ANALYSIS OF THE ENZYMOLOGICAL PROPERTIES OF PROLYL-TRNA

SYNTHETASES IN PLANTS FOCUSING ON THE MISACTIVATION OF THE

PROLINE ANALOG AZETIDINE-2-CARBOXYLIC ACID

by

JIYEON LEE

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Plant Biology

written under the direction of

Professor Thomas Leustek

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2009

ABSTRACT OF THE DISSERTATION

ANALYSIS OF THE ENZYMOLOGICAL PROPERTIES OF PROLYL-TRNA

SYNTHETASES IN PLANTS FOCUSING ON THE MISACTIVATION OF THE

PROLINE ANALOG AZETIDINE-2-CARBOXYLIC ACID

By JIYEON LEE

Dissertation Director:
Professor Thomas Leustek

Azetidine-2-carboxylic acid (A2C), a structural analogue of L-proline (Pro),

inhibits the growth of bacterial, plant and animal cells. Toxicity is thought to occur when

A2C is incorporated into proteins in place of Pro resulting from misrecognition by prolyl-

tRNA synthetases (ProRS). *Arabidopsis thaliana* seedlings are highly sensitive to A2C

resulting in growth inhibition. To explain this result, the activity of the two forms of

ProRS from *A. thaliana* was studied. One form of ProRS is localized to

chloroplasts/mitochondria and the other form to the cytosol. Both forms were expressed

as His-tagged recombinant proteins in *Escherichia coli*. Purified enzymes were

functionally active in the ATP-PP$_i$ exchange, and aminoacylation assays demonstrated

similar K_m values for Pro. A major difference was observed in the specificity for A2C.

The organellar form showed a 340-fold greater $K_{m[A2C]}$ than the cytosolic enzyme. These

results suggest that A2C-sensitivity of *A. thaliana* is primarily due to the inability of

cytosolic ProRS to distinguish between Pro and A2C. A similar result was obtained with

the ProRSs from *Zea mays* suggesting that the difference in substrate specificity is a conserved feature of plant cytosolic and organellar ProRSs. The tRNA-specificity of the organellar and cytosolic enzymes was also examined. The organellar ProRS At5g52520 was able to complement a conditional ProRS mutant strain of *E. coli* but the cytosolic ProRS At3g62120 did not. At5g52520 was able to efficiently aminoacylate both *E. coli* and *Saccharomyces cerevisiae* tRNA. In contrast, At3g62120 was able to aminoacylate *S. cerevisiae* tRNA more efficiently than *E. coli* tRNA. The interest in ProRS arose initially from the observation that some plant species, including *Convallaria majalis* and *Polygonatum multiflorum,* two members of the Liliopsida, accumulate high levels of A2C. In 1972, Norris *et el.* reported that ProRS from *C. majalis* can discriminate between Pro and A2C (Norris and Fowden, 1972). To gain a better understanding of the mechanism of A2C resistance, the cDNAs encoding ProRS were isolated from *C. majalis* and *Polygonatum pubescens*. Based on sequence homology, the cDNA encoding organellar ProRS was successfully cloned. However, attempts to produce a functional ProRS from the cloned cDNAs were unsuccessful.

Acknowledgement

I appreciate and dedicate this work to God.

I appreciate my advisor, Professor Thomas Leustek, who supported and guided me to complete this work. I also appreciate my committees, Professor Barbara Zilinskas, Professor Michael Lawton, Professor Peter Kahn, and Professor Randy Kerstetter. I appreciate my master advisor, Professor Sangyeol Lee, who first made me interested in biochemistry.

I appreciate Dr. Melinda Martin who helped and shared many things with me in same office over 6 years, and all current and former Leustek lab members, including Minsang Lee, Andre Hudson, Patricia Zafian, Varinnia Gomes, Rita Pasini, and Susan Jones-Held.

I thank Barbara Mulder for her helping me collect the *C. majalis* samples from her garden. I also appreciate Dr. Jong-Chan Chae for his help in my adjustment to life in Foran Hall, and Dr. Young-beom Ahn and Dr. Joong-Wook for their help in discussing my experiments.

 I'd like to share this pleasure with my family within God' love: Eulsong Lee and Sokyoung Lee, Jonghun Lee and Nayeon Kim and Youngjae Choi and Nara Lee, Eunkyu Son and Sangtaek Jong, TJ Choi and Jeongkyu Song, Soonyong Park and Youngjoo Nam, Jongbeom Lee and Youngjoo Lee and Songsu Lee.

First of all, I appreciate my mother and father who believe and support me until now. I also appreciate my family, Mijong Lee, Sinae Lee, Kwang Lee, Hohyeong Cho, Jongsik Kang and Kunjoo Lee. I love my parent and family too much. They are my power and my life.

Table of Contents

Lists of Tables

List of Figures

1. Introduction

The fidelity of protein synthesis depends on two steps operating in sequence. First, the correct amino acid is joined to the corresponding tRNA to form an aminoacyl-tRNA. Then, the correct amino acid-linked tRNA is used at the codon-programmed ribosomal A site (Jakubowski and Goldman, 1992). Aminoacyl-tRNA synthetase (AARS) catalyzes the formation of aminoacyl-tRNA. This process is critical for the accurate translation of genetic information. In spite of the importance of AARSs in protein synthesis, they are poorly characterized in higher plants. To better understand the process, my research project focused on amino acid specificity and tRNA recognition for one of the twenty AARSs, ProRS which is responsible for Pro recognition and selection.

1.1 A2C is a toxic Pro analogue

A2C is a four-membered ring molecule that is a structural analogue of L-Pro (Fig. 1). A2C is toxic and inhibits the growth of most bacterial, plant and animal cells which do not produce this analogue (Nomura et al., 2003). When A2C is incorporated into proteins in place of Pro the resulting proteins are malformed and non-functional (Fowden et al., 1963; Fowden and Richmond, 1963; Hewitt and Notton, 1967). Presumably, due to the structural similarity between A2C and Pro, the ProRS of A2C-sensitive species is unable to discriminate between A2C and Pro resulting in mis-incorporation of A2C into proteins during translation. Moreover, in microorganisms and plants it is known that Pro biosynthesis is feedback-inhibited by Pro and also by A2C (Smith et al., 1984; Sugiura et al., 1985; Omori et al., 1991; Omori et al., 1992). Thus, it is possible that another mechanism for A2C toxicity is the inhibition of Pro biosynthesis.

1.1.1 A2C resistance

ProRSs of most organisms recognize A2C as a substrate and are able to transfer it to tRNA$_{Pro}$ (Fowden et al., 1963; Hewitt and Notton, 1967; Nomura et al., 2003). However, A2C occurs naturally in some plants belonging to the Convallariaceae family, notably *Convallaria majalis* (lily-of-the-valley) and *Polygonatum multiflorum* (Solomon's seal) (Baum et al., 1975). These plants not only produce high levels of A2C, but also are resistant to it. Although A2C has been reported to comprise up to 6% of the dry weight of *C. majalis* and to exist in the free amino acid fraction, it is not present in the proteins of *C. majalis* leaves (Fowden and Bryant, 1959). The mechanism of A2C resistance is unknown but is likely to be multifaceted. For example, A2C producing species may sequester A2C in the vacuole away from the translational apparatus. The subcellular localization of A2C has never been examined. One resistance mechanism that has been reported is the ability of the ProRS of *C. majalis* to discriminate between Pro and A2C (Norris and Fowden, 1972).

Resistance to A2C has been reported in microbial species, animals, and plants (Dunnill and Fowden, 1965; Wasmuth and Caskey, 1976; Gallori et al., 1978; Vanlerberghe and Brown, 1987; Verbruggen et al., 1996). Mutants selected for resistance to A2C fall into 2 classes, mutations in a regulatory site that leads to Pro overproduction and mutations in amino acid transporters that lead to an inability to take up A2C. Naturally occurring A2C resistance has also been reported and falls into two categories. A2C-resistance in *Saccharomyces cerevisiae* is due to an acetyl transferase that modifies A2C preventing it from being incorporated into protein (Nomura et al., 2003). Resistance of Enterobacter species results from the ability to degrade A2C (Yeung et al., 1998).

Yet a third naturally occurring resistance mechanism has been identified in members of the Convallariaceae. These species appear to contain a form of ProRS that is able to preferentially use Pro over A2C (Norris and Fowden, 1972). The identification of a 'resistant' form of ProRS in the Convallariaceae was reported in 1972, shortly after the discovery of AARS. The activity was identified in total protein extracts. Since that time, nothing more has been reported. The current study was undertaken to study the ProRSs from *Arabidopsis thaliana* in order to understand the basis for A2C sensitivity in this species, and as a prelude to the study of ProRS from the Convallariaceae.

1.1.2 A2C biosynthesis

Although A2C is a close structural analog of Pro, it does not appear to be synthesized by a modification of the Pro biosynthesis pathway. A survey of possible precursors found that radiolabeled-methionine is incorporated into A2C in *C. majalis*, suggesting that this compound is the precursor of A2C (Leete et al., 1986). Leete et al. (1986) proposed that A2C is synthesized via S-adenosylmethionine (SAM) by the intramolecular displacement of methylthioladenosine. Even if SAM, produced from Met, is the most efficient precursor of A2C, little is known about the biosynthesis and metabolic pathway of A2C in plants, including the enzymes necessary for its synthesis. SAM is also the precursor of nicotianamine (NA) (Higuchi et al., 1999). Nicotianamine synthase (NAS) catalyzes the trimerization of SAM following an initial intramolecular displacement of methylthioladenosine that effectively produces A2C as a part of the NA molecule (Fig. 2). Considering the biosynthetic similarities between A2C and NA, it is tempting to speculate that a putative A2C synthase may be related to NAS.

1.2 AARS

1.2.1 The role of aminoacyl-tRNA synthesis and AARS enzymes

Aminoacyl-tRNA synthesis is an essential cellular process that leads to the utilization of the appropriate amino acids for ribosomal translation of messenger RNA during protein synthesis (Fig. 3). AARSs ensure the correct selection and attachment of an amino acid to its corresponding tRNA in a highly specific two-step reaction (Fig. 4). In the first step, the cognate amino acid is activated with ATP to form an aminoacyl-adenylate intermediate by AARS (Step 1). In the second step, the activated amino acid is transferred to the 2' or 3'-hydroxyl of the terminal adenosine of the cognate tRNA, yielding the specific aminoacyl-tRNA and AMP as the leaving group (Step 2). Most AARSs do not require tRNA for the amino acid activation step, although some enzymes, notably, GlnRS, ArgRS, GluRS, and LysRS isoform I are tRNA-dependent (Ibba and Soll, 2000).

In most living cells, there are 20 members of AARS enzyme family that are divided into two distinct classes of 10 enzymes each based on structural features and the topology of their active site (Ibba and Soll, 2000). The catalytic domain of class I enzymes is made up of the Rossman fold (five-handed parallel β-sheet) containing two motif sequences, HIGH and KMSKS. These enzymes attach the amino acid to the 2'-OH of the 3'-terminal ribonucleotide of the tRNA. Class II enzymes share an anti-parallel β-sheet (seven-stranded β-structure) structure with flanking α-helices, and attach the amino acid to the 3'-OH of the 3'-teminal ribonucleotide of the tRNA chain (Stehlin et al., 1997; Ibba and Soll, 2000). The catalytic domain of class II enzymes is made up of three distinct motifs

that are conserved in all class II enzymes (Zhang et al., 2006). Most class I AARSs are monomeric enzymes, while class II enzymes include dimeric or multimeric forms.

In addition to providing translational fidelity during aminoacyl-tRNA synthesis, AARSs function as important factors in several other cellular processes (Fig. 5) (Szymanski et al., 2000). Specific AARSs lead to tRNA processing, RNA splicing, RNA trafficking, and transcriptional and translational regulation in response to stress, apoptosis, and embryo development (Martinis et al., 1999; Szymanski et al., 2000; Hausmann and Ibba, 2008) Many questions, however, remain unanswered related to AARSs structure and function. Recent studies of AARS have focused on three aspects of function: the mechanism of amino acid recognition and chemical activation, the specificity of tRNA recognition, and the origin and evolution of AARS (Szymanski et al., 2000).

1.2.2 The substrate specificity of AARS

AARSs must maintain the faithful relationship between anticodon and amino acid for high fidelity during the translation process because errors in this process will lead to abnormal proteins which will ultimately lead to aging or cell death (Freist et al., 1998; Wong et al., 2003). Aminoacylation depends on the substrate specificity of AARSs for both cognate amino acid and tRNA substrates. Inaccuracy is more frequent in the selection of the amino acid than in the selection of tRNAs due to the small surface area of the amino acid and fewer distinguishing features that serve as discriminating factors (Beuning and Musier-Forsyth, 2001). AARSs have proofreading mechanisms that increase the accuracy of aminoacylation and reduce errors in protein synthesis. Many

AARSs have evolved editing mechanisms known as "pretransfer" editing and "posttransfer" editing (Splan et al., 2008). In pretransfer editing, misactivated aminoacyl-adenylates are hydrolyzed to yield free amino acid and AMP. In posttransfer editing, mischarged aminoacyl-tRNAs are hydrolyzed. This "double-sieve" mechanism is a well described editing model (Fersht and Dingwall, 1979; Ahel et al., 2002; Tang and Tirrell, 2002). Cognate amino acid selection is based primarily on the size of the molecule and secondarily by specific chemical features. The first sieve, the "coarse sieve", acts at the active site for aminoacylation and activates the amino acids of the same or smaller size than the correct one. The second sieve, the "fine sieve" acts at the editing site. Here the product of activation is hydrolyzed if it is an amino acid smaller than the desired one. Editing activities have been studied for many tRNA synthetases including prolyl-tRNA synthetase (ProRS) (Beuning and Musier-Forsyth, 2000; Dock-Bregeon et al., 2000).

1.2.3 Protein synthesis in plants

In plant cells, protein synthesis takes place in three subcellular compartments the cytosol, mitochondria, and chloroplasts. Thus, a full set of AARSs and tRNAs must be present within each compartment. Most AARSs are encoded by nuclear genes and posttranslationally localized into their appropriate compartments via a transit peptide sequence at the N-terminus of the protein. In plants, the organization of AARS genes became evident after the *A. thaliana* genome was sequenced. In this species, the ProRS genes are readily identified as the cytosolic and chloroplast forms. But there is not a separate gene for an AARS targeted to the mitochondria. This paradox was recently solved when Duchene et al. (2005) showed that most organellar AARSs in *A. thaliana*

including ProRS show dual targeting specificity. A single gene produces an AARS that is targeted to both the mitochondria and chloroplasts. Two polypeptides are produced from mRNAs expressed from alternation of transcriptional start sites that differ at their N-terminus (Akashi et al., 1998).

In some but not all cases, tRNAs are encoded within their respective compartments. The chloroplast genome encodes the necessary tRNAs for translation in the chloroplast. The nuclear genome encodes the tRNA genes for cytosolic translation. However, the mitochondrial genome lacks a full set of tRNA genes. Some mitochondrial tRNAs are encoded by the nuclear genome and are transported into the mitochondria (Menand et al., 1998). In plants, it is known that cytosolic and organellar AARSs aminoacylate specific classes of tRNAs. The organellar enzymes are known to use tRNAs that more closely resemble *E. coli* tRNAs and cytosolic AARSs aminoacylate tRNAs that are specific to eukaryotes (Reger et al., 1970; Imbault et al., 1981). For example, bean (*Phaseolus vulgaris*) cytosolic LeuRS is able to aminoacylate a yeast tRNA preparation, but not chloroplast or *E. coli* tRNA (Gillemaut and Weil, 1975; Menand et al., 1998). On the other hand, plastid AARS can aminoacylate both *E. coli* tRNA and yeast tRNA preparations, presumably because yeast tRNA contains both cytosolic and mitochondrial tRNAs (Nakai et al., 2007). A few exceptions to this rule have been noted. Spinach chloroplast HisRS has been reported to use both cytosolic and chloroplast tRNAs, whereas the cytosolic HisRS was able to use only cytosolic tRNAs (Steinmetz and Weil, 1986; Akashi et al., 1998).

1.3 ProRS

ProRS activates Pro and forms prolyl-tRNA during protein biosynthesis. ProRS contains class II specific motifs as well as an anticodon binding domain.

1.3.1 The classification of ProRS

ProRSs from many organisms cluster into two phylogenetically distinct groups: a "prokaryote-like group" consisting of eubacterial and eukaryotic mitochondrial enzymes and a "eukaryote/archaea-like group" including the cytosolic ProRS from eukaryotes, archaea, and some bacteria (Yaremchuk et al., 2000; Beuning and Musier-Forsyth, 2001; Woese et al., 2000). Both types of ProRSs have a class II specific motif 1 responsible for dimer formation, a motif 2 and 3 responsible for amino acid activation, and an anticodon binding domain (Eriani et al., 1990; Ruff et al, 1991). Prokaryote-like ProRSs have a large insertion domain (INS) (~ 180 amino acids) of unknown function between motif 2 and 3 in the catalytic domain (Fig. 6) (Wong et al., 2002; Hati et al., 2006). Eukaryote/archaea-type enzymes contain a variable length C-terminal extension beyond the anticodon-binding domain (Yaremchuk et al., 2000). The C-terminal extension in the eukaryotic-type group bears no resemblance to the prokaryotic insertion domain. In addition to the pattern of insertion/extension domains, ProRS in the two groups have specific differences in their catalytic sites for aminoacylation (Liu et al., 1995). ProRSs from the two groups have different mechanisms for recognizing the nucleotides in the tRNA acceptor stem and the anticodon sequence to distinguish its cognate tRNA substrate. Whereas the prokaryote-like ProRSs recognize the bases in both a specific tRNA acceptor stem and the anticodon, the eukaryote/archaea-like ProRSs only recognizes nucleotide bases in the anticodon (Beuning and Musier-Forsyth, 2001; Burke

et al., 2001). The main point of these detailed mechanistic studies is that ProRSs are segregated into two evolutionarily divergent groups based on differences in recognition of the tRNA accepter stem.

Several studies provide evidence that the insertion domain (INS) of prokaryote-like ProRS is responsible for an editing function (Jakubowski and Goldman, 1992; Beuning and Musier-Forsyth, 2001; Hati et al., 2006; Splan et al., 2008). When the INS domain was expressed independently as a protein, it showed deacylation activity of a mischarged Ala-microhelix^{Pro} variant as an editing mechanism in ProRS (Wong et al., 2003). The ProRS editing domain is also found as an N-terminal extension and as a free-standing editing module (in both N-terminal and C-terminal extensions) in ProRSs of lower eukaryotes (Zhang et al., 2000; Wong et al., 2003; Ahel et al., 2003; SternJohn et al., 2007). Surprisingly, eukaryote/archaea-like ProRS does not contain an INS-like editing domain. Unlike bacterial enzymes, eukaryote/archaea-like ProRS contains a carboxyl-terminal appended domain that does not resemble the INS of prokaryote-like ProRS. It has been observed that the carboxyl-terminal domain in one of the eukaryote/archaea-like ProRSs binds one atom of zinc to interact with the aminoacylation active site (Yaremchuk et al., 2000). However, the functional role this binding is not known yet. Although the prokaryote-like ProRS INS was recently characterized, the function of the carboxyl-terminal domain is still unknown.

1.3.2 Dual-specificity of ProRS

It has been assumed that a set of 20 AARSs, one for each canonical amino acid, is essential for protein synthesis in each organism. However, this concept has been

overturned by recent discoveries. Bacteria and archaea do not appear to use a full complement of 20 canonical AARSs (Ibba and Soll, 2000). For example, these organisms lack both GlnRS and AsnRS and are thus unable to form these aminoacyl-tRNAs directly. Instead Gln-tRNA and Asn-tRNA are formed by amidation of Glu-tRNA and Asp-tRNA respectively (Tumbula et al., 2000). Furthermore, dual specificity AARSs have been identified. The complete genome sequences of archaeal species such as *Methanococcus jannaschii, Methanococcus maripaludi*s, and *Methanobacterium thermoautotrophicum* appear to lack a canonical cysteinyl-tRNA synthetase. Instead, a dual-specificity prolyl-cysteinyl-tRNA synthetase (ProCysRS) for the synthesis of both cysteinyl-tRNA$_{Cys}$ and prolyl-tRNA$_{Pro}$ was recently demonstrated (Stathopoulos et al., 2000). Furthermore, the binding sites of *M. jannaschii* ProCysRS for proline and cysteine (Cys) overlap. Surprisingly, although the genome from the lower eukaryote *Giardia lamblia* contains an open reading frame encoding cysteinyl-tRNA synthetase (CysRS), its ProRS is able to synthesize both cysteinyl-tRNA$_{Cys}$ and prolyl-tRNA$_{Pro}$ *in vivo* and *in vitro* (Bunjun et al., 2000). Cys activation was also observed for human, *Thermus thermophilus* and prokaryote-like group *E. coli* ProRS *in vitro* (Beuning and Musier-Forsyth, 2001; Feng et al., 2002; Jacquin-Becker et al., 2002). These studies raise the question of whether all ProRS enzymes have the potential to activate Cys.

1.4 Objectives of this study

One tool that has proven useful for studying the mechanism of amino acid selection and utilization by AARSs has been the use of amino acid analogs. One such analogue is A2C. As the first step to study the role of A2C and the substrate selectivity of plant

ProRS, it is essential to characterize and compare the enzymological properties of ProRS enzymes in A2C-sensitive species. The current study focused on the misactivation of the Pro analog A2C by ProRSs from *A. thaliana* and *Z. mays* in order to understand the basis for A2C sensitivity in these species.

2. Material and Method

2.1 General methods and specific supplies

The preparations of bacterial media and nucleic acid techniques were performed as described by Sambrook et al. (Sambrook et al., 1989). Nucleotide sequencing was carried out using an Applied Biosystems Model 373A Automated Sequencer (Applied Biosystems, Foster City, CA). The analysis of nucleotide and protein sequences was carried out using the programs, DNASTAR and Molecular Evolutionary Genetic Analysis (Mega) software version 4.0. Proteins were expressed in *E. coli* using the pET-30(a) system (Novagen Inc., Darmstadt, Germany) and purified by Ni-agarose affinity chromatography (Clontech Inc., Mountain View, CA). The concentration of protein was measured using a BioRad kit (Hercules, CA) based on the method of Bradford (Bradford, 1976).

All organic compounds and enzymes were purchased from Sigma, Inc. (St. Louis, MO). The DH5α strain of *E. coli* was used for plasmid amplification, and *E. coli* BL21-Codon Plus (DE3) RIL strain (Invitrogen Inc., Carlsbad, CA) was used for protein expression.

2.2 Studies of ProRSs in *A. thaliana* and *Z. mays*

2.2.1 Specific supplies

$[^{32}P]$pyrophosphate (PP_i) (84 Ci/mmol) and $[^3H]$Pro (85 Ci/mmol) were purchased from PerkinElmer Life Sciences. Pro, A2C, Met, Cys, *yeast* and *E. coli* tRNA were purchased from Sigma.

E. coli strains UQ27 and UQ818 were obtained from the *E. coli* Genetic Stock Center (New Haven, CT). UQ27 *(proS127 (ts), lacZ4, λ⁻, argG75)*, an *E. coli* temperature-sensitive *proS* mutant strain, and UQ818 (*lacZ4*, *λ⁻*, *gyrA222 (NalR)*, *aroE24, metB⁻, rpoB⁻ (Rifʳ), and cysS818 (ts)*), an *E. coli* temperature-sensitive *cysS* mutant strain, were used for functional complementation studies. *E. coli* XL1-Blue and BM25.8 strains, included in the SMART™ cDNA Library Construction Kit, were purchased from BD Biosciences (Palo Alto, CA).

A. thaliana Col7 (ABRC #CS3731) seed was obtained from the Arabidopsis Biological Resource Center (Columbus, OH). Full-length open reading frame (ORF) cDNA clones of At3g62120 (ABRC Stock # AY128720) and At5g52520 (clone # SQ040d11) were obtained from the Arabidopsis Biological Resource Center (Columbus, OH) and the Kazusa DNA Research Institute (Chiba, Japan), respectively. *Z. mays* cDNA clones, Zm_0027G07 (GenBank accession #AY104927) and ES13257 (GenBank accession #AY104268) were obtained from the Plant Gene Expression Center (Albany, CA) and the Arizona Genomics Institute (Tucson, AZ), respectively.

2.2.2 Plant growth conditions and measurement of root length

A. thaliana was grown on Noble-agar-solidified nutrient medium containing Murashige and Skoog (MS) nutrients (Life Technologies, Rockville, MD) and 0.8% ($^w/_v$) Bacto agar in the presence of different concentrations of A2C. Seeds were sterilized in 0.05% ($^v/_v$) Triton X-100 for 10 min, 95% ($^v/_v$) ethanol containing 0.05% ($^v/_v$) Triton X-100 for 5 min, and 30% ($^v/_v$) bleach containing 0.05% ($^v/_v$) Triton X-100 for 12 min. Sterilized seeds were rinsed three times with distilled water and mixed with 0.1% ($^w/_v$)

agarose. The seeds were sown on the appropriate medium and plates were positioned on edge to direct root growth along the surface of the medium. The plates were incubated at 24ºC at a diurnal cycle of 16 hours light and 8 hours dark. Light intensity was approximately 100 μmol photons $m^{-2}s^{-1}$. After 9 days of growth, the plants were transferred to fresh MS-agar medium for imaging. Root length was measured using ImageJ (Image Processing and Analysis in Java; http://rsbweb.nih.gov/ij/). When branched roots systems were present the length of the longest root was measured.

2.2.3 Sequence analysis

At3g62120 and At5g52520 have previously been identified as cytosolic and organellar forms of ProRS (Duchene et al., 2005). Two possible full-length ProRS cDNAs (GenBank accession # AY104927 and AY104268) were identified from *Z. mays* that are homolog of At3g62120 and At5g52520. The AY104268 translation product contains a possible organellar targeting sequence based on analysis with TargetP (http://www.cbs.dtu.dk/services/TargetP). No targeting sequences were identified in the AY104927 translation product. Based on these criteria, the protein products are referred to hereafter as the cytosolic and organellar forms of *Z. mays* ProRS, ZmProRS-Cy (AY104927) and ZmProRS-Org (AY104268) although localization has not been experimentally determined.

The phylogenetic relationship of ProRSs was analyzed. The sequence alignment and phylogenetic analysis were performed using Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 (Tamura et al., 2007). The sequences for ProRSs were analyzed using the NCBI web-based Blast server, and sequence alignments were

performed using CLUSTAL W. Phylogeny construction of combined sequences was done using the Bootstrapped Neighbor-Joining method and evolutionary distances were calculated using Mega4. The amino acid sequences of ProRSs included in the analysis were obtained from National Center for Biotechnology Information (NCBI). The species of origin and gene bank numbers are as follows; *Alkaliphilus oremlandii* (YP_001512023.1), *Aquifex aeolicus* (AAC06648.1), *Arabidopsis thaliana*-At5g52520-Org (AAQ65189.1), *Arabidopsis thaliana*-At3g62120-Cy (NP_850736.1), *Archaeoglobus fulgidus* (28664), *Bacillus subtilis* (CAB13530), *Borrelia burgdorferi* (AAC66767.1), *Caenorhabditis elegans*-Cy (AAA50660.1), *Caenorhabditis elegans*-Mt (CAB04884.1), *Chlamydia trachomatis* (P36431), *Chlorobium limicola* (YP_001943717.1), *Clostridium novyi* (YP_878835.1), *Deinococcus radiodurans* (AAF10837.1), *Drosophila melanogaster* (AAC47469.1), *Escherichia coli* (P16659.4), *Giardia lamblia* (AAG23138.2), *Homo sapiens* (P07814.3), *Homo sapiens*-Mt (CAB55948.1), *Methanococcus jannaschii* (Q58635), *Methanococcus maripauldis* (NP_987816), *Methanothermobacter thermoautotrophicus* (O26708.1), *Mycobacterium leprae* (CAB36573.1), *Mycobacterium tuberculosis* (O05814.1), *Mycoplasma genitalium* (P47525), *Myxococcus xanthus* (YP_634772.1), *Onion yellows* phytoplasma (NP_950498.1), *Oryza sativa*-Org (BAC79747.1), *Oryza sativa*-Cy (ABA97740.1), *Pyrococcus horikoshii* (BAA30103.1), *Saccharomyces cerevisiae*-Cy (P38708), *Saccharomyces cerevisiae*-Mt (P39965.1), *Schizosaccharomyces pombe*-Cy (CAA19574.1), *Schizosaccharomyces pombe*-Mt (CAA21147.1), *Synechocystis sp.* (P73942.2), *Sulfolobus solfataricus* (CAB57731.1), *Thermotoga maritima* (AAD35599.1), *Treponema pallidum* (O83195.1), *Ureaplasma parvum* (AAF30864.1),

Z. may-Org (AY104268) and *Z. may*-Cy (AY104927). Abbreviations are as follows: mitochondrial (Mt), organellar (Org), and cytoplasm (Cy). The designations are those reported in the definitions within the GenBank accession.

The phylogenetic relationship of tRNA$_{Pro}$ was analyzed. The sequences of tRNA$_{Pro}$ were obtained from public databases, including NCBI and Saccharomyces genome databases. Sequence alignment was performed on tRNA$_{Pro}$ from *A. thaliana, S. cerevesiae* and *E. coli*. These included 66 tRNA$_{Pro}$ contained in the nuclear genome of *A. thaliana* and one each contained in mitochondria and chloroplast genomes; 12 tRNA$_{Pro}$ contained in the nuclear genome and one from mitochondrial genome of *S. cerevesiae* and 3 tRNA$_{Pro}$ from the genome of *E. coli*.

2.2.4 Plasmid construction

Plant ProRSs were cloned for expression in *E. coli*, protein purification and complementation. All plant genes were amplified by PCR and cloned into the appropriate vectors. The PCR primers, their intended purpose, and the cloning restriction sites for each construct are listed in Table 1. In the case of organellar ProRSs, At5g52520 and ZmProRS-Org, the segment encoding the mature protein lacking the signal/transit peptide was expressed. All constructs were prepared in pET30a (Novagen Inc). This vector places a His$_6$-tag at the N-terminus of the recombinant protein for use in affinity purification. For complementation, plant ProRS and *E. coli cysS* genes were amplified by PCR using primers that included a ribosomal binding site sequence placed 10 nucleotides upstream of the translational initiation codon. The PCR products were cloned into pBAD33 (Guzman et al., 1995), a plasmid that carries an arabinose-inducible and

glucose-repressible promoter for control of transgene expression. The sequence data for

E. coli cysS (M59381), obtained from NCBI, was used to design the primers (Table 1).

The PCR reactions (50 μL) contained 1 μg of genomic DNA, 0.5 mM of each dNTP, 3.8

pM primers, 1x PCR buffer (manufacturer supplied, Invitrogen, Inc.), and 1 U of Extaq

DNA polymerase (Invitrogen, Inc.). The PCR conditions were 1 cycle for 10 min at 94ºC

followed by 30 cycles of 30 sec at 94ºC, 30 sec at 55ºC, and 2 min at 72ºC. The PCR

products were sequenced to confirm that they were free of errors/mutations.

2.2.5 Complementation of *proS*^{ts} and cysS^{ts} mutants

Complementation was studied in the *proS*^{ts} and *cysS*^{ts} *E. coli* mutants UQ27 and

UQ818, respectively. The strains were transformed with the pBAD33 constructs carring

AtProRSs or *E. coli cysS* indicated in Table 1. Transformants were recovered on rich

medium containing 34 μg/mL chloramphenicol. Colonies were replica plated onto M-9

minimal medium supplemented with the 20 L-amino acids each at a concentration of 25

μg/mL and 0.2% ($^{W}/_{v}$) arabinose. The plates were incubated at 30ºC and 42ºC. Growth

was compared to a transformant carrying the empty vector pBAD33. A growth curve was

performed in M9 minimal liquid medium supplemented with the 20 L-amino acids and

arabinose or glucose at 30ºC and 42ºC. Optical density was measured at 600 nm every 2

hours.

2.2.6 Overexpression in *E. coli*

The pET30a constructs of AtProRS and ZmProRS were transformed into *E. coli*

BL21-Codon Plus (DE3) RIPL. Cultures were grown at 30ºC to an OD$_{600}$ of 0.5 and

protein expression was induced with 1 mM isopropyl-1-thio-β-D-glactopyranoside (IPTG) for 3 hours at 30ºC. Cells were collected by centrifugation at 4ºC for 10 min at 12,000 ×g and lysed by sonication in the purification buffer containing 137 mM NaCl, 10 mM sodium phosphate (pH 7.4) and 2.7 mM KCl. The soluble proteins were mixed with Ni-NTA (Clontech, Inc) for 4 hours and the resin was washed with protein purification buffer followed by wash buffer (pH 6.0) containing 120 mM sodium phosphate, 500 mM NaCl and 10 mM imidazole. The bound protein was eluted with wash buffer containing 200 mM imidazole. The concentration of the eluted protein was determined. The protein was concentrated using an Amicon filtration device. The purity of protein was determined by Coomassie Blue stained SDS-PAGE gels. The purified proteins were stored at -70ºC until use for kinetic experiments.

2.2.7 ATP-PP$_i$ exchange assay

Reactions were performed as described by Stathopoulos (Stathopoulos et al., 2000) and Ahel (Ahel et al., 2002). The basic scheme of the assay is shown in Figure 7. The reaction mixture, contained in a final volume of 200 μL, consisted of 50 mM HEPES-NaOH, pH 7.2, 15 mM MgCl$_2$, 50 mM KCl, 5 mM dithiothreitol, 1 mM potassium fluoride, 2 mM ATP, 0.5 mM L-Pro or A2C, 2 mM [^{32}P]pyrophosphate (specific activity of 3.12 TBq/mmol), 1 mg/mL tRNA and variable ProRS protein. Incubation was at 30ºC for varying times as indicated in the text. The temperature for incubation was chosen as described by previous papers (Dietrich et al., 1996; Pujol et al., 2008; Hogg et al., 2008; Igloi and Schiefermayr, 2009). Forty μL aliquots of the reaction were mixed with 200 μL acid-washed Norit (a 1% suspension (^{w}/$_v$) in 0.4 M sodium pyrophosphate and 15% (^{v}/$_v$)

perchloric acid). The Norit was recovered on Whatmann GF/C filter disks, washed with 15 mL of water and 5 mL of ethanol, and added to vials with 3 mL scintillation fluid (Ready-Safe; Beckman). Radioactivity was measured by liquid scintillation counting. The values of the kinetic constants were fitted to the Michaelis-Menten equation using SigmaPlot 8 Software (Systat Software Inc.).

2.2.8 Aminoacylation assay

Assays were performed as described by Ambrogelly (Ambrogelly et al., 2002) and Ahel (Ahel et al., 2002). The basic scheme of the aminoacylation assay is shown in Figure 8. The reaction buffer, contained in a final volume of 300 μL, consisted of 50 mM HEPES-KOH pH 7.2, 50 mM KCl, 15 mM $MgCl_2$, 5 mM dithiothreitol, 10 mM ATP, 20 μM [^{3}H]Pro (85 Ci/mmol), 1 mg/mL tRNA and 0.1 μg ProRS protein. Incubation was at 30ºC. Aliquots of 75 μL were transferred to 500 μL ice-cold 10% (^{w}/$_v$) trichloroacetic acid and filtered onto Whatmann 3MM filter disks. After washing three times with ice-cold 5% (^{w}/$_v$) trichloroacetic acid and 95% (^{v}/$_v$) ethanol, the filter disks were added to vials with 3 mL scintillation fluid (Ready-Safe; Beckman). Radioactivity was measured by liquid scintillation counting. The effect of A2C on aminoacylation was examined by adding various concentrations of A2C to an aminoacylation reaction containing 20 μM [^{3}H]Pro (85 Ci/mmol). The inhibition of activity in the presence of A2C was expressed as a percentage of the activity in a reaction lacking A2C.

2.3 Construction of cDNA libraries from *C. majalis* and *P. pubescens*

2.3.1 The harvesting of plant samples

The plant samples were harvested from landscape plantings in New Jersey gardens: *C. majalis* from East Brunswick, and *P. pubescens* from Colts Neck, during April and May. *C. majalis* root, stem, shoot, young leaf, mature leaf, young flower and mature flower tissues were harvested separately. *P. pubescens* tuber, root, stem, shoot, young leaf, mature leaf, young flower, and mature flower tissues were harvested individually. The samples were immediately frozen in dry ice and then stored at -70ºC until use.

2.3.2 Analysis of free amino acid content of plant tissues by HPLC

The free amino acids were extracted in 20 mM HCl and derivatized with 6-aminoquinolyl-N-hydroxysuccinimide carbamate (AQC) reagent using an AccQ-Fluor Reagent Kit (Waters, Inc., Milford, MA) as described in Kim et al. (2002). Amino acids were extracted, derivatized, resolved by HPLC and quantified from seven different *C. majalis* tissues: root (including rhizome and root hairs), stem, shoot (meaning emerging young stem from rhizome), young leaf, mature leaf, young flower, and mature flower. Amino acids were extracted, derivatized, resolved, and quantified in eight *P. pubescens* tissues including tuber, root, stem, shoot, young leaf, mature leaf, young flower and mature flower. Frozen tissue was ground to a fine powder under liquid nitrogen with a mortar and pestle. To the ground powder was added 20 mM HCl at a ratio of 10 μL HCl for each mg of tissue. The HCl also contained 20 pmol/μL of norvaline or norleucine as internal standards, and homogenization was continued. The crude tissue extract was centrifuged at 14,000g for 10 min at 4ºC. Amino acids were derivatized by mixing 5 μL of supernatant from the plant extract, 35 μL borate buffer and 10 μL of AccQ-Fluor

reagent. The mixture was incubated at room temperature for 1 min and diluted by adding 50 µL of 20% (^{v}/$_v$) acetic acid. 20 µL of samples was injected on to the HPLC column.

