S39-002

Patterns of gene expression in developing *Arabidopsis* seeds during storage lipid synthesis

SA Ruuska[1], C Benning[2] and JB Ohlrogge[1]

[1]*Michigan State University, Department of Plant Biology, East Lansing MI 48824 USA, FAX + (517) 353 1926, email ruuska@pilot.msu.edu*

[2] *Michigan State University, Department of Biochemistry, East Lansing MI 48824 USA*

Keywords: functional genomics, lipid synthesis, microarrays, seeds

Introduction

Seeds provide most of the economic value in agricultural products, and a great amount of research has been directed to improve traits associated with seed crops. The properties and uses of crop seeds are largely determined by the different amounts and proportions of storage compounds they have accumulated. Although the biosynthetic pathways responsible for the synthesis of main storage compounds (such as starch, oil and protein) are largely known, it is still unclear what factors determine the overall partitioning of seed reserves to the main storage components.

With its fully sequenced genome, *Arabidopsis thaliana* has now become the dominant model system in dicot plant science. Since the main storage reserve in *Arabidopsis* seeds are lipids, it can be used as a model to study fatty acid synthesis in oilseed crops. To this end, we have produced microarrays containing about 5000 clones selected from a developing *Arabidopsis* seed-specific cDNA library. These microarrays have already proven useful in identifying seed-specific genes (Girke et al 2000) and in this study we have explored the changes in the mRNA levels of thousands of genes during the early seed filling period. In this paper we discuss some selected results of this vast dataset, concerning storage compound synthesis.

Material and methods

Microarray experiments

Arabidopsis thaliana (Col-2) were grown in a growth chamber with 16 h light at 80-100 µE intensity and day/night temperature of 22/20ºC. For harvesting of siliques of defined age, individual flowers were tagged with colored threads on the day of the flower opening. Developing seeds were dissected from siliques between 5 and 13 days after flowering (DAF), frozen immediately and stored at –80ºC before RNA extractions. Fluorescent probes were synthesized from 3 µg of total seed RNA using 3DNA Submicro Expression Array Detection Kits (Genisphere Inc, Montvale, NJ, Cy 3 or 5 fluors). Seed RNA isolation, probe synthesis and array hybridization procedures are described in detail in Ruuska and Ohlrogge (2001). The microarrays used for this study contained a total of 6000 DNA elements, which consisted of ESTs collected from a cDNA library of developing Arabidopsis seeds, additional cDNA clones from genes of lipid and carbohydrate pathway, as well as sensitivity and cross-hybridization controls. Amplification of cDNAs and printing of the arrays was described previously (Girke et al 2000). The hybridized arrays were scanned with Scanarray 4000 (GSI Lumonics, Billerica, MA) at a resolution of 10 µm per pixel. The mean fluorescence intensities for each cDNA clone were determined by using the ScanAlyze v. 2.44 software. The fluorescence intensities from the two fluors were normalized using a linear stepwise

normalization procedure. First, the intensities from both Cy3 and Cy5 channels were sorted separately, and divided into 10 subgroups. The average signal intensities of both channels were calculated for each subgroup, and the ratio between them was used to normalize the Cy5 intensities.

Experimental setup and data presentation

As a common reference for the time course, we chose a 1:1 mixture of RNA extracted from 9 and 10 days old seeds. At this stage the seeds are metabolically very active, and maximal amount of genes should be expressed and so give a reliable reference reading. Three different time points were compared to this reference. The earliest time point was a 1:1 mixture of RNA from 5 and 6 days old seeds (referred to as 5+6 DAF, Days After Flowering). This stage was prior to the active storage compound accumulation. The next time point was 11 DAF, at the early phase of the storage lipid synthesis. The last point was 1:1 mixture of RNA from 12 and 13 days old seeds (referred to as 12+13 DAF), at the maximal oil accumulation stage. The averaged expression ratios from two technical replicates (X DAF / 9+10 DAF) were calculated for each clone for the complete time course. A clone was deemed reliable only if its average signal intensity was more than 2 times above the average background. To simplify the presentation and interpretation of the results, the time course data was re-scaled such that the lowest expression ratio was set to 1, and the other ratios adjusted to accordingly. This modification makes it easier to visualize the shape and magnitudes of the expression profiles in the figures.

Lipid and protein assays

Total lipid content was measured from 20-50 seeds by GC-MS after direct esterification of fatty acids (Browse et al 1986). For total protein assays, 5-10 seeds were homogenized in 250 µl of 25 mM Tris-HCl pH 8, 125 mM NaCl, 0.5 mM EDTA, 0.5% (w/v) SDS. Protein content was measured from 200 µl of the crude homogenate with the Bio-Rad DC protein assay system using γ-globulin as a standard.

