The Plant Cell, Vol. 6, 265–275, February 1994 

# Light Regulation of Chlorophyll Biosynthesis at the Level of 5-Aminolevulinate Formation in Arabidopsis

**Lawrence L. Ilag,[1] A. Madan Kumar,[1] and Dieter Söll[2]**

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

**5-Aminolevulinic acid (ALA) is the universal precursor of tetrapyrroles, such as chlorophyll and heme. The major control of chlorophyll biosynthesis is at the step of ALA formation. In the chloroplasts of plants, as in *Escherichia coli*, ALA is derived from the glutamate of Glu-tRNA via the two-step $C_5$ pathway. The first enzyme, Glu-tRNA reductase, catalyzes the reduction of Glu-tRNA to glutamate 1-semialdehyde with the release of intact tRNA. The second enzyme, glutamate 1-semialdehyde 2,1-aminomutase, converts glutamate 1-semialdehyde to ALA. To further examine ALA formation in plants, we isolated Arabidopsis genes that encode the enzymes of the $C_5$ pathway via functional complementation of mutations in the corresponding genes of *E. coli*. The Glu-tRNA reductase gene was designated *HEMA* and the glutamate 1-semialdehyde 2,1-aminomutase gene, *GSA1*. Each gene contains two short introns (149 and 241 nucleotides for *HEMA*, 153 and 86 nucleotides for *GSA1*). The deduced amino acid sequence of the HEMA protein predicts a protein of 60 kD with substantial similarity (30 to 47% identity) to sequences derived from the known *hemA* genes from microorganisms that make ALA by the $C_5$ pathway. Purified Arabidopsis HEMA protein has Glu-tRNA reductase activity. The *GSA1* gene encodes a 50-kD protein whose deduced amino acid sequence shows extensive homology (55 to 78% identity) with glutamate 1-semialdehyde 2,1-aminomutase proteins from other species. RNA gel blot analyses indicated that transcripts for both genes are found in root, leaf, stem, and flower tissues and that their levels are dramatically elevated by light. Thus, light may regulate ALA, and hence chlorophyll formation, by exerting coordinated transcriptional control over both enzymes of the $C_5$ pathway.**

## INTRODUCTION

Tetrapyrrole compounds are essential components of life. They include hemes, which serve as prosthetic groups of respiratory enzymes, and chlorophylls, which are the major photosynthetic light-harvesting pigments (Senge, 1993). All carbon and nitrogen atoms of tetrapyrroles are derived from 5-aminolevulinic acid (ALA) whose formation provides the basic control point in the multistep pathway of tetrapyrrole biosynthesis, especially in the case of chlorophyll during dark/light transitions (Beale and Weinstein, 1990). Exogenous addition of ALA to dark-grown seedlings leads to the accumulation of protochlorophyllide (Sisler and Klein, 1963; Nadler and Granick, 1970; Castelfranco et al., 1974), suggesting that all of the enzymes required for the synthesis of protochlorophyllide from ALA are present in significant quantities in etiolated tissue and that ALA formation is the rate-limiting step in chlorophyll biosynthesis in dark-grown plants. When etiolated seedlings are first illuminated, the initial photoreduction of protochlorophyllide to chlorophyllide is not immediately followed by maximal chlorophyll synthesis. Instead, there is a lag phase during which the enzyme system responsible for ALA synthesis appears to be formed de novo. Indeed, after the lag phase, the rate of chlorophyll synthesis is enhanced and accompanied by a commensurate increase in level of ALA-forming activity (Castelfranco et al., 1974).

In plants, algae, and certain eubacteria, including *Escherichia coli*, ALA is formed via the $C_5$ pathway, as shown in Figure 1. The initial metabolite is the normal Glu-tRNA$^{Glu}$ that is converted by the action of a unique enzyme, Glu-tRNA reductase (GluTR), to glutamate 1-semialdehyde (GSA) with the concomitant release of tRNA$^{Glu}$. GSA is then converted to ALA by a specific transaminase, GSA-2,1-aminomutase (GSA-AM). Thus, GSA is the first committed precursor of porphyrin synthesis in organelles and organisms that use the $C_5$ pathway. In plants, tRNA$^{Glu}$ is chloroplast encoded, whereas the genes for GluTR and GSA-AM are believed to be nuclear encoded (reviewed in Jahn et al., 1992). Glu-tRNA has a dual function: it provides glutamate for protein biosynthesis and GSA for porphyrin synthesis (see Jahn et al., 1992).

GluTR is an unusual enzyme that catalyzes the NADPH-dependent reduction of Glu-tRNA$^{Glu}$ to GSA with the release of free tRNA$^{Glu}$. The enzyme recognizes the tRNA cofactor in a nucleotide sequence–specific fashion and does not discriminate on the basis of the activated glutamate. The

[1] Lawrence L. Ilag and A. Madan Kumar contributed equally to the work described in this study.

[2] To whom correspondence should be addressed at: Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, 266 Whitney Avenue, New Haven, CT 06520-8114.

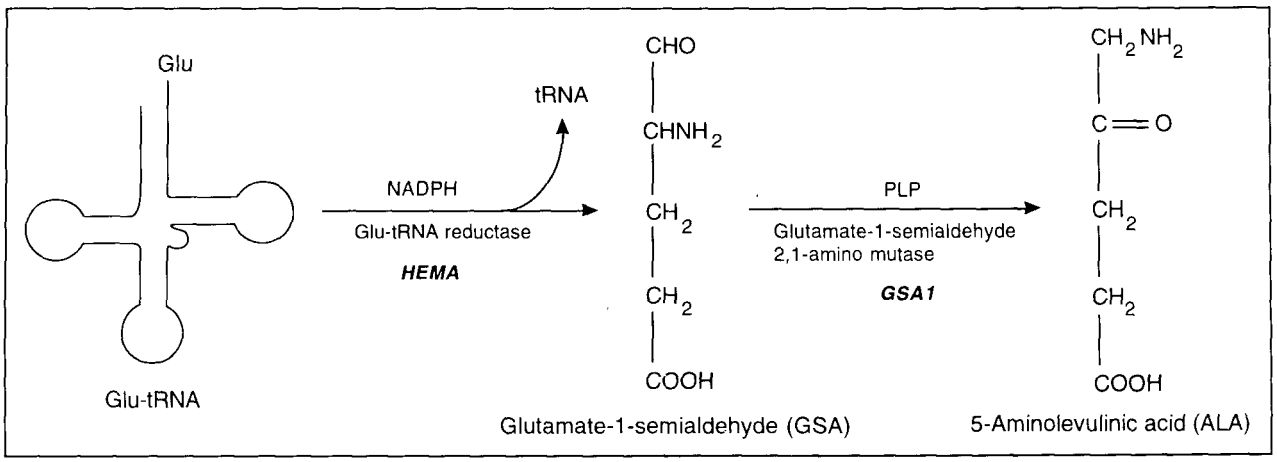


**Figure 1.** The $C_5$ Pathway of ALA Formation in Arabidopsis Chloroplasts.

Glu-tRNA reductase (encoded by the *HEMA* gene) reduces normal Glu-tRNA$^{Glu}$ to glutamate 1-semialdehyde with the release of intact tRNA$^{Glu}$. GSA is converted to ALA by glutamate 1-semialdehyde 2,1-aminomutase (encoded by the *GSA1* gene).

biochemical mechanism of this unprecedented reduction reaction is still insufficiently studied due to the scarcity of purified enzyme, which has been found only in low abundance in those organisms tested (reviewed in Jahn et al., 1992). In an attempt to delineate the mechanism of action, genes encoding GluTR have been isolated by complementation of the *E. coli hemA* mutation from *E. coli* (Drolet et al., 1989; Li et al., 1989; Verkamp and Chelm, 1989), *Salmonella typhimurium* (Elliott, 1989), *Bacillus subtilis* (Petricek et al., 1990), *Chlorobium vibrioforme* (Majumdar et al., 1991), and *Synechocystis* sp strain PCC 6803 (Grimm, 1992; Verkamp et al., 1992). However, corresponding genes from higher plants have not been reported.

