MOLECULAR CLONING AND CHARACTERIZATION OF REGULATORY ENZYMES IN THREONINE BIOSYNTHETIC PATHWAY FROM SOYBEAN

A Thesis presented to the Faculty of the Graduate School at the University of Missouri-Columbia

---

In Partial Fulfillment of the Requirements for the Degree

Master of Science

---

by

SRINIVASA RAO YANAMADALA

Dr. Hari B Krishnan, Thesis Supervisor

AUGUST 2007

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

MOLECULAR CLONING AND CHARACTERIZATION OF REGULATORY ENZYMES IN THREONINE BIOSYNTHETIC PATHWAY FROM SOYBEAN

presented by **Srinivasa Rao Yanamadala,**

a candidate for the degree of master of science

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Hari B. Krishnan

Dr. Dale G. Blevins

Dr. Joseph C. Polacco

ACKNOWLEDGMENTS

First, I would like to thank my thesis supervisor Dr. Hari Krishnan for his guidance and encouragement during the entire course of this project. I also would like to thank Drs. Dale Blevins and Joseph C. Polacco for their help in reviewing this thesis. I thank my parents and the rest of my family for their support and encouragement. To my friends, Ping Huang, Satish, Ganesh, and Ranjitha, thanks for the help and friendship. Thanks to current and past members of the Krishnan's lab for all the interaction I had with them.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1

Figure **Page**

Chapter 2

Chapter 3

MOLECULAR CLONING AND CHARACTERIZATION OF REGULATORY ENZYMES IN THREONINE BIOSYNTHETIC PATHWAY FROM SOYBEAN

SRINIVASA RAO YANAMADALA

Dr. Hari B Krishnan, Thesis Supervisor

ABSTRACT

Soybeans (*Glycine max* [L.] Merr.) are a good source of protein and oil. In addition to its importance in the human diet, this crop plays a major role as an important component of animal feed, along with corn. Despite being a very good source of protein, soybeans nutritional value is limited by low proportion of sulfur amino acids methionine and cysteine. Hence, achieving a favorable amino acid composition of soybean protein has been an objective of researchers for a long time. Recently, synthetic amino acids are being added to soy/corn based animal feeds in order to supplement the limited amino acids. The cost involved aside, other essential amino acids, called second-tier amino acids, are also becoming limiting in animal feed due to addition of synthetic amino acids. Threonine is one of those second-tier amino acids and improving its proportion in soybean seeds can improve the nutritional value of soy-based animal feed. The manipulation of key regulatory enzymes in the biosynthesis pathway of this

amino acid is one method used to increase the proportion of free threonine in soybean seeds. One prerequisite for such an approach is the isolation and characterization of the genes encoding the key enzymes. The objective of this project was to isolate and characterize the homoserine dehydrogenase (HSDH) and threonine synthase (TS) genes in the soybean, in order to increase the proportion of threonine in free amino acid pool.

The genes coding both HSDH and TS have been isolated and named *gmhsdh* and *gmts,* respectively. The *gmhsdh* seems to code a mono-functional, cytosolic HSDH and is interrupted by 11 introns. The *gmts* encodes a chloroplast localized TS and this gene has no introns. The cDNA encoding HSDH was also isolated and expressed in *E.coli* to verify the functionality of the predicted coding region of HSDH and used to raise antibodies. Multiple sequence analysis of the HSDH amino acid sequence leads to a proposition that this protein is feedback insensitive to threonine. Verification of this fact could make this a very good candidate for achieving a favorable proportion of not only threonine but also methionine and isoleucine in the free amino acid pool. The promoter region of *gmts* also was isolated and sequenced, which could reveal novel mechanisms in the regulation of plant TS.

CHAPTER 1

INTRODUCTION AND REVIEW OF LITERATURE

Economic Importance of Soybeans

Soybean (*Glycine max* [L.] Merr.) is an economically important legume. Seeds of this crop are comprised of carbohydrates (12%), oils (20%) and proteins (30-40%) based on cultivars and agronomic practices (Liu, 1997). Soybean is a rich source of both oil and protein and is considered to be the oilseed crop with the highest protein content. According to Food and Agricultural Organization(FAO), 225.84 million acres of soybeans were harvested all over the world in the year 2005.

The United States (US) ranks first in the world in cultivated area and in total production of soybeans. This crop was grown in 75.5 million acres in the year 2006. Soybean production for the same year was 86.7 million metric tons with an average yield of 42.7 bushels per acre (Soy Stats Guide, 2007). The soybean is a very important crop for the US economy as the estimated export of soybean for the year 2006 is 1.12 billion bushels (Economic Research Service, 2007).

With this crop grown in 5.15 million acres, the state of Missouri ranks fifth in cultivated area for soybean in the year 2006 after Iowa, Illinois, Minnesota, and Indiana. However, Missouri stands only seventh in terms of production with 5.2 million metric tons. This crop is important to this state as the total value of this crop in 2006 was $1.22 billions (Soy Stats Guide, 2007).

Soybean as a Food Source

Though soybeans were introduced into the US in the late 1700s, the cultivation of this crop was started only in the late 1800s. Initially, soybean was grown as a forage crop with a slow rise in acreage for seed production. By the year 1941, the acreage for soybean grown for seed exceeded that grown for forage, and since then the crop has been cultivated almost exclusively for seed production (Probst and Judd, 1973). Today, soybeans are grown mainly for their protein and the oil components of seeds.

Soybean is a major source of food for both humans and animals. While it has been one of the major food sources for humans in Asian countries, human consumption of soybeans is gaining momentum in North American countries recently. According to Soyfoods Association of America (1995), 26 million Americans consume foods made from

soybeans. Soy-derived foods have also been proven to be beneficial for human health by helping to prevent diseases such as cancer, heart disease, and osteoporosis. In a significant development, a health claim that consumption of 25 grams of soybean protein per day can help lower cholesterol and prevent heart disease was approved by the United States Food and Drug Administration (FDA) in the year 1999 (Bachmann, 2001). By the year 2003, at least one in six Americans consumed soy foods or soy beverages once a week or more (Rao *et al.,* 2002). Because of the proven health benefits, soy foods are becoming readily available in large supermarkets in the USA. Today, various types of soy foods like fermented whole soybean products, roasted soybean products, soymilk, tofu, soy cheese, soy yogurt, nondairy frozen deserts and okara are available to consumers (Kroll, 2000).

The soybean has been the most important protein source in animal feed. Most of the soy protein is used as animal feed in the form of defatted meal. Soybean meal contains 44 to 48% of protein, depending on the type of processing used to produce the meal (Maier *et al.,* 1998). The importance of soy meal in animal feed can be understood by the fact that 83% of total protein in animal feed consumed in 2004 came from soybean meal as against 7% from all other oilseed

meals together, 3% from grain protein, and 7% from animal protein (Fig. 1).

Limiting Amino Acids

Despite the importance of soybean seed protein, the nutritional value of this protein is limited by the low proportions of sulfur containing amino acids methionine and cysteine. Soy/Corn based rations have been the chief food source for animal feed industry. Since soy proteins are limited in sulfur amino acids and corn proteins are limited in lysine (Beauregard and Hefford, 2006), one of the major focuses of plant biologists has been improving the content of these limited amino acids in legumes and cereals. While these efforts are still ongoing, the practice of adding synthetic methionine and lysine to the feed has become popular. This practice not only costs more than $100 million annually (Imsande, 2001), but also causes other amino acids to be limited in soy/corn based rations. From the results of recent animal feeding experiments, a group of essential amino acids called second-tier amino acids are becoming limited in soybean and corn based feed when synthetic amino acids are added. These second-tier amino acids include threonine, tryptophan, isoleucine, arginine and valine (Rapp ***et al.,*** 2003). Since compensating for the deficiency of all these essential amino acids with

synthetics is more expensive, enhancing the content of these second-tier amino acids in soybean seeds is important.

Approaches to Improve Amino Acid Content

Amino acid content in the economic parts of crop plants can be enhanced by the following approaches.

- Conventional Plant Breeding
- Genetic Engineering Approaches

Conventional Plant Breeding

Traditional plant breeders employed various selection and/or hybridization methods to improve soybean seed protein quality (Wilcox, 1998). Various selection methodologies have been successful in increasing the proportion of seed protein in soybeans (Brim and Burton, 1979). However, these efforts were not very successful in increasing the content of specific amino acids for two reasons. First, the major focus was increasing total seed protein *per se* rather than looking to a particular amino acid. This was challenging, owing to the negative correlation with seed yield and seed oil content. Second, though there have been efforts to increase content of a particular amino acid, poor variability for the levels of target amino acids limited any progress. This made

breeders look to various approaches like mutation breeding and tissue culture methods in order to create variability. Though a spontaneous maize mutant, *opaque-2* has high lysine content, the kernels of mutant plants showed poor texture (Mertz *et al.,* 1964). Subsequent efforts to retrieve normal seed phenotype met partial success and these mutants with a near normal phenotype are called 'High-Quality Maize' (Gibbon and Larkins, 2005).

Though different tissue culture techniques helped plant breeders in various aspects of crop improvement, one major problem associated is regeneration of crop plants. Upon treatment with ethionine – a chemical analog of methionine – soybean cell cultures resulted in lines that have methionine 8.7 times more than the parental line. However, these lines were not able to regenerate soybean seedlings (Madison and Thompson, 1988). One significant effort of mutation breeding was the chemical mutagenesis of soybean seeds using ethyl methane sulfonate (EMS) followed by screening for high methionine lines. Of the many lines thus screened, one showed 20% higher levels of methionine and cysteine (Imsande, 2001). This investigation not only achieved higher levels of target amino acids but also pioneered the efforts to increase the contents of specific amino acids rather than total seed protein. Another

approach was to derive mutants for altered composition of different seed proteins to conveniently elevate levels of target amino acids. The proportion of 7S globulins to the total seed protein of soybean determines the content of methionine and cysteine in soybean seeds. Approximately 70% of the total seed protein in soybean is constituted by glycinin (11S) and β-conglycinin (7S) multi-subunit proteins (Meinke ***et al.,*** 1981). While methionine and cysteine account for 3 to 4.5% of amino acid residues of glycinin (Nielson ***et al.,*** 1989; Fukushima, 1991), their contribution to β-conglycinin is only 1% (Harada ***et al.,*** 1989). Since the manipulation of the expression of these two proteins in the seed could result in higher levels of methionine and cysteine, many breeders tried to produce mutants to create altered composition in these two proteins. While one recessive and one dominant mutation resulting in the complete elimination of the expression of β-conglycinin were characterized, the recessive mutants showed developmental abnormalities and reproduction failures (Kitagawa ***et al.,*** 1991). However, the dominant mutant was successfully used to incorporate this trait into soybean cultivars (Hajika ***et al.,*** 1996; Teraishi ***et al.,*** 2001).

Genetic Engineering Approaches

With the accumulation of knowledge about seed storage proteins, genes that code these proteins and expression of the genes opened new avenues for protein quality improvement through genetic engineering. Reasonable success in the development of plant transformation systems for a variety of crop plants is another encouraging factor for the importance of genetic engineering approaches. Various genetic engineering approaches can be broadly classified into four categories

1. Expression of Heterologous Proteins
2. Expression of Homologous Proteins
3. Protein Engineering, and
4. Manipulation of Amino Acid Biosynthesis Pathways

Expression of Heterologous Proteins

One way to increase the concentration of specific amino acids is the heterologous expression of proteins. Expression of methionine rich maize or any other plant protein in soybean or lysine rich soybeans or other legume proteins in maize may increase methionine and lysine in soybean and maize, respectively. Heterologous expression of lysine rich proteins in cereal crops has been attempted with notable success. When β-phaseolin from beans was

expressed in transgenic rice, a high level of expression, to an extent of 4% of total seed proteins, was achieved (Zheng *et al.,* 1995). Since methionine and cysteine are limiting amino acids in the seed proteins of legumes, expression of heterologous methionine rich proteins has been the major object in crops like soybean. Brazil nut 2S albumin was one such methionine-rich protein which, when expressed in soybeans, resulted in an increase of methionine by 15 to 40% (Townsend and Thomas, 1994). However, this protein was later characterized as a major allergen (Nordlee *et al.,* 1996) and hence no serious efforts were employed to develop transgenic cultivars. Expression of a methionine rich 2S albumin protein from sunflower in lupin seeds resulted in a two-fold improvement in methionine content (Tabe *et al.,* 1993). However, it has not been tested thoroughly to see if it is similarly allergenic. Different zein proteins of maize have been identified as rich sources of methionine. The 10 kD zein has 30% methionine (Kirihara *et al.,* 2001), a novel 11 kD zein has 25% combined methionine and cysteine (Krishnan, 2005) and the 15 kD zein has 15% methionine (Pedersen *et al.,* 1986). Expression of the 15 kD delta zein protein in soybeans resulted in an increase of 12-20% of methionine and 15-35% of cysteine (Dinkins *et al.,* 2001). In a similar

effort, a 50-70% increase in methionine content of alcohol-soluble protein fractions was reported when an 11 kD delta zein protein was over expressed in soybean seeds (Kim and Krishnan, 2004).

Expression of Homologous Proteins

Another strategy to increase the concentration of specific amino acids is to up or down regulate levels of endogenous proteins. Whereas lysine is co-limiting with tryptophan in the seed proteins of maize, lysine and threonine are the limiting amino acids in all other cereal crops (Beauregard and Hefford, 2006). Prolamines and glutelins are the predominant proteins in all cereal plants. Prolamines account for approximately 50% of the total seed protein, but the lysine fraction of these proteins is very low. Even though glutelins have a high lysine fraction, the content of glutelin is small thus making cereal seed proteins in general limited in lysine. In maize, the zein proteins are major prolamines, but they have as little as 0.1g lysine/100 g of protein compared to 3.2 g lysine/100g of glutelin protein (Nelson, 1969). This contributes to the poor nutritional quality of corn. The focus of researchers working on maize has been the down-regulation of highly expressed zein genes so that the lysine rich glutelins proportion increases in seeds. Using

chimeric double stranded RNA constructs, Huang *et al.,* (2006) achieved a higher content of lysine and tryptophan by repressing the expression of 19 and 22 kD zeins.

In soybeans, methionine rich proteins (MRPs) have a high proportion of methionine but the concentration of protein as such is poor in seeds (de Lumen *et al.,* 1999). Elevated expression of these proteins in seeds should theoretically increase the levels of sulfur amino acids. However, the structure and regulation of these proteins have not been characterized enough to be genetically engineered.

Protein Engineering

Modification of already existing abundant proteins is another approach to improving the levels of specific amino acids. The hypervariable region (HVR) of glycinin, an abundant protein in soybeans, has a high variability for the amino acids it codes among its multi-gene family and hence was proposed to tolerate the insertion of methionine and cysteine residues (Nielsen *et al.,* 1990). When one of the genes coding for this protein was engineered by inserting oligonucleotides coding for methionine and expressing it in tobacco and potatoes, the protein accumulated normally without any problem of folding (Gidamis *et al.,* 1995; Takaiwa *et al.,* 1995). These

experiments showed that the insertion of coding regions for new amino acids is a possible way of genetic engineering for the improvement of target amino acid concentrations. A prerequisite for this approach is a thorough characterization of protein structures and identifying the right coding region that can tolerate new insertions. Rapp *et al.,* (2003) also succeeded in introducing oligonucleotides coding for 10 to 20 residues of isoleucine, another second tier amino acid, in the coding region of β-conglycinin without compromising high level expression of the protein.

Manipulation of Biosynthetic Pathways

The manipulation of biosynthetic pathways has been gaining attention as one method for improving the free amino acid pool in seeds. Free amino acids account for 5% of the total amino acids available in plant seeds and hence modification of composition of the amino acid pool should result, theoretically, in an improvement of the nutritional value of the crops (Ghislain *et al.,* 1995). Since these free amino acids are a part of the total amino acids available for protein synthesis in cells, they can also determine, at least to some extent, the type and amount of proteins synthesized in plant cells. Hence, increasing the levels of limiting amino acids in the free amino acid pools

can potentially remove at least a few limitations to protein synthesis (Beauregard and Hefford, 2006). For example, according to Krishnan (2005), an inadequate availability of sulfur amino acids in developing seeds of the soybean is limiting the synthesis of sulfur-rich proteins. Efforts to improve the availability of these amino acids in developing seeds can also result in a higher expression of such proteins.

