Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Phytochemistry 63 (2003) 753–763
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Metabolic engineering to increase isoflavone biosynthesis in soybean seed

Oliver Yu[1], June Shi, Aideen O. Hession, Carl A. Maxwell, Brian McGonigle*, Joan T. Odell

E.I. du Pont de Nemours & Company, Inc., Crop Genetics, Experimental Station, PO Box 80402, Wilmington, DE 19880-0402, USA

Received 6 February 2003; received in revised form 3 April 2003

Abstract

Isoflavone levels in *Glycine max* (soybean) were increased via metabolic engineering of the complex phenylpropanoid biosynthetic pathway. Phenylpropanoid pathway genes were activated by expression of the maize C1 and R transcription factors in soybean seed, which decreased genistein and increased the daidzein levels with a small overall increase in total isoflavone levels. Cosuppression of flavanone 3-hydroxylase to block the anthocyanin branch of the pathway, in conjunction with C1/R expression, resulted in higher levels of isoflavones. The combination of transcription factor-driven gene activation and suppression of a competing pathway provided a successful means of enhancing accumulation of isoflavones in soybean seed.

Keywords: *Glycine max*; Fabaceae; Isoflavone; Transcription Factor; Flavanone 3-Hydroxylase

1. Introduction

Isoflavones are synthesized predominantly in legumes, where they are an important class of compounds that mediate multiple plant-microbial interactions. In *Glycine max* (L.) Merr. (soybean), isoflavones both attract Rhizobial bacteria and induce *nod* gene expression to initiate nitrogen-fixing root nodule formation (Van Rhijn and Vanderleyden, 1995; Pueppke, 1996). Several isoflavone compounds have antifungal activity (Rivera-Vargas et al., 1993) and/or are precursors to major phytoalexins (Blount et al., 1992; Graham, 1995; Garcia-Arenal et al., 1978). Consumption of isoflavones is also associated with human health benefits such as decreased risk of heart disease, reduced menopausal symptoms, and reduced risk of some hormone-related cancers (Davis et al., 1999; Messina, 1999; Watanabe et al., 2002; Clarkson, 2002).

In soybean, the isoflavones daidzein **1,** genistein **2**, and glycitein **3** are synthesized via the phenylpropanoid pathway and stored in the vacuole as glucosyl- and malonyl-glucose conjugates (Graham, 1991). The pathway to daidzein **1** branches from the phenylpropanoid pathway, that is common to most plants, following the chalcone synthase catalyzed reaction (Fig. 1) through a legume-specific enzyme, chalcone reductase (CHR). Glycitein **3** synthesis is not yet clearly defined, but is likely to be derived from isoliquiritigenin (Latunde-Dada et al., 2001). Genistein **2** synthesis shares the naringenin intermediate with the flavonoid/anthocyanin branch of the phenylpropanoid pathway. In all cases the unique aryl migration reaction to create the isoflavones is mediated by isoflavone synthase (IFS), the encoding gene of which has been identified (Akashi et al., 1999; Steele et al., 1999; Jung et al., 2000).

Genes in the phenylpropanoid pathway are activated in maize cells by the maize transcription factors C1 and R leading to the synthesis and accumulation of anthocyanins (Dooner et al., 1991; Grotewold et al., 1998). C1 is a myb-type transcription factor that requires interaction with a basic helix-loop-helix-containing R myc-type factor to be effective. A fusion in which the R protein is inserted between the DNA binding and activation

* Corresponding author. Tel.: +1-302-695-9176; fax: +1-302-695-4509.
E-mail address: brian.mcgonigle@usa.dupont.com (B. McGonigle).
[1] Current address: Donald Danforth Plant Science Center, St. Louis, MO, USA.

0031-9422/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0031-9422(03)00345-5

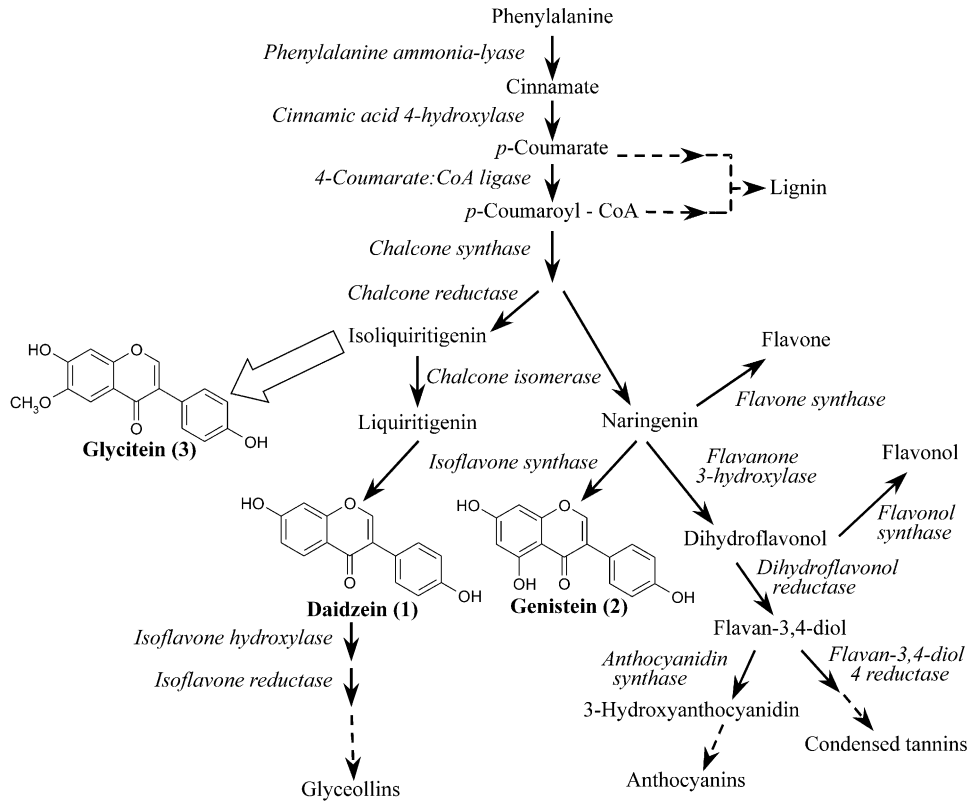

Fig. 1. A partial diagram of the phenylpropanoid pathway. Intermediates and enzymes involved in isoflavone synthesis, as well as some branch pathways, are shown. Dotted arrows represent multiple steps; the block arrow represents speculative steps.

domains of C1, called CRC, is also effective in causing anthocyanin production in maize cells (Bruce et al., 2000). Expression of maize *C1* and the *Lc* allele of *R* in stably transformed *Arabidopsis thaliana* plants produced anthocyanins in tissues that do not normally synthesize these pigments, while in *Nicotiana tabacum* transformants containing these genes the anthocyanin levels in already pigmented tissues were increased (Lloyd et al., 1992). Transient expression of *C1* and the *B-Peru* homologue of *R* produced anthocyanins in some tissues of the legumes *Trifolium repens* (white clover) and *Pisum sativum* (pea; De Majnik et al., 1998). These results indirectly show, by using anthocyanin production as a marker, that the maize C1 and R transcription factors can activate phenylpropanoid pathway genes in some dicots.