In contrast, GSA-AM has been purified from many species and characterized in considerable detail. In barley (Grimm et al., 1989), pea (Nair et al., 1991), *Synechococcus* (Grimm et al., 1989), and *E. coli* (Ilag et al., 1991), GSA-AM is a homodimer consisting of 40- to 46-kD subunits, whereas in *Chlamydomonas*, a monomer of 43 kD has been found (Jahn et al., 1991). Like other aminotransferases, GSA-AM utilizes pyridoxal phosphate as a cofactor. Sequence information is available for GSA-AM genes from *E. coli* (Grimm et al., 1991), *B. subtilis* (Hansson et al., 1991), *S. typhimurium* (Elliott et al., 1990), *Synechococcus* (Grimm et al., 1991), barley (Grimm, 1990), and soybean (Sangwan and O'Brian, 1993). Comparison of the deduced amino acid sequences revealed that GSA-AM proteins are very similar to one another; e.g., the *Synechococcus* and barley enzymes are 70% identical. This degree of similarity is greater than that observed among the GluTR proteins.

We know less about the detailed mechanisms and the regulation of the enzymes in the $C_5$ pathway; however, our knowledge of the Shemin pathway of ALA synthesis is quite comprehensive. This route involves a one-step condensation of glycine and succinyl-coenzyme A by ALA synthase and operates in animals, yeast, and certain eubacteria (May et al., 1990). Because ALA formation is the key regulatory step in tetrapyrrole biosynthesis, we decided to explore the regulation of the pathway in higher plants using Arabidopsis as a model. In this study, we describe the isolation, characterization, and light response of genes that encode GluTR and GSA-AM, the enzymes involved in ALA formation.

## RESULTS

### Isolation and Characterization of Genes Encoding GluTR and GSA-AM

With the knowledge that in *E. coli* the *hemA* and *hemL* genes encode GluTR (Verkamp et al., 1992) and GSA-AM (Ilag et al., 1991), respectively, a likely route in isolating the homologous Arabidopsis genes was through functional complementation of the *E. coli hemA* and *hemL* mutations. Thus, *E. coli* strains containing the *hemA* (SASX41B; Sasarman et al., 1968) or *hemL* (GE1377; Ilag et al., 1991) mutation were transformed with DNA from cDNA libraries of Arabidopsis (ecotype Columbia) leaf mRNA. Transformants prototrophic for ALA were selected by plating the cells on minimal medium lacking ALA. Two clones, pAH1 from the *hemA* selection and pGA1 from the *hemL* selection, were chosen for further study. Corresponding genomic clones were isolated by plaque filter hybridizations using the cDNAs as probes and sequenced. Comparison of the complete nucleotide sequence of the cDNAs and the corresponding genomic DNAs revealed that the Arabidopsis *HEMA* contains a 1.6-kb coding sequence that is interrupted by two short introns (149 and 241 bp). Analogously, the *GSA1* gene of Arabidopsis also contains two short introns (153 and 86 bp) within the open reading frame of 1664 bp. The introns found in both of the genes carry the expected initial GT and final AG dinucleotides as boundaries (Breathnach and Chambon, 1981). Also, a potential

polyadenylation signal, AATAAG, is found 12 bases downstream of the UAA translation termination codon of the *HEMA* gene. A similar sequence, AAGAAA, is present 181 bases downstream of the UAG stop codon in the *GSA1* gene.

Genomic DNA gel blot analyses (data not shown) suggested that the *GSA1* gene exists as a single copy in the Arabidopsis genome. This conclusion was drawn from the observation that the *GSA1* probe hybridized to a single genomic fragment in each of four restriction enzyme digests. On the other hand, the *HEMA* probe hybridized very strongly to fragments of the expected size but also weakly to one or two larger fragments in each digest. This pattern was found to be reproducible. Thus, we cannot at present define the number of *HEMA* genes in the Arabidopsis genome. However, given the weak signal, the cross-hybridizing DNA fragments presumably have some sequence divergence. This also means that the strong signal detected on RNA gel blots with a *HEMA* cDNA probe is the corresponding mRNA (see below).

The transcription start site for *GSA1* has been mapped to 82 nucleotides upstream of the translation initiation codon ATG by primer extension and confirmed by S1 nuclease protection analysis (data not shown). There is a putative TATA box 33 nucleotides upstream of the transcription start site. Nucleotides −103 to −77 represent a region that is exclusively A+T, part of which shares significant similarity with AATATTTTTATT, the consensus sequence for the AT-1 box found in light-responsive promoters (Datta and Cashmore, 1989; Gilmartin et al., 1990).

The exact transcription start site for the *HEMA* gene could not be determined because primer extension experiments yielded multiple products. Nevertheless, a number of motifs similar to the consensus sequence (PyAPyPy) around the transcription start site (usually adenine) (Breathnach and Chambon, 1981) were found. The other characteristic motifs present in the *HEMA* gene are numbered with respect to the translation start site. A TATA box (−339), a CCAAT box (−440), and several light-responsive elements, such as 3AF1 (−341), and a GATA box (−287, −466, and −474) were found in the 5′ untranslated region of the *HEMA* gene. It is interesting to note that the 3AF1 box was juxtaposed to the TATA box as described previously (reviewed in Gilmartin et al., 1990).

### Comparison of the *HEMA*- and *GSA1*-Encoded Proteins with Corresponding Proteins

The amino acid sequence deduced from the open reading frame indicated that the *HEMA* gene product consists of 544 amino acids with a molecular mass of 60 kD, as shown in Figure 2, whereas the *GSA1* gene encodes a protein of 474 amino acids with a calculated molecular mass of ~50 kD, as shown in Figure 3. The deduced amino acid sequences of the encoded proteins are shown (Figures 2 and 3) aligned with the corresponding enzymes from other sources. The *GSA1* gene exhibits strong sequence conservation over the entire length. The lysine important for binding the cofactor pyridoxal phosphate (Grimm et al., 1992; Ilag and Jahn, 1992) and the surrounding amino acids are highly conserved in the seven GSA-AM protein sequences known to date (Figure 3). The degree of similarity is consistent with the species' phylogenetic relationships. The Arabidopsis enzyme is most similar to the soybean GSA-AM (79% identity); next in similarity is the barley homolog (78% identity) followed by the enzyme from the cyanobacterium *Synechococcus* (69% identity); least similar are the *B. subtilis*, *S. typhimurium*, and *E. coli* proteins (55 to 58% identity). In the case of the *HEMA* gene product, sequence conservation is much lower. The *HEMA*-derived Arabidopsis protein sequence shows considerable similarity (e.g., 32% identity with the *E. coli* protein) including a stretch of 23 amino acids with 52% identity with *hemA*-encoded proteins from other sources. This region contains the Cys→Tyr change that is the cause of the *B. subtilis* HemA phenotype (Petricek et al., 1990).