Separation and quantification of derivatized amino acids was performed using a Waters Alliance System with a 2690 separation module, a Waters fluorescence detector with excitation wavelength set at 250 nm. Emission was monitored at 395 nm, and Millenium 32 software version 3.2. A Waters AccQTagC$_{18}$ column (3.9 x 150 mm), coupled with a Nova-Pak C$_{18}$ guard column, was used for separation and analysis. Mobile phase A consisted of sodium acetate buffer and triethylamine at pH 5.8, mobile phase B was acetonitrile:water (30:70), and mobile phase C was acetonitrile:water (60:40). Mobile phase A was purchased as a concentrated premix from water and its pH was adjusted to 5.8 with 50% (^{w}/$_w$) sodium hydroxide. The elution was performed as follows: 0 to 0.5 min, 100% A; 0.5 to 1.5 min, linear gradient to 6.2% B; 1.5 to 32 min, linear gradient to 7.3% B; 32 to 51 min, linear gradient to 28% B; 51 to 66 min, linear gradient to 37% B; 66 to 84 min, linear gradient to 65% B; 84 to 97 min, linear gradient to 100% B; 97 to 101 min, linear gradient to 50% C; 101 to 104 min, linear gradient to 100% C; 104 to 106 min, linear gradient to 100% B; 106 to 108 min, linear gradient to 100% A; 108 to 114 min, 100% A. The flow rate was 1.5 mL min^{-1}, and amino acid standards were measured at various concentrations to establish a detector response factor and standard curves.

2.3.3 Isolation of total RNA and mRNA

All procedures used RNase-free plastic ware and reagents. Five grams of frozen leaves of *C. majalis* or *P. pubescens* each were ground to a powder with liquid N$_2$, and

total RNA was extracted using Concert™ Plant RNA Reagent (Invitrogen, Inc.). Five mL of cold Plant RNA Reagent were added to each 1 g of frozen, ground tissue and incubated at room temperature for 5 min. RNA was precipitated in 0.9 volume of isopropyl alcohol and washed with 75% (^{v}/$_v$) ethanol. After centrifugation at 12,000 rpm for 10 min, the RNA pellet was air dried and resuspended with 0.1% DEPC-treated water. The quality of RNA was examined by electrophoresis on a 1% (^{w}/$_v$) agarose gel, and the concentration of RNA was determined using a UV spectrophotometer at 260 nm.

mRNA was purified from total RNA by using NuceloTrap mRNA Purification Kit (BD Bioscience, Inc.) according to the manufacturer's protocol. Total RNA was mixed with NucleoTrap mRNA matrix based on oligo (dT)-latex beads and incubated at room temperature for 10 min. The NucleoTrap matrix was collected on a NucleoSpin Microfilter and washed with 75% ethanol. mRNA was eluted from NucleoTrap mRNA matrix using pre-warmed eluting solution at 65ºC (0.1% DEPC-treated water).

2.3.4 Construction of cDNA libraries

cDNA libraries from *C. majalis* and *P. pubescens* were prepared using the SMART™ cDNA Library Construction Kit (BD Bioscience, Inc.) according to the manufacturer's protocol. The vector used for cloning was λTriplEX2, a lambda phage from which a plasmid can be excised by CRE/LOX recombination. cDNA was synthesized from mRNA by primer extension, and cDNA was size-fractionated by gel filtration using CHROMA SPIN-400. The profiles of the fractions were analyzed by agarose/ethidium bromide electrophoresis. Ligation of cDNA to λTriplEX2 was carried out at three different ratios of cDNA to phage vector. Incubation was performed

overnight at 16 ºC. Packaging of recombinants was performed *in vitro* using MaxPlax Lambda Packaging Extracts (EPICENTRE, Inc, Madison, WI). The titer of unamplified libraries was measured to estimate the number of independent clones. The percentage of recombinants in libraries was determined using blue/white screening in *E. coli* XL1-Blue. To determine the average insert size, PCR was performed using primers against the flanking vector DNA. Randomly, 27 independent clones from each library were tested. Finally, the libraries were amplified, and the titer was determined.

2.3.5. Isolation of ProRSs from the libraries of *C. majalis and P. pubescens*

2.3.5.1. Functional complementation

To study the substrate specificity of ProRSs in *C. majalis* and P. *pubescens*, the primary focus of the research was to identify full-length cDNAs encoding ProRS from these libraries. I initially made an attempt to clone these cDNAs by functional complementation of the *E. coli* temperature sensitive ProRS mutant of UQ27. A plasmid based cDNA library was prepared from the phage libraries by mass-excision according to the manufacturer's instruction (BD Bioscience, Inc). The plasmid-based cDNA library was transformed into UQ27 by electroporation (Bio-Rad Gene Pulser, following the manufacturer's protocol) using a cuvette with 0.2cm electrode gap and the following conditions, 25μF, 200W, and 2.5kV. The transformants were recovered on rich medium containing 100 μg/ml ampicillin and complementing cDNAs were screened by replica plating of the transformants onto M9 minimal medium with 100 μg/ml ampicillin. The ability to complement was retested by retransforming each plasmid into the same mutant. Those plasmids that tested positive in the re-test were analyzed further by nucleotide

sequencing. The sequences were used to search the nucleotide databases to assess function based on homology with known sequences.

2.3.5.2 PCR amplification and sequence analysis

The disadvantage of the functional complementation screen is that it is likely to produce only the chloroplast ProRS cDNA. Ultimately, I would like to examine both the chloroplast and cytosolic ProRS, so additional cloning strategies are necessary. Another strategy was to use PCR method to amplify ProRS from *C. majalis* and *P. pubescens*. Primers were designed from highly conserved regions between the different members of the plant ProRSs (Fig. 29). Primers to isolate the organellar ProRS gene were ProRS-Org (M) F (5'-AGCAACATGTACTTCCC-3') and ProRS-Org (M) R (5'-AAATTGAACTTCCATCC-3'). Primers to isolated cytosolic ProRS gene were ProRS-Cy (M) F (5'-AACTTCGGGGARTGGTAYTCY-3') and ProRS-Cy (M) R (5'-CYTCATCACACCADGGAGCCA-3'). The PCR reaction (50 μL) contained 1 μg of each phage cDNA library, 0.5 mM of each dNTP, 3.8 pmol 5' and 3' ProRS-Org primers, 1x PCR buffer (manufacturer supplied), and 1 U of Extaq DNA polymerase (Invitrogen, Inc.). The PCR conditions were 1 cycle for 10 min at 95ºC followed by 30 cycles of 30 sec at 95ºC, 30 sec at 55ºC, and 2 min at 72ºC. For cytosolic ProRS, the PCR reaction (50 μL) contained 1 μg of each phage cDNA library, 0.5 mM of each dNTP, 3.8 pmol each of 5' and 3' ProRS-Cy degenerate primers, 1x PCR buffer (manufacturer supplied, Invitrogen, Inc.), and 1 U of Extaq DNA polymerase (Takara, Inc.). The PCR conditions to amplify cytosolic ProRS was 1 cycle for 10 min at 94ºC followed by 30 cycles of 30 sec at 94ºC, 30 sec at 45ºC, and 2 min at 72ºC. PCR products were analyzed using gel

electrophoresis. From the amplification of organellar ProRS, one PCR fragment from each *C. majalis* and *P. pubescens* library was further amplified. These PCR products were eluted from the gel and subcloned into pGEM-T Easy Vector (Invitrogen, Inc.). Subcloned cDNAs were sequenced and confirmed the identity by sequence alignment with other ProRSs.

2.3.5.3 Nested PCR

To clone 5' and 3' ends of ProRS from *C. majalis* and *P. pubescens*, a nested PCR method was used, and three pairs of specific primers were designed for each clone based on the sequence of the cloned cDNA. Figure 9 shows a diagram of nested PCR. 5' and 3' end nested primers were designed from cloned ProRS sequences in section 2.3.5.2. Primers are listed in Table 2. The nested PCR reaction (50 μL) contained 1 μg of phage cDNA library, 0.5 each dNTP, 10 μM of 5' or 3' gene specific primer, 10 μM 5' or 3' sequencing primer of pTriplEx2, 1x PCR buffer, and 1 U of Extaq DNA polymerase. The PCR conditions were 1 cycle for 10 min at 95ºC followed by 35 cycles of 30 sec at 95ºC, 30 sec at 60ºC, and 2 min at 72ºC. Three pairs of gene specific primers were used for each cDNA. The first reaction with each set of 5' or 3' sequencing and nested primer 1 amplified the cDNA, and the second reaction amplified the cDNA within the first PCR product by each set of 5' or 3' sequencing and nested primer 2. The second PCR product is shorter than the first PCR product. The possibility that the wrong cDNA was amplified is very low, because it would be amplified a third time by a third pairs of primers, each set of 5' or 3' sequencing and nested primer 3.

2.3.5.4 Rapid amplification of cDNA ends (RACE)

To clone the full-length cDNA sequence from a partial PpProRS cDNA, 5' RNA ligase mediated-rapid amplification of cDNA ends (RLM-RACE) was performed using the RLM-RACE kit (Ambion, Inc, Foster City, CA) according to the manufacturer's protocol. Poly (A) RNA was used as template. Calf intestine alkaline phosphatase (CIP) was used to treat poly (A) selected RNA to remove free 5'-phosphates from degraded mRNA, tRNA and genomic DNA. After removal of the phosphatase by phenol:chloroform extraction, tobacco acid pyrophosphatase (TAP) was then used to remove the cap structure found on intact 5'end of full-length mRNA. The cap structure is not affected by CIP. 5' RACE adapter oligonucleotide was ligated to decapped mRNA using T4 RNA ligase. The ligated RNA was reverse transcribed by using the FirstChoice RLM-RACE kit (Ambion Inc.). The reaction mixture (20 μL) for reverse transcription contained 50 ng of poly (A) RNA, 0.5 μg of 5' RACE adapter (5'-GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUCAUGAAA-3'), 1 U of placental RNase inhibitor, 0.5 mM each dNTP, 1x RT buffer (manufacturer supplied), and 10U of molony murine leukemia virus (M-MLV) reverse transcriptase. The mixture was incubated at 42ºC for one hour and primary, secondary, and third PCR was then performed. Three nested gene specific PCR primers were designed to perform nested PCR for 5' RLM-RACE PCR. Primes were Tail PCR 3'-1 (5'-AACTGAGTTCCAAAGGCACGA-3'), Tail PCR 3'-2 (5'-TGGCTGGTTCCAGCCTGTA-3'), and Tail PCR 3'-3 (5'-TTTCCCATCTTGTGACATTT -3'). The 5'-RLM-RACE PCR reaction (50 μL) contained 1 μL of RT reaction, 0.5 each dNTP, 10 μM of 5' RACE gene specific outer

(5'-GCTGATGGCGATGAATGAACACTG-3')   /inner   (5'-
CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG-3') primer, 10 μM of 5'
RACE outer/inner primer, 1x PCR buffer (manufacturer supplied), and 1 U of
thermostable DNA polymerase. The PCR conditions were 1 cycle for 5 min at 94ºC
followed by 35 cycles of 30 sec at 94ºC, 30 sec at 60ºC, 2 min at 72ºC and a final step of
7 min at 72ºC. The PCR product was run on a 2% ($^w/_v$) high resolution agarose gel and
shown several bands. Each DNA band was eluted from the gel and cloned into pGEM T-
Easy Vector (Promega, Madison, WI). Recombinant plasmids were sequenced. Sequence
analysis revealed that one of the cDNA sequences was homologous to other plant ProRS
genes.

2.3.5.5 Isolation of a genomic DNA

Inverse PCR with genomic DNA was another strategy attempted to clone the 5'-
end of the coding sequence bored on the truncated PpProRS cDNA. Since isolation of
high-quality DNA from wild plants is particularly difficult, genomic DNA was isolated
by two different methods. One was the DNeasy Plant Maxi Kit (QIAGEN, Valencia, CA)
and the other was a modified CTAB (Hexadecyltrimethylammonium bromide) procedure
(Xu et al., 2004). For the DNeasy method, one gram of frozen leaves of *C. majalis* or *P.
pubescens* each were ground to a powder with liquid N_2, and genomic DNA was
extracted according to the manufacturer's protocol. For the CTAB method, five grams of
frozen leaves of *C. majalis* or *P. pubescens* each were ground to a powder with liquid N_2,
and genomic DNA was extracted using a modified CTAB procedure described by Xu et
al. (2004). The quality of DNA was examined by electrophoresis on a 1% ($^w/_v$) agarose

gel, and the concentration of DNA was determined using a UV spectrophotometer at 280 nm. These genomic DNA preparations were used for inverse PCR.

2.3.5.6 Inverse PCR

To obtain flanking regions of 5'-end sequence of PpProRS, inverse PCR was performed using procedure described by Ochman et al. (1988). Restriction digests were carried out using 5 µg of DNA treated with 10 units of *BglII*, *EcoRV* and *XbaI*. The restriction fragment was circularized by ligation reaction. The ligated sample was treated with phenol:chloroform mixture and precipitated with ethanol. Then primary, secondary, and third PCR was performed. Three gene specific 5' primers and one 3' primer were designed to perform inverse PCR. Primes were IPCR 5'-1 (5'-ATCTTGTGACATTTGCCCACT-3') and IPCR 5'-2 (5'-AACAGATGTAGCAGCATT-3'), IPCR 5'-3 (5'-TAGACACAACTACACTCC-3'), and IPCR 3' (5'-TTTCCCATGAGTGTTAAATTA -3'). The PCR was performed in reaction containing 0.1 µg of circularized DNA as described at section 2.3.5.4. The PCR conditions were 1 cycle for 5 min at 94ºC followed by 30 cycles of 1.5 min at 94ºC, 1.0 min at 48ºC, 4 min at 72ºC and a final step of 7 min at 72ºC.

2.3.5.7 Overexpression of PpProRS in *E. coli* and ATP-PP$_i$ exchange assay

The PpProRS, coding sequence was amplified with primers PpProRS5'-*KpnI* (5'-CC**GGTACC**ATGAACACTGCGTTTGCT-3'), and PpProRS3'-*SalI* (5'-CC**GTCGAC**TTAATAAGATCTTGCAAA-3'). The PCR product was cloned into the pGEM-T Easy (Promega, Inc.) and the fidelity of the sequence confirmed. The *KpnI-SalI*

fragment was subcloned into pET30a and the construct was transformed into *E. coli* BL21-Codon Plus (DE3) RIPL. The protein was expressed and purified as described above (Section 2.2.6). The concentration of protein was determined by using the BioRad assay, and the purity of protein was determined by Coomassie stained SDS-PAGE gels. ATP-PPi exchange activity was measured as described in section 2.2.7.

3. Results

3.1 Characterization of ProRSs from *A. thaliana*

A number of plants, including *A. thaliana*, have been shown to be sensitive to A2C (Verbruggen et al., 1996). Therefore, analysis of the ProRSs from *A. thaliana* and other A2C-sensitive species provides a basis for comparison with ProRS from *P. pubescens* and *C. majalis*. An earlier study suggested that the resistance of *C. majalis* to A2C may be due to expression of ProRS with altered substrate specificity (Norris and Fowden, 1972). Toward this end, ProRSs from *A. thaliana* were expressed in *E. coli* for analysis of their catalytic properties and substrate specificity.

3.1.1 A2C inhibits *A. thaliana* root growth

A2C is known to inhibit the growth of *A. thaliana* (Verbruggen et al., 1996). To further examine this phenomenon, *A. thaliana* was germinated on media with varying concentrations of A2C. Figure 10A shows that root growth is more severely inhibited than shoot growth. At 3 μM A2C and higher, root length was severely inhibited and the root system became branched. Measurement of root length revealed that the concentration of A2C producing 50% inhibition of root length is approximately 2.5 μM (Fig. 10C). The inhibition by A2C is completely reversed by addition of 100 μM Pro to the growth medium (Fig. 10B and C). Growth inhibition by A2C caused result from the mis-incorporation of A2C into protein and/or from the feedback inhibition of Pro synthesis by down regulation of pyrroline-5-carboxylate reductase. The growth experiment shown in Fig. 10 did not distinguish between these two mechanisms.

3.1.2 Analysis of AtProRSs from *A. thaliana*

A. thaliana contains two genes that may encode ProRSs, locus At5g52520 and At3g62120. Both genes are known to be expressed as evidenced by the existence of full-length cDNAs. The protein encoded by At5g52520 has an N-terminal extension with features characteristic of a transit peptide. A peculiarity of *A. thaliana* is that there is not a separate gene for a mitochondrial-targeted ProRS. This paradox was recently solved by Duchene et al. (2005) who showed that the At5g52520 product is targeted to both chloroplasts and mitochondria. Indeed, most organellar AARSs in *A. thaliana* show dual targeting specificity (Duchene et al., 2005). The same is likely also true for most plant AARSs. The subcellular localization of At3g62120 has not been determined experimentally; however, it has been predicted to be localized in the cytosol (Duchene et al., 2005).

Analysis of the amino acid sequences of At5g52520 and At3g62120 revealed that they share 41% similarity. Analysis of the At5g52520 and At3g62120 genes revealed that they have a divergent genomic structure. At3g62120 contains 12 exons and 11 introns and the coding sequence is 530 codons, whereas At5g52520 contains 10 exons and 9 introns with a coding sequence of 543 codons, the first 61 codons of which encode the transit peptide. The positions of two introns in each of the ProRS genes are exactly conserved, but all other intron positions are not conserved. The position of intron 3 in At3g62120 matches the position of intron 1 in At5g52520, and the position of intron 6 in At3g62120 matches the position of intron 5 in At5g52520 (Fig. 11). The conservation of intron position in the two AtProRS genes is evidence of common ancestry during evolution.

A phylogenetic comparative analysis of ProRSs from many organisms was performed to gain insight in to the evolutionary relationship of plant organellar and cytosolic ProRSs. The results shown in Figure 12 reveal that both of the plant ProRS types, cytosolic and organellar appear to share a common ancestry with the eukaryote/archaea clade. None of the plant ProRS sequences clustered with the mitochondrially localized forms of ProRS from animal and fungal sources, or the prototypical bacterial ProRS, typified by that from *E. coli* and cyanobacteria. These results support the notion that plant cytosolic and organellar ProRSs are closely related to cytosolic forms of ProRS from other eukaryotes and that plant probably did not retain the prokaryote-like ProRS that accompanied the endosymbiotic ancestor of mitochondria or chloroplasts.

Prokaryotic and eukaryotic ProRS shows the structural difference as described at section 1.3.1. To demonstrate the structural features, AtProRS amino acid sequences were aligned with several plants (Fig. 33) and *E. coli* ProRSs (data not shown). The result shows that the AtProRS share the structural features of eukaryotic ProRS. AtProRS does include three distinct motifs of the catalytic domain (indicated with dotted lines in Fig. 33) and the anticodon binding motif (indicated with dots in Fig. 33). In addition, AtProRS bears C-terminal extension beyond the anticodon-binding domain instead of the prokaryotic insertion domain between motif 2 and 3.

3.1.3 Only the organellar AtProRS is able to complement an *E. coli* temperature-sensitive ProRS mutant

The function of At3g62120 and At5g52520 was examined under *in vivo* conditions by testing whether each sequence was able to functionally complement UQ27, an *E. coli* temperature-sensitive ProRS mutant strain. This strain is able to grow at the permissive temperature of 30°C, but is unable to grow at the non-permissive temperature of 42°C. At3g62120 and At5g52520 cDNAs were cloned into pBAD33, a plasmid that carries an arabinose-inducible and glucose-repressible promoter for control of transgene expression. Figures 13A and 13B show that all the strains were able to grow at 30°C but only At5g52520 was able to grow at 42°C. Neither the vector nor a CysRS-expressing plasmid was able to complement UQ27 (Fig. 13). To confirm the complementation result, a growth curve was carried out in liquid culture medium. The results show that all transformed UQ27 strains grew well in liquid minimal medium at 30°C, regardless of the carbon source. However, only the strain carrying At5g52520 was able to grow at 42°C in the presence of arabinose but not in the presence of glucose (Fig. 14). This result confirms that complementation of UQ27 results from arabinose-dependent expression of At5g52520 protein from pBAD33-At5g52520.

Dual-specificity prolyl-cysteinyl-tRNA synthetases (ProCysRSs) were recently demonstrated from archaeal species including *Methanococcus jannaschii*, *Methanococcus maripauldis* and *Methanobacterium thermoautotrophicum* (Stathopoulos et al., 2000) and from the lower eukaryote *Giardia lamblia* (Bunjun et al., 2000). Since the plant ProRSs clustered with the dual specificity enzymes (Fig. 12), it was of interest to determine whether *A. thaliana* ProRSs have CysRS activity. To examine this question, functional complementation of the temperature-sensitive *E. coli cysS*[ts] mutant UQ818 was attempted. As a positive control complementation was also examined with the wild-

type CysRS from *E. coli*. Neither *A. thaliana* enzyme was able to complement the mutant. However, in the same experiment, native CysRS of *E. coli* was able to complement the mutant (Fig. 15B). Based on this experiment, it is tentatively concluded that at least At5g52520 does not display CysRS activity at 42°C. It is not possible to conclude that At3g62120 does not have CysRS activity because no activity from this cDNA was demonstrated in either complementation experiment (Figs. 13 and 14).

3.1.4 Both AtProRSs are active in ATP-PP$_i$ exchange

At5g52520 and At3g62120 proteins were expressed in *E. coli*, purified and their ATP-PP$_i$ exchange activity measured. The purity of the proteins was determined by SDS-PAGE followed by staining with Coomassie Brilliant Blue (Fig. 16). The calculated molecular weight of the At3g62120 fusion proteins is 61 kDa and the calculated molecular weight of the At5g52520 is 54 kDa. The gels in Figure 16 show that proteins matching the predicted molecular mass are expressed and a purifiable by Ni-agarose of affinity chromatography. The At3g62120 protein was nearly pure, but the At5g52520 protein contained a few minor contaminating proteins.

To determine if the purified proteins are functional enzymes, they were assayed for ATP-PP$_i$ exchange activity. Figures 17 and 18 show that both of AtProRSs are functional in ATP-PP$_i$ exchange activity. The activity was directly proportional to the amount of enzyme in the reaction (Figs. 17A and 18A). And the reactions progressed steadily over a period of 30 minutes (Figs. 17B and 18B). No activity was detected if Pro was omitted from the reaction or if Pro was replaced with either Met or Cys (data not shown). These results indicate that both the cytosolic and the organellar forms of ProRS from *A.*

thaliana show robust Pro-dependent PP$_i$ exchange activity. The fact that At3g62120 protein shows ATP-PP$_i$ exchange activity with Pro demonstrates that the inability to complement UQ27 is not due to the production of a catalytically inactive protein or enzyme that is incapable of forming a Pro-adenylate.

3.1.5 Cytosolic and organellar AtProRSs have differing tRNA substrate specificity

The Pro adenylylation activity of both *A. thaliana* ProRSs was demonstrated *in vitro*, yet only At5g52520 was capable of complementing the *proS* mutation of UQ27. Since complementation would require that the plant enzyme carry out both adenylylation and aminoacylation reactions, the simplest explanation for the inability of At3g62120 to complement UQ27 is that it cannot aminoacylate *E. coli* tRNAs. Sequence divergence during the course of eukaryotic evolution of AARS has generated considerable structural differences in the AARS and in the tRNA substrate between cytosolic and organellar compartments. It is known that organellar AARSs are unable, in general, to use cytosolic tRNAs as substrate, and cytosolic AARSs are generally unable to use organellar tRNAs (Reger et al., 1970; Imbault et al, 1981). Organellar tRNAs are structurally closer to prokaryotic tRNAs (Marechal-Drouard et al, 1993). To address this question, the tRNA substrate specificity of At3g62120 and At5g52520 was tested by examining the ability to amino-acylate *E. coli* tRNA (prokaryotic type tRNA) and yeast tRNA (eukaryotic type tRNA). This approach was justified by the close homology of *A. thaliana* mitochondrial tRNA with *E. coli* and *S. cerevisiae* mitochondrial tRNA. Fig. 20 is a phylogenetic tree showing the sequence relationship of *E. coli*, *S. cerevisiae* and *A. thaliana* tRNAs. The results showed that At5g52520 was equally capable of aminoacylating either *E. coli* or

yeast tRNA (Fig. 19A). This result is consistent with the complementation result and also with the likelihood that the yeast tRNA preparation is a mixture of both cytosolic and mitochondrial tRNA. In contrast, At3g62120 was significantly more efficient at charging yeast tRNA than *E. coli* tRNA (Fig. 19B). This result demonstrates conclusively that At3g62120 is a functional ProRS. However, the fact that it was capable of charging *E. coli* tRNA, although less efficiently than yeast tRNA, is not consistent with the complementation result. It should be noted, however, that the concentration of *E. coli* tRNA used for the assay was based on a standard protocol. It is not known whether the same concentration exists in the *E. coli* protoplasm.

3.1.6 Two AtProRSs have different specificity for A2C

To investigate the amino acid specificity of AtProRSs, A2C was tested in place of Pro. Both At5g52520 and At3g62120 were capable of A2C-dependent PP_i exchange activity. To further explore the quantitative aspects of this activity, either Pro or A2C was titrated in the assay to determine the K_m value for each substrate. The substrate specificity of AtProRSs for these amino acids is shown in Figures 21A and 21B. The concentration of Pro or A2C was varied from 0.01-5 mM. Whereas the K_m value of At5g52520 is 0.10 mM for Pro, and 13.55 mM for A2C, the K_m value of At3g62120 is 0.06 mM for Pro and 0.04 mM for A2C (Table 3). Interestingly, both AtProRSs have similar K_m for Pro, but At3g62120 shows a 340-fold higher affinity for A2C, than does At5g52520. To determine the enzyme efficiency, the K_{cat}/K_m parameters for Pro and A2C were revealed (Table 3).

To further examine the difference in substrate specificity of the AtProRSs, the ability of A2C to compete with Pro in the aminoacylation assay was examined. It is important to note that in the PP$_i$ exchange assay the ability of the enzyme to acylate A2C was assessed. In the present experiment, the ability of A2C to inhibit incorporation of [^{3}H]Pro into tRNA is examined. It could be predicted that if there is indeed a difference in specificity, A2C would be a more effective competitive inhibitor of the Pro aminoacylation reaction for At3g62120 than for At5g52520. Figure 22 shows that A2C is indeed a better inhibitor of At3g62120 than At5g52520. Whereas At3g62120 is inhibited by 50% at approximately 100 μM A2C, At5g52520 was not inhibited at this concentration of A2C. Indeed, at 200 μM A2C, the highest concentration tested, At3g62120 was inhibited by only 25%. These results confirm the greater specificity At5g52520 for Pro compared to A2C.

3.1.7 Two ZmProRSs also have different specificity for A2C

The significant difference in $K_{m[A2C]}$ between *A. thaliana* ProRS isoforms indicates that they must have significant active site differences despite having evolved from a common ancestor. The result also raises the question of whether the difference in A2C specificity is a conserved feature of plant cytosolic and organellar ProRSs. Analysis of the DNA sequence databases identified *Z. mays* and *O. sativa* sequences that are closely related to both *A. thaliana* ProRS types. The two clones from *Z. mays* were used for protein expression and analysis. ZmProRS enzymes were expressed in *E. coli* and purified using NI-NTA chromatography. The proteins were judged to be 95% pure by SDS-PAGE followed by staining with Coomassie Brilliant Blue (Fig. 23). The induced *E.*

coli cell culture accumulated a 58 kDa protein for ZmProRS-Cy and a 55 kDa protein for ZmProRS-Org (lanes 2, 3, 5 and 6). The molecular masses of both ZmProRS proteins expressed from the pET30a construct were similar to that predicted by their amino acid sequences. ZmProRS-Cy was nearly pure, but ZmProRS-Org contained a few minor contaminating proteins. To determine if the purified proteins are functional, they were assayed by ATP-PP$_i$ exchange activity. Figure 24 shows the rate of exchange activity of ZmProRS-Cy as a function of enzyme concentration and time. The concentration of ZmProRS-Cy was varied from 0.1-1 µg in (A), and the Pro concentration was 0.5 mM in (A) and (B). There was no activity in the absence of Pro or enzyme, and the activity was unaffected by addition of tRNA to the assay (data not shown). Figure 25 shows the exchange activity of ZmProRS-Org protein under the same conditions. Both ZmProRSs exhibited robust PP$_i$ exchange activity. The substrate specificity of ZmProRSs for amino acids is shown in Figures 26A and 26B. The concentration of Pro or A2C was varied from 0.01-5 mM. Whereas the K_m value of ZmProRS-Org is 0.13 mM for Pro, and 1.51 mM for A2C, the K_m value of ZmProRS-Cy is 0.09 mM for Pro and 0.35 mM for A2C (Table 4). ZmProRS-Cy shows a 10-fold higher affinity for A2C, than does the organellar enzyme.

3.2 Construction of cDNA libraries of *C. majalis* and *P. pubescens*

The initial objective of the project was to examine whether ProRS from *C. majalis* has significantly altered substrate specificity when compared with ProRS from non-A2C accumulating species. Based on the work of Norris et al. (1972), *C. majalis* ProRS would be predicted to show a far lower affinity for A2C than the comparable ProRS from a non-

A2C-accumulator. In order to examine *C. majalis* ProRS, a cDNA library was produced. An additional experiment was also initiated to examine whether the amino acid specificity of ProRS is a general feature of A2C-accumulating species. For this reason, a cDNA library was also prepared from *P. pubescens*. This species has never before been examined for A2C-accumulation nor has the ProRS from this species been studied.

3.2.1 Identification of local populations of *C. majalis* and *P. pubescens* and analysis of A2C content

    C. majalis and *P. pubescens* are different species within the Liliopsida (the monocotyledonous plants) in the family Convallariaceae. The common name of *C. majalis* is 'Lily-of-the-Valley' and *P. pubescens* is 'Hairy Solomon's Seal'. Both of these species grow in an area of full to partial shade and medium moisture, and are commonly cultivated as garden herbaceous perennial plants in many areas of North America. *C. majalis* is of European origin, and *P. pubescens* is a North American native species. Both species produce rhizomes from which new leaves and short fibrous roots emerge each year (Fig. 27). Leaves are produced in early-late spring and flowering occurs in early-mid summer. *C. majalis* is an attractive garden plant that produces white bell-shaped flowers with a sweet scent. Colonies of *C. majalis* were identified in Colts Neck, New Jersey, growing in a moist and open forest location; and in East Brunswick, NJ, growing in an open sunny location in hard red clay. A colony of *P. pubescens* was identified growing in the vicinity of the *C. majalis* colony in Colts Neck. Plant samples were harvested during April and May. A2C analysis of *C. majalis* and *P. pubescens* tissues was carried out using HPLC, as described in Materials and Methods (Table 5). The results demonstrate

that A2C in *P. pubescens* is 1.85% of the dry weight, and in *C. majalis* it is 0.66%. Interestingly, *P. pubescens* tissues showed a 3-fold higher level of A2C than *C. majalis*.

3.2.2 Identification of the temporal and spatial accumulation of A2C in the *C. majalis* and *P. pubescens* plants

To determine in which tissues and at what time in development A2C accumulates, A2C was resolved and quantified in various tissues of *C. majalis* and *P. pubescens* by HPLC. The results of amino acid analysis were shown in Tables 6 and 7. All tested tissues of *P. pubescens* and *C. majalis* showed high levels of A2C. There were tissue-specific differences between *P. pubescens* and *C. majalis*. While *C. majalis* accumulated more A2C in leaves and flowers, *P. pubescens* accumulated a higher level of A2C in stems and rhizomes. The A2C concentration in roots was not as high as in other tissues in either species.

The levels of proline and Met were quantified in the same tissue samples used to measure A2C (Tables 6 and 7). Since Pro is the structural analog of A2C and Met is proposed to be the precursor of A2C (Leete et al., 1986), it may be useful to compare the levels of A2C with Pro and Met. The HPLC method for measurement of A2C also provides a facile assay for other soluble amino acids. From all tissues of *C. majalis* and *P. pubescens,* the Pro level was 10- to 50-times lower than A2C. The level of Met was nearly 600-fold lower compared to A2C in both species. Interestingly, the HPLC analysis revealed the presence of a major amine compound in leaf and flower tissues of *C. majalis* that is equal in abundance to that of A2C (Fig. 28). Although the identity of this

compound was not explored, based on its abundance it is tempting to speculate that the unknown amine compound may be an A2C derivative.

3.2.3 Preparation of cDNA libraries from *C. majalis* and *P. pubescens*

Since the genomes of *C. majalis* and *P. pubescens* are not sequenced, it was necessary to prepare cDNA libraries that can be used to screen for the desired sequences using functional and hybridization assays. Based on the results of HPLC, the tissue with the highest level of A2C is young leaf of *C. majalis* and the rhizome of *P. pubescens*. Thus cDNA libraries were prepared from young leaf, stem, and rhizome tissues of *C. majalis* and *P. pubescens* using the SMART^{TM} cDNA Library Construction Kit (BD Biosciences, Inc.). The frequency of white (recombinant) plagues to blue (non-recombinant) was estimated at >90% in both libraries indicating that 90% of the plasmids are recombinant. The titer of the un-amplified libraries was determined to be >10^6, meaning that the libraries contained at least this number of independent clones. Finally, the libraries were successfully amplified to a very high titer (~10^{10} pfu/mL). To determine the average insert size, PCR was performed using primers to the regions flanking the cDNA insertion site of the vector. Twenty seven randomly selected clones from each library were tested. The average inset size was found to be about 1150 bp from the young leaf library of *C. majalis* and the young leaf library of *P. pubescens*. The libraries from stem and tuber gave an average insert size below 1000 bp. Therefore, only cDNA libraries from leaf tissues were used for screening since they were of higher quality.

3.3 Cloning of known genes to test the quality of the cDNA libraries

The challenges of isolating functional DNA, RNA, and proteins from "wild" plant samples are well known. Thus, since the average insert size was rather low for both cDNA libraries, it was necessary to examine whether it would be possible to clone full-length cDNAs for genes that are in the expected size range of approximately 1500 bp for known plant ProRSs. Moreover, it was of interest to determine whether functional cDNAs could be cloned from the libraries. To examine whether full-length, functional cDNAs could be cloned from these libraries, a well-established complementation assay was used to screen for cDNAs encoding 5'-adenylylsulfate reductase (APR) and cystathionine γ-synthase (CGS). A functional complementation assay, whereby plant cDNAs that can restore the growth of auxotrophic mutant *E. coli* strains has previously been used by the Leustek lab to clone these cDNAs from plant and algal species. APR is a plant gene necessary for Cys synthesis, and CGS is necessary for Met synthesis. The *A. thaliana* cDNAs for APR are approximately 1400 bp and CGS is 1200 bp. An additional advantage to cloning the CGS cDNA from *C. majalis* and *P. pubescens* is that if Met is the precursor of A2C, the CGS cDNA could prove useful later if gene expression related to A2C synthesis were studied. Full-length cDNAs encoding both enzymes were successfully cloned from both *C. majalis* and *P. pubescens*, demonstrating the usefulness of the cDNA libraries. Analysis of the cloned APR and CGS cDNAs is described in Appendix A.

3.4 Cloning of ProRS cDNAs from *C. majalis* and *P. pubescens*

3.4.1. Functional complementation

An initial attempt was made to clone cDNAs encoding ProRS from *C. majalis* and *P. pubescens* by functional complementation of the *E. coli* temperature sensitive ProRS mutant of UQ27. Unfortunately, this mutation showed such a high reversion frequency that it was impossible to identify true complementing cDNAs. 1,000,000 transformants were screened and hundreds of positives were identified. DNAs of approximately 35 clones were analyzed by nucleotide sequencing. The sequences were used to search the nucleotide databases to assess function based on homology with known sequences. However, none of positive clones showed homology with ProRS sequences.

3.4.2 PCR strategy

As another strategy, PCR was performed to clone the ProRS cDNA from *C. majalis* and *P. pubescens* phage cDNA libraries. Oligonucleotide primers were designed corresponding to highly conserved regions of plant ProRS examples in the sequence databases. The organellar type ProRS sequences from *A. thaliana, Z. mays, O. sativa* and *R. sativus* were aligned using ClustalW as were cytosolic ProRS sequences from *A. thaliana*, *Z. mays* and *O. sativa*. This analysis revealed several highly conserved regions to which primers were designed (Fig. 29). Only the primer design strategy for organellar ProRS is shown in Fig. 29 since PCR cloning of the cytosolic ProRS was unsuccessful. The ProRS-specific primers were used with *C. majalis* and *P. pubescens* cDNA library as a template. Figure 30A shows the strategy for cloning full length ProRS from *C. majalis* and *P. pubescens*. The phage cDNA libraries were used as template for PCR. A PCR product of 756 bp was amplified from *C. majalis* cDNA library and a 712 bp product was amplified from *P. pubescens*. The PCR products were sequenced and found to be

homologous with plant organellar ProRS sequences. The highest homology was found with *Ricinus communis* ProRS, with which the PCR products showed 80% identity.

