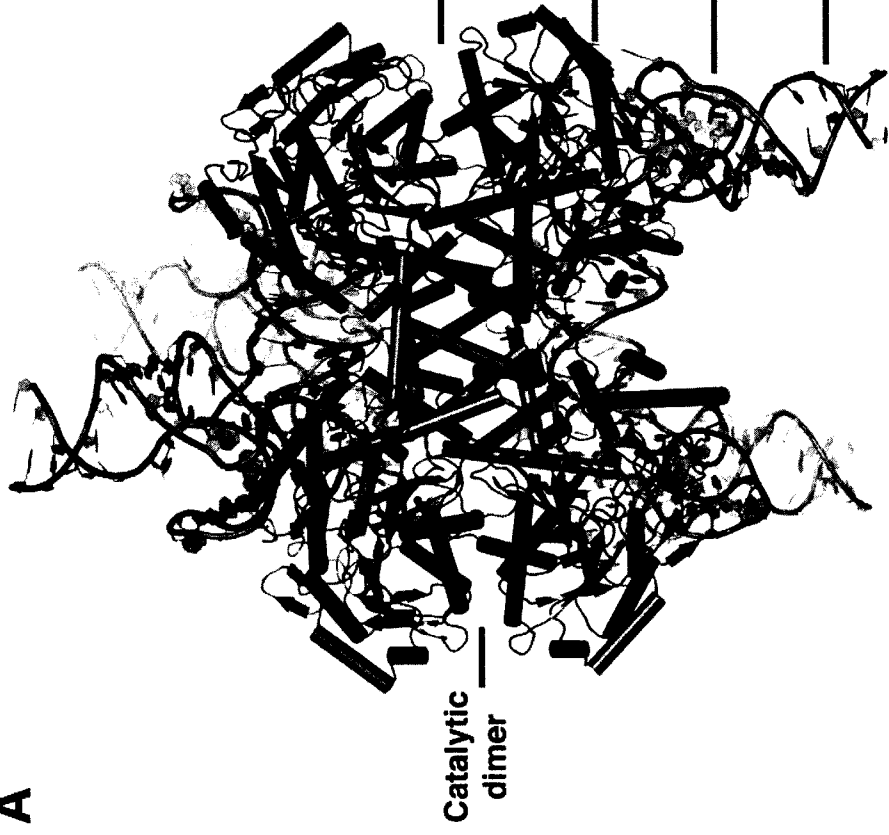
Crystal	
Space group	P3₁12
Cell dimensions	$a=b=166.8 \text{ \AA}$, $c=236.3 \text{ \AA}$, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$
Data collection	
Resolution limit (\AA)	40.0-2.8
Unique reflections	89,336
Completeness (overall / last shell; %)	97 (94)
R_{sym} (overall / last shell; %)	17.0 (94.0)
I/σ I (overall / last shell)	6 (1)
Redundancy (overall / last shell)	5.4 (4.3)
Refinement	
Average B-factor (\AA^2)	
Protein	56.5
Nucleic acid	94.3
Number of atoms (protein and nucleic acid)	17,150
Number of atoms (ligand)	141
Number of solvent molecules	301
R_{cryst} ($F > 0.5$; %)	20.3
R_{free} ($F > 0.5$; %)	23.8
R.m.s. deviations from ideality	
Bond lengths (\AA)	0.006
Bond angles ($^\circ$)	1.126

