Regulation of lysine catabolism in higher plants

Paulo Arruda, Edson L. Kemper, Fabio Papes and Adilson Leite

Lysine is an essential amino acid for mammals but its concentration in cereals, one of our main food sources, is low. Research over the past 40 years has unraveled many biochemical and molecular details of the aspartic acid pathway, which is the main route of lysine biosynthesis in plants. However, genetic manipulation of this pathway has not been successful at producing high-lysine seeds. This is because lysine, instead of being accumulated, is degraded via the saccharopine pathway. Recent work has increased our knowledge of this pathway, including both the enzymes involved and their regulation.

Lysine is one of the most important essential amino acids, owing to its low concentration in cereals, an important human and animal food source[1]. Because of its nutritional relevance, extensive investigations have focused on understanding the regulatory mechanisms that control lysine accumulation in seeds. This involves complex processes including synthesis, incorporation into proteins and degradation.

Lysine is synthesized in plants by a specific branch of the aspartate-family pathway[1]. This pathway is regulated by end-product feedback inhibition, with lysine inhibiting aspartate kinase (AK) and dihydrodipicolinate synthase (DHDPS)[2]. AK is the first enzyme in the pathway and is feedback inhibited by lysine and threonine, whereas DHDPS is specific for the lysine branch and is inhibited only by lysine. Several research groups have isolated mutants of AK and/or DHDPS that are less sensitive to feedback inhibition by lysine in an attempt to increase the concentration of this amino acid in vegetative tissues and seeds. Mutant plants with AK that is less sensitive to lysine inhibition overproduce threonine but not lysine[1]. The failure to accumulate lysine has been attributed to DHDPS, which is much more sensitive to lysine inhibition than AK (Ref. 2). Transgenic plants overexpressing bacterial, feedback-insensitive AK and DHDPS overproduce threonine[3] and accumulate free lysine, but they also display increased lysine degradation[4]. These findings have led to the consensus that lysine catabolism is an important factor to consider when engineering high-lysine plants. However, high free-lysine levels might be toxic to cells and therefore this amino acid might need to be incorporated into lysine-containing proteins.

In cereals, synthesis through the aspartate pathway[5] is not the only source of lysine for developing seeds; lysine is also translocated in appreciable amounts (~5% of the translocated amino acid pool) from vegetative tissues to developing seeds[6]. However, the amount of lysine-containing proteins in cereal seeds is low, in contrast with the high content of lysine-devoid storage proteins, the prolamins[7]. This could lead to an excess of free lysine but this does not occur because cereal seeds have extensive lysine degradation ability through the saccharopine pathway[8,9]. Lysine degradation is not only important for controlling free-lysine levels in plant tissues. Recently, several lines of investigation have revealed that lysine degradation might be related to other physiological processes. In this article, we focus on the regulation of lysine catabolism and its implications for the control of lysine content in seeds, as well as its possible role in plant growth, development and response to environmental changes.

Lysine catabolism through the saccharopine pathway

The saccharopine pathway (Fig. 1) is generally regarded as the major metabolic route for lysine degradation in plants. Feeding experiments on cereal seeds using ^{14}C-lysine showed significant isotope incorporation into α-amino adipic acid and glutamic acid[10,11] (Fig. 1). This pathway has been confirmed by the demonstration of lysine–ketoglutaric acid reductase (LKR, also referred to as lysine 2-oxoglutaric acid reductase) activity in the immature maize endosperm[8].

The first two enzymatic steps of the saccharopine pathway are catalyzed by LKR and saccharopine dehydrogenase (SDH), which are separate domains of a bifunctional polypeptide[12] (LKR–SDH). LKR condenses lysine and α-ketoglutaric acid to form saccharopine, which is then hydrolyzed by SDH, giving rise to α-aminoadipic-δ-semialdehyde and glutamic acid (Fig. 1). These two enzymatic steps can be viewed as an atypical transamination reaction in which the ϵ-amino group of lysine is transferred to α-ketoglutaric acid to form glutamic acid (Fig. 1). A second glutamic acid is generated in a reaction catalyzed by α-aminoadipic acid aminotransferase (AAA), in which the lysine skeleton's α-amino group is transferred from α-aminoadipic acid to α-ketoglutaric acid (Fig. 1).

The ^{14}C-labeling of glutamic acid observed in cereals[10,11] is not due to the direct production of glutamic acid in the reactions catalyzed by LKR–SDH and AAA, because these steps do not transfer carbon atoms from lysine to glutamic acid (Fig. 1). Instead, the carbon atoms from lysine are directed to acetyl-CoA at the end of the pathway, which then enters the citric acid cycle, generating α-ketoglutaric acid. This can then be used to produce another glutamic acid molecule via amino acid transamination (Fig. 1). Thus, the saccharopine pathway seems to channel the whole lysine skeleton to the production of glutamic acid. This amino acid might have regulatory functions not only in plant growth and development but also in responses to environmental changes, as will be discussed later.

Regulation of lysine catabolism in seeds

Prolamins are the most abundant storage proteins of cereal seeds[7] and are devoid of lysine[13]. Thus, the demand for free lysine during seed development is likely to be low. Lysine translocation from vegetative tissues[6] provides more lysine to the seed than is actually required, which could lead to an accumulation of excess lysine. Nevertheless, at least in the developing maize endosperm, the concentration of free lysine is maintained at low levels[14]. By contrast, seeds of dicots synthesize large amounts of lysine-containing

1360 - 1385/00/$ – see front matter PII: S1360-1385(00)01688-5

storage proteins[7] and they, therefore, demand large amounts of free lysine.