Transformed maize cell culture lines expressing soybean IFS could produce genistein **2** when CRC was also expressed, but genistein **2** was undetectable without CRC expression (Yu et al., 2000). In transgenic *Arabidopsis* plants expressing the soybean IFS gene, the level of genistein **2** produced was enhanced when the phenylpropanoid pathway was activated by high UV-light. In IFS transgenic tobacco plants, genistein **2** accumulated to higher levels in anthocyanin-producing flowers than in leaves. All of these results establish a correlation between the activity of the phenylpropanoid pathway and the level of genistein **2** produced in non-legume IFS transgenic plant tissues. In soybean, isoflavones are constitutively expressed in some tissues, such as seed and roots, while in other tissues such as hypocotyls and leaves they may be present constitutively, but are further induced in response to elicitor treatment or pathogen attack (Graham and Graham, 1991). Transcription factors endogenous to legumes that specifically mediate this response are unknown. However, general activation of the phenylpropanoid pathway has the potential to raise isoflavone levels, based on our observation that the soybean IFS was able to compete for the naringenin pathway intermediate in the maize cell culture, *Arabidopsis*, and tobacco systems.

In the present study, we tested whether CRC has an effect on seed isoflavone levels and isoflavone pathway intermediates, and whether genes of the phenylpropanoid pathway are activated by expression of CRC in soybean seed. The results suggest a strategy for developing lines of soybeans that accumulate isoflavones to much higher levels than in wild-type seed. By combining gene activation using the transcription factor, and blocking a competing branch pathway, isoflavone synthesis was engineered such that accumulation was up to four times higher than in wild-type seed. These high isoflavone soybeans would be useful for the production of soy foods providing potentially greater health benefits to consumers.

2. Results

2.1. Expression of CRC in soybean seed affects individual isoflavone levels

A chimeric gene (CRC) containing a fusion of maize nucleotide sequences encoding C1 and the Lc allele of R (Bruce et al., 2000), under control of the seed-specific phaseolin promoter and termination signals, was introduced into soybean embryogenic suspension cultures by particle gun bombardment. Transformed lines carrying the CRC gene were selected, identified by PCR, and regenerated. Single seed of independent CRC positive transformation events were analyzed by HPLC. Amounts of the isoflavone aglycones as well as the glucose-conjugates (daidzin, genistin, and glycitin) and malonyl-glucose conjugates were determined for each individual isoflavone and summed to give the total amounts for daidzein **1**, genistein **2**, and glycitein **3**.

The isoflavone component profile, shown for a control transformant that was PCR negative for the CRC gene, contains genistein **2** as the most abundant of the isoflavones (Fig. 2). Daidzein **1** is generally the next highest-level component with glycitein **3** lowest. Levels of individual isoflavones, as well as the total isoflavone levels varied among individual seed from the same plant; in this example the total isoflavone levels varied over a 3-fold range. Nevertheless the isoflavone component profile changed in seed from four of the 13 independent transformation events analyzed. The R1 seed from the hemizygous primary transformants would be expected to segregate for the transgene. As shown in Fig. 2 for the two independent transformed lines 2590-1-1 and 2590-1-2, there are individual R1 seed with the control isoflavone component profile, and seed with greatly increased levels of daidzein **1** and greatly reduced levels of genistein **2**. In the altered phenotype seed of the 2590-1-1 and 2590-1-2 events, daidzein **1** contributes 65–75% of the total isoflavones, while daidzein **1** is generally 25–35% of the total in control and wild-type segregating seed. In the same altered seed the genistein **2** component is close to zero, while in control and wild-type segregating seed it makes up about 50–60% of the total isoflavones.

In seed with altered isoflavone content, as in control seed, the majority of each isoflavone was about equally divided between glucose and malonyl–glucose conjugates, with very little aglycone present. These results indicate that sufficient enzyme activities are present to conjugate daidzein **1,** whose levels are much higher than those found in control seed, as well as genistein **2** and glycitein **3**. While the total isoflavone levels vary among

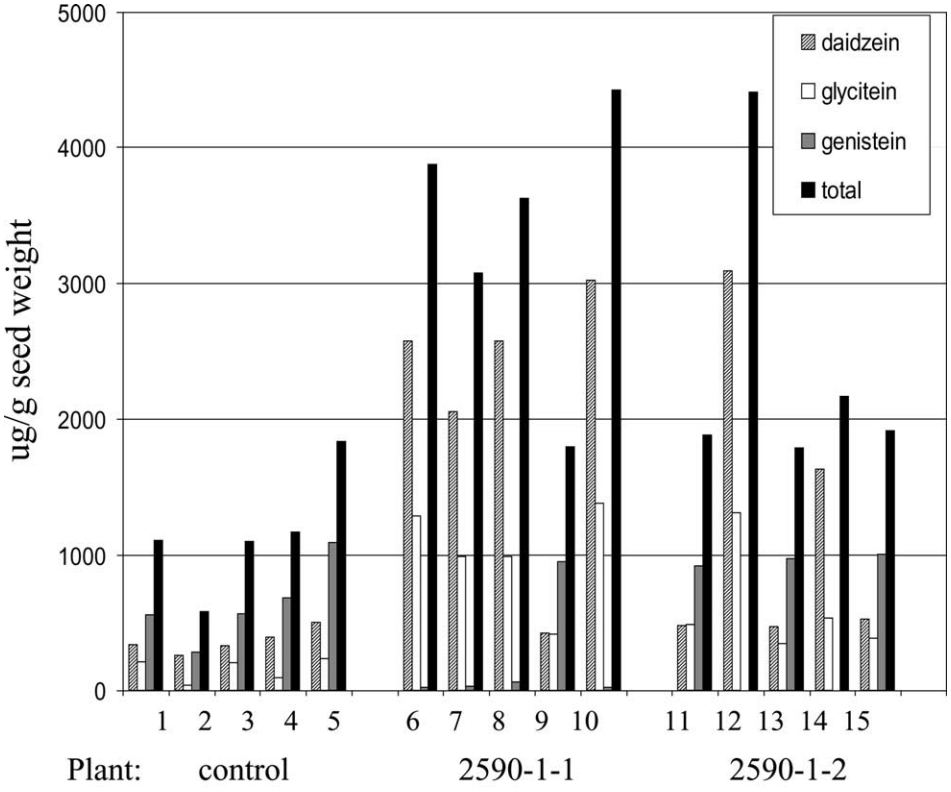

Fig. 2. Isoflavone components of CRC transformant R1 soybean seed. Plots of totals of individual isoflavone aglycone equivalents, as well as the totals of all isoflavones, determined by HPLC analysis of extracts prepared from individual R1 seed from primary transformants and reported in μg/g seed weight. Individual seed are numbered 1–15. The control plant is a transformant that is negative for the CRC gene. Two individual lines positive for CRC are labeled 2590-1-1 and 2590-1-2.

individual seed, CRC did in some cases raise the level about 2-fold (Fig. 2).