Supporting Figure 1

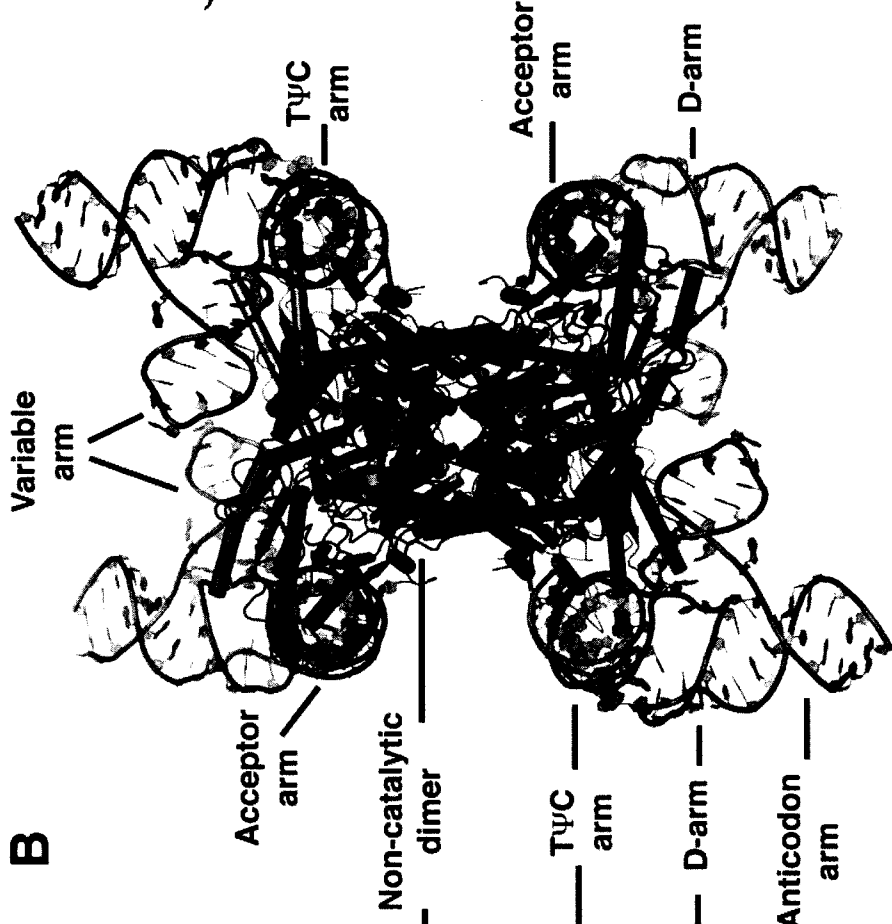


Supporting Figure 2

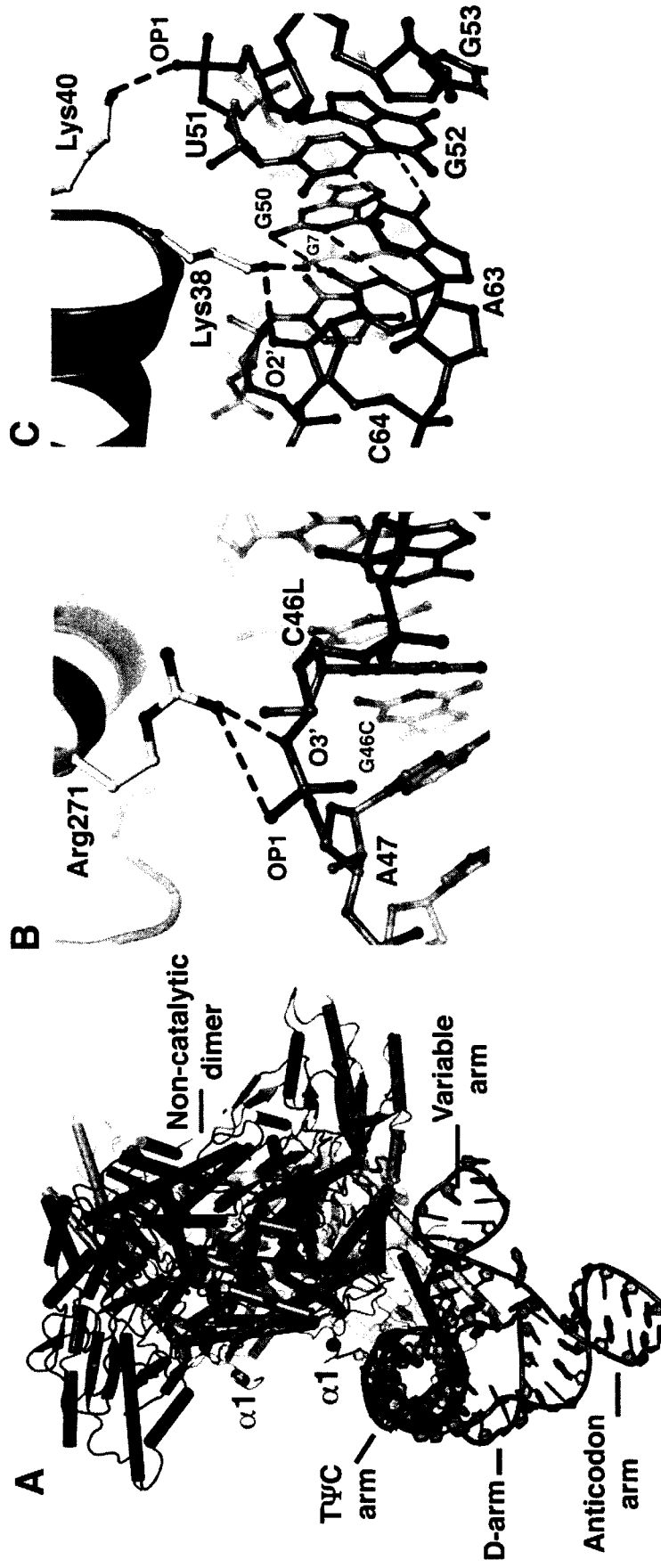
A



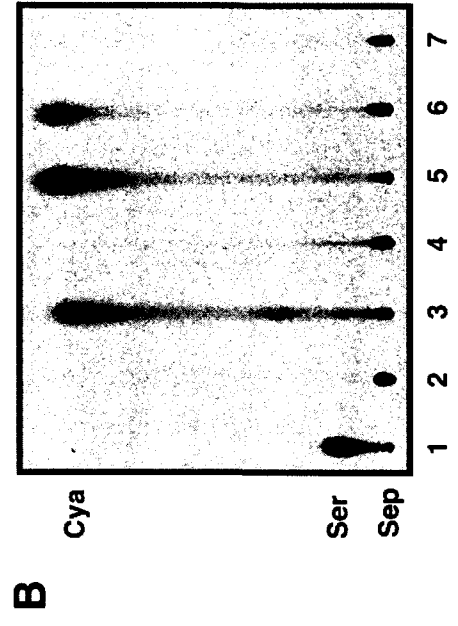
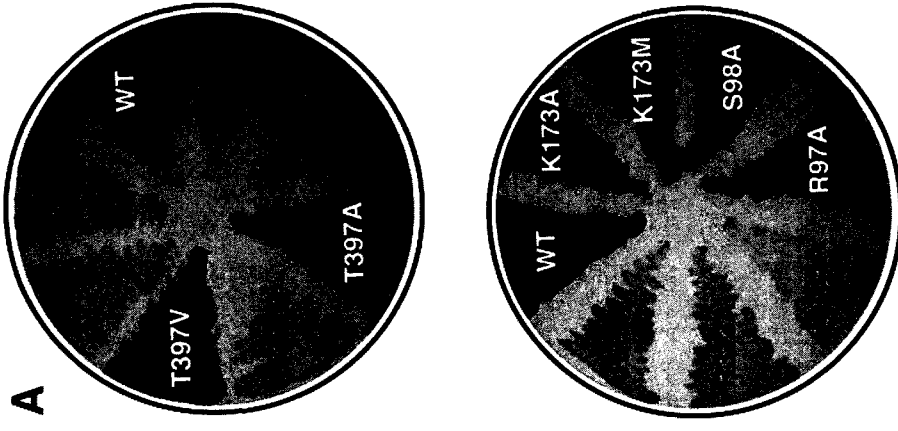
B



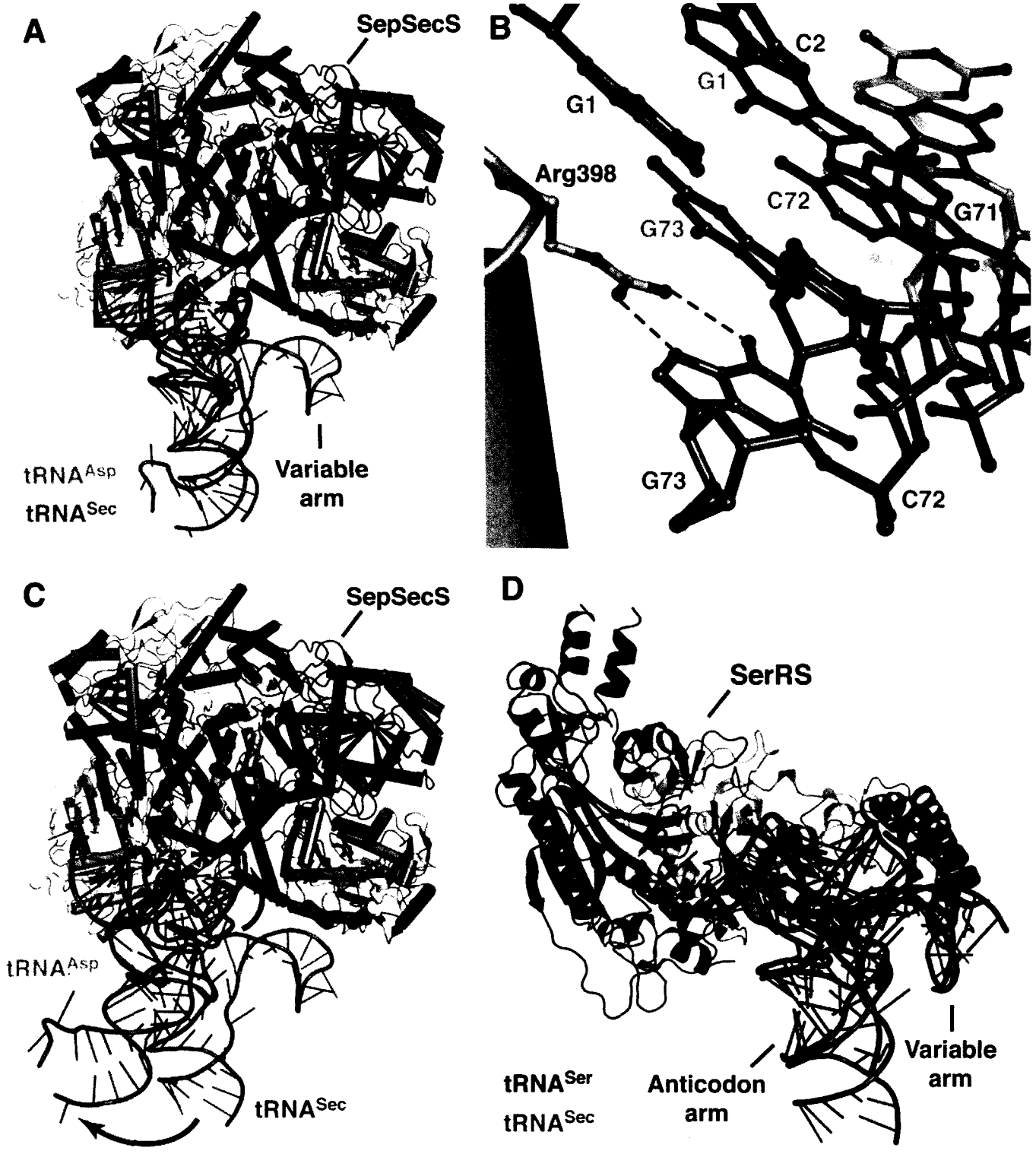
Supporting Figure 3



Supporting Figure 4



Supporting Figure 5





Review

Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems

Jing Yuan^{a,1,*}, Patrick O'Donoghue^{a,1,*}, Alex Ambrogelly^c, Sarath Gundllapalli^c, R. Lynn Sherrer^a, Sotiria Palioura^a, Miljan Simonović^d, Dieter Söll^{a,b}

^a Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114, USA

^b Department of Chemistry, Yale University, New Haven, CT 06520-8114, USA

^c Schering-Plough Corporation, Kenilworth, NJ 07033-0530, USA

^d Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, USA

ARTICLE INFO

Article history:

Received 21 October 2009

Revised 3 November 2009

Accepted 4 November 2009

Available online 11 November 2009

Edited by Michael Ibba

Keywords:

o-phosphoseryl-tRNA^{Sec} kinase

Sep-tRNA:Sec-tRNA synthase

Pyrrolysyl-tRNA synthetase

Stop codon re-coding

Natural suppression

ABSTRACT

Selenocysteine and pyrrolysine, known as the 21st and 22nd amino acids, are directly inserted into growing polypeptides during translation. Selenocysteine is synthesized via a tRNA-dependent pathway and decodes UGA (opal) codons. The incorporation of selenocysteine requires the concerted action of specific RNA and protein elements. In contrast, pyrrolysine is ligated directly to tRNA^{Pyl} and inserted into proteins in response to UAG (amber) codons without the need for complex re-coding machinery. Here we review the latest updates on the structure and mechanisms of molecules involved in Sec-tRNA^{Sec} and Pyl-tRNA^{Pyl} formation as well as the distribution of the Pyl-decoding trait.

© 2009 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Selenocysteine biogenesis

Selenocysteine (Sec) is the major biological form of the element selenium, which in trace amounts is essential for human health. Sec is incorporated into polypeptides to form selenoproteins during translation. The 21st amino acid is typically found in catalytic centers of selenoproteins where it plays a functionally essential role. Unlike most amino acids, Sec is universally synthesized on its cognate tRNA [1–4]. During translation, selenocysteinyl-tRNA^{Sec} (Sec-tRNA^{Sec}) is delivered to the ribosome by a specific translation factor that requires a characteristic stem-loop structure in the mRNA to actively recode an in-frame UGA from stop codon to Sec sense codon. The human genome encodes only 25 selenoproteins [5], yet variations in these Sec-containing proteins or their synthetic machinery is linked to a range of human disorders including cancer and numerous diseases affecting the nervous, immune, and endocrine systems [6].

The machineries to synthesize Sec and incorporate it into selenoproteins are divergent in bacteria compared to archaea and

eukaryotes. In bacteria, serine (Ser) as the precursor of Sec is initially attached to tRNA^{Sec} by seryl-tRNA synthetase (SerRS). The resulting Ser-tRNA^{Sec} is then converted to Sec-tRNA^{Sec} by selenocysteine synthase (Sela) in the presence of the selenium donor selenophosphate. This pathway has been well characterized in *Escherichia coli* [7] and extensively reviewed before [8,9].