Both cereals (e.g. maize[8,14] and rice[9]) and dicots (e.g. *Arabidopsis*[15], tobacco[16] and soybean[17]) catabolize lysine through the saccharopine pathway but the physiological role of lysine catabolism might differ between the two plant families. In maize endosperm and tobacco seeds, LKR–SDH activity is temporally coordinated with the rate of storage-protein accumulation and total nitrogen input into the seed[14,16]. In maize, this agrees with the hypothesis that, because zein (the maize prolamin) synthesis does not demand lysine, the excess of this amino acid should be catabolized. Indeed, the genes encoding members of the zein family, including the abundant 22-kDa zein class, and the gene encoding the maize LKR–SDH have been shown to be under control of Opaque-2, a transcription factor belonging to the basic-domain–leucine-zipper family[18–20]. Thus, as the rate of zein synthesis increases, owing to the transcriptional activation of their corresponding genes by Opaque-2, LKR activity concomitantly increases, resulting in degradation of the excess lysine in the endosperm cells (Fig. 2).

The *opaque2* mutant has decreased levels of LKR–SDH transcript and protein, and therefore has decreased enzyme activity[18,21,22]. This might increase the free lysine available to be used for lysine-containing-protein synthesis. A double mutant homozygous for *opaque2* and *ask1* (which encodes an AK that is less sensitive to feedback inhibition by lysine[1]) has more free lysine, a lower zein content and more lysine-containing proteins than the *opaque2* single mutant[23]. These results suggest that an increase in free lysine owing to the feedback-insensitive characteristics of the AK encoded by *Askl* and decreased LKR–SDH activity has a direct effect on the synthesis of lysine-containing proteins[23] (Fig. 2). Thus, in normal maize, free-lysine availability might be a limiting factor for the synthesis of lysine-containing proteins.

In dicots, LKR activity has been shown to be coordinated with AK activity during seed development[16]. Transgenic tobacco plants overexpressing genes encoding AK and DHDPS enzymes that are less sensitive to feedback inhibition by lysine have shown little effect on free-lysine accumulation because of a dramatic induction of LKR activity[16]. Increased lysine degradation through the saccharopine pathway has also been observed in transgenic canola (*Brassica napus*) and soybean plants overexpressing genes that encode feedback-insensitive AK and DHDPS (Ref. 4). Taken together, these results suggest that lysine degradation by LKR–SDH in seeds of both dicots and monocots is likely to fine tune the regulation of free-lysine levels. In addition, the products of lysine degradation might exert some regulatory functions in seed development as well.