Because GluTR and GSA-AM proteins are chloroplast-targeted enzymes in photosynthetic eukaryotes, the Arabidopsis enzymes are expected to have a transit peptide at the amino terminus that is absent in the corresponding prokaryotic proteins. Examination of the Arabidopsis sequences (Figures 2 and 3) shows this to be the case. The transit peptide cleavage site has been empirically determined for the barley GSA-AM (Grimm, 1990), and the corresponding site in Arabidopsis GSA-AM is predicted to be after the 37th amino acid based on the alignment in Figure 3. Interestingly, the barley and Arabidopsis GSA-AM proteins are least similar in their transit peptides in both amino acid sequence and composition. This disparity in transit peptide sequence is not totally unexpected because transit peptides are in general highly variable in sequence and are thought to fold into a common three-dimensional motif that is recognized by the chloroplast import machinery (von Heijne et al., 1989). In the alignment in Figure 3, the transit peptides are 25% identical. The barley transit peptide has a preponderance (62%) of nonpolar amino acids (A, L, I, V, P, F, W, and M), whereas the Arabidopsis transit peptide contains many (54%) polar amino acids (S, T, Y, Q, C, and G). The Arabidopsis transit peptide is similar to most other transit peptides when the fraction of serines and arginines is considered (von Heijne et al., 1989). Another difference between the two transit peptides is their cleavage sites. The cleavage site for the barley enzyme perfectly fits the consensus sequence, (V/I)X(A/C)↓A (von Heijne et al., 1989), whereas in *GSA1*, the proposed cleavage site or any surrounding sequence lacks resemblance to the consensus.

The *HEMA*-deduced amino acid sequence also displays a long amino-terminal extension with a characteristic amino acid composition of chloroplast transit peptide sequence. This region shows two amino acid sequences (NAAS) and (NSAA) that are similar to the signal peptidase recognition sequence. Although GluTR binds tRNA in a sequence-specific fashion, a property shared with aminoacyl-tRNA synthetases, no similarities in amino acid sequence were found. Probably this reflects the fact that there are no distinguishing characteristics of a tRNA binding site (unlike that for ATP) in the synthetase family. However, GluTR enzymes display a characteristic fingerprint

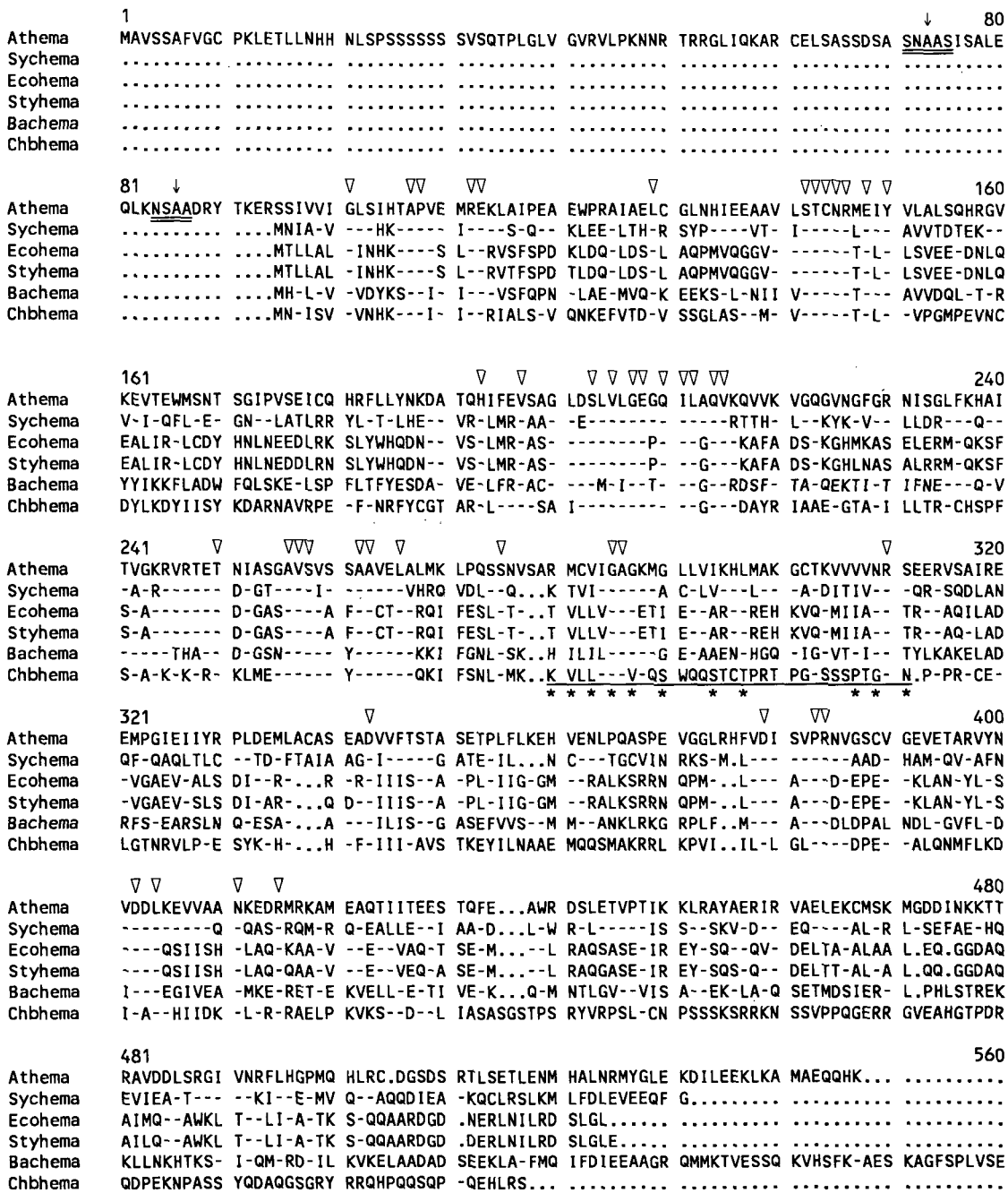

**Figure 2.** Sequence Alignment of Various GluTR Proteins.

The predicted amino acid sequence of the Arabidopsis *HEMA* protein (Athema) was aligned with the predicted amino acid sequences of the GluTR proteins from *E. coli* (Ecohema; Verkamp and Chelm, 1989), *B. subtilis* (Bachema; Petricek et al., 1990), *Synechocystis* sp strain PCC 6803 (Sychema; Verkamp et al., 1992), *S. typhimurium* (Styhema; Elliott, 1989), and *C. vibrioforme* (Chbhema; Majumdar et al., 1991) using the Pileup program in the sequence analysis software of the Wisconsin Genetics Computer Group (Devereux et al., 1984). The complete amino acid sequence of Arabidopsis HEMA is presented; the dots in the amino acid sequences were introduced for optimal alignment. A dash indicates an amino acid identical to Arabidopsis, and open triangles above the Athema sequence indicate positions of identical amino acids in all six sequences. Double-underlined amino acid sequences are the predicted signal peptidase recognition motifs, and arrows indicate the predicted cleavage site that will yield mature protein. The underlined amino acids are in the region that can potentially fold to form βαβ structure for ADP binding. Asterisks indicate the positions of amino acids found in the consensus fingerprint sequence of ADP binding proteins (Wierenga et al., 1986).