The pathway for Sec biosynthesis in archaea and eukaryotes (Fig. 1) was revealed only in the last few years. The missing component was an archaeal/eukaryotic analog of Sela, since no clear sequence-based homolog could be found. An additional enzymatic step is involved in Sec biosynthesis in archaea and eukaryotes. *O*-phosphoseryl-tRNA kinase (PSTK) [10] catalyzes the phosphorylation of Ser-tRNA^{Sec} to form *O*-phosphoseryl-tRNA^{Sec} (Sep-tRNA^{Sec}). The Sep-tRNA:Sec-tRNA synthase (SepSecS), an independently evolved protein that is distantly related to Sela [11], then forms the final product Sec-tRNA^{Sec} from selenophosphate and Sep-tRNA^{Sec} [2–4]. Phylogenetic analysis indicated that PSTK and SepSecS co-evolved and are restricted to the archaeal and eukaryotic domains [2,12]. An interesting similarity, as reviewed previously [13], exists between the archaeal and eukaryotic Sec biosynthetic pathway and archaeal tRNA-dependent cysteine biosynthesis, which also proceeds via a Sep-tRNA intermediate [14]. In the first section of this paper, we focus on recent biochemical and structural work that further elucidated the mechanism and specificity

* Corresponding authors.

E-mail addresses: jing.yuan@yale.edu (J. Yuan), patrick.odonoghue@yale.edu (P. O'Donoghue).

¹ Both authors contributed equally.

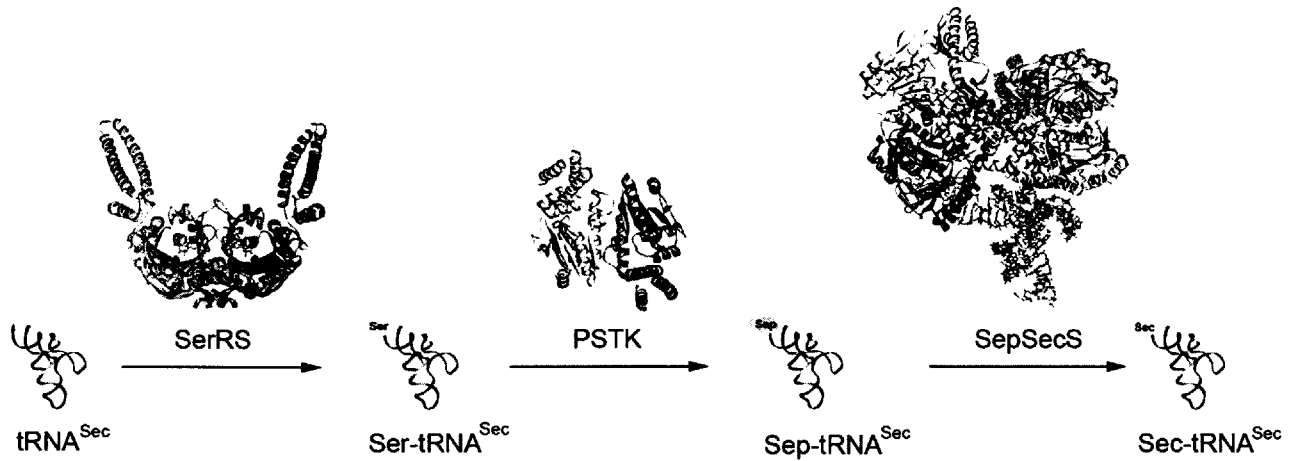


Fig. 1. The Sec biosynthesis pathway in archaea and eukaryotes. Sec is synthesized on tRNA^{Sec} in three steps. (1) The unacylated tRNA^{Sec} is serylated by SerRS; (2) the resulting Ser-tRNA^{Sec} is phosphorylated by PSTK forming Sep-tRNA^{Sec}; (3) the phosphorylated intermediate is converted to the final product Sec-tRNA^{Sec} by SepSecS. The crystal structures of tRNA and enzymes in this pathway are presented: tRNA^{Sec} (ribbon) from *Homo sapiens* [19], SerRS (ribbon) with AMP (stick) from *Pyrococcus horikoshii* [64], PSTK (ribbon) from *M. jannaschii* [26], and SepSecS (ribbon) with tRNA^{Sec} (stick, only one tRNA is shown) from *Homo sapiens* [18].

of the tRNA and enzymes involved in Sec biosynthesis in archaea and eukaryotes.

2. tRNA^{Sec} has a distinct structure

tRNA^{Sec} was identified more than two decades ago [1]. It is the longest tRNA with an extended acceptor stem resulting from an abnormal RNase P cleavage specificity [15]. tRNA^{Sec} has an 8-bp acceptor stem and 5-bp T-stem (a 8/5 secondary structure) in bacteria and a 9/4 arrangement in archaea and eukaryotes, and both can fold into a 13-bp long acceptor-TΨC helix (Fig. 2A). In contrast,

canonical tRNAs typically have a 7-bp acceptor stem and a 5-bp T-stem forming a 12-bp acceptor-TΨC helix. Several other features of tRNA^{Sec}, including an elongated D-stem (6-bp instead of 4-bp), a smaller D-loop (4-bp instead of 8-bp), a long variable arm, and the absence of the highly conserved U8 residue, make tRNA^{Sec} distinct from canonical tRNAs.

The tertiary structures of *E. coli* and eukaryotic tRNA^{Sec} were investigated by chemical and enzymatic probing during the 1990s [16,17]. Very recently, the first tRNA^{Sec} crystal structure was solved. The unacylated human tRNA^{Sec} transcript was crystallized in complex with human SepSecS [18] and in an unbound state

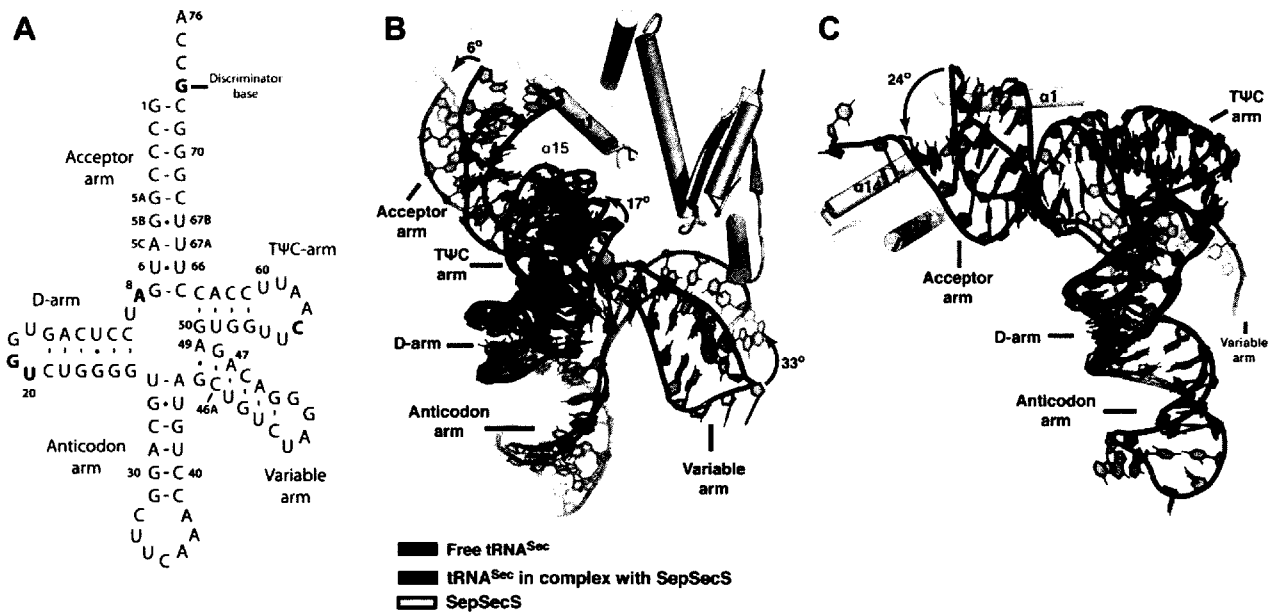


Fig. 2. Binding to SepSecS promotes a conformational change in tRNA^{Sec}. (A) The secondary structure of human tRNA^{Sec} with bases mentioned in the text highlighted in bold. The correct tRNA^{Sec} sequence is shown here (Fig. 2C in Ref. [18] omitted a base from the D-arm). (B) Superposition of the sugar-phosphate backbone of both the D- and anticodon arms of free tRNA^{Sec} (blue) on the corresponding atoms of tRNA^{Sec} complexed with SepSecS (red) reveals a conformational change in the tRNA molecule on binding to the enzyme. The variable arm of tRNA^{Sec} rotates by 33°, the T arm swings 17° and the acceptor arm rotates by 6° around the axis that runs parallel to both the D- and anticodon arms. The free tRNA^{Sec} conformer cannot bind to SepSecS because the tip of its acceptor arm would clash with the helix $\alpha 1$ and the variable arm would be positioned away from the helix $\alpha 9$. (C) The acceptor arm also slides down towards the anticodon arm on binding to SepSecS through a 24° rotation around the axis that is parallel to the variable arm. This movement positions G73 to interact with Arg398 in the helix $\alpha 14$ and orients the CCA end toward the active site. Free tRNA^{Sec} is blue, tRNA^{Sec} complexed with SepSecS is red and SepSecS is beige. Only the secondary structure elements of SepSecS that interact with tRNA are shown. The view is first rotated $\sim 30^\circ$ clockwise around the horizontal axis, and then another 30° anticlockwise around the vertical axis relative to panel B.