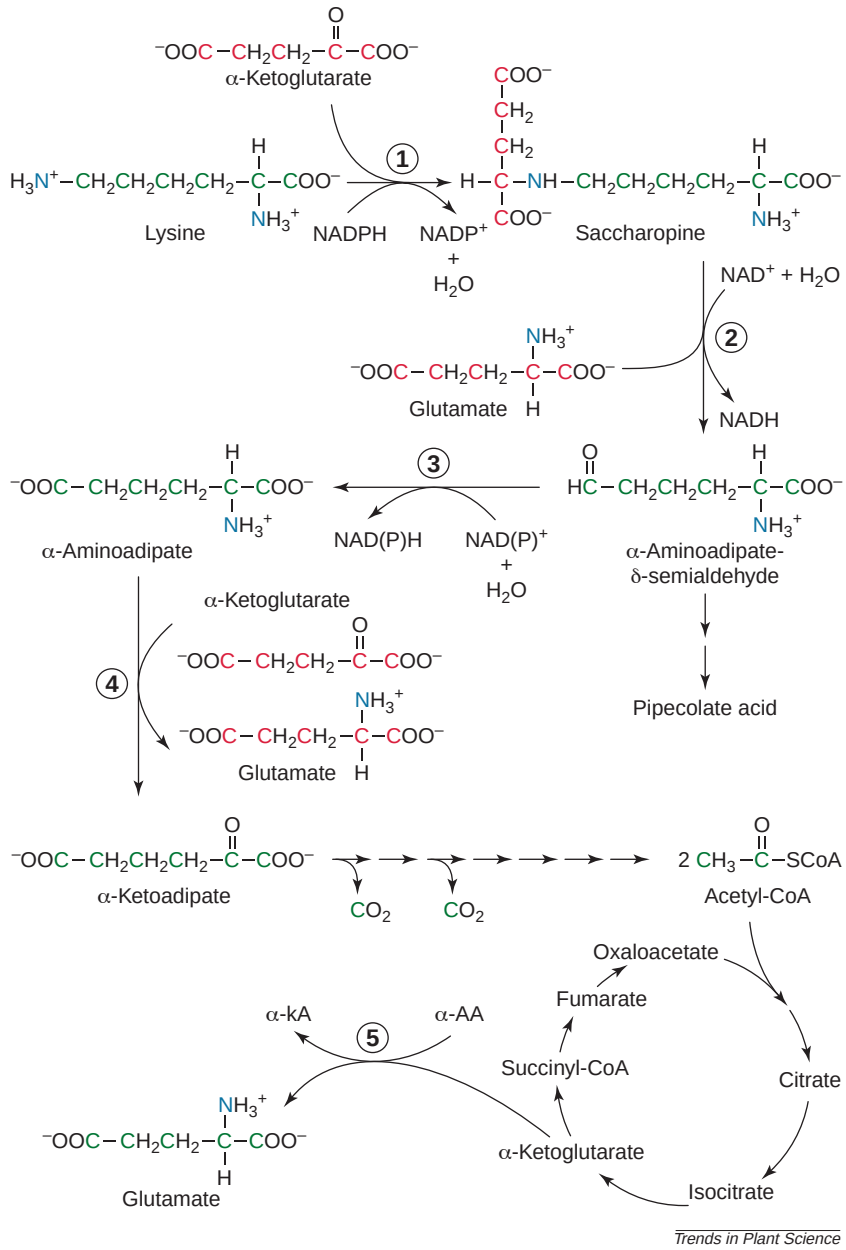

Fig. 1. The pathway of lysine degradation in plants. The enzymes indicated are: (1) lysine–ketoglutaric acid reductase (LKR); (2) saccharopine dehydrogenase (SDH); (3) aminoadipic acid semialdehyde dehydrogenase (AADH); (4) aminoadipic acid aminotransferase (AAA); (5) general α-amino acid transaminase. LKR–SDH and AAA activities incorporate lysine nitrogen atoms (blue) into two molecules of glutamic acid. Lysine and α-ketoglutaric acid carbon atoms are green and red, respectively. Lysine carbon atoms are converted to two carbon dioxide and two acetyl-CoA molecules. These carbon atoms might eventually be incorporated into glutamic acid by other metabolic pathways. Abbreviations: α-kA, α-ketoacid; α-AA, α-aminoacid.

The saccharopine pathway appears to be under complex regulation, particularly at the post-translational level. In tobacco seeds, exogenous lysine administration induces LKR activity by a mechanism involving intracellular Ca^{2+} and protein-phosphorylation cascades[24]. Moreover, the LKR enzymes of soybean[17] and maize (Fig. 3) are phosphorylated *in vitro* by casein kinase. Dephosphorylation of the native enzyme leads to a decreased LKR activity[17]. However, whereas phosphorylation increases LKR activity in a lysine-dependent manner in maize (Fig. 3), it is dephosphorylation that depends on lysine in soybean[17]. It is possible that, in

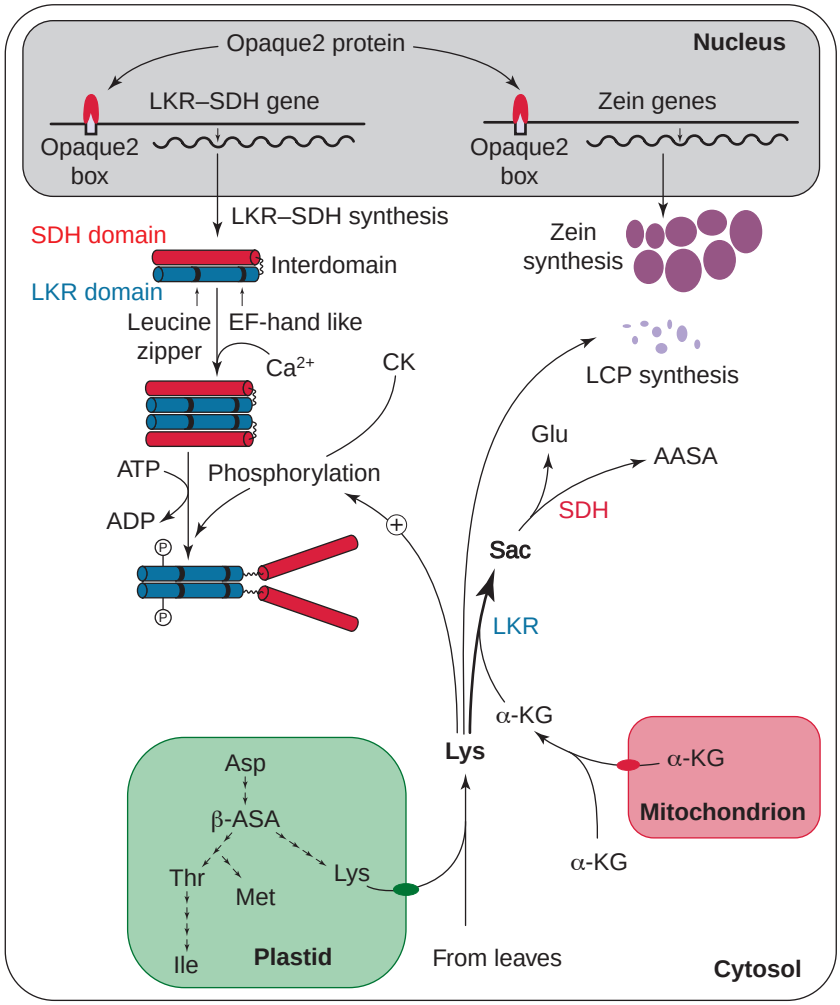

Fig. 2. A proposed model for the regulation of lysine catabolism in a cereal endosperm cell. Developing seeds have two sources of lysine: transport to endosperm cells from vegetative tissues and synthesis in plastids via the aspartate pathway. Part of the lysine is incorporated in lysine-containing proteins (LCP) but the major storage proteins are prolamins, which are devoid of lysine. In maize, the genes encoding zeins are controlled by the transcriptional activator Opaque2, as is the gene encoding the bifunctional enzyme lysine–ketoglutaric acid reductase–saccharopine dehydrogenase (LKR–SDH). The LKR–SDH enzyme is regulated by Ca^{2+}, which is involved in enzyme dimerization, and by phosphorylation by casein kinase (CK) in a lysine-dependent manner. As the pool of lysine increases, LKR activity increases owing to lysine-dependent phosphorylation. Phosphorylation of the LKR domain might derepress the enzyme, which is inhibited by the SDH domain and/or the interdomain of the bifunctional enzyme. In this process, a large proportion of free lysine is catabolized, giving rise to glutamic acid and α-aminoadipic-δ-semialdehyde (AASA). Abbreviation: α-KG, α-ketoglutarate.

maize, lysine accumulation favors the phosphorylated state of LKR and thus activates its own degradation, whereas, in soybean, lysine accumulation stimulates phosphorylation (probably by inducing a protein kinase) but that binding of lysine to the LKR active site favors dephosphorylation, which would prevent lysine depletion from the tissue[17].