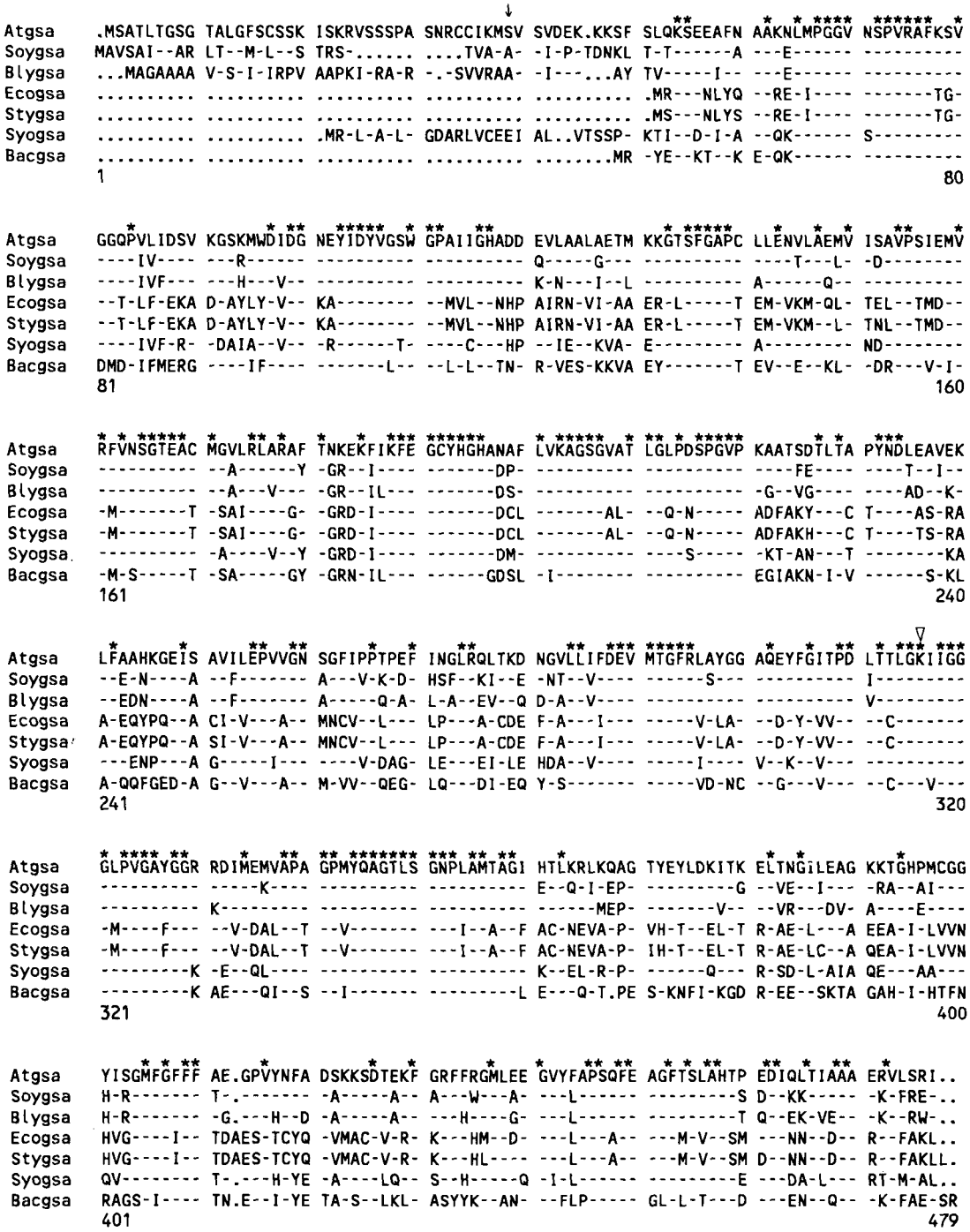


**Figure 3.** Sequence Alignment of Various GSA-AM Proteins.

The predicted amino acid sequence of the Arabidopsis GSA1 protein (Atgsa) was aligned with the predicted amino acid sequences of the GSA-AM proteins from soybean (Soygsa; Sangwan and O'Brian, 1993), barley (Blygsa; Grimm, 1990), *E. coli* (Ecogsa; Grimm et al., 1991), *S. typhimurium* (Stygsa; Elliott et al., 1990), *Synechococcus* (Syogsa; Grimm et al., 1991), and *B. subtilis* (Bacgsa; Hansson et al., 1991) using the Pileup program in the sequence analysis software of the Wisconsin Genetics Computer Group (Devereux et al., 1984). A dash represents an amino acid identical to that in the Arabidopsis enzyme at that particular position, whereas an asterisk indicates a position with identical amino acid residues in all seven sequences, and a dot within a sequence represents a gap that was introduced to maximize homology. The lysine residue implicated in binding pyridoxal phosphate (Grimm et al., 1992; Ilag and Jahn, 1992) is indicated by an open triangle; the arrow indicates the start of the mature protein in barley (Grimm, 1990).

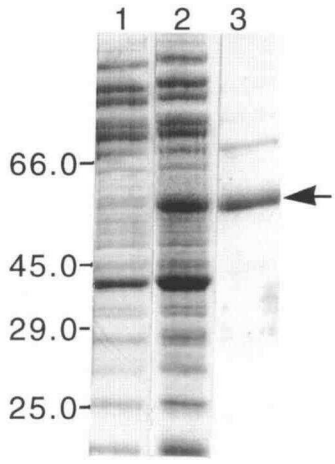


**Figure 4.** Purification of Arabidopsis GluTR.

The entire *HEMA* cDNA sequence was cloned in pET15B for T7 RNA polymerase overexpression in *E. coli* as a histidine-tagged protein. Cells were transformed with the construct and induced with isopropyl-β-D-thiogalactopyranoside (IPTG). Approximately 5 μg of protein from the extract was analyzed on a 10% SDS–polyacrylamide gel by electrophoresis. Lane 1 contains cell extract before IPTG induction; lane 2, cell extract after induction with IPTG; lane 3, eluate from an Ni-NTA agarose column. The lower molecular mass (given in kilodaltons) results from the cloning strategy that truncated part of the leader sequence. Arrow indicates the 53-kD protein.

structure (see Figure 2) found in ADP binding proteins (Wierenga et al., 1986). This structure is composed of a set of 11 amino acids that folds to form a βαβ fold for the binding of ADP moiety.

### The Arabidopsis *HEMA* Gene Encodes a Functional GluTR

Biochemical studies suggested that GluTR is the product of the *E. coli hemA* gene (Verkamp et al., 1992). The ability of the Arabidopsis *HEMA* gene to bestow viability on an *E. coli hemA* strain and the significant homology between the bacterial and Arabidopsis genes suggested that the cloned *HEMA* gene encodes a functional GluTR enzyme. To demonstrate that *HEMA* encodes GluTR, we constructed a histidine-tagged *HEMA* cDNA gene in a T7 RNA polymerase vector (Studier et al., 1990) and expressed it in *E. coli*. A 53-kD protein was very abundant only in the S-100 extract of the transformed strain and could be purified significantly by an Ni-NTA affinity column, as shown in Figure 4. When GluTR activity was tested under standard assay conditions with *E. coli* Glu-tRNA, the extract of the transformed cells had a specific activity of 3.4 pmol of ALA per mg (versus 0.18 for the untransformed strain), while the purified protein had a specific activity of 12 pmol/mg. Thus, the Arabidopsis *HEMA* gene encodes a functional GluTR that recognized *E. coli* $tRNA^{Glu}$.

### Expression Patterns of *HEMA* and *GSA1* mRNA

The tissue distribution of *HEMA* and *GSA1* mRNA was examined through RNA gel blot analysis. As shown in Figure 5, the transcript of *GSA1* (~1.8 kb) is present in all tissues examined (root, stem, leaf, and flower) in approximately the same abundance, whereas the expression of *HEMA* (with a transcript length of 1.85 kb) was found to be slightly more in leaves.

The effect of light on the steady state level of *HEMA* and *GSA1* mRNAs was measured in etiolated seedlings. In response to light, the *GSA1* mRNA level initially dipped and then increased, such that seedlings exposed to 24 hr of light exhibited a *GSA1* mRNA level noticeably greater than that exhibited by dark-grown seedlings, as shown in Figure 6. In contrast, *HEMA* mRNA remained at undetectable levels throughout the course of the experiment (data not shown).