Results and discussion

Lipid and protein content

Youngest seeds (5+6 DAF) contained about 0.1 µg of fatty acids (FAs) per seed. The FA content increased to 2.4 µg in the 13 DAF seeds, which corresponded to 50-60% of the total lipid content of mature seeds grown under the same conditions. The major accumulation of storage lipids started around 9-10 DAF, as indicated by an increase in the total FA content, and appearance of long-chain FAs (20:1). The total protein content of the developing seeds increased more steadily through the time course. The protein content at 5 DAF was about 1.25 µg per seed, and it increased to 3.8 µg per seed at 13 DAF, which corresponds to about half of the total protein content of mature seeds (Data not shown).

General gene expression changes

In the microarray experiments, altogether 4544 of the >6000 clones had an average signal intensity above 2 times the background and were included for further analysis. About 30% of these reliable clones changed more than 1.9-fold across the time course. For the most part, the gene expression changes during the time course were moderate. Only 21 clones changed more than 10-fold, 180 more than 4-fold and the rest 1000 or so changed 1.9-4 fold during the experimental period. Thus, although major changes in metabolism occurred between 5 and 13 DAF, the majority of genes represented on this array changed very little during this time.

In some part, the reasons for the small expression differences between 5 and 13 DAF could reflect technical issues. Array data may compress differentials (Hihara et al 2001), and on the other hand, the development of embryos within a silique is not synchronous (Bowman 1994). However, by and large our observations agree with the earlier results obtained using

hybridization kinetics to study different mRNA classes in developing soybeans (Goldberg et al 1981). The majority of mRNAs that were found in mid-maturation stage soybean embryos during storage protein synthesis were already present in earlier cotyledon-stage. Moreover, a large fraction of the transcripts were present throughout the whole seed development and were even stored in the dry seeds. These results, and our preliminary data from *Arabidopsis* mutant vs. wild-type comparisons indicate that large changes in metabolism are possible with relatively minor changes in transcription of enzymes involved.

Fatty acid synthesis-related genes

The arrays used in this study contained about 250 genes involved in lipid metabolism. Altogether, about 40 % of the lipid-clones changed more than 1.9-fold during the time course. Generally the clones that changed could be divided into 2 main categories. The first group contained several genes encoding central enzymes in the fatty acid synthesis (FAS) pathway whose expression was lowest in the youngest (5+6 DAF) seeds, clearly increased by 11 DAF and, surprisingly, decreased at 12+13 DAF. This is illustrated in Figure 1A, which presents the expression patterns of 2 different subunits of plastidial heteromeric acetyl-CoA carboxylase (ACCase), the first committed enzyme of FAS. Of all the enzymes of the fatty acid synthase complex, only 3 ketoacyl-ACP-synthase I (KAS I) changed more than 2-fold, following the expression pattern of ACCase (Fig 1B, compared with KAS III). Other enzymes with similar expression profile included 2 isoforms of acyl carrier protein (ACP) and 2 acyl-ACP-thioesterases (data not shown). The second group contained some fatty acid modifying enzymes such as linoleate desaturase (FAD3) and fatty acid elongase 1 (FAE1) as well as several oil body proteins, oleosins (Figs 2C,D). Their expression of was lowest in young seeds and increased by 11 DAF as well, but was not down-regulated at 12+13 DAF.