In maize, the LKR enzyme is activated by Ca^{2+} and inhibited by the SDH domain and/or the interdomain region of LKR–SDH (Ref. 25). Ca^{2+} appears to be associated with enzyme dimerization, whereas phosphorylation of the LKR domain probably derepresses LKR by releasing the SDH and/or interdomain; after modulation, the enzyme would undergo a conformational change, exposing the LKR catalytic domain for substrate binding (Fig. 2). The post-translational regulation of LKR activity is in keeping with the idea of LKR–SDH fine-tuning the regulation of lysine levels and, in addition, indicates that the enzyme might be involved in some signaling process.

The LKR–SDH gene, mRNAs and proteins

Genomic and cDNA clones encoding the bifunctional LKR–SDH enzyme have been isolated from *Arabidopsis*[15,26] and maize[18] (Fig. 4). A cDNA encoding a monofunctional SDH isoform has also been isolated from *Arabidopsis*[15]. A distinct, shorter mRNA that is transcribed from the same gene as the bifunctional enzyme[15] encodes this monofunctional SDH. Shorter SDH mRNAs have also been detected in small amounts in maize tissues but these mRNAs do not appear to be translated[18].

Little is known about the role of the monofunctional SDH. This isoform might be necessary to enhance saccharopine degradation because SDH has a nonphysiological optimum pH (~9.0); this might mean that more SDH units have to be synthesized for proper saccharopine hydrolysis *in vivo*. Indeed, transgenic canola (a crucifer that presumably expresses a monofunctional SDH, like *Arabidopsis*) overexpressing lysine-insensitive AK and DHDPS accumulates α-aminoadipic acid[27], but transgenic plants that do not have a monofunctional SDH, such as soybean[17], accumulate saccharopine[27]. In maize endosperm, which contains only the bifunctional form, there is no net accumulation of saccharopine[27], which might reflect the large amount of LKR–SDH needed to cope properly with saccharopine hydrolysis.

The genes encoding LKR–SDH in *Arabidopsis* and maize are very large and complex: the maize gene has 26 exons and the *Arabidopsis* gene has 25 exons[26] (Fig. 4). Except for the second exon of the maize gene, which is absent in *Arabidopsis*, the exons are highly conserved in size and sequence (Fig. 4). However, the introns are diverse between the two species, being larger in maize and accounting for the almost-double size of the maize gene (Fig. 4). In *Arabidopsis* and maize, the genes for LKR–SDH encode predicted proteins of 116 and 125 kDa, respectively[15,18]. The polypeptides have N-terminal LKR and C-terminal SDH domains, predicted from similarities to the yeast monofunctional enzymes lysine-forming SDH (encoded by the gene *LYS1*) and glutamic-acid-forming SDH (encoded by the gene *LYS9*)[28]. Both the *Arabidopsis* and the maize bifunctional LKR–SDH have an interdomain region ~100 residues long[15,18,26].

Sequence analyses have revealed that the maize and *Arabidopsis* genes for LKR–SDH contain CCAAT and TATA boxes in a good promoter context in the promoter of the bifunctional gene and also in an internal region (Fig. 4). This might mean that there are two promoters, one for the production of transcripts encoding bifunctional polypeptides and the other for the production of transcripts

encoding monofunctional SDH. In addition, GCN4-like sequences, which are involved in the transcriptional activation of genes involved in nitrogen metabolism in yeast[29] and plants[30], can be found in both the upstream and the internal promoters of the maize gene, and in the internal promoter of the *Arabidopsis* gene (Fig. 4).

However, Opaque-2-binding sites similar to those found in the promoters of prolamin genes are also present in the upstream and internal promoters of the *Arabidopsis* gene for LKR–SDH, but only in the upstream promoter of the maize LKR–SDH gene (Fig. 4). It is possible that both Opaque-2 and GCN4-like target sequences are involved in the transcriptional regulation of maize and *Arabidopsis* genes for LKR–SDH; if this is true, the absence of an Opaque-2 target sequence in the internal promoter of the maize gene might explain why this species expresses only the bifunctional polypeptide even though *Arabidopsis* expresses bifunctional LKR–SDH and monofunctional SDH from the same gene.

Implication of the saccharopine pathway in growth and development

In mammals, the saccharopine pathway is involved in growth and development. LKR–SDH activity has been detected during embryonic central nervous system development[31] and mutations in genes encoding LKR–SDH have been associated with a metabolic disorder known as familial hyperlysinemia, whose symptoms include severe developmental abnormalities, such as mental retardation[32]. Whether the saccharopine pathway has any particular role beyond that of excess lysine degradation in plants remains to be elucidated.