3.4.3. Cloning of full-length ProRS cDNAs

Having positively identified a PCR product for ProRS from *C. majalis* and *P. pubescens* in the next step an effort was made clone a full-length cDNA from each species using a nested PCR strategy (see Fig. 9 and section 2.3.5.4 for details on the nested PCR procedure). PCR products were isolated corresponding to 200 bp from the 5'-end and 600 bp from the 3'-end of the *P. pubescens* ProRS, and 500 bp from the 5'-end and 750 bp fragment from 3'-end of ProRS from *C. majalis*. The secondary PCR product sequences were identical to the primary PCR products in the overlap regions confirming that they were derived from the same cDNA (Figs. 30 and 31). In addition, the secondary PCR products from the 3' region of both *C. majalis* and *P. pubescens* contained poly (A) sequences further confirming that they were derived from cDNA. In order to clone the open reading frames of a single unit ORF-specific G and H primers were designed (Figs. 30A and 31A, IV). A contiguous 1,368 bp PCR product was obtained from *P. pubescens*. The nucleotide sequence and deduced amino acid sequence of the *P. pubescens* PCR product is presented in Figure 32. The ORF is 438 codons in length. The deduced ProRS sequence amino acid sequence shows the highest homology with a *Z. mays* ProRS sequence which it shows 74% identity. An alignment of PpProRS amino acid sequence with several plant ProRS is shown in Fig. 33. The result shows that the cloned PpProRS most likely encodes the organellar-type ProRS.

The alignment also reveals that PpProRS lacks an amino terminal extension present in the other ProRSs (Fig. 33), suggesting that the clone is missing 5'-region of the coding sequence. The PpProRS clone does encode the three distinct motifs of the catalytic domain (indicated with dotted lines in Fig. 33), and the anticodon binding motif (indicated with dots in Fig. 33). Given that the PpProRS sequence is very likely a 5'-truncated clone, two additional efforts were made to clone the 5'-region including 5' RLM-RACE PCR using isolated total RNA and inverse PCR using genomic DNA from P. pubescens as template. Neither technique proved successful.

3.4.4. Overexpression of PpProRS and ATP-PPi exchange assay

The PpProRS protein was functionally tested. The protein was expressed in *E. coli* as an amino terminal a fusion with a His-tag. The protein was purified by Ni-affinity chromatography and the purity of the preparation examined by SDS-PAGE, showed that a ~ 50 kDa and 40 kDa protein was purified. The 50 kDa protein the molecular weight expected of the fusion protein. The large quantity of the 40 kDa protein suggests that it may be a cleavage on truncated form of PpProRS, but it was not further examined. Figure 34 shows PpProRS protein had no ATP-PP_i exchange activity. There are two possible reasons for no activity. One is that truncated PpProRS lacks a domain at the amino terminus that is critical for activity. Another possibility is that the contaminating proteins may inhibit the activity of PpProRS protein (Fig. 34).

4. Discussion

4.1 The evolution of ProRS in plants

In plant cells protein synthesis takes place in three subcellular compartments: the cytosol, mitochondria, and plastids. Thus a full set of AARSs and tRNAs must be present within each compartment. Neither the plastid or mitochondrial genomes encode AARSs, rather, these enzymes are encoded in the nuclear genome, are translated on cytoplasmic ribosomes, and are then transported in to the organelles. A recent discovery showed that most plant AARSs targeted to mitochondria and chloroplasts are encoded by the same nuclear gene (Duchene et al., 2005; Rokov-Plavec et al., 2002; Rokov-Plavec et al., 2008). Among 24 organellar-targeted AARSs in *A. thaliana,* only two examples exist of enzymes that are uniquely targeted to plastids; 17 exhibit dual-targeting to mitochondria and plastids, and five are targeted to both the cytosol and mitochondria (Duchene et al., 2005). The organellar-targeted form of ProRS is among the examples of mitochondrial and plastid co-targeted AARS.

The *A. thaliana* nuclear genome contains two genes encoding ProRS. One of these genes encodes the cytosolic enzyme and the other encodes the dual-targeted organellar form. Based on amino acid sequence homology, it can be predicted that the same catalog of ProRS genes and dual-targeting generally occurs in plants. The phylogenetic analyses presented in Figure 12 shows that both of the *A. thaliana* ProRS share a common evolutionary origin with archaeal ProRS and the cystosolic ProRS of heterotrophic eukaryotes, all of which form a monophyletic group. The mitochondrial ProRS from heterotrophic eukaryotes and the eubacterial ProRS, including all the cyanobacterial ProRSs, belong to a second, distinct monophyletic group. ProRS orthologs

of the later monophyletic group are not found in plants. Assuming that a eubacterial type ProRS gene must have been brought into plants with the endosymbiont ancestors of mitochondria and plastids [presuming that the endosymbionts were eubacteria as has been proposed (Gray et al., 2001; Gould et al., 2008)], such a gene must, therefore, have been lost during the evolution of plants. In contrast, such a loss did not occur during the evolution of fungi and animals in which a eubacterial type ProRS exists that is targeted to mitochondria. The significant sequence difference between the archaeal and eubacterial ProRS types suggests that convergent evolution could not account for the present catalog of ProRS genes in plants and that a more likely scenario is that the archaeal/eukaryotic cytosolic ProRS was duplicated and one copy was co-opted as the source of both the mitochondrial and plastid enzymes. No sequence or biological information exists that point to the basis for the evolution of the organellar ProRS in the plant lineage. One possibility may be that dual-targeting of a single form of ProRS (and other AARSs for that matter) might provide a means of coordinating protein synthesis in the two organellar compartments.

4.2 Biochemical properties of plant ProRS- amino acid specificity and implications for A2C toxicity

Until now all plant ProRS genes were annotated based exclusively on sequence homology. The work reported here is the first experimental demonstration of the function of *A. thaliana* and *Z. mays* ProRS. The results reveal that both cytosolic and organellar enzymes are able to catalyze proline adenylylation and the proline aminoacylation of tRNA. Although a complete kinetic characterization of the plant ProRSs was not

performed (because it was not the primary objective of my project), the limited analysis revealed that both the cytosolic and organellar forms of ProRS show kinetic constants that are within the range previously reported for other prokaryotic and eukaryotic species (Beuning and Musier-Forsyth, 2001). Thus, despite the difference in evolutionary origin, all ProRS, including those from *A. thaliana* and *Z. mays* appear to share similar kinetic properties.

A major difference was observed between the cystosolic and organellar ProRS with respect to the misutilization of the proline structural analog A2C. Whereas the cytosolic form of ProRS showed a $K_{m[A2C]}$ from 4.7 to 13 times greater than for proline (depending on the species source), the organellar form showed $K_{m[A2C]}$ 16 to 135 times greater than for proline (Tables 3 and 4). This difference was confirmed for the *A. thaliana* ProRSs by the finding that A2C is a more efficient inhibitor of proline aminoacylation activity of the cytosolic enzyme than the organellar enzyme (Fig 22), and providing a second line of evidence that the organellar enzyme does indeed have a greater ability to distinguish proline from A2C. This result provides new information on the comparative properties of ProRSs. For example, Beuning and Musier-Forsyth (2001) reported that whereas ProRSs from *E. coli* and *M. jannaschii* are capable of efficient A2C mis-utilization, the cytosolic ProRS from human does not utilize A2C under the conditions of their assay (250 mM A2C). Thus, there appears to be a major difference in the ability of plant and animal cytosolic ProRS to distinguish between proline and A2C. Indeed, with regard to A2C misutilization, the organellar plant ProRS more closely resembles the human cytosolic enzyme. All three ProRSs, human cytosolic and plant cystosolic and organellar forms, are closely related and belong to the archaeal/eukaryotic

cytosolic clade. The difference in affinity for A2C must, therefore, reflect differences in the amino acid binding site of the enzyme. Thus, it may be that the amino acid binding site of plant cytosolic ProRS was altered through mutation such that it now can accept A2C, a compound with one fewer carbon in the ring moiety than proline (see Fig 1).

The difference in the ability of cytosolic and organellar ProRS to utilize proline and A2C impacts our understanding of the mechanism for A2C toxicity. As was shown in Figure 10, *A. thaliana* is highly sensitive to A2C and the initial growth inhibitory effect is greater for roots than for shoots. The misincorporation of A2C in to proteins has been implicated as primary mechanism for A2C toxicity (Fowden and Richmond, 1963; Rubenstein et al., 2006), although other means have also been proposed, such as mis-regulation of proline synthesis (Hoshikawa et al., 2003). The implication for the difference in amino acid substrate specificity is that proteins translated in the cytosol will have a higher A2C incorporation rate than proteins translated in the organellar compartments. Therefore, it is possible that the differential response of root and shoot growth to A2C exposure may result from the greater proportion of total gene expression and translational activity of plastids in leaves than in roots (Isono et al, 1997). It is important to point out that other mechanisms could also contribute to the greater sensitivity of roots to A2C. Whereas roots are in immediate contact with A2C in the medium, it must be translocated to leaves, thus the A2C concentration might be lower in leaves under the conditions used in the experiment shown in Fig. 10. The free proline concentration in leaves of Arabidopsis is known to be much greater than in roots (Lee et al., 2005). Therefore, another possibility is that the higher proline concentration in leaves could serve as an antidote for A2C. Finally, it is possible that there is a difference in the

ability of roots and leaves to catabolize A2C, although there is currently no evidence indicating that plants have the ability to degrade A2C.

4.3 Biochemical properties of plant ProRS- tRNA specificity

At3g62120 and At5g52520 also displayed a difference in the ability to utilize different forms of tRNA. ProRS catalyzes the aminoacylation of a set of four cognate tRNAs that differ in their anticodon sequence. However, recognition of the tRNA is only partly due to the anticodon sequence and also includes determinants in the structure of the tRNA. The ability of At3g62120, the organellar enzyme, to complement the *proS E. coli* mutant demonstrated that it is able to recognize *E. coli* tRNA^{Pro}. Indeed, this enzyme was able to utilize both *E. coli* and *S. cerevisiae* tRNA preparations (the *S. cerevisiae* source including both cytosolic and mitochondrial tRNAs). In contrast, At5g52520, the cytosolic enzyme, was unable to complement the *proS* mutant, despite the fact that it was fully functional in proline adenylylation and aminoacylation activities *in vitro*, suggesting that this enzyme was unable to utilize *E. coli* tRNA^{Pro}. Surprisingly, At5g52520 was able to utilize both *E. coli* and *S. cerevisiae* tRNA preparations when it was assayed *in vitro*, although less efficiently with the *E. coli* source, in seeming contradiction to the *in vivo* result.

In general, organellar tRNAs are thought to be structurally similar to eubacterial tRNAs (Reger et al., 1970; Imbault et al., 1981). Both *S. cerevisiae* and *A. thaliana* organellar tRNA^{Pro} sequences more closely resemble *E. coli* tRNA^{Pro} than they do cytosolic tRNA^{Pro} (Fig 20). All cytosolic tRNA^{Pro} sequences from eukaryotes display a conserved C73 and a G1˙C72 base pair, whereas all eubacterial tRNA^{Pro} sequences have a

conserved A73 and the C1˙G72 base pair (Stehlin et al., 1998; Cusack et al., 1998; An et al., 2008). The differences in tRNA structure are thought to result in a barrier to cross tRNA recognition. Therefore, the utilization of *E. coli* tRNA^{Pro} by the *A. thaliana* organellar enzyme but not the cytosolic enzyme was expected, but the utilization of *A. thaliana* cytosolic ProRS was unexpected. The result could be explained by postulating that at the high tRNA concentration used to measure *in vitro* activity At5g52520 is able to inefficiently utilize *E. coli* tRNA^{Pro}, but that the tRNA^{Pro} concentration in the *E. coli* cytoplasm is low to support activity by At5g52520. This question was not further explored due to the prohibitive cost of the radiolabelled proline substrate used to perform the aminoacylation assay.

An interesting observation that arose from the analysis of *A. thaliana* tRNA^{Pro} sequences was that organellar-like sequences are encoded in the nuclear genome. Indeed, the mitochondrial genomes of most eukaryotes are known to be deficient in certain tRNAs, including tRNA^{Pro} (Menand et al., 1998; Kumar et al., 1996), in contrast to the genomes of plastids, which encode a full complement of the necessary tRNAs (Kumar et al., 1996). This observation indicates that certain tRNAs encoded in the nucleus must be post-transcriptionally imported in to the mitochondria, and this was borne out in the analysis of *A. thaliana* tRNA^{Pro} shown in Figure 20.

4.4 ProRS encoding cDNAs from *P. pubescens* and *C. majalis*

By using a PCR-based strategy, the coding sequence of organellar ProRS genes from *P. pubescens* and *C. majalis* were successfully isolated. Alignment of the theoretical translation products from these sequences revealed close homology with the organellar

ProRSs from *A. thaliana* and *Z. mays*. Since the source of template DNA for the initial PCR reactions was the cDNA libraries, which were themselves prepared from wild-grown samples, the possibility exists that the isolated sequences were derived from a contaminating organism. However, this is unlikely to be the case since the deduced amino acid sequences showed the highest homology with the *Z. mays* ProRS, as would be expected given that *Z. mays*, *P. pubescens* and *C. majalis* are all from the monocotyledonous lineage.

Follow on work with the *P. pubescens* clone revealed that it was able to drive production of a recombinant protein of the expected molecular weight. This protein could be purified by Ni^+-affinity chromatography, but it did not show ProRS activity. There are several possible explanations for the lack of catalytic activity. First, the clone appears to not include a segment of the 5' end of the gene. This conclusion is based on the fact that the translation product is missing an approximately 50 codon stretch that is present in other plant ProRS sequences (Fig 33). There are a number of conserved residues in this region that might be critical for catalysis or biogenesis of plant ProRS. Another possibility is that a debilitating point mutation was introduced during the three rounds of PCR that was necessary to construct the complete sequence. Taq polymerase lacks a 3' to 5' exonuclease proofreading activity so its error rate is higher than polymerase with the proofreading activity. This might account for the lack of activity from the *P. pubescens* ProRS clone.

The frustrating inability to isolate a full-length coding sequence for either *P. pubescens* or *C. majalis* ProRS was likely due to a number of factors originating with the source material for RNA and DNA isolation. Both *P. pubescens* and *C. majalis* tissue

samples were from native plants harvested from the wild. Many plant tissues are notorious for production of polysaccharides that interfere with subsequent *in vitro* manipulations of the nucleic acids. Despite the fact that full-length cDNA for APS reductase and CGS were successfully isolated from the *P. pubescens* and *C. majalis* cDNA libraries, both libraries were of marginal quality with a low average insert size.

4.5 How might A2C-resistance have evolved in the Convallariaceae?

The discovery of A2C accumulation in *C. majalis* in 1959 (Fowden and Bryant, 1959), followed by the discovery that A2C toxicity results from its misincorporation into proteins in place of Pro (Fowden and Richmond 1963), prompted the question of how *C. majalis* resists the toxin that it accumulates. Shortly after the discovery of AARS, L. Fowden and co-workers showed that the ProRS from *C. majalis* has the capacity to discern proline from A2C, whereas the ProRS from bean (a species that does not accumulate A2C) cannot (Norris and Fowden, 1972). It was this early publication that formed the basis for my thesis. I set out to determine using molecular tools whether *C. majalis* ProRS had indeed adapted to its A2C-accumulating host by evolving the ability to distinguish proline from A2C. Although, I did not succeed in answering this question with regard to ProRS from *C. majalis* or *P. pubescence* my work did uncover the finding that a significant difference exists between plant cytosolic and organellar ProRS with respect to the ability to mis-incorporate A2C. This finding has implications for how A2C resistance may have evolved. Since A2C-resistant ProRS (the organellar form) appears to exist in plant species that do not accumulate A2C, raises the prospect that the initial observation by Norris et al. (1972), may have resulted from the specific conditions of

their experiment and that *C. majalis* ProRS does not differ from that in other plant species. In their experiment Norris et al. (1972) did not discern between cytosolic and organellar ProRS (indeed, at the time of their paper there was not yet an indication that different isoforms of ProRS may exist in subcellular compartments). Moreover, their experimental observation was made using partially pure protein preparations, the tissue and developmental stage of the source tissue was not carefully reported, and the tRNA substrate used for their experiments was purified from plant leaves and most surely consisted of both cytosolic and organellar tRNA^{Pro}. Therefore, the possibility exists that the A2C-resistant form of ProRS that they characterized was the organellar ProRS that happened to be enriched in their partially pure preparation. If this were the case ProRS from *C. majalis* may not be biochemically different the enzyme in plant species that do not accumulate A2C.

However, if Norris et al. observation is indeed correct, that *C. majalis* contains an A2C-resistant form of ProRS, the present findings suggest that A2C-resistant ProRS may already exist in the plant kingdom, and that *C. majalis* ProRS may have evolved through the innate ability of plant organellar ProRS to discern between proline and A2C. The differences between cytosolic and organellar plant ProRS with regard to amino acid and tRNA specificity may be a function of the primary structure of these enzymes. All ProRSs contain conserved class II AARS-specific motifs 1, 2, and 3, as well as an anticodon binding domain. Experimental analysis has revealed that motif 1 functions in dimer formation and motif 2/3 function in amino acid activation (Wong at al., 2002). The anticodon binding domain functions to recognize the anticodon of its cognate tRNA. Plant ProRS show conservation with in motif 2 and motif 3, but they also show

significant differences (Fig 33), (Stehlin et al., 1998). The key differences include R205, T206, K207, T212, and L213 in motif 2 and T316, F324, V325, G327, and I328 in motif 3 (residue numbering is of the organellar ProRS from *A. thaliana*), which are perfectly conserved in the organellar ProRS, but are not shared with the cytosolic ProRS. These amino acid residues are perfectly conserved among the organellar ProRSs. Finally, examination of the differences in the conserved ProRS domains in PpProRS, G215 is highlighted as a possible candidate for further examination with respect to A2C specificity of the amino acid binding site (Fig. 33).

In consideration of how *C. majalis* resists its own toxin, A2C, it important to point out that it is extremely unlikely that A2C-resistance could be the sole function of a specialized form of A2C. The extremely high tissue concentration of A2C suggests that this compound is also sequestered away from the protein synthesis machinery, perhaps in the vacuole or in the extracellular space. Indeed, in other examples of plants that accumulate toxic defense compounds, in almost all known cases the defense compound is sequestered in an inactive form that is released after tissue damage occurs. These examples include the accumulation of glucosinolates in the Brassicacea, which are activated by myrosinase after tissue damage by herbivores (Hopkins et al., 2009).

4.6 Conclusions and prospects for further study

Although the present work has not definitively answered the question posed at the outset of the project, the resulting data provide intriguing prospects for future research and may provide clues on the possible structure and function relationship of amino acid substrate recognition by ProRS in plants.

Appendix A

1. Introduction

1.1 5'-Adenylylsulfate reductase (APR)

Sulfate assimilation is an essential process for the biosynthesis of the amino acids Cys, and Met and other sulfur containing metabolites in plants and microorganisms. When sulfate is taken up into plant cells, adenosine 5'-phosphosulfate (APS) is formed by ATP sulfurylase and then reduced to sulfite from APS by APR (Setya et al., 1996; Bick et al., 2000). Sulfate assimilation is highly regulated and APR is a metabolic regulation point in sulfur assimilation and reduction in plant (Martin et al., 2005; Koprivova et al., 2008). Plant APR has a novel two domain structures. The amino terminal domain contains conserved Cys residues that function as FeS cluster center ligands, substrate, and in dithiol/disulfide reaction. The C-terminal domain functions as a thioredoxin or glutaredoxin.

1.2 Cystathionine γ-synthase (CGS)

Met is an essential dietary amino acid for animals (Chiba et al., 1999). Met functions as a precursor of S-adenosylmethionine (SAM), which is the primary methyl donor in many transmethylation reactions. It is also a precursor of the phytohormone ethylene in plants. Furthermore, Leete et al (1986) suggest that methionine is a precursor of A2C and that the most likely mechanism for such a reaction would be cyclization of S-SAM and methylthioladenosine displacement. CGS catalyzes the condensation of *O*-phosphohomoserine and Cys to form cystathionine as the first committed step of Met biosynthesis in higher plants. This reaction is a key regulatory step of the biosynthetic

pathway of Met. It was previously shown that the stability of the CGS mRNA is feedback-regulated in response to applied Met. The MTO (methionine over-accumulation) region is a stretch of 11-13 amino acid residues that is necessary and sufficient for the posttranscriptional regulation of CGS. The MTO region is conserved most higher plant forms of CGS (Chiba et al., 1999).

2. Materials and Method

2.1 Specific supplies

E. coli strains JM96 and LE392 were obtained from the *E. coli* Genetic Stock Center (New Haven, CT). *E. coli* strain JM96 (*thr-1, leuB6, fhuA2, lacY1, supE44, gal-6, λ⁻, trp-1, hisG1, cysH56, galP63, gltB31, rpsl9, malT1* (λ^R), *xyl-7, mtl-2, argH1, thi-1*), a cysteine auxotroph mutant strain, was used to screen APR by functional complementation. *E. coli* strain LE392 (*hsdR574 (r_K^-, M_K^+), supE44, supF58, lacy1, galK2, galT22, metB1, trpR55*), a methionine auxotroph mutant strain, was used to screen for CSG. Phosphoadenosine phosphosulfate [^{35}S] (1.95 Ci/mmol) was purchased from PerkinElmer Life Sciences.

2.2 Cloning of APRs and CGSs

2.2.1 Complementation

Functional complementation was used to screen for APR and CGS cDNAs from *C. majalis* and *P. pubescens* and was carried out as described by Setya et al. for cloning of *A. thaliana* cDNAs (1996). A plasmid derivative of the cDNA libraries was transformed into mutant strain, JM96, a Cys auxotroph carrying the *cysH* mutation (for PAPS reductase), or LE392, a Met auxotroph carrying the *metB* mutation. The JM96 strain can grow only if Cys is added to the growth medium and the LE392 strain can grow only if Met is added to the growth medium. However, the mutant strains would be able to grow prototrophically when expressing a cDNA encoding the functional enzyme needed to replace the defective, endogenous mutant enzyme. A plasmid based cDNA library was prepared from the phage libraries by mass-excision according to the manufacturer's

instruction (BD Bioscience, Inc). Briefly, λTriplEx2 clone was converted to a pTriplEx2 plasmid by *in vivo* excision. The phage libraries were used to infect *E. coli* BM25.8, a strain expressing CRE recombinase. The infection was plated onto LB with 100 mg/mL ampicillin and growing colonies were used for plasmid purification. Plasmid DNA was prepared using NucleoBond Plasmid Maxi kit (BD Bioscience, Inc) according to the manufacturer's protocol. Lysed cells were centrifuged to clear the lysate. The supernatant was added to a NucleoBond column and washed. Plasmid DNA was eluted from the NucleoBond column with elution buffer (manufacturer supplied) and the concentration of DNA was determined using a UV spectrophotometer. The plasmid-based cDNA library was transformed into the Cys requiring *E. coli* strain JM96 or the Met requiring *E. coli* strain LE392 by electroporation (Bio-Rad Gene Pulser, following the manufacturer's protocol) using a cuvette with 0.2 cm electrode gap and the following conditions, 25 μF, 200 W, and 2.5 kV. The transformants were recovered on rich medium containing 100 μg/mL ampicillin and complementing cDNAs were screened by replica plating of the transformants onto M9 minimal medium supplemented with 0.2% glucose, 100 μg/mL ampicillin, and 20 amino acids (each at a concentration of 25 μg /mL). After isolating putative complementing cDNAs, the ability to complement was retested by retransforming each plasmid into the same mutant. Those plasmids that tested positive in the re-test were analyzed further. Restriction analysis of the positive cDNAs revealed that three different types of APR cDNAs and one type of CGS cDNA was isolated from *P. pubescens*. One type of APR and one type of CGS cDNA were isolated from *C. majalis*. The cDNAs were named PpAPR1, PpAPR2, PpAPR3, and CmAPR. CGS cDNAs were named PpCGS and CmCGS.

2.2.2 Rapid amplification of cDNA ends (RACE)

To clone the full-length cDNA sequence from a partial PpCGS cDNA, 5' RNA ligase mediated-rapid amplification of cDNA ends (RLM-RACE) was performed using the RLM-RACE kit (Ambion, Inc, Foster City, CA) according to the manufacturer's protocol. Poly (A) RNA was used as template. Calf intestine alkaline phosphatase (CIP) was used to treat poly (A) selected RNA to remove free 5'-phosphates from degraded mRNA, tRNA and genomic DNA. After removal of the phosphatase by phenol:chloroform extraction, tobacco acid pyrophosphatase (TAP) was then used to remove the cap structure found on intact 5'end of full-length mRNA. The cap structure is not affected by CIP. 5' RACE adapter oligonucleotide was ligated to decapped mRNA using T4 RNA ligase. The ligated RNA was reverse transcribed by using the FirstChoice RLM-RACE kit (Ambion Inc.). The reaction mixture (20 μL) for reverse transcription contained 50 ng of poly (A) RNA, 0.5 μg of 5' RACE adapter (5'-GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUCAUGAAA-3'), 1 U of placental RNase inhibitor, 0.5 mM each dNTP, 1x RT buffer (manufacturer supplied), and 10U of molony murine leukemia virus (M-MLV) reverse transcriptase. The mixture was incubated at 42°C for one hour and primary, secondary, and third PCR was then performed. Three nested gene specific PCR primers were designed to perform nested PCR for 5' RLM-RACE PCR. Primes were Tail PCR 3'-1 (5'-ATGAACCAAAAGGCTCGCATC-3'), Tail PCR 3'-2 (5'-AGCTAAACACCGCATCCAACC-3'), and Tail PCR 3'-3 (5'-CAACCGGCGAGATCCGACCCT-3'). The 5'-RLM-RACE PCR reaction (50 μL) contained 1 μL of RT reaction, 0.5 each dNTP, 10 μM of 5' RACE gene specific outer

(5'-GCTGATGGCGATGAATGAACACTG-3') /inner (5'-CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG-3') primer, 10 μM of 5' RACE outer/inner primer, 1x PCR buffer (manufacturer supplied), and 1 U of thermostable DNA polymerase. The PCR conditions were 1 cycle for 5 min at 94ºC followed by 35 cycles of 30 sec at 94ºC, 30 sec at 60ºC, 2 min at 72ºC and a final step of 7 min at 72ºC. The PCR product was run on a 2% ($^w/_v$) high resolution agarose gel and shown several bands. Each DNA band was eluted from the gel and cloned into pGEM T-Easy Vector (Promega, Madison, WI). Recombinant plasmids were sequenced. Sequence analysis revealed that one of the cDNA sequences was homologous to other plant CGS genes.

2.2.3 Isolation of a genomic clone of PpCGS

Based on the cDNA sequences, a genome walking procedure was performed by PCR. Six primers were used to anneal to the coding region of the putative gene and three different fragments were produced using three sets of primers. The sets of primer for each fragment were PpCGS5' (5'- CC GGA TCC ATG GCC GTC TCA TCC-3') and PpCGS L1 (5'-AGCACCGCGCAGCTGGCGTAC-3'), PpCGS R2 (5'-TTCAGCCTTGGTTCCTGCTGGT-3') and PpCGS L2 (5'-ATCGTTGTGTCCAGCAATGAA-3), PpCGS R3 (5'-GTTATTGCAGGTTGTGTC-3') and PpCGS L4 (5'-AAACTGAAGCTAACATCAATATT-3') PpCGS3'(Xho I); 5'- CC CTC GAG TTA TAT AGC TTC CAG -3'. PCR was performed in a 50 μL reaction containing 0.5 mg of genomic DNA, 0.5 mM each dNTP, 3.8 pM 5' and 3' primers, 1x PCR buffer (manufacturer supplied), and 1 U of DNA polymerase

(Invitrogen Inc.). The PCR conditions were 1 cycle for 10 min at 94ºC followed by 30 cycles of 30 sec at 94ºC, 30 sec at 60ºC, and 4 min at 72ºC. The PCR products were cloned into pGEM T-Easy Vector and both strands of resulting recombinant plasmids were sequenced.

2.3 DNA sequencing and sequence analysis

Each cloned cDNA from *C. majalis* and *P. pubescens* was completely sequenced on both strands with plasmid vector primers **a**nd the internal primers designed to extend deletion subclones on the parent vectors, using an Applied Biosystems Model 373A Automated Sequencer (Applied Biosystmes, Foster City, CA). Amino acid sequences were derived from DNA sequences using DNASTAR program. Sequence analyses were carried out using protein-protein BLAST (blastp).

The phylogenetic relationship of APRs and CGSs were analyzed. The sequence alignment and phylogenetic analysis was performed using Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 (Tamura et al., 2007). The sequences for APRs were analyzed using the NCBI web-based Blast server and the sequence alignments were performed using CLUSTAL W. Phylogeny construction of combined sequences was done using Bootstrapped Neighbor-Joining method and evolutionary distances were calculated using Mega4. The other amino acid sequences included in the phylogenetic analysis were obtained from National Center for Biotechnology Information (NCBI).

2.3.1 APRs

The gene bank numbers of various APRs are as follows; *Arabidopsis thaliana*-APR1 (NP_192370), *Arabidopsis thaliana*-APR2 (NP_176409), *Arabidopsis thaliana*-APR3 (NP_193930), *Archaeoglobus fulgidus* (CAA45030), *Bathymodiolus thermophilus* (ABV80103), *Brassica juncea* (CAA04611), *Catharanthus roseus* (AAB05871), *Chlorobium chlorochromatii* (YP_379883), *Enteromorpha intestinalis* (AAC26855), *Escherichia coli* (BAE76827), *Myxococcus xanthus* (YP_630560), *Nitrosomonas eutropha* (YP_747407), *Oryza sativa* (BAC83952), *Ostreococcus tauri* (CAL56791), *Pyrobaculum aerophilum* (NP_560100), *Saccharomyces cerevisiae* (J05591), *Solanum tuberosum* (CAD44841), *Wolinella succinogenes* (CAE10112), *Zea may*-APSR1 (NP_001105764), *Zea may*-APSR2 (NP_001105805).

2.3.2 CGSs

The gene bank numbers of various CGSs are as follows; *Agrobacterium tumefaciens* (NP_356816), *Arabidopsis thaliana* (AAB41235), *Bacillus amyloliquefaciens* (YP_001420788), *Chlamydomonas reinhardtii* (XP_001695742), *Escherichia coli* (BAE77371), *Glycine max* (AAD34548), *Leucaena leucocephala* (BAF80449), *Medicago sativa* (ABI34092), *Mesembryanthemum crystallinum* (AAC19395), *Oryza sativa* (AAL82522), *Ostreococcus lucimarinus* (ABO98015), *Pseudomonas fluorescens* (YP_260914), *Schizosaccharomyces pombe* (CAA20484), *Solanum tuberosum* (AAD31520), *Xanthomonas campestris* (YP_242341), *Zea mays* (AAB61347).

2. 4 Extraction of APR

DNA was isolated and purified from *E. coli* JM96 expressing the APR clones using a plasmid purification kit (Clontech, Palo Alto, CA) and transformed into DH5α competent cell by heat shock at 42ºC for 45 sec. *E. coli* DH5α cell expressing APR were grown at 37ºC to an OD600 of 0.5 in LB broth with 100 μg/mL ampicillin and induced with 1mM isopropyl-1-thiol-β-D-glactopyranoside (IPTG) at 37ºC for 3 hours. The cells were harvested by centrifugation at 4ºC for 10 min at 7,000X. The harvested cell pellet was frozen at – 70ºC and then was resuspended in protein extraction buffer containing 50 mM Tris-HCl (pH8.0). Cell extract was prepared by sonication and centrifugation at 10,000 X for 15 min. The supernatant was used for APR activity assay. The concentration of protein was measured by the method of Bradford using a Biorad kit (BioRad, Hercules, CA). Crude cell extracts were stored at -70ºC until used.

2. 5 APS reductase activity measurement

APS reductase activity was measured as described previously (Setya et al., 1996). The assay contained 50 mM Tris-HCl (pH 8.5), 1 mM EDTA, 5 mM dithiothreitol (DTT), 500 mM Na_2SO_3, 25 μM [^{35}S]APS, and the experimental enzyme. The [^{35}S]SO_3^{2-} from the radiolabeled sulfonucleotide was conversed to volatile $^{35}SO_2$ by addition of acid. 3M H_2SO_4 was added to the uncapped reaction tubes and the tubes placed into a scintillation vial containing 1ml trioctylamine. Vials were capped tightly and incubated overnight at room temperature, resulting in volatilized SO_2 to be absorbed by the trioctylamine (Trisubstituted-amine by long straight chains). The reaction tubes were removed from vials and then 3 ml scintillation fluid added to vials (Ready-Safe; Beckman). Radioactivity was measured by scintillation counting.

3. Results

3.1 APR

3.1.1 Cloning of APR genes by complementation

Plasmid based cDNA libraries were prepared from the phage library of both *C. majalis* and *P. pubescens* by mass-excision using CRE/LOX recombination. The plasmid-based cDNA library was transformed into JM96, a cysteine auxotroph carrying the *cysH* mutation. Transformants were recovered on rich medium containing ampicillin, and complementing cDNAs were screened by replica plating of the transformants onto minimal medium lacking the required amino acid, cysteine. Plasmid DNA was recovered from the growing colonies. The ability to complement was retested by transforming each complementing plasmid into the same mutant. Those plasmids that tested positively in the re-test were analyzed further. About 1,000,000 transformants were screened and 15 positives were identified, giving a frequency of one positive clone per 67,000 colonies screened. Figure A1 shows the functional complementation of APRs in *E. coli* mutant strain, JM96 on M-9 minimal medium lacking Cys supplementation. Restriction analysis of the positive cDNAs revealed that three different types of APRs cDNA were isolated from *P. pubescens*. The longest cDNA was about 1400 bp. Only one type of APR cDNA was isolated from *C. majalis*. Each of these cDNAs was sequenced and used to search for homologous sequences in the nucleotide database. These results confirmed the identity of all the cloned cDNAs. The *P. pubescens* APR cDNAs were completely sequenced and were then deposited in GenBank under accessions AAZ75685, AAZ75684, and AAZ75683.

3.1.2 Sequence analysis of APR cDNAs

As noted above, three classes of cDNA clone from *P. pubescens* and one cDNA clone from *C. majalis* were isolated by complementation of *E. coli* strain JM96. Table A1 shows the major characteristics of the APR clones and their similarity to each other. A BLASTP search of GenBank database with the translational products of the PpAPR and CmAPR clones revealed high homology with APRs from other organisms. All four APR clones had the highest homology with *Solanum tuberosum* (81.8% identical with CmAPR, 80.9% with PpAPR2, 80.9% with PpAPR4, and 82.7% with PpAPR5) and also had homology with most monocot APRs between 80 and 66%. The highest homology *with A. thaliana* APRs is with APR3 for all clones (69.7% with CmAPR, 71.5% with PpAPR2, 78.4% with PpAPR4, and 69.9% with PpAPR5). A comparison of amino acid sequence with APRs from *C. majalis* and *P. pubescens* is shown in Fig. A2. The four APR enzymes have all of the hallmarks of plant APRs, having nearly perfect identity with regions shown to be essential for the catalytic function of these enzymes: the unique sequence of APR (indicated by dots and asterisks in Fig. A2) (Bick and Leustek, 1998), a region shared in all of the assimilatory (P)APRs (indicated by open dots in Fig. A2) (Berendt et al., 1995), and homology with thioredoxin, glutaredoxin, and protein disulfide isomerase including the active site sequence (indicated by square in Fig. A2). Three of the APRs, CmAPR, PpAPR2, and PpAPR5 have full-length open reading frames (ORF) including translational initiation and termination codons. All four APRs show high homology each other. CmAPR was highly homologous with PpAPR4 and PpAPR5 (91% of identity). The identity between APR2 and APR4 is 99.5%, although the N-terminus of APR4 is 174 bp shorter than APR2 indicating the possibility that APR4 is a partial cDNA

of APR2. A phylogenetic tree was constructed to illustrate evolutionary lineages between the APR enzymes (Fig. A3). All PpAPRs and CmAPR were clustered with other plant APR enzymes.

3.1.3 Enzyme activity of PpAPRs and CmAPR

Additional confirmation of the function of the APR cDNAs came from measuring APS reductase enzyme activity in extracts of the *E. coli* strains expressing the APR cDNAs. The results are presented in Table A2. JM96 carrying the empty expression plasmid showed little or no sulfite formation. JM96 carrying PpAPRs and CmAPR had comparatively high APR reductase activities, ranging from approximately 100 to 400 x 10^{-3} units mg^{-1}.

3.2 CGS

3.2.1 Cloning of CGS genes by complementation and RACE

A plasmid based cDNA library was transformed into mutant strain LE392, a methionine auxotroph carrying the *metB* mutation. Transformants were recovered on rich medium containing ampicillin, and complementing cDNAs were screened by replica plating of the transformants onto minimal medium lacking the required amino acid, methionine. The frequency with which confirmed complementing cDNAs were isolated was 4 in 1,000,000 colonies screened. The plasmids carried by the positive colonies were isolated and the ability to complement was retested by transforming each plasmid into the same mutant. All screened CGSs were retested by transforming each plasmid into the LE392 strain and grown well on minimal medium lacking methionine, compared with

untransformed strain as negative control. Figure A4 shows the functional complementation of CGSs in *E. coli* mutant strain, LE392. One type CGS cDNA was isolated from each *C. majalis* and *P. pubescens* library and named PpCGS and CmCGS. Table A3 shows the major characteristics of CGS clones and their similarity to each other. Although a BLASTP search of GenBank database with translation products of PpCGS and CmCGS revealed significant homology with other different CGSs, both clones of PpCGS and CmCGS were predicted to be partial cDNAs lacking several hundred bases from the N-terminus.

A full length-DNA sequence of the PpCGS gene was isolated by using 5' RLM-RACE PCR method. The PCR fragment by 5' RACE was obtained about 500 bp longer and appeared to be full length. This fragment contains both initiation codon (ATG) and overlapping region with 5' end sequences of partial PpCGS cDNA screened from library. The overlap with 5' end sequence of the cDNA fragment indicates that 5' RACE PCR fragment is derived from the same mRNA as the PpCGS cDNA fragment. Complementary DNA fragments of PpCGS gene was composed by using Seqman program of DNASTAR and this result predicted the sequence of 1674 bp in length containing complete open reading frame (Fig. A5).