An ***E.coli*** serine acetyl transferase (SAT) gene was fused to the sequence coding chloroplast transit peptide of Ribulose-1, 5-bisphosphate carboxylase/oxygenase (RuBisCo) and expressed in the potato in order to obtain up to a threefold increase of free cysteine and glutathione (Den and Bock, 1987). Similarly, the introduction of bacterial cysteine γ-synthase into the potato resulted in a threefold increase in cysteine levels (Hesse ***et al.,*** 2003). When ***Arabidopsis*** cystathionine γ-synthase(CGS) gene, under the control of RuBisCo small unit promoter, was expressed in alfalfa, a significant increase in methionine content and a two - five fold increase in free cysteine levels were achieved (Avraham ***et al.,*** 2005). However, similar results were not obtained when the native potato gene was over-expressed in the potatoes (Kreft ***et al.,*** 2003).

Aspartate Pathway

In bacteria and plants, threonine as well as lysine, methionine, and isoleucine are derived from the aspartate pathway (Fig. 2). This pathway consists of three branches: the first of which leads to lysine, the second to threonine and isoleucine and the third, to methionine (Azevedo *et al.,* 1997). While various mechanisms have been reported for the regulation of enzymes involved in the bacterial pathway, plants appear to utilize feedback inhibition of branch point enzymes (Galili, 1995; Azevedo *et al.*, 1997). The first committed step in the aspartate pathway is ATP-mediated conversion of aspartate to β-aspartyl phosphate catalyzed by aspartate kinase (AK; EC 2.7.2.4). Homoserine dehydrogenase (HSDH) is the first branch point enzyme in threonine/isoleucine/methionine specific pathway competing with dihydrodipicolinate synthase (DHPS), the first committed enzyme in lysine specific pathway, for the common substrate dihydrodipicolinate. The second branch point enzyme in this pathway is threonine synthase (TS), which competes with CGS of the methionine biosynthetic pathway for homoserine phosphate. Since the partitioning of carbon into biosynthesis of threonine is regulated by HSDH and TS, over-expression of these two enzymes may result in an improved content of free threonine in seeds of soybeans. In

order to over-express enzymes, it is necessary to clone and characterize genes coding these enzymes in soybeans. The focus of this research project is to clone and characterize genes coding HSDH and TS enzymes in the soybean.

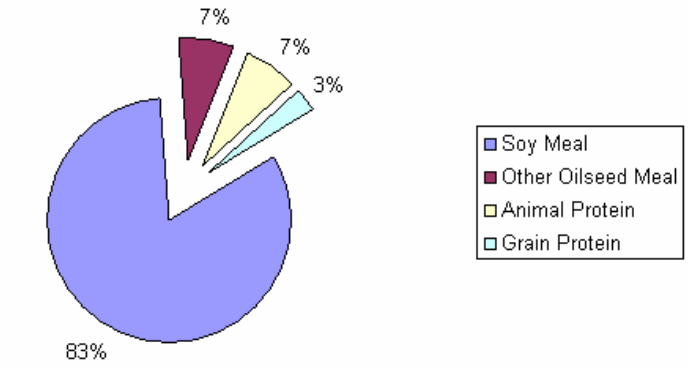

Fig. 1. Consumption of Various Protein Sources in Animal Feed in the year 2004.
Source: Market and Trade Economics Division 2006

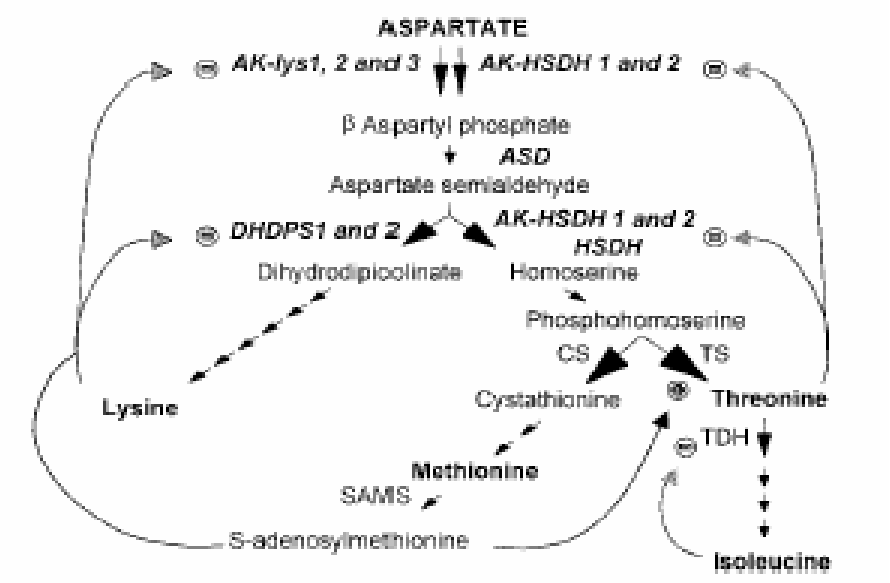

Fig. 2. Key regulatory enzymes in aspartate pathway
Source: Rognes *et al.*, 2002

References

Avraham T, Badani H, Galili S, Amir R (2005) Enhanced levels of methionine and cysteine in transgenic alfalfa (*Medicago sativa* L.) plants over-expressing the *Arabidopsis* cystathionine γ-synthase gene. Plant Biotechnol J 3: 71–79

Azevedo RA, Arruda P, Turner WL, Lea PJ (1997) The biosynthesis and metabolism of the aspartate derived amino acids in higher plants. Phytochemistry **46:**395-419

Bachmann J (2001) Soyfoods: Adding Value to Soybeans [Online]. National Sustainable Agriculture Information Center (ATTRA), University of Arkansas, Fayetteville. Available at http://attra.ncat.org/attra-pub/soyfoods.html (Verified 10 Jun 2007).

Beauregard M, Hefford MA (2006) Enhancement of essential amino acid contents in crops by genetic engineering and protein design Plant Biotechnol J 4:561-574

Brim CA, Burton JW (1979) Recurrent selection in soybeans.II. Selection for increased percent protein in seeds. Crop Sci 19: 494–498

de Lumen BO, Galvez AF, Revilleza MJ, Krenz DC (1999) Molecular strategies to improve the nutritional quality of legume proteins. Adv Exp Med Biol 464: 117–126

Denk D, Böck A (1987) L-cysteine biosynthesis in *Escherichia coli*: nucleotide sequence and expression of the serine acetyltransferase (*cysE*) gene from the wild-type and a cysteine-excreting mutant. J Gen Microbiol 133: 515–525

Dinkins RD, Reddy MSS, Meurer CA, Yan B, Trick H, Thibaud-Nissen F, Finer JJ, Parrott WA, Collins GB (2001) Increased sulfur amino acids in soybean plants over-expressing the maize 15 kD zein protein. In Vitro Cell Dev Biol Plant 37: 742–747

Economic Research Service (2007) [Online]. Available at http://www.ers.usda.gov/News/soybeancoverage.htm (Verified 10 Jun 2007).

Food and Agricultural Organizations of the United Nations. (2005). FAO statistical tables, agriculture [Online]. Available at http://faostat.fao.org/?alias=faostat (Verified 1 Nov 2006).

Fukushima D (1991) Recent progress of soybean protein foods: Chemistry, technology, and nutrition. Food Rev Int 7: 323–351

Galili G (1995) Regulation of lysine and threonine synthesis. Plant Cell 7: 899-906

Ghislain M, Frankard V, Jacobs M (1995) A dinucleotide mutation in DHDPS of *Nicotiana sylvestris* leads to lysine overproduction. Plant J 18: 733–743

Gibbon BC, Larkins BA (2005) Molecular genetic approaches to developing quality protein maize. Trends Genet 21: 227–233

Gidamis AB, Wright P, Haque ZU, Katsube T, Kito M, Utsumi S (1995) Modification tolerability of soybean proglycinin. Biosci Biotech Biochem 59: 1593–1595

Hajika M, Takahashi M, Sakai S, Igita M (1996) A new genotype of 7S globulin (conglycinin) detected in wild soybean (Glycine soja Sieb. et Zucc.). Breed Sci 46: 385–386

Harada JJ, Barker SJ, and Goldberg RB (1989) Soybean betaconglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. Plant Cell 1: 415-425

Hesse H, Harms K, and Hofgen R (2003) Means and methods for enhancing the content of sulfur compounds in plants. US Patent 6,608,239

Huang S, Frizzi A, Florida CA, Kruger DE, Luethy MH (2006) High lysine and high tryptophan transgenic maize resulting from the reduction of both 19- and 22-kD alpha-zeins. Plant Mol Biol 61: 525-535

Imsande J (2001) Selection of soybean mutants with increased concentrations of seed methionine and cysteine. Crop Sci 41: 510–515

Kim WS, and Krishnan HB (2004) Expression of an 11 kD methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. Plant Biotechnol J 2: 199–210

Kirihara JA, Hibberd KA, Janice A (2001) Method for altering the nutritional content of plant seed. US Patent 6,326,527

Kitagawa S, Ishimoto M, Kikuchi F, Kitamura K (1991) A characteristic lacking or remarkably decreasing 7S globulin subunits induced with gamma-ray irradiation in soybean seeds. Jpn J Breed 41: 460–461

Kreft O, Höfgen R, Hesse H (2003) Functional analysis of cystathionine- synthase in genetically engineered potato plants. Plant Physiol 131: 1843–1854

Krishnan HB (2005) Engineering soybean for enhanced sulphur amino acid content. Crop Sci 45 : 454–461

Kroll, D (2000) Soyfoods: Trends and Developments. Updated Edition GA-099R. Business Communications Company, Inc., Norwalk, CT.

Liu K (1997) Soybeans: Chemistry, Technology, and Utilization. p. 1-4. Chapman &Hall, New York, NY.

Madison JT, Thompson JF (1988) Characterization of soybean tissue culture cell lines resistant to methionine analogs. Plant Cell Rep 7: 473–476

Maier DE, Reising J, Briggs JL, Day KM, Christmas EP (1998) High value soybean composition [Online]. Available at http://www.agcom.purdue.edu/AgCom/Pubs/GQ/GQ-39.html (Verified 10 Jun 2007)

Market and Trade Economics Division (2006) [Online]. Available at http://www.library.uiuc.edu/agx/agnic/soynic/info.php?id=9 (Verified Jun 10 2007)

Mertz ET, Bates LS, Nelson OD (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145: 279–280

Meinke DW, Chen J, Beachy RN (1981) Expression of storage protein genes during soybean seed development. Planta 153:130-139

Nelson OE (1969) Genetic modification of protein quality. Adv Agron 21: 171–194.

Nielsen NC, Dickinson CD, Cho T, Thanh VH, Scallon BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB (1989) Characterization of the glycinin family in soybean. Plant Cell 1: 313–328

Nielsen NC, Scott MP, Lago WJP (1990) Assembly properties of modified subunit family. p. 635–640.

Nordlee JA, Taylor SL, Townsend JA, Thomas LA, Bush RK (1996) Identification of Brazil-nut allergen in transgenic soybeans. N Engl J Med 334: 688–692

Pedersen K, Argos P, Naravana SV, Larkins BA (1986) Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of Mr 15000. J Biol Chem 261: 6279–6284

Probst AH, Judd RW (1973) Origin, U.S. history and development, and world distribution. p. 1–15. In B.E. Caldwell (ed.) Soybeans: improvement, production, and uses. ASA, CSSA, and SSSA, Madison, WI.

Rapp BW, Weaver L, Wang Q, Crow L, Ream J, Hill B, Brown W, Oulmassov T, Gruys K (2003) Enhancing the nutritional value of soybean. Plant Biology 2003, Annual meeting of the American Society of Plant Biologists, abstract 38001

Rao MSS, Mullinix BG, Rangappa M, Cebert E, Bhagsari AS, Sapra VT, Joshi JM, Dadson RB (2002) Genotype X environment interactions and yield stability of food-grade soybean genotypes. Agron J 94:72-80

Rognes SE, Dewaele E, Aas SF, Jacobs M, Frankard V (2002) Transcriptional and biochemical regulation of a novel *Arabidopsis thaliana* bifunctional aspartate kinase-homoserine dehydrogenase gene isolated by functional complementation of a yeast *hom6* mutant. Plant Mol Biol 51: 281-294

Soyfoods Association of America (1995) A Consumer Survey of the Soyfoods Shopper [Online]. Walnut Creek, CA. Available at http://www.soyfoods.org (verified 10 Jun 2007)

Soy Stats Guide (2007) [Online]. Available at http://www.soystats.com/2007/Default-frames.htm (Verified 10 Jun 2007)

Tabe LM, Higgins CM, McNabb WC, Hoggins TSV (1993) Genetic engineering of grain and pasture legumes for improved nutritive value. Genetica 90:181–200

Takaiwa F, Katsube T, Kitagawa S, Hisago T, Kito M, Utsumi S (1995) High level accumulation of soybean glycinin in vacuole-derived protein bodies in endosperm tissue of transgenic tobacco seed. Plant Sci 111: 39–49

Teraishi M, Takahashi M, Hajika M, Matsunaga R, Uematsu Y, Ishimoto M (2001) Suppression of soybean-conglycinin genes by a dominant gene, Scg-1. Theor Appl Genet 103: 1266–1272

Townsend JA, Thomas LA (1994) Factors which influence the *Agrobacterium*- mediated transformation of soybean. J Cell Biochem Suppl 18A: 78

Wilcox JR (1998) Increasing Seed Protein in Soybean with Eight Cycles of Recurrent Selection. Crop Sci 38: 1536–1540

Zheng Z, Sumi K, Tanaka K, Murai N (1995) The bean seed storage protein phaseolin is synthesized, processed, and accumulated in the vacuolar type 11 protein bodies of transgenic rice endosperm. Plant Physiol 109: 777–786

CHAPTER 2

ISOLATION AND CHARACTERIZATION OF HOMOSERINE DEHYDROGENASE

Abstract

Threonine, one of the second-tier amino acids, can limit the nutritional value of soybean/corn based animal feeds. Over-expression of a feedback resistant homoserine dehydrogenase (HSDH), one of the regulatory enzymes in threonine biosynthesis, is a possible method for enhancing free threonine in soybean seeds. HSDH operates as a bi-functional enzyme (AK-HSDH) in association with aspartate kinase (AK) in gram-negative bacteria. It functions as a mono-functional enzyme in gram-positive bacteria and yeast. To date, cloning efforts in plants have revealed only bi-functional and feedback-sensitive HSDH. In this study, the cloning and characterization of mono-functional soybean HSDH is reported. Screening of a soybean seedling cDNA library with an EST [Accession # BM308805] as probe revealed a full-length cDNA clone which was found to encode a protein that shows sequence homology with other mono-functional HSDHs. Soybean HSDH consists of 376 amino acids and has a molecular mass of 40.6 kD. The soybean HSDH lacks

coding sequence for a transit peptide causing speculation that this enzyme is cytosolic, unlike most of the enzymes in the aspartate pathway. Sequence analysis also shows that this enzyme lacks the C-terminal extension, which is implicated in feedback inhibition of HSDH by threonine. Heterologous expression of the cDNA in ***E.coli*** resulted in the accumulation of a 41 kD protein which was purified and used to raise antibodies. Southern analysis of soybean genomic DNA indicated that HSDH exists as a single copy in the genome. A genomic clone that has the entire coding sequence of HSDH was also isolated. The soybean HSDH coding gene was named as ***gmhsdh1,*** and this gene is interrupted by 11 introns.

Introduction

Homoserine dehydrogenase (EC 1.1.1.3) catalyzes the third step of threonine biosynthesis, a NADH- or NADPH-mediated reduction of aspartate semialdehyde to homoserine. HSDH is well characterized in both bacteria and yeast. It operates as a bi-functional enzyme (AK-HSDH) in association with AK in gram-negative bacteria such as ***E.coli***, and as a mono-functional enzyme in gram-positive bacteria and yeast (Cohen and Saint-Giron, 1987). In ***E.coli***, two bi-functional isoforms (AKI-HSDHI and AKII-HSDHII) and one mono-functional isoform of AK (AKIII) were reported. AKI-

HSDHI is encoded by the *thrA* gene and its activity is inhibited by threonine (Truffa-Bachi *et al.,* 1974). AKII-HSDHII is encoded by *metL* gene but not inhibited by the end product (Patte *et al.,* 1967). In *Corynebacterium glutamicum* and *Bacillus subtilis*, gram-positive bacteria, a single HSDH was reported. In both cases, the HSDH activity was inhibited by threonine (Archer *et al.,* 1991, Reinscheid *et al.,* 1991; Parsot and Cohen, 1988). HSDH in yeast is encoded by the *hom6* gene and its activity is not inhibited by the end products of the pathway (Thomas *et al.,* 1993). The first HSDH cloned in plants was that of the carrot and was shown to be a bi-functional AK-HSDH (Weisemann and Matthews, 1993), even though its sensitivity to feedback inhibition was not studied. To date, cloning efforts in plants reveal only bi-functional HSDH. Three different cDNAs of AK-HSDH from maize (Muehlbauer *et al.,* 1994) and two from soybeans (Gebhardt *et al.,* 1993) were also cloned. While all three AK-HSDH of maize were sensitive to feedback inhibition by threonine, the sensitivity of soybean AK-HSDHs was not characterized. In *A. thaliana*, two bi-functional AK-HSDH coding genes sensitive to threonine (Ghislain *et al.,* 1994; Rognes *et al.,* 2002) were reported in addition to three different mono-functional AKs (Frankard *et al.,* 1997; Tang *et al.,* 1997; Yoshioka *et al.,*

2001). Sainis *et al.,* (1981) reported a threonine resistant cytosolic form of HSDH in addition to a sensitive isoform localized in chloroplast from both barley and peas though the gene was not cloned. Moreover, all bi-functional AK-HSDHs and mono-functional AKs reported to date are localized in the chloroplast and are either threonine or lysine sensitive. These observations point to the existence of an HSDH that is resistant to feedback inhibition from threonine. Since over-expression of threonine resistant HSDH would be more advantageous compared to a threonine sensitive form, the cloning of genes coded for a threonine resistant form would be a reliable approach to improving the amount of free threonine in soybean seeds through genetic engineering.