Analysis of the spatial and temporal patterns of LKR–SDH production in maize has revealed that the gene is highly expressed in the endosperm and that this expression is temporally coordinated

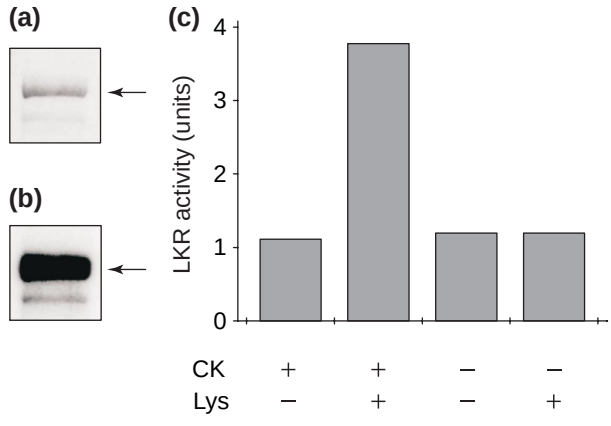

Fig. 3. Phosphorylation of the bifunctional enzyme lysine–ketoglutaric acid reductase–saccharopine dehydrogenase (LKR–SDH) *in vitro*. Aliquots of purified LKR–SDH were subjected to phosphorylation *in vitro* with casein kinase (CK) in a reaction mixture containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT and 20 μM [γ-^{32}P]ATP (1000 cpm pmole^{-1}) in a final volume of 20 μl. After 30 min incubation at 30°C, the samples were separated on a 7% SDS-PAGE gel and the gel stained with Coomasie Brilliant Blue (a) and autoradiographed (b). An aliquot of purified LKR–SDH was dephosphorylated with alkaline phosphatase, leading to the complete loss of LKR activity. The dephosphorylated sample was chromatographed through a Superdex 200 gel filtration column and the protein was phosphorylated with CK in the presence (+) and absence (−) of 11 mM lysine (Lys). After phosphorylation, LKR activity was assayed (c).

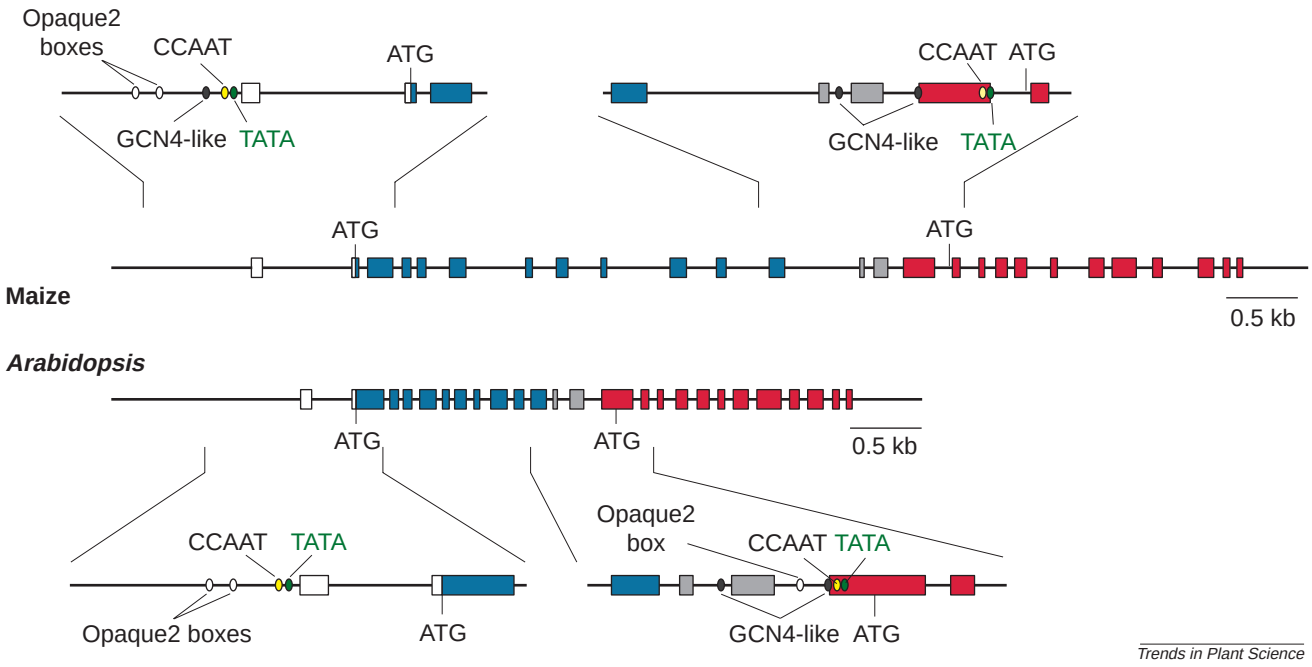

Fig. 4. Structure of coding and regulatory sequences of genes from maize and *Arabidopsis* encoding the bifunctional enzyme lysine–ketoglutaric acid reductase–saccharopine dehydrogenase (LKR–SDH). The gene structure is shown at the center; exons are represented as boxes and introns as lines. Exons belonging to the LKR and SDH regions are blue and red, respectively. Gray boxes represent exons from the interdomain region. The first exons of both the maize or the *Arabidopsis* genes are noncoding (white). The schemes shown above and below are enlargements of selected regions of the maize and *Arabidopsis* genes showing the regulatory TATA and CCAAT sequences, and putative Opaque2 and GCN4 binding sites, in the upstream (left) and internal (right) promoters. The GenBank Accession numbers for the *Arabidopsis* and maize genomic sequences are U95758 and AF271636, respectively.