[19]. The tertiary structure confirms the predicted distinct features of the acceptor, T and D-stems. Compared to canonical tRNAs, the overall structure of tRNA^{Sec} is less compact due to the lack of tertiary interactions from the D-arm and the variable arm, i.e., absence of the G15:C48 base pair, which creates a hole in the tertiary core [18,19]. The D-stem does not stack perfectly along the axis formed by the anticodon stem and the T-loop and shifts to the minor groove side of the tRNA. The long variable arm protrudes from the tertiary core on the side of tRNA^{Sec} opposite from the D-arm [18,19]. Remarkably, despite different secondary and tertiary interactions, the relative position of the T-stem and the variable arm in the free tRNA^{Sec} is almost indistinguishable to that in tRNA^{Ser}. The variable arm is one of the major identity elements for the recognition of tRNA^{Ser} and tRNA^{Sec} by seryl-tRNA synthetase (SerRS), and these observations in part explain the dual specificity of SerRS for tRNA^{Sec} and tRNA^{Ser} [18]. Another model of the SerRS-tRNA^{Sec} complex suggests that SerRS may also recognize both the G19:C56 base pair and the discriminator base of tRNA^{Sec} [19]. Thus, all the major identity elements for tRNA serylation are present and can be recognized in tRNA^{Sec}.

Uridine at position 8 is highly conserved and forms a tertiary base pair with A14 in all other mature tRNAs. This conserved U8:A14 base pair defines the elbow region of the tRNA [20] and U8 acts as a sensor for the response mechanism to UV exposure [21]. The tRNA^{Sec} has an adenosine at position 8 instead. An A8:A14:U21 base triple was suggested from the chemical and enzymatic probing data [17], yet A8 does not appear to interact with other bases in the two recently reported tRNA^{Sec} crystal structures [18,19]. A tertiary base interaction change is observed when comparing the SepSecS-bound and the free tRNA^{Sec} crystal structures. In the unbound tRNA^{Sec}, the base of U20 stabilizes interactions between the D and T Ψ C loops by forming a base triple with the G19:C56 Watson–Crick base pair [19]. In the SepSecS-bound state the base of U20 does not participate in the base triple. Its electron density is weak, suggesting increased mobility in this part of the D-loop. Thus, the base triple stabilizes an unbound conformation of tRNA^{Sec}, whereas on binding to SepSecS the base triple breaks apart. Besides canonical tertiary base interactions (i.e., G18:U55 and G19:C56), unique tertiary interactions are present in tRNA^{Sec} such as the U16:U59 base pair, which stacks on top of A15:A20a and forms a 13-bp stacking helix together with the D and anticodon stems.

Further analysis, based on superimposing the two solved structures of tRNA^{Sec}, reveals that while the individual arms of tRNA^{Sec} adopt the same structure, their relative orientations differ in the SepSecS-bound state and unbound state. Indeed, the acceptor, T- and variable arms undergo a conformational change upon binding to SepSecS (Fig. 2B): (i) the variable arm makes a 33° rotation around the axis that is parallel to the anticodon arm, (ii) the T-arm moves in the same direction, but to a somewhat lesser extent (17°), and (iii) the acceptor arm makes a 6° rotation around the anticodon stem as well as a 24° rotation around the axis that is parallel to the variable arm (Fig. 2C). The conformational change facilitates the interaction between tRNA^{Sec} and SepSecS and presumably orients the CCA end toward the active site. It will be interesting to see if tRNA^{Sec} undergoes a similar conformational change as it binds to other components of the Sec biosynthetic apparatus and the translating ribosome.

tRNA^{Sec} plays a pivotal role in the Sec biosynthesis pathway. It interacts with Sec specific proteins including PSTK and SepSecS in archaea and eukaryotes. The distinct structural features of tRNA^{Sec} are proven to be major recognition elements throughout Sec biosynthesis as detailed below. It is unknown, however, if these features of tRNA^{Sec} are also important for Sec incorporation, which requires the participation of multiple protein factors in Sec decoding during translation.

3. PSTK is a tRNA^{Sec}-dependent kinase

In archaea and eukaryotes, PSTK catalyzes the second step of Sec biosynthesis (Fig. 1, center), the phosphorylation of Ser-tRNA^{Sec} to Sep-tRNA^{Sec} in a reaction requiring ATP and magnesium. Such kinase activity was first observed in rat and rooster liver lysate almost 40 years ago [22,23]. In 2004, the identity of this elusive kinase as PSTK was finally revealed in mouse via a comparative genomic approach [10]. Subsequently the archaeal homolog was shown to have the same activity [24], and the biochemical and structural properties of the *Methanocaldococcus jannaschii* PSTK (MjPSTK) have been extensively studied [12,24–26].

MjPSTK is a homodimeric enzyme with each monomer consisting of a N-terminal kinase domain and a C-terminal domain that is putatively involved in tRNA binding. PSTK is a member of the P-loop kinase family with the conserved Walker A and B motifs and the RX₃R motif that are responsible for the recognition of the Mg²⁺-ATP complex [12,26]. Mutations to either the Walker A or B motifs ablate PSTK activity both in vitro and in vivo [26]. Mutagenesis also revealed that the distal Arg residue in the RX₃R is critical for catalysis both in vivo and in vitro. Interestingly, the kinase activity is not highly specific for the phosphate donor, since all four rNTPs and even dATP are competent substrates [12]. Structural analysis of the PSTK:AMPPNP complex supported this observation since the adenosine ring makes few specific contacts, with the exception of a cation- π interaction involving the RX₃R motif [26].

The conserved Asp residue (Asp41) in the Walker B motif plays an essential role in MjPSTK catalysis. Mutation of Asp41 to Asn leads to total loss of activity [26]. The structure shows that Asp41 stabilizes the ATP bound Mg²⁺ ion via a water mediated hydrogen bond, while modeling of the Ser substrate suggests direct contact between Asp41 and the hydroxyl group of Ser [26]. Based on these reports, we can propose a plausible reaction mechanism for PSTK. The reaction begins when both substrates, Ser-tRNA^{Sec} and Mg²⁺-ATP, are bound to the enzyme. Asp41 then attracts a proton from the hydroxyl group of the Ser in the substrate. The deprotonated Ser nucleophilically attacks the γ -phosphate of the ATP yielding a pentavalent transition-state intermediate. Finally, the phosphodiester bond between the β and γ phosphate in the ATP is broken to complete the transphosphorylation reaction, and the product Sep-tRNA^{Sec} is formed. Mg²⁺ aids in the departure of the leaving group ADP.

The enzyme activity of PSTK is strictly tRNA^{Sec}-dependent. PSTK does not hydrolyze ATP in the absence of tRNA nor in the presence of Ser-tRNA^{Ser}. The binding of tRNA^{Sec}, however, promotes ATP hydrolysis [12]. This suggests that tRNA^{Sec} might play an essential role in positioning the Ser moiety for initiating phosphoryl transfer. Compared to aminoacyl-tRNA synthetases, PSTK has approximately 20-fold higher affinity toward its substrate, Ser-tRNA^{Sec} ($K_m = 40$ nM) [12], which may compensate for the low abundance of tRNA^{Sec} in vivo. The concentration of tRNA^{Sec} in vivo is at least 10-fold lower than tRNA^{Ser} in tRNA^{Sec}-rich tissues such as liver, kidney and testis in rat [27].

The crystal structure of PSTK reveals an unusual large central groove that is formed in the MjPSTK dimer interface, a feature not observed in other P-loop kinases previously. The central groove exposes the monomer active sites at either end, and contains many positively charged residues. Computational docking suggested that the central groove provides a complementary surface for the two tRNA^{Sec} substrates. The model places archaeal PSTK identity elements (G2:C71 and the C3:G70 [25]) within contact of the protein dimer interface. Interestingly, the second base pair in the acceptor stem is highly conserved as C2:G71 in eukaryotic tRNA^{Sec}, and mutation of G2:C71 to C2:G71 in archaeal tRNA^{Sec} resulted in a Ser-tRNA^{Sec} variant that is phosphorylated inefficiently [25].

Moreover, the eukaryotic PSTK has been reported to recognize the unusual D-arm of tRNA^{Sec} as the major identity element for phosphorylation [28], and phylogeny indicates a deep divide between the archaeal and eukaryotic PSTK proteins [12]. Given these observations, it is quite likely that the eukaryotic PSTK/tRNA complex may involve interactions that are distinct from its archaeal counterpart. Future co-crystal structures of tRNA^{Sec} with the archaeal and eukaryotic PSTKs will provide a definitive picture of these differences.

4. SepSecS requires Sep-tRNA^{Sec} as the precursor for Sec biosynthesis

The conversion of phosphoseryl-tRNA^{Sec} (Sep-tRNA^{Sec}) to selenocysteiny-tRNA^{Sec} (Sec-tRNA^{Sec}) is the last step of Sec biosynthesis in both archaea and eukaryotes, and it is catalyzed by O-phosphoseryl-tRNA:selenocysteiny-tRNA synthase (SepSecS). SepSecS forms its own branch in the phylogenetic tree of the fold type I family of the pyridoxal phosphate (PLP)-dependent enzymes and displays a distinct homotetrameric ([α_2]₂) quaternary structure [11]. Two active sites are formed at each homodimer interface and each active site contains a PLP-binding pocket. Thus, the tetrameric SepSecS has four active sites and four PLP molecules bound. The PLP cofactor forms a reversible Schiff-base linkage (internal aldimine) with a highly conserved Lys284. The PLP-dependent enzymes form a functionally diverse group, which is responsible for more than 140 distinct activities [29]. Remarkably, almost all the PLP-dependent enzymes utilize a catalytic mechanism that proceeds through a carbanion intermediate, which is, in turn, stabilized by PLP [30]. Based on both *in vitro* and *in vivo* activity assays, a catalytic mechanism of SepSecS that is dependent on PLP was recently proposed [11,18,31].