Materials and Methods

Plant Material

Five days after sowing (DAS) seedlings of soybean cv. Williams 82. were used to isolate RNA. Plants of Williams 82. were also grown at the Bradford Research and Extension Center on a Mexico silt loam soil in order to collect developing seeds for expression analysis. Developing seeds collected from the pods of these plants were frozen in

liquid nitrogen and stored at -80^0 C. They were later used for RNA isolation.

Probe Preparation

The total RNA from hypocotyls of 5 DAS seedlings was isolated through a lithium chloride precipitation procedure (Lizzardi, 1983). In order to isolate sequences that would later be used as a probe for library screenings, primers (Forward: 5′ GCT GGT CTT CCT GTG ATA GCA TCA and Reverse: 5′ CCA TTT GAA GCT GCC TTC TCA ACT C) were synthesized on the basis of sequence information obtained from an EST clone (Genbank accession # BM308805). A 394 bp fragment was amplified using these primers through the use of RT-PCR following the protocol given by the manufacturer (Stratagene, La Jolla, CA). This fragment was cloned into a sequencing vector pGEMT-EASY and sequenced with a *Taq* Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) at the DNA core. After confirming the sequence, the fragment was purified by agarose gel electrophoresis and radiolabeled with [^{32}P] dCTP (PerkinElmer Life Sciences Inc., Boston, MA) using a random labeling kit (Takara Bio Inc., Shiga, Japan). This probe was used to screen the soybean seedling cDNA and genomic libraries.

cDNA Isolation and Sequence Analysis

The above-mentioned probe was used to screen a seedling cDNA library that was constructed in λ Zap II by Dr. Joseph Polacco, Department of Biochemistry. Two positive clones were isolated after colony hybridization with a radiolabeled probe following the standard method for three times (Sambrook *et al.,* 1989). One of the two clones was used to recover plasmid, using the Rapid Excision Kit (Stratagene, La Jolla, CA). The plasmid was sequenced with a *Taq* Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) at the DNA core. Analysis of the cDNA sequence was done using different publicly available software tools. BLASTX at National Center for Biotechnology Information (NCBI) site was used for the homology search, CLUSTALW (Thompson *et al.,* 1994) and BOXSHADE (http://workbench.sdsc.edu/) were used for multiple sequence analysis. The CLUSTALW tool was also used for phylogenetic analysis.

Heterologous Expression

The cDNA was used to amplify the coding region of HSDH through PCR using specific primers (Forward: 5′ CATATG ATG AAG AAT ATT CCT CTG ATT CTA ATG G; Reverse: 5′ CTCGAG TCA AGG AAA TAA GTC TTG AAT ATC TAC G). These primers were incorporated with terminal restriction sites *Nde I* and *Xho*

I in N- and C- terminals, respectively, to facilitate cloning into expression vector. The purified PCR product was digested with ***Nde I*** and ***Xho I*** followed by ligation into the corresponding restriction sites of the ***E. coli*** expression vector pET 28(a)+ (Calbiochem– Novabiochem, San Diego, CA, USA) using the ExTaq ligase kit (Takara Mirus Bio). The recombinant vector pHSDH was mobilized into the ER2566 ***E. coli*** strain (New England Biolabs, Beverly, MA, USA) by heat shock treatment at 42 °C for 45 sec. Overnight cultures inoculated with a single colony of the ER2566 ***E. coli*** strain with the pSHSDH plasmid were used to inoculate 100 ml of Luria Broth medium with 100 µg/ml kanamycin. The culture was allowed to grow overnight at 37 °C, with 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) added when the optical density of the culture reached 0.85 (A600). To purify the protein, the cells were harvested by centrifugation at 6,000 rpm for 10 minutes at 4 °C. After removing the supernatant completely, 5ml of buffer B (100 mM NaH_2Po_4, 10 mM Tris-Cl, 8 M Urea, pH 8.3) was added to re-suspend the pellet followed by 10 minutes of incubation at room temperature. After spinning at 12000 rpm for 10 minutes at 4 °C, the supernatant was run through a nickel-nitrilotriacetic acid (Ni-NTA) agarose column. The column was washed with the buffer C (100 mM NaH_2Po_4, 10 mM Tris-Cl,

8 M Urea, pH 6.3) twice with full volume of the column. Initial elution was done with 1 ml buffer E (100 mM NaH_2Po_4, 10 mM Tris-Cl, 8 M Urea, pH 4.5) and the final elution was done with 1 ml 2% SDS. Antibodies of the purified protein were raised in rabbits following the procedure described previously (Krishnan and Okita, 1986).

Expression Analysis

Total RNA was extracted from six stages of developing seeds of soybean cv. Williams 82, following the standard lithium chloride precipitation procedure (Lizzardi, 1983). Quantification of RNA was done by measuring the optical density (A260/A280). Equal amounts of RNA were fractionated in a 1.2% agarose-formaldehyde gel, transferred to a Hybond-N+ membrane and immobilized by UV cross-linking. The cDNA of soybean HSDH isolated from intermediate vector pBS was purified from an agarose gel using Ultrafree-DA columns (Millipore Corporation, Bedford, MA, USA). The purified insert was labeled with [^{32}P] dCTP using the Ladderman labeling kit (PanVera Corporation, Madison, WI, USA). Six hours of prehybridization was performed at 65 $^{\circ}$C in 7% SDS, 191 mM Na_2HPO_4, 58 mM NaH_2PO_4, 1% bovine serum albumin (BSA) and 100 µg/mL denatured salmon sperm DNA. Hybridization was performed overnight using the same buffer at 65 $^{\circ}$C. After hybridization, the membrane was washed with 0.5 × SSC (1 ×

SSC is 150 mM NaCl, 15 mM sodium citrate) and 0.1% SDS at 65 °C and exposed to an X-ray film at -80^0 C. RNA molecular weight markers were used to estimate the sizes of hybridizing transcripts.

Southern Analysis

Soybean genomic DNA was isolated from the leaves of the Williams 82 cultivar using the cetyltrimethyl ammonium bromide (CTAB) method (Saghai-Maroof, 1984). After quantification, 8 µg of DNA was digested along with different restriction enzymes overnight at 37 °C. The digested DNA was electrophoresed on a 0.8% agarose gel and transferred to nylon membranes by a capillary transfer using 0.4 M of NaOH. The same DNA fragment used to screen the cDNA library was labeled with ^{32}P-dCTP using the Ladderman kit (Takara Bio Inc. Otsu, Shiga, Japan) and used as the probe for Southern analysis. After allowing six hours of prehybridization, overnight hybridization was done at 65 °C using 6X SSPE buffer, (1X SSPE is 0.1 M NaCl, 0.01 M Na2PO4, and 0.001 M EDTA), 5X Denhardt's solution, 0.5% SDS, and 50 µg/mL of salmon sperm DNA. Following hybridization, the membranes were washed three times in wash solution I (2X SSPE and 0.5% SDS) for 10 min at room temperature before two washes in wash solution II (0.1X SSPE and 0.1% SDS) at 65 °C for 30 min. After washing, the

nylon membrane was exposed to X-ray film overnight at -80 °C.

Isolation of Genomic Clone

The same probe used to screen the cDNA library was used for screening a commercially available soybean genomic library (Stratagene La Jolla, CA). Seven positive clones were isolated through colony hybridization with a radiolabeled probe. Hybridization was done three consecutive times following the standard method (Sambrook *et al.,* 1989). DNA restriction, followed by southern analysis of one of the seven clones, revealed hybridization with two ***Hind III*** inserts of 3.2 and 4.2 kb and one 0.6 kb ***Eco RI*** insert. These inserts were gel isolated, purified and subcloned into an intermediate vector pBS before sequencing with a ***Taq*** Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) at the DNA core. The genomic sequence and amino acid sequence deduced from the cDNA were used to determine the number and positions of introns and exons in the soybean HSDH gene using the Wise2 tool publicly available at http://www.ebi.ac.uk/Wise2/index.html.

Results

Screening the soybean seedling cDNA library with an EST sequence as the probe revealed a full-length cDNA clone (pHSDH). The putative cDNA clone was sequenced at the DNA core facility and analysis of the sequence revealed a 1472 bp DNA insert containing a single open reading frame (ORF) of 1128 bp. The cDNA has a 90 bp long, 5′ un-translated region (UTR). It also has a 251 bp long UTR on the 3′ end. The physical map of this cDNA clone is shown in Fig. 1. The sequence of soybean HSDH cDNA was submitted to the NCBI database [Genbank # DQ172918]. The predicted ORF encodes a protein of 376 amino acids (Fig. 2) with a molecular weight of 40.6 kD and an isoelectric point of 6.19.

BLAST analysis showed that this sequence has significant identity with mono-functional HSDH from different bacteria and yeast. A pairwise amino acid comparison of the soybean HSDH using the BLASTX tool revealed an 82% identity with that of ***Medicago truncatulata*** [Genbank # AC146743], 66% identity with that of ***Arabidopsis thaliana*** [Genbank # BT004002], 62% with the HSDH of ***Oryza sativa*** [Genbank # DP000011]. Multiple sequence alignment (Fig. 3) showed that the mono-functional soybean HSDH lacks any transit peptide, which is not a common feature of the enzymes in the aspartate pathway. The multiple sequence

alignment also revealed that the 376 aa long soybean HSDH, like that of yeast (359 aa), lacks a C-terminal extension that is present in ***Corynebacterium glutamicum*** and ***Bacillus subtilis,*** both of which are gram-positive bacteria in which mono-functional HSDH is well characterized. Phylogenetic analysis (Fig. 4) revealed that soybean HSDH is more closely related to that of yeast rather than the HSDH of various gram-positive bacteria.

HSDH Genomic Clone

Sequencing of two ***Hind III*** (3.2 kb and 4.2 kb) fragments and one ***EcoRI*** (0.6 kb) fragment of one of the seven clones obtained from genomic library screening revealed the entire coding region of soybean HSDH. Out of the 7997 bp sequenced, the cDNA of HSDH begins at position 1461 and ends at position 7763. The start codon of the cDNA maps to position 1537 and the stop codon is located at 7760 of the genomic sequence. The nucleotide sequence of the genomic clone, along with the deduced amino acid sequence from the full length DNA, was used to determine the position and span of introns and exons. The Soybean HSDH protein coding sequence is interrupted by 11 introns of varying length (Fig. 5). While the length of the shortest intron is 107 bp, that of the longest is 1637 bp. The entire 7997 bp of the genomic sequence with deduced

amino acids of 11 exons is presented here (Fig. 6). The genomic sequence was also submitted to the NCBI database [Genbank # DQ788566].

Heterologous Expression

To characterize soybean HSDH and to isolate the purified protein in adequate quantity, the coding region was subcloned into the pET28a vector and expressed in the ER2566 strain of ***E.coli*** under the T7 promoter. Six histidine residues are introduced to the N-terminus region of soybean HSDH during this process and expression of 6X His-tagged recombinant HSDH was induced by adding 1 mM IPTG to the medium when the optical density of the culture reached 0.85. SDS-PAGE analysis of denatured total soluble and insoluble protein preparation from IPTG induced cultures revealed an abundant 41 kD protein (Fig.7). Owing to the 6X His-tag, the molecular weight of the recombinant protein is slightly higher than the deduced mass from sequence analysis. The recombinant protein was purified by Ni-affinity column chromatography and used to raise antibodies.

Expression Analysis

In order to determine the HSDH gene expression during seed development, northern analysis was performed using RNA isolated from six stages of developing seeds. 5 µg of

total RNA was resolved on formaldehyde/agarose gel and mobilized to a nylon membrane. Hybridization was carried out with a ^{32}P labeled cDNA insert. The soybean HSDH showed temporal expression during seed development (Fig. 8). HSDH transcript was abundant during the mid-stages of seed development and declined during the terminal stages. In contrast to such an expression pattern showed by the HSDH, expression of the 18S RNA was similar at all stages of seed development.

Genomic Southern Analysis

Southern analysis with six different restriction enzymes was carried out to determine the copy number of HSDH in the soybean genome. Autoradiography resulted in more than one band for all the restriction enzymes (Fig. 9). However, except for ***Eco RI*** (lane 3), strong hybridization was seen for one band, while the other bands showed weak hybridization. The obtained results indicate that soybean HSDH is encoded by a low copy gene.

Discussion

The focus of researchers aiming the quality of soybeans has been on the improvement for sulfur amino acids, methionine and cysteine. While these efforts are ongoing, the deficiency of methionine in the soy/corn based

ration is being compensated by synthetic methionine in the animal feed industry. This practice is not only expensive, but also causes other amino acids, such as threonine, isoleucine, tryptophan, valine, and arginine, which are collectively called second-tier amino acids, to be limited in animal feed (Rapp *et al.*, 2003). One way of improving the amount of threonine in soybean seeds is through the over-expression of key enzymes in the threonine biosynthesis pathway. HSDH is the first committed reaction in the branch of the aspartate pathway that leads to the biosynthesis of methionine, threonine, and isoleucine. Hence, the isolation and characterization of the gene coding this enzyme can be helpful in improving the proportion of free threonine found in soybean seeds.

HSDH exists as a bi-functional enzyme (AK-HSDH) in gram-negative bacteria and as a mono-functional form in gram-positive bacteria. The HSDH isolated and characterized from various plants to date are only bi-functional. The coding sequence of the soybean HSDH cDNA showed significant homology with mono-functional HSDH of yeast and gram-positive bacteria. Hence, this study is the first to report cloning of mono-functional HSDH not only in soybeans but also in other plants. Apart from the full-length cDNA, a genomic clone was also isolated that has the entire

coding region of soybean HSDH and this gene is named *gmhsdh*. Multiple sequence analysis of HSDH from soybeans, yeast and gram-positive bacteria (Fig. 3) shows that soybean HSDH cDNA does not have sequences that code for a transit peptide. Though not characterized, sequence information of HSDH in *Arabidopsis*, *Medicago*, and *Oryza* species is available in Genbank. From the multiple sequence analysis including those sequences (Fig. 3), it can be observed that the sequence coding for transit peptides is not available in HSDH of these three species also. So, it is speculated that soybean mono-functional HSDH is cytosolic. Sainis *et al.,* (1981) also reported a threonine resistant cytosolic form of HSDH in addition to a threonine sensitive isoform localized in chloroplast from both barley and pea. However, there have been no reports elucidating the function of such a cytosolic HSDH in amino acid biosynthesis. In plants, all the enzymes of aspartate pathway have a chloroplast transit peptide and the function of these enzymes in biosynthesis of amino acids is associated with their localization to chloroplast. Though this study resulted in the cloning of a mono-functional and possibly cytosolic HSDH, the function of cytosolic HSDH still remains to be seen.

One major difference between mono-functional HSDH of yeast and those of gram-positive bacteria is that the yeast HSDH is resistant to feedback inhibition. HSDH of *Corynebacterium* (445 aa) and *Bacillus* (433 aa) contain the C-terminal extension, which is implicated in feedback inhibition by threonine. A single amino acid (Gly 378) in HSDH of *Corynebacterium* has been identified as the principal residue involved in interaction with threonine (Reinscheid *et al.,* 1991). This residue is conserved across all gram-positive bacterial HSDHs, but is absent in yeast (Thomas *et al.,* 1993). The soybean HSDH also lacks this C-terminal extension along with the glycine residue, causing speculation that it is also a threonine resistant form. However, enzyme analysis needs to be done in order to determine exactly whether the soybean HSDH is feedback resistant or sensitive.