2.2. A visual phenotype segregates with the altered isoflavone composition

In CRC events producing seed with high daidzein **1** and low genistein **2**, a portion of the seed had a dark brown stripe around the median on the side opposite to the hilum, parallel to the cotyledon axis (Fig. 3), as opposed to the overall light tan wild-type phenotype. Some of the striped seed were smaller than control seed, and some were slightly wrinkled. Removing the seed coat showed that the brown pigmentation did not extend into the cotyledons, but was limited to the external coat.

To investigate a potential correlation between the visual phenotype and isoflavone profile, plants from either tan or striped R1 seed produced by primary transformants were grown in a growth chamber. Plants grown from tan seed were PCR negative for the CRC gene while plants grown from striped seed were PCR positive for the CRC gene. Plants of the 2590-1-1 and 2590-1-2 lines that were grown from tan seed produced only tan seed. Plants of the 2590-1-1 and 2590-1-2 lines that were grown from striped seed all produced segregating striped and tan R2 seed. For the 2590-1-2 line the segregation ratio was 3:1 striped to tan, indicating a single gene dominant trait. For the 2590-1-1 line, the segregation ratio was 2–1, suggesting either lower penetrance of the trait or a possible association with a recessive seed lethal phenotype. Isoflavone levels were analyzed in individual striped and tan seed. In every case, the striped seed showed the high daidzein **1** and low genistein **2** phenotype described above, while tan seed had the control profile. This correlation provides a means of identifying CRC homozygotes and wild-type segregants.

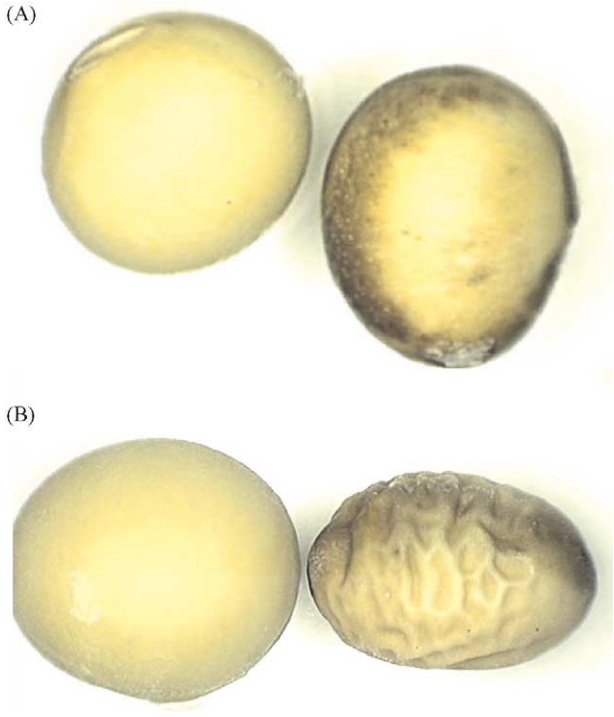

Fig. 3. Seed phenotypes from CRC and control plants. (A) and (B) Control seed are on the left and CRC seed on the right. The brown stripe, and in some cases wrinkling (B), characterize the CRC seed phenotype.

2.3. Identification of intermediates that accumulate in CRC seed

Expression of CRC in soybean seed had effects on levels of many compounds other than the isoflavones, as determined by differences in the HPLC profiles between striped and wild-type segregant seed extracts (data not shown). Three peaks that are almost undetectable in wild-type segregant seed, but present in the striped seed, were determined using mass spectrometry, as each having an m/z of 505 but differing HPLC retention times of 15.46, 21.29, and 21.75 min. (Fig. 4). MS2 analysis produced one major fragment with an m/z of 257 for each of the compounds. This mass indicates the loss of a fragment with a mass of 248, consistent with the mass of malonyl-glucose. The 257 ion was further fragmented to 239, 147, and 137 species again for all three unknowns. Liquiritigenin and isoliquiritigenin, intermediates in daidzein **1** synthesis (Fig. 1), both have a mass of 256 and MS2 fragments of 239, 147, and 137. The UV spectrum of liquiritigenin matches that of the unknown

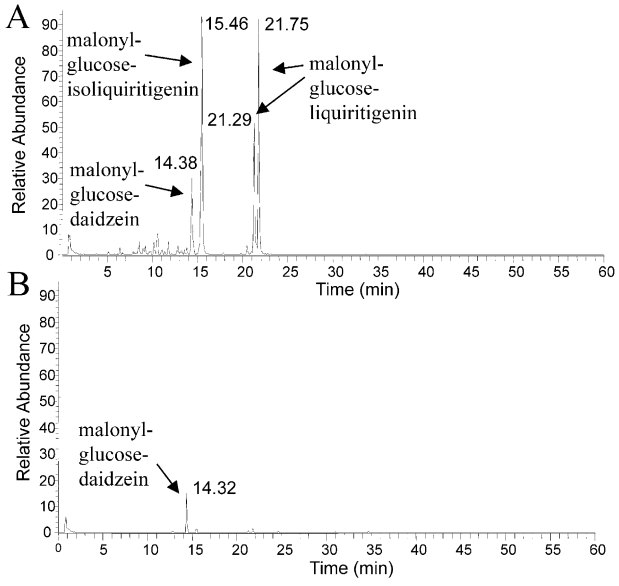

Fig. 4. LC-MS2 mass chromatograms of soybean seed extracts showing accumulation of intermediates. Ion chromatograms of m/z 504.6 to 505.6 of extracts of: (A) R3 seed from CRC line 2590-1-1 and (B) control seed. The peaks are labeled as identified by further analysis that is described in the text.

with 15.46 min. retention time, while the spectra of the unknowns at 21.29 and 21.75 both are similar to the isoliquiritigenin UV spectrum (data not shown). The differences between the HPLC retention times of the unknowns and those of liquiritigenin and isoliquiritigenin are also similar to the differences between retention times for other flavonoid aglycones and their corresponding malonyl-glucose conjugates. All results suggest that the 15.46 min. peak is a malonyl–glucose conjugate of liquiritigenin, and the unknowns at 21.29 and 21.75 min. are malonyl-glucose conjugates of isoliquiritigenin.

Neither flavones nor anthocyanins typically accumulate in commodity soybean seeds and we were unable to detect either of those classes of compounds in CRC seeds. Flavonol levels were increased in CRC seed, and the results are presented below.