The tetrameric SepSecS is distinct from all other members of its family including O-phosphoseryl-tRNA:cysteiny-tRNA synthase (SepCysS). SepCysS acts on a tRNA-based substrate (Sep-tRNA^{Cys}), but it does so as a dimer [11]. Moreover, SepSecS is a highly specific enzyme that acts on Sep-tRNA^{Sec}, and not on Sep-tRNA^{Cys}, Ser-tRNA^{Sec}, or free Sep. Crystal structures of both the archaeal and murine apo-SepSecS [11,31] and that of the human SepSecS-tRNA^{Sec} complex [18] have provided insights into the substrate specificity and the catalytic mechanism of SepSecS.

SepSecS binds both the unacylated tRNA^{Sec} and Sep-tRNA^{Sec} with comparable affinities *in vitro* [4], which presumably allowed crystallization of the human complex between SepSecS and unacylated tRNA^{Sec}. The SepSecS/tRNA^{Sec} complex structure revealed that one SepSecS homodimer interacts with the sugar-phosphate backbone of both the acceptor-T Ψ C and the variable arms of tRNA^{Sec}, while the other homodimer interacts specifically with the tip of the acceptor arm through interaction between the conserved Arg398 and the discriminator base G73 of tRNA^{Sec} [18]. Thus, one homodimer measures both the length of the acceptor-T Ψ C arm and the distance between the variable arm and the CCA end, whereas the other homodimer ensures that the correct tRNA is bound to the enzyme. The latter homodimer also provides the catalytic site that would act on the Sep moiety. Finally, structural modeling revealed that canonical tRNAs, including tRNA^{Ser} (which is the closest structural homolog of tRNA^{Sec}), cannot bind to SepSecS due to the shorter acceptor-T Ψ C arm. These observations explain why only the SepSecS tetramer can bind tRNA^{Sec} and why SepSecS cannot bind canonical tRNAs.

Unlike its bacterial counterpart Sela, SepSecS cannot directly convert Ser-tRNA^{Sec} to Sec-tRNA^{Sec} and has a significantly reduced affinity for Ser-tRNA^{Sec} *in vitro* [4]. This argues that the phosphate group of Sep plays an important role in substrate binding. Indeed, the crystal structure of SepSecS/tRNA^{Sec} complex in presence of

free Sep shows extensive hydrogen bonding between the active site and the phosphate group of the free Sep [18]. This suggests that the phosphate group may serve as an anchor to bring Sep in close proximity of the cofactor PLP. However, the α -amino group of the free Sep appears quite flexible in the structure and it is not oriented properly for the nucleophilic attack on the Schiff-base [18]. Thus, the covalent attachment of Sep to tRNA^{Sec} may restrain the conformation of the Sep moiety as well as facilitate the proper orientation of the α -amino group for efficient PLP-dependent catalysis to occur. This would explain why SepSecS could not act on either free Sep or Ser-tRNA^{Sec}, as both molecules bear only one anchoring moiety.

Sec biosynthesis in all domains of life is achieved by RNA-dependent amino acid modification. As such it is reminiscent of other pathways of tRNA-dependent amino acid synthesis that form Asn-tRNA, Gln-tRNA and Cys-tRNA in many organisms [32]. The individual reactions in this stepwise process may actually be carried out in a large tRNA:protein complex. One example is the transamidosome [33], a complex of tRNA^{Asn}, aspartyl-tRNA synthetase, and Asp-tRNA^{Asn} amidotransferase, that provides an efficient route to bacterial Asn-tRNA^{Asn} formation. The complex architecture should also aid protein quality control, as it may prevent release of mis-acylated aminoacyl-tRNA that would be used in protein synthesis. A similar complex possibly involving tRNA^{Sec}, SerRS, PSTK and SepSecS may exist to facilitate efficient selenocysteine formation.

5. Pyrrolysine, the uncommon 22nd amino acid

Evidence for the genetic encoding of pyrrolysine (Pyl) is only present in 1% of all sequenced genomes thus far. The rarity of Pyl is matched by its unique chemical structure, a lysine in N^ε-linkage to a pyrroline ring. Pyl was first observed in a crystal structure at the active site of the monomethylamine methyltransferase (MtmB1) from the methanogenic archaeon *Methanosarcina barkeri* [34]. Pyl has now been characterized in three distinct methylamine methyltransferases from the *Methanosarcinaceae* that allow these methanogens to utilize the unique growth substrates mono-, di- and trimethylamines [35]. Although each of the genes encoding these methyltransferases contains an in-frame amber (TAG) codon (a stop codon in most organisms) the methylamine methyltransferase genes are read-through to the next opal or ochre stop codon and Pyl is inserted in-response to the in-frame amber codon. Due to its proximity to the active site, Pyl was hypothesized to be a key catalytic residue in these Pyl-proteins [36].

While some work has attempted to address the mechanism of stop codon suppression in Pyl-decoding organisms [37–39], it remains unclear whether all amber codons in these organisms encode Pyl or whether some are translational stop signals. In recent work [40], Pyl was also found in the tRNA^{His} guanylyltransferase (Thg1) from *Methanosarcina acetivorans*. Thg1 is responsible for addition of the key identity element for His-tRNA formation, a G1 residue ligated to the 5'-end of immature tRNA^{His}. Interestingly, Pyl does not play a role in the catalytic action of Thg1 and represents the first example of a dispensable Pyl residue. Efficient read-through of the in-frame amber codon in Thg1, without the apparent presence of re-coding signals, indicates that Pyl insertion may be more similar to natural stop codon suppression [41] than to the more elaborate re-coding of UGA codons that is known to occur for selenocysteine [9].

Previous reviews focused on how Pyl enters the genetic code [42,43], namely by the catalytic action of a pyrrolysyl-tRNA synthetase (PylRS) that ligates pyrrolysine to the amber codon decoding tRNA^{Pyl} [44,45]. Since that time the most dramatic advances have been the structural characterization of PylRS and tRNA^{Pyl}. This

work, which is reviewed in detail below, revealed the unique nature of the Pyl binding pocket and provided atomic detail for the PylRS/tRNA^{Pyl} interaction which underlies the specific and exclusive association between the protein and tRNA. This section of the review concludes with an overview of an expanded number of possible Pyl-decoding organisms uncovered from recently sequenced genomes and metagenomic surveys.

6. Recognition of Pyl by PylRS

While PylRS was first crystallized in 2006 [46], the first structure was solved independently the following year [47]. This structure of PylRS was determined in complex with Pyl and the Pyl analog *N*-ε-[(cyclopentylloxy)carbonyl]-L-lysine (Cyc). As for all PylRS structures reported to date the highly insoluble N-terminal domain was not included (annotated as pylSn below and in Fig. 3). This and other structures reported subsequently describe the unique organization of the binding pocket for the large Pyl side chain, thus elucidating the principles governing Pyl recognition by PylRS [47,48]. The core catalytic domain contains the typical features of the class II aaRS family, and structural phylogeny suggested the divergence of PylRS from an ancestral version of

PheRS early in evolution, i.e., during the time of the last common ancestor of all life on earth [47].

The structure also revealed that hydrophobic interactions account for most of the contact surface between PylRS and the Pyl or Cyc substrates, but specific recognition is derived from hydrogen bonds formed between PylRS and the amino acid substrate (Fig. 3A). The interactions of two residues (Asn346 and Arg330 in *Methanosarcina mazei* numbering) with the primary and secondary carbonyl of the substrate were noticed in complexes with Pyl or Cyc. Although the cyclic components of both substrates were bound in a roughly equivalent position in the hydrophobic pocket of PylRS, subtle differences in the interactions are likely responsible for a significantly greater charging efficiency of tRNA^{Pyl} with Pyl as opposed to Cyc [45,49]. The residue Tyr384 in PylRS is believed to play an important role in orienting Pyl into the binding pocket through its hydrogen bond to the pyrrole ring nitrogen. Tyr384 is located on a mobile loop that only appears ordered in the presence of the cognate substrate Pyl [47,48]. A report of the first bacterial PylRS crystal structure revealed a somewhat more sterically constrained active site that can nevertheless accommodate Pyl [50]. The large active site of PylRS has been shown to accommodate several Pyl analogs [51–53], including some that