3.2.2 Cloning and analysis of genomic PpCGS

Based on the cDNA sequence information the genomic sequence and organization of the PpCGS gene was elucidated using PCR. The length of total genomic sequence was determined to be 5149 bp. Figure A5 shows the nucleotide sequences and the deduced amino acid sequence of PpCGS gene. PpCGS consists of 11 exons (nucleotide positions 1

- 532, 640 - 743, 829 - 902, 998 - 1166, 2551 - 2610, 2837 - 3040, 3162 - 3248, 3634 - 3740, 4476 - 4571, 4770 - 4888, and 5029 - 5149) and 10 introns (nucloetide positions 533 - 639, 744 - 828, 903 - 997, 1167 - 2550, 2609 - 2836, 3041 - 3161, 3249 - 3633, 3741 - 4475, 4572 - 4469, and 4889 - 5028). Most introns had conserved (Shapiro and Senapathy, 1987) and contained high content of A + T. The positions of the splice junction were confirmed by comparison with the cDNA sequence.

3.2.3 Sequence analysis PpCGS and CmCGS

The PpCGS sequence was translated to a protein of 558 amino acid residues with a calculated molecular mass of 59.3 kDa and isoelectric point of 6.6 (Table A3). The CmCGS gene contains 458 codons with a molecular mass of 48.9 kDa (Table A3). The amino acid sequence of PpCGS and CmCGS was used for search of GenBank database using Blastp program. Both of PpCGS and CmCGS share most similarity with CGS from *Zea may* (68% identity with PpCGS and 72% identity with CmCGS). Identity with the Arabidopsis CGS was 60% with PpCGS and 67% with CmCGS. PpCGS showed 85% of the homology with CmCGS. Phylogenetic analysis of CGS enzymes shows that PpCGS and CmCGS were derived from same evolutionary lineages with as other plant species (Fig. A6). PpCGS and CmCGS have a high overall homology with CGSs from *Z. mays* and *O. sativa*.

CGS is a key regulatory point in the pathway of Met biosynthesis in plants. The molecular mechanisms for regulation of methionine biosynthesis were studied using the *A. thaliana* mutant, *mto1*, that overaccumulates soluble Met (Chiba et al, 1999). PpCGS

has the highly conserved MTO region that is found in other plant CGS (Fig. A7). CmCGS was predicted to lack MTO region.

4. Discussion

To examine whether it was be possible to clone full length cDNAs from the *P. pubescens* and *C. majalis* cDNA libraries a well developed complementation cDNA cloning strategy was used to clone APR and CGS. In this study, four cDNAs encoding APR and two cDNA encoding CGSs were isolated from *C. majalis* and *P. pubescens* cDNA libraries. Comparison of the deduced amino acid sequences indicated that each cDNA show striking homology in conserved regions with other examples of APR and CGS (Figs. A2 and A7). The enzyme activity of PpAPRs and CmAPR explains why these cDNAs are able to complement *E. coli* mutant (Table A2). PpAPR4 shows 99% identity with PpAPR2 and includes all essential regions for catalytic function. The PpAPR4 ORF is 58 amino acids shorter than PpAPR2. The cloning results demonstrate that it is possible to clone full length cDNAs from the *P. pubescens* and *C. majalis* cDNA libraries.

Biosynthesis and metabolism of Cys and Met is connected each other. The schematic diagram of these amino acid biosynthesis and metabolism is shown in figure A8. APR produces sulfite from adenylylsulfate (APS) during biosynthesis of Cys and functions a key enzyme for the division of the reduction pathway. CGS catalyzes the condensation of O-phosphohomoserine and Cys to form cystathionine and present a key regulatory step of the biosynthesis pathway of Met. The stability of the CGS mRNA is feedback-regulated in response to applied Met. MTO region of CGS enzyme is necessary and sufficient for its posttranscriptional regulation. Ideally, If Met is a precursor of A2C, APR or CGS enzymes are able to participate in the regulation of A2C biosynthesis.

However, this idea is not certain because the mechanism of regulation among these amino acids is not determined yet.

Table 1. Construction of expression plasmids

Restriction sites are indicated in bold letters, initiation and termination codons are underlined.

Enzyme	Intended purpose	Primers	Restriction enzymes Cloning vector
At5g52520	Complementation	5'-CA**GGTACC**GAAGGAGATATAACC<u>ATG</u>TCAGACCGAGCCGTTACG-3' 5'-CA**GTCGAC**<u>TTA</u>ATATGACTTGGCAAAGATT-3'	*KpnI SalI* pBAD33
At3g62120	Complementation	5'-CA**GGTACC**GAAGGAGATATAACC<u>ATG</u>GCTGATCCCAGTGAACAA-3' 5'-CA**GTCGAC**<u>TTA</u>GTAACTCCTGCCCCAATA-3'	*KpnI SalI* pBAD33
At5g52520	Protein expression	5'-CC**GGTACC**<u>ATG</u>TCAGACCGAGCCGTTACG-3' 5'-CA**GTCGAC**<u>TTA</u>ATATGACTTGGCAAAGATT-3'	*KpnI SalI* pET30a
At3g62120	Protein expression	5'-CC**GGTACC**<u>ATG</u>GCTGATCCCAGTGAACAA-3' 5'-CA**GTCGAC**<u>TTA</u>GTAACTCCTGCCCCAATA-3'	*KpnI SalI* pET30a
ZmProRS-Org	Protein expression	5'-GATATC**GGATCC**<u>ATG</u>GACAGGGAGGGGCAGGTC-3' 5'-GAATTC**CTCGAG**<u>CTA</u>GTATGACTTTGCGAAAAT-3'	*BamHI XhoI* pET30a
ZmProRS-Cy	Protein expression	5'-GAGCTC**GGATCC**<u>ATG</u>TCGACCAACAAGGGCAGC-3' 5'-CTCGAG**AAGCTT**<u>TCA</u>GTAGCTGCGGCCCCAGAA-3'	*BamHI HindIII* pET30a

| EcCysS | Complementation | 5'-CAGAAGGAGATATAACCATGCTAAA AATCTTCAATACT-3' 5-CCTCTAGATTACGACGCCAGGTGGTC CCT-3' | *EcoRI XbaI* pBAD33 |

Table 2. The primers for nested PCR to clone ProRSs from *P. pubescens* and *C. majalis*

Clone name	Direction	Primer #	Primer sequence
PpProRS	5'-end	1	5'-TGTATTGGTGCTAATCTGGGA-3'
		2	5'-AACTGAGTTCCAAAGGCACGA-3'
		3	5'-CTGGCTTCCTTCTCTATAAAT-3'
	3'-end	1	5'-AGCTGCAATACCTGTTATTGT-3'
		2	5'-GATGAAAATGGAGAAAGA-3'
		3	5'-GTGGTCAAGTTGGATGACTCA-3'
CmProRS	5'-end	1	5'-CGATATCCTGCGAATGGA-3'
		2	5'-GATCCGAAGAAGTCTTCT-3'
		3	5'-TCCTGCCTACAATAACAGGTA-3'
	3'-end	1	5'-AGAGCTATCGTGAACTTCCTC-3'
		2	5'- TAAGGACTCTCGAGTTTCTAT-3'
		3	5'- TTCTTAATGCTGCTACATCTG -3'

Table 3. Kinetic parameters of AtProRSs

	At5g52520-Org		At3g62120-Cy	
Amino acid	Proline	A2C	Proline	A2C
V_{max} (µmol min^{-1} mg protein^{-1})	13.07 (±0.58)	22.81 (±4.66)	69.53 (±3.47)	67.56 (±9.36)
K_m (mM)	0.10 (±0.02)	13.55 (±5.39)	0.01 (±0.01)	0.13 (±0.10)
K_{cat} / K_m (M^{-1}·min^{-1})	1.30×10^5	1.68×10^3	9.44×10^5	5.02×10^5

Table 4. Kinetic parameters of ZmProRSs

Amino acid	ZmProRS-Org		ZmProRS-Cy	
	Proline	A2C	Proline	A2C
V_{max} (μmol min^{-1} mg protein^{-1})	32.27 (±1.53)	22.24 (±1.03)	5.35 (±0.20)	6.27 (±0.23)
K_m (mM)	0.20 (±0.05)	3.14 (±0.36)	0.07 (±0.02)	0.33 (±0.05)
K_{cat} / K_m (M^{-1}·min^{-1})	8.07×10^5	2.25×10^3	1.09×10^6	5.76×10^4

Table 5. HPLC analysis of A2C of leave tissue from *C. majalis* and *P. pubescens*

Total soluble amino acids were extracted from pools of three independent samples and A2C measured by HPLC analysis. The variation from analysis of 3 independent measurements is given in standard deviation in parentheses.

	FW/DW	nmole mg^{-1} fresh weight	% of dry weight
Convallaria majalis	3.9	16.4 (±4.7)	0.66
Polygonatum pubescens	4.9	86.1 (±15.0)	1.85

Table 6. HPLC analysis of amino acids of various tissues from *C. majalis*

The content of A2C, Pro, and Met was determined in various tissues. The level of amino acid was calculated as the peak areas from the chromatograms divided by the fresh weight of the plant sample. The unit of amino acid content is pmoles mg^{-1}.

	A2C	Pro	Met
Root	3,379	112	126
Stem	3,728	146	18
Shoot	9,395	718	70
Young leaf	14,345	726	23
Mature leaf	11,266	322	256
Young flower	13,269	306	39
Mature flower	13,657	364	38

Table 7. HPLC analysis of amino acid of various tissues from *P. pubescens*

The content of A2C, Pro, and Met was determined in various tissues. The level of amino acid was calculated as the peak areas from the chromatograms divided by the fresh weight of the plant sample. The unit of amino acid content is pmoles mg^{-1}.

	A2C	Pro	Met
Tuber	84,940	593	155
Root	12,030	322	237
Stem	47,000	956	362
Shoot	38,769	1739	214
Young leaf	38,922	2471	592
Mature leaf	26,230	635	156
Young flower	21,620	2236	752
Mature flower	9,156	1321	466

Table A1. Characteristics of APR clones

Characteristic	CmAPR	PpAPR2	PpAPR4	PpAPR5
Size (bp)	1404	1272	1098	1404
codons	468	424	366	468
Mass (kDa)	51.8	67.4	53.4	83.2
pI	6.9	9.4	8.2	9.2
Amino acid identity				
CmAPR	100			
PpAPR2	81.6	100		
PpAPR4	91.3	99.5	100	
PpAPR5	91.2	84.9	90.2	100

Table A2. APS reductase activity of PpAPRs and CmAPR

PaAPR is the *P. aeruginosa CysH* gene expressed from pB-PaAPR. APS reductase activity of all cloned APRs was measured with 5 mM DTT as the electron donor and 0.1 μg of cell lysate protein. The reaction mixture of PaAPR contained 0.25 ng of pure protein. The reaction mixtures were incubated at 30 ºC for 20 min.

Proteins	Activity ($\mu mol^{.}min^{-1.}mg^{-1}$)
-APR	0
CmAPR	0.11
PpAPR2	0.39
PpAPR4	0.18
PpAPR5	0.19
PaAPR	0.18

Table A3. Characteristics of CGS clones

Characteristic	CmCGS	PpCGS
Size (bp)	1404	1674
Codons	468	557
Mass (kDa)	51.8	59.6
PI		6.6
Amino acid identity		
CmCGS	100	
PpCGS	86.1	100

Figure 1. The structures of Pro and A2C
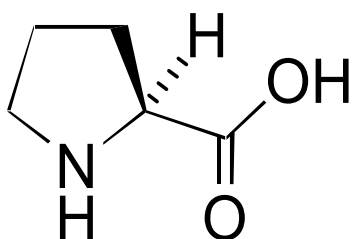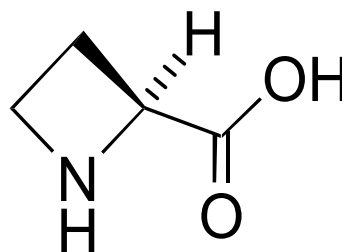
L-proline   Azetidine-2-carboxylic acid

Figure 2. Biosynthesis of nicotianamine from L-Met

Nicotianamine synthase (NAS) catalyzes the trimerization of A2C via S-adenosylmethionine (SAM) from L-Met to synthesize one molecule of nicotianamine (NA). This figure was modified from Ma et al. (1995) and Higuchi et al. (1999).

Nicotianamine (NA)

Figure 3. Aminoacyl-tRNA synthesis and the role of AARSs in protein translation

AARSs select the corresponding amino acid from the cellular amino acid pools and transfer it to the proper tRNA coupled with ATP hydrolysis. Aminoacyl-tRNA then interacts with a translation elongation factor and is delivered to the ribosomal A site. The anticodon of tRNA interacts with the corresponding codon of mRNA in the ribosomal A site. The diagram was redrawn from Ibba et al. (2001).
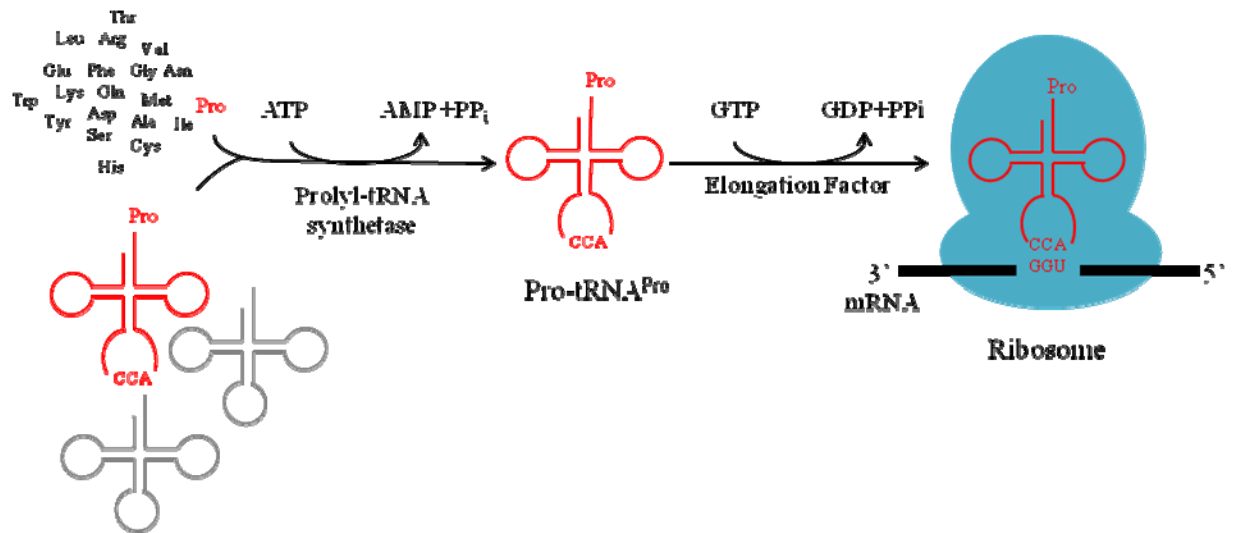
Figure 4. Two step reaction by AARS in aminoacyl-tRNA synthesis

This diagram was redrawn from Szmanski and Barciszewski (1999).

Step1.

Step2.

$$AA \ + \ ATP \ \underset{\longleftarrow}{\overset{AARS}{\longrightarrow}} \ AA\text{-}AMP \ + \ PP$$

$$AA\text{-}AMP \ + \ tRNA \longrightarrow AA\text{-}tRNA \ + \ AMP$$
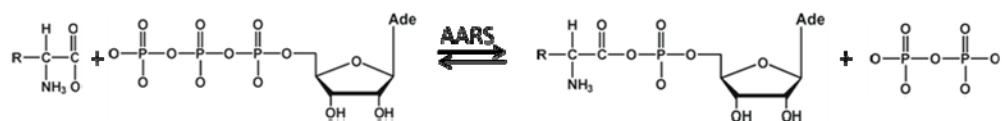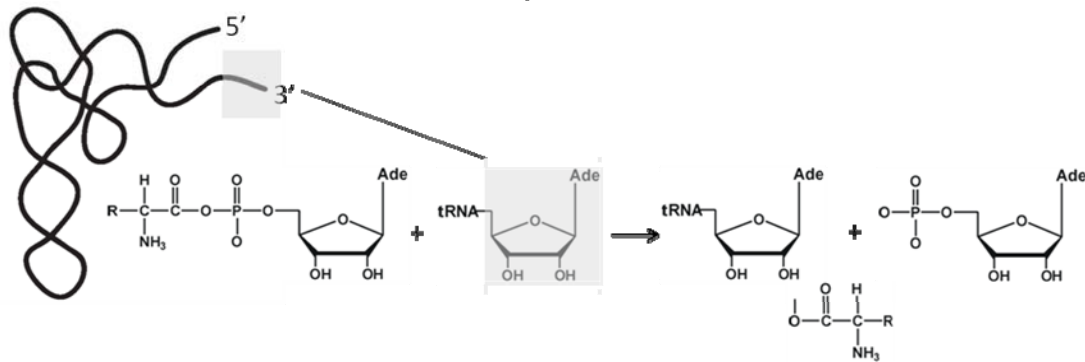
Figure 5. Schematic illustration of the function of AARS in several other cellular processes

In addition to providing translational fidelity during aminoacyl-tRNA synthesis, AARSs function as important factors in several other cellular processes such as tRNA processing, RNA splicing, RNA trafficking, transcriptional and translational regulation in response to stress, apoptosis, and embryo development.

tRNA aminoacylation

RNA-dependent amino
acid edting

tRNA maturation
proofreading

Cytokine-like
activity

Mitochondrial
RNA spilcing

AARS

Aminoacylation of
tRNA-like structures

Chlorophyll
biosynthesis

Translational
regulation

Histidine
biosynthesis

Transcriptional
regulation
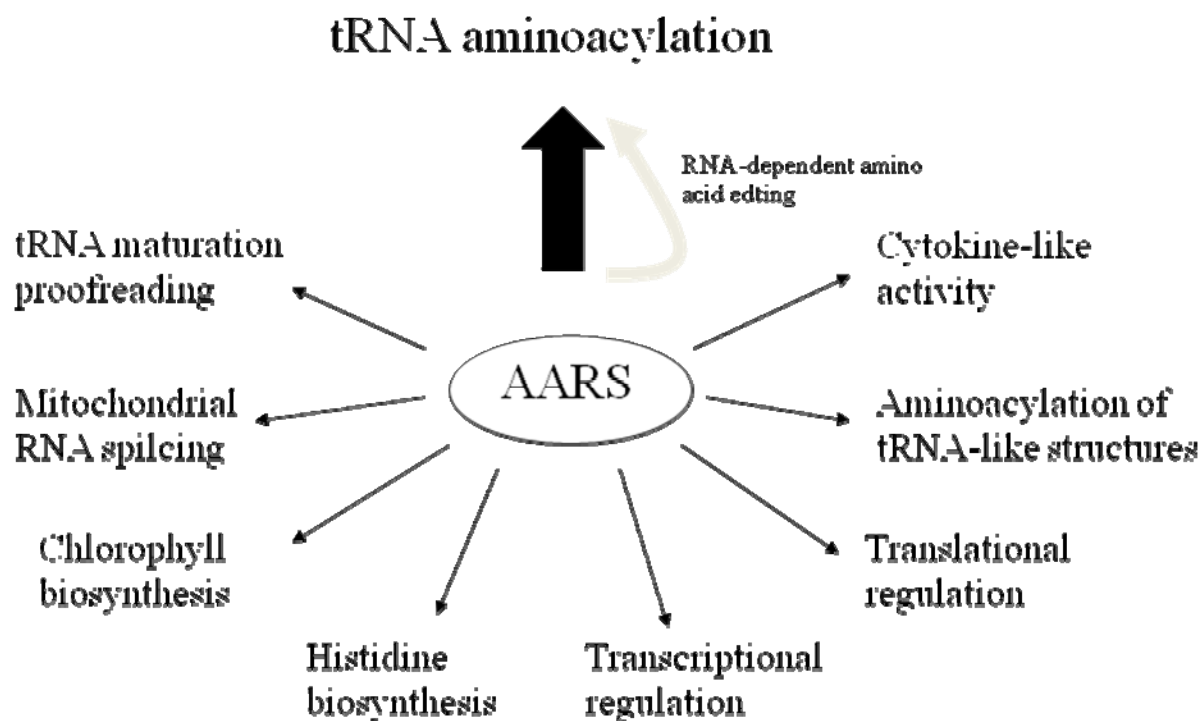
Figure 6. Schematic representation of the structural differences between prokaryote-like group (A) and eukaryote/archaea-like group (B) ProRS proteins

Three motifs class II-defining represented in each group of ProRS. Ab indicates the anticodon binding domain.

A. Prokaryote- like ProRS

B. Eukaryote/archaea - like ProRS

Figure 7. The scheme of the ATP-PP$_i$ exchange assay

ATP-PP$_i$ exchange depends on the rate of synthesis of the enzyme bound amino acid adenylylate. The reaction is freely reversible and results in the incorporation of [^{32}P] pyrophosphate into ATP, which is measured by scintillation counting.

ProRS +ATP + Pro $\rightleftarrows$ ProRS-Prolyl-AMP + PP$_i$

$\downarrow\uparrow$ +^{32}PP$_i$

ProRS + Pro + [^{32}P] ATP
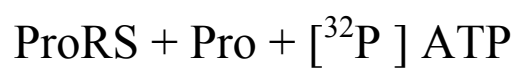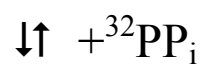
Figure 8. The scheme of the aminoacylation assay

Aminoacylation activity is detected by the incorporation of radiolabeled Pro into tRNA.

The incorporation of $[^3H]$Pro into tRNA is measured by scintillation counting.

$$\text{ProRS} + [^3\text{H}]\text{-Pro} + \text{ATP} \leftrightarrows \text{ProRS-}[^3\text{H}]\text{-Pro-AMP} + \text{PP}_i$$

$$\downarrow \ + \text{tRNA}$$

$$\text{ProRS} + [^3\text{H}]\text{-Pro-tRNA} + \text{AMP}$$

$$\text{ProRS} + [^3\text{H}]\text{-Pro} + \text{ATP} \leftrightarrows \text{ProRS-}[^3\text{H}]\text{-Pro-AMP} + \text{PP}_i$$
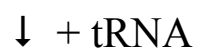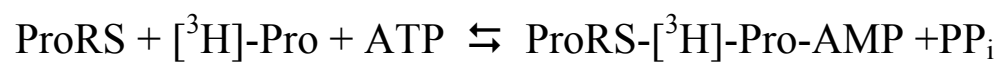
Figure 9. The strategy to clone ProRS from *C. majalis* and *P. pubescens* using nested PCR

Plasmid primers were coupled with ProRS-specific primers described in section 2.3.5.2. *C. majalis* and *P. pubescens* cDNA libraries were used as template for the first-round PCR, then the first-round product was used as template for second-round PCR and then the second-round product was used as template for third-round PCR.

pTripleEx2
including ProRS
from libraries
3.6 kb

ATG   Cloned ProRS sequence   Stop

5' λTripleEx2
Sequencing
Primer

5'-end nested
Primer1

3'-end nested
Primer1

3' λTripleEx2
Sequencing
Primer

↓ 1st Nested PCR

5' λTripleEx2
Sequencing
Primer

5'-end nested
Primer2

3'-end nested
Primer2

3' λTripleEx2
Sequencing
Primer

1st PCR Product

↓ 2nd Nested PCR

5' λTripleEx2
Sequencing
Primer

5'-end nested
Primer3

3-end nested
Primer3

3' λTripleEx2
Sequencing
Primer

2nd PCR Product

↓ 3rd Nested PCR

3rd PCR Product

Figure 10. Effect of A2C on the growth of *A. thaliana*

(A) *A. thaliana* seeds were germinated and grown for 9 days on MS-agar medium in the presence of increasing concentrations of A2C (0-10 μM). (B) and (C) Pro functions as an antidote to growth inhibition by A2C. *A. thaliana* seeds were germinated and grown for 9 days on MS-agar medium at increasing concentrations of A2C (0-10 μM) with or without supplementation with 100 μM Pro (C). The data are presented as the mean ±SD of three independent plants.

(A)

A2C
(μM)

| 0 | 0.5 | 1 | 2 | 3 | 5 | 10 |

(B)

| None | Pro | A2C | A2C+Pro |
| | 100 μM | 5 μM | 5 +100 μM |

(C)

Root length (mm)

Concentration of A2C (μM)

● - Pro
○ + 0.1 mM Pro
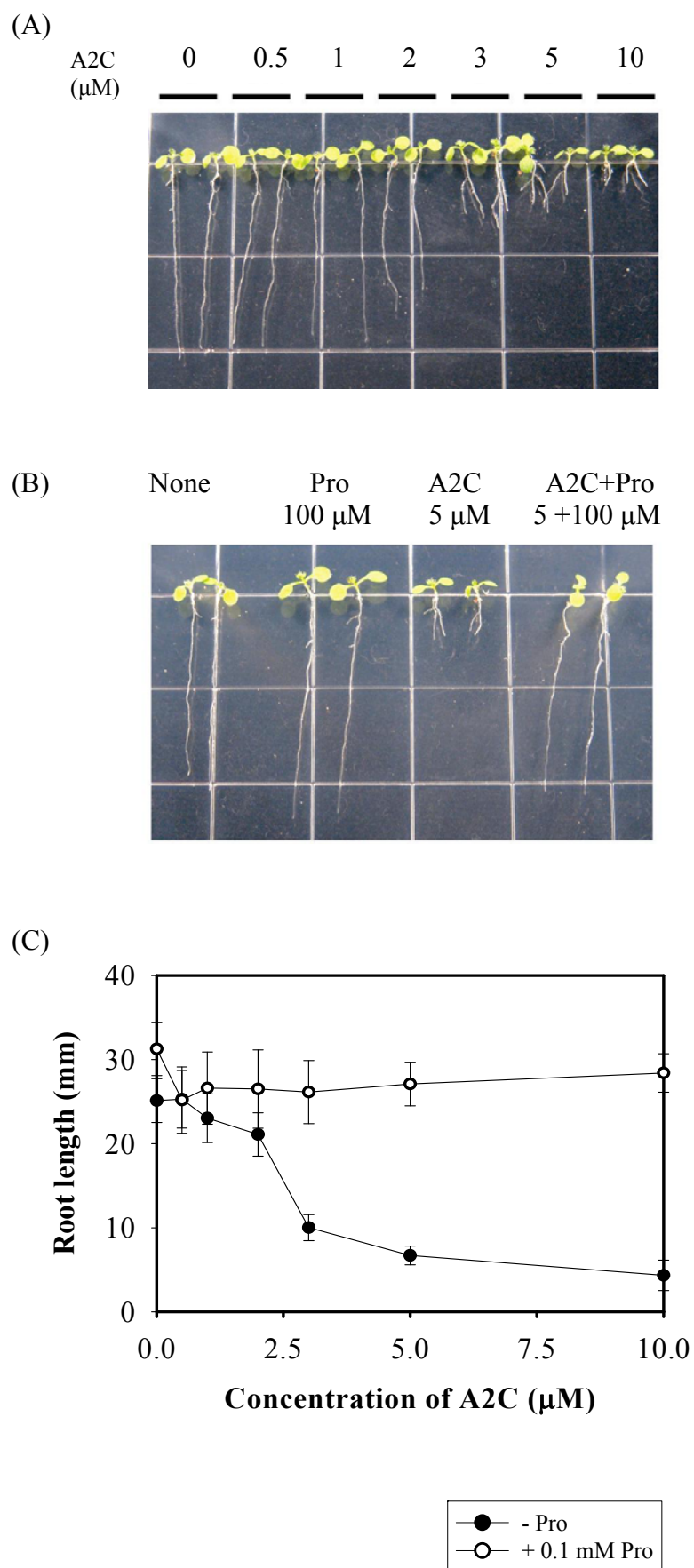
Figure 11. The alignment of AtProRS proteins

The BLAST alignment of At3g62120 (At3g) and At5g52520 (At5g) is shown. In addition, the intron positions in the coding sequence are indicated with arrows. The intron numbers are shown above (At3g) and below (At5g) their marked positions.

```
                         I2                           I3
At3g:  54  KDENFGEWYSEV-CKQDMIEYYDISGCYILRPWSMAIWEIMQIFFDAEIKKMKVKNCYFP  112
           + ++F  WY +V     ++ +Y + G   ++RP+ AIWE Q + + + K+    N YFP
At5g:  69  RSQDFNAWYLDVIASAELADYGPVRGTMVIRPYGYAIWEAIQDYLNVKFKETGHSNMYFP  128
                                                  I1

                                    I4
At3g: 113  LFVSPGVLEKEKDHIEGFAPEVAWVTKSGKSDLEVPIAIRPTSETVMYPYYSKWIRGHRD  172
           F+    +EKE  H+EGF+PE+A VT  G  +LE + +RPTSET++   +++WI +RD
At5g: 129  QFIPYSFIEKEASHVEGFSPELALVTVGGGKELEEKLVVRPTSETIVNHMFTQWIHSYRD  188
           I2                                       I3

                              I5                      I6
At3g: 173  LPLKLNQWCNVVRWEFSNPTPFIRSREFLWQEGHTAFATKAEADEEVLQILELYRRIYEE  232
           LPL +NQW NV RWE     PFIR+ EFLWQEGHTA AT   EA++E Q++E+Y R   E
At5g: 189  LPLMINQWANVTRWEMRT-KPFIRTLEFLWQEGHTAHATPEEAEKEAKQMIEIYTRFAFE  247
               I4                                      I5

                                  I7
At3g: 233  YLAVPVVKGMKSENEKFAGGLYTTSVEAFIPNTGRGVQGATSHCLGQNFAKMFEINFENE  292
            A+PV+ G KS+ E FAG   T ++EA + + + +Q  TSH LGQNF++ F   F +E
At5g: 248  QTAIPVIPGRKSKLETFAGADITYTIEAMMGDR-KALQAGTSHNLGQNFSRAFGTQFADE  306
                                                              I6

                           I8
At3g: 293  KAETEMVWQNSWAYSTRTIGVMIMTHGDDKGLVLPPKVASVQVVVIPVPYKDANTQGIYD  352
           E + VWQ SWA STR +G +IMTHGDD GL+LPPK+A +QVV++P+   KD    G+
At5g: 307  NGERQHVWQTSWAVSTRFVGGIIMTHGDDTGLMLPPKIAPIQVVIVPIWKKDTEKTGVLS  366
                                                    I7

                                                      I9
At3g: 353  ACTATASALCEAGIRAEEDLRDNYSPGWKYSDWEMKGVPLRIEIGPRDLENDQVRTVRRD  412
           A ++   AL  AG+R + D  D  +PGWK++ WEMKG+PLRIEIGPRD+ ++ V   RRD
At5g: 367  AASSVKEALQTAGVRVKLDDTDQRTPGWKFNFWEMKGIPLRIEIGPRDVSSNSVVVSRRD  426
                                                      I8

At3g: 413  NGVKEDIPRG------SLVEHVKELLEKIQQNMYEVAKQKREACVQEVKTWDEFIKALNE  466
              K     G      +LV +VKE L++IQ ++ E A   R++ + +V ++ E   A++
At5g: 427  VPGKAGKVFGISMEPSTLVAYVKEKLDEIQTSLLEKALSFRDSNIVDVNSYAELKDAISS  486
                                                      I9

                     I10                        I11
At3g: 467  KKLILAPWCDEEEVERDVKARTKGETGAAKTLCSPFDQPELPEGT-LCFASGKPAKKWTY  525
           K    PW   + E+   R K ETGA    C PF+Q    +GT  C  +G PA++
At5g: 487  GKWARGPWSASDADEQ----RVKEETGAT-IRCFPFEQ---TQGTKTCLMTGNPAEEVAI  538


At3g: 526  WGRSY    530
           + +SY
At5g: 539  FAKSY    543
```

Figure 12. Phylogenetic relationship between ProRSs from several organisms

The phylogenetic relationship of ProRSs was analyzed. Abbreviations are as follows; mitochondrial (Mt), organellar (Org), and cytoplasm (Cy).
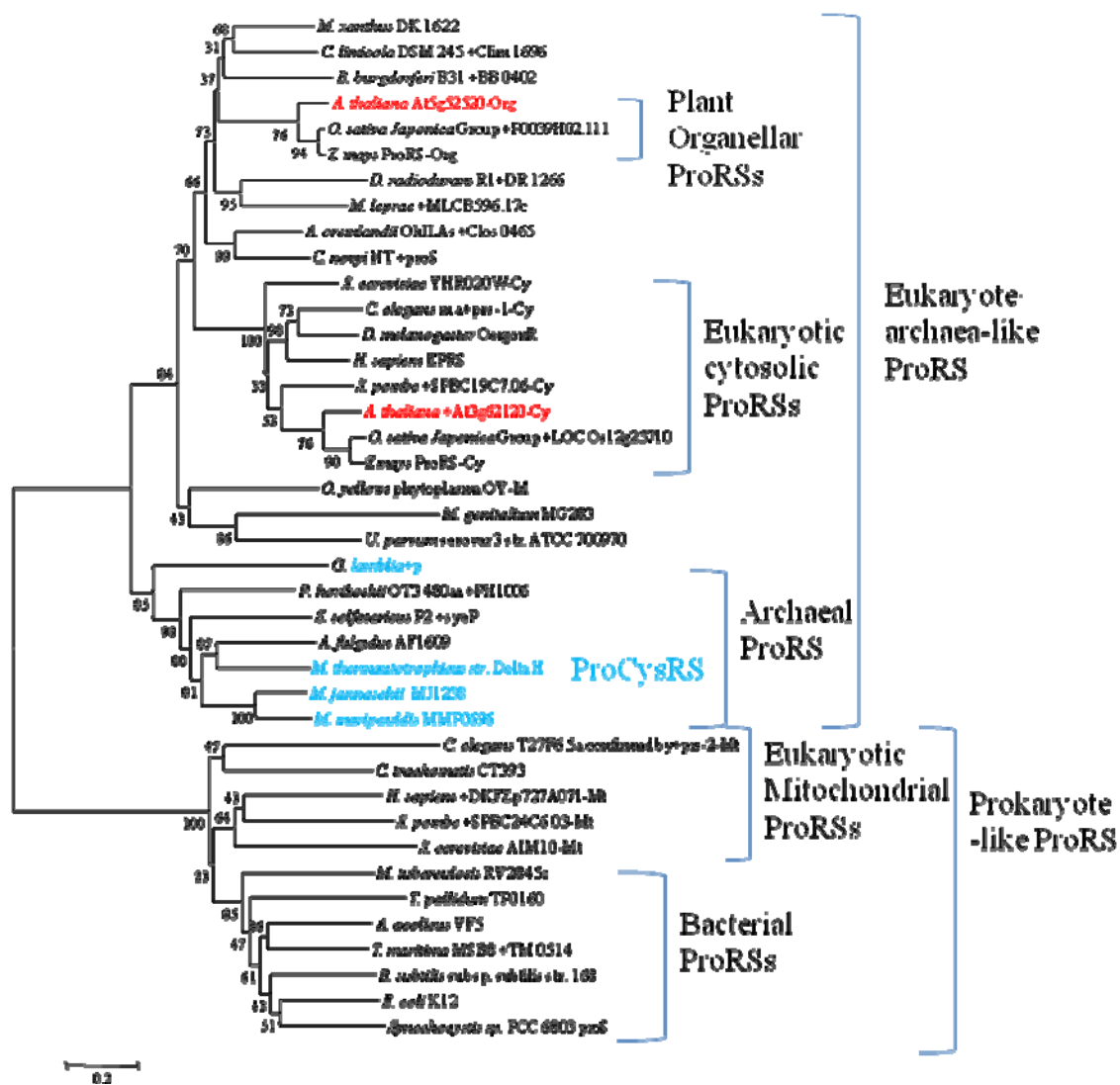
Figure 13. Functional complementation of *E. coli* ProRS temperature sensitive mutant strain UQ27

UQ27 transformed with pBAD33, pBAD33-*E. coli* CysRS, pBAD33-At3g62120, and pBAD33-At5g52520 was plated onto minimal agar medium containing chloramphenicol, 20 amino acids, and arabinose as described in the Materials and Methods. The cultures were incubated for 2 days at the permissive temperature of 30°C (A) or the nonpermissive temperature of 42°C (B).
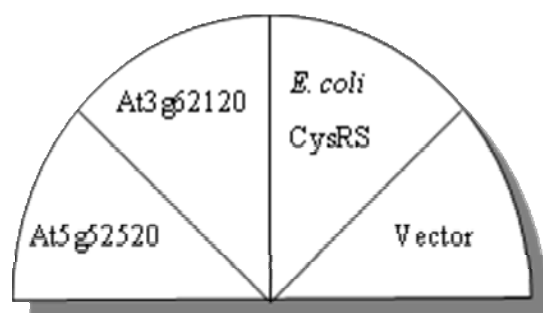
(A)
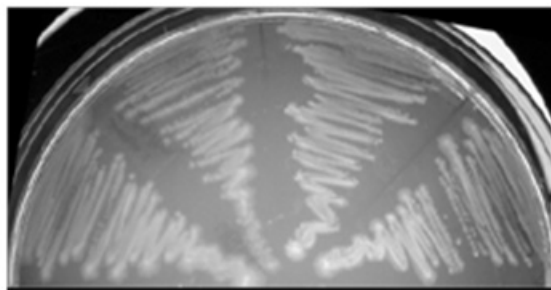
(B)
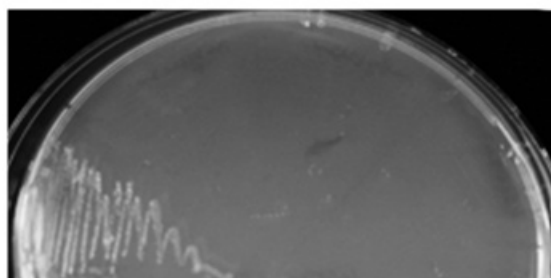
Figure 14. Growth curve of *E. coli* ProRS temperature sensitive mutant strain UQ27

UQ27 transformed with pBAD33, pBAD33-*E.coli* CysRS, pBAD33-At3g62120, and pBAD33-At5g52520 was inoculated into liquid minimal medium containing chloramphenicol, 20 amino acids, and glucose or arabinose. The graphs correspond to medium with arabinose incubated at 30°C (A), with arabinose incubated at 42°C (B), and with glucose incubated at 30°C (C) with glucose incubated at 42°C (D). The cultures were grown for 16 hours and absorbance was measured at 600 nm every 2 hours.