2.4. Effects of CRC on expression levels of phenylpropanoid pathway genes

A homozygous CRC line from the 2590-1-1 event was identified as a plant producing only brown striped seed. R3 seed from this line continued to have high daidzein **1** and low genistein **2** levels as in previous generations. Plants were grown in a growth chamber from this line, as well as from a wild-type segregant line of the same event. Immature R4 seed were harvested at weights of approximately 150 mg and 250 mg, which correspond to approximately 10 and 20 days after fertilization, and used for protein and RNA preparations. Seed from the CRC homozygotes were confirmed to be expressing the CRC transgene by RT-PCR.

A Western blot of protein extracts from the control and CRC seed using an antiserum raised against chalcone synthase (CHS) showed that the level of CHS protein increases dramatically in seed with the CRC transgene at both stages (Fig. 5A). In contrast, an IFS antiserum detected no change in the amount of IFS protein at either stage.

For analysis of RNA produced from different phenylpropanoid pathway genes, probes were made from clones in the DuPont soybean EST libraries that were identified by sequence homology with published sequences to encode phenylalanine ammonia-lyase (PAL), cinnamic acid 4-hydroxylase (C4H), chalcone isomerase (CHI), chalcone reductase (CHR), IFS, flavanone 3-hydroxylase (F3H), dihydroflavonol reductase (DFR), flavonol synthase (FS), and isoflavone reductase (IFR). Results from Northern blots of RNA prepared from seed of the two stages of development using these probes showed that expression of PAL, C4H, CHI, CHR (the branch point for daidzein synthesis), F3H, DFR and FS (all three in the flavonol/anthocyanin branch) was increased in the CRC seed (Fig. 5B and C). In agreement with the Western results, the IFS RNA levels remained similar to the control levels in both

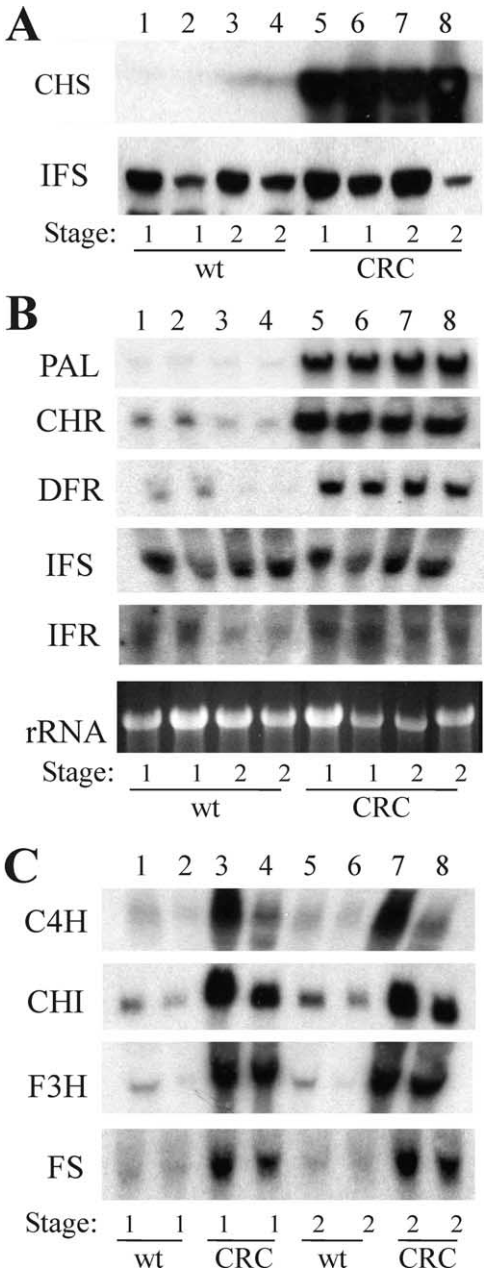

Fig. 5. Gene expression analysis of CRC seed. (A) Western blot of protein extracts of: lanes 1 and 2, 150 mg seed (stage 1) from two control plants; lanes 3 and 4, 250 mg seed (stage 2) from the same control plants; lanes 5 and 6, 150 mg seed from two plants homozygous for CRC, events 2590-1-1 and 2590-1-2, respectively; lanes 7 and 8, 250 mg seed from the same CRC plants. Separate blots were probed with antisera to CHS and IFS and the protein bands corresponding to the correct sizes labeled. (B) Northern blot of RNA extracted from seed in the same order as described in (A). Probes made to specific genes were hybridized to separate blots and the bands representing the mRNA for each gene are labeled. The stained rRNA panel shows the RNA loading. (C) Northern blot of the same RNA samples as in B, but in the following order: lanes 1 and 2, 150 mg control; lanes 3 and 4, 150 mg CRC 2590-1-1 and 2590-1-2; lanes 5 and 6, 250 mg control; lanes 7 and 8, 250 mg CRC 2590-1-1 and 2590-1-2.

stages of seed development. Expression of IFR, an enzyme involved in the synthesis of glyceollins from daidzein **1**, was not increased in the younger seed and had only a slight increase in the older seed.

2.5. Expression of CRC in conjunction with cosuppression of flavanone 3-hydroxylase causes high levels of isoflavone accumulation

The enzyme flavanone 3-hydroxylase (F3H; EC 1.14.11.9) catalyzes the conversion of flavanones to dihydroflavonols, which are intermediates in the biosynthesis of flavonols, anthocyanidins, catechins and proanthocyanidins. In soybeans, F3H and IFS compete for the naringenin substrate. The increased expression of F3H in CRC transformant seed combined with reduced levels of genistein suggested that suppression of F3H might increase naringenin availability for IFS, thereby allowing higher levels of genistein **2** to be synthesized. To test this possibility, soybean transformants were generated that contain the CRC gene and a gene designed to promote cosuppression of F3H: a soybean F3H coding region flanked by nucleotide sequences that promote formation of a stem-loop structure and expressed using the *Kti3* promoter and termination signals. The presence of both genes in these transformed lines was verified by PCR. Transformants with only the F3H construct were generated independently.

Bulk R1 seed samples from independent transformation events containing both CRC and the F3H construct, as well as events containing the F3H construct alone, were assayed for isoflavone composition and level. Seed extracts were treated with base to convert the malonyl-glucose conjugates to glucose conjugates, and the three glucose conjugates (genistin, daidzin, and glycitin) were quantified. While the aglycone is not measured by this method, the amount of aglycone is present presumably in such low quantities as to not affect the final results. The isoflavone numbers are reported as μg of glucosyl-conjugate per gram of seed weight. The bulk seed samples show much reduced variation as compared to single seed samples. Rather than the 3-fold variation between individual seed as shown in the Fig. 2 control, bulk seed samples from controls grown under similar conditions showed only 1.6-fold variation.