Induction of *HEMA* and *GSA1* mRNA levels by light was also examined in mature Arabidopsis plants. The various treatments employed prior to RNA extraction are summarized in Figure 7, and the *HEMA* and *GSA1* mRNA levels in the plants after the various treatments are shown in Figures 8A and 8B, respectively. *HEMA* and *GSA1* transcripts were present in plants that were grown in light. Exposure of plants to dark resulted in the complete disappearance of the *HEMA* transcript, whereas the *GSA1* transcript was expressed at a considerably lower level than in the original plants. Transcripts of both genes reappeared when dark-grown plants were transferred to light. Reexposure of plants to the dark resulted in complete inhibition of *HEMA* transcription, whereas the transcript of *GSA1* was significantly diminished. On the other hand, the mRNA levels of the *HEMA*

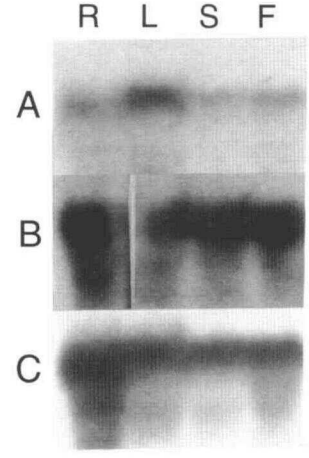


**Figure 5.** Expression of *HEMA* and *GSA1* Genes in Various Tissues of Arabidopsis as Detected by RNA Gel Blots.

Total RNA (20 μg per lane) isolated from different parts (R, roots; L, leaf; S, stem; F, flower) of 5-week-old Arabidopsis plants was blotted and hybridized with $^{32}P$-labeled probes.

**(A)** *HEMA* gene (600-bp NheI fragment of pAH1).

**(B)** *GSA1* gene (1.6-kb ApaI-KpnI fragment of pGA1).

**(C)** Arabidopsis 18S rDNA (1-kb EcoRI-BamHI fragment of pPD-55).

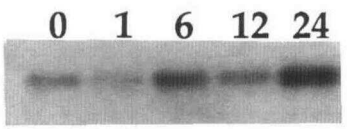


**Figure 6.** Light Induction of *GSA1* mRNA Expression in Etiolated Seedlings.

Poly(A)$^+$ RNA (2 μg per lane) from whole Arabidopsis etiolated seedlings exposed to dim white light (0.7 μE $m^{-2}$ $sec^{-1}$) for the indicated number of hours was hybridized with a $^{32}$P-labeled 1.6-kb ApaI-KpnI fragment of pGA1. Equal loading of RNA in each gel lane was made possible by spectrophotometric determination of RNA concentration in each sample.

and *GSA1* genes in plants that remained exposed to light throughout the experiment did not exhibit any dramatic change. This observation implied that light induces the *HEMA* and *GSA1* gene expression in Arabidopsis.

## DISCUSSION

### *HEMA* and *GSA1* Encode Functional GluTR and GSA-AM

Heterologous complementation in *E. coli* has been exploited previously in cloning genes from plants (e.g., Snustad et al., 1988; Kumar and Söll, 1992; Senecoff and Meagher, 1993). Aside from being a powerful strategy for gene isolation, complementation simultaneously provides evidence of authenticity and functionality for the isolated gene. Hence, it can be concluded that due to the ability of *HEMA* and *GSA1* to complement *E. coli hemA* and *hemL* mutations, respectively, and the presence of homologies at the amino acid level with corresponding genes from other sources, the Arabidopsis *HEMA* and *GSA1* genes encode GluTR and GSA-AM, respectively, in Arabidopsis. GluTR is a unique enzyme that catalyzes a reduction requiring a specific tRNA as a "cofactor." Earlier studies have shown that GluTR preferentially recognizes the homologous tRNA (reviewed in Jahn et al., 1992). Nevertheless, the functional complementation in *E. coli* by the Arabidopsis *HEMA* gene implies that Arabidopsis GluTR does interact with *E. coli* $tRNA^{Glu}$, a result also borne out by our in vitro studies. The less than perfect interaction of the Arabidopsis GluTR with *E. coli* $tRNA^{Glu}$ is also evident from the reduced growth rate (by approximately one half) of the *E. coli hemA* strain transformed with the *HEMA* plasmid. In contrast, the growth rate of the *GSA1*-transformed *E. coli hemL* strain is indistinguishable from the wild type. Obviously, these growth characteristics reflect the much closer sequence homology of the Arabidopsis and *E. coli* GSA-AM enzymes compared to that of the GluTR proteins (Figures 2 and 3).

### Light Induction of *HEMA* and *GSA1*

RNA gel blot analyses indicated that *GSA1* mRNA accumulation is induced by light in etiolated and dark-adapted plants. In etiolated seedlings, initial light exposure caused a decline in the *GSA1* mRNA level; this level then steadily increased upon prolonged exposure to light to reach a level that modestly exceeded the *GSA1* mRNA level in the original etiolated seedlings. This small increase is consistent with that observed in soybean (Sangwan and O'Brian, 1993). In dark-adapted plants,

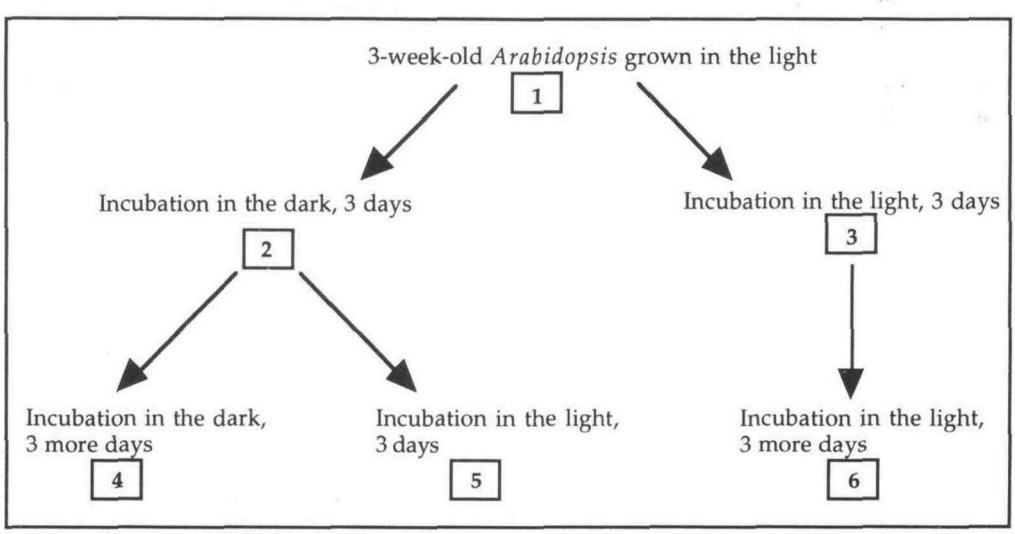


**Figure 7.** Induction of *HEMA* and *GSA1* mRNA Levels by Light.

This schematic illustrates the regimes of dark and light transitions conducted to study *HEMA* and *GSA1* gene expression by RNA gel blot analysis.

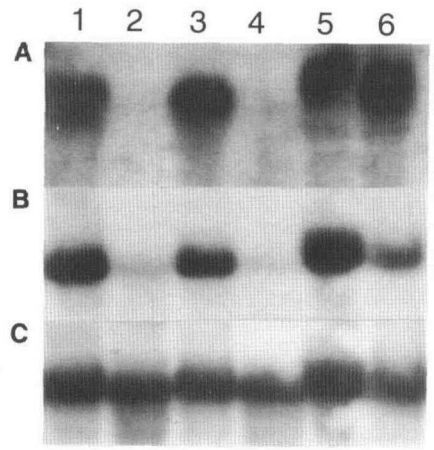


**Figure 8.** RNA Gel Blot Analysis of *HEMA* and *GSA1* mRNAs.

The mRNAs for this analysis were isolated from plants exposed to different light regimes (see Figure 7). Total leaf RNA (50 μg per lane) was used.