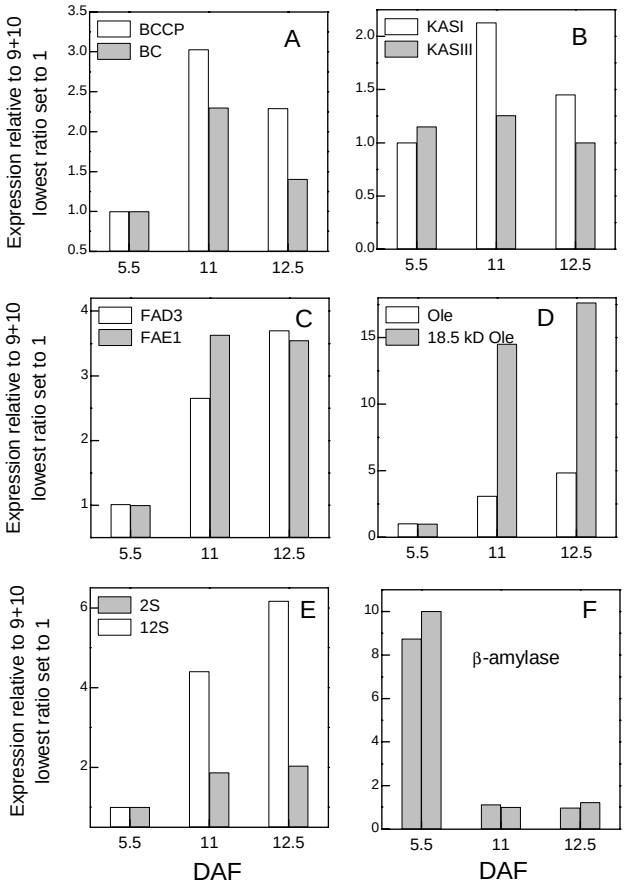

Fig. 1. Expression patterns of:
A. Biotin carboxyl carrier protein (BCCP), and biotin carboxylase (BC) subunits of plastidial acetyl-CoA-carboxylase,
B. 3 ketoacyl-ACP-synthase (KAS) I and III,
C. Linoleate desaturase (FAD3) and fatty acid elongase 1 (FAE1),
D. Two oleosin isoforms,
E. 12S and 2S storage proteins and
F. Two clones of β-amylase.
Expression is related to 9+10 DAF, and the ratios (average of two replicates) were rescaled by setting the lowest ratio to 1.

These results agree with earlier reports showing that the transcript levels and activities of many FAS enzymes increase during lipid accumulation in oilseeds (Post-Beittenmiller et al 1993). The reason for the modest changes (2-3 fold) in the transcription of many FAS enzymes may reflect the fact that they also are ubiquitously needed for membrane biosynthesis in all cells and so must be present all the time. Array data further established that desaturation and elongation of the fatty acids, as well as oil body formation is likely to be under different control from the core FAS enzymes.

Storage protein and starch-related genes

The two major storage proteins in *Arabidopsis* are 12S (cruciferin-like) and 2S (napin-like) proteins. Throughout the time course, the 2S transcripts were amongst the most highly expressed (in terms of their hybridization signal) and their expression increased about 2-fold during the time course. On the other hand, 12S transcripts gave weaker signals, but their expression was strongly induced in the maturing seeds (Fig. 1E). These results further confirm that there are distinct mechanisms regulating the oil and protein accumulation in seeds (see Post-Beittenmiller et al 1993).

Young *Arabidopsis* seeds transiently accumulate starch, which later disappears and presumably is used as a carbon source for fatty acid and amino acid synthesis. The expression pattern of β-amylase changed strongly during the time course (Fig 1F), being highest in young seeds. Another starch-hydrolysis related enzyme, isoamylase, followed the same time course, but changed only about 2-fold (data not shown). These profiles correlated well with the reported changes in seed starch content (Focks and Benning 1998) and suggest that in developing *Arabidopsis* seeds the induction of starch breakdown is important in feeding the temporarily stored carbon towards oil synthesis. Since β -amylase was also found to be strongly seed-specific (Girke et al 2000), it may have a regulatory role in the conversion of starch to other products in developing seeds.