Expression analysis of soybean HSDH in six stages of developing seeds showed temporal expression. Expression of this RNA was lower in early and terminal stages but was abundant in the middle stages. Since mono-functional HSDH has not been studied in plants before, this expression pattern cannot be compared with any other study. Though more than one band resulted from southern analysis of the soybean genomic DNA with the HSDH probe, development of the

weak bands may be caused by the hybridization of the probe with related genes in the genome. There are at least two genes coding for bi-functional AK-HSDH genes in the soybean genome (Gebhardt *et al.,* 1993), so there is a good possibility that such hybridization can occur. Hence, it can be proposed that HSDH exists as a single copy gene in soybeans. Together with the two different genes coding for bi-functional AK-HSDH from soybeans isolated previously (Gebhardt *et al.,* 1993); there are at least three genes in the HSDH family in the genome of the soybean.

Over-expression of regulatory enzymes is achieved by either deactivating the portion of the enzyme that mediates feedback inhibition, or by isolating feedback resistant forms and expressing them in the target tissues without going through the tedious process of identifying and deactivating the feedback inhibition mediating region of the enzyme. Thus, feedback resistant forms have an advantage over feedback sensitive forms. Since the mono-functional HSDH reported here seems to be feedback resistant, it can be advantageous in manipulating the aspartate pathway for an increase in both threonine and methionine. In addition, biosynthesis of isoleucine, another second-tier amino acid can also be manipulated.

The findings of this research are significant in that a full-length soybean cDNA that encodes mono-functional HSDH has been isolated for the first time in the plant kingdom. The structure of the gene coding mono-functional HSDH has also been elucidated. Though it needs to be experimentally confirmed, the sequence analysis of soybean HSDH indicates that this can be a cytosolic and feedback resistant form.

Fig. 1. A partial restriction map of soybean cDNA encoding the mono-functional HSDH. The black box indicates the location of the open-reading frame (ORF).

.

```
1     gaattcgccacgagtaaaaccctagtaaatccaattggggatttttttttctagaagag  60
                                  M  K  N  I  P  L  I  L  M  G
61    gattgaaagtgaaaaaaaggaaataacaaaatgaagaatattcctctgattctaatgggt  120
      C  G  G  V  G  R  Q  L  L  Q  H  I  V  S  C  R  S  L  H  F
121   tgtggaggagttggtcgtcaacttctccaacacattgtctcgtgtcgttctcttcacttt  180
      T  Q  G  L  C  L  R  V  V  G  V  G  D  S  K  S  L  V  V  T
181   acacaggggctgtgcttgagagttgtaggagtcggtgatagtaaatctttggtggttacg  240
      E  D  L  L  H  E  G  L  N  D  G  F  L  L  E  L  C  R  V  K
241   gaggatttgctgcatgaggggttgaatgatggcttcttgttagaactttgccgggtcaag  300
      S  V  G  E  S  L  L  K  L  L  D  F  G  K  C  Q  A  F  V  H
301   agtgtcggcgaatctctattaaaacttcttgattttgggaaatgccaggcatttgtgcat  360
      P  E  S  Q  G  K  I  L  E  I  A  F  Q  L  G  K  K  T  G  L
361   ccggagtcacaaggaaagattctagagattgcatttcaacttggtaaaaagacaggtttg  420
      V  F  V  D  C  S  A  S  S  D  T  V  A  V  L  K  Q  A  I  D
421   gtatttgtagattgctctgctagctctgacactgttgcggtgctaaagcaagcgattgat  480
      M  G  C  C  A  V  M  A  N  K  K  P  L  T  S  T  M  E  D  F
481   atgggttgttgtgctgttatggcaaataagaagcctcttacatctacaatggaggatttt  540
      K  K  L  F  I  Y  P  R  R  I  R  H  E  S  T  V  G  A  G  L
541   aaaaaacttttcatctatccacgtcgtattcggcatgagtcaactgtaggcgctggtctt  600
      P  V  I  A  S  L  N  R  I  I  S  S  G  D  P  V  H  Q  I  I
601   cctgtgatagcatcactgaatcgcataatctcttctggtgatcccgttcatcaaattatt  660
      G  S  L  S  G  T  L  G  Y  V  M  S  E  V  E  D  G  K  P  L
661   gggagtttgagcgggacattgggttatgtaatgagtgaggttgaagatggaaagccattg  720
      S  Q  V  V  R  D  A  K  S  L  G  Y  T  E  P  D  P  R  D  D
721   agccaagttgttagagatgctaaaagtttggggtatactgaaccagatccacgtgatgac  780
      L  G  G  M  D  V  A  R  K  A  L  I  L  A  R  I  L  G  H  Q
781   cttggtggaatggacgtcgctagaaaggctttaatacttgctcgaatacttggccatcaa  840
      I  N  L  D  S  I  Q  I  E  S  L  Y  P  K  E  M  G  P  G  V
841   attaatctggatagcattcagattgagagcttgtatcccaaagaaatggggcctggcgta  900
      M  T  V  E  D  F  L  N  R  G  L  L  L  L  D  K  D  I  Q  D
901   atgactgttgaagacttcttgaaccgtggactcttgttgcttgataaagatattcaagat  960
      R  V  E  K  A  A  S  N  G  N  V  L  R  Y  V  C  V  I  M  G
961   agagttgagaaggcagcttcaaatggaaatgtgttgcgctatgtctgtgtgatcatgggt  1020
      S  R  C  E  V  G  I  Q  E  L  P  K  N  S  P  L  G  R  L  R
1021  tcaaggtgtgaagttggtattcaggagcttccaaagaattctcccttgggaagactaaga  1080
      G  S  D  N  V  L  E  I  Y  T  R  C  Y  S  K  Q  P  L  V  I
1081  ggaagtgataatgtgttggaaatatatactcgatgctatagtaaacaaccactggttatt  1140
      Q  G  A  G  A  G  N  D  T  T  A  A  G  V  L  A  D  I  V  D
1141  caaggtgctggagctggaaatgacactactgctgctggggttcttgctgatatcgtagat  1200
      I  Q  D  L  F  P  *
1201  attcaagacttatttccttgaagggaaatttttgtgagccatcctaattaaagaatgtag  1260

1261  aaggtacagctgctaaaggtcttgccctggattgtagctgatgcgcgtttctctttggtt  1320

1321  gtgcattgtgaaagattggatacacattgttctgaagcaagaataaataagtagctttta  1380

1381  tgtttttttgggatatttgacatgaaatttgtgtaaacgtccatgcagaaccagacacact  1440

1441  gaaataaaatgaattctgctcgtgccgaattc  1472
```

Fig. 2. Nucleotide sequence and deduced amino acid sequence of soybean HSDH cDNA. The sequenced region covers 1472 nucleotides. The ORF for HSDH begins at position 91 and ends at position 1222 encoding a 40.6 kD protein.

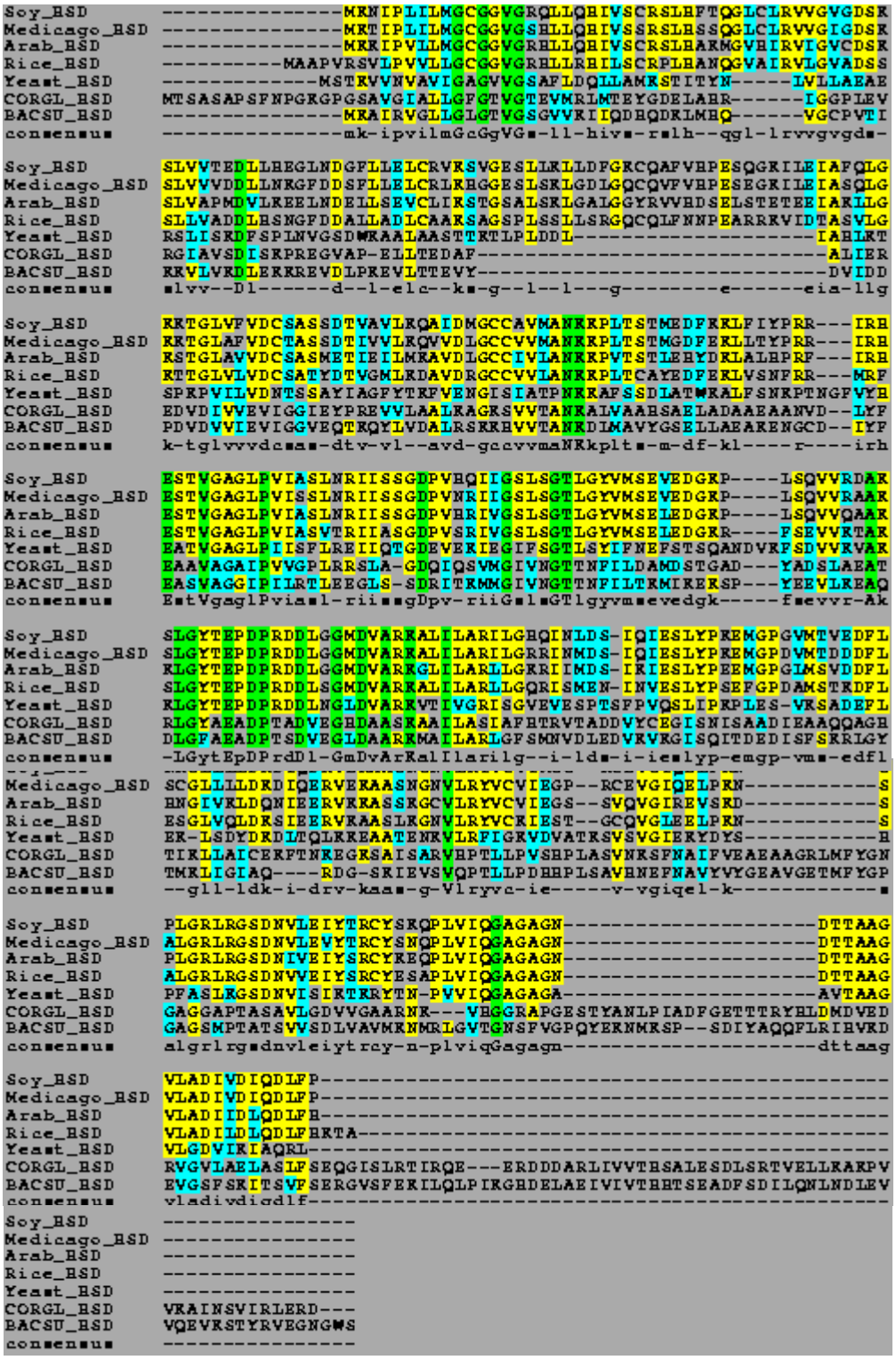

Fig. 3. Multiple sequence analysis of mono-functional HSDH from Soybean, *Medicago*, *Arabidopsis*, Rice, Yeast, *Corynebacterium glutamicum* and *Bacillus subtilis*. Completely conserved residues are in green background color, identical residues are in yellow color, and similar residues are in cyan color.

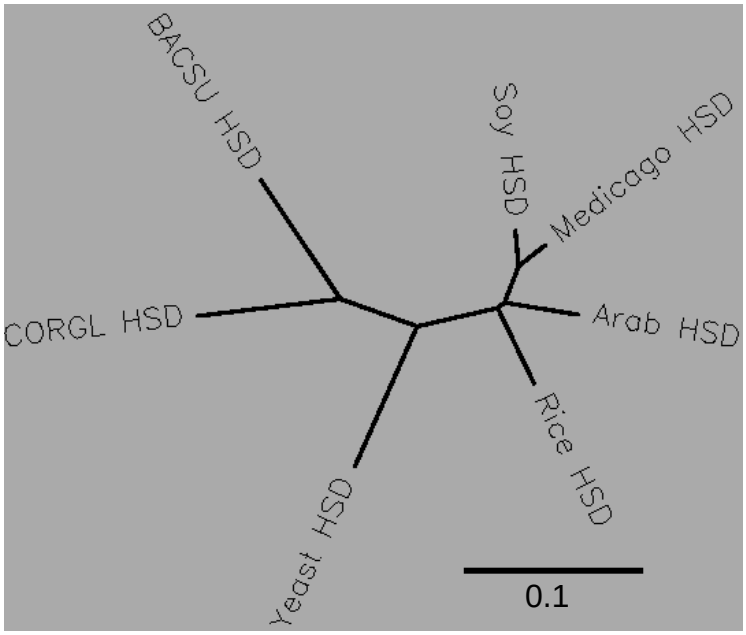

Fig. 4. Phylogenetic analysis of mono-functional HSDH from plants, yeast and gram-positive bacterial species. CLUSTALW software was used to draw the unrooted phylogenetic tree.

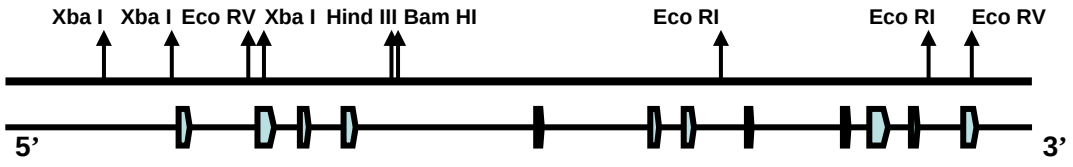

Fig. 5. A partial restriction map of soybean HSDH genomic clone and gene structure. The boxed portions indicate exons and the connecting lines show introns. 5’ end and 3’ end are labeled.

```
1     aagcttctaaaatttttttgaaggaagagagagaagatttgggcttttaaaaaaaatgat  60
61    tttcttttcttttttcattttcttttctaaaagcaaatgccacatgtcctcttttaagtgg  120
121   agcaaaagggcccacctttttccttttgattagacatcattctcagccataaggaagaaaa  180
181   aattgaaccttttttggatgctgaaatcctgcctcagtctgcgtgccgcctctctggttcc  240
241   agttcctcgcgtttctctgcacccgtcggggcccgttttcaaaagtaggcaatatatata  300
301   atatcaaaacgctcagaatgagaccctgagcgtggttcaaaggttggttttgttaaattt  360
361   taagttgcacgcaaaatgattatttttagactaattaattgcgaattactctataaccgt  420
421   tcggttatgggttactctttgttaattagtctaacccgcgtatctttcctccaatgtaca  480
481   tacttctaccaagaatatatatatataattatttagatgtaacacttacaaaatttcgag  540
541   tagaaattctaggatgtcacacatctggcattgggggtctctttcgagatacctcgagtc  600
601   aatggctgcttggttattcaggaaattgtggctatacaacaatcatgaatgcgaaattac  660
661   aaggaatttataatggccttatcttaaattggaatgttggttataagcaagtaatttgtg  720
721   aaactaattctttattaacactcaaacttataacatgaactatttttctttggcaccctc  780
781   atggttgtgtcataaatcacatcaaaaggttaatggagaaggaatgacaaatcactttca  840
841   ctcacaccttgtgagaaggaaaccaacgtgtggattggttggcaaaacatgggagccacc  900
901   aacattgctccttagtagtattggattcttgtccctcgcctatagctcccttactgcaag  960
961   ctaatgctatgcagatagccttctctagaagttaatttgttatgtaagttttacatgata  1020
1021  aaatatatatatatataattataattgtatcaacactattatttatgtataaaaaattta  1080
1081  catattatgaattatttaatatttttttaatatagattgagttatgttgatcaaccatta  1140
1141  acttgttgggttaatttttgacccaatgattaaatatcttaactaaatcaataatcgaca  1200
1201  tgagtgacataaatataatttttaggtttggtaaggaatgagagggagagtgaaagaattt  1260
1261  ccaattgcaccttacttttattttgttttaaatcaataattacttttttatttttaattt  1320
1321  actgagttgtcaaattttcattggatgtgtgcaaatctggcatgttaggttatgctgtac  1380
1381  gcaagattaactcagagtttcatgtttcacagtcccatccctgctcacactgctccagtg  1440
1441  ctacttcctcttccatcacataaaaccctagtaaatccaattggggatttttttttctta  1500
                                          M  K  N  I  P  L  I  L
1501  gaagaggattgaaagtgaaaaaaaggaaataacaaaatgaagaatattcctctgattcta  1560
      M  G  C  G  G  V  G  R  Q  L  L  Q  H  I  V  S  C  R  S  L
1561  atgggttgtggaggagttggtcgtcaacttctccaacacattgtctcgtgtcgttctctt  1620
      H  F  T  Q
1621  cactttacacaggtaatgtccctttagtttcacggtttttgattcttatgttttcagatg  1680
1681  atgttttatcttccttatttgtcgttgctgtaacctgaaaatgaaaactttgtggtttct  1740
1741  gataacccatttgtcttttgaacatttttttatgggttattgtttgggacgtacgtttgc  1800
```