(A)

(B)

(C)

(D)

pBAD33
pBAD33+At3g62120
pBAD33+At5g52520
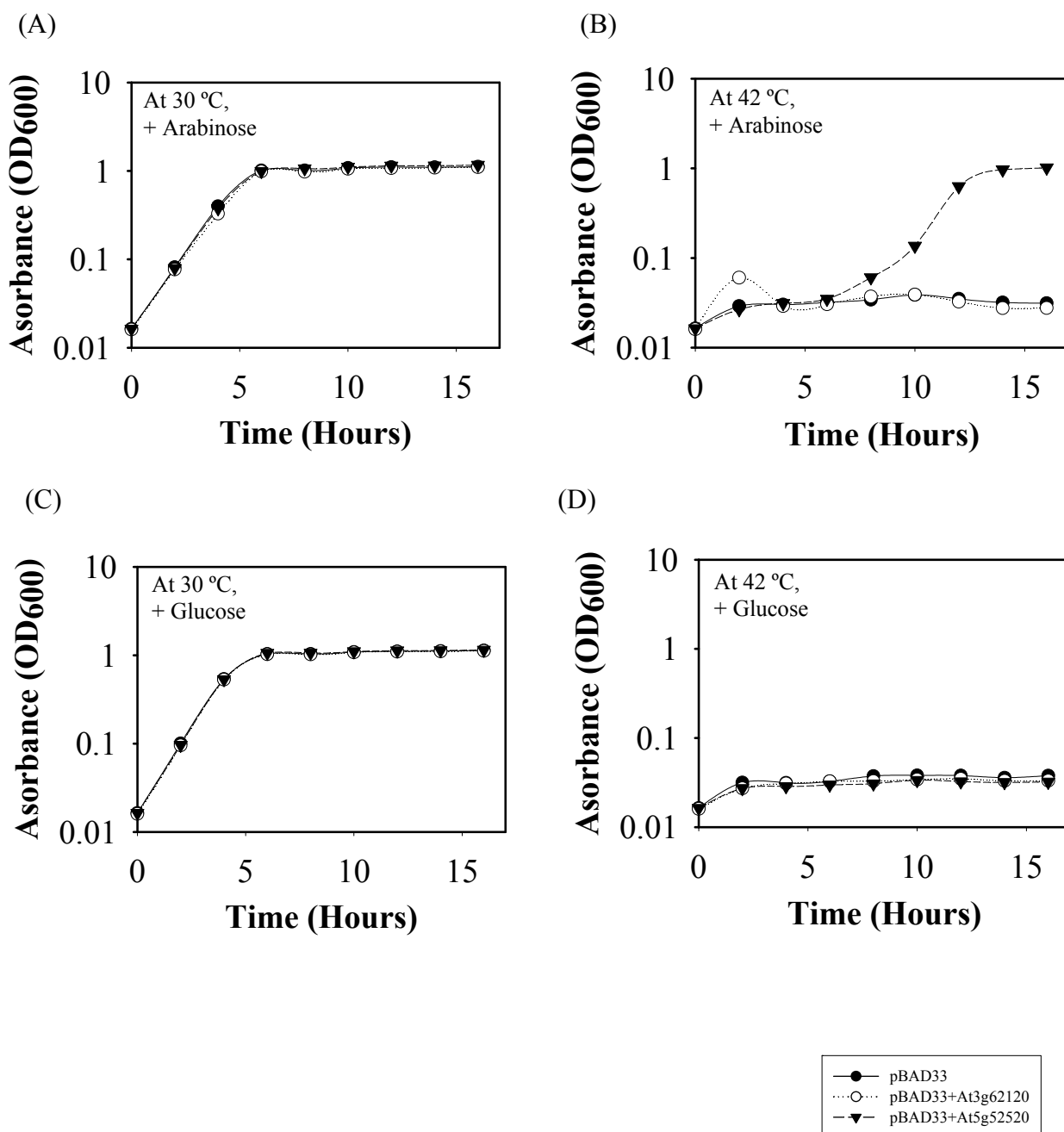
Figure 15. Functional complementation of *E. coli* CysRS temperature sensitive mutant strain UQ818

UQ818 transformed with pBAD33, pBAD33-*E.coli* CysRS, pBAD33-At3g62120, and pBAD33-At5g52520 were plated onto minimal agar medium containing chloramphenicol, 20 amino acids, and arabinose as described in Materials and Method. The cultures were incubated for 2 days at the permissive temperature of 30°C (A) or at the nonpermissive temperature of 42°C (B).
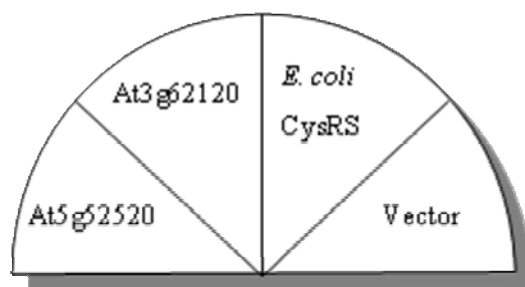
(A)
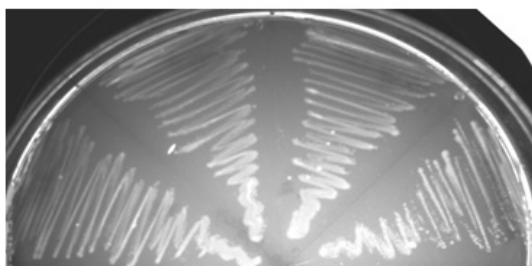
(B)
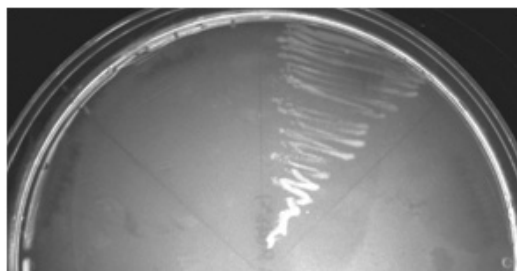
Figure 16. SDS-PAGE of AtProRS proteins followed by staining with Coomassie Brilliant Blue

The protein expressed from At3g62120 (A) and At5g52520 (B). Parallel cultures were grown and sampled prior (lane 1 and 4) and after induction of gene expression with 1 mM IPTG for 3 hours (lanes 2 and 5). AtProRS purified by Ni-affinity chromatography was also analyzed (lane 3 and 6). Fifteen micrograms proteins were loaded in all lanes. Kaleidoscope pre-stained standards (Bio-Rad) were loaded into the lane labeled M.

(A)
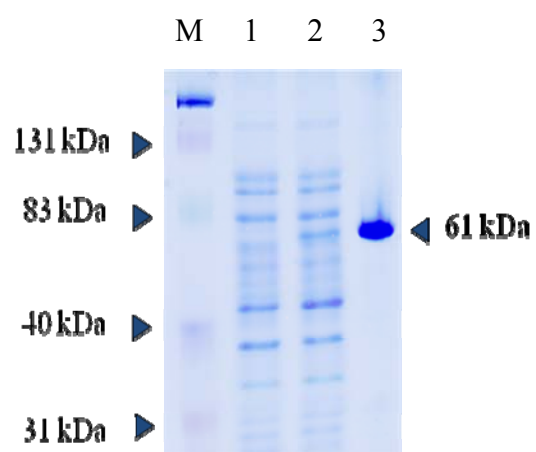
(B)
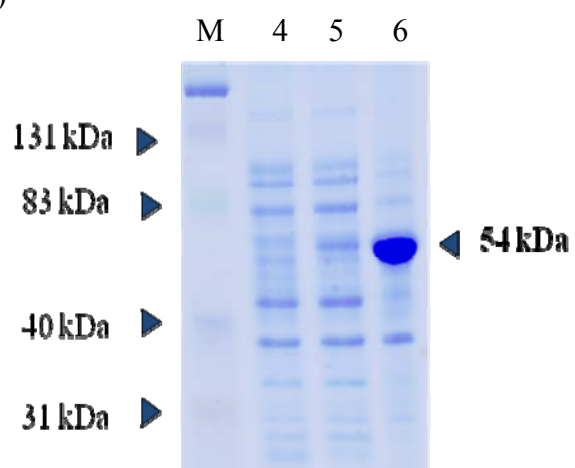
Figure 17. Prolyl-AMP synthesis of At5g52520 as measured in the ATP-PP$_i$ exchange reaction

The rate of Pro-dependent PP$_i$ exchange activity (●) is shown as a function of enzyme concentration (A) or time (B) at 30°C. The Pro concentration used to measure the enzyme dependent activity was 0.5 mM. No activity was observed in the absence of Pro (○). The background activity was measured with a reaction lacking enzyme. The experiment in graph A was assayed for 10 min. The amount of enzyme added to the reaction shown in graph B was 0.1 µg. The data are presented as the mean ±standard deviation of two independent experiments performed with a single protein preparation. For the data points without visible standard deviation bars the deviation was less than the diameter of the symbol.

(A)
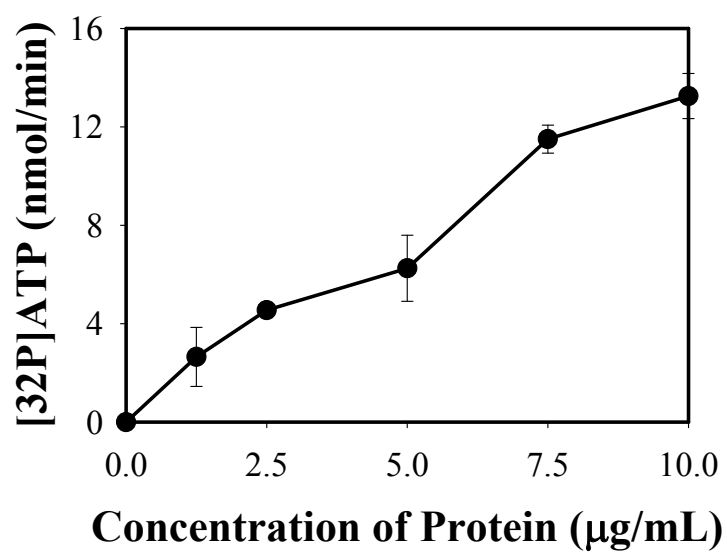
(B)

Figure 18. Prolyl-AMP synthesis of At3g62120 as measured in the ATP-PP$_i$ exchange reaction

The rate of Pro-dependent PP$_i$ exchange activity (●) is shown as a function of enzyme concentration (A) or time (B) at 30°C. The Pro concentration used to measure the enzyme dependent activity was 0.5 mM. No activity was observed in the absence of Pro (○). The background activity was measured with a reaction lacking enzyme. The experiment in graph A was assayed for 10 min. The amount of enzyme added to the reaction shown in graph B was 0.1 µg. The data are presented as the mean ±standard deviation of three independent experiments performed with a single protein preparation. For the data points without visible standard deviation bars the deviation was less than the diameter of the symbol.

(A)

(B)

Figure 19. Aminoacylation activity is dependent on tRNA substrate

Aminoacylation was carried in a reaction with yeast tRNA (●), E. coli (○) or no tRNA

(▼) at 30°C. The reactions were carried out as described in Materials and Methods with

0.1 μg At5g52520 (A) or At3g62120 (B). The graph shows the result of single

measurement. The prohibitive cost of [^{3}H]Pro did not allow for multiple measurements.

(A)

(B)

Yeast tRNA
E.coli tRNA
- tRNA
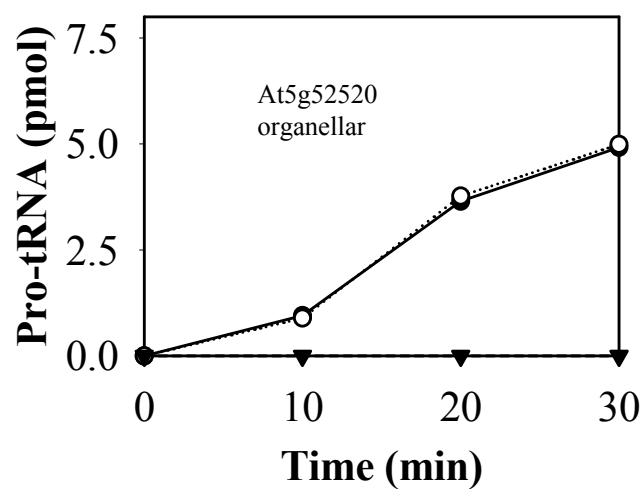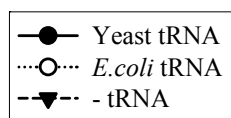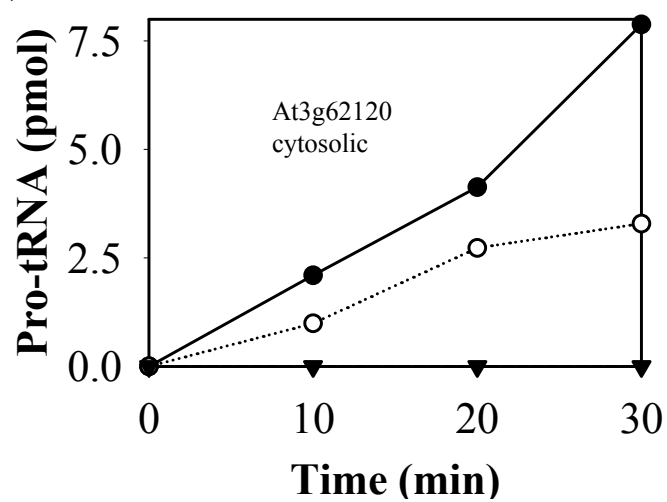
Figure 20. Phylogenetic comparison tRNA$_{PRO}$ of *A. thaliana*, *S. cerevisiae* and *E. coli*

The nucleotide sequences of tRNA$_{PRO}$ were analyzed by ClustalW. The nuclear encoded tRNA$_{PRO}$ are indicated by their locus tags (At nuclear) and the organellar encoded tRNA$_{PRO}$ by the Arth prefix followed by Mt for the mitochondrial encoded form or Ct for the chloroplast encoded form. *S. cerevisiae* tRNA$_{PRO}$ are indicated on the diagram and the mitochondrial encoded form is indicated by TP(UGG)Q, all others are nuclear encoded. *E. coli* tRNA$_{PRO}$ are indicated on the diagram. The mitochondrial encoded tRNA$_{PRO}$ cluster and the *E. coli* tRNA$_{PRO}$ cluster are indicated on the diagram. The cluster includes nuclear encoded tRNA$_{PRO}$ that may be imported into mitochondria.

Figure 21. Kinetic analysis of AtProRS

Pro or A2C were titrated into a PPi exchange assay at 30°C. The saturation curves for Pro

(●) and A2C (○) are presented in graph A (At5g52520) and graph B (At3g62120). This

assay was repeated three times with three independently isolated enzyme preparations.

Independent replicates were not averaged because enzyme from the three purifications

differed in specific activity. The other two data sets are shown in appendix B (Fig B1 and

B2).

(A)
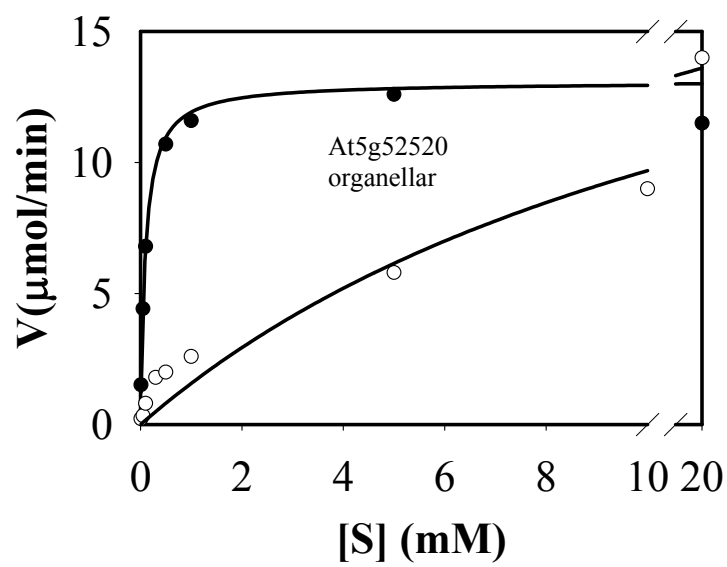
(B)
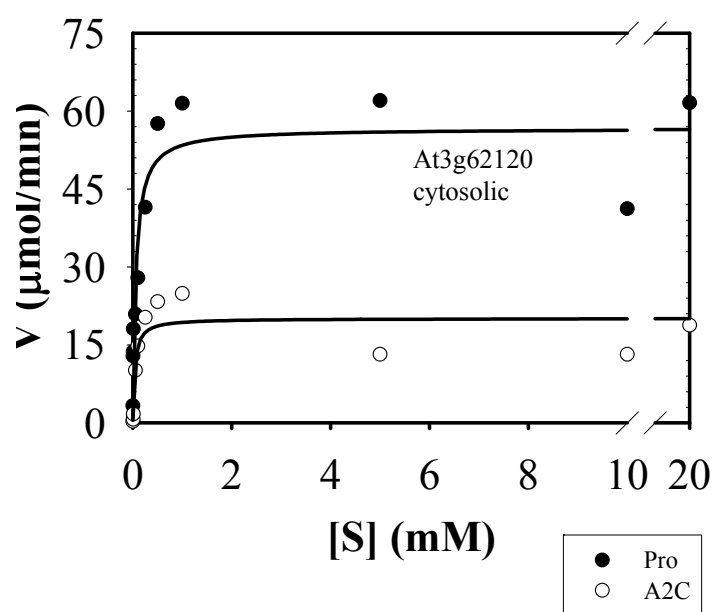
Figure 22. Inhibition of Pro aminoacylation by A2C

A2C was titrated into an aminoacylation assay containing 20 µM [^{3}H]Pro. The reactions were incubated at 30ºC for 30 min. Inhibition was assayed for At3g62120 (●) or At5g52520 (○). The assays were performed as a single measurement.
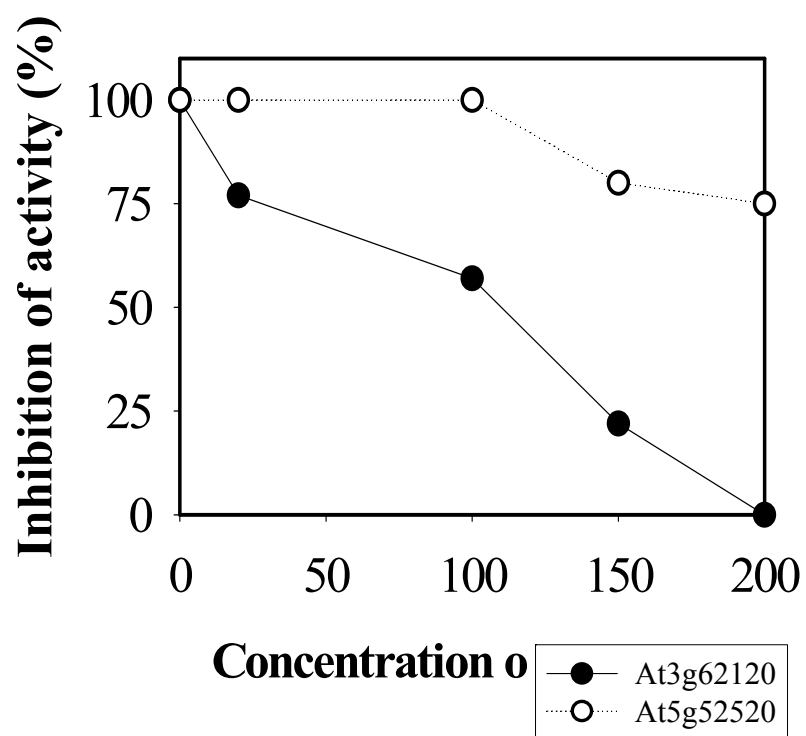
Figure 23. SDS-PAGE of ZmProRS proteins followed by staining with Coomassie Brilliant Blue

The protein expressed from ZmProRS-Cy and ZmProRS-Org (B). Parallel cultures were grown and sampled prior (lane 1 and 4) and after induction of gene expression with 1 mM IPTG for 3 hours (lanes 2 and 5). ZmProRS purified by Ni-affinity chromatography was also analyzed (lane 3 and 6). Fifteen micrograms protein was loaded in lane 3 and 10 µg protein was loaded in lane 1, 2, 4, 5, and 6. Bench marker protein ladder (Invitrogen) was loaded into the lane M in (A) and kaleidoscope pre-stained standard (Bio-Rad) was loaded into the lane labeled M in (B).
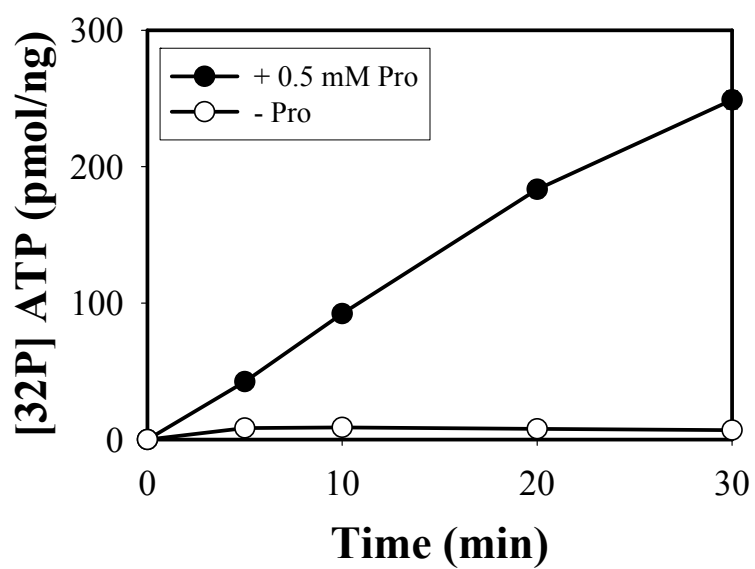
(A)
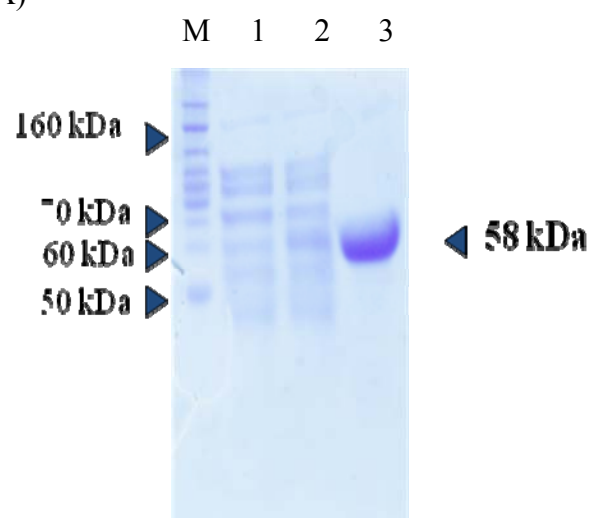
(B)
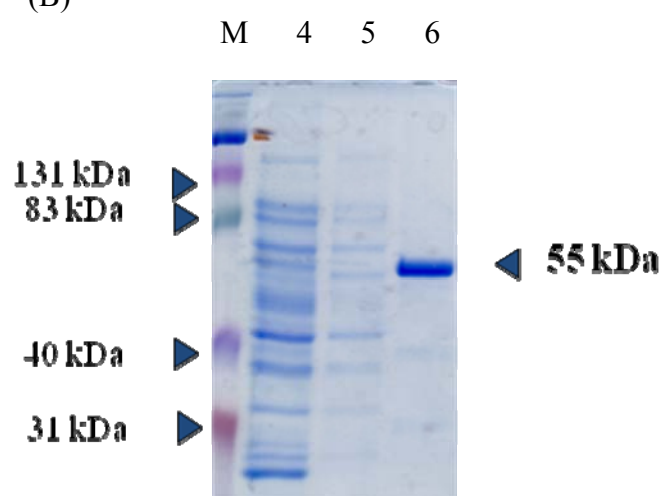
Figure 24. Prolyl-AMP synthesis of ZmProRS-Cy as measured in the ATP-PP$_i$ exchange reaction

The rate of Pro-dependent PP$_i$ exchange activity (●) is shown as a function of enzyme concentration (A) or time (B) at 30°C. The Pro concentration used to measure the enzyme dependent activity was 0.5 mM. No activity was observed in the absence of Pro (○). The background activity was measured with a reaction lacking enzyme. The experiment in graph A was assayed for 10 min. The amount of enzyme added to the reaction shown in graph B was 0.1 µg. The data are presented as the mean ±standard deviation of three independent experiments performed with the same protein preparation. For the data points without visible standard deviation bars the deviation was less than the diameter of the symbol.
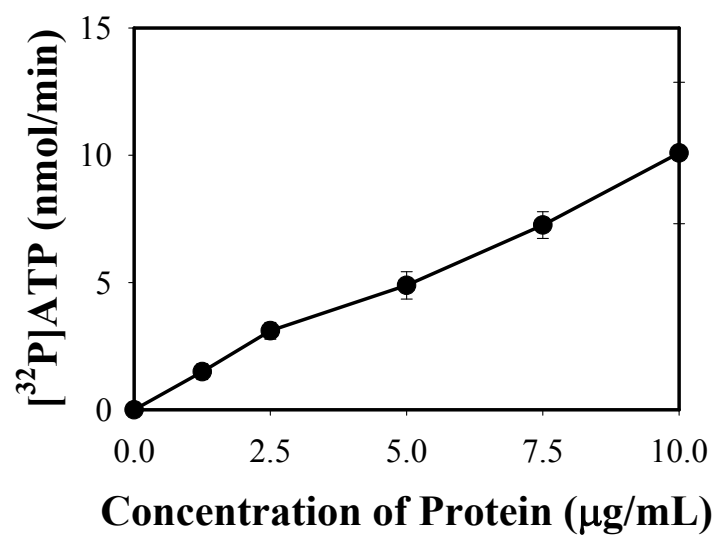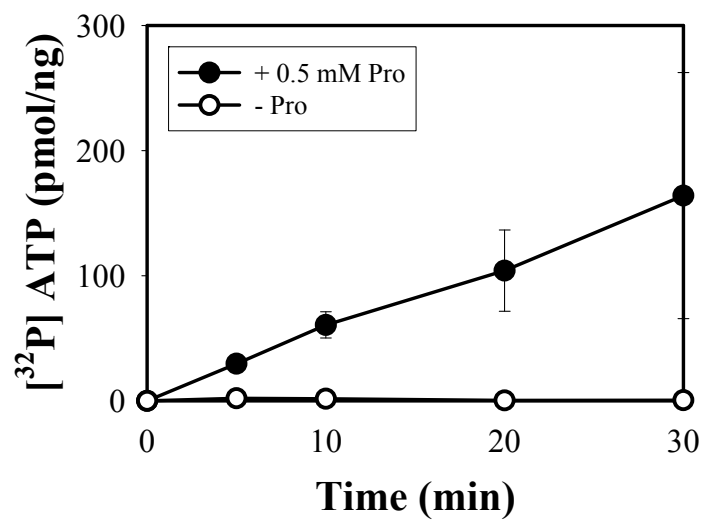
(A)

(B)

Legend:
- ● + 0.5mM Proline
- ○ - Proline
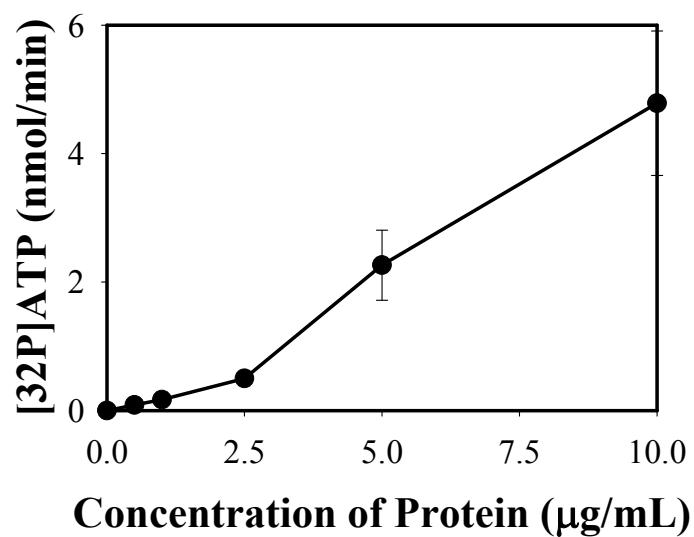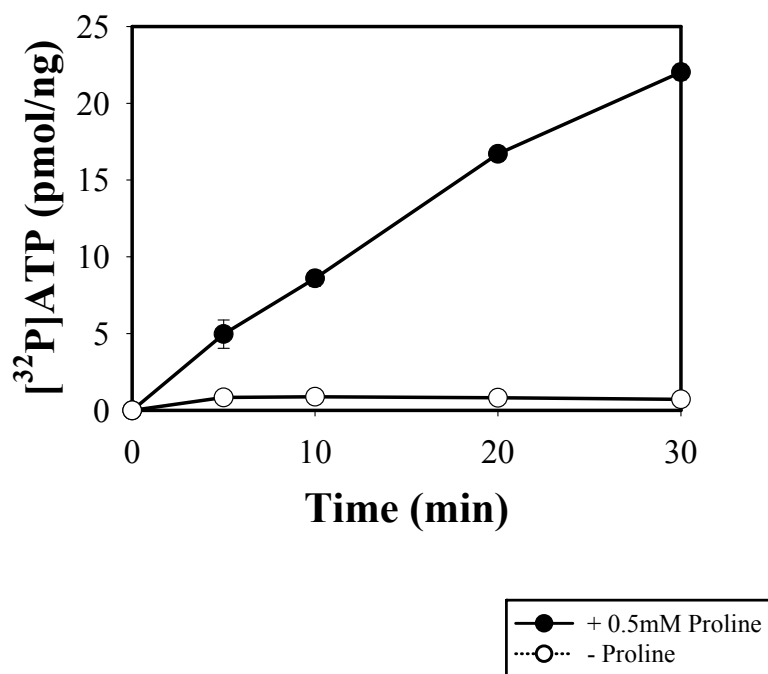
Figure 25. Prolyl-AMP synthesis of ZmProRS-Org as measured in the ATP-PP$_i$ exchange reaction

The rate of Pro-dependent PP$_i$ exchange activity (●) is shown as a function of enzyme concentration (A) or time (B) at 30°C. The Pro concentration used to measure the enzyme dependent activity was 0.5 mM. No activity was observed in the absence of Pro (○). The background activity was measured with a reaction lacking enzyme. The experiment in graph A was assayed for 10 min. The amount of enzyme added to the reaction shown in graph B was 0.1 µg. The data are presented as the mean ±SD of two independent experiments performed with same protein preparation.

(A)
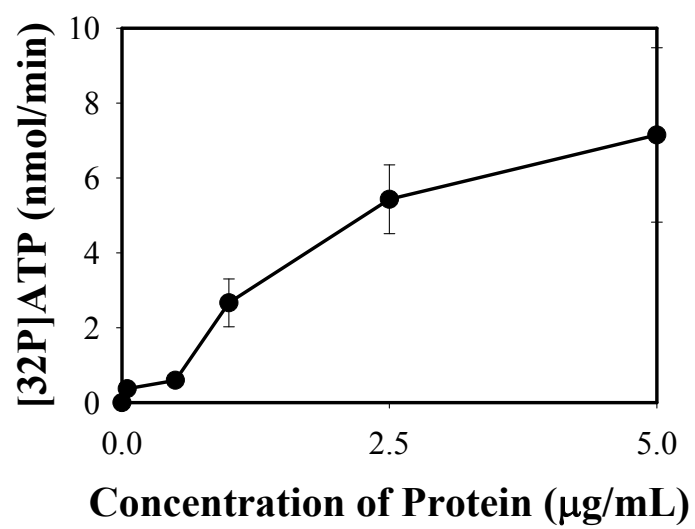
(B)
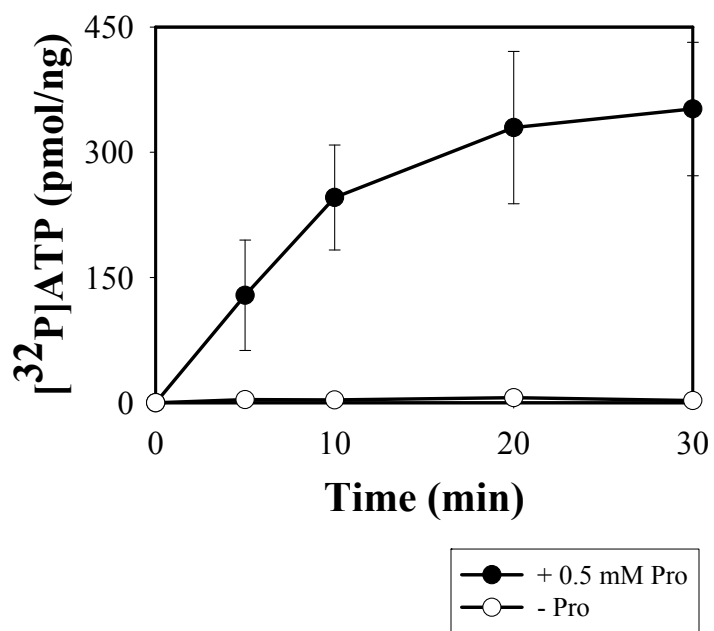
Figure 26. Kinetic analysis of ZmProRS

Pro or A2C was titrated into a PPi exchange assay at 30°C. The saturation curves for Pro

(●) and A2C (○) are presented in graph A (ZmProRS-Org) and graph B (ZmProRS-Cy).

This assay was repeated two times with two independently isolated enzyme preparations.

The second data set is shown in appendix B (Fig. B3).