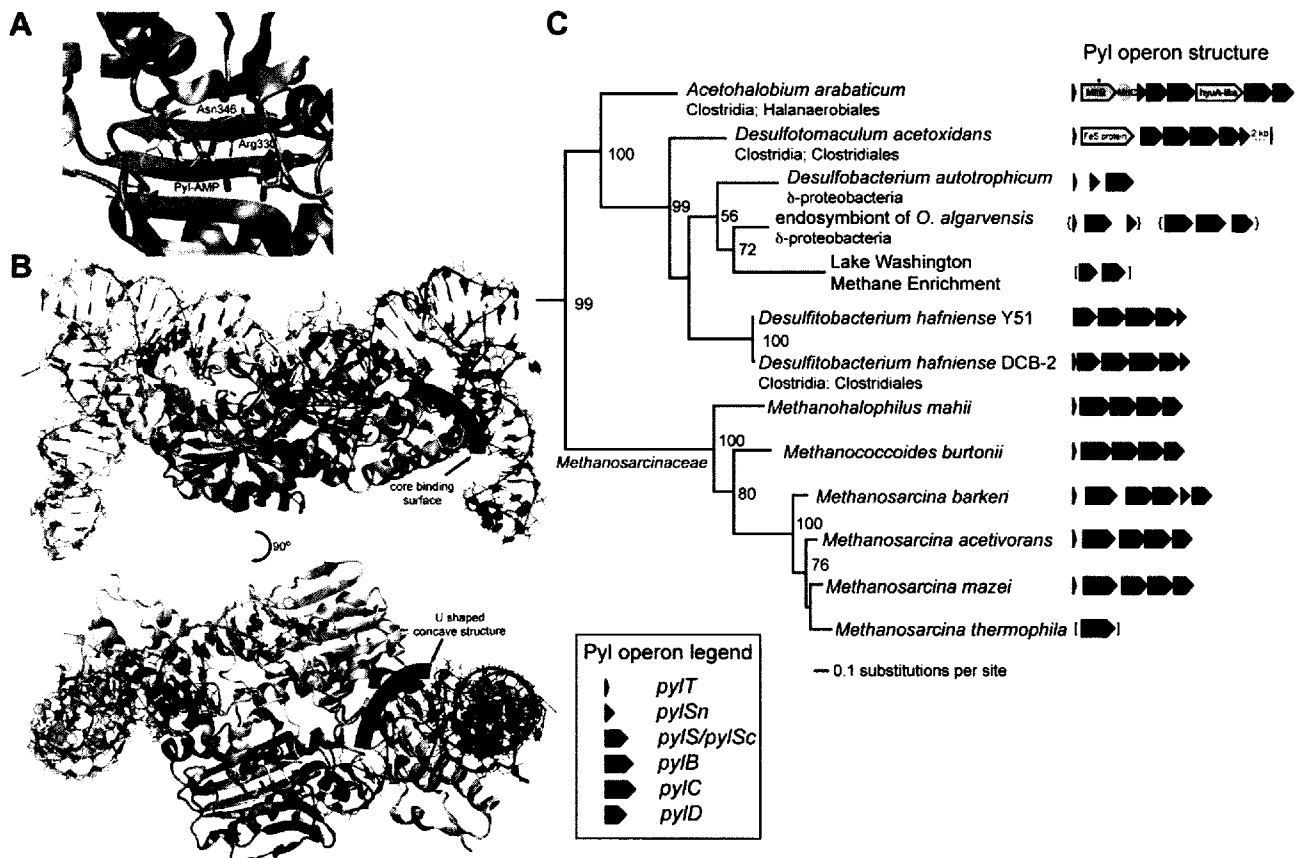


Fig. 3. Structure and evolution of PylRS and the Pyl operon. (A) Key hydrogen bond interactions in the active site of *M. mazei* PylRS with pyrrolysyl-adenylate (Pyl-AMP) bound. (B) Two views of the *D. hafniense* PylRS/tRNA^{Pyl} complex showing the tRNA binding domain 1 (cyan), the catalytic core domain (tan), the bulge domain (yellow), the C-terminal tail domain (green), and α-helix 6 (blue) emerging from the other protomer (gray). The tRNA core binding surface (magenta highlight) and U shaped concave surface (purple highlight) are also illustrated. Pyl-AMP from the *M. mazei* PylRS structure was superimposed onto the DhPylRS structure for reference. (C) The PylRS maximum likelihood phylogenetic tree includes all known putative Pyl-decoding organisms to date, and the organization of the Pyl operon in each genome sequence is shown adjacent to the tree. Square brackets indicate partial sequence data from a short sequence read, and curly brackets indicate that the Pyl operon is split between different contigs from a metagenomic survey. * indicates an in-frame TAG codon. Sequence data were downloaded from the Integrated Microbial Genomes database [65]. The tree was calculated with PHYML [66] using a BioNJ starting tree and SPR tree search followed by NNI branch swapping to optimize the tree. All alignment positions were considered, likelihood parameters were estimated from the alignment, and the JTT + Γ model with eight rate categories was applied. Bootstrap values (only those >50 are shown) were computed according to the Shimodaira–Hasegawa re-estimation of log-likelihood test implemented in PHYML.

can be chemically altered post-translationally using “click chemistry” to specifically attach fluorescent probes [54,55].

7. Molecular basis for the PylRS:tRNA^{Pyl} orthogonality

As is typical for class II aaRSs, PylRS is a dimer of two identical subunits, each of which bind one tRNA molecule from the major groove side. Each tRNA^{Pyl} interacts with one subunit and makes few specific contacts with the second protomer. Since PylRS and tRNA^{Pyl} display no significant cross reactivity with other aaRS/tRNA pairs in native or heterologous contexts [38,56,57], they are said to be orthogonal to other aminoacylation systems. Several unique structural features in both the protein and its substrate RNA have emerged in the course of evolution that explain the specificity of PylRS for its tRNA (Fig. 3B).

In both archaea and bacteria, the amber suppressor tRNA^{Pyl} has several unique structural features when compared to canonical tRNAs. These include an anticodon stem of six base pairs instead of five, a short variable loop of only three bases, a single base separating the acceptor and D-stems, a small D-loop with only five bases, the absence of the almost universally conserved G18G19 sequence in the D-loop and the T54Ψ55C56 sequence in the T-Loop, and a small number of post-transcriptional base modifications. *M. barkeri* tRNA^{Pyl} was found to have only a 4-thiouridine at position 8 and a 1-methyl-pseudouridine at position 50 [45].

Crystallization of the *Desulfitobacterium hafniense* PylRS (DhPylRS) in complex with its cognate tRNA provided the first view of a naturally evolved orthogonal aaRS/tRNA pair. The structure supports what was predicted from sequence, that PylRS/tRNA^{Pyl}, unlike engineered orthogonal pairs, is highly distinct from even the most closely related aaRS/tRNA pair [57]. The structure showed that while the unusual features of tRNA^{Pyl} do not prevent the formation of the expected L-shape tertiary structure fold, they result in a unique core more compact than those seen in canonical tRNAs. To match the structural peculiarity of its tRNA substrate, PylRS has evolved unique protein domains from which residues emerge and engage in specific interactions with tRNA^{Pyl} core nucleotides. The tRNA core comes in contact with one side of the protein hence termed core binding surface. This surface groups the tRNA binding domain 1 (N-terminus of pylSc domain), the C-terminal tail and helix α6 on the second protomer (Fig. 3B, upper).

Structural data identified the tRNA core nucleotides involved in specific interactions with the protein. Nucleotide G9 and D-stem nucleotide pairs G10:C25 and A11:U24 are engaged in a series of hydrogen bonding interaction with PylRS amino acid side chains. Earlier biochemical evidence had indicated the importance of these particular nucleotides for tRNA^{Pyl} aminoacylation [49,58]. Their mutation resulted in dramatic decrease in aminoacylation efficiency both in vitro and in vivo. It was, however, unclear whether the deficient aminoacylation resulted from destabilization of the tRNA structure or the direct disruption of specific contacts with PylRS. Structural data now unambiguously identify these nucleotides as key elements contributing to the orthogonality of PylRS/tRNA^{Pyl}.

The tRNA binding domain 1, C-terminal tail and bulge domain of the opposite subunit form a U shaped concave structure complementary to the acceptor helix of the tRNA that directs its 3'-terminus toward the PylRS catalytic site (Fig. 3B, lower). The structural data also show how PylRS selectively recognizes, through a network of specific hydrogen bonding interactions, the discriminator base and the G1:C72 base pair at the top of the acceptor stem. Compelling biochemical evidence demonstrating the critical importance of the same nucleotides for the aminoacylation of tRNA^{Pyl} with pyrrolysine in vitro and in vivo were also obtained [49,58]. The discriminator base G73 and the G1:C72 base pair along

with those of the core domain define the complete tRNA^{Pyl} identity set. Transfer of these nucleotides is able to convert a catalytically non-competent tRNA substrate into a pyrrolysine accepting tRNA [49]. Although biochemical studies have shown that tRNA^{Pyl} anticodon nucleotides are dispensable for the formation of PylRS/tRNA^{Pyl} complex and tRNA aminoacylation, the two bases adjacent to the anticodon (U33 and A37) were shown to be identity elements for the *M. mazei* PylRS [49]. These two bases are possibly recognized in a sequence-specific manner by residues from the N-terminal domain, a domain lacking in the DhPylRS and all other PylRS crystal structures.

The strict orthogonality of the PylRS/tRNA^{Pyl} system, which has been demonstrated [57], and available crystal structures have already been exploited for the genetic encoding, and thus site-specific incorporation of typically post-translationally modified amino acids directly into proteins. PylRS variants were selected to generate a mutant that specifically charges tRNA^{Pyl} with *N*-acetyl-lysine [59]. Wild type PylRS was employed to aminoacylate tRNA^{Pyl} with a reactive Pyl analog which was chemically converted to the common histone modification *N*-methyl-lysine [60].