**(A)** *HEMA* gene (600-bp NheI fragment of pAH1).

**(B)** *GSA1* gene (1.6-kb ApaI-KpnI fragment of pGA1).

**(C)** Arabidopsis 18S rDNA (1-kb EcoRI-BamHI fragment of pPD-55).

light induction of *GSA1* mRNA accumulation is much more pronounced. The initial decline of *GSA1* mRNA level in etiolated seedlings upon exposure to light was reproducible. Further experiments are required to elucidate the molecular basis of this decline.

In etiolated seedlings, increase in *GSA1* mRNA concentration did not completely coincide with induction of chlorophyll synthesis by light (data not shown), which was suggestive of regulatory mechanisms of ALA synthesis other than at the level of *GSA1* mRNA synthesis. Under the light conditions used, chlorophyll was synthesized at steadily increasing rates by greening etiolated seedlings after a lag phase of ~4 hr during the 24-hr time course. During the same time period, the *GSA1* mRNA level did not increase as steadily. *GSA1* mRNA levels in greening seedlings after 6 and 12 hr of illumination were similar (Figure 6), suggesting that accumulation of *GSA1* mRNA was not rate limiting for chlorophyll synthesis at this time. Because the rate of chlorophyll synthesis increased during this time period, it is safe to assume that GSA-AM activity rose as well. This could have been accomplished even if the *GSA1* mRNA level did not increase through other mechanisms that range from increased translation of the *GSA1* transcript to increased specific activity of GSA-AM. The increase in *GSA1* mRNA level in greening seedlings (two- to threefold) also appeared modest; this would not account for the observed increase in chlorophyll content (~72-fold), a further indication that chlorophyll (and ALA) synthesis is regulated through other mechanisms as well.

White light was used in our experiments to assay for light inducibility of *HEMA* and *GSA1*. Determination of the particular wavelength that stimulates expression of these genes should permit us to infer the identity of the photoreceptor involved.

### Are *HEMA* and *GSA1* Transcriptionally Coregulated?

In this study, we demonstrated that the expression of *HEMA* and *GSA1* transcripts in mature plants is strongly stimulated by light. This suggested that light coordinately regulates *HEMA* and *GSA1* expression. Whether the molecular basis of the light-induced stimulation of expression—a light-activatable promoter or increased transcript stability in the light or both—is the same for *HEMA* and *GSA1* remains to be demonstrated. Although several motifs similar to light-responsive elements are found in the 5′ untranslated region of *HEMA*, their contribution in the observed effect of light on *HEMA* expression is not known. It can also be seen that a particular region in the *HEMA* and *GSA1* proximal promoter sequences exhibited considerable (71%) identity, as shown in Figure 9. Albeit with no striking similarity to previously identified light-responsive promoter elements (Gilmartin et al., 1990), this may potentially serve as a *trans*-acting factor binding site to coordinate light regulation of *HEMA* and *GSA1*. In transient expression experiments in Arabidopsis mesophyll protoplasts, deletion of a fragment of the *GSA1* promoter containing this region results in reduced promoter activity (Ilag, 1993).

Interestingly, light does not have as dramatic an effect on *HEMA* and *GSA1* transcript accumulation in young, etiolated seedlings. At the peak of light-induced chlorophyll synthesis during greening, the *GSA1* mRNA level was only moderately augmented, whereas *HEMA* mRNA remained at extremely low levels (A.M. Kumar and D. Söll, unpublished data). Post-transcriptional regulatory mechanisms are therefore likely to be involved in large part to effect maximal ALA synthesis during greening (Castelfranco et al., 1974).

### Light Regulation of Other Chlorophyll Biosynthetic Genes

Our studies have shown that both enzymes of the $C_5$ pathway are regulated by light. Because the synthesis of


**Figure 9.** Comparison of *HEMA* and *GSA1* Promoter Sequences.

The −358 to −335 region (relative to the translation start site) of the *HEMA* promoter is aligned with the −221 to −198 region (relative to the transcription start site) of the *GSA1* promoter. A vertical bar represents identity between the two sequences at that particular position.

protochlorophyllide from exogenous ALA occurs even in the dark, the enzymes involved in these steps in chlorophyll biosynthesis appear to be constitutively expressed in the dark and in the light. Another known enzyme of chlorophyll biosynthesis whose mRNA level is light regulated is NADPH-protochlorophyllide oxidoreductase, which catalyzes the reduction of protochlorophyllide to chlorophyllide. However, light has an opposite effect on this enzyme. In Arabidopsis and all other species studied to date, protochlorophyllide reductase protein and activity levels are drastically reduced by illumination (Forreiter et al., 1991). Illumination causes a decline in the corresponding mRNA level, although the extent varies from species to species (Forreiter et al., 1991); in Arabidopsis, the decline is only slight (Benli et al., 1991). Our understanding of this step is still incomplete, because there might be a second protochlorophyllide reductase involved (G. Armstrong and K. Apel, personal communication).

## Overall Regulation of ALA Synthesis

Based on many physiological studies, it is evident that the regulation of ALA formation is complex. Besides transcriptional and translational regulatory mechanisms, feedback inhibition by metabolites of the chlorophyll pathway and possibly regulation of mRNA and protein stability of GluTR or GSA-AM may operate in the formation of ALA (reviewed in Beale and Weinstein, 1990; Jahn et al., 1992). Heme inhibition of ALA synthesis has been observed in vivo (reviewed in Beale and Weinstein, 1990) and in vitro. Biochemical studies with purified Arabidopsis GluTR will need to be conducted to address this question. However, data from *Chlamydomonas* mutants suggest that heme also regulates the expression of the enzymes catalyzing ALA formation (Huang and Wang, 1986). Thus, heme may have an effect on *hemA* transcription. Because heme controls the availability of free iron, which may act via iron-responsive elements, we searched for such elements. However, the 5′ untranslated region of *HEMA* did not contain CAGUGN (N=C,U,A), an iron-responsive element consensus sequence that operates in ALA synthase (May et al., 1990). In addition, as seen with the light response, different mechanisms may operate in different species. The availability of cloned genes and antibodies will greatly aid the elucidation of the molecular basis underlying the effects of various environmental stimuli (e.g., hormones, heme, and light) on ALA synthesis.

The availability and the subsequent expression of the Arabidopsis *HEMA* and *GSA1* genes will permit the establishment of an in vitro Glu→ALA conversion assay with purified enzymes (Chen et al., 1990; Lloyd et al., 1993). This might prove useful in screening for inhibitors of ALA formation; these inhibitors may have potential as herbicides. Because mammals do not utilize the $C_5$ pathway for ALA biosynthesis, such compounds may not be toxic to them.

## METHODS

### Plant Material and Growth Conditions

*Arabidopsis thaliana* plants (Columbia ecotype) were grown under standard conditions (22°C, 60% relative humidity, and with a regime of 16-hr white light [90 μE $m^{-2}$ $sec^{-1}$] and 8-hr dark cycle). In cases when Arabidopsis plants were to be aseptically germinated, germination medium contained 1 × Murashige and Skoog basal salt mixture (Sigma), 10 g sucrose, 0.1 g myo-inositol, 2.14 mL 1 M K-Mes (pH 5.7), and 1 mL of 1000 × vitamin stock solution per liter. The 1000 × vitamin stock solution contained vitamin B1 at 20 mg/mL, and nicotinic acid and pyridoxine hydrochloride at 1 mg/mL. Medium was autoclaved at 121°C for 15 min. Vitamin stock solution (1000 ×) was added after the medium had cooled to 50°C.