Fig. 6. Sequence of soybean HSDH genomic clone with deduced amino acid residues of exons.

```
1801  tatgccaaatgattctaaggattaagcactgtgaaagatttgatttttttttttcagacat  1860

1861  tgatggaaaatgggttcataaaatagagtgatgtttagttggttgatgtttttatttttg  1920
                          G  L  C  L  R  V  V  G  V  G  D  S  k  s
1921  ttatgtattttggtggcaggggctgtgcttgagagttgtaggagtcggtgatagtaaatc  1980
       L  V  V  T  E  D  L  L  H  E  G  L  N  D  G  F  L  L  E  L
1981  tttggtggttacggaggatttgctgcatgaggggttgaatgatggcttcttgttagaact  2040
       C  R  V  K  S  V  G  E  S  L  L  K  L  L  D  F
2041  ttgccgggtcaagagtgtcggcgaatctctattaaaacttcttgattttggttgatatct  2100

2101  atgcctaacttgtctaggtggtgatatgaaaatttgaatgtctttcttaatgcttacttg  2160
                       G  K  C  Q  A  F  V  H  P  E  S  Q  G  K  I
2161  ttagatgttatttttaagggaaatgccaggcatttgtgcatccggagtcacaaggaaagat  2220
       L  E  I  A  F  Q  L  G  K  K  T  G
2221  tctagagattgcatttcaacttggtaaaaagacaggtatgctgaatttttgttatgtcgt  2280

2281  gcattacatgcctcctggggttgattttaaaatttaaagcaaattcaatataagagtgag  2340

2341  cctacatgccttttcaatcactgagactgatgttatctctgaatatacttttttttggtta  2400
            L  V  F  V  D  C  S  A  S  S  D  T  V  A  V  L  K  Q  A
2401  taggtttggtatttgtagattgctctgctagctctgacactgttgcggtgctaaagcaag  2460
         I  D  M  G  C  C  A  V  M  A  N  K  K  P  L  T  S  T  M
2461  cgattgatatgggttgttgtgctgttatggcaaataagaagcctcttacatctacaatgg  2520

2521  taagatcatcttaaccgttttataataattgttgaaatgggagagtaggggatacatact  2580

2581  tttttgtaaaatcaagaaaaactatttattcatgattttatttttaccttctttattcaac  2640

2641  ccttggctctgatatagacataataagaaagaaaatgtccaatttattgcctccaagtta  2700

2701  cggccatgtacattttggtcatgaaacttaaaaaaagtgctgtctgatccttaaactctc  2760

2761  taatttgttcatttttatccttctgtttgtttattgtcaaattctaacggaagttgtcat  2820

2821  tgtggaatcccattttcaaacttcaagactgttcgctcatgacccttttgtaaagtggact  2880

2881  tttcatgttatgaaacttggtacaagcaatggtcacataatatgggcatgacataatatc  2940

2941  agtgtgatgaaaaatcaaatgaaagggcaaaacacgcattagaaagtttagcaattcttt  3000

3001  gtttaggtacttttcctgttataattatgggttataattagatgtatctttagaatagat  3060

3061  ttatctgtttttatcttgttgtagaagacaacagtgtgtcttgtcttgtcaaagccagcc  3120

3121  aagctgtctctctccacattgtgtatatatatgttatttgaatgaaataaaagctaagct  3180

3181  agtgagtgaacagttttctcctcacaaattcaaagcttctcctcaagattatcatccttc  3240

3241  ttcaatcaatcctcaagtcatggatcctgctacctctcattcctctgattcagattgtcc  3300

3301  cattgccatccaaaaagatactcgatccactcgtaatcctcatcctatctataactttct  3360

3361  gagttatcattgtttgtctccttcatattcttcctttgttttctctctatcttcgcattc  3420

3421  tgtccctgctaatattcatgaggcacttagtcatcctggatggcgacaggttatgattga  3480

3481  tgaaatgcaggctctcgagcatagtggtacttgggaacttgtccctcttccacctggcaa  3540

3541  gaagactgtgggttgtagatggttctacgcagtaaaagttggacctaatggtgagattga  3600
```

```
3601  tcgacttaaggctcgattagtggctaaaggctatactcagatatatggccttgattattg  3660

3661  tgacaccttctcttcggtagctaagatcactactgtccgtctgtttcttgctatggctgc  3720

3721  tatgcgtcattggcctctccatcagcttgatattaaaaatgcattccttcatggtgacct  3780

3781  tgaggaggaaatttatatggagcaacctcctgggtttgttgctcaggggagtatggcct  3840

3841  tgtttgtaagctatgtcgatctctctatgggttgaagcaatcccctcgggcttggtttga  3900

3901  taaattcagtcatattgttcaacttttgggttgaaacggagtgaggctgatcattatgt  3960

3961  ttttgattgtcattcatcccctgggaaatatgtttatttgatagtatatgttgataataa  4020

4021  gtatcactgctgtctttgtttaggtacttttccctaataaggtttcataagttttggtt  4080

4081  tcttccctgtcactagtttcttttctggtagttatgcttgcttcaaatactaaataatac  4140
                      E   D   F   K   K   L   F   I   Y   P   R   R   I   R   H
4141  atcttatatgtaacaggaggattttaaaaaacttttcatctatccacgtcgtattcggca  4200
       E   S   T
4201  tgagtcaactgtaagtgatagttatttcttgccttaactagagtatatcactttatcagc  4260

4261  atattaatagcaatgcaacatttgtgacggtatggaaaatagttattgaaatgtgtattt  4320

4321  tcaacaaactggaatataaaatgctgcattattgcttgaatttctttgcttggttatgct  4380

4381  attccttgaactctgatctccttgtaaccaaattagagcctatagaagttctgaatgtct  4440

4441  tgtatggatcaataatctacaagctgttgctttgcttagattctttaagcaatctttacc  4500

4501  ttgataaagaaggaaaatgaagtccatggaaaagctgaaccgcatgtggtgtggtttgcg  4560

4561  agaagtagcttgatggaatgacttagtataatctcatcttccccaataaagtaaggaggg  4620

4621  gaagatcatgaatttgaggtgtgtagatgacttatgattatggaaatctaaatatgaaga  4680

4681  aattctataaagaaatgctttaacacatactagcatccatgaagaaattattctaattac  4740

4741  tcatttcagtgtgaaaccaatacccaagaagccattgctccagaaaggaaaaacacagaa  4800

4801  gcagaaacaagaaccatgacaatataatatttgagacacatataaacctcccttgcagtc  4860

4861  cacactgaaaattttcagtccctttctccattcctctccttacctttaacctgttact  4920

4921  caaattgattccactacctgtcacatgggcaggattttacaacttacaaatttacaaact  4980

4981  gttttccctgctcatgctctcctcattgtacacttgtgtgcaatggatcacctgccactc  5040

5041  ttccatttatctatgcggtacaccttagatataggagagtctgtcatattgttattttat  5100

5101  ttaaattggaatattgaattttcaatggattgatgaactattattgaactggttattata  5160
                V   G   A   G   L   P   V   I   A   S   L   N   R   I   I   S   S
5161  tttttacaggtaggcgctggtcttcctgtgatagcatcactgaatcgcataatctcttct  5220
      G   D   P   V   H   Q   I   I   G   S   L   S
5221  ggtgatcccgttcatcaaattattgggagtttgagcggtatatatatctttcatgatcct  5280

5281  ataattacatatttttttggttttccatcatgcattcttgttaggaagaagaaattgtat  5340
```

```
                                                  G  T  L  G  Y  V  M  S
5341  ttttgatgttgtattttatgctacttgctcttattagggacattgggttatgtaatgagt  5400
      E  V  E  D  G  K  P  L  S  Q  V  V  R  D  A  K  S  L  G  Y
5401  gaggttgaagatggaaagccattgagccaagttgttagagatgctaaaagtttggggtat  5460
      T  E  P
5461  actgaaccaggttagtgaatgctacatttgaaattgtattgttgacatgtgaaactcttt  5520

5521  gatctgtgtaatttctatttctgttttacgtcaacctttttaatgtggacatattttct  5580

5581  atgtatcagtattagtgggctttggcttgtactagtactagtgtttgacctaattattag  5640

5641  catatataggaccattgaatcattattttgtaccacattatcaacacaccatatcaatct  5700

5701  aattattattcattgtctttatcctcatgtttttccttccctacctaaatactatgattc  5760

5761  caatttccaacaattttaagaattccaagtccaacatttaattgacatctacttttttcct  5820

5821  ttaaacctttctatttgaatgaggtatttacattttaaaatagattcttagtttgggcat  5880

5881  tggcatgattaatcttcatgttcttgctcattcccctcttgggattttgctaattgctaa  5940

5941  acagtttgttaatgataacttatcttttaggtgtttagatgatttatacatactatcctt  6000
            D  P  R  D  D  L  G  G  M  D  V  A  R  K
6001  tattgcagatccacgtgatgaccttggtggaatggacgtcgctagaaaggttcaatgctg  6060

6061  tttgttttatgtctgaaaattaattttgtatctgattctgctctatgtaaaagctcataa  6120

6121  tgtcttaaaatcaacagaatgctttctggtctcataagaacactgcacataagtgaagca  6180

6181  taaacttaaatgatgtccatttttggccaccaatgaatgatgcccatttttttataatagga  6240

6241  atatcagtttatccttgtagaaactaatctaagtttctttggacaaagtttgtaaaattc  6300

6301  atgtatgtccatttgatgattttggttatgaaaaatataggccaaattaaaattttggtc  6360

6361  tcccttcctaatccgcgattttggtacccatttttttcttgcgattttggtccccatatt  6420

6421  ttggaaaatctgcaatttggtccaatcctcagttttgcatttttttttttggtccaattta  6480

6481  atgctggatctaacatattcacttaagataccataaaaaattatttaattaattaaaata  6540

6541  ttaaaaaaagaaatgcatgcaaaattaaagattggacgaaaattgctgatttttaaaata  6600

6601  gggggaccaaaattgcaaaaataaaatcaggggagcaaaatcacggattaagaattatag  6660

6661  ggggaccaaaattgcaattgagccaaaaatataatgtgttgttgaaagaagtgttacacc  6720

6721  atgtgatgctaatggtggagtatgagttttcatttcttgggttgtgatattttacttttat  6780

6781  caacttttaattcttatgacctttagtaaaattttgtaatactgttggttgtgtactata  6840
                                                  A  L  I  L  A  R
6841  aatgtgctaatactaacttttgtgtttgggccaatatgacaggctttaatacttgctcga  6900
      I  L  G  H  Q  I  N  L  D  S  I  Q
6901  atacttggccatcaaattaatctggatagcattcaggtttgtatgggcatttattatatt  6960

6961  ccatttatgtgtccctttttcttttttctgtcaaaattatctctatgttattagttatatttt  7020
```

```
                                I  E  S  L  Y  P  K  E  M  G  P  G  V
7021  ataattttatatgctttttaaagattgagagcttgtatcccaaagaaatggggcctggcg  7080
        M  T  V  E  D  F  L  N  R  G  L  L  L  L  D  K  D  I  Q  D
7081  taatgactgttgaagacttcttgaaccgtggactcttgttgcttgataaagatattcaag  7140
        R  V  E  K  A  A  S  N  G  N  V  L  R  Y  V  C  V  I  M  G
7141  atagagttgagaaggcagcttcaaatggaaatgtgttgcgctatgtctgtgtgatcatgg  7200
        S  R  C  E
7201  gttcaaggcatcttcgtttccctccatttgcacatatatttttctatttgttcattttca  7260

7261  attcaatcttcgtataggatttggtgatgtttattgtttaaaattattcatatccaatat  7320
                                                              G  I
7321  tgtggttctttctatgctaattataacttttttttctcttcataggtgtgaagttggtat  7380
       Q  E  L  P  K  N  S  P  L  G  R  L  R  G  S  D  N  V
7381  tcaggagcttccaaagaattctcccttgggaagactaagaggaagtgataatgtggtaag  7440

7441  cttcttttgctttctcaatgtttgtgttatgtatttttatttagtatttttttccattta  7500

7501  aagtcatgactttcaccatatgcttttgtcatgcgtcttgtttgctttgctatattgcaa  7560

7561  aagttcagctctcctaagtaataacttcaaaggcctgactgtattctattttattgtatg  7620
                       L  E  I  Y  T  R  C  Y  S  K  Q  P  L  V  I
7621  ttgatacaaatggcagttggaaatatatactcgatgctatagtaaacaaccactggttat  7680
       Q  G  A  G  A  G  N  D  T  T  A  A  G  V  L  A  D  I  V  D
7681  tcaaggtgctggagctggaaatgacactactgctgctggggttcttgctgatatcgtaga  7740
       I  Q  D  L  F  P
7741  tattcaagacttatttccttgaagggaaatttttgtgagccatcctaattaaagaatgta  7800

7801  gaaggtacagctgctaaaggtcttgccctggattgtagctgatgcgcgtttctctttggt  7860

7861  tgtgcattgtgaaagattggatacacattgttctgaagcaagaataaataagtagctttt  7920

7921  atgtttttgggatatttgacatgaaatttgtgtaaacgtccatgcagaaccagacacac  7980

7981  tgaaataaaatgaattc  7997
```

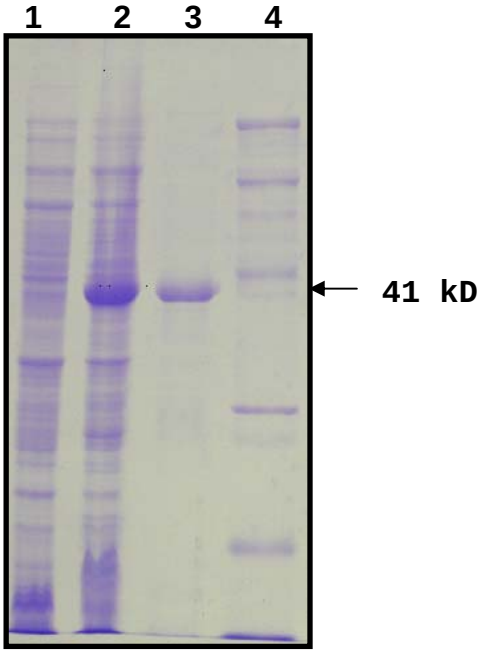

Fig. 7. Heterologous expression of soybean HSDH in *E.coli*.
1. uninduced protein 2. Induced 3. Induced and purified
4. Protein Marker

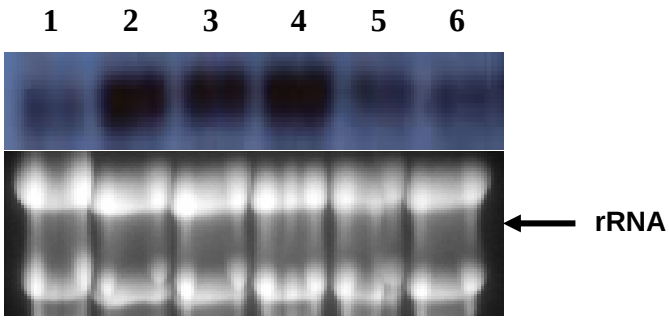

Fig 8. Expression analysis of soybean HSDH in developing seeds. Lanes1-6 represent developmental stages 1-6.