No transformants containing the F3H transgene alone in 27 independently generated events had significant increases in total isoflavone levels as compared to the 3000 μg g^{-1} levels in wild-type plants (data not shown). However, seed from 11 out of 91 (12%) independently generated events containing both CRC and the F3H transgene accumulated up to 4-fold higher levels of isoflavones than controls (Fig. 6). Furthermore, these soybeans had increased levels of genistein **2** as compared to the CRC alone seed, and there was no brown stripe or wrinkling. In the next generation, the R2 seed showed further increases in genistein **2** to regain a profile similar to that of controls, and retained the high total isoflavone level (Fig. 6) and normal phenotype.

The level of flavonols that accumulate varies in different cultivars of soybean. Analysis of flavonols in R4 CRC seeds derived from event 2590-1-1 showed a substantial accumulation of the flavonol kaempferol (740 μg g^{-1}) as compared to wild type Jack soybeans (2.9 μg g^{-1}). We then analyzed R2 seed from eleven plants derived from event 277-6-3-1 containing both CRC and the F3H transgene and found levels of kaempferol that ranged between 30 and 150 μg g^{-1} soy, with an average accumulation of 68 μg g^{-1}. This is evidence that the increase in isoflavone levels observed in CRC/F3H plants is due to the down-regulation of flavone 3-hydroxylase, since lower amounts of flavonols accumulate. Presumably if F3H could be totally suppressed so that flavonol levels accumulate similarly to levels found in Jack soybeans, total isoflavone levels could be raised even further.

3. Discussion

The CRC maize transcription factor fusion, which increases anthocyanin levels in maize tissues (Bruce et al., 2000), caused major changes in the levels of isoflavone components of soybean seed: total daidzein **1** was greatly increased, while total genistein **2** was decreased. One explanation for the change in relative levels of daidzein **1** and genistein **2** is diversion of intermediates away from genistein **2** into the daidzein **1** branch through increased CHR activity (Fig. 1). Levels of transcripts encoding CHR, a legume specific enzyme, were substantially increased in the CRC seed. Increased expression of three genes from the flavonol/anthocyanin branch, F3H, DFR and FS, also suggests that intermediates are directed away from genistein **2** into the synthesis of other compounds in this pathway. In fact, increased F3H would directly compete for the naringenin intermediate with IFS, which does not show enhanced expression. IFS activity appears to be limiting for the synthesis of daidzein **1**, since we identified liquiritigenin and isoliquiritigenin as intermediates that accumulate in CRC seed but not in controls. Accumulation of daidzein **1** may not be limited by further metabolism, since the unchanged expression of IFR suggests that the genes encoding the enzymes for biosynthesis of glyceollins from daidzein **1** are not induced.

Genes in the phenylpropanoid pathway, including PAL, C4H and CHS, were shown to have increased expression in CRC seed. Expression of another phenylpropanoid pathway gene, CHI, is increased in CRC seed unlike in Black Mexican Sweet maize cells (Grotewold et al., 1998; Bruce et al., 2000). However, the accumulation of isoliquiritigenin indicates that the CHI

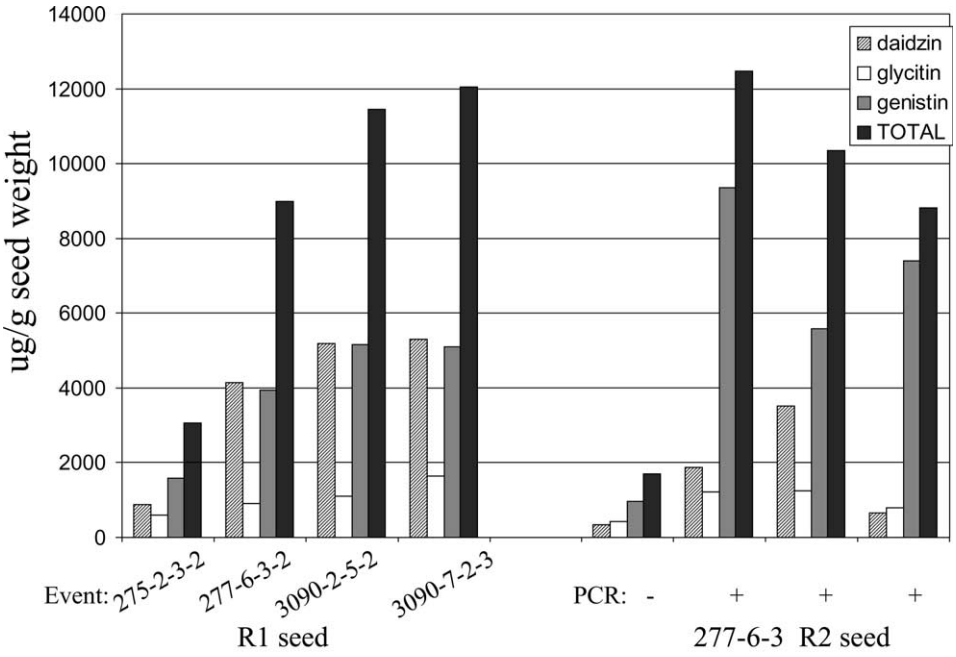

Fig. 6. Isoflavone components of CRC and F3H transformant R1 and R2 soybean seed. Plots of glucose conjugates of individual isoflavones, as well as the totals of all isoflavones, determined by HPLC analysis of extracts prepared from a bulk R1 seed sample from each of four transformation events on the left (events are named 275-2-3, 277-6-3, 3090-2-5, and 3090-7-2). Event 275-2-3 is PCR positive for both genes but shows a phenotype similar to wild type soybeans. On the right is shown the isoflavone analysis of a bulk R2 seed sample from each of four individual plants grown from the 277-6-3 R1 seed. The 277-6-3 R2 seed samples are labeled as negative or positive for PCR analysis of the two transgenes in the parent plant.

enzyme activity is still limiting for daidzein **1** production. This increased expression of phenylpropanoid pathway genes should lead to increased flow of intermediates through the pathway, as reflected in the accumulation of liquiritigenin and isoliquiritigenin, and in elevated total isoflavone levels in some seed. Anthocyanins do not appear to accumulate in the CRC seed, but kaempferol and associated conjugates do as well as unidentified peaks from HPLC analysis, many of which may represent newly accumulating phenylpropanoid pathway compounds. Accumulation of new compounds may be related to the shrunken and wrinkled appearance of some CRC seed, and to the brown stripe in the seed coat that is correlated with the high daidzein **1** and low genistein **2** phenotype. Compound(s) causing the brown stripe must be transported from the embryo to the seed coat, since this phenotype does not show maternal inheritance as would be expected for a seed coat-expressed trait. Since a brown stripe does not develop in the CRC/F3H seed, it is suggestive that in the CRC seeds the stripe may be composed of flavonols or proanthocyanidins. We have used vanillin-hydrochloric acid as a histochemical test for proanthocyanidins (Gardner, 1975), and have found that proanthocyanidins are present at higher levels in CRC seed. However, the proanthocyanidins accumulation does not appear to co-localize with the brown stripe, instead the vanillin-hydrochloric acid causes a dark red stain formed around the hilum.