Table 1. Plant ESTs homologous to the maize lysine-α-ketoglutaric acid reductase and saccharopine dehydrogenase

Nr hits[a]	Species	Variety	Organ	Developmental stage	Accession no.[b]
13	Tomato	TA496	Carpel	5 days pre-anthesis to 5 days post-anthesis	AI486200, AI488742, AI485129, AI486726, AI483615, AI771941, AI486762, AI771935, AI899372, AI488387, AI487518, AI486763, AI771382
7	Tomato	TA496	Callus	25–40-day-old	AI894899, AW034280, AI894874, AW035261, AW035300, AI896768, AW031018
5	Tomato	Rio Grande PtoR	Leaf	4–6-week-old plants treated with mixed elicitors	AW037965, AW093830, AI782310, AW038858, AW096237
4	Tomato	TA496	Pericarp	Red ripe (7–20 days post-breaker)	AW442160, AW224200, AW441656, AW224318
1	Maize	Ohio43	Anthers and pollen	Premeiotic anthers to pollen shed	AW057000
1	Sugarcane	SP3280	Root tip	5 days of stem culturing	*
1	Sugarcane	NA	Leaf roll	NA	AA577639
1	Rice	Milyang23	Immature seed	5 days after pollination	AA753786
1	Rice	Nipponbare	Rice panicle	Flowering stage	C72468
1	Soybean	Williams	Seedling cotyledons	3- and 7-day-old	AW201969
1	Soybean	NA	Mature flowers, field grown plants	NA	AW432287
3	Cotton	NA	Boll abscission zone	NA	AI054604, www.genome.clemson.edu*
6	*Arabidopsis*	Columbia	Pool of several tissues	NA	AA585912, T04246, T45802, T13618, T88051, N37568
1	*Arabidopsis*	Columbia	Seedling hypocotyl	3-day-old	W43474

[a]Number of non-redundant EST clones.
[b]All EST clones, except clones marked with an asterisk, were retrieved from the dbEST using the tblastx algorithm (www.ncbi.nlm.nih.gov/BLAST).
Abbreviations: Nr, non-redundant; NA, data not available.

with endosperm growth and storage-protein deposition[14,18]. Moreover, immunohistochemical data have shown a strong expression of LKR–SDH in the subaleurone layer of the developing maize endosperm, which contains actively dividing cells[18].

In *Arabidopsis*, LKR–SDH is expressed in several regions of the reproductive organs and tissues in which cells are actively dividing[15]. This pattern also appears to be valid in other plants. Recently, worldwide public studies of plant ESTs have sequenced clones corresponding to genes for LKR–SDH. Careful analysis of the tissues used for the construction of the cDNA libraries has provided some clues about LKR–SDH production in several plant species (Table 1). Most of the ESTs were obtained from libraries of reproductive or other growing tissues such as root tip and immature leaves (the leaf rolls). Leaves of tomato plants treated with a mixture of elicitors also produced LKR–SDH (Table 1), as did leaf discs of oilseed rape submitted to osmotic stress or treated with abscisic acid[33].

These data suggest that there is a link between the operation of the saccharopine pathway and developmental processes in plants and animals. How they relate to each other is not known as yet but it is possible that the products of lysine degradation through the saccharopine pathway are needed to regulate growth, development and response to environmental changes. Two potential candidates for such a role are the products of the SDH reaction, α-aminoadipic-δ-semialdehyde and glutamic acid. The former has been shown to be involved in the Lys14-dependent transcriptional activation of some yeast genes involved in lysine biosynthesis[34]. A similar mechanism could operate in plants, in which α-aminoadipic-δ-semialdehyde would transcriptionally regulate the expression of genes involved in nitrogen metabolism and thus in growth and development.

In animals, glutamic acid is involved in cell–cell communication in the central nervous system and is required for normal neuronal-synapse development[35]. Plants also have glutamic acid receptors with strong sequence similarity to animal glutamic acid receptors[36]. It might be that, over the course of evolution, glutamic acid has been preserved as a biologically active amino acid, functioning in development and cellular communication. In this sense, the saccharopine pathway is unique among amino acid degradation pathways in that it generates at least two molecules of glutamic acid per lysine molecule oxidized (Fig. 1). This pathway might be involved in the transient synthesis of glutamic acid, which then functions as messenger between cells during organ

development or in response to environmental changes. In the case of the osmotic-stress response[33], glutamic acid generated through lysine degradation might also act as a precursor of proline, a well-known osmolyte in plants[37].

Conclusions and future prospects

Lysine catabolism might play a central role in controlling free lysine levels in plant cells. The saccharopine pathway, the best characterized and apparently the main route of lysine degradation, appears to be under complex regulation at the transcriptional and post-translational levels. The structure and regulatory properties of the gene for LKR–SDH reveal that it can encode both bifunctional LKR–SDH and monofunctional SDH, a process that involves a putative internal promoter in the middle of the gene. The observation that the possession bifunctional and monofunctional isoforms is species specific suggests that the enzyme has different physiological roles in different taxa. Moreover, the existence of Opaque-2 and GCN4-like binding sites in both the upstream and internal promoters suggests that LKR–SDH might be involved in the regulation of nitrogen balance in plants, by linking lysine degradation to protein synthesis.

It is also clear that these processes are related to seed development and other developmental processes such as root elongation and floral development. The production of this enzyme in processes such as abscission and response to abiotic stress is also important. It is not yet known whether these processes are related to the regulation of free lysine levels. It is possible that, in senescing and stressed tissues, protein hydrolysis leads to transient increases in free lysine concentration, which then needs to be degraded. Alternatively, the saccharopine pathway might be responsible for the synthesis of regulatory molecule(s) involved in developmental processes from root growth to leaf senescence to respond to biotic or abiotic stress.

Detailed studies on LKR–SDH activities in mutants with modified endosperm development, in specific mutants for the aspartic acid pathway and in LKR–SDH mutants or knockouts will soon help to elucidate the steps controlling amino acid and protein synthesis in seeds and determine the precise role of LKR–SDH in plant nitrogen balance and development. Furthermore, the expression of recombinant bifunctional LKR–SDH or the separate LKR and SDH domains will help to determine the enzymatic properties of each domain as well as to identify the inhibitory domain of the enzyme. If the putative regulatory *cis*-acting elements identified in the promoters of the gene for LKR–SDH are proved to be functional, and if *trans*-acting factors (other than Opaque2) that can bind to such *cis*-acting elements are identified, this would help us to understand the signaling process involved in the regulation of lysine catabolism.