Fig. 9. Southern analysis of soybean HSDH. 1.*Bam HI* 2. *Dra I* 3. *Eco RI* 4. *Eco RV* 5. *Hind III* 6. *Xba I*

References

Archer JAC, Solow-Coredero DE, Sinskey AJ (1991) A C-terminal deletion in *Corynebacterium glutamicum* homoserine dehydrogenase abolishes allosteric inhibition by L-threonine. Gene 107: 53-59

Cohen NG, Saint-Giron I (1987) Biosynthesis of threonine, lysine and methionine. In: Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE (eds) *Escherichia coli* and *Salomonella typhonium*: Cellular and Molecular Biology , Vol I. American Society for Microbiology, Washington DC, pp 429-444

Frankard V, Vauterine M, Jacobs M (1997) Molecular characterization of an *Arabidopsis thaliana* cDNA coding for a mono-functional aspartate kinase. Plant Mol Biol 34: 233-242

Gebhardt JS, Weisemann JM, Matthews BF (1993) Molecular analysis of the aspartate kinase-homoserine dehydrogenase gene family in soybean. Plant Physiol (Suppl) 102: 69

Ghislain M, Frankard V, Vandenbossche D, Matthews BF, Jacobs M (1994) Molecular analysis of the aspartate kinase –homoserine dehydrogenase gene from *Arabidopsis thaliana.* Plant Mol Biol 24: 835-851

Krishnan HB, Okita TW (1986) Structural Relationship among the Rice Glutelin Polypeptides. Plant Physiol 81: 748–753

Lizzardi PM (1983) Methods for the preparation of messenger RNA. Methods Enzymol 96: 24–38

Muehlbauer GJ, Somers DA, Matthews BF, Gengenbach BG (1994) Molecular genetics of the maize (*Zea mays* L.) aspartate kinase-homoserine dehydrogenase gene family. Plant Physiol 106: 1303-1312

Parsot C, Cohen GN (1988) Cloning and nucleotide sequence of the *Bacillus subtilis hom* gene coding for homoserine dehydrogenase. Structural and evolutionary relationships with *Escherichia coli* aspartokinases-homoserine dehydrogenases I and II. J Biol Chem 263: 14654-14660

Patte JC, Le Bras G, Cohen GN (1967) Regulation by methionine of the synthesis of a third aspartokinase and of

a second homoserine dehydrogenase in *Escherichia coli* K 12. Biochem Biophys Acta 136: 245-257

Rapp BW, Weaver L, Wang Q, Crow L, Ream J, Hill B, Brown W, Oulmassov T, Gruys K (2003) Enhancing the nutritional value of soybean. Plant Biology 2003, Annual meeting of the American Society of Plant Biologists. Abstract 38001

Reinscheid DJ, Eikmanns BJ, Sahm H (1991) Analysis of a *Corynebacterium glutamicum hom* gene coding for a feedback-resistant homoserine dehydrogenase. J Bacteriol 173: 3228-3230

Rognes SE, Dewaele E, Aas SF, Jacobs M, Frankard V (2002) Transcriptional and biochemical regulation of a novel *Arabidopsis thaliana* bi-functional aspartate kinase-homoserine dehydrogenase gene isolated by functional complementation of a yeast *hom6* mutant. Plant Mol Biol 51: 281-294

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA sepacer-length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81: 8014-8019

Sainis JK, Mayne RG, Wallsgrove RM, Lea PJ, Miflin BJ (1981) Localization and characterization of homoserine dehydrogenase isolated from barley and pea leaves. Planta 152: 491-496

Sambrook J, EF Fritsch, Maniatis T (1989) Molecular cloning: A laboratory manual, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Tang G, Zhu-shimoni JX, Amir R, Zchori IB-T, Galili G (1997) Cloning and expression of an *Arabidopsis thaliana* cDNA encoding a mono-functional aspartate kinase homologous to the lysine-sensitive enzyme of *Escherichia coli*. Plant Mol Biol 34: 287-294

Thomas D, Barbey R, Surdin-Kerjan Y (1993) Evolutionary relationships between yeast and bacterial homoserine dehydrogenases. FEBS 323: 289-293

Thompson JD, Higgins DG, Gibson TJ (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap

penalties and weight matrix choice." Nucleic Acids Res 22: 4673-4680

Truffa-Bache P, Veron M, Cohen G (1974) Structure, function, and possible origin of a bi-functional allosteric enzyme, *Escherichia coli* aspartokinase I-homoserine dehydrogenase I. CRC Crit Rev Biochem 2: 379-415

Weisemann JM, Matthews BF (1993) Identification and expression of a cDNA from *Daucus carota* encoding a bifunctional aspartokinase-homoserine dehydrogenase. Plant Mol Biol 22: 301-312

Yoshioka Y, Kurei S, Machida M (2001) Identification of a mono-functional aspartate kinase gene of *Arabidopsis thaliana* with spatial and temporally regulated expression. Gene Genet Syst 76: 189-198

CHAPTER 3

ISOLATION AND CHARACTERIZATION OF THREONINE SYNTHASE

Abstract

Threonine synthase (TS) catalyzes the last step of threonine biosynthesis. TS competes with cystathionine γ-synthase (CGS), the first committed enzyme specific to methionine biosynthesis, for the branch point substrate L-homoserine phosphate (HSerP). To date, the only known way of regulating TS in plants is an allosteric activation by S- adenosylmethionine (SAM), a product of methionine. Thus, accumulation of methionine leads to an activation of TS, which in turn results in a flow of HSerP into the threonine specific pathway. The cDNA coding this enzyme was isolated in *Arabidopsis* and the mechanism of activation was well studied. Isolation of the TS coding region, along with the promoter sequence in the soybean may not only help in exploring other possible regulatory mechanisms but also in over-expression of the enzyme in seeds. In this study, a complete genomic clone was isolated and sequenced. The clone was shown to contain the entire coding sequence of TS. The coding region showed significant identity with

other plant and bacterial counterparts of soybean TS. Northern analysis revealed a temporal expression of TS in developing seeds and Southern analysis indicated the existence of TS as a low copy gene in the soybean genome. Along with the coding region, 1096 bp upstream to the translation start codon was also sequenced and deposited to the GenBank. The coding sequence of TS can be helpful in determining the possibility of increasing threonine proportions in soybean seeds without a reduction in methionine. The promoter region could be useful to study the possible regulation of TS through methods other than activation by SAM.

Introduction

In plants, threonine is synthesized from homoserine phosphate, not from homoserine as in bacteria (Umbarger, 1978). Thus the last step in the biosynthesis of threonine is different between bacteria and plants. Threonine synthase (TS; EC 4.2.3.1) catalyzes the last step of a pyridoxyl phosphate dependent conversion of L-homoserine phosphate (HSerP) and homoserine to threonine. Because of the difference in the substrate catalyzed by this enzyme, plant TS is unique in many ways compared to the bacterial counterparts. Laber ***et al.,*** (1999) proposed major differences in their studies with recombinant ***Arabidopsis***

TS. First, in plants, TS exists as a dimer. It exists as a monomer in *E.coli* and yeast. Second, TS activity is increased by S- adenosylmethionine (SAM) in plants but not in bacteria. Third, HSerP, which is a substrate for plant TS, inhibits bacterial TS.

HSerP is also a substrate for cystathionine γ-synthase (CGS), the first committed enzyme in the pathway of methionine biosynthesis (Fig.1). Unlike HSDH and DHPS, the branch point enzymes of Thr/Ile or Met and Lys pathways, TS and CGS are not regulated by a feed back inhibition of the end products (Kreft *et al.,* 1994; Curien *et al.,* 1996), but by an activation of TS by SAM, a downstream product of methionine (Madison and Thompson, 1976). Curien *et al.,* (1998) proposed that the accumulation of methionine and thereby SAM results in the activation of TS, which directs homoserine phosphate into the threonine pathway. Allosteric activation of TS by SAM is of interest for two reasons. First, SAM has been reported to activate only one other enzyme, the mammalian cystathionine β-synthase (Roper and Kraus, 1992). Second, while stimulation of an enzyme by the product of another pathway is specific to plants, only a few such enzymes exist in the plant kingdom. One example is the activation of chorismate mutase by tryptophan (Schmidheini *et al.,* 1990).

Recent analyses using mutants and antisense constructs of TS and CGS clearly indicated competition between these two enzymes for the same substrate (Gakiere *et al.,* 2000; Kim and Leustek, 2000; Bartlem *et al.,* 2000; Zeh *et al.,* 2001). With this background knowledge, over expression of TS in the soybean would not be beneficial, given that methionine is already a limiting amino acid in the soybean and that efforts are currently underway to improving methionine in soybean seed proteins. However, Muhitch (1997) expressed bacterial TS in tobacco and found a 5-fold increase in threonine levels with no reduction of methionine. Also, the regulation of metabolic flux through the methionine pathway is more complex than previously thought, as post-transcriptional and post-translational mechanisms are involved (Amir *et al.,* 2002). So it would be advantageous to isolate TS coding gene in soybeans and over-express in developing seeds. Though TS has been well characterized in other plants, its promoter region has not been isolated and characterized. So the objective of this study is the isolation of the gene coding TS in soybeans and also a characterization of the promoter region.

Materials and Methods

Plant Material

In order to isolate the RNA used for probe preparation, soybean cv. Williams 82. seedlings were grown in a greenhouse and used five days after sowing. To isolate the RNA from soybean developing seeds, soybean plants were grown at the Bradford Research and Extension Center. Developing seeds collected from the pods were frozen in liquid nitrogen and stored at -80^0 C .

Probe Preparation

The total RNA from hypocotyls of 5 days after sowing (DAS) seedlings was isolated through the lithium chloride precipitation procedure (Lizzardi, 1983). Primers (Forward: 5’ TCC AAG GCT AGT GTG TGC TCA GGC and Reverse: 5’ GAT AAC CCC ACG ATT CCT CAG CT) were synthesized on the basis of sequence information obtained from that of an EST clone (Genbank # BE822813). These primers were used to amplify a 328 bp fragment by RT-PCR following the protocol given by the manufacturer (Stratagene, La Jolla, CA). This fragment was sequenced with a *Taq* Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) after being cloned into a sequencing vector pGEMT-EASY. The sequence obtained was confirmed with that of the published sequence.

The 328 bp RT-PCR product was purified and radiolabeled with [^{32}P] dCTP (PerkinElmer Life Sciences Inc., Boston, MA) using a random labeling kit (Takara Bio Inc., Shiga, Japan). This probe was used to screen the soybean genomic library and also for expression analysis.

Isolation of Genomic Clone

A commercially available soybean genomic library (Stratagene La Jolla, CA) was screened to isolate a TS genomic clone. Colony hybridization was carried out with a radiolabeled probe for three consecutive times following the standard method (Sambrook ***et al.,*** 1989) resulted in three positive clones. DNA restriction followed by a southern analysis of one of the three clones revealed hybridization with two ***Bam HI*** inserts of 3 and 4.5 kb. These inserts were gel isolated, purified and sub cloned into an intermediate vector pBS before sequencing. Analysis of the genomic sequence employed different publicly available software. BLASTX (NCBI) was used for the homology search; CLUSTLAW and BOXSHADE were used for multiple sequence analysis and phylogenetic analysis.

Expression Analysis

Total RNA was extracted from six stages of the developing seeds of soybean cv. Williams 82, following the standard $LiCl_3$ precipitation procedure (Lizzardi, 1983) and

was quantified by measuring the optical density (A260/A280) using a spectrophotometer. 5 µg of RNA was fractionated in a 1.2% agarose-formaldehyde gel, transferred to a Hybond-N+ membrane and immobilized by UV cross-linking. The 328 bp RT-PCR product was purified from an agarose gel using Ultrafree-DA columns (Millipore Corporation, Bedford, MA, USA) and labeled with [^{32}P] dCTP using the Ladderman labeling kit (PanVera Corporation, Madison, WI, USA). Prehybridization was done at 65 °C in 7% SDS, 191 mM Na_2HPO_4, 58 mM NaH_2PO_4, 1% bovine serum albumin (BSA) and 100 µg/mL denatured salmon sperm DNA for six hours. Hybridization was performed overnight using the same buffer at 65 °C. Following hybridization, the membrane was washed with 0.5 × SSC (1 × SSC is 150 mM NaCl, 15 mM sodium citrate) and 0.1% SDS at 65 °C. and exposed to an X-ray film an intensifying screen at −80 °C. An RNA molecular weight marker was used to estimate the sizes of hybridizing transcripts.

Southern Analysis

The cetyltrimethyl ammonium bromide (CTAB) method of Saghai-Maroof (1984) was followed in order to isolate the genomic DNA from leaves of the Williams 82 cultivar. Eight µg of DNA was fractionated with different restriction enzymes overnight at 37 °C. After electrophoresis on a 0.8%

agarose gel, the DNA was transferred to a nylon membrane by capillary transfer using 0.4 M NaOH. The same DNA fragment used to screen the genomic library was labeled with [^{32}P] dCTP using the Ladderman kit (Takara Bio Inc. Otsu, Shiga, Japan) and used as the probe for southern analysis. After six hours of prehybridization, overnight hybridization was done at 65 °C using 6X SSPE buffer, (1X SSPE is 0.1 M NaCl, 0.01 M Na_2PO4, and 0.001 M EDTA), 5X Denhardt's solution, 0.5% SDS, and 50 ug/mL of salmon sperm DNA. After hybridization, the membrane was washed three times in wash solution I (2X SSPE and 0.5% SDS) for 10 min at room temperature before two washes in wash solution II (0.1X SSPE and 0.1% SDS) at 65 °C for 30 min. Following the washes, the nylon membrane was exposed to X-ray film overnight at -80 °C.

Results

Screening the soybean genomic library with an EST sequence as the probe resulted in three positive clones. Two ***Bam HI*** inserts of 3 and 4.5 kb were sequenced and obtained sequence information of a total of 3635 nucleotides. BLAST analysis of this sequence revealed a single open reading frame (ORF) of 1545 bp, encoding a predicted protein of 515 amino acids. The putative protein

has a predicted molecular mass of 56.3 kD and an isoelectric point of 6.68. The physical map of this cDNA clone is shown in Figure 2. Out of the 3635 bp sequenced, the ORF of TS begins at position 1097 and ends at position 2641. The soybean TS protein coding sequence does not contain any introns. The entire 3635 bp of the genomic sequence with deduced amino acids is given here (Fig 3). The sequence of the soybean TS genomic clone was submitted to the NCBI database [Genbank # DQ275353].

BLAST analysis showed that this sequence has significant identity with TS from other plant species. An amino acid comparison of the soybean TS revealed an 73% identity with that of ***Medicago truncatulata*** [Genbank # AC146308], 79% identity with that of ***Arabidopsis thaliana*** [Genbank # AB027151], 76% identity with ***Solanum tuberosum*** [Genbank # AF082894], and 69% with the TS of ***Oryza sativa*** [Genbank # AC135925]. Multiple sequence alignment (Fig. 4) showed that the soybean TS has a transit peptide of at least 34 amino acids. Phylogenetic analysis (Fig 5) revealed that soybean TS is more closely related to those of other plant species than its bacterial counterparts.

Expression Analysis

To determine the expression of TS transcript during seed development, northern analysis was performed using RNA isolated from six stages of the developing seeds. Five µg of RNA was resolved on a formaldehyde/agarose gel and mobilized to a nylon membrane. Hybridization was carried out with a ^{32}p labeled EST insert. The soybean TS showed temporal expression during seed development (Fig. 6). Expression of TS is less abundant in stages 1 and 5 and very poor in stage 6. However, the transcript of TS showed abundance during stages 2, 3, and 4. The expression of the 18S RNA, which was used as control, was similar in all stages of seed development.

Genomic Southern Analysis

Southern analysis with six different restriction enzymes was carried out to determine the copy number of TS in the genome of the soybean. Autoradiography resulted in more than two bands for all the restriction enzymes. However, for all the enzymes studied, while one band was very strong, other bands showed weak hybridization. The obtained results indicate that soybean HSDH is encoded by a low copy gene (Fig. 7).

Discussion

Threonine is one of the limiting second tier amino acids in soybean seeds (Rapp *et al.,* 2003). Over-expression of key regulatory enzymes in the threonine biosynthesis pathway can improve the level of threonine in seeds of this crop. TS is one of the key enzymes in the regulation of biosynthesis for this amino acid as it competes for CGS from the methionine specific pathway for the common substrate HSerP. Though initial experiments suggested that over-expression of TS can reduce methionine, which is very important for the nutritional value of soybeans, subsequent studies (Muhitch, 1997) raised questions about this idea. Hence, the primary objective of this project was to isolate the TS coding gene from soybean. Since a characterization of the promoter region can give novel information on the transcriptional regulation of this gene, isolation of the promoter region was also an objective.

Efforts to screen the soybean genomic library resulted in the isolation of a gene encoding TS. The coding sequence of this gene showed significant identity with TS of other plants and microorganisms. Though the TS-coding gene has been isolated and thoroughly characterized in *Arabidopsis* (Curien *et al.,* 1996), this study is the first to report the soybean ortholog, named *gmts*. From the multiple amino

acid sequence sequence comparisons of TS from plants and bacteria, I propose that soybean TS has a transit peptide. All enzymes in the aspartate pathway have such a transit peptide since they are localized to chloroplast. In addition, studies with other plant TS also reported the existence of such a transit peptide (Curien ***et al.,*** 1996). The ***gmts*** has no introns. The genes coding this enzyme in ***Arabidopsis*** [Genbank # AY099629] and ***Oryza*** [Genbank # XM_475849] also do not have introns. Hence, I propose that, in plants, TS is coded by intron-less genes.

Expression analysis of soybean TS in six stages of developing seeds showed temporal expression (Fig 6). Expression of this protein was lower in stages 1 and 5 and lowest in stage 6. It was abundant in stages 2, 3, and 4. Though TS has been well- characterized biochemically in Arabidopsis, its expression in various tissues has not been reported in comparison with the expression of soybean TS. Southern analysis revealed more than one band, indicating the existence of a multiple number of TS genes in the genome. Curien ***et al.,*** (1996) reported that two copies of TS genes in the ***Arabidopsis*** genome.

The significance of this study lies in the isolation of the TS coding gene in the soybean. This gene can be used to determine whether its over-expression can increase the

level of threonine in soybean seeds without causing a reduction of methionine. The promoter sequences isolated can be used to study the possible regulation of TS in ways other than SAM mediated allosteric activation.