To determine the effect of light on mRNA accumulation in etiolated seedlings, Arabidopsis was germinated and grown aseptically in liquid germination medium in complete darkness at 22°C for 7 days. The seedlings were then exposed to dim white light (0.7 μE $m^{-2}$ $sec^{-1}$) for 1, 6, 12, and 24 hr, and RNA was extracted from plant tissue at each of these time points. For the zero time point, tissue was harvested and frozen in liquid nitrogen under a dim green safelight to minimize any effects of illumination on gene expression.

To determine the effect of light on mRNA accumulation in dark-adapted plants, 3-week-old Arabidopsis plants grown on a 16-hr white light (90 μE $m^{-2}$ $sec^{-1}$) and 8-hr darkness cycle were put in complete darkness for 3 days and then returned to light (90 μE $m^{-2}$ $sec^{-1}$) for 3 more days. RNA was extracted from leaves before and after the complete darkness treatment and after the 3-day light treatment. RNA was also prepared from plants put in complete darkness for 6 days and plants that remained in the light throughout the duration of the experiment.

### *Escherichia coli* Strains and Growth Media

*E. coli* SASX41B (Sasarman et al., 1968) is a 5-aminolevulinic acid (ALA) auxotroph carrying a *hemA* mutation. *E. coli* K12 strain GE1377 (Ilag et al., 1991) harbors the *hemL207* mutant allele. LI102 (Ilag and Jahn, 1992) contains a tetracycline resistance gene disrupting the *hemL* gene. DH5α was used for routine DNA subcloning, XL-1 Blue was used for phagemid rescue, and LE392 and NM621 were the host strains used for plating Arabidopsis genomic libraries.

*E. coli* was routinely grown in Luria–Bertani medium. Minimal medium with glycerol was used for the complementation experiments. Ampicillin, tetracycline, and ALA, when used, were at final concentrations of 50 to 200 μg/mL, 15 μg/mL, and 50 μg/mL, respectively.

### Nucleic Acid Methods

DNA and RNA analyses were standard (Ausubel et al., 1987). RNA was extracted from Arabidopsis as described previously (Chomczynski and Sacchi, 1987) with minor modifications. Tissue was ground into a fine powder in liquid nitrogen using a mortar and pestle and then added to denaturing solution. The slurry was further homogenized on ice using a Brinkmann homogenizer for 45 sec before proceeding. Poly(A)$^+$ RNA was isolated through chromatography on oligo(dT) cellulose columns.

RNA gel blot analysis was performed using standard procedures.

Equal loading of RNA samples was ensured by spectrophotometric determination of RNA concentration in samples, ethidium bromide staining of the gel, or hybridization of the blot to an 18S rDNA probe. Intensity of signals in RNA blot autoradiograms was quantitated through densitometric scanning.

The 5′ ends of the mRNAs were mapped by primer extension with SUPERSCRIPT RNase H⁻ reverse transcriptase (400 units; GIBCO BRL).

### Isolation and Sequence Analysis of Genomic Clones Encoding Glutamyl tRNA Reductase or Glutamate 1-Semialdehyde 2,1-Aminotransferase

To isolate *HEMA* genomic clones, an Arabidopsis genomic library in λGEM11 (kindly provided by J.T. Mulligan and R. Davis, Stanford University, Stanford, CA) was screened with a 600-bp NheI fragment of pAH1. One genomic clone, λGH1, contained a 2.5-kb HpaI-HpaII fragment that included the entire gene. However, another clone was needed to acquire more sequence information toward the 5′ flanking region of the gene. Sequencing was by the dideoxy method. The 2879-nucleotide-long sequence has been submitted to GenBank as accession number U03774.

The same procedures were followed starting from a library of non-size-selected fragments from a partial Sau3A digest of Arabidopsis (Columbia wild-type) DNA in λGEM11 (kindly provided by Carrie Schneider and Chris Somerville, Michigan State University, East Lansing). The 3204-nucleotide-long sequence was submitted to GenBank as accession number U03773.

### Overexpression of HEMA

As described in the legend to Figure 4, the *HEMA* cDNA sequence was inserted into the pET15 vector, transformed into BL21, and expressed (Studier et al., 1990). Assays for Glu-tRNA reductase (GluTR) activity were modified from the method of Chen et al. (1990) as described by Lloyd et al. (1993) by using precharged *E. coli* $^{14}$C-Glu-tRNA as substrate and converting the produced GSA to ALA by the action of added *E. coli* GSA 2,1-aminomutase (GSA-AM).

## ACKNOWLEDGEMENTS

We would like to thank Adrian Lloyd for help with the GluTR assays and Alison DeLong and Elizabeth Verkamp for many discussions, Gary O'Neill for deoxy-oligonucleotides, and Ron Davis, Xing-Wang Deng, John Mulligan, Richard Meagher, Chris Somerville, and Michael Thomashow for gifts of libraries and clones. This work was supported by a grant from the U.S. Department of Energy.

Received November 8, 1993; accepted November 23, 1993.

## REFERENCES

**Ausubel, F.M., Brent, R., Kingston, R.E., Moose, D.D., Smith, J.A., Seidman, J.G., and Struhl, K.,** eds (1987). Current Protocols in Molecular Biology. (New York: Green Publishing Associates and Wiley-Interscience).

**Beale, S.I., and Weinstein, J.D.** (1990). Tetrapyrrole metabolism in photosynthetic organisms. In Biosynthesis of Heme and Chlorophylls, H.A. Dailey, ed (New York: McGraw-Hill), pp. 287–391.

**Benli, M., Schulz, R., and Apel, K.** (1991). Effect of light on the NADPH-protochlorophyllide oxidoreductase of *Arabidopsis thaliana*. Plant Mol. Biol. **16,** 615–625.

**Breathnach, R., and Chambon, P.** (1981). Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50,** 349–383.

**Castelfranco, P.A., Rich, P.M., and Beale, S.I.** (1974). The abolition of lag phase in greening cucumber cotyledons by exogenous δ-aminolevulinic acid. Plant Physiol. **53,** 615–618.

**Chen, M.-W., Jahn, D., O'Neill, G.P., and Söll, D.** (1990). Purification of the glutamyl-tRNA reductase from *Chlamydomonas reinhardtii* involved in δ-aminolevulinic acid formation during chlorophyll biosynthesis. J. Biol. Chem. **265,** 4058–4063.

**Chomczynski, P., and Sacchi, N.** (1987). Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162,** 156–159.

**Datta, N., and Cashmore, A.R.** (1989). Binding of a pea nuclear protein to promoters of certain photoregulated genes is modulated by phosphorylation. Plant Cell **1,** 1069–1077.

**Devereux, J., Haeberli, P., and Smithies, O.** (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12,** 387–395.

**Drolet, M., Peloquin, L., Echelard, Y., Cousineau, L., and Sasarman, A.** (1989). Isolation and nucleotide sequence of the *hemA* gene of *Escherichia coli* K12. Mol. Gen. Genet. **216,** 347–352.

**Elliott, T.** (1989). Cloning, gene characterization, and nucleotide sequence of the *hemA-prfA* operon of *Salmonella typhimurium*. J. Bacteriol. **171,** 3948–3960.

**Elliott, T., Avissar, Y.J., Rhie, G.-E., and Beale, S.I.** (1990). Cloning and sequence of the *Salmonella typhimurium hemL* gene and identification of the missing enzyme in *hemL* mutants as glutamate-1-semialdehyde aminotransferase. J. Bacteriol. **172,** 7071–7084.