Activation of genes by CRC may be through direct binding and activation of promoters of these genes. However, some genes may be activated through feedback of accumulating intermediates. For example, accumulation of the *p*-coumarate intermediate stimulates expression of a CHS gene in alfalfa protoplasts (Loake et al., 1992). Thus the increased expression of CHS in soybean seed may be due to activation of a CHS gene promoter directly by CRC, or to accumulation of *p*-coumarate following activation of PAL and C4H gene expression. Expression of genes encoding other enzymes in the pathway may also be affected by altered levels of intermediates.

A goal of this project was to develop high isoflavone soybean seeds for use in producing soy-food products that may provide enhanced health benefits to consumers. The results of the analysis of the CRC transformed lines, showing reduced genistein **2** and increased F3H transcript, suggested that suppression of F3H in addition to CRC expression might further raise the levels of isoflavones. Through this metabolic engineering strategy, it was possible to produce soybean seed that accumulate up to 4-fold higher levels of isoflavones than in wild-type. Liu et al. (2002) recently reported that expression of IFS in the *tt6*/*tt3* Arabidopsis mutant, which is impaired in expression of both F3H and DFR, allowed the accumulation of genistein **2** to much higher levels than IFS expression in wild type Arabidopsis. In our experiments in soybean seed, where IFS expression is the wild type state, suppression of F3H had no significant

effect on isoflavone levels. Thus flow of intermediates through F3H does not appear to be a limiting step for isoflavone production in soybean seed, as it was found to be in Arabidopsis. It is only in the presence of CRC, where expression levels of multiple genes are enhanced, including F3H, that suppression of F3H leads to increased accumulation of isoflavones in soybean seed.

The high isoflavone seed have a profile of individual isoflavone components in R2 seed similar to wild-type and a normal phenotype. The levels of isoflavone accumulation are higher than in any wild-type cultivar of soybean of which we are aware. Thus the combination of transcription factor activation of multiple pathway genes and cosuppression of a single gene, F3H, the product of which competes with the biosynthesis of target compounds, has provided an effective targeted metabolic engineering strategy for producing high levels of isoflavones in soybean seed.

4. Experimental

4.1. Vector construction

To prepare the chimeric CRC gene a SmaI fragment containing the CRC fusion was prepared from pDP7951, a precursor of PHP6680 (Bruce et al., 2000) and inserted between the French bean (*Phaseolus vulgaris*) β-phaseolin promoter and 3′ polyadenylation signal region (Doyle et al., 1986; Slightom et al., 1991) to create pOY135. The entire Ph/P-CRC-Ph 3′ gene was transferred to pZBL102, a vector containing a hygromycin phosphotransferase coding region expressed from the CaMV 35S promoter, to create pOY203.

A *Glycine max* (soybean) clone with homology to known genes encoding F3H was identified from the DuPont EST collection (McGonigle et al., 2000) by a BLAST search (Altschul et al., 1990, 1997). The encoded protein has a BLAST X pLog score of 168 to flavanone 3-hydroxylase from *Daucus carota* [AF184270 (Hirner et al., 2001)]. A 500 bp portion of coding region of the gene was amplified and a Not I site was added with the primers GCGGCCGCATGGCACCAACAGCCAAG and GCGGCCGCATCCGTGTGGCGCTTCAG using advantage 2 polymerase and GC melt reagent (Clontech). The resulting amplification product was cloned into pCR2.1 using the TOPO TA cloning kit according to the manufacture's instructions (Invitrogen). The NotI fragment was isolated and inserted between synthetic inverted repeat sequences of 40 bp (Glassman et al., 2002) that are bounded by the *Kti3* promoter and polyadenylation signal sequences in a vector containing a hygromycin phosphotransferase coding region expressed from the CaMV 35S promoter to create pAC21.

4.2. Generation of soybean transformants

Soybean (cv Jack) embryogenic suspension cultures were transformed with plasmid pOY203 by particle gun bombardment (Klein et al., 1987) following previous protocols (Kinney and Fader, 2001). The bombarded DNA sample also included a plasmid containing the alpha′ β-conglycinin promoter-IFS gene, but in this study only lines without this construct, as determined by PCR, were chosen. Lines with the CRC gene were identified by PCR using primers: 5′-AGGCGGAAGAACTGCTGCAACG and 5′ - AGGTCCATTTCGTCGCAGAGGC. Transformed embryos were germinated and grown to maturity according to above protocols. All plants were allowed to self.

Soybean cultures were similarly transformed with pOY135 and pAC21 in a 10:1 ratio and lines containing the F3H and CRC transgenes were identified by PCR. The F3H transgene was detected using primers 5′ - TCCTCAGTCACCGATCTCCACCC and 5-CGGATATAATGAGCCGTAAACA or primers 5′-TGGATGGACGCAGAAGAGAGATTTG and 5′-CCGATTCTCCCAACATTGCTTATTC. CRC was detected using the primers above. Lines containing the F3H transgene alone were prepared by transforming with pAC21 only.

4.3. Isoflavone analysis

Individual R1 or R2 seed of CRC and control lines were ground, extracted with MeOH–H_2O (4:1) and analyzed by HPLC (model 1100, Agilent, Wilmington, DE) equipped with a diode array detector and a Luna C18 column (4.6×150 mm, 3 micron, Phenomenex, Torrence, CA) maintained at 25 °C. The column was eluted at 1 ml/min with a linear solvent gradient from 95% water and 5% methanol with 0.1% trifluoroacetic acid (TFA) to 100% methanol with 0.1% TFA over 16 min. The quantity of daidzein **1**, glycitein **3**, genistein **2**, and their conjugate derivatives was calculated by comparison with standard curves prepared from authentic compounds (Indofine Chemical Co., Somerville, NJ; Fujico Co., Japan) at 262 nm.

Seed of F3H-CRC or F3H lines were ground in batches of five to eight per plant. A 1 gram sample was extracted with MeOH–H_2O (4:1) and then incubated with 2 N NaOH at room temperature to hydrolyze malonyl and acetyl esters to the corresponding glucosides. Acetic acid was used to neutralize pH, samples were filtered and then assayed by HPLC (model 2690, Waters, Milford, MA) equipped with UV detector (model 486, Waters) and with a Luna C18 column (3 micron, 4.6×50 mm, Phenomenex) maintained at 30 °C. The column was eluted with 90% A and 10% B (A as 1% acetic acid in water and B as 1% acetic acid in acetonitrile) for 5 min at 1 ml/min, 10% B to 22% B from 5

to 11 min at 1 ml/min, 22% B to 100% B from 11 to 12 min at 1 ml/min to 2 ml/min and 100% B from 12 to 14.5 min at 2 ml/min. The amounts of daidzin, glycitin and genistin were determined by comparison with authentic compounds as above.