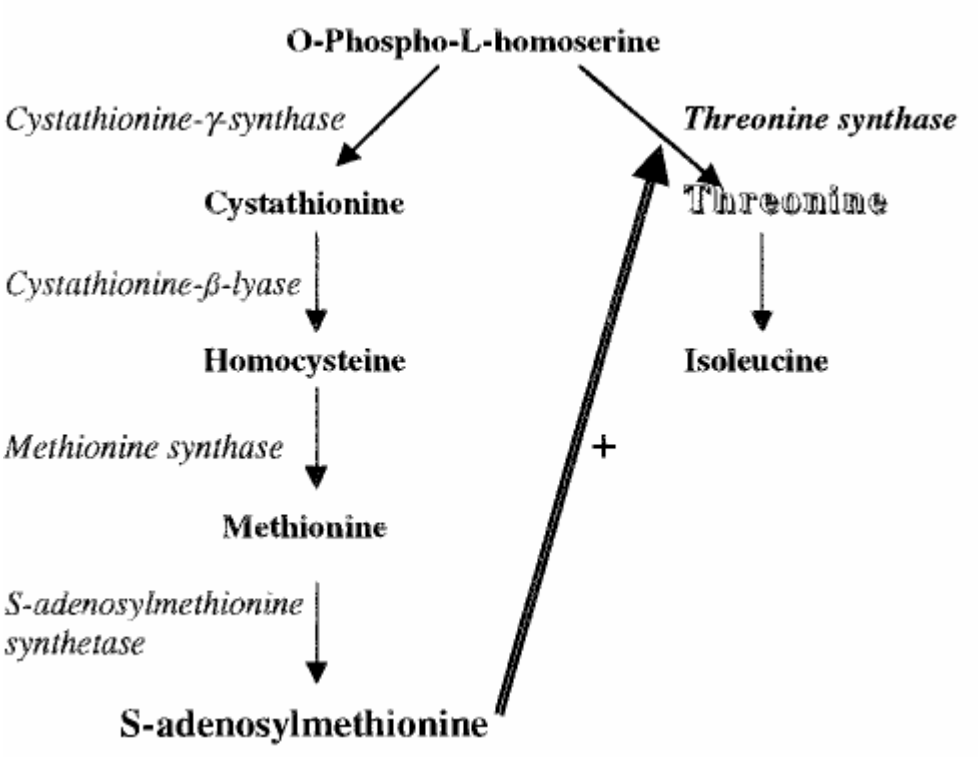

Fig. 1. A pathway from O-phosphohomoserine to threonine, methionine, and SAM. Enzyme names are written in italics, and compounds are written in bold. The thick arrow indicates the activation of TS by SAM.
Source: Thomazeau *et al.*, 2001

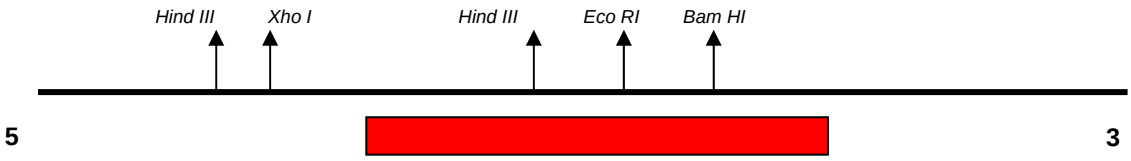

Fig. 2. A partial restriction map of soybean genomic sequence encoding the TS. The red box indicates the location of the open-reading frame (ORF).

```
2     gaattaaaatatttttttatgagttttttttttttaaagataatgttttaggagttgctatc  61
62    tgcctatcattccaaattgtactacatgtcttatcctaattttatacgttgaacatttag  121
122   ttttatgttttattccttttaaatatacttcttcataatttgaatttgaattttgatttg  181
182   atttaataatgaccagcatttttttttctctatcatctcaaatcactttgctttttatctt  241
242   tatctatatctgtgtagttgattctttattaatcacttatctctttctccctctcttttt  301
302   tttttaagaaacttatctctttcaaataatcatgctttcccatctttatttatagattga  361
362   attttaatatttttttaaaaattatcaataaataaatatagacctttatcacaatatgaat  421
422   tacctatcataaatttaaaataaaatttaaatacttaaaatatttgttatacattttagc  481
482   tatataaaaaaaatatttatctcacagtttaaaatattagtataaatgtgattgaagaaat  541
542   gtgtaatatgaaagcactgagaaaatttccttacaaagccatgccaaaacaccgtaagct  601
602   tcatgcaccttcttgccacaatccagcggaacctgactcctgcaggataaattaaactgt  661
662   aaaggaatccagattttattttattattgatggtgagacacgacaatattgtcatttcca  721
722   cctcaggctcatgcatatataccctccctgaacgattaattatccatttacatttatctc  781
782   gagcacatcaatttatcaagagcaaacgttacttaaaacaacagtaattatcactaattt  841
842   ctcgggaaaccaatattttattatctttcaatttaagaaaattgaaacaaccaaattgat  901
902   atgacatgattagttgattacacacagtttctttatttttttttttttgcaacgttttgag  961
962   agttaaactacacgtgcgcatcaacctgagattcccttgtccctccttcgataaataaca  1021
1022  tcagaaccattcatatttcatcagtgtttttctcttctccaaattccaatccttcaattc  1081
                    M   L   S   S   L   L   N   P   S   F   T   T   T   L   P
1082  aataatttagtaaaaatgttgtcttctctgctgaacccttccttcaccaccacccttcca  1141
      I   H   S   I   L   I   P   N   N   R   R   A   S   T   V   V   S   C   T   S
1142  atccattcaattttaatccctaataaccgccgcgcctccaccgtagtatcatgcacctcc  1201
      S   H   L   T   T   N   N   H   S   H   V   S   S   P   P   N   I   K   D   E
1202  tcacacctcaccaccaacaaccactcccatgtttcttcgccgccgaacatcaaggacgag  1261
      A   R   R   R   P   A   A   E   N   D   F   T   A   K   Y   V   P   F   N   A
1262  gcccgccgccgccccgcggcggagaacgacttcacggcgaagtacgtccccttcaacgcc  1321
      G   F   D   S   P   E   T   Y   S   L   D   E   I   V   Y   R   S   R   S   G
1322  ggcttcgactcgccggaaacctactcgctcgacgaaatcgtgtaccgcagccgttccggc  1381
      G   L   L   D   V   Q   H   D   M   E   A   L   G   R   F   D   G   A   Y   W
1382  ggactcctcgacgtgcagcatgacatggaggctttagggcgcttcgacggcgcgtactgg  1441
      R   A   L   F   D   S   R   V   G   K   T   T   W   P   Y   G   S   G   V   W
1442  cgcgcgctgttcgactcgcgcgtggggaagacgacgtggccgtacgggtccggcgtgtgg  1501
```

Fig. 3. Nucleotide sequence and deduced amino acid sequence of soybean TS genomic DNA. The sequenced region covers 3635 nucleotides. The ORF for TS begins at position 1097 and ends at position 2641 encoding a 56.3 kD protein.

```
      S  K  K  E  W  V  L  P  E  I  D  P  D  D  I  V  S  A  F  E
1502  agcaagaaggagtgggtcctaccagaaatagaccctgacgacatcgtttccgccttcgaa  1561
      G  N  S  N  L  F  W  A  E  R  F  G  K  Q  F  V  G  M  N  D
1562  ggcaactcgaatctcttctgggcggaacgcttcgggaaacagttcgtgggaatgaacgat  1621
      L  W  V  K  H  C  G  I  S  H  T  G  S  F  K  D  L  G  M  T
1622  ttgtgggtgaagcactgcggaatcagccacaccggaagcttcaaggatctcggcatgacc  1681
      V  L  V  S  Q  V  N  R  L  R  K  M  N  R  P  V  V  G  V  G
1682  gttctcgtcagccaggtcaaccgcctccgcaaaatgaaccgccccgtcgtcggcgtcggc  1741
      C  A  S  T  G  D  T  S  A  A  L  S  A  Y  C  A  S  A  G  I
1742  tgcgcctccaccggcgatacttccgccgcgctctccgcctactgcgcttccgccggaatc  1801
      P  S  I  V  F  L  P  A  N  K  I  S  T  A  Q  L  I  Q  P  V
1802  ccctccatcgtcttcctccccgccaacaagatctccaccgcgcaactaatccaaccggtt  1861
      S  N  G  S  L  V  L  S  I  N  T  D  F  D  G  C  M  K  L  I
1862  tcaaacggctctttagtcttaagcatcaacaccgattttgacggctgcatgaagctgatt  1921
      R  E  I  T  A  E  L  P  I  Y  L  A  N  S  L  N  S  L  R  L
1922  cgtgaaatcaccgccgaattacccatctatttggcgaattcgttgaacagtttgcgtctg  1981
      E  G  Q  K  T  A  A  I  E  I  L  Q  Q  F  N  W  E  V  P  D
1982  gagggtcaaaaaaccgctgcaattgagattctgcagcagttcaactgggaggtgccggat  2041
      W  V  I  V  P  G  G  N  L  G  N  I  Y  A  F  Y  K  G  F  K
2042  tgggttattgtacccgggggcaatctgggaaatatttatgcttttacaagggttttaag  2101
      M  C  K  E  L  G  L  V  E  R  I  P  R  L  V  C  A  Q  A  A
2102  atgtgcaaggaattagggcttgtggagaggattccaaggctagtgtgtgctcaggctgcg  2161
      N  A  N  P  L  Y  L  H  Y  K  N  G  F  K  D  F  N  A  V  K
2162  aatgcaaacccttttgtatttgcactacaagaatgggtttaaggattttaacgctgtgaag  2221
      A  E  T  T  F  A  S  A  I  Q  I  G  D  P  V  S  I  D  R  A
2222  gctgagactacttttgcttctgctattcagataggggatcctgtttcaattgacagggca  2281
      V  H  A  L  R  N  T  E  G  I  V  E  E  A  T  E  E  E  L  M
2282  gtgcatgctctgaggaacacggaggggatcgtggaggaggcaacagaggaggagctgatg  2341
      D  A  M  V  Q  A  D  S  T  G  M  F  I  C  P  H  T  G  V  A
2342  gatgcaatggtgcaggctgattccactggaatgttcatatgtccacacactggggtggct  2401
      L  A  A  L  I  K  L  R  N  R  G  V  I  G  A  G  E  R  V  V
2402  ctggcggcgcttattaagctgaggaatcgtggggttatcggtgccggtgagagggttgtg  2461
      V  V  S  T  A  H  G  L  K  F  A  Q  S  K  I  D  Y  H  S  G
2462  gtggtgagcactgcacatggattgaagtttgcacagagcaagattgattatcattctggg  2521
      L  I  P  G  M  G  R  Y  A  N  P  L  V  S  V  K  A  D  F  G
2522  ctcattcctggaatgggccgctatgctaacccgctggtttcggttaaggcggattttgga  2581
      S  V  M  D  V  L  K  D  F  L  H  N  K  S  P  D  F  N  K  S
2582  tcggtcatggatgttctcaaggatttcttgcacaacaagtcccccgactttaacaagtct  2641
      *
2642  tgacggttgccaagtaagttttagttcggggttttttctgattaagatgttttaacatg  2701

2702  tttgtgttcactttcggtcgttattattggatttgtaagatttggtccaagtattcgagg  2761
```

```
2762  gttttgatttcaacaacatgcttctggtgacgcaatgcatattttcgtgcataacatcat  2821
2822  tgtcgaagatggatctcgatcgatgaatacttgtgtggcaagtaaatgagaagaaaaata  2881
2882  gtgctacttgtacaagagatttttaaatgttttattttcaagtgtaatcacatcgaataa  2941
2942  atcttgaattggagctattaatgtagcttagcgttgcatgaaatgaacaattgagattga  3001
3002  agtttaagattttatgcaacggatttttttgttcatgtttctcactatcccatctcatgct  3061
3062  ctggctagttctttctttttgtgttttaccagaaacgatcttctgctgaaaaaccgggga  3121
3122  cattgatatgtgatcatctccatattagaaatatctacttttttctcttcttgtcatatca  3181
3182  tatccaaggtgtctatctatctttttccttttttagtttcttatacagctcatacaccatc  3241
3242  acgcagcaatgctattattgacaatttttaatttgcaaaaaatcttgattatgttgaagt  3301
3302  caaacataattatttctttgacaatacaatgaggtggaagatttcatagttttgtttttc  3361
3362  tttttgctacgaacttttgaatataaaatttttccatcgtatactgtttgatttgcttcaa  3421
3422  tttaaactaagtttaaaaaacaaaaattattcgtaaccatttttgtaatgatcacacaaa  3481
3482  cgtacaacgttctcattgtgacagttagcaggagatttgatgctttttccatacggcttt  3541
3542  acaaccactagcaaaccaaatacaatatagttaagtaatacgctgtcggatgccaatttt  3601
3602  qctqtqcacqqtaqcccqaaaqcatcqtcttcca  3635
```