8. Expanding the Pyl-decoding biosphere

Data from newly sequenced genomes and metagenomic projects has doubled the potential number of Pyl-decoding organisms. The Pyl-decoding trait is associated with the presence of the Pyl operon, which includes genes for the tRNA^{Pyl} (*pylT*) and the PylRS (*pylS*) followed by genes that encode the putative Pyl biosynthetic machinery (*pylBCD*). The Pyl operon is sufficient to encode the Pyl-decoding trait as shown by transformation of the operon into a heterologous *E. coli* context that then supported translational read-through of amber codons [38,56]. Earlier reports [37] found the Pyl-decoding trait confined to five archaeal organisms from the *Methanosarcinaceae* family and one bacterial organism (*D. hafniense*), while later studies [47,58,61] identified components of the Pyl operon in a δ -proteobacterial endosymbiont of the marine worm *Olavius algarvensis*, based on sequence data from a metagenomic survey [62].

There are now four additional examples of the bacterial version of the Pyl-trait, and one additional *Methanosarcinaceae* (*Methanohalophilus mahii*) is shown to have the complete Pyl operon (Fig. 3C). For reasons unknown, only *M. barkeri* among the archaea appears to deviate from the otherwise strict operonal organization and includes a small (100 codon) ORF that is a mutated duplication of the 3'-end of *pylC*. The bacterial examples show greater diversity in Pyl operon structure. In keeping with this observation, an updated phylogeny of the PylRS sequences continues to show a clear divide between the bacterial and archaeal type PylRSs, with a statistically significant root placed between the bacterial and archaeal clades according to an alignment with other subclass IIC aaRSs. While the archaeal Pyl-decoding trait is confined to the *Methanosarcinaceae*, bacterial examples appear in only a few members of the phyla Clostridia and δ -proteobacteria.

A key difference between the archaeal and bacterial PylRS is that the bacteria encode the N-terminal domain (about 100 amino acids) of PylRS as a separate gene, often at the 3'-end of the operon, leading to the organization (*pylT*, *pylSc*, *pylBCD*, *pylSn*). While the evolutionary scenario for the *pylS* gene split and operonal rearrangement was previously unclear, the newly sequenced and deepest branching bacterial Pyl-decoding representative *Acetohalobium arabaticum*, appears to display a transitional form. In this organism, *pylSn* terminates 11 bases upstream of and is thus out-of-frame with *pylSc*. The phylogeny indicates that the split *pylS* evolved after the divergence of the archaeal and bacterial *pylS*

genes and that the re-arrangement (placement of *pylSn* at the 3'-end of the operon) occurred later in bacterial evolution.

Only two organisms interrupt the Pyl operon with other genes. *A. arabaticum* encodes the Pyl-containing trimethylamine methyltransferase (MttB) and cognate corrinoid protein (MttC) in the operon as well as a gene related to *N*-methylhydantoinase (*hyuA*) and 5-oxoprolinase between *pylB* and *pylC*. The gene has close homologs in most Pyl-decoding organisms and various bacteria. Proteins from this superfamily (*hyuA*) are typically involved in the breakdown of cyclic amines (similar to the pyrroline ring of Pyl), and this gene could be the first example of a catabolic Pyl enzyme. *Desulfotomaculum acetoxidans* encodes an unknown iron-sulfur protein between *pylT* and *pylSc* and an exact copy of the 3' 30 bases of tRNA^{Pyl} downstream of the operon, which is an indication that the Pyl operon is on a mobile genetic element [63]. Finally, the complete genome of *Desulfobacterium autotrophicum* contains the first example of a degraded Pyl operon, only *pylSc* and *pylB* are present. Gene loss has been assumed to explain in part the sparsity of the Pyl-operon, and this example provides the first evidence of a transitional form towards loss of the Pyl-decoding trait.

9. The 21st and 22nd amino acids promise exciting new directions

The enzymes involved in the archaeal and eukaryotic Sec biosynthesis pathways achieve specificity for tRNA^{Sec} by recognizing distinct structural features of the tRNA substrate. While a similar theme is observed in the Pyl-decoding system, Sec and Pyl followed different evolutionary and metabolic routes to enter the genetic code. Sec relied on a specialized tRNA that is recognized by a canonical aaRS and on the evolution of enzymes that specifically modify the aminoacyl-tRNA to ultimately form Sec-tRNA^{Sec}. In contrast, Pyl entered the code as a result of the establishment of a tRNA-independent biosynthetic pathway and co-evolution of a aaRS/tRNA pair that is quite different from its closest known relatives. Despite their differences, both systems will be important for future biotechnological applications. Just as detailed biochemical and structural investigation of the PylRS:tRNA^{Pyl} system facilitated the engineering of this orthogonal pair to incorporate other non-canonical amino acids, the archaeal/eukaryotic Sec-tRNA^{Sec} synthesis machinery, specifically the intermediate Sep-tRNA^{Sec}, provides another possible means to incorporate phosphoserine or possibly other phosphorylated amino acids directly into a growing polypeptide chain.

Acknowledgments

We thank Ilka Heinemann for advice and Lennart Randau for suggestions and help locating bacterial *pylT* genes. Work in the authors' laboratory was supported by grants (to D.S.) from the Department of Energy, the National Science Foundation, and the National Institute for General Medical Sciences.

References

- Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M.A. and Böck, A. (1988) Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* 331, 723–725.
- Yuan, J., Palioura, S., Salazar, J.C., Su, D., O'Donoghue, P., Hohn, M.J., Cardoso, A.M., Whitman, W.B. and Söll, D. (2006) RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc. Natl. Acad. Sci. USA* 103, 18923–18927.
- Abe, K., Mihara, H., Tobe, R. and Esaki, N. (2007) Characterization of human selenocysteine synthase involved in selenoprotein biosynthesis. *Biomed. Res. Trace Elements* 19, 80–83.
- Xu, X.M., Carlson, B.A., Mix, H., Zhang, Y., Saira, K., Glass, R.S., Berry, M.J., Gladyshev, V.N. and Hatfield, D.L. (2007) Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol.* 5, e4.
- Kryukov, G.V., Castellano, S., Novoselov, S.V., Lobanov, A.V., Zehtab, O., Guigo, R. and Gladyshev, V.N. (2003) Characterization of mammalian selenoproteomes. *Science* 300, 1439–1443.
- Bellinger, F.P., Raman, A.V., Reeves, M.A. and Berry, M.J. (2009) Regulation and function of selenoproteins in human disease. *Biochem. J.* 422, 11–22.
- Ehrenreich, A., Forchhammer, K., Tormay, P., Veprek, B. and Böck, A. (1992) Selenoprotein synthesis in *E. coli*. Purification and characterisation of the enzyme catalysing selenium activation. *Eur. J. Biochem.* 206, 767–773.
- Böck, A., Thanbichler, M., Rother, M. and Resch, A. (2005) Selenocysteine in: *The Aminoacyl-tRNA Synthetases* (Ibba, M., Francklyn, C.S. and Cusack, S., Eds.), pp. 320–327. Landes Bioscience, Georgetown, TX.
- Yoshizawa, S. and Böck, A. (2009) The many levels of control on bacterial selenoprotein synthesis. *Biochim. Biophys. Acta* 1790, 1404–1414.
- Carlson, B.A., Xu, X.M., Kryukov, G.V., Rao, M., Berry, M.J., Gladyshev, V.N. and Hatfield, D.L. (2004) Identification and characterization of phosphoseryl-tRNA^{Ser} kinase. *Proc. Natl. Acad. Sci. USA* 101, 12848–12853.
- Araiso, Y., Palioura, S., Ishitani, R., Sherrer, R.L., O'Donoghue, P., Yuan, J., Oshikane, H., Dornae, N., DeFranco, J., Söll, D. and Nureki, O. (2008) Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic Acids Res.* 36, 1187–1199.
- Sherrer, R.L., O'Donoghue, P. and Söll, D. (2008) Characterization and evolutionary history of an archaeal kinase involved in selenocysteinyl-tRNA formation. *Nucleic Acids Res.* 36, 1247–1259.
- Su, D., Hohn, M.J., Palioura, S., Sherrer, R.L., Yuan, J., Söll, D. and O'Donoghue, P. (2009) How an obscure archaeal gene inspired the discovery of selenocysteine biosynthesis in humans. *IUBMB Life* 61, 35–39.
- Sauerwald, A., Zhu, W., Major, T.A., Roy, H., Palioura, S., Jahn, D., Whitman, W.B., Yates 3rd, J.R., Ibba, M. and Söll, D. (2005) RNA-dependent cysteine biosynthesis in archaea. *Science* 307, 1969–1972.
- Burkard, U. and Söll, D. (1988) The unusually long amino acid acceptor stem of *Escherichia coli* selenocysteine tRNA results from abnormal cleavage by RNase P. *Nucleic Acids Res.* 16, 11617–11624.
- Baron, C., Westhof, E., Böck, A. and Giege, R. (1993) Solution structure of selenocysteine-inserting tRNA^{Sec} from *Escherichia coli*. Comparison with canonical tRNA^{Ser}. *J. Mol. Biol.* 231, 274–292.
- Sturchler, C., Westhof, E., Carbon, P. and Krol, A. (1993) Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA^{Sec}. *Nucleic Acids Res.* 21, 1073–1079.
- Palioura, S., Sherrer, R.L., Steitz, T.A., Söll, D. and Simonović, M. (2009) The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. *Science* 325, 321–325.
- Itoh, Y., Chiba, S., Sekine, S.I. and Yokoyama, S. (2009) Crystal structure of human selenocysteine tRNA. *Nucleic Acids Res.* 37, 6259–6268.
- Westhof, E., Dumas, P. and Moras, D. (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.* 184, 119–145.
- Caldeira de Araujo, A. and Favre, A. (1986) Near ultraviolet DNA damage induces the SOS responses in *Escherichia coli*. *EMBO J.* 5, 175–179.
- Mäenpää, P.H. and Bernfield, M.R. (1970) A specific hepatic transfer RNA for phosphoserine. *Proc. Natl. Acad. Sci. USA* 67, 688–695.
- Sharp, S.J. and Stewart, T.S. (1977) The characterization of phosphoseryl tRNA from lactating bovine mammary gland. *Nucleic Acids Res.* 4, 2123–2136.
- Kaiser, J.T., Gromadski, K., Rother, M., Engelhardt, H., Rodnina, M.V. and Wahl, M.C. (2005) Structural and functional investigation of a putative archaeal selenocysteine synthase. *Biochemistry* 44, 13315–13327.
- Sherrer, R.L., Ho, J.M.L. and Söll, D. (2008) Divergence of selenocysteine tRNA recognition by archaeal and eukaryotic O-phosphoseryl-tRNA^{Sec} kinase. *Nucleic Acids Res.* 36, 1871–1880.
- Araiso, Y., Sherrer, R.L., Ishitani, R., Ho, J.M.L., Söll, D. and Nureki, O. (2009) Structure of a tRNA-dependent kinase essential for selenocysteine decoding. *Proc. Natl. Acad. Sci. USA* 106, 13.
- Diamond, A.M., Choi, I.S., Crain, P.F., Hashizume, T., Pomerantz, S.C., Cruz, R., Steer, C.J., Hill, K.E., Burk, R.F., McCloskey, J.A. and Hatfield, D.L. (1993) Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA^{Ser}. *J. Biol. Chem.* 268, 14215–14223.
- Wu, X.Q. and Gross, H.J. (1994) The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosphorylation. *EMBO J.* 13, 241–248.
- Percudani, R. and Peracchi, A. (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep.* 4, 850–854.
- Eliot, A.C. and Kirsch, J.F. (2004) Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* 73, 383–415.
- Ganichkin, O.M., Xu, X.M., Carlson, B.A., Mix, H., Hatfield, D.L., Gladyshev, V.N. and Wahl, M.C. (2007) Structure and catalytic mechanism of eukaryotic selenocysteine synthase. *J. Biol. Chem.* 283, 5849–5865.
- Yuan, J., Sheppard, K. and Söll, D. (2008) Amino acid modifications on tRNA. *Acta Biochim. Biophys. Sin. (Shanghai)* 40, 539–553.
- Bailly, M., Blaise, M., Lorber, B., Becker, H.D. and Kern, D. (2007) The transamidosome: A dynamic ribonucleoprotein particle dedicated to prokaryotic tRNA-dependent asparagine biosynthesis. *Mol. Cell* 28, 228–239.
- Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A. and Chan, M.K. (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 296, 1462–1466.
- Soares, J.A., Zhang, L., Pitsch, R.L., Kleinholz, N.M., Jones, R.B., Wolff, J.J., Amster, J., Green-Church, K.B. and Krzycki, J.A. (2005) The residue mass of L-pyrrolysine in three distinct methylamine methyltransferases. *J. Biol. Chem.* 280, 36962–36969.