(A)
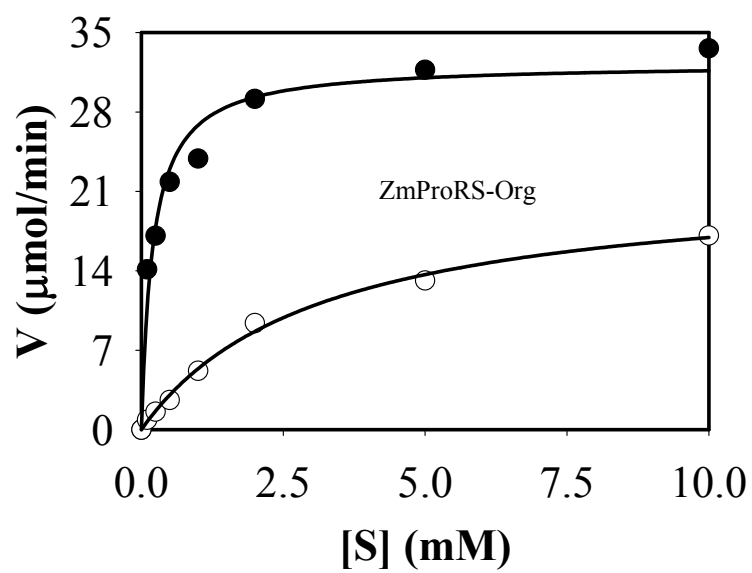
(B)
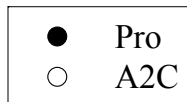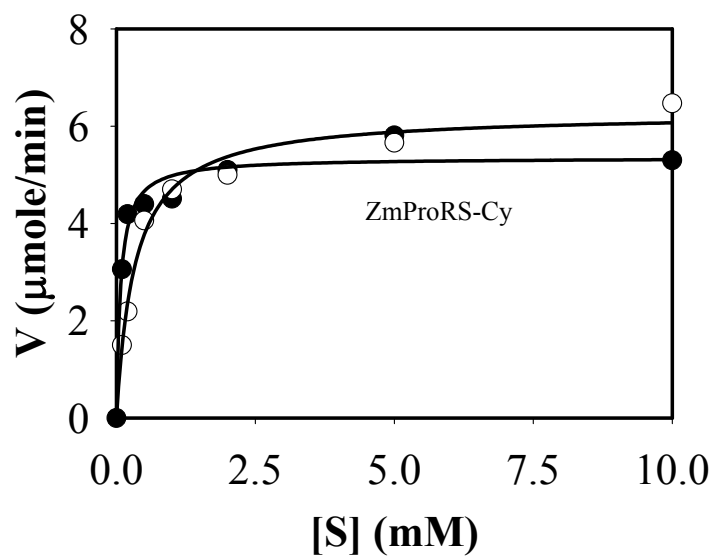
Figure 27. *C. majalis* **(A and B) and** *P. pubescens* **(C and D)**

A

B

C

D

Figure 28. The profile of the amino acids from leaves of *C. majalis*

The chromatogram shows amino acids profile from the leaves of *C. majalis*. The major

amino acids are labeled. Norleucine (Norleu) was marked as internal standard.
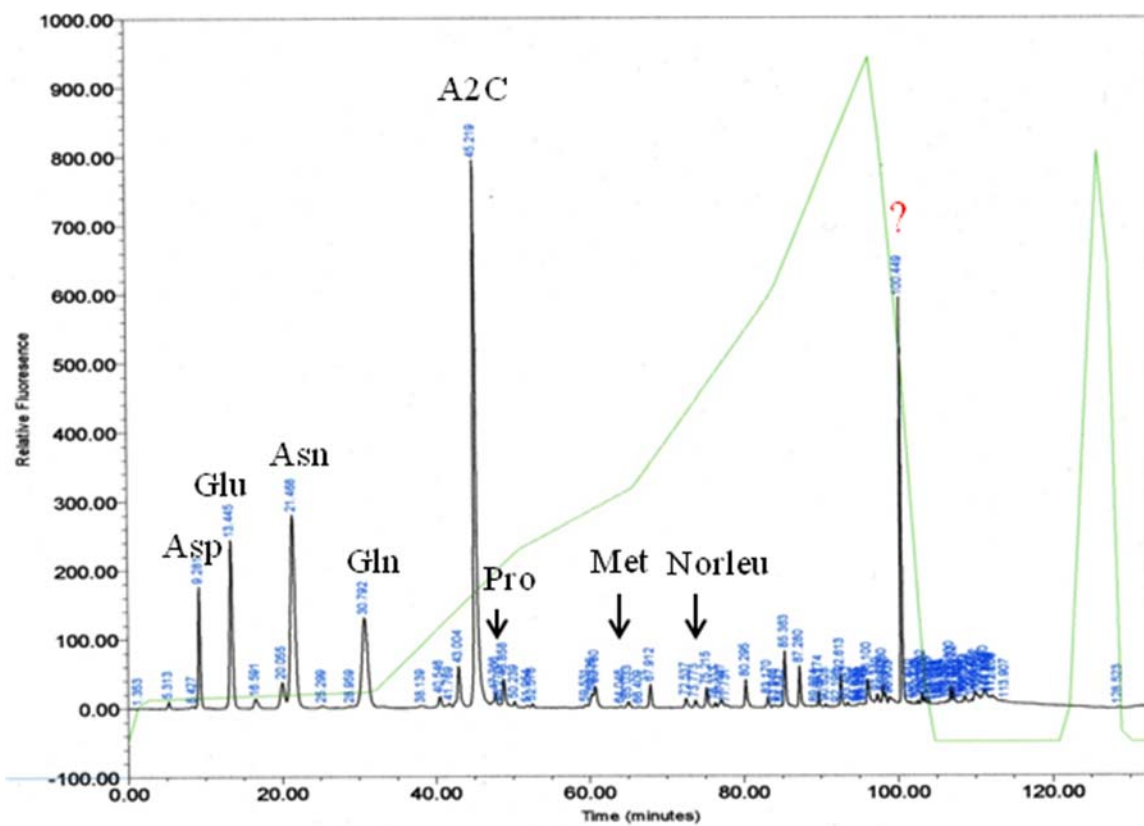
Figure 29. Design of oligonucleotide primers to clone ProRS cDNAs from *C. majalis* and *P. pubescens*

The cDNA sequences of the plant ProRSs were aligned by ClustalW. The highly conserved sequences are underlined. The accession numbers for the sequences in the alignment are as follows; AtProRS (NM_180405), OsProRS (AK106594) and RsProRS (AB097406).

```
AtProRS  362    GGTCATAGCAACATGTACTTCCCACAGTT ............ GAACCCCAGGATGGAAGTTCAATTTCTGGG  1199
OsProRS  372    GGGCACAGCAACATGTACTTCCCTCAGTT ............ GGACACCAGGATGGAAGTTCAATTTTTACG  1205
RsProRS  369    GGGCATAGCAACATGTACTTCCCACAGTT ............ GAACTGCAGGATGGAAGTTCAATTTCTGGG  1202
```

Consensus
DNAsequence   AGCAACATGTACTTCCC   GGATGGAAGTTCAATTT

Primer   Primer A 5'-AGCAACATGTACTTCCC-3'   Primer B 5'- AAATTGAACTTCCATCC-3'

Figure 30. The strategy for cloning the ProRS gene from *P. pubescens*

An initial PCR reaction using primers A and B (Fig. 29) from conserved regions of organellar plant ProRSs was used to amplify product I. Based on the sequence of product I, two additional primers (C and D) were designed and used in combination with plasmid-specific primers (E and F) to amplify the 5' (II) and 3' (III) regions of the ProRS sequence. Then, based on the sequences of the flanking amplicons, primers G and H were designed to clone the ProRS coding sequence. The diagram of the cloning strategy is given in panel (A). The sequences of products I, II, and III products from *P. pubescens* are given in panel (B). The overlap sequences of the products are given in panel (C). The overlap regions are indicated in red and blue in panel (B) and (C). The poly A tail is indicated as shaded letters, and the vector sequences are indicated as bold shading.

(A)
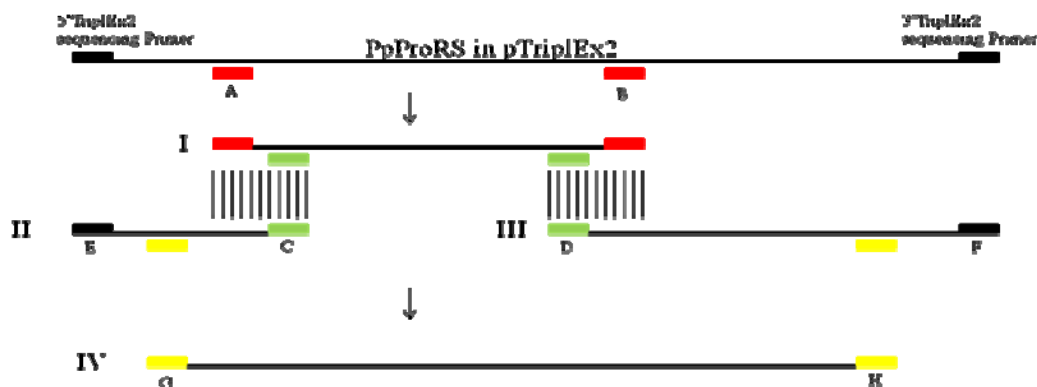
(B)

II CGCCATTGTGTTGGTACCCGGGAATTCGGCCATTACGCGGATCCGAACACTGCGTTTGCTGGCTTTGATGG
GGCTGATAGCGATGAATGAACACTAGCAATATGTACTTTCCTCAGTTCATACCATTCTCATTTATAGAGAA
GGAAGCCA

I AGCAATATGTACTTTCCTCAGTTCATACCATTCTCATTTATAGAGAAGGAAGCCAGCCACATTGAAGGCTT
CAGCCCAGAGCTAGCACTAGTTACGATTGGGGGAGGAAAGGAACTCGAGGAAAAACTTGTGGTACGACCAA
CAAGTGAAACCATTGTGAATCACATGTTTGCACAATGGATTCAGAGCTATCGCGAACTTCCTCTAATGGTT
AACCAGTGGGCAAATGTCACAAGATGGGAAAGCAGGACAAAGCCATTTATAAGGACTCTCGAGTTTCTATG
GCAGGAGGGCCATACAGCTCATGCAACACCTGAAGAGGCAGAAARGAAGGCACTACAAATGATCAATGTAT
ATAAGAAATTTGCTTATGAGCAAGCTGCAATACCTGTTATTGTAGGCAGGAAATCAAGGGCTGAATCATTT
ACTGGTTCTATTTGTaCATATaCAATTGAAGCGATGATGGGTGaCAGGAAGGCTCTACAGGCTGGAACCAG
CCATAACCTTGGCCAAACCTTTTCTCGTGCCTTTGGAACTCAGTTCACTGATGAAAATGGAGAAAGACGAT
ATGTATGGCAGACCTCCTGGGGTGTCAGCACTCGTTTCGTTGGTGGTATTATCATGACTCATGGGGATGAT
TCTGGTTTAATGCTTCTTCCCAGATTAGCACCAATACAGGTGGTAATTGTTCCCATTTGGAAGAAAGCTGA
TGAAAAAATGGGGGTTCTTAATGCTGCTACATCTGTTGAGAAAATACTGAAAACAGCTGGAATCATGGTCA
AGTTGGATGACTCAGAACAGAGGACCCCAGGATGGAAATTCAATTT

III ATGGTCAAGTTGGATGACTCAGAACAGAGGACCCCAGGATGGAAATTCAATTTCTGGGAGATGAAAGGCGT
TCCTTTGAGGACAGAAGTGGGACCTCGCGATGTTGCCAGTGGGAGTGTAGTTGTGTCTAGAAGAGATATCC
CCGGAAAACTTGGAAAGGAGTTTGGGGTATCCATGGAGCCTTTAGTTCTGGAAGCCCATGTGAAGAACAGG
TTGGAGGAGGTCCAGGCTTCATTATTGCAGAGAGCTACAACATTCCGTGATAGTAACATTGTGGACGTTAG
CTCCTATGAGGAGTTGAAGGGTGCTATTTCTCAAGGGAAATGGGCAAGAGGTCCATGGCCAGCCAGTGGTG
CCGAGGAGTTGAAGGTGAAGGAGGAGACAGGGGCGACAATTCGCTGCTTCCCATTTGAACAGCCCCAAGGT
CCAAAGACTTGCTTCATGACTGGTAACCCAGCTGAAGAAGTTGCAATTTTTGCAAGATCTTATTAAGAAGT
TTTATTTTCTTCCTCCCCATCTGAAAGTTTCCACATTCATTTTAAGCTTCTCCTTGCTAATTAATATGTAG
AGTTATGTGATGTGATTTGTTCATATTTTTCGAGTGATCCTAATTTATATGCAATGCTTCTGATAAATTTT
GTTTCCCATGAGTGTTAAATTATATAAATTGTGTTGGGAAAAACCGCCAGCATTTGAAAGAATTTTATTTG
TATAAATTATATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACATGTCGGCCGCCTCGGCCCAGTTGACTTT
AGACTCGAGCAAGCTTATGCATGCGGCCGCAATTCGAGCTCACTTGGCCAATTCG**CCCTATAGTGAGTCAA
TCG**

(C) **I and II conserved region**

    I   AGCAATATGTACTTTCCTCAGTTCATACCATTCTCATTTATAGAGAAGGAAGCCA
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
    II   AGCAATATGTACTTTCCTCAGTTCATACCATTCTCATTTATAGAGAAGGAAGCCA

II and III conserved region

    II   ATGGTCAAGTTGGATGACTCAGAACAGAGGACCCCAGGATGGAAATTCAATTT
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
    III   ATGGTCAAGTTGGATGACTCAGAACAGAGGACCCCAGGATGGAAATTCAATTT

Figure 31. The strategy for cloning the ProRS gene from *C. majalis*

An initial PCR reaction using primers A and B (Fig. 29) from conserved regions of organellar plant ProRSs was used to amplify product I. Based on the sequence of product I, two additional primers (C and D) were designed and used in combination with plasmid-specific primers (E and F) to amplify the 5' (II) and 3' (III) regions of the ProRS sequence. Then, based on the sequences of the flanking amplicons, primers G and H were designed to clone the ProRS coding sequence. The diagram of the cloning strategy is given in panel (A). The sequences of products I, II, and III products from *C. majalis* are given in panel (B). The overlap sequences of the products are given in panel (C). The overlap regions are indicated in red and blue in panel (B) and (C). The poly A tail is indicated as shaded letters, and the vector sequences are indicated as bold shading.

(A)
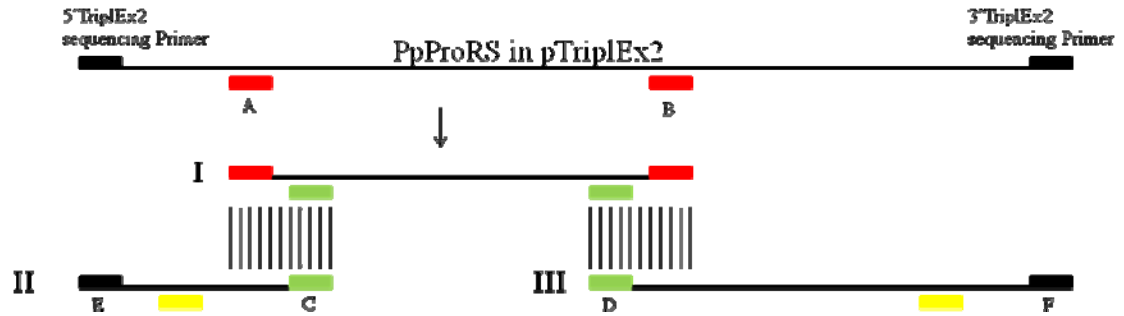
(B)

II   **CTCCGAGATCTGGACGAGC**ATGCTTTTGCTGGCTTTGATGAAAAGCACGATCCAATGGGAGCTCTTCGCCTC
CCATCTCTCGCTCTCCTCCTCTCCCCATCCAAAACCCTAACCCTAACCCTGAAACCCACTTCGCGCTACCAT
CTTCGCCCCTCCATTCGCAGGATATCGGCCTCGATCTCCGGCGACGGCGGCCGGCGGCGGTTCACCGCCCGC
AGCCTCGCGAAGGAGGAGGCGGCGGCGCCTGCTACTCCGCCTCCTCCTACAGCGGTGGCGGACCAGAAGACT
TCTTCGGATCAAGAGCGAGCGGTGACTCCGAGGTCGGTGGACTTCAGTGCTTGGCGCCTTGATGTCATCGCG
AGTGCCGAGCTGGCTGATTTATGGTCCGGTTCGTGGGACGATGGTTATTCGTTCTTATGGTTACGCTATTTG
GGAATCAATTCAGGAATATTTGAATGTGGAGTTTAGGGAAACAGGGGCCAGCAATATGTACTTTCCTCAGTT
CATACCATACTCATTTATAGAGAAGGAAGCCAGCCATATTGAAGGCTTCAGTCCAGAGCTAGCACTGGTTAC
AATTGGAGGAGGAAAGGAACTCGAGGAAAAACTCGTGGGTCTGGCAGGTAAGACCAACAAGTGAAACCATTG
TGAATCACATGTTTGCACAATGGATTCAGAGCTATCGTGAACTTCCTCTCATGATTAACCAGTGGGCAAATG
TCACAAGATGGGAAAGCTGGACAAAGCCATTTATAAGGACTCTCGAGTTTCTAT

I   AGCAATATGTACTTTCCTCAGTTCATACCATACTCATTTATAGAGAAGGAAGCCAGCCATATTGAAGGCTTC
AGTCCAGAGCTAGCACTGGTTACAATTGGAGGAGGAAAGGAACTCGAGGAAAAACTCGTGGGTCTGGCAGGT
AAGACCAACAAGTGAAACCATTGTGAATCACATGTTTGCACAATGGATTCAGAGCTATCGTGAACTTCCTCT
CATGATTAACCAGTGGGCAAATGTCACAAGATGGGAAAGCTGGACAAAGCCATTTATAAGGACTCTCGAGTT
TCTATGGCAGGAGGGCCATACAGCTCATGCAACACCTGACGAGGCAGAAGAGAAGGCACTACAGATGATCAG
ATGTCTATAAGAAATTTGCTTATGAGCAAGCTGCAATACCTGTTATTGTAGGCAGGAAATCAAGGGCTGAAG
CATTTACTGGTTCTGTTTGTACATATACAATTGAAGCGATGATGGGTGACAGGAAGGCTCTACAGGCTGGAA
CCAGCCACAACCTTGGCCAAACCTTTTCTCGTGCCTTTGGAACTCTGTTCACTGATGAAAATGGAGAAAGAC
AATATGTATGGCAGACCTCCTGGGGTGTCAGCACGCGTTTCGTTGGTxxGGxTATTATCGTGACTCATGGGG
ATGATGCTGGTTTAATGCTTCCTCCCAGATTAGCACxCAATACAGGTGGTAATTGTTCCCATTTGGAAGAAA
GCTGATGAAAAAATAGGGATTCTTAATGCTGCTACATCTGTTGAGAAAATACTGAAAACAGCTGGAATCAAG
GTCAAGCTTGATAACTCAGAACAGAGGACGCCAGGATGGAAATTCAATTT

III TTCTTAATGCTGCTACATCTGTTGAGAAAATACTGAAAACAGCTGGAATCAAGGTCAAGCTTGATAACTCAG
AACAGAGGACGCCAGGATGGAAATTCAATTTTTGGGAGATGAAAGGCGTTCCTTTGAGGATAGAAGTGGGAC
CTCGCGATGTCACCAGTGGGAGTGTAGTTGTGTCTAGAAGAGATATTCCCGGAAAGCCTGGAAAGGAGTTTG
GGGTTTCCATGGAGCCTTCAGTTCTGGAAGCCCATGTGAAGAACAGGTTGGAGGACGTCCAGGCTTCATTAT
TGCAGAGGGCTACAACATTCCGTGATTGTAACATTGTGGACGTTAGCTCCTATGAGGAGTTGAAGGATGCTA
TTTCTCAAGGGAAATGGGCAAGAGGTCCATGGTCAGCCAGTGATGCTGAGGAGTTGAAGGTGAAAGAGGAGG
CAGGCGCGACAATTCGCTGCTTCCCATTTGAACAGCCCCAAGGCCCAAAAATTTGCTTCATGACTGGTAACC
CGGCTGACGAAGTTGCAATTTTTGCGAAATCTTATTAAGAAGTTCCGTTTTCTTCCACTCCATCTGAAAGTT
TCCACATTCATTTTAAGCTTCTCCTTGCTTAATATGTAGAGTTATGTGATGTAATTTGTTCATATTTTTCGA
ATGATCMTAATTTATATGTGATGCTTCYGATAAATTTTGTTTCCAAAAAAAAAAAAAAA

(C)

I and II conserved region

```
 I    AGCAATATGTACTTTCCTCAGTTCATACCATACTCATTTATAGAGAAGGAAGCCAGCCATATTGAAGGC
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 II   AGCAATATGTACTTTCCTCAGTTCATACCATACTCATTTATAGAGAAGGAAGCCAGCCATATTGAAGGC

 I    TTCAGTCCAGAGCTAGCACTGGTTACAATTGGAGGAGGAAAGGAACTCGAGGAAAAACTCGTGGGTCTG
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 II   TTCAGTCCAGAGCTAGCACTGGTTACAATTGGAGGAGGAAAGGAACTCGAGGAAAAACTCGTGGGTCTG

 I    GCAGGTAAGACCAACAAGTGAAACCATTGTGAATCACATGTTTGCACAATGGATTCAGAGCTATCGTGA
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 II   GCAGGTAAGACCAACAAGTGAAACCATTGTGAATCACATGTTTGCACAATGGATTCAGAGCTATCGTGA

 I    ACTTCCTCTCATGATTAACCAGTGGGCAAATGCACAAGATGGGAAAGCTGGACAAAGCCATTTATAATG
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 II   ACTTCCTCTCATGATTAACCAGTGGGCAAATGCACAAGATGGGAAAGCTGGACAAAGCCATTTATAATG

 I    GACTCTCGAGTTTCTAT
      |||||||||||||||||
 I    GACTCTCGAGTTTCTAT
```

II and III conserved region

```
 II   TTCTTAATGCTGCTACATCTGTTGAGAAAATACTGAAAACAGCTGGAATCAAGGTCAAGCTTGATA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 III  TTCTTAATGCTGCTACATCTGTTGAGAAAATACTGAAAACAGCTGGAATCAAGGTCAAGCTTGATA

 II   ACTCAGAACAGAGGACGCCAGGATGGAAATTCAATTT
      |||||||||||||||||||||||||||||||||||||
 III  ACTCAGAACAGAGGACGCCAGGATGGAAATTCAATTT
```

Figure 32. Nucleotide and deduced amino acid sequence of the *P. Pubescens* cDNA encoding ProRS

The cDNA is 1368 bp in length and was obtained by amplification using primers PpProRS5'-*KpnI* (G) and PpProRS3'-*SalI* (H) (Fig.30A, IV). Numbers on the left represent nucleotide positions and those on the right represent amino acid positions.

```
1                                                   ATGAACACT
                                                    M  N  T          3
10   GCGTTTGCTGGCTTTGATGTGAAGTTTAAGGAAACAGGGGTCAGCAATATGTACTTTCCT
     A  F  A  G  F  D  V  K  F  K  E  T  G  V  S  N  M  Y  F  P    23
70   CAGTTCATACCATTCTCATTTATAGAGAAGGAAGCCAGCCACATTGAAGGCTTCAGCCCA
     Q  F  I  P  F  S  F  I  E  K  E  A  S  H  I  E  G  F  S  P    43
130  GAGCTAGCACTAGTTACGATTGGGGGAGGAAAGGAACTCGAGGAAAAACTTGTGGTACGA
     E  L  A  L  V  T  I  G  G  G  K  E  L  E  E  K  L  V  V  R    63
190  CCAACAAGTGAAACCATTGTGAATCACATGTTTGCACAATGGATTCAGAGCTATCGCGAA
     P  T  S  E  T  I  V  N  H  M  F  A  Q  W  I  Q  S  Y  R  E    83
250  CTTCCTCTAATGGTTAACCAGTGGGCAAATGTCACAAGATGGGAAAGCAGGACAAAGCCA
     L  P  L  M  V  N  Q  W  A  N  V  T  R  W  E  S  R  T  K  P    103
310  TTTATAAGGACTCTCGAGTTTCTATGGCAGGAGGGCCATACAGCTCATGCAACACCTGAA
     F  I  R  T  L  E  F  L  W  Q  E  G  H  T  A  H  A  T  P  E    123
370  GAGGCAGAAARGAAGGCACTACAAATGATCAATGTATATAAGAAATTTGCTTATGAGCAA
     E  A  E  X  K  A  L  Q  M  I  N  V  Y  K  K  F  A  Y  E  Q    143
430  GCTGCAATACCTGTTATTGTAGGCAGGAAATCAAGGGCTGAATCATTTACTGGTTCTATT
     A  A  I  P  V  I  V  G  R  K  S  R  A  E  S  F  T  G  S  I    163
490  TGTACATATACAATTGAAGCGATGATGGGTGACAGGAAGGCTCTACAGGCTGGAACCAGC
     C  T  Y  T  I  E  A  M  M  G  D  R  K  A  L  Q  A  G  T  S    183
550  CATAACCTTGGCCAAACCTTTTCTCGTGCCTTTGGAACTCAGTTCACTGATGAAAATGGA
     H  N  L  G  Q  T  F  S  R  A  F  G  T  Q  F  T  D  E  N  G    203
610  GAAAGACGATATGTATGGCAGACCTCCTGGGGTGTCAGCACTCGTTTCGTTGGTGGTATT
     E  R  R  Y  V  W  Q  T  S  W  G  V  S  T  R  F  V  G  G  I    223
670  ATCATGACTCATGGGGATGATTCTGGTTTAATGCTTCTTCCCAGATTAGCACCAATACAG
     I  M  T  H  G  D  D  S  G  L  M  L  L  P  R  L  A  P  I  Q    243
730  GTGGTAATTGTTCCCATTTGGAAGAAAGCTGATGAAAAAATGGGGGTTCTTAATGCTGCT
     V  V  I  V  P  I  W  K  K  A  D  E  K  M  G  V  L  N  A  A    263
790  ACATCTGTTGAGAAAATACTGAAAACAGCTGGAATCATGGTCAAGTTGGATGACTCAGAA
     T  S  V  E  K  I  L  K  T  A  G  I  M  V  K  L  D  D  S  E    283
850  CAGAGGACCCCAGGATGGAAATTCAATTTCTGGGAGATGAAAGGCGTTCCTTTGAGGACA
     Q  R  T  P  G  W  K  F  N  F  W  E  M  K  G  V  P  L  R  T    303
910  GAAGTGGGACCTCGCGATGTTGCCAGTGGGAGTGTAGTTGTGTCTAGAAGAGATATCCCC
     E  V  G  P  R  D  V  A  S  G  S  V  V  V  S  R  R  D  I  P    323
970  GGAAAACTTGGAAAGGAGTTTGGGGTATCCATGGAGCCTTTAGTTCTGGAAGCCCATGTG
     G  K  L  G  K  E  F  G  V  S  M  E  P  L  V  L  E  A  H  V    343
1030 AAGAACAGGTTGGAGGAGGTCCAGGCTTCATTATTGCAGAGAGCTACAACATTCCGTGAT
     K  N  R  L  E  E  V  Q  A  S  L  L  Q  R  A  T  T  F  R  D    363
1090 AGTAACATTGTGGACGTTAGCTCCTATGAGGAGTTGAAGGGTGCTATTTCTCAAGGGAAA
     S  N  I  V  D  V  S  S  Y  E  E  L  K  G  A  I  S  Q  G  K    383
1150 TGGGCAAGAGGTCCATGGCCAGCCAGTGGTGCCGAGGAGTTGAAGGTGAAGGAGGAGACA
     W  A  R  G  P  W  P  A  S  G  A  E  E  L  K  V  K  E  E  T    403
1210 GGGGCGACAATTCGCTGCTTCCCATTTGAACAGCCCCAAGGTCCAAAGACTTGCTTCATG
     G  A  T  I  R  C  F  P  F  E  Q  P  Q  G  P  K  T  C  F  M    423
1270 ACTGGTAACCCAGCTGAAGAAGTTGCAATTTTTGCAAGATCTTATTAA
     T  G  N  P  A  E  E  V  A  I  F  A  R  S  Y  .                443
```

Figure 33. Alignment of deduced amino acid sequence of plant ProRS homologs

ProRS sequences of *P. pubescens* were compared with those from several plant species: At55ProRS (At5g52520), At3ProRS (At3g62120), ZmProRS-Org (TC288824), and ZmProRS-Cy (AY104927). The sequences were aligned with ClustalW by using the DNA star program. Solid boxes indicate identical amino acid residues in at least 50 % of the sequences in the alignment. Intermediate or light boxes are those with less similarity to consensus. The arrow indicates the transit peptide cleavage site for At5g52520. Three motifs are indicated by dotted line, and the anticodon binding domain is indicated by dots. The motifs are as defined by Eriani et al. (1990), Cusack et al. (1998), Ibaa and Soll (2000), and Pesaresi et al (2006). Invariant residues of ProRSs are indicated by asterisks (Stehlin et al., 1998).

-- Motif 1 --

```
AtProRS-Cy   1   --------------MADPSEQKEKVVKEKKEKVKEKVVKEKVAKASSSGQKKKDVKKETGLGLSVKKDENGGEWTSEVCK-QDMIEWYDISGCYILEPWSMAIWEIMGIFFDAEIKKMKVKNGYFPLF
ZmProRS-Cy   1   ------------------------MSTNKGSAAKGGGGEKKEVKKETKLGMAYKKDDNEGEWTSEVVVNSEMIEWYDISGCYILEPWNAMEIWELLKEFFDAEIDLLLKLKPIYEPLF
AtProRS-Org  1   MVSSSLRLPSLTSLLFPATTRYPATLRRTVCLRNRPLSGFATAPSGTASPETWSSEVDRLFSDRAVTPRSQDRNATLLVVIASABUANLGPVRGTMVIRPYGTAIWBAIQDYLNVKFKETGHSNMYPPQF
ZmProRS-Org  1   ----------------------------------ETEGGDFEE----EQVTPRSADPNATWTTDVIAAABLALWGPVRGTMVIRPYGTAIWBAIQDYLNVKFKETGHSNMYPPQF
PpProRS      1   --------------------------------------MNTAPAG-----------F----------D-VKFKETGTSNMYPPQF
```

```
AtProRS-Cy   115  VSPGVLEKEIGHIEGEAPEDWVKSEKSILEVPIAIRPTSERVMYPYYSKWIRGHEDLPLKINQNVTRWEFSNPTPRIISEEFLWQEGHTAPATKAEADEBYLGLLBIIPERIYEBYILEPVWIPEKS
ZmProRS-Cy   93   VTENVLYKEIGHIEGEAPEDWVKSEKSILEAPIAIRPTSERVMYPYPGSKWIRGREDLPLRQNQVNVTRWEFSNPTPRIISEEFLWQEGHTAPANTEEADEBVLGILBLWEIYEBELAVGEVSIGRKKS
AtProRS-Org  131  IPTGSFIEKEASHIEGFSPELALVTIGGGKELEBKLVVRPTSBTIVNHMEEWKWIPSYRDLPLMANQWANVTRWBMGR-TKPPIRTLEFLWQEGETAEHATPEEABFEALQMIEWIREAFEQWAIPVIFGRKS
ZmProRS-Org  76   IPLGSFIEKEASHVEGFSPELALVTIGGGKELEBKLVVRPTSBTIVNHMEEWKWIWSYRDLPLMANQWANVTRWBR-TKPPIRTLEFLWQEGETAEHATPEEABFEALQMIDWIIEAFEBWAAIPVIFGRKS
PpProRS      26   IPTGSFIEKEASHIEGFSPELALVTIGGGKELEBKLVVRPTSBTIVNHMEEACWIWSYRBLPLMANQWANVTRWEBSR-TKPPIRTLEFLWQEGETAEHATPEEABXYALQMINVWIFEAFEQWAIPVIFGRKS
```

---------- Motif 2 ----------

---- Motif 3 ----

```
AtProRS-Cy   245  ENBYFAGGLYDTSVEAFIPNTGRGVTQGRAFSEFLGQNEAAKMGEIWEBNBXABEFVYWQNSWAYSTRTFGVIWTHGDDIGLWLPPRDTASYQVVVTFVPYKEANTQGTIYDACTATASAPCBAGIEAEEEELRD
ZmProRS-Cy   223  EMBYFAGGLYDTSVEAFIPNTGRGIQGAFSEFLGQNEAAKMGDITEBNSYGYEBTWQNSWAYDTHSIGTVWMYTHGDDIGLWLPPKDLAPIQVVIVSVPYKDADTTAIKGACBSTVYIFDDQSGIFADQFTPF
AtProRS-Org  260  KLEDFAGADIDTYTIEANMGDR-IALQAGTSEHLGQNFSRAFGTQPADENGGEWQVWQTSWANDSTRFVGGIIMTHGDDIGLMLPPKLAPIQVVIVPIWKKDFETFGVLSAASSVKEALQITAGVBRVELDDTD
ZmProRS-Org  205  KVEDFAGANFDTYTIEANMGDK-IALQAGTSEHLGQNFSRAFGTQELDENGGIBEVWQTSWALSTRFVGGIIMTHGDDAGLMLPPKVAPIQVEIVPIWKKGIBANVVSBAVTSVQKILKEBAGIRVKIDDSE
PpProRS      155  RAESFTFESIGTYTIEANMGDR-IALQAGTSEHLGQTFSRAFGTQPFDENGGEWEVWQTSWGVSTRFVGGIIMTHGDDIGLMLPPKLAPIQVVIVPIWKKAIGEKMGVLMAATSVEKILDYTAGTGTVKEDDSE
```

-- Anticodon binding domain --------------------

```
AtProRS-Cy   375  NYSPGWYKYSIWEMKGVPLRIBIGPEDLENQVRTYRRDNGVF------EDIPRGSLVEBEVVEBLLEYGQNMYEVEKQIFBACVQBWFTTWEBPIKALNBFKLILAFTCDEEBEYBRDVKAFBTKGBTGAAKTL
ZmProRS-Cy   353  NYSPGWYSEWEMKGVPLRIBIGPEDLAANKQVFEVWRRDNGAK------VDIPVTNLVEBKVYLLDBIGKNEFKTAQEYRBACVEYVONTWEBPTTALNNEKLILAFKCDEBEBEKDVKIFAKGBLGAAKTL
AtProRS-Org  389  QRTPGWKFNEWEMKGHPLRIBIGPRDVSSNSVVVSRRDYFGCAGYDFGHSMEBSTLVAYVKERLDEICTSLEBALSEFRDSNIVDVNSYABLTEAISGKWARGPWSASQAEBDVKEETQ-------ATIR
ZmProRS-Org  334  LRTPGWKFNETEMKGVPWTRIBIGPRDVSNSVVVSRRDYFGKKGKBFGVSMEBSILVDBIIGELBLTCALLQFAIIFRDSNIVDYSSYBELKBAISEGKWARGPWSASACEBLKVKEEBS------ATIR
PpProRS      284  QRTPGWKFNEWEMKGVPLRTBTVGPRDVASGSVVVSRRDVFGKIGKBFGVSMEBPLVEAEVANHFEBYCASLLQFAFIDFRDSNIVDVSSYBBLKACAISKGKWARGPWPASCAEBLKVKEEBS------ATIR
```

```
AtProRS-Cy   499  CSPBLQPELBBGTLCBASCIPAKKNTYWGRSY
ZmProRS-Cy   477  CTPPEQPBLBBGTLCBASCTPAKKWSPWGRSY
AtProRS-Org  514  CBPFEQTCGTK--TCLMTGNPAEBVAIFARSY
ZmProRS-Org  459  CVPFEQPEGTK--TCPMTGNPAEBVAIPARSY
PpProRS      409  CBPFEQPFGPK--TCPMTGNPAEBVAIPARSY
```

Figure 34. SDS-PAGE of PpProRS protein followed by staining with Coomassie Brilliant Blue

Parallel cultures were grown and sampled prior to induction of gene expression with IPTG (lane 1), and after induction for 3 hours with 1 mM IPTG (lane 2). PpProRS was purified by Ni-affinity chromatography, and the purified protein was analyzed (lane 3). Fifteen µg protein was loaded in each lane. Kaleidoscope pre-stained standards (Bio-Rad) were loaded into the lane labeled M.
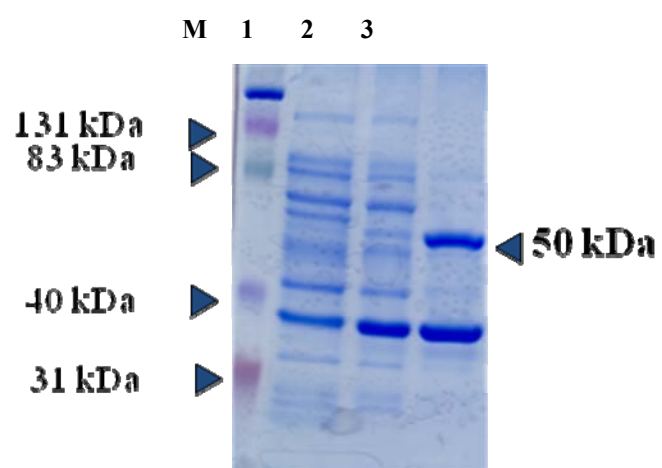
Figure 35. Prolyl-AMP synthesis of PpProRS as measured in the ATP-PP$_i$ exchange reaction

The rate of Pro-dependent PP$_i$ exchange activity (●) is shown as a function of time. The Pro concentration used to measure the enzyme dependent activity was 0.5 mM. The amount of enzyme added to the reaction shown was 1 μg.
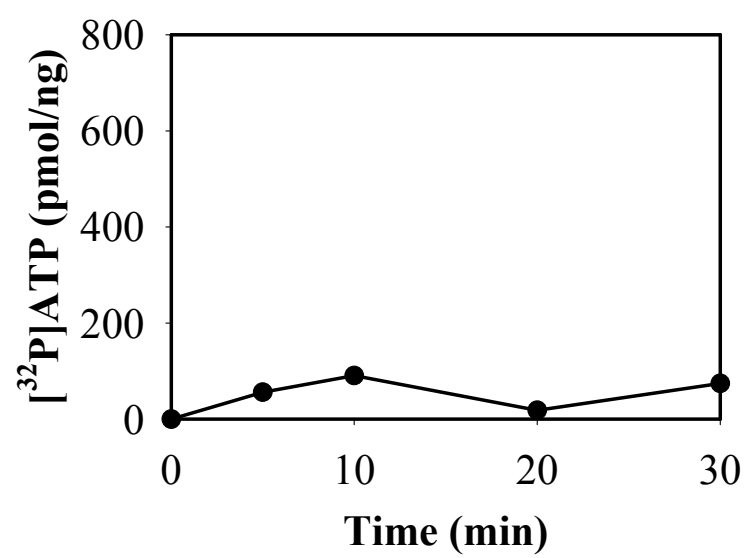
Figure A1. Functional complementation of *cysH56* mutant strain, JM96

JM96 strain transformed with CmAPR or PpAPRs was grown in minimal agar medium with (A) or without Cys (B) at 30°C for 2 days. Untransformed strains were used as negative control. Cloned genes are indicated as Pp2 (PpAPR2), Pp4 (PpAPR4), Pp5 (PpAPR5) and Cm10 (CmAPR).
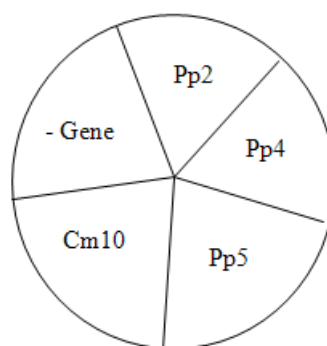
(A)
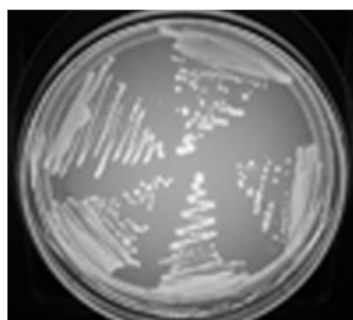
(B)
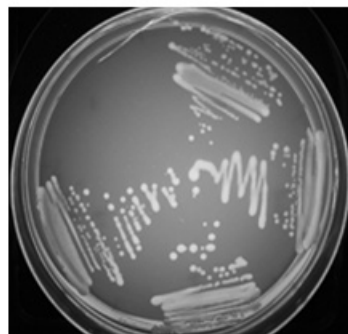
Figure A2. **Alignment of deduced amino acid sequence of PpAPRs and CmAPR**