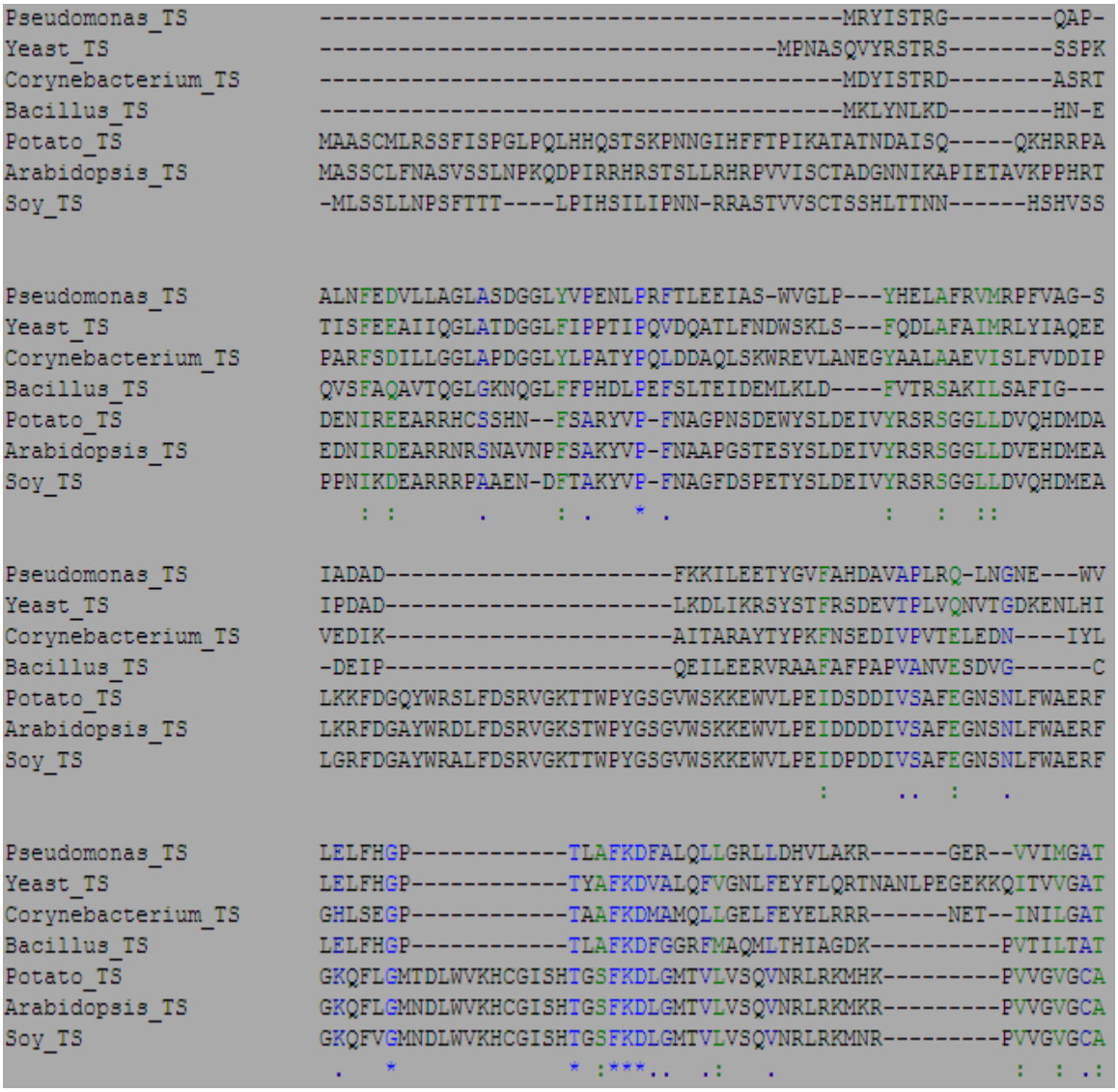

Fig. 4. Multiple sequence analysis of TS from Soybean, *Medicago*, *Arabidopsis*, Potato, Yeast, *Corynebacterium* and *Pseudomonas*. CLUSTALW software was used for multiple alignment. * indicates single, fully conserved residue, : indicates conservation of strong groups, and . indicates conservation of week groups .

```
Pseudomonas_TS          SGDTGSAAIEGCRRCDNVDIFIMHPHNRVSEVQRRQMTTILGDNIHNIAIEGNFDDCQEM
Yeast_TS                SGDTGSAAIYGLRGKKDVSVFILYPTGRISPIQEEQMTTVPDENVQTLSVTGTFDNCQDI
Corynebacterium_TS      SGDTGSSAEYAMRGREGIRVFMLTPAGRMTPFQQAQMFGLDDPNIFNIALDGVFDDCQDV
Bacillus_TS             SGDTGAAVAHAFYGLPNVKVVILYPRGKISPLQEKLFCTLGG-NIETVAIDGDFDACQAL
Potato_TS               STGDTSAALSAYCASAGIPSIVFLPANKISMAQLVQPIAN---GAFVLSIDTDFDGCMQL
Arabidopsis_TS          STGDTSAALSAYCASAGIPSIVFLPANKISMAQLVQPIAN---GAFVLSIDTDFDGCMKL
Soy_TS                  STGDTSAALSAYCASAGIPSIVFLPANKISTAQLIQPVSN---GSLVLSINTDFDGCMKL
                        * .  ::.  .    .:  .:: * .:::  *         .   :::   ** *  :

Pseudomonas_TS          VKASFADQGFLKGTRLVAVNSINWARIMAQIVYYFHAALQL--GAPHRSVAFSVPTGNFG
Yeast_TS                VKAIFGDKEFNSKHNVGAVNSINWARILAQMTYYFYSFFQATNGKDSKKVKFVVPSGNFG
Corynebacterium_TS      VKAVSADAEFKKDNRIGAVNSINWARLMAQVVYYVSSWIRTTTSNDQK-VSFSVPTGNFG
Bacillus_TS             VKQAFDDEELKVALGLNSANSINISRLLAQICYYFEAVAQLP-QETRNQLVVSVPSGNFG
Potato_TS               IREVTAELPIYLAN------SLNSLRLEGQKTAAIEILQQFDWEVP---EWVIVPGGNLG
Arabidopsis_TS          IREITAELPIYLAN------SLNSLRLEGQKTAAIEILQQFDWQVP---DWVIVPGGNLG
Soy_TS                  IREITAELPIYLAN------SLNSLRLEGQKTAAIEILQQFNWEVP---DWVIVPGGNLG
                        ::    :  :          *:*  *: .*    .    :           . ** **:*

Pseudomonas_TS          DIFAGY----LARNMGLP--VSQLIVATNRNDILHRFMSGNRYDK---DTLHPSLSPSMD
Yeast_TS                DILAGY----FAKKMGLP--IEKLAIATNENDILDRFLKSGLYER--SDKVAATLSPAMD
Corynebacterium_TS      DICAGH----IARQMGLP--IDRLIVATNENDVLDEFFRTGDYRVRSSADTHETSSPSMD
Bacillus_TS             DLTAGL----LAKSLGLP--VKRFIAATNVNDTVPRFLHDGQWSP---KATQATLSNAMD
Potato_TS               NIYAFYKGFQMCKELGLVDRIPRLVCAQAANANPLYLHYKSGWKDFKPVKANTTFASAIQ
Arabidopsis_TS          NIYAFYKGFKMCQELGLVDRIPRMVCAQAANANPLYLHYKSGWKDFKPMTASTTFASAIQ
Soy_TS                  NIYAFYKGFKMCKELGLVERIPRLVCAQAANANPLYLHYKNGFKDFNAVKAETTFASAIQ
                        :: *      :.:.:**   : ::  *   *     :   . :          : : :::

Pseudomonas_TS          IMVSSNFERLLFDLH------G---RNGKAVAELLDAFKASGKLSVEDQR---WTEARKL
Yeast_TS                ILISSNFERLLWYLAREYLANGDDLKAGEIVNNWFQELKTNGKFQVDKSI---IEGASKD
Corynebacterium_TS      ISRASNFERFIFDLLG---------RDATRVNDLFGTQVRQGGFSLADDANFEKAAAEYG
Bacillus_TS             VSQPNNWP-----------------RVEELFRRKIWQLKELGYAAVDDET----------
Potato_TS               IG------------------------DPVSIDRAVFALQQCNGIVEEAT----------
Arabidopsis_TS          IG------------------------DPVSIDRAVYALKKCNGIVEEAT----------
Soy_TS                  IG------------------------DPVSIDRAVHALRNTEGIVEEAT----------
                        :                              .   .
Pseudomonas_TS          FDSLAVSDEQTCETIAEVYRSSGE----LLDPHTAIGVRAAR----ECRRSLSVPMVTLG
Yeast_TS                FTSERVSNEETSETIKKIYESSVNPKHYILDPHTAVGVCATERLI-AKDNDKSIQYISLS
Corynebacterium_TS      FASGRSTHADRVATIADVHSRLD----VLIDPHTADGVHVAR----QWRDEVNTPIIVLE
Bacillus_TS             -------TQQTMRELKELGYTS--------EPHAAVAYRALR-----DQLNPGEYGLFLG
Potato_TS               -------EEELMDAMAQADSTG-----MFICPHTGVALTALFKLRNSGVIAPTDRTVVVS
Arabidopsis_TS          -------EEELMDAMAQADSTG-----MFICPHTGVALTALFKLRNQGVIAPTDRTVVVS
Soy_TS                  -------EEELMDAMVQADSTG-----MFICPHTGVALAALIKLRNRGVIGAGERVVVVS
                                :    : .             **:. .  .               : :

Pseudomonas_TS          TAHPVKFPEAVEKAGIG------QAPALPAHLADLFEREERCTVLP-NELAKVQAFVSQH
Yeast_TS                TAHPAKFADAVNNALSGFSNYSFEKDVLPEELKKLSTLKKKLKFIERADVELVKNAIEEE
Corynebacterium_TS      TALPVKFADTIVEAIG-------EAPQTPERFAAIMDAPFKVSDLP-NDTDAVKQYIVDA
Bacillus_TS             TAHPAKFKESVEAILG-------ETLDLPKELAERADLPLLSHNLP-ADFAALRKLMMNH
Potato_TS               TAHGLKFTQSKIDYHSK---------EIKDMECRFANPPVEVKADFGSVMDVLKSYLLSQ
Arabidopsis_TS          TAHGLKFTQSKIDYHSN---------AIPDMACRFSNPPVDVKADFGAVMDVLKSYLGSN
Soy_TS                  TAHGLKFAQSKIDYHSG---------LIPGMG-RYANPLVSVKADFGSVMDVLKDFLHNK
                        **   ** ::                                          ::  : .

Pseudomonas_TS          GNRGKPL
Yeast_TS                LAKMKL-
Corynebacterium_TS      IANTSVK
Bacillus_TS             Q------
Potato_TS               NSKL---
Arabidopsis_TS          TLTS---
Soy_TS                  SPDFNKS
```

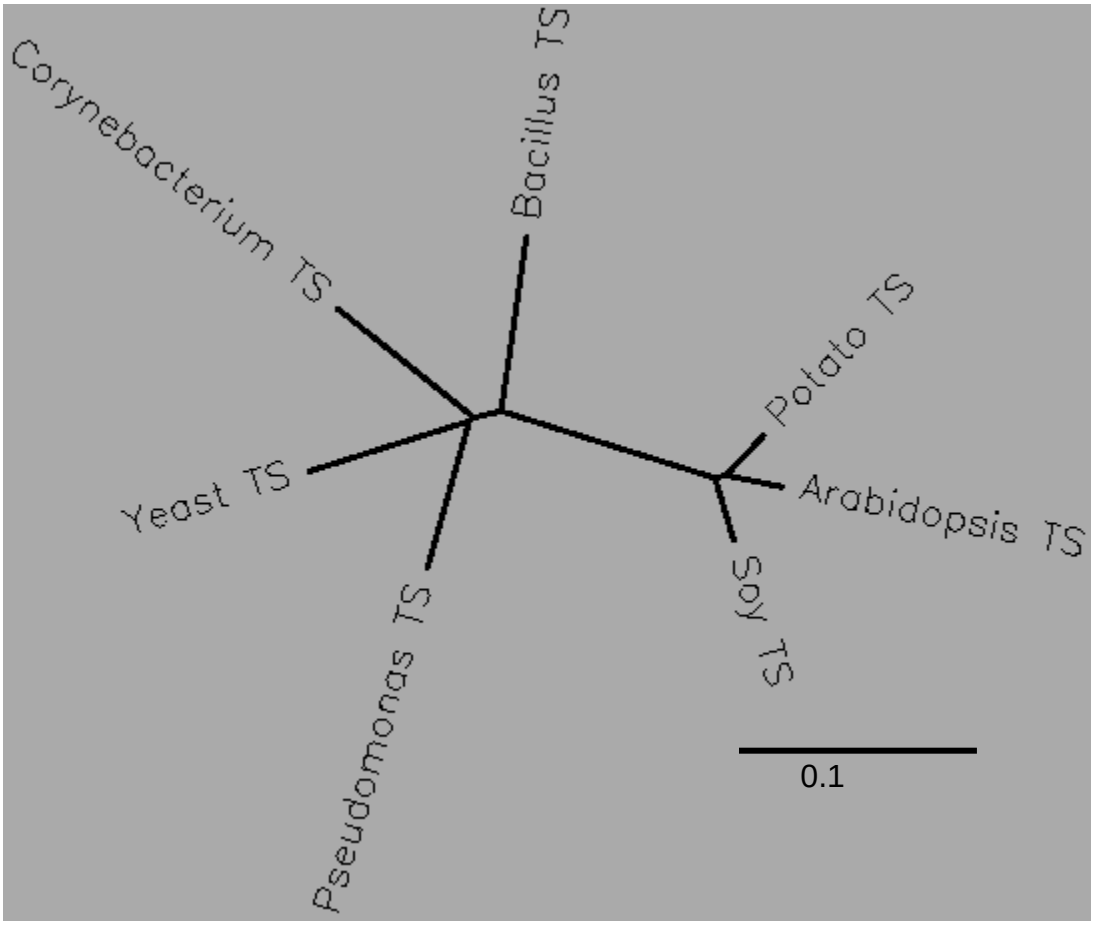

Fig. 5. Phylogenetic analysis of TS from plants, yeast and gram-positive bacterial species. CLUSTALW software was used to draw the unrooted phylogenetic tree.

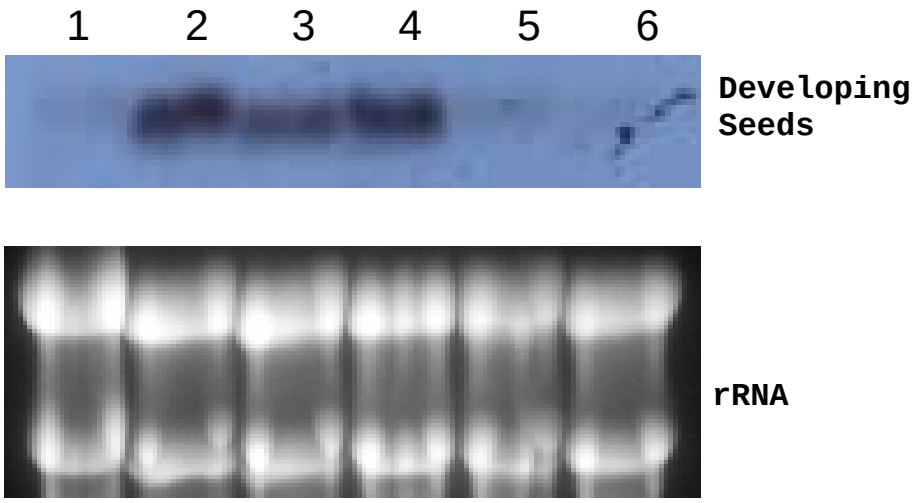

Fig. 6. Expression analysis of soybean TS in developing seeds. Lanes1-6 represent developmental stages 1-6.

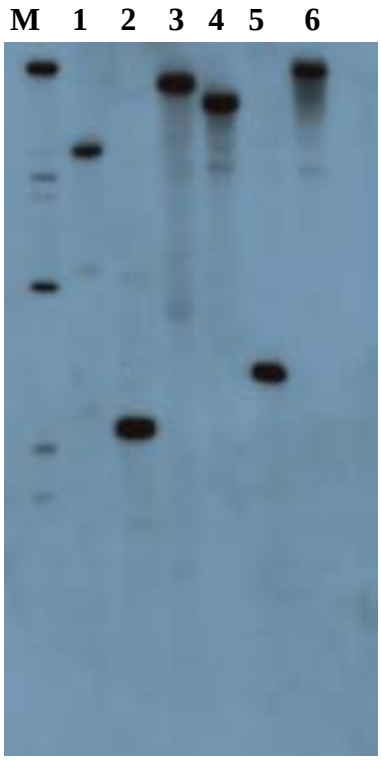

Fig. 9. Southern analysis of soybean TS. 1.*Bam HI* 2. *Dra I* 3. *Eco RI* 4. *EcoRV* 5. *Hind III* 6. *Xba I*

References

Amir R, Hacham Y, Galili G (2002) Cystathionine γ-synthase and threonine synthase operate in concert to regulate carbon flow towards methionine in plants. Trends in Plant Sci 7: 153-156

Bartlem D, Lambein I, Okamoto T, Itaya A, Uda Y, Kijima F, Tamaki Y, Nambara E, Naito S (2000) Mutation in the threonine synthase gene results in an over-accumulation of soluble methionine in *Arabidopsis*. Plant Physiol 123: 101-110

Curien G, Dumas R, Ravanel S, Douce R (1996) Characterization of an *Arabidopsis thaliana* cDNA encoding an S-adenosylmethionine-sensitive threonine synthase. Threonine synthase from higher plants. FEBS Lett 390: 85-90

Curien G, Job D, Douce R, Dumas R (1998) Allosteric activation of *Arabidopsis* threonine synthase by S-adenosylmethionine. Biochemistry 37: 13212-13221

Gakiere B, Ravanel S, Droux M, Douce R, Job D (2000) Mechanisms to account for maintenance of the soluble methionine pool in transgenic *Arabidopsis* plants expressing antisense cystathionine γ-synthase cDNA. C R Acad Sci III. 10: 841- 51

Kim J, Leustek T (2000) Repression of cystathionine γ-synthase in *Arabidopsis thaliana* produces partial methionine auxotrophy and developmental abnormalities. Plant Sci 151: 9-18

Kreft BD, Townsend A, Pohlenz HD, Laber B (1994) Purification and Properties of Cystathionin γ-Synthase from Wheat (Triticum aestivum L.). Plant Physiol 104: 1215-1220

Laber B, Maurer W, Hanke C, Grafe S, Ehlert S, Messerschmidt A, Clausen T (1999) Characterization of recombinant *Arabidopsis thaliana* threonine synthase Eur J Biochem 263: 212-221

Lizzardi PM (1983) Methods for the preparation of messenger RNA. Methods Enzymol 96: 24–38

Madison JT, Thompson JF (1976) Threonine synthetase from higher plants: stimulation by S-adenosylmethionine and inhibition by cysteine. Biochem Biophys Res Commun 71: 684-91

Muhitch M (1997) Effects of expressing ***E.coli*** threonine synthase in tobacco ***Nicotiana tabacum*** L. suspension culture cells on free amino acid levels, aspartate pathway enzyme activities and uptake aspartate into the cells. J Plant Physiol 150: 16-22

Roper MD, Kraus JP (1992) Rat cystathionine _-synthase: Expression of four alternatively spliced isoforms in transected cultured cells. Arch Biochem Biophys 298: 514–521

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA sepacer-length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81: 8014-8019

Sambrook J, Fritsch EF, and Maniatis T (1989) Molecular cloning: A laboratory manual, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Schmidheini T, Mosch HU, Evans JN, Braus G (1990) Yeast allosteric chorismate mutase is locked in the activated state by a single amino acid substitution. Biochemistry 29: 3660–3668

Thomazeau K, Curien G, Dumas R, Biou V (2001) Crystal structure of threonine synthase from ***Arabidopsis thaliana*** Protein Sci 10: 638-648

Umbarger HE (1978) Amino acid biosynthesis and its regulation. Annu Rev Biochem 47: 532-606

Zeh M, Casazza AP, Kreft O, Roessner U, Bieberich K, Willmitzer L, Hoefgen R, Hesse H (2001) Antisense inhibition of threonine synthase leads to high methionine content in transgenic potato plants. Plant Physiol 127: 792-802