**Forreiter, C., van Cleve, B., Schmidt, A., and Apel, K.** (1991). Evidence for a general mechanism of light-dependent control of the NADPH-protochlorophyllide oxidoreductase in higher plants. Planta **183,** 126–132.

**Gilmartin, P.M., Sarokin, L., Memelink, J., and Chua, N.-H.** (1990). Molecular light switches for plant genes. Plant Cell **2,** 369–378.

**Grimm, B.** (1990). Primary structure of a key enzyme in plant tetrapyrrole synthesis: Glutamate-1-semialdehyde aminotransferase. Proc. Natl. Acad. Sci. USA **87,** 4169–4173.

**Grimm, B.** (1992). Identification of a *hemA* gene from *Synechocystis* by complementation of an *E. coli hemA* mutant. Hereditas **117,** 195–197.

**Grimm, B., Bull, A., Welinder, K.G., Gough, S.P., and Kannangara, C.G.** (1989). Purification and partial amino acid sequence of the glutamate 1-semialdehyde aminotransferase of barley and *Synechococcus*. Carlsberg Res. Commun. **54,** 67–79.

**Grimm, B., Bull, A., and Breu, V.** (1991). Structural genes of glutamate-1-semialdehyde aminotransferase for porphyrin synthesis in a cyanobacterium and *Escherichia coli*. Mol. Gen. Genet. **225,** 1–10.

**Grimm, B., Smith, M.A., and von Wettstein, D.** (1992). The role of Lys272 in the pyridoxal 5-phosphate active site of *Synechococcus*

glutamate-1-semialdehyde aminotransferase. Eur. J. Biochem. **206,** 579–585.

**Hansson, M., Rutberg, L., Schröder, I., and Hederstedt, L.** (1991). The *Bacillus subtilis hemAXCDBL* gene cluster, which encodes enzymes of the biosynthetic pathway from glutamate to uroporphyrinogen. J. Bacteriol. **173,** 2590–2599.

**Huang, D.D., and Wang, W.Y.** (1986). Genetic control of chlorophyll biosynthesis: Regulation of δ-amino levulinate synthesis in *Chlamydomonas.* Mol. Gen. Genet. **205,** 217–220.

**Ilag, L.L.** (1993). Genes in *Escherichia coli* and *Arabidopsis thaliana* That Encode Glutamate-1-Semialdehyde 2,1-Aminomutase, an Enzyme in Tetrapyrrole Biosynthesis. PhD Dissertation (New Haven, CT: Yale University).

**Ilag, L.L., and Jahn, D.** (1992). Activity and spectroscopic properties of the *Escherichia coli* glutamate-1-semialdehyde aminotransferase and the putative active site mutant K265R. Biochemistry **31,** 7143–7151.

**Ilag, L.L., Jahn, D., Eggertsson, G., and Söll, D.** (1991). The *Escherichia coli hemL* gene encodes glutamate 1-semialdehyde aminotransferase. J. Bacteriol. **173,** 3408–3413.

**Jahn, D., Chen, M.W., and Söll, D.** (1991). Purification and functional characterization of glutamate-1-semialdehyde aminotransferase from *Chlamydomonas reinhardtii.* J. Biol. Chem. **266,** 161–167.

**Jahn, D., Verkamp, E., and Söll, D.** (1992). Glutamyl-transfer RNA: A precursor of heme and chlorophyll biosynthesis. Trends Biochem. Sci. **17,** 215–218.

**Kumar, A.M., and Söll, D.** (1992). *Arabidopsis* alternative oxidase sustains *Escherichia coli* respiration. Proc. Natl. Acad. Sci. USA **89,** 10842–10846.

**Li, J.M., Russell, C.S., and Cosloy, S.D.** (1989). Cloning and structure of the *hemA* gene of *Escherichia coli* K12. Gene **82,** 209–217.

**Lloyd, A.J., Weitzman, P.D., and Söll, D.** (1993). Incomplete citric acid cycle favors $C_5$ pathway of 5-aminolevulinic acid synthesis in a methylotroph. J. Gen. Microbiol. **139,** 2931–2938.

**Majumdar, D., Avissar, Y.J., Wyche, J.H., and Beale, S.I.** (1991). Structure and expression of the *Chlorobium vibrioforme hemA* gene. Arch. Microbiol. **156,** 281–289.

**May, B.K., Bhasker, C.R., Bawden, M.J., and Cox, T.C.** (1990). Molecular regulation of 5-aminolevulinate synthetase. Mol. Biol. Med. **7,** 405–421.

**Nadler, K., and Granick, S.** (1970). Controls on chlorophyll synthesis in barley. Plant Physiol. **46,** 240–246.

**Nair, S.P., Harwood, J.L., and John, R.A.** (1991). Direct identification and quantification of the cofactor in glutamate semialdehyde aminotransferase from pea leaves. FEBS Lett. **283,** 4–6.

**Petricek, D., Rutberg, L., Schröder, I., and Hederstedt, L.** (1990). Cloning and characterization of the *hemA* region of the *Bacillus subtilis* chromosome. J. Bacteriol. **172,** 2250–2258.

**Sangwan, I., and O'Brian, M.R.** (1993). Expression of the soybean (*Glycine max*) glutamate 1-semialdehyde aminotransferase gene in symbiotic root nodules. Plant Physiol. **102,** 829–834.

**Sasarman, A., Surdeanu, M., Szegli, G., Horodniceanu, T., Greceanu, V., and Dumitrescu, A.** (1968). Hemin deficient mutants of *Escherichia coli* K-12. J. Bacteriol. **96,** 570–572.

**Senecoff, J.F., and Meagher, R.B.** (1993). Isolating the *Arabidopsis thaliana* genes for *de novo* purine synthesis by suppression of *Escherichia coli* mutants. I. 5′-Phosphoribosyl-5-aminoimidazole synthetase. Plant Physiol. **102,** 387–399.

**Senge, M.O.** (1993). Recent advances in the biosynthesis and chemistry of chlorophylls. Photochem. Phytobiol. **57,** 189–206.

**Sisler, E.C., and Klein, W.H.** (1963). The effect of age and various chemicals on the lag phase of chlorophyll synthesis in dark grown bean seedlings. Physiol. Plant. **16,** 315–322.

**Snustad, D.P., Hunsperger, J.P., Chereskin, B.M., and Messing, J.** (1988). Maize glutamine synthetase cDNAs: Isolation by direct genetic selection in *Escherichia coli.* Genetics **120,** 1111–1124.

**Studier, W.F., Rosenberg, A.H., Dunn, J.J., and Dudendorf, J.W.** (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **85,** 60–89.

**Verkamp, E., and Chelm, B.K.** (1989). Isolation, nucleotide sequence, and preliminary characterization of the *Escherichia coli* K-12 *hemA* gene. J. Bacteriol. **171,** 4728–4735.

**Verkamp, E., Jahn, M., Jahn, D., Kumar, A.M., and Söll, D.** (1992). Glutamyl-tRNA reductase from *Escherichia coli* and *Synechocystis* 6803. J. Biol. Chem. **267,** 8275–8280.

**von Heijne, G., Steppuhn, J., and Herrmann, R.G.** (1989). Domain structure of mitochondrial and chloroplast targeting peptides. Eur. J. Biochem. **180,** 535–545.

**Wierenga, R.K., Terpstra, P., and Hol, W.G.J.** (1986). Prediction of occurrence of the ADP binding βαβ fold in proteins, using an amino acid sequence finger print. J. Mol. Biol. **187,** 101–107.