Sequences were aligned with the program ClustalW. Black boxes indicate identical amino acid residues at that position. The conservatives are shaded in gray boxes. Conserved motifs with plant APR are indicated with dots and asterisks. Open circles represents the conserved motifs of all APRs. The square indicates residues conserved with Trx and Grx including the active site sequence.

```
CmAPSR     1  MASATAAVSGSIPSHSLLHRDLKAAQIGNFVTGVQFRESSLRKRWAVHVVEPGKRNDSAVAAAAASAVAE
PpAPSR2    1  -------MRWSPASG-----------T------TRWLPRRPRRRWRS----------------------
PpAPSR4    1  --------------------------------------------------------------------
PpAPsR5    1  MASATAAISGSISSHSLLDRDLKSAQIGNLATGVKFRESSLRKRWAVNVMEPGKRND---SVAAASAVAD

CmAPSR    71  LAAVVEEKTRAPEKEVEVVDYEKMAEELQNASPLEIMDRALAEFGNDIAIAFSGAEDVALIRYARLTGRP
PpAPSR2   24  -PAVVEEKARAPDPEEEVVDYEKMAEELQNASPLEIMDRALAEFGNDIAIAFSGAEDVALIRYARLTGRP
PpAPSR4    1  -----------------------------------MDRALAEFGNDIAIAFSGAEDVALIRYARLTGRP
PpAPsR5   68  LAAVVEEKARALEKEVEVADYEKTAVELQNASPLEIMDRALAEFGNEIAIAFSGAEDVALIRYARLTGRP
                                                                              ..
CmAPSR   141  PRVFSLDTGRLNPETYRYPNEVETHYNIHIEYMFPDAVEVQALVRNKGLFSFYEDGHQECCRVRKVRPLR
PpAPSR2   93  PRVFSLDTGRLNPETYRYPNEVEKHYNIHIEYMFPDAVEVQALVRNKGLFSFYEDGHQECCRVRKVRPLR
PpAPSR4   35  PRVFSLDTGRLNPETYRYPNEVEKHYNIHIEYMFPDAVEVQALVRNKGLFSFYEDGHQECCRVRKVRPLR
PpAPsR5  138  PRVFSLDTGRLNPETYRYPDEVERHYNIHIEYMFPDAVEVQALVRTKGLFSFYEDGHQECCRVRKVRPLR

CmAPSR   211  RALKGLRAWITGQRKDQSPGTRAHIPVVQVDPSFEGLDGGAASLIKWNPVANVDGKDVWNFLRTLEVPVN
PpAPSR2  163  RALKGLRAWVTGQRKDQSPGTRAHIPVVQVDPTFEGSDGGAGSLIKWNPVANVDGKDVWNFLRTMEVPVN
PpAPSR4  105  RALKGLRAWVTGQRKDQSPGTRAHIPVVQVDPTFEGSDGGAGSLIKWNPVANVDGKDVWNFLRTMEVPVN
PpAPsR5  208  RALKGLRAWITGQRKDQSPGTRAHIPVVQVDPSFEGLDGGAGSLIKWNPVANVDGKDVWNFLRTMEVPVN
                    ********                               ooooooo
CmAPSR   281  SLHAKGYVSIGCEPCTRPVLPGQHEREGRWWWEDAKAKECGLHKGNISQD---EAGVNGNG-AAAMNSVN
PpAPSR2  233  SLHAKGYVSIGCEPCTRPVLPGQHEREGRWWWEDAKAKECGLHKGNIAQDESQKAGSDGNGDGVAMSSVN
PpAPSR4  175  SLHAKGYVSIGCEPCTRPVLPGQHEREGRWWWEDAKAKECGLHKGNIAHDESQKAGSDGNGDGVAMSSVN
PpAPsR5  278  SLHTQGYVSIGCEPCTRPVLPGQHEREGRWWWEDAKAKECGLHKGNISEDELQKAGVNGNG-AASMNSVN
                                                            ■■■■■■■■
CmAPSR   347  GTTDIFKSQAIVNLSRPGIENLLKLENRQDPWVVVLYAPWCRFCQAMEGSYVELAEKLSGSGIKVAKFRA
PpAPSR2  303  GTTDIFKSQAIVNLSRPGIENLLKLENRQDPWLVVLYAPWCRFCQAMEGSYIELAEKLSLSGVKVAKFRA
PpAPSR4  245  GTTDIFKSQAIVNLSRPGIENLLKLENRQDPWLVVLYAPWCRFCQAMEGSYIELAEKLSLSGVKVAKFRA
PpAPsR5  347  GTTDIFKSQDVVSLSRPGIENLLKLENRQDPWLVVLYAPWCRFCQAMEGSYVELAEKFSGSGIKVAKFRA

CmAPSR   417  DGDQKPFAQQELQLGSFPTILFPPKHASRPIKYPSEKRDVDSLQAFINALR
PpAPSR2  373  DGDEKPFAQQELQLGSFPTILFPPKHASTHIKYPSEKRDVDSLLTFVNALR
PpAPSR4  315  DGDEKPFAQQELQLGSFPTILFPPKHASTHIKYPSEKRDVDSLLTFVNALR
PpAPsR5  417  DGDQKPFAQQELQLGSFPTILFPPRRASRPIKYPSEKRDVDSLLAFIDALR
```

Figure A3. **Phylogenetic analysis of APR proteins from various organisms**

The sequence alignments were performed using ClustalW. Phylogenetic trees were constructed using Bootstrapped Neighbor-Joining method and evolutionary distances were calculated using Mega4.
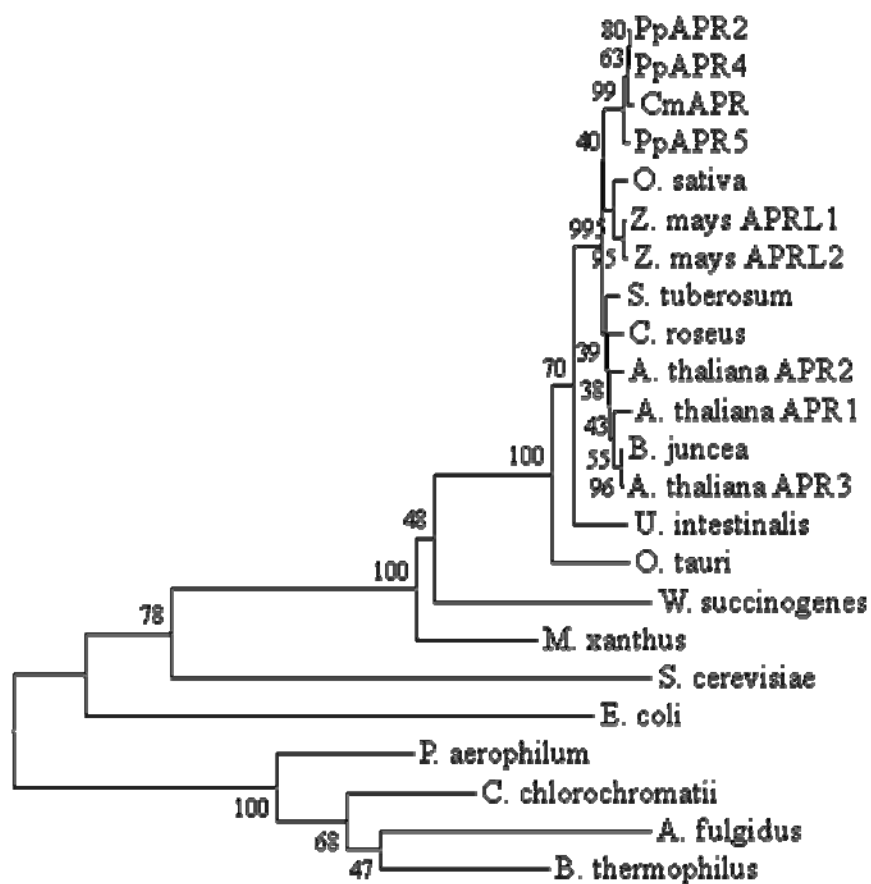
Figure A4. Functional complementation of *metB* mutant strain, LE392

LE392 strain transformed with CmCGS and PpCGS was grown in the minimal agar plates with (A) and without methionine (B) at 30°C for 2 days. Untransformed strains were used as negative control. PpCGS gene was indicated as Pp2.
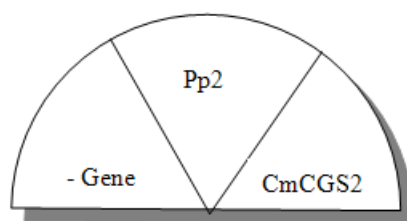
(A)
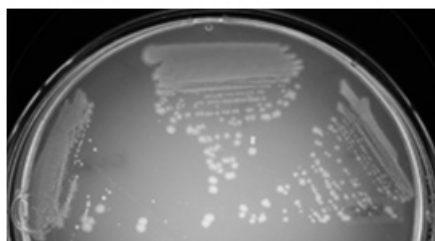
(B)
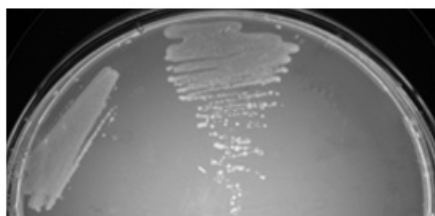
Figure A5. Nucleotide and deduced amino acid sequence of *P. pubescens* cDNA encoding CGS

The length of genomic CGS DNA is 5149 bp. The cDNA of coding region is 1674 bp in length. Lower case letters indicates intron sequences and exon is shown in upper case letters. Number on the left represent nucleotide position and that on the right shows amino acid position. It has a consensus translational initiation (**ATG**) and termination (**TAA**) sequence.

```
1         ATGGCCGTCTCATCCGCTTGCTTCCACTGTAACCCTAACCCTAATTTCTCCGCCCGCTCC
          M  A  V  S  S  A  C  F  H  C  N  P  N  P  N  F  S  A  R  S
61        CCATTTGCCGGCTCCGCCTCCGCCGCCGCCGGCCGTTCTAAACCCCAATCCGCCCCCAAA      20
          P  F  A  G  S  A  S  A  A  A  G  R  S  K  P  Q  S  A  P  K
121       CCCTACTCCTTCCTCAACAAGAAGGCTTCGGGATCGGGATCTTTCCCCAACTATGGCGGC      40
          P  Y  S  F  L  N  K  K  A  S  G  S  G  S  F  P  N  Y  G  G
181       CTCTCCTCGGCCCTCATCCTCAAGTTCCCGCCCAACTTCGTCCGCCAGCTCAGCAACAAA      60
          L  S  S  A  L  I  L  K  F  P  P  N  F  V  R  Q  L  S  N  K
241       GCCCGCCGGAACTGCGGCAACATCGGCGTCGCGCAGGTCGTCGCGGCTTCGTGGACCAAC      80
          A  R  R  N  C  G  N  I  G  V  A  Q  V  V  A  A  S  W  T  N
301       TCCCCGGCCAACGATGAGACCCCCTTCTCCTCCTCCGCTGCTGCGGCCGCTGCGGCGGTC      100
          S  P  A  N  D  E  T  P  F  S  S  S  A  A  A  A  A  A  A  V
361       GACGGTGCTGCGGCCGGTTGTGAACCCCGATGCGGCGTCGGAGGAGGAGGCCGCTGTCGCC      120
          D  G  A  A  A  V  V  N  P  D  A  A  S  E  E  E  A  A  V  A
421       GTCGCTGGAAGCGGCGACGGCGAGGTTGTAGGTTGTGTACAGGGTCGGATCTCGCCGGTT      140
          V  A  G  S  G  D  G  E  V  V  G  C  V  Q  G  R  I  S  P  V
481       GGCGGGGTTGGATGCGGTGTTTAGCTCCGATGCGAGCCTTTTGGTTCATGCGGgtaaagtt      160
          G  G  L  D  A  V  F  S  S  D  A  S  L  L  V  H  A
541       ttggacctgagatattggataaaaaaacccatctttttcccctttaaattaattttttccc      177
601       cccttaaataaatgagaatggatctaatggctgaaatagGCGAGAGATATGGGCGCGGCA
          G  E  R  Y  G  R  G
661       TAAAGACGGATACGATCACCACGCCTGTTTTCAACACTTCTGCTTACTGGTTCAAGGACT      184
          I  K  T  D  T  I  T  T  P  V  F  N  T  S  A  Y  W  F  K  D
721       GCAATGAGCTGATTGATTTTAAGgtaaaaaaatcatctttttgaaataaagaattcgctat      204
          C  N  E  L  I  D  F  K
781       tgaaatcattttcaaaaaaaaaaaattaatcttttggtattaatacagGAAGGGAGGCAT      212
          E  G  R  H
841       GCGAGCTTTGAGTACGGGCGATACGGGAATCCGACTACCGAGGTATTGGAGAAGAAGATG      216
          A  S  F  E  Y  G  R  Y  G  N  P  T  T  E  V  L  E  K  K  M
901       AGgtgtgtgcgctctctctcttttgtgtattcttttgttgttaggcttttttaactgatgt      236
          S
961       ttattgattgtggtaatggtttgtgttttgatcgcagTGCGCTGGAGAGAGCTGAAACTA      237
          A  L  E  R  A  E  T
1021      CTTTGTTCGTGTCATCTGGCATGTACGCCAGCTGCGCGGTGCTTTCAGCCTTGGTTCCTG      244
          T  L  F  V  S  S  G  M  Y  A  S  C  A  V  L  S  A  L  V  P
1081      CTGGTGGGCATATTGTTCCACCACCGATTGCTACAGGCCGGACCAGGATGTTCATTGAGA      264
          A  G  G  H  I  V  P  P  P  I  A  T  G  R  T  R  M  F  I  E
1141      ATGAGCTCCCAAAGAAGGGAATATCGgtaatttcttgctgtttcgtcagatttataacat      284
          N  E  L  P  K  K  G  I  S
1201      acaaattcttgcagcattcttttaccaagtgctgataagcgacgttttctaaaattttga      293
1261      cctttcaggtcaagcaaatacaatttaagttcataatcaatttttacctgataccagaga
1321      agttttattttctagttaattaatacacatatgattgccaatgttacgtgggaatttga
1381      atattacaattgtgactttccttaggatgttcttgtacatatatataaataatgaccttg
1441      tcacttaatatgaaaaatattgtataaactatgctatctggttgctgaatttggttgacg
1501      ggaactaatttgtacacagttactacaacttcttatcaaatatcttgaaaaatagtagt
1561      tttcagttggcaatactgacttggaagtgtttttggtctgactaaataacttgtatttgac
1621      caagtgttttaaacatatgcataactcttgaattgatgtggctgaccctctgtttgacta
1681      attaaattttgtaatcttttgattttgttcgtgagccattctagtaagttgttactgtc
1741      ccaggaaaactgatttacaacttaaaggtttaaatactaagaaagagcaaacatatggag
1801      tattatattctacttcatggatatccaggattttttttttgcaaaaaatatctagtcatat
1861      ctccaagcatcatgctagaatttcgcttacgttacgttttggacaactcattagcatatt
1921      tacccaccttttttttggcttttctatttagtctttcttaacttagaattgaactgaaca
1981      gaagaaaaggattcatgtagccgctcccaaaatttgaaatgaagactttgttgttgttg
2041      tttaaattttctttgtcacgggcccacaatgctaatacatgtatttgcgaacatgtcaatt
2101      gtgtaatattttgagtagatctatctatctagggactgatcttgaaatcacttttaaaat
2161      acaaatagcgattgaaaggtttgtgctacctttgctcacaaagtgatgaaattttgaaag
2221      aatttgaattggctctacaagaatgttcttgtaagtatctcctttttttctcaaatagtaaa
2281      gtggaatagttagtagcttcactacaatttaatattttgctacaagagctgtatcttgtc
2341      ttttatgtaatgaaccgccatgttttttgtagaaaatctcactgaatgatctattacgttc
2401      atcacctatttaatttagatggatcactagtgcatacaccattttctatatataataagat
2461      taatcgtccttcattgtaatgactaaacctttagatctgaaatgtttgtttgtgtgtgt
2521      tactcatatttcactatcttcaaattgtagGCCACTGTCATTGATCCTGCTGATATGGAC
          A  T  V  I  D  P  A  D  M  D
2581      TCACTTGAAAGAGCATTGGATCAACATAAGgtcagtgcacaccattgctttgttggtctc      303
          S  L  E  R  A  L  D  Q  H  K
2641      tccctttgttgcgttcactttttattggtactcttactattattaattccacaatgtaat      313
2701      ccttaatggtagttgggaaagggaaagcgtgtcattttttttttttgtgtaggcgggtggtg
2761      ttctgcctgagcttatatgcacatattcatttattttacatggctctttttgtgcctgata
2821      cattaagtttactcagGTGTCTCTTTTCTTCACTGAGTCTCCAACTAATCCCTTCCTCCG
          V  S  L  F  F  T  E  S  P  T  N  P  F  L  R      328
```

```
2881    ATGCATCGATGTCAAGTTTGTTTCAGAGCTTTGCCACAGTCATGGAGCTCTGGTTTGCAT
         C  I  D  V  K  F  V  S  E  L  C  H  S  H  G  A  L  V  C  I      348
2941    CGATGGTACTTTTGCCCCACCTGTTAATCAGAAACCCTTAACTATGGGGGCTGATCTAGT
         D  G  T  F  A  P  P  V  N  Q  K  P  L  T  M  G  A  D  L  V       368
3001    CGTTCATTCTGCAACCAAGTTCATTGCTGGACACAACGATgtgagttacctttgcaactt
         V  H  S  A  T  K  F  I  A  G  H  N  D                           381
3061    tacttactggaaaaatatgggggttagtgttaatttggcatgtttaagttaacctgattga
3121    attactattattttgaagatgtattcgatttttgttttcagGTTATTGCAGGTTGTGTCA
                                                   V  I  A  G  C  V       387
3181    GTGGTCCAGAGGAATTGATTTCTAGAGTTCGCCAGTATCATCATATTATTGGTGGTGTTC
         S  G  P  E  E  L  I  S  R  V  R  Q  Y  H  H  I  I  G  G  V       408
3241    TTGATCCTgtaagtaatacctttttataatgtttgacgatttcaatagtgtctgtcaaagt
         L  D  P                                                         410
3301    tgttgcatgccttaaaggtagcccttatcttcatgtatctgtgtatctgtctatgttgat
3361    ggcatgtatgtacctctgtgttgatctgattatcagctttccattacactgaaaatcagt
3421    tgacaacatcatctgtcaagaacatcaccggtatcatgttattggtgatgttcttaatta
3481    tttaagtattcgtcttacattaattgataattcccataatggctgtaaatgctgattgta
3541    actgttaagagtatctcttaataatctctgtatctgattattagcttctccattatgctt
3601    aagtaaatcaattgacaaaatcatctttacagAATGCTGCTTACATGATCTCCCGGGGTA
                                         N  A  A  Y  M  I  S  R  G        419
3661    TGAAGACACTGCATCTTCGTGTACAGTATCAGAATAGCATAGCATTGAGGATGGCCCAGT
         M  K  T  L  H  L  R  V  Q  Y  Q  N  S  I  A  L  R  M  A  Q       439
3721    TTTTAGAGGAACATCCCAAGgtacattaacctcccccacctcccccccgcgcgcgcgcgcg
         F  L  E  E  H  P  K                                             446
3781    cacacacacacacactgtggaagatgttctggtaagttgtacaagtcaacaaactcattt
3841    ttcttagcggggggagaaaataattagagaatacgagttaaataacctgaggtccccagc
3901    tttgggtctcatttctatatgagctaacgatcaagtaggagctgttcctgaatttttttt
3961    atatgctagtccatattaatattagtccatgatcattctgatctctcgacgggttttaaa
4021    ttcttctcatattgtagtattcttagtgttgtttcttggctttccccccagcctatgagt
4081    gtaggttaccaggatgccaacctttttgaattatgagaagtctatttcttgtctatacttc
4141    ctgtggccaagaagttggccttaattttaggtgaggagtactgtcagtatgatgccgtttg
4201    agattttgtgaaattttttcatggagcatttttaccccatgtatcctatcccattttaatgg
4261    ccaatgacttgggggatgatgatgatatggtgacgctgcctcataccgactagaaagagta
4321    ggtggacaacataacatcacaacaattgtaatttcgcctctgtaacgaaactgttcacaa
4381    tattctatcttttgttccaaaaaaatgatcatgtaatatttgttgcgatcatatttctgt
4441    tcttctaatagagttttgtgatgttatattgccagATTATACGAGTATATTATCCAGGCC
                                           I  I  R  V  Y  Y  P  G         454
4501    TTCCGAGTCATCCGGAGCATCACATTGCCAAGAAACAAATGACTGGTTTTGGTGGTGTGG
         L  P  S  H  P  E  H  H  I  A  K  K  Q  M  T  G  F  G  G  V       474
4561    TTAGTTTTGAGgtgagaaatgtttatccattcattgccattatttactattagttcccag
         V  S  F  E                                                      478
4621    ggaggcccctttagcttgtgttgctgctttagtacgtatttgtatcatagaggattgggtg
4681    catatatcatcagagtaggttagttttaccgatattgttcgtattgagagattttgagct
4741    tatattgtttgatgtatttggtggtgtagCTTTCCGGAGGCTTGTTGGAGACCATAAAGT
                                     L  S  G  G  L  L  E  T  I  K         488
4801    TCATTGATTCTTTGAAGATACCATACCTTGCCCCTTCTTTTGGAGGTTGCGAGAGTATAG
         F  I  D  S  L  K  I  P  Y  L  A  P  S  F  G  G  C  E  S  I       508
4861    TTGACCCACCGGCCATTATGTCTTACTGgtaatgttcccgacttgtatatacatatttga
         V  D  P  P  A  I  M  S  Y  W                                    518
4921    tattatttctgtttaatgccatctttgccagtcatttatagcataaggtaactacacta

4981    agcaatcaaggtcctctgacaagggtgctttcttaatgaatggaacagGGATAGCGTGTC
                                                         D  S  V  S       522
5041    GGAGAGAGCTAAGTACGGCATAGGTGACAACCTGATCAGGTTCAGTTTTGGATTAGAGGA
         E  R  A  K  Y  G  I  G  D  N  L  I  R  F  S  F  G  L  E  D       542
5101    TTTTGAGGACGTGAAGGCTGATATCGCTAAAGCTCTGGAAGCTATCTAA
         F  E  D  V  K  A  D  I  A  K  A  L  E  A  I  *                   548
```

Figure A6. **Phylogenetic analysis of CGS proteins from various organisms**

The sequence alignments were performed using ClustalW. Phylogenetic trees were constructed using Bootstrapped Neighbor-Joining method and evolutionary distances were calculated using Mega4.
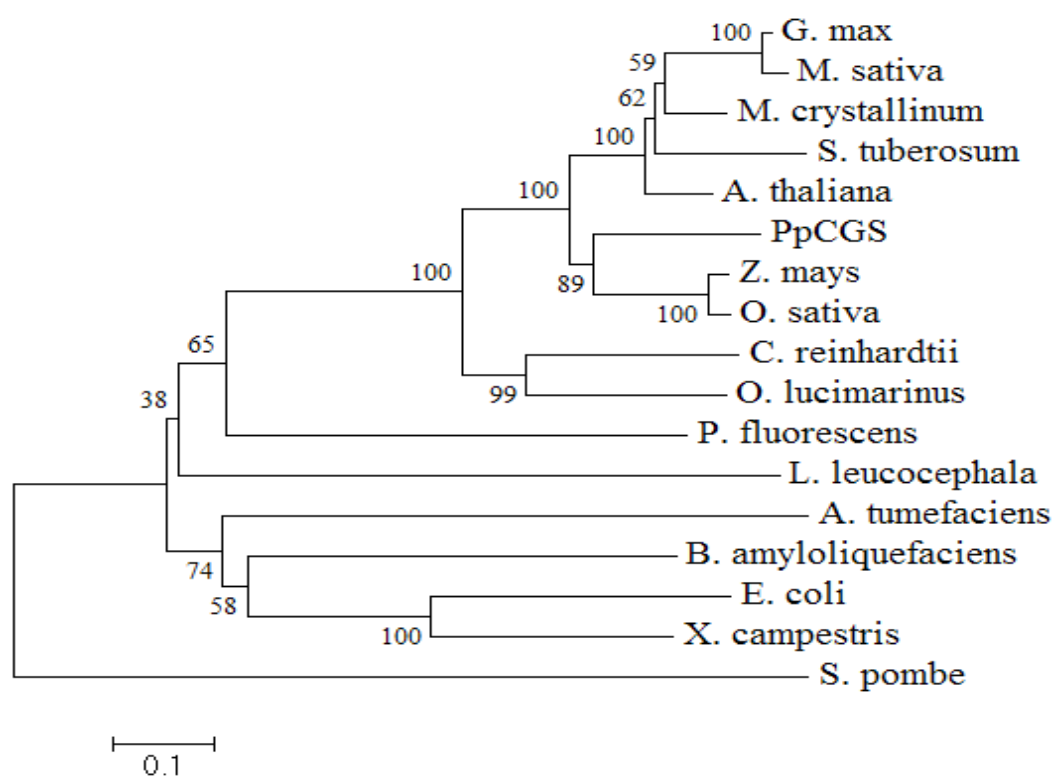
Figure A7. Alignment of CGS amino acid sequence from various plant species

Sequences were aligned with the program ClustalW. Black boxes indicate identical amino acid residues at that position. The conservatives are shaded in gray boxes. MTO region are marked with asterisks. At, Arabidopsis; Cm, Convallaria; Gm, soybean; Mc, common ice plant; Ms, Medicago; Os, rice; Pp, Polygonatum; St, solanum; Zm, maize,

```
                                                                                    **********
CmCGS     1  -----------------------------------------------------GRG---------------------------ANSPANDET
StCGS     1  ---MAVSSYAR--AFPS------FECRSEP-DFSGSLPHPKAGVRLSGKYNSRENRSQV---YGL-SSLIFFFPPPNFVRQLSIKARRNCSNIGVACVVVAASWSNNQAG--P
AtCGS     1  ---MAVSSPQCPTIFSSSSISGFQCRSDP-DLVG-SPVGGSSRRRVHASAGISSSFTGD--AGL-SSRILRFPPPNFVRQLSTKARRNCSNIGVAQIVAAAWSNNPSSALP
GmCGS     1  ---MAVSSSHM--RFT------FECRSDP-DFS--PPPPSFDNLRRRNFRSSADSGAAF-HGI-SSLILRFPPPNFVRQLSTKARRNCSNIGVAQIVAASWSNNSDN---
McCGS     1  MAATTSSCCAR--VLPP------FECRPEPADFSTAAAQEKNIPKS-PFGGANGRLFS---SSS-SSLILRFPPPNFVRQLSTKARRNCSNIGVAQVVAASWSNNSDAGAT
MsCGS     1  ---MAVSSPHR--VFT------FECRSDP-DFTSLPSTDNHHHNRRRHFPTTSLSSAA----AI-SSPILRFPPPNFVRQLSTKARRNCSNIGVAQIVAASWSNEGTG---
OsCGS     1  ---MATVSSLP------------------------S--P-A-----FLAADPAAALP-------SATILRFPPPNFVRQLSTKARRNCSNIGVAQIA------------
PpCGS     1  ---MAVSSACFHCNPNPNFSARSPFAGSASAAAGRSKPQSAPKPYSFLNKKASGSGSFPPNYGGLSSALILMFPPPNFVRQLSTKARRNCSNIGVAQVVAASWTNSPANDET
ZmCGS     1  ---MATVS-LT------------------------P--Q-A-----VFSTESGGALA-----SATILRFPPPNFVRQLSTKARRNCSNIGVAQIVAAAWSDCPAAR-P


CmCGS    13  PFSS---ANTAAVDGALVVVSPDAABEE---VAVASDGHGGVVDGVPSRISPACG-----------LDAVFASDASLLVHAGEIYGTRLKIDSLTTPVVTNSAYNFKDC
StCGS    94  EPT----PAANAVDSSASALAVTSIGITAGDEEVAVVEN--ADCNDQNEQITGTCV-----------KYASFLNSDGSVATHAGERLGRGIVTDAITTPVVNTSAYFFTKL
AtCGS   103  SAAAA---ANTSSASAVSSAASAAAASSAAAAPVAAAPPVVLKSVDEEVVVAEEGIRRKIGSVQLTDSKHSFLSSDGSLTVHAGERLGRGIVTDAITTPVVNTSAYFFKKT
GmCGS    91  S------PAAGAPAPPAATATDAATVPLPVVVAANEDVVVSAAADENGAVQLNSS-----------SYSSFLKSDASKTIHAAERLGRGIETDGITTPVVNTSAYFFKKT
McCGS    99  SCGGV--PAATAVDAAAIAAVDHVEIDGGAVVEDSNDGGVKELSSNNSNTLGLCD-----------PRPPFLSSDGSVAIHAGERLGRGIVTDAITTPVVNTSAYFFKKT
MsCGS    91  N------PSGIAGSDSAGCCICCXCNDPAGAVRPRRRCRRWDVVENGAVQTNFS-----------SYSSFLKSDASKTIHAAERLGRAIVTDGITTPVVNTSAYFFKKT
OsCGS    55  -------------------------------------------LLGSDASLAVHAGERRIATDAITTPVVNTSAYWFNNS
PpCGS   108  PFSSSAAAAAAAAVDGALAVVNPDAASEBEAAVAVAGSGDGBVVGCVQGRISPVCG-----------LDAVFSSDASLLVHAGERYGRCIKTDLITTPVFNTSAYWFKDC
ZmCGS    66  HLGGGG-RFARGVASSHAALASAAAAASAAAADVSAIPN--AKVAQPSAVVLAEKN-----------LLGSDASLAVHAGERLGRFIATDAITTPVVNTSAYWFNNS


CmCGS   105  NELIDFKEGRH-AGFEYGTYGNPTTEVLEQKMSALERAESTLFVSSGMYASCAVLSALAPAGGHIVTTTDCYERDQGCSLRMSSPRKGISAITVIDPADMDSTERALDQHK
StCGS   187  PDLIDFKERKTCKVLNLEARYGNPTTVVLEEKISALEGAESTLIVASGMCASDVMFLALDAAGGHIVTTTDCYRKTT-VFIETNILPKMGITAATVIDPADMGALELTINQKK
AtCGS   211  AELIDFKENRQ-VSFEYGRYGNPTTVVLEDKISALEGAESTLVAASGMCASTVMLLALVPAGGHIVTTTDCYRKTR-IFMENFLPKLGITVTVIDPADIACLEAAVNEFK
GmCGS   184  ADLIDFKENRQ-VSFEYGRYGNPTTVVLEDKISALEGAESTVIMASGMCASVVLFMALVPAGGHIVTTTDCYRKTR-IFIETFLPKMGITTTVIDPADVGALESALEQHN
McCGS   196  ADLIDFKEGRQ-GSFEYGRYGNPTSVVLEDKISALEGAESTILLAASGMCASILLMLALVPAGGHIVTTTDCYRKTR-IFIETFLPKMGITATVIDPADVEGLIKSALDENN
MsCGS   184  ADLIDFKEKRQ-TSFEYGRYGNPTSTVLEEKISBLEGAESTILMASGMCASVVLFMALVPAGGHLVTTTDCYRKTR-IFIETFLPKMGITTTVIDPADVGALESALBCHN
OsCGS    95  QELIDFKENRH-AGFEYGRYGNPTTEALEEKKMSALEKAESTVFVASGMTASVAMLSALVPAGGHVVTTTDCYRKTR-ILMETELPKRGITMTVIDPADMDALQNALDNNN
PpCGS   206  NELIDFKEGRH-AGFEYGRYGNPTTBVLEBKMSALERAETLLFVSSGMYASCAVLSALVPAGGHIVPPPIAIGETR-MFIENELPKKGISATVIDPADMDSLERALDQHK
ZmCGS   158  QBLIDFKEGRH-AGFEYGRYGNPTTBALEKKMSALEKAESTVFVASGMYAAVAMLSALVPAGGHIVTTTDCYRKTR-IIMENELPKRGISMTVIPPADMDALQNALDNNN


CmCGS   214  VSLFFTESPTNPFLRCIDVKFVSELCHSKGALVCIDGTFASPVNQKPDTMGADLVVHSATKFIAGHNDVIAGCVSGSEELVSRVPQYHEIIGGVLDPNAAYMISPGMKTL
StCGS   296  VDLFFTESPTNPFLRCVDIELVSNLCFEKGALVCIDGTFATPLNQKALALGADLVVHSATKFTGGHNDVLAGCISGPERLVSVTRNLHEIDGGALNPNAAYLAIRGRGLN
AtCGS   319  VSLFFTESPTNPFLRCVDIEVLSKICHKFGTLVCIDGTFATPLNQKALALGADLVVHSATKYIGGHNDVLAGCICGSLKLVSEIRNLHHVLGGTLNPNAAYLIIRGMKTL
GmCGS   292  VSLFFTESPTNPFLRCVDIKLVSELCHKKGILICIDGTFATPLNQKALALGADLIEHSLTKYMGGEHDVLGGCISGSIKVVSQIRTLHEVLGGTLNPNAAYLEIRGMKTL
McCGS   304  VTLFFTESPTNPFLRCVDIKLVSELCHKKGTLICIDGTFATPLNQKALALGADEVVHSMTKFIGGHNDVLGGCISGSIKVVSQIRTLHHVLGGTLNPNAAYLMIRGMKTL
MsCGS   292  VSLFFTESPTNPFLRCVDIKLVSELCHKKGILICIDGTFATPLNQKALALGADLIEHSLTKYIGGEHDVLGGCISGSIKVVSQIRTLHHVLGGTLNPNAAYLFIRGMKTL
OsCGS   203  VSLFFTETPTNPFLRCLDIDLVSKMCHSKGALLCIDSTFASPTNQKALTLGADLVVHSATKYIAGHNDVTGGCISGRDELVSKVPIYHEVVGGVLNPNAAYLIRGMKTL
PpCGS   314  VSLFFTESPTNPFLRCIDVKFVSELCHSHGALVCIDGTFABPVNQKPLTLGADLVVHSATKFIAGHNDVIAGCVSGPEBLISRVPQYHEIIGGVLDPNAAYWISRGMKTL
ZmCGS   266  VSLFFTEPTPTNPFLRCIDIEHVSNMCHSKGALLCIDSTFASPINQKALTLGADLVIHSATKYIAGHNDVIGGCVSGRDBLVSKVPIYHEVVGGVLNPNAAYLILRGMKTL


CmCGS   324  HLTVQYQNSIALRMAQFLEEHPKIIRVYYPGLPKSSGASHCQEANDWFGGVVSFELACGGLLETTKFIDSLRIPYLAPSFGGCESIVDPPAIMSYWD-SVSERAKYGIGDN
StCGS   406  HLRDSTKIQLHLRMAEILEAHPKVKHVYYPGLPSHPBYHLAKKQMTGFGGVVSFEVDGDPLTTAAKFVDALRIPKIAPSEGCKSIVDQPAIMSYWDLSQSURAKYGILDN
AtCGS   429  HLRVQQONSTAFRMAEILEAHPKVSHVYYPGLPSHPEHELAKRQMTGFGGVVSFEIDGDLETTIKFVDSLKIPYIAPSFGGCESIVDQPAIMSYWDLPQEERLKYGIKDN
GmCGS   402  HLRVQQONSTGMRMAKLLEAHPKVKTVYYPGLPSHPEHELAKRQMTGFGGVVSFEIDGDLHTTIKFIDSLKIPYIAMSFGGCISGSIVDWDLPSERAKYKIYDN
McCGS   414  HLRVQQONSTGMRMAEILEAHPKVKHVYYPGLPSHPEKLAKRQMTGFGGVVSFEVDGDPLTTAAKFVDSLKILAYIATSFGGCESIVDQPAIMSYWDLSPSDRLKYGIKDN
MsCGS   402  HLRVQQONSTGMRMAKLLEAHPKVKHVYYPGLPSHPEHELAMRQMTGFGGVVSFEIDGDLTTTIKFIDSLKIPYIAMSFGGCISGSKIPYIATSFGGCESIVDWDLPASERAKYKIYDN
OsCGS   313  HLRVQQONNTAMRMAQFLEEHPKIARVYYPGLPSHPEHLIAKSQMTGFGGVLSFEVAGDFDALREFIDSVKIPYHAPSFGGCESILDQPAIMSYWD-SKEQRETYGIKDN
PpCGS   424  HLRVQQONSIALRMAQFLEEHPKIIRVYYPGLPSHPEHHIAKKQMTGFGGVVSFELSGGLLETTKFIDSLKIPYLAPSFGGCESIVDPPAIMSYWD-SVSERAKYGIGDN
ZmCGS   376  HLRVQQONDTALRMAQFLEEHPKIARVYYPGLPSHPEHHIAKSQMTGFGGVVSFEVAGDFDAREKFIDSWKIPYHAPSFGGCESILDQPAIMSYWD-SKEQRDIYGIKDN


CmCGS   433  LIRFSFGLEGFEDVKADIAKALEAI-------------------------------------------
StCGS   516  LVRFSFGVEDFEDVKADVLQALDSIRLPHLQQFFFTWFHPDLVVRTCVITDFHHVANCLNVMKGSSSR
AtCGS   539  LVRFSFGVEDFEDVKADVLQALEAI-------------------------------------------
GmCGS   512  LVRFSFGVEDFEDLKADVLQALEAI-------------------------------------------
McCGS   524  LVRFSFGVEDFEDLKADVLQALEAI-------------------------------------------
MsCGS   512  LVRFSFGVEDFEDLKADVLQALEAI-------------------------------------------
OsCGS   422  LIRFSLGVEDFEDLKQDVVQALDEI-------------------------------------------
PpCGS   533  LIRFSFGLEDFEDVKADIAKALEAI-------------------------------------------
ZmCGS   485  LIRFSLGVEDFEDLKQDLVQALEKI-------------------------------------------
```

A8. Schematic of Cys and Met biosynthesis and relationship among metabolism of amino acids

Sulfate

APS   APR

Sulfite

Cys

Cystathionine   CGS   O-Phosphohomoserine

Homocysteine

Met

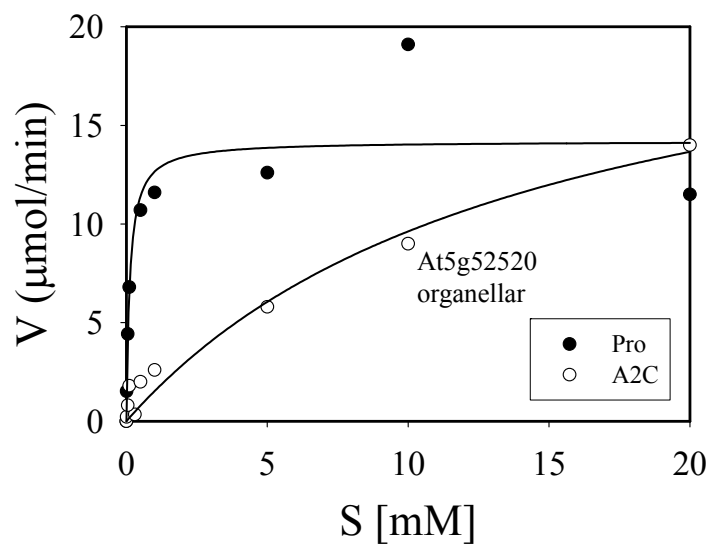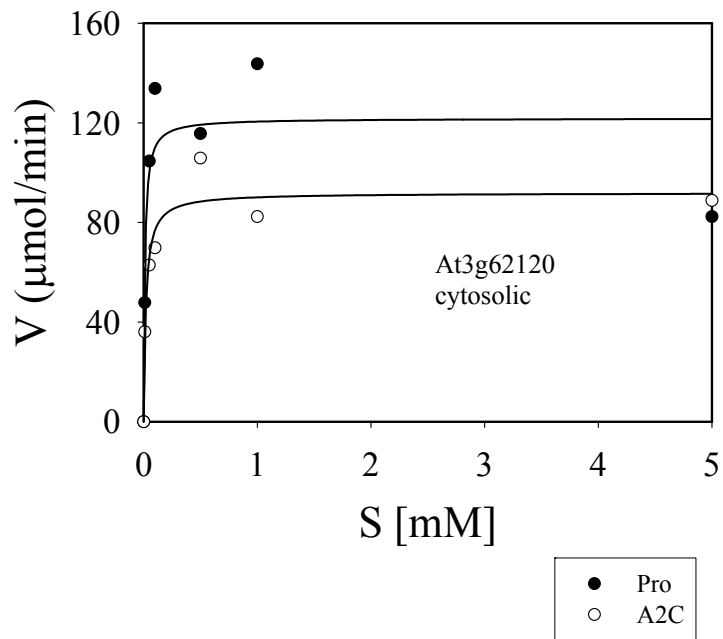SAM

?