Acknowledgements

This work was supported by grants to P.A. and A.L. from Fundação de Amparo à Pesquisa do Estado de São Paulo/FAPESP. P.A. and A.L. received research fellowships from CNPq. Postgraduate fellowships were granted to E.L.K. and F.P. by FAPESP.

References

1 Azevedo, R.A. *et al.* (1999) The biosynthesis and metabolism of the aspartate derived amino acids in higher plants. *Phytochemistry* 46, 395–419

2 Galili, G. (1995) Regulation of lysine and threonine synthesis. *Plant Cell* 7, 899–906

3 Shaul, O. and Galili, G. (1993) Concerted regulation of lysine and threonine synthesis in tobacco plants expressing bacterial feedback-insensitive aspartate kinase and dihydrodipicolinate synthase. *Plant Mol. Biol.* 23, 759–768

4 Falco, S.C. *et al.* (1995) Transgenic canola and soybean seeds with increased lysine. *Biotechnology* 13, 577–582

5 Sodek, L. (1976) Biosynthesis of lysine and other amino acids in the developing maize endosperm. *Phytochemistry* 15, 1903–1906

6 Arruda, P. and Silva, W. (1979) Amino acid composition of vascular sap of maize ear peduncle. *Phytochemistry* 18, 409–410

7 Shewry, P.R. and Casey, R. (1999) Seed proteins. In *Seed Proteins* (Casey, R. and Shewry, P.R., eds), pp. 1–10, Kluwer Academic Publishers

8 Arruda, P. *et al.* (1982) Lysine–ketoglutaric acid reductase activity in developing maize endosperm. *Plant Physiol.* 69, 988–989

9 Gaziola, S.A. *et al.* (1997) The enzymology of lysine catabolism in rice seeds – isolation, characterization, and regulatory properties of a lysine 2-oxoglutarate reductase/saccharopine dehydrogenase bifunctional polypeptide. *Eur. J. Biochem.* 247, 364–371

10 Sodek, L. and Wilson, C.M. (1970) Incorporation of leucine-C^{14} and lysine-C^{14} into protein in the developing endosperm of normal and *opaque2* corn. *Arch. Biochem. Biophys.* 140, 29–38

11 Brandt, A.B. (1975) *In vivo* incorporation of lysine-C^{14} into the endosperm of wild type and high lysine barley. *FEBS Lett.* 52, 288–291

12 Gonçalves-Butruille, M. *et al.* (1996) Purification and characterization of the bifunctional enzyme lysine–ketoglutarate reductase–saccharopine dehydrogenase from maize. *Plant Physiol.* 110, 765–771

13 Shewry, P.R. and Tatham, A. (1999) The characteristcs, structures and evolutionary relationships of prolamins. In *Seed Proteins* (Casey, R. and Shewry, P.R., eds), pp. 11–33, Kluwer Academic Publishers

14 Arruda, P. and Silva, W.J. (1983) Lysine–ketoglutaric acid reductase activity in maize: its possible role on lysine metabolism of developing endosperm. *Phytochemistry* 22, 206–208

15 Tang, G. *et al.* (1997) Regulation of lysine catabolism through lysine–ketoglutaric acid reductase and saccharopine dehydrogenase in *Arabidopsis*. *Plant Cell* 9, 1305–1316

16 Karchi, H. *et al.* (1994) Lysine synthesis and catabolism are coordinately regulated during tobacco seed development. *Proc. Natl. Acad. Sci. U. S. A.* 91, 2577–2581

17 Miron, D. *et al.* (1997) *In vitro* dephosphorylation inhibits the activity of soybean lysine–ketoglutaric acid reductase in a lysine-regulated manner. *Plant J.* 12, 1453–1458

18 Kemper, E.L. *et al.* (1999) The role of Opaque2 on the control of lysine degrading activities in developing maize endosperm. *Plant Cell* 11, 1981–1994

19 Schmidt, R.J. *et al.* (1992) Opaque2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. *Plant Cell* 4, 689–700

20 Cord Neto, G. *et al.* (1995) The involvement of Opaque2 on β-prolamin gene regulation in maize and Coix suggests a more general role for this transcriptional activator. *Plant Mol. Biol.* 27, 1015–1029

21 Brochetto-Braga, M.R. *et al.* (1992) Partial purification and characterization of lysine–ketoglutarate reductase activity in normal and *opaque2* maize endosperms. *Plant Physiol.* 98, 1139–1147

22 Gaziola, S.A. *et al.* (1999) Quality protein maize: a biochemical study of enzymes involved in lysine metabolism. *J. Agric. Food Chem.* 47, 1268–1275

23 Azevedo, R.A. *et al.* (1990) Biochemical genetics of the interaction of the lysine plus threonine resistant mutant *Ltr*1* with *opaque-2* maize mutant. *Plant Sci.* 70, 81–90

24 Karchi, H. *et al.* (1995) The lysine-dependent stimulation of lysine catabolism in tobacco seed requires calcium and protein-phosphorylation. *Plant Cell* 7, 1963–1970

25 Kemper, E.L. *et al.* (1998) Structure and regulation of the bifunctional enzyme lysine–oxoglutarate reductase–saccharopine dehydrogenase in maize. *Eur. J. Biochem.* 253, 720–729

26 Epelbaum, S. *et al.* (1997) Lysine–ketoglutaric acid reductase and saccharopine dehydrogenase from *Arabidopsis thaliana*: nucleotide sequence and characterization. *Plant Mol. Biol.* 35, 735–748