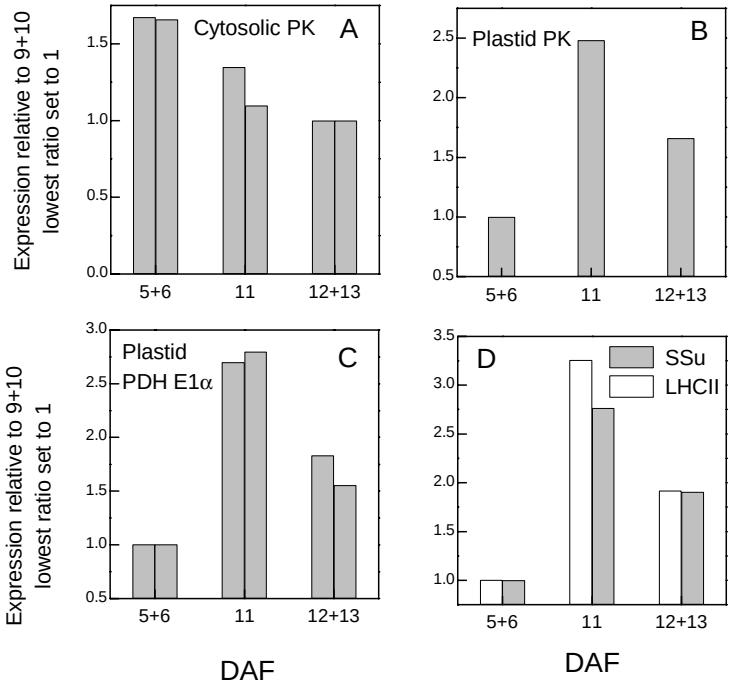

Fig 2. Expression patterns of
A. Two clones of cytosolic pyruvate kinase (PK),
B. Plastic pyruvate kinase,
C. E1 α subunit of plastid pyruvate dehydrogenase (PHD E1α),
D. PSII light harvesting complex (LHCII) and Rubisco small subunit (SSu).
Expression is related to 9+10 DAF, and the ratios (average of two replicates) were rescaled by setting the lowest ratio to 1.

The pathway of carbon from sucrose to fatty acids

In addition to cytosolic glycolytic pathway, plastids of oil seeds also have a complete set of glycolytic enzymes (Kang and Rawsthorne 1994), but it is still unclear how these two pathways are utilized during seed filling. Results from the *Arabidopsis* seed EST project (White et al 2000) suggested that the major route of carbon into oil involves cytosolic glycolysis up to PEP, and PEP is then imported to plastids and converted to pyruvate and AcCoA. The microarray results agree with this notion. As the lipid synthesis increased, the expression of cytosolic pyruvate kinase decreased (Fig 2A), whereas the expression of the plastid form was similar to many FAS enzymes (Fig 2B). The E1 α subunit of the plastid pyruvate dehydrogenase, the last enzyme in the pathway of carbon to acetyl-CoA, behaved similarly as well (2C). These results support the model of an active cytosolic glycolysis where PEP is imported into plastids and processed there.

Developing *Arabidopsis* embryos are green, contain functioning chloroplasts and are photosynthetically competent. It is still unknown how much seed photosynthesis contributes to seed filling, either in producing reducing equivalents or refixing CO_2 liberated by FAS. We found that the expression of several genes from both photosynthetic light (such as LHC II, PSII oxygen evolving complex) and dark reactions (Rubisco, phosphoribulokinase,) was lowest in the youngest seeds but increased strongly by 11 DAF and then decreased again, following the profile of several FAS enzymes (Fig. 2D). This suggests that the embryo's maximal photosynthetic capacity coincides with the oil accumulation period and thus has the potential to contribute to the carbon acquisition (King et al 1998).

Summary

We have analyzed the expression patterns of approximately 4000 genes during seed development. About 1200 genes, whose expression changed significantly during the seed filling, were identified. Shifts in metabolism during seed development were associated with changes in transcription profiles of many enzymes/proteins involved in storage product accumulation. Although not discussed in this paper, these experiments also identified many transcription factors, kinases and phosphatases whose expression correlated with FAS enzyme expression, and so may participate in the regulation of the pathway. These regulatory factors represent good candidates for future molecular biology studies. Since the whole *Arabidopsis* genome is fully sequenced, it is also possible to study whether the genes that share the same expression pattern also share common regulatory sequences in their promoter regions.

Acknowledgements

This work was supported in part by grants from the Consortium for Plant Biotechnology Research and Dow AgroSciences. We also acknowledge the Michigan Agricultural Experiment Station for its support of this research.

References

Bowman J (Ed.), (1994). *Arabidopsis: an atlas of morphology and development*, New York : Springer-Verlag, pp 352-353

Browse J, McCourt P, Somerville CR (1986) *Analytical Biochemistry* **152**, 141-145.

Focks N, Benning C (1998) *Plant Physiology* **118**, 91-101.

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge JB (2000) *Plant Physiology* **124**, 1570-1582.

Goldberg RB, Hoschek G, Tam ST, Ditta G, Breidenbach RW (1981) *Developmental Biology* **83**:201-217

Hihara Y, Kamei A, Kanehisa M, Kaplan A, Ikeuchi M (2001) *Plant Cell* **13**, 793-806.

Kang F, Rawsthorne S (1994) *Plant Journal* **6**:795-805

King SP, Badger MR, Furbank RT(1998) A*ustralian Journal of Plant Physiology* 25, 377-386.

Post-Beittenmiller D, Ohlrogge JB, Somerville CR (1993) In: *Control of Plant Gene Expression*, Ed. Verna S, Boca Raton, CRC Press, pp 157-174.

Ruuska S, Ohlrogge JB (2001) *Biotechniques*, In Press

White J, Todd J, Newman T, Focks N, Girke T, Martinez de Ilarduya O, Jaworski JJ, Ohlrogge JB, Benning C (2000) *Plant Physiology* **124**, 1582-1594.