Maize Glutamine Synthetase cDNAs: Isolation by Direct Genetic Selection in *Escherichia coli*

D. Peter Snustad,* John P. Hunsperger,* Barbara M. Chereskin† and Joachim Messing†

**Department of Genetics and Cell Biology and Plant Molecular Genetics Institute, University of Minnesota, St. Paul, Minnesota 55108, and †Waksman Institute, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08855*

Manuscript received May 2, 1988
Accepted September 8, 1988

ABSTRACT

Maize glutamine synthetase cDNA clones were isolated by genetic selection for functional rescue of an *Escherichia coli* Δ*glnA* mutant growing on medium lacking glutamine. The Black Mexican Sweet cDNA library used in this study was constructed in pUC13 such that cDNA sense strands were transcribed under the control of the *lac* promoter. *E. coli* Δ*glnA* cells were transformed with cDNA library plasmid DNA, grown briefly in rich medium to allow phenotypic expression of the cDNAs and the pUC13 *amp*^{r} gene, and challenged to grow on agar medium lacking glutamine. Large numbers of glutamine synthetase cDNA clones have been identified in individual 150-mm Petri dishes; all characterized cDNA clones carry complete coding sequences. Two cDNAs identical except for different 5′ and 3′ termini have been sequenced. The major open reading frame predicts a protein with an amino acid sequence that exhibits striking similarity to the amino acid sequences of the predicted products of previously sequenced eukaryotic glutamine synthetase cDNAs and genes. In addition, the maize glutamine synthetase cDNAs were shown to contain a 5′ mini-ORF of 29 codons separated by 37 nucleotide pairs from the major ORF. This mini-ORF was shown not to be essential for the functional rescue of the *E. coli* Δ*glnA* mutant. Expression of the cDNAs in *E. coli* is presumed to be due to the function of a polycistronic hybrid *lac* messenger RNA or translational fusions encoded by the pUC plasmids. Proteins of the expected sizes encoded by two different pUC clones were shown to react with antibodies to tobacco glutamine synthetase.

SPECIFIC eukaryotic genes or cDNAs have been isolated most frequently by screening genomic or cDNA libraries for sequences that cross-hybridize with specific nucleic acid probes or by screening expression libraries for the production of proteins that cross-react with specific antibody probes. GRUNSTEIN and HOGNESS (1975) initially developed procedures for hybridization screening of up to 10^3 recombinant plasmid-containing colonies per 150-mm plate. BENTON AND DAVIS (1977) devised protocols for screening up to 2×10^4 recombinant phage λ plaques per 150-mm plate. Subsequently, HANAHAN and MESELSON (1980) described colony-lift hybridization procedures that permitted 10^5 recombinant clones to be screened per 150-mm plate. The applicability of these techniques to the isolation of DNA sequences which were previously genetically undefined and were thus not subject to genetic manipulations has proven especially significant.

Clearly, one disadvantage of these colony- and plaque-lift procedures is the limited number of recombinant clones that can be conveniently screened. By comparison, the resolving power of genetic selection, as so elegantly demonstrated in the neoclassical studies of BENZER (1957), is orders of magnitude greater—10^8 to 10^9 cells or viruses can be screened on a single plate. This comparison indicates that genetic selection should be the screening tool of choice whenever it is applicable. Indeed, genetic selection has been used quite successfully to isolate prokaryotic genes and cDNAs (for example, see BONDARYK and PAULUS 1985; BAZZICALUPO *et al.* 1987). In contrast, it seems to have been largely overlooked in similar studies of eukaryotes. Although the presence of introns in most eukaryotic genes would often exclude the possibility of using genetic selection in bacteria to identify genomic clones, this obviously would not be a factor in screening eukaryotic cDNA libraries.

Our decision to attempt to use genetic selection to screen maize cDNA libraries was based on the following premises. (1) There is a vast repertoire of *Escherichia coli* auxotrophic mutants carrying null mutations in genes of interest. (2) In most, if not all cases, folding and subunit associations of proteins are specified by their