4.4. Identification of liquiritigenin and isoliquirtigenin by LC-MS2

R4 seed were ground, extracted with MeOH–H_2O (4:1) and analyzed by LC-MS-MS on an HPLC (model 1100, Agilent) with an ion-trap mass spectrometer (LCQ, Thermo Finnigan, San Jose, CA). Each sample was separated on a Zorbax C18 column (4.6×75 mm, 3.5 micron, Agilent) maintained at 40 °C. The column was eluted at 1 ml/min with a gradient of 95% A and 5% B to 50% B (A as 0.1% formic acid in water and B as 0.1% formic acid in acetonitrile) from 0 to 45 min, then 50% B to 100% B from 45 to 50 min and held at 100% B from 50 to 60 min. The solvent stream was split post-column with 300 μl per min delivered to an electrospray ion source, operating in positive ion mode, and the remaining 700 μl flowed through a diode array detector with spectra collected from 200 to 400 nm. Ions were monitored in full scan MS1 from m/z 150 to 2000 and subsequently in full scan m/z 100–800 with MS2 performed on the most abundant ion. Additionally, extracts were analyzed using source collision-induced dissociation (CID) and monitored full scan m/z 100 to 800 with MS2 performed on the most abundant ion. Source CID, MS2 parameters and ion optics were optimized by infusion with daidzin (Indofine).

4.5. Northern blot analyses of genes of the phenylpropanoid pathway

RNA was extracted from R3 seed (5) using a modified Trizol method (Gibco BRL, Life Technologies, Rockville, MD). RNA (30 μg) was loaded per lane of a precast Rilant RNA Gel (FMC, Rockland, ME). The RNA components were separated by electrophoresis and transferred to a membrane following standard protocols.

Probes were prepared from soybean cDNA clones in the DuPont collection with insert coding regions having 95% or greater homology to soybean sequences with the NCBI Accession numbers: PAL: S46988, C4H: X92437, CHI: AF276302, CHR: X55730, IFR: AJ003245 and two clones encoding DFR: AF167556. The F3H clone was described above. The FS sequence has a BLAST N pLog score of 115.74 to AF240764 from *Eustoma grandiflorum*. IFS was described previously (Jung et al., 2000). Probes were prepared using the Random Primers DNA Labeling System from GIBCO-BRL, Life Technologies. The entire plasmid was used as template for all probes, except for IFS, where the template was a PCR product containing the IFS coding region (5′-TTGCTGGAACTTGCACTTGGT-3′ and 5′-GTATATGATGGGTACCTTAATTAAGAAAGGAG - 3′) amplified from pOY204 (Yu et al., 2000). The entire random primer reaction was used for hybridization in PerfectHyb Buffer (Sigma-Aldrich, St. Louis, MO) at 68 °C overnight. Membranes were washed twice with 2 X SSC (GIBCO BRL, Life Technologies) and once with 0.1 X SSC for 15 min each at 68 °C then exposed to film.

4.6. Immunoblot analyses of genes of the phenylpropanoid pathway

Approximately 100 mg of ground R3 seed were extracted with 500 μl of 50 mM Tris-hydrochloride, pH 7.5, 10 mM 2-mercapto-ethanol, and 0.1% SDS. Antibodies to CHS and CHR were prepared by Covance (Richmond, CA) to protein purified from *E. coli* expressing the CHS or CHR coding region from the above EST clones using standard methods. The IFS antibody was prepared to a synthetic peptide of the IFS protein as described previously (Yu et al., 2000). Standard protocols were used for immunoblot analyses with anti-CHS, anti-CHR, or anti-IFS antisera. The secondary antibody was anti-rabbit IgG conjugated to horseradish peroxidase (Amersham Life Science, Arlington Heights, IL), with visualization using Supersignal West Pico chemiluminescent substrate (Pierce, Rockford, IL), or Supersignal West Femto for IFS.

4.7. Flavonol quantification

Flavonol levels were determined by a modification of the method presented by Nuutila et al. (2002). Batches of 8 seeds were ground and 100 mg extracted with acetonitrile–H_2O (3:2). Ascorbic acid was added to the extract and the flavonol conjugates were hydrolyzed with 1.2 N HCl at 80 °C for 1 h. Samples were analyzed by LC-MS on an HPLC (model 2690, Waters) with an ion-trap mass spectrometer (LCQ, Thermo Finnigan). Each sample was separated on a Luna C18 column (3 micron, 4.6×75 mm, Phenomenex) maintained at 40 °C. The column was eluted at 0.8 ml/min with 90% A and 10% B (A as 0.1% formic acid in water and B as 0.1% formic acid in acetonitrile) to 20% B from 0 to 0.5 min, 20% B from 0.5 to 6 min, 20% B to 50% B from 6 to 8 min, 50% B to 95% B from 8 to 10 min and 95% B from 10 to 12 min. Kaempferol levels were determined by comparison with a standard curve prepared with authentic kaempferol (Indofine).

5. Novel materials

Novel materials described in this publication may be available for non-commercial research purposes upon

acceptance and signing of a material transfer agreement. In some cases such materials may contain or be derived from materials obtained from a third party. In such cases, distribution of material will be subject to the requisite permission from any third-party owners, licensors or controllers of all or parts of the material. Obtaining any permissions will be the sole responsibility of the requestor.

Acknowledgements

We would like to thank Laura Szymanski, Chris Hazel, Cheryl Caster and the DuPont soybean transformation group, Tony Guida for field and greenhouse efforts, Scott Tingey, Guo-Hua Miao, and Maureen Dolan for the DuPont EST program, and the CR&D DNA sequencing group. We thank Sharon Ellis, Mehran Moghaddam, Chris Vlahakis, Michaela Owens, and Jan Hazebroek for isoflavone analysis; and Wes Bruce and Ben Bowen for the CRC construct. We also thank Gary Fader, Enno Krebbers and Bill Hitz for many discussions and suggestions.

References

Akashi, T., Aoki, T., Ayabe, S., 1999. Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. Plant Physiol 121, 821–828.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Blount, J.W., Dixon, R.A., Paiva, N.L., 1992. Stress response in alfalfa (*Medicago sativa L.*). XVI. Antifungal activity of medicarpin and its biosynthetic precursors: implications for the genetic manipulation of stress metabolites. Physiol. Mol. Plant Pathol. 41, 333–349.

Bruce, W., Folkerts, O., Garnaat, C., Crasta, O., Roth, B., Bowen, B., 2000. Expression profiling of the maize flavonoid pathway genes controlled by estradiol-inducible transcription factors CRC and P. Plant Cell 12, 65–79.