- [36] Krzycki, J.A. (2004) Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr. Opin. Chem. Biol.* 8, 484–491.
- [37] Zhang, Y., Baranov, P.V., Atkins, J.F. and Gladyshev, V.N. (2005) Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* 280, 20740–20751.
- [38] Namy, O., Zhou, Y., Gundllapalli, S., Polycarpo, C.R., Denise, A., Rousset, J.P., Söll, D. and Ambrogelly, A. (2007) Adding pyrrolysine to the *Escherichia coli* genetic code. *FEBS Lett.* 581, 5282–5288.
- [39] Longstaff, D.G., Blight, S.K., Zhang, L., Green-Church, K.B. and Krzycki, J.A. (2007) *In vivo* contextual requirements for UAG translation as pyrrolysine. *Mol. Microbiol.* 63, 229–241.
- [40] Heinemann, I.U., O'Donoghue, P., Madinger, C., Benner, J., Randau, L., Noren, C.J. and Söll, D. (2009) The appearance of pyrrolysine in tRNA^{His} guanylyltransferase by neutral evolution. *Proc Natl Acad Sci USA*, doi:10.1073/pnas.0912072106.
- [41] Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R. and Gross, H.J. (1984) UAG readthrough during TMV RNA translation: isolation and sequence of two tRNAs with suppressor activity from tobacco plants. *EMBO J.* 3, 351–356.
- [42] Ambrogelly, A., Palioura, S. and Söll, D. (2007) Natural expansion of the genetic code. *Nat. Chem. Biol.* 3, 29–35.
- [43] Krzycki, J.A. (2005) The direct genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* 8, 706–712.
- [44] Blight, S.K., Larue, R.C., Mahapatra, A., Longstaff, D.G., Chang, E., Zhao, G., Kang, P.T., Green-Church, K.B., Chan, M.K. and Krzycki, J.A. (2004) Direct charging of tRNA_{CUA} with pyrrolysine *in vitro* and *in vivo*. *Nature* 431, 333–335.
- [45] Polycarpo, C., Ambrogelly, A., Berube, A., Winbush, S.M., McCloskey, J.A., Crain, P.F., Wood, J.L. and Söll, D. (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. USA* 101, 12450–12454.
- [46] Yanagisawa, T., Ishii, R., Fukunaga, R., Nureki, O. and Yokoyama, S. (2006) Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 62, 1031–1033.
- [47] Kavran, J.M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D. and Steitz, T.A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. USA* 104, 11268–11273.
- [48] Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K. and Yokoyama, S. (2008) Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* 378, 634–652.
- [49] Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C. and Söll, D. (2007) Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci. USA* 104, 3141–3146.
- [50] Lee, M.M., Jiang, R., Jain, R., Larue, R.C., Krzycki, J. and Chan, M.K. (2008) Structure of *Desulfitobacterium hafniense* PylSc, a pyrrolysyl-tRNA synthetase. *Biochem. Biophys. Res. Commun.* 374, 470–474.
- [51] Kobayashi, T., Yanagisawa, T., Sakamoto, K. and Yokoyama, S. (2009) Recognition of non- α -amino substrates by pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* 385, 1352–1360.
- [52] Li, W.T., Mahapatra, A., Longstaff, D.G., Bechtel, J., Zhao, G., Kang, P.T., Chan, M.K. and Krzycki, J.A. (2009) Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *J. Mol. Biol.* 385, 1156–1164.
- [53] Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K. and Yokoyama, S. (2008) Adding ϵ -lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* 371, 818–822.
- [54] Fekner, T., Li, X., Lee, M.M. and Chan, M.K. (2009) A pyrrolysine analogue for protein click chemistry. *Angew. Chem. Int. Ed. Engl.* 48, 1633–1635.
- [55] Nguyen, D.P., Lusic, H., Neumann, H., Kapadnis, P.B., Deiters, A. and Chin, J.W. (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA_{CUA} pair and click chemistry. *J. Am. Chem. Soc.* 131, 8720–8721.
- [56] Longstaff, D.G., Larue, R.C., Faust, J.E., Mahapatra, A., Zhang, L., Green-Church, K.B. and Krzycki, J.A. (2007) A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc. Natl. Acad. Sci. USA* 104, 1021–1026.
- [57] Nozawa, K., O'Donoghue, P., Gundllapalli, S., Araisio, Y., Ishitani, R., Umehara, T., Söll, D. and Nureki, O. (2009) Pyrrolysyl-tRNA synthetase-tRNA^{Pyl} structure reveals the molecular basis of orthogonality. *Nature* 457, 1163–1167.
- [58] Herring, S., Ambrogelly, A., Polycarpo, C.R. and Söll, D. (2007) Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase. *Nucleic Acids Res.* 35, 1270–1278.
- [59] Neumann, H., Peak-Chew, S.Y. and Chin, J.W. (2008) Genetically encoding N^ε-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* 4, 232–234.
- [60] Nguyen, D.P., Alai, M.M., Kapadnis, P.B., Neumann, H. and Chin, J.W. (2009) Genetically encoding N^ε-methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.* 131, 14194–14195.
- [61] Zhang, Y. and Gladyshev, V.N. (2007) High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic delta-proteobacterium of gutless worm *Olavius algarvensis*. *Nucleic Acids Res.* 35, 4952–4963.
- [62] Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloeckner, F.O., Boffelli, D., Anderson, I.J., Barry, K.W., Shapiro, H.J., Szeto, E., Kyrpides, N.C., Mussmann, M., Amann, R., Bergin, C., Ruehlend, C., Rubin, E.M. and Dubilier, N. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443, 950–955.
- [63] Hacker, J. and Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* 54, 641–679.
- [64] Itoh, Y., Sekine, S., Kuroishi, C., Terada, T., Shirouzu, M., Kuramitsu, S. and Yokoyama, S. (2008) Crystallographic and mutational studies of seryl-tRNA synthetase from the archaeon *Pyrococcus horikoshii*. *RNA Biol.* 5, 169–177.
- [65] Markowitz, V.M. and Kyrpides, N.C. (2007) Comparative genome analysis in the integrated microbial genomes (IMG) system. *Methods Mol. Biol.* 395, 35–56.
- [66] Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.