A2C

NAS

NA

Asp

Thr

Ile
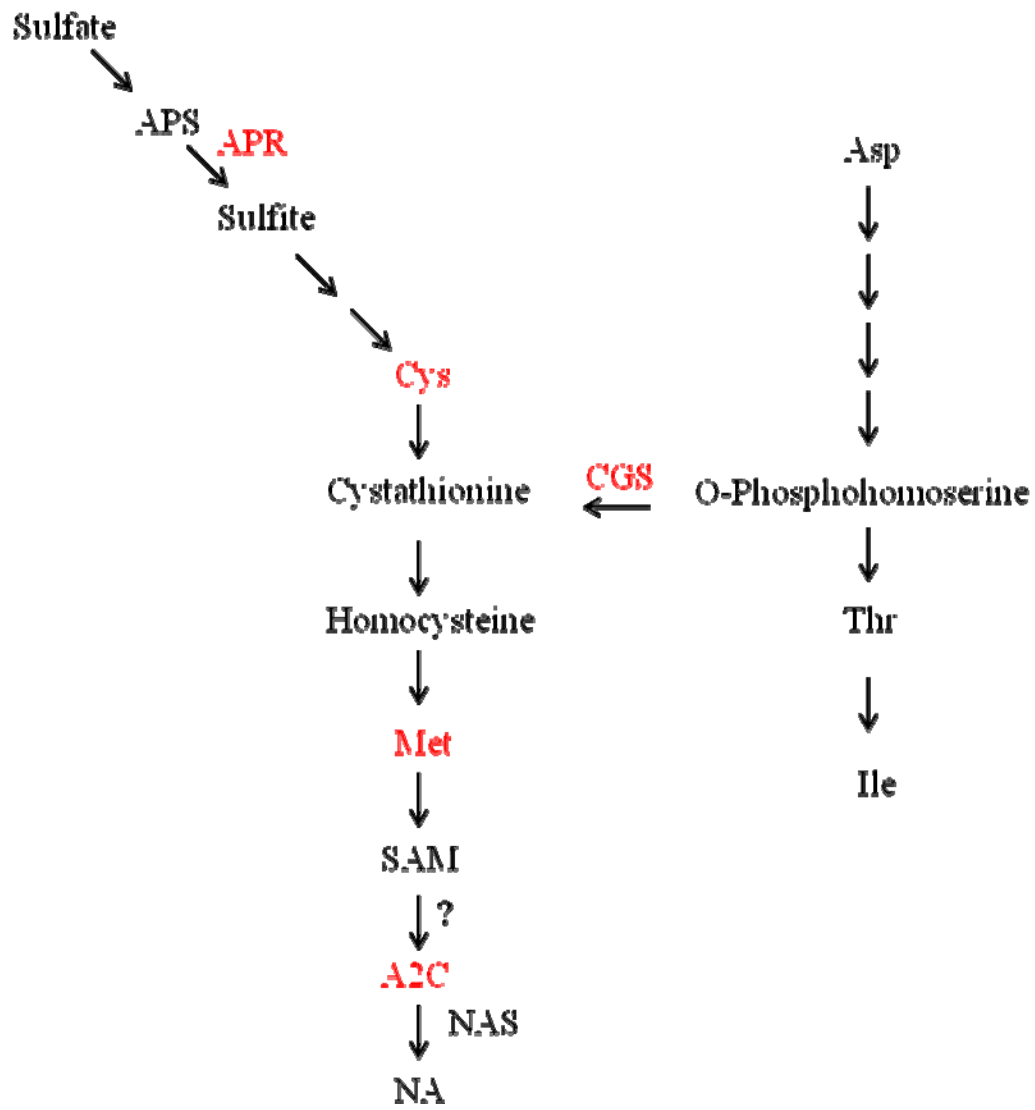
Appendix B

Figure B1. Kinetic analysis of AtProRS

Pro or A2C were titrated into a PPi exchange assay at 30°C. The saturation curves for Pro

(●) and A2C (○) are presented in graph A (At5g52520) and graph B (At3g62120).
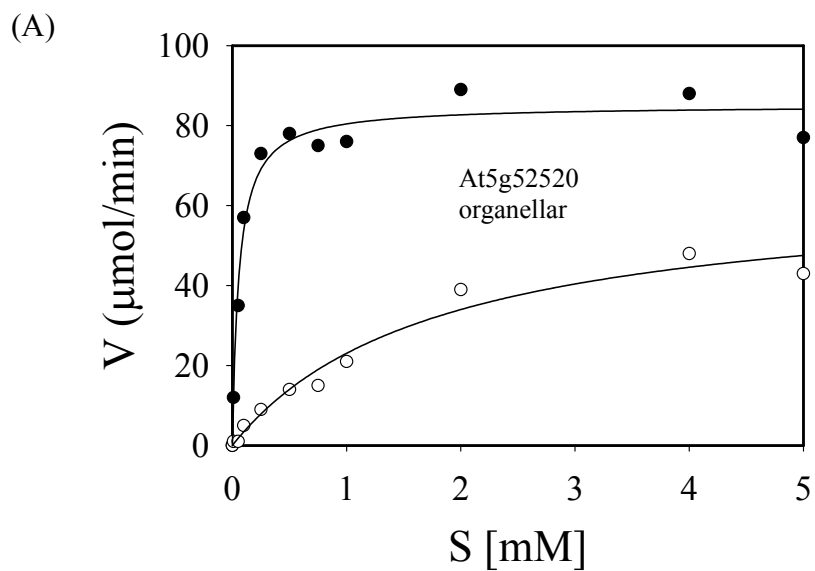
(A)

(B)

	At5g52520		At3g62120	
Amino acid	Proline	A2C	Proline	A2C
V_{max} (µmol min^{-1} mg protein^{-1})	14.20 (±1.24)	23.52 (±5.11)	121.77 (±12.99)	91.84 (±5.68)
K_m (mM)	0.12 (±0.06)	14.44 (±5.97)	0.01 (±0.01)	0.02 (±0.01)

Figure B2. Kinetic analysis of AtProRS

Pro or A2C were titrated into a PPi exchange assay at 30°C. The saturation curves for Pro

(●) and A2C (○) are presented in graph A (At5g52520) and graph B (At3g62120).

(A)

(B)
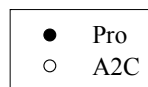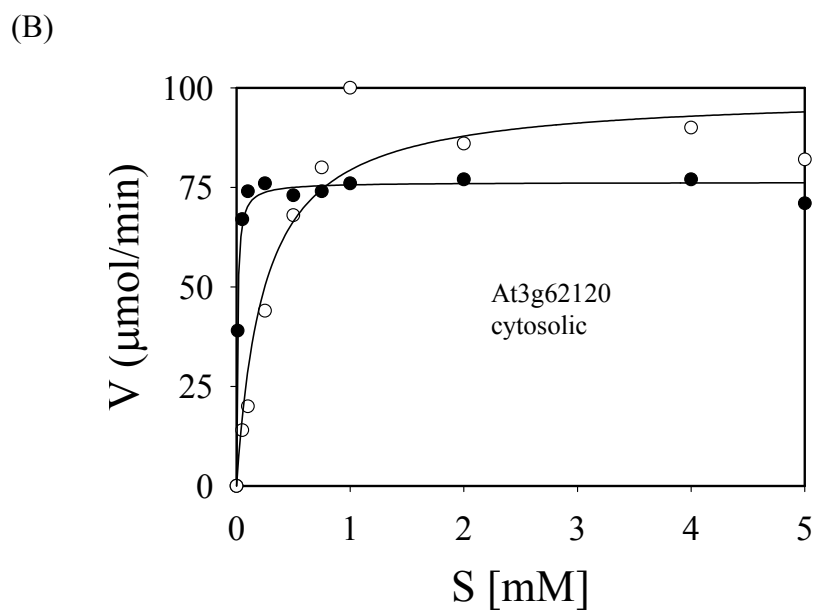
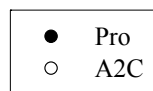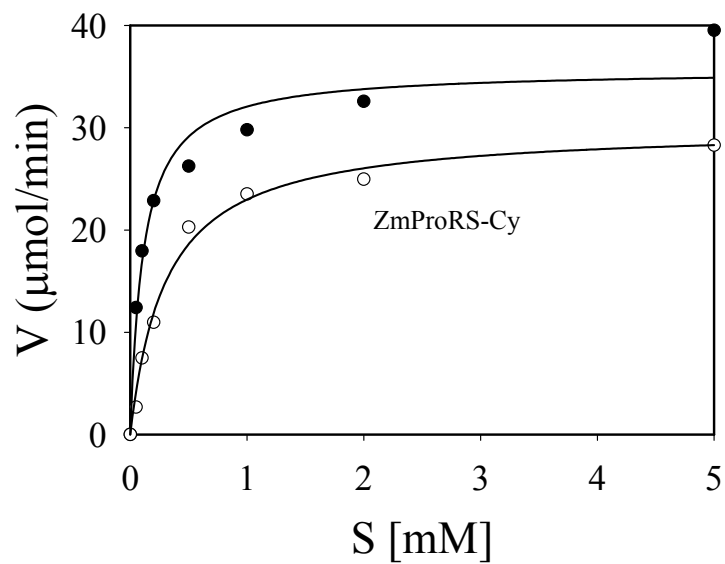|	At5g52520		At3g62120	
Amino acid	Proline	A2C	Proline	A2C
V_{max} (µmol min^{-1} mg protein^{-1})	85.08 (±2.23)	64.70 (±6.30)	85.70 (±1.76)	79.22 (±1.13)
K_m (mM)	0.06 (±0.01)	1.81 (±0.41)	0.07 (±0.01)	0.01 (±0.01)

Figure B3. Kinetic analysis of ZmProRS

Pro or A2C was titrated into a PPi exchange assay at 30°C. The saturation curves for Pro

(●) and A2C (○) are presented in graph A (ZmProRS-Org) and graph B (ZmProRS-Cy).

(A)
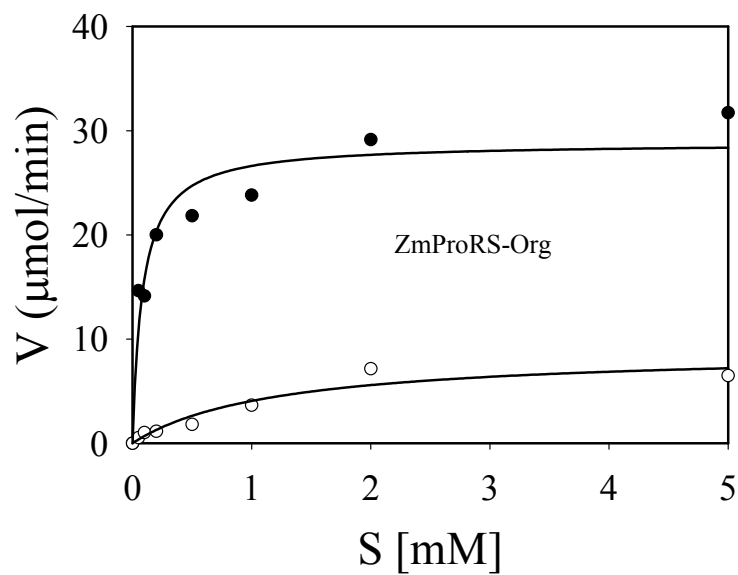
(B)

	ZmProRS-Org		ZmProRS-Cy	
Amino acid	Proline	A2C	Proline	A2C
V_{max} (μmol min^{-1} mg protein^{-1})	28.84 (±178)	8.92 (±1.53)	30.04 (±1.02)	35.67 (±1.74)
K_m (mM)	0.08 (±0.03)	1.20 (±0.53)	0.31 (±0.04)	0.11 (±0.03)

5. References

Ahel I, Korencic D, Ibba M, Soll D (2003) Trans-editing of mischarged tRNAs. Proc Natl Acad Sci U S A 100: 15422-15427

Ahel I, Stathopoulos C, Ambrogelly A, Sauerwald A, Toogood H, Hartsch T, Soll D (2002) Cysteine activation is an inherent in vitro *property* of prolyl-tRNA synthetases. J Biol Chem 277: 34743-34748

Akashi K, Grandjean O, Small I (1998) Potential dual targeting of an Arabidopsis archaebacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. FEBS Lett 431: 39-44

Ambrogelly A, Ahel I, Polycarpo C, Bunjun-Srihari S, Krett B, Jacquin-Becker C, Ruan B, Kohrer C, Stathopoulos C, RajBhandary UL, Soll D (2002) Methanocaldococcus jannaschii prolyl-tRNA synthetase charges tRNA(Pro) with cysteine. J Biol Chem 277: 34749-34754

An S, Barany G, Musier-Forsyth K (2008) Evolution of acceptor stem tRNA recognition by class II prolyl-tRNA synthetase. Nucleic Acids Res 36: 2514-2521

Andersson JO, Sarchfield SW, Roger AJ (2005) Gene transfers from nanoarchaeota to an ancestor of diplomonads and parabasalids. Mol Biol Evol 22: 85-90

Baum BJ, Johnson LS, Franzblau C, Troxler RF (1975) Incorporation of L-Azetidine-2-carboxylic Acid into Hemolgobin in Rabbit Reticulocytes in *Vitro*. The Journal of Biological Chemistry 250: 1464-1471

Berendt U, Haverkamp T, Prior A, Schwenn JD (1995) Reaction mechanism of thioredoxin: 3'-phospho-adenylylsulfate reductase investigated by site-directed mutagenesis. Eur J Biochem 233: 347-356

Beuning PJ, Musier-Forsyth K (2000) Hydrolytic editing by a class II aminoacyl-tRNA synthetase. Proc Natl Acad Sci U S A 97: 8916-8920

Beuning PJ, Musier-Forsyth K (2001) Species-specific differences in amino acid editing by class II prolyl-tRNA synthetase. J Biol Chem 276: 30779-30785

Bick JA, Dennis JJ, Zylstra GJ, Nowack J, Leustek T (2000) Identification of a new class of 5'-adenylylsulfate (APS) reductases from sulfate-assimilating bacteria. J Bacteriol 182: 135-142

Bick JA, Leustek T (1998) Plant sulfur metabolism--the reduction of sulfate to sulfite. Curr Opin Plant Biol 1: 240-244

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248-254

Bunjun S, Stathopoulos C, Graham D, Min B, Kitabatake M, Wang AL, Wang CC, Vivares CP, Weiss LM, Soll D (2000) A dual-specificity aminoacyl-tRNA synthetase in the deep-rooted eukaryote Giardia lamblia. Proc Natl Acad Sci U S A 97: 12997-13002

Burke B, Lipman RS, Shiba K, Musier-Forsyth K, Hou YM (2001) Divergent adaptation of tRNA recognition by Methanococcus jannaschii prolyl-tRNA synthetase. J Biol Chem 276: 20286-20291

Chiba Y, Ishikawa M, Kijima F, Tyson RH, Kim J, Yamamoto A, Nambara E, Leustek T, Wallsgrove RM, Naito S (1999) Evidence for autoregulation of cystathionine gamma-synthase mRNA stability in Arabidopsis. Science 286: 1371-1374

Cusack S, Yaremchuk A, Krikliviy I, Tukalo M (1998) tRNA(Pro) anticodon recognition by Thermus thermophilus prolyl-tRNA synthetase. Structure 6: 101-108

Dietrich A, Marechal-Drouard L, Carneiro V, Cosset A, Small I (1996) A single base change prevents import of cytosolic tRNA(Ala) into mitochondria in transgenic plants. Plant J 10: 913-918

Dock-Bregeon A, Sankaranarayanan R, Romby P, Caillet J, Springer M, Rees B, Francklyn CS, Ehresmann C, Moras D (2000) Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem. Cell 103: 877-884

Duchene AM, Giritch A, Hoffmann B, Cognat V, Lancelin D, Peeters NM, Zaepfel M, Marechal-Drouard L, Small ID (2005) Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in Arabidopsis thaliana. Proc Natl Acad Sci U S A 102: 16484-16489

Dunnill PM, Fowden L (1965) Azetidine-2-carboxylic acid breakdown by soil micro-organism. Phytochemistry 4: 445-451

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347: 203-206

Feng L, Stathopoulos C, Ahel I, Mitra A, Tumbula-Hansen D, Hartsch T, Soll D (2002) Aminoacyl-tRNA formation in the extreme thermophile Thermus thermophilus. Extremophiles 6: 167-174

Fersht AR, Dingwall C (1979) Establishing the misacylation/deacylation of the tRNA pathway for the editing mechanism of prokaryotic and eukaryotic valyl-tRNA synthetases. Biochemistry 18: 1238-1245

Fowden L (1962) Amino-acid Analogues and the Growth of Seedlings. Journal of experimental botany 14: 387-398

Fowden L, Bryant M (1959) Nitrogenous compounds and nitrogen metabolism in the Liliaceae. 5. The metabolism of azetidine-2-carboxylic acid. Biochem J 71: 210-217

Fowden L, Neale S, Tristram H (1963) Effect Of 3,4-Dehydro-Dl-Proline On Growth And Protein Synthesis. Nature 199: 35-38

Fowden L, Richmond MH (1963) Replacement of proline by azetidine-2-carboxylic acid during bioxynthesis of protein. Biochim. Biophys. Acta. 71: 459-461

Freist W, Sternbach H, Pardowitz I, Cramer F (1998) Accuracy of protein biosynthesis: quasi-species nature of proteins and possibility of error catastrophes. J Theor Biol 193: 19-38

Gallori E, Bazzicalupo M, Parisi B, Pedaggi G, Polsinelli M (1978) Resistance to (L)-azetidin-2-carboxylic acid in Bacillus subtilis. Biochem Biophys Res Commun 85: 1518-1525

Gendre D, Czernic P, Conejero G, Pianelli K, Briat JF, Lebrun M, Mari S (2007) TcYSL3, a member of the YSL gene family from the hyper-accumulator Thlaspi caerulescens, encodes a nicotianamine-Ni/Fe transporter. Plant J 49: 1-15

Gillemaut P, Weil JH (1975) Aminoacylation of Phaseolus vulgaris cytoplasmic, chloroplastic and mitochondrial tRNAsMet and of Escherichia coli tRNAsMet by homologous and heterologous enzymes. Biochim Biophys Acta 407: 240-248

Gould SB, Waller RF, McFadden GI (2008) Plastid evolution. Annu Rev Plant Biol 59: 491-517

Gray MW, Burger G, Lang BF (2001) The origin and early evolution of mitochondria. Genome Biol 2: REVIEWS1018

Gross C, Felsheim R, Wackett LP (2008) Genes and enzymes of azetidine-2-carboxylate metabolism: detoxification and assimilation of an antibiotic. J Bacteriol 190: 4859-4864

Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177: 4121-4130

Hati S, Ziervogel B, Sternjohn J, Wong FC, Nagan MC, Rosen AE, Siliciano PG, Chihade JW, Musier-Forsyth K (2006) Pre-transfer editing by class II prolyl-tRNA synthetase: role of aminoacylation active site in "selective release" of noncognate amino acids. J Biol Chem 281: 27862-27872

Hausmann CD, Ibba M (2008) Aminoacyl-tRNA synthetase complexes: molecular multitasking revealed. FEMS Microbiol Rev 32: 705-721

Hewitt EJ, Notton BA (1967) Inhibition by L-azetidine-2-carboxylic acid of induction of ntirate reductase in plants and its reversal by L-proline. Pytochemistry 6: 1329-1335

Higuchi K, Suzuki K, Nakanishi H, Yamaguchi H, Nishizawa NK, Mori S (1999) Cloning of nicotianamine synthase genes, novel genes involved in the biosynthesis of phytosiderophores. Plant Physiol 119: 471-480

Hogg J, Schiefermayr E, Schiltz E, Igloi GL (2008) Expression and properties of arginyl-tRNA synthetase from jack bean (Canavalia ensiformis). Protein Expr Purif **61:** 163-167

Hopkins RJ, van Dam NM, van Loon JJ (2009) Role of glucosinolates in insect-plant relationships and multitrophic interactions. Annu Rev Entomol 54: 57-83

Hoshikawa C, Shichiri M, Nakamori S, Takagi H (2003) A nonconserved Ala401 in the yeast Rsp5 ubiquitin ligase is involved in degradation of Gap1 permease and stress-induced abnormal proteins. Proc Natl Acad Sci U S A 100: 11505-11510

Ibba M, Soll D (2000) Aminoacyl-tRNA synthesis. Annu Rev Biochem 69: 617-650

Igloi GL, Schiefermayr E (2009) Amino acid discrimination by arginyl-tRNA synthetases as revealed by an examination of natural specificity variants. Febs J **276:** 1307-1318

Imbault P, Colas B, Sarantoglou V, Boulanger Y, Weil JH (1981) Chloroplast leucyl-tRNA synthetase from Euglena gracilis. Purification, kinetic analysis, and structural characterization. Biochemistry 20: 5855-5859

Isono K, Niwa Y, Satoh K, Kobayashi H (1997) Evidence for transcriptional regulation of plastid photosynthesis genes in Arabidopsis thaliana roots. Plant Physiol 114: 623-630

Jacquin-Becker C, Ahel I, Ambrogelly A, Ruan B, Soll D, Stathopoulos C (2002) Cysteinyl-tRNA formation and prolyl-tRNA synthetase. FEBS Lett 514: 34-36

Jakubowski H, Goldman E (1992) Editing of errors in selection of amino acids for protein synthesis. Microbiol Rev 56: 412-429

Kim J, Lee M, Chalam R, Martin MN, Leustek T, Boerjan W (2002) Constitutive overexpression of cystathionine gamma-synthase in Arabidopsis leads to accumulation of soluble methionine and S-methylmethionine. Plant Physiol 128: 95-107

Koprivova A, North KA, Kopriva S (2008) Complex signaling network in regulation of adenosine 5'-phosphosulfate reductase by salt stress in Arabidopsis roots. Plant Physiol 146: 1408-1420

Kumar R, Marechal-Drouard L, Akama K, Small I (1996) Striking differences in mitochondrial tRNA import between different plant species. Mol Gen Genet 252: 404-411

Lee M, Martin MN, Hudson AO, Lee J, Muhitch MJ, Leustek T (2005) Methionine and threonine synthesis are limited by homoserine availability and not the activity of homoserine kinase in Arabidopsis thaliana. Plant J 41: 685-696

Leete N, Lousters LL, Rao H, Prakash S (1986) Biosynthesis of azetidine-2-carboxylic acid in Convallaria majalis: studies with N-15 labelled precursors. Phytochemistry 25: 2753-2758

Liu H, Peterson R, Kessler J, Musier-Forsyth K (1995) Molecular recognition of tRNA(Pro) by Escherichia coli proline tRNA synthetase in vitro. Nucleic Acids Res 23: 165-169

Ma JF, Nomoto K (1994) Incorporation of Label from 13C-, 2H-, and 15N-Labeled Methionine Molecules during the Biosynthesis of 2[prime]-Deoxymugineic Acid in Roots of Wheat. Plant Physiol 105: 607-610

Maréchal-Drouard L, Weil JH, Dietrich A (1993) Transfer RNAs and transfer RNA genes in plants. Annu Rev Plant Physiol Plant Mol Biol 44:13–32

Martin MN, Tarczynski MC, Shen B, Leustek T (2005) The role of 5'-adenylylsulfate reductase in controlling sulfate reduction in plants. Photosynth Res 86: 309-323

Martinis SA, Plateau P, Cavarelli J, Florentz C (1999) Aminoacyl-tRNA synthetases: a family of expanding functions. Mittelwihr, France, October 10-15, 1999. Embo J 18: 4591-4596

Menand B, Marechal-Drouard L, Sakamoto W, Dietrich A, Wintz H (1998) A single gene of chloroplast origin codes for mitochondrial and chloroplastic methionyl-tRNA synthetase in Arabidopsis thaliana. Proc Natl Acad Sci U S A 95: 11014-11019

Mizuno D, Higuchi K, Sakamoto T, Nakanishi H, Mori S, Nishizawa NK (2003) Three nicotianamine synthase genes isolated from maize are differentially regulated by iron nutritional status. Plant Physiol 132: 1989-1997

Nakai Y, Nakai M, Lill R, Suzuki T, Hayashi H (2007) Thio modification of yeast cytosolic tRNA is an iron-sulfur protein-dependent pathway. Mol Cell Biol 27: 2841-2847

Nomura M, Nakamori S, Takagi H (2003) Characterization of novel acetyltransferases found in budding and fission yeasts that detoxify a proline analogue, azetidine-2-carboxylic acid. J Biochem 133: 67-74

Norris D, Fowden L (1972) Substrate discrimination by Prolyl-tRNA synthetase from various higher plants. Pytochemistry 11: 2921-2935

Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. Genetics **120:** 621-623

Omori K, Suzuki SI, Imai M, Komatsubara S (1992) Analysis of the mutant proBA operon from a proline-producing strain of *Serratia marcescens*. J. Gen. Microbiol. 138: 693-699

Omori K, Suzuki SI, Imai Y, Komatsubara S (1991) Analysis of the *Serratia marcescens* proBA operon and feedback control of proline biosynthesis. J. Gen. Micorbiol. 137: 509-517

Pesaresi P, Masiero S, Eubel H, Braun HP, Bhushan S, Glaser E, Salamini F, Leister D (2006) Nuclear photosynthetic gene expression is synergistically modulated by rates of protein synthesis in chloroplasts and mitochondria. Plant Cell 18: 970-991

Pujol C, Bailly M, Kern D, Marechal-Drouard L, Becker H, Duchene AM (2008) Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and chloroplastic Gln-tRNAGln synthesis in plants. Proc Natl Acad Sci U S A **105:** 6481-6485

Reger BJ, Fairfield SA, Epler JL, Barnett WE (1970) Identification and origin of some chloroplast aminoacyl-tRNA synthetases and tRNAs. Proc Natl Acad Sci U S A 67: 1207-1213

Ribas de Pouplana L, Brown JR, Schimmel P (2001) Structure-based phylogeny of class IIa tRNA synthetases in relation to an unusual biochemistry. J Mol Evol 53: 261-268

Rokov-Plavec J, Dulic M, Duchene AM, Weygand-Durasevic I (2008) Dual targeting of organellar seryl-tRNA synthetase to maize mitochondria and chloroplasts. Plant Cell Rep 27: 1157-1168

Rokov-Plavec J, Lesjak S, Landeka I, Mijakovic I, Weygand-Durasevic I (2002) Maize seryl-tRNA synthetase: specificity of substrate recognition by the organellar enzyme. Arch Biochem Biophys 397: 40-50

Rubenstein E, Zhou H, Krasinska KM, Chien A, Becker CH (2006) Azetidine-2-carboxylic acid in garden beets (Beta vulgaris). Phytochemistry 67: 898-903

Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry JC, Moras D (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). Science 252: 1682-1689

Sambrook S, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press

Setya A, Murillo M, Leustek T (1996) Sulfate reduction in higher plants: molecular evidence for a novel 5'-adenylylsulfate reductase. Proc Natl Acad Sci U S A 93: 13383-13388

Shapiro MB, Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15: 7155-7174

Shichiri M, Hoshikawa C, Nakamori S, Takagi H (2001) A novel acetyltransferase found in Saccharomyces cerevisiae Sigma1278b that detoxifies a proline analogue, azetidine-2-carboxylic acid. J Biol Chem 276: 41998-42002

Smith CJ, Deutch AH, Rushlow KE (1984) Purification and characteristics of a γ-glutamyl kinase involved in *Escherichia coli* proline biosynthesis. J. bacteriol. 157: 545-551

Splan KE, Ignatov ME, Musier-Forsyth K (2008) Transfer RNA modulates the editing mechanism used by class II prolyl-tRNA synthetase. J Biol Chem 283: 7128-7134

Stathopoulos C, Li T, Longman R, Vothknecht UC, Becker HD, Ibba M, Soll D (2000) One polypeptide with two aminoacyl-tRNA synthetase activities. Science 287: 479-482

Stehlin C, Burke B, Yang F, Liu H, Shiba K, Musier-Forsyth K (1998) Species-specific differences in the operational RNA code for aminoacylation of tRNAPro. Biochemistry 37: 8605-8613

Stehlin C, Heacock DH, 2nd, Liu H, Musier-Forsyth K (1997) Chemical modification and site-directed mutagenesis of the single cysteine in motif 3 of class II Escherichia coli prolyl-tRNA synthetase. Biochemistry 36: 2932-2938

Steinmetz A, Weil JH (1986) Isolation and characterization of chloroplast and cytoplasmic tRNAs. Methods in Enzymology (A. Weissbach and H. Weissbach eds) Academic Press, New York. 118: 212-231

SternJohn J, Hati S, Siliciano PG, Musier-Forsyth K (2007) Restoring species-specific posttransfer editing activity to a synthetase with a defunct editing domain. Proc Natl Acad Sci U S A 104: 2127-2132

Sugiura M, Takagi T, Kisumi M (1985) Proline production by regulatory mutants of Serratia marcescens. Appl. Microbiol. Biotechnol. 21: 213-219

Szymanski M, Barciszewski J (1999) The aminoacyl-tRNA Synthetase Data Bank (AARSDB). Nucleic Acids Res 27: 332-335

Szymanski M, Deniziak M, Barciszewski J (2000) The new aspects of aminoacyl-tRNA synthetases. Acta Biochim Pol 47: 821-834

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599

Tang Y, Tirrell DA (2002) Attenuation of the editing activity of the Escherichia coli leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. Biochemistry 41: 10635-10645

Tumbula DL, Becker HD, Chang WZ, Soll D (2000) Domain-specific recruitment of amide amino acids for protein synthesis. Nature 407: 106-110

Vanlerberghe GC, Brown LM (1987) Proline overproduction in cells of the green alga *Nannochloris Bacillaris* resistant to azetidine 2 carboxylic acid. Plant Cell Environ. 10: 251-257

Verbruggen N, Borstlap AC, Jacobs M, Van Montagu M, Messens E (1996) The raz1 mutant of Arabidopsis thaliana lacks the activity of a high-affinity amino acid transporter. Planta 200: 247-253

Wasmuth JJ, Caskey CT (1976) Biochemical characterization of azetidine carboxylic acid-resistant Chinese hamster cells. Cell 8: 71-77

Woese CR, Olsen GJ, Ibba M, Soll D (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol Mol Biol Rev 64: 202-236

Wolf YI, Aravind L, Grishin NV, Koonin EV (1999) Evolution of aminoacyl-tRNA synthetases--analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Res 9: 689-710

Wong FC, Beuning PJ, Nagan M, Shiba K, Musier-Forsyth K (2002) Functional role of the prokaryotic proline-tRNA synthetase insertion domain in amino acid editing. Biochemistry 41: 7108-7115

Wong FC, Beuning PJ, Silvers C, Musier-Forsyth K (2003) An isolated class II aminoacyl-tRNA synthetase insertion domain is functional in amino acid editing. J Biol Chem 278: 52857-52864

Yaremchuk A, Cusack S, Tukalo M (2000) Crystal structure of a eukaryote/archaeon-like protyl-tRNA synthetase and its complex with tRNAPro(CGG). Embo J 19: 4745-4758

Yaremchuk A, Cusack S, Tukalo M (2000) Crystallization and preliminary X-ray diffraction analysis of Thermus thermophilus prolyl-tRNA synthetase. Acta Crystallogr D Biol Crystallogr 56: 195-196

Yaremchuk A, Kriklivyi I, Cusack S, Tukalo M (2000) Improved crystals of Thermus thermophilus prolyl-tRNA synthetase complexed with cognate tRNA obtained by crystallization from precipitate. Acta Crystallogr D Biol Crystallogr 56: 197-199

Yeung KF, Lee KM, Woodard RW (1998) Isolation and Identification of Two l-Azetidine-2-carboxylic Acid-Degrading Soil Microorganisms, Enterobacter agglomerans and Enterobacter amnigenus. J Nat Prod 61: 207-211

Xu Q, Wen, X., and Deng, X. (2004) A Simple Protocol for Isolation Genomic DNA from Chestnut Rose (*Rosa roxburghii Tratt*) for RFLP and PCR Analyses. Plant Molecular Biology Reporter **22:** 301a-301g

Zhang CM, Perona JJ, Ryu K, Francklyn C, Hou YM (2006) Distinct kinetic mechanisms of the two classes of Aminoacyl-tRNA synthetases. J Mol Biol 361: 300-311

Zhang H, Huang K, Li Z, Banerjei L, Fisher KE, Grishin NV, Eisenstein E, Herzberg O (2000) Crystal structure of YbaK protein from Haemophilus influenzae (HI1434) at 1.8 A resolution: functional implications. Proteins 40: 86-97

Curriculum Vitae
Jiyeon Lee

Education

2002-2009   Rutgers University, New Brunswick, NJ
              Ph.D. in Plant Biology, October 2009
1999-2001   Gyeongsang National University, Gyeongsangnam-Do, South of Korea
              M.A. in Biochemistry, February 2001
1994-1999   Gyeongsang National University, Gyeongsangnam-Do, South of Korea
              B.A degree in Biochemistry, February 1999

Research Experience

2001-2002   Intern Researcher
            Korea Science and Engineering Foundation (KOSEF), Taejon, Korea
            Affiliated Research Task under KOSEF
1999-2002   Co-researcher, Department of Biochemistry,
            Graduate School, Gyeongsang National University, Chinju, South of Korea
1999-2001   Co-researcher, Department of Biochemistry,
            Graduate School, Gyeongsang National University, Chinju, South of Korea

Publications

1.   Lee, M., Martin M., Hudson A., **Lee, J**., Muhitch, M., Leustek, T. (2005) Methionine and threonine synthesis are limited by homoserine availability and not the activity of homoserine kinase in *Arabidopsis thaliana*. The Plant Journal. (41) 685-696

2.   Seung Sik Lee, Kyun Oh Lee, Bae Gyo Jung, Young Hun chi, Ji Young Yoo, **Ji Yeon Lee**, Jung Ro Lee, Soo kwon Park, Soon Suk Kang, Ho Hee Jang and Sang Yeol Lee (2001) Molecular characterization of a Chinese cabbage cDNA enconding thioredoxin-h which is predominantly expressed in flowers. J. Biochem. Mol. Biol. (34;4) 334-341

3.   Kyun Oh Lee, Ho Hee Jang, Bae Gyo Jung, Young Hun Chi, **Ji Yeon Lee**, Yeon Ok Choi, Jung Ro Lee, Chae Oh Lim, Moo Je Cho, and Sang Yeol Lee (2000) Rice 1Cys-peroxiredoxin over-expressed in transgenic tobacco does not maintain dormancy but enhances antioxidant activity. FEBS (486) 103-106

4.   Kyun Oh Lee, Min Gab Kim, Ho Hee Jang, **Ji Yeon Lee**, Sun Chang Kim, and Sang Yeol Lee (1999) Biochemical properties of a Chitin-Binding Class Ⅲ Chitinase in Pumpkin Leaves J. Biochme. Mol. Biol. (32;2) 541-546