Clarkson, T.B., 2002. Soy, soy phytoestrogens and cardiovascular disease. J. Nutr. 132, 566. S-569S.

Davis, S.R., Dalais, F.S., Simpson, E.R., Murkies, A.L., 1999. Phytoestrogens in health and disease. Recent Prog. Horm. Res. 54, 185–210.

De Majnik, J., Tanner, G.J., Joseph, R.G., Larkin, P.J., Weinman, J.J., Djordjevic, M.A., Rolfe, B.G., 1998. Transient expression of maize anthocyanin regulatory genes influences anthocyanin production in white clover and peas. Aust. J. Plant Physiol. 25, 335–343.

Dooner, H.K., Robbins, T.P., Jorgensen, R.A., 1991. Genetic and developmental control of anthocyanin biosynthesis. Annu. Rev. Genet. 25, 173–199.

Doyle, J.J., Schuler, M.A., Godette, W.D., Zenger, V., Beachy, R.N., Slightom, J.L., 1986. The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. J. Biol. Chem. 261, 9228–9238.

Garcia-Arenal, F., Fraile, A., Sagasta, E.M., 1978. The multiple phytoalexin response of bean (*Phaseolus vulgaris*) to infection by *Botrytis cinerea*. Physiol. Plant Path. 13, 151–156.

Gardner, R.O., 1975. Vanillin-hydrochloric acid as a histochemical test for tannin. Stain Technol. 50, 315–317.

Glassman, K.F., Gordon-Kamm, W.J., Kinney, A.J., Lowe, K., Stecca, K.L., Nichols, S.E., inventors. January 3, 2002. Recombinant constructs and their use in reducing gene expression. PCT patent publication No. WO0200904.

Graham, T.L., 1991. Flavonoid and isoflavonoid distribution in developing soybean seedling tissues and in seed and root exudates. Plant Physiol. 95, 594–603.

Graham, T.L., Graham, M.Y., 1991. Glyceollin elicitors induce major but distinct shifts in isoflavonoid metabolism in proximal and distal soybean cell populations. Mol. Plant–Microbe Interact. 4, 60–68.

Graham, T.L., 1995. Cellular biochemistry of phenylpropanoid responses of soybean to infection by *Phytophthora sojae*. In: Daniel, M., Purkayastha, R.P. (Eds.), Handbook of Phytoalexin Metabolism and Action. Marcel Dekker, New York, pp. 85–116.

Grotewold, E., Chamberlin, M., Snook, M., Siame, B., Butler, L., Swenson, J., Maddoc, S., Clair, G.S., Bowen, B., 1998. Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. Plant Cell 10, 721–740.

Hirner, A.A., Veit, S., Seitz, H.U., 2001. Regulation of anthocyanin biosynthesis in UV-A-irradiated cell cultures of carrot and in organs of intact carrot plants. Plant Sci. (Shannon, Irel.) 161, 315–322.

Jung, W., Yu, O., Lau, S.-M.C., O'Keefe, D.P., Odell, J., Fader, G., McGonigle, B., 2000. Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. Nat. Biotechnol. 18, 208-212.; erratum Jung, W., Yu, O., Lau, S.C., O'Keefe, D.P., Odell, J., Fader, G., McGonigle, B., 2000. Nat. Biotechnol. 18, 559.

Kinney, A.J., Fader, G.M., inventors. August 2, 2001. Suppression of specific classes of soybean seed protein genes. US issued patent No. US6362399.

Klein, T.M., Wolf, E.D., Wu, T., Sanford, J.C., 1987. High-velocity microprojectiles for delivery of nucleic acids to living cells. Nature (London, U.K.) 85, 70–73.

Latunde-Dada, A.O., Cabello-Hurtado, F., Czittrich, N., Didierjean, L., Schopfer, C., Hertkorn, N., Werck-Reichart, D., Ebel, J., 2001. Flavonoid 6-hydroxylase from soybean (*Glycine max* L.), a novel plant P-450 monooxygenase. J. Biol. Chem. 276, 1685–1688.

Liu, C.-J., Blount, J.W., Steele, C.L., Dixon, R.A., 2002. Bottlenecks for metabolic engineering of isoflavone glycoconjugates in *Arabidopsis*. Proc. Natl. Acad. Sci. U.S.A. 99, 14578–14583.

Lloyd, A.M., Walbot, V., Davis, R.W., 1992. *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators R and C1. Science (Washington, DC, U.S.) 258, 1773–1775.

Loake, G.J., Faktor, O., Lamb, C.J., Dixon, R.A., 1992. Combination of H-box [CCTACC(N)7CT] and G-box (CACGTG) cis elements is necessary for feed-forward stimulation of a chalcone synthase promoter by the phenylpropanoid-pathway intermediate *p*-coumaric acid. Proc. Natl. Acad. Sci. U.S.A. 89, 9230–9234.

McGonigle, B., Keeler, S.J., Lau, S.-M.C., Koeppe, M.K., O'Keefe, D.P., 2000. A genomics approach to the comprehensive analysis of the glutathione S-transferase gene family in soybean and maize. Plant Physiol. 124, 1105–1120.

Messina, M.J., 1999. Legumes and soybeans: overview of their nutritional profiles and health effects. Am. J. Clin. Nutr. 70, 439S–450S.

Nuutila, A.M., Kammiovirt, K., Oksman-Caldentey, K.M., 2002. Comparison of methods for the hydrolyis of flavonoids and phenolic aids from onion and spinach for HPLC analysis. Food Chem. 76, 519–525.

Pueppke, J.L., 1996. The genetic and biochemical basis for nodulation of legumes by rhizobia. Crit. Rev. Biotechnol. 16, 1–51.

Rivera-Vargas, L.I., Cshmitthenner, A.F., Graham, T.L., 1993. Soybean flavonoid effects on and metabolism by *Phytophthora sojae*. Phytochemistry 32, 851–857.

Slightom, J.L., Chee, P.P., inventors. September 19, 1991. Protein expression via seed specific regulatory sequences. PCT patent publication No. WO9113993.

Steele, C.L., Gijzen, M., Qutob, D., Dixon, R.A., 1999. Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. Arch. Biochem. Biophys. 367, 146–150.

Van Rhijn, R., Vanderleyden, J., 1995. The *Rhizobium*-plant symbiosis. Microbiological Rev. 59, 124–142.

Watanabe, S., Uesugi, S., Kikuchi, Y., 2002. Isoflavones for prevention of cancer, cardiovascular diseases, gynecological problems and possible immune potentiation. Biomed. Pharmacother. 56, 302–312.

Yu, O., Jung, W., Shi, J., Croes, R.A., Fader, G.M., McGonigle, B., Odell, J.T., 2000. Production of the isoflavones genistein and daidzein in non-legume dicot and monocot tissues. Plant Physiol. 124, 781–793.