27 Mazur, B. *et al.* (1999) Gene discovery and product development for grain quality traits. *Science* 285, 372–375

28 Ramos, F. *et al.* (1988) Control of enzyme synthesis in the lysine biosynthetic pathway of *Saccharomyces cerevisiae*: evidence for a regulatory route of LYS14. *Eur. J. Biochem.* 171, 171–176

29 Hinnebusch, A.G. (1988) Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. *Microbiol. Rev.* 52, 248–273

30 Müller, M. and Knudsen, S. (1993) The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. *Plant J.* 4, 343–355
31 Rao, V.V. *et al.* (1992) Developmental changes of L-lysine–ketoglutaric acid reductase in rat brain and liver. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 103, 221–224
32 Markovitz, P.J. *et al.* (1984) Familial hyperlysinemias. *J. Biol. Chem.* 259, 11643–11646
33 Deleu, C. *et al.* (1999) Three new osmotic stress-regulated cDNAs identified by differential display polymerase chain reaction in rapeseed leaf discs. *Plant Cell Environ.* 22, 979–988
34 Feller, A. *et al.* (1994) Repression of the genes for lysine biosynthesis in *Saccharomyces cerevisiae* is caused by limitation of Lys14-dependent transcriptional activation. *Mol. Cell. Biol.* 14, 6411–6418
35 Verhage, M. *et al.* (2000) Synaptic assembly of the brain in the absence of neurotransmitter secretion. *Science* 287, 864–869
36 Lam, H.M. *et al.* (1998) Glutamic acid receptor genes in plants. *Nature* 396, 125–126
37 Nanjo, T. *et al.* (1999) Biological functions of proline in morphogenesis and osmotolerance revealed in antisense transgenic *Arabidopsis thaliana*. *Plant J.* 18, 185–193

Paulo Arruda*, Edson L. Kemper, Fabio Papes and Adilson Leite are at the Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, 13083-970, Campinas, SP, Brazil. Paulo Arruda is also in the Departamento de Genética e Evolução, IB, Universidade Estadual de Campinas, 13083-970, Campinas, SP, Brazil.

*Author for correspondence (tel +55 19 788 1137; fax +55 19 788 1089; e-mail parruda@unicamp.br).

Recent progress in reconstructing angiosperm phylogeny

Robert K. Kuzoff and Charles S. Gasser

In the past year, the study of angiosperm phylogeny has moved from tentative inferences based on relatively small data matrices into an era of sophisticated, multigene analyses and significantly greater confidence. Recent studies provide both strong statistical support and mutual corroboration for crucial aspects of angiosperm phylogeny. These include identifying the earliest extant lineages of angiosperms, confirming *Amborella* as the sister of all other angiosperms, confirming some previously proposed lineages and redefining other groups consistent with their phylogeny. This phylogenetic framework enables the exploration of both genotypic and phenotypic diversification among angiosperms.

Understanding the phylogenetic relationships among the principal lineages, or clades (Box 1), of angiosperms is essential for elucidating the evolutionary events that underlie the diversification and ascension of this ecologically dominant plant group. We also need to reconstruct flowering-plant phylogeny to facilitate comparative studies of plant development, metabolism, reproduction, pathology and genomics. For these and other reasons, reconstructing angiosperm phylogeny has been a major goal of plant systematists.

The state of knowledge before 1999

Attempts to unravel the overall phylogeny of angiosperms through cladistic analysis date back more than a decade[1,2]. Goals of such studies include identifying the composition of major lineages, the relationships among them and the earliest lineages (first-branching clades) of flowering plants. Analyses reported before 1999 were typically based on relatively small non-molecular[2,3] or single-gene[4–6] data matrices, with some exceptions[7,8]. Many results generated during this period constituted noteworthy advances that were largely upheld by subsequent work. For example, several clades were identified, including the eudicots, rosids and asterids; some previously proposed groups, including the Hamameliidae and Dilleniidae, were also shown to be assemblages of distantly related species[2,4–6,8,9]. However, although a potentially accurate picture of angiosperm phylogeny was taking shape, the plant-systematic and larger biological communities did not place great confidence in it.

In addition to obvious instances of conflict among the earlier studies, systematists were aware of several other problems that tempered their enthusiasm. One major concern was that statistical support for putative clades and the relationships among them was generally low, if investigated. A second concern was that earlier studies relied exclusively on parsimony as an optimality criterion in data analysis. However, in parsimony analyses of DNA sequences, long branches in a tree separated by short internodes can attract each other artifactually because of chance substitutions of identical nucleotides at homologous sequence positions[10,11]. Such long-branch attraction can be engendered by using distantly related outgroups. This is because the branch leading to the outgroups attracts another long branch to the base of the ingroup (Box 1). Alternatively it can be engendered by insufficient taxon sampling, because taxonomically large groups are represented only by sparse, long branches in an analysis[9,12–14].

A third concern about these earlier studies was that the available analysis protocols and computer programs employed were not well suited to analysing complex phylogenies (those with large numbers of taxa[5,15,16]). Consequently, analyses of some complex phylogenies had to be stopped by the investigators before they could be completed[4–6]. Finally, it became clear that the amount of data being analyzed was not sufficient to resolve the phylogenetic problems addressed, both because there were too few phylogenetically informative characters[9,12,15] and because some of the apparently informative characters were potentially biased and misleading[9,17].

Breakthroughs during the past year

Beginning in late 1999, several more-rigorous, multigene studies have been published that address phylogenetic relationships among

1360 - 1385/00/$ – see front matter PII: S1360-1385(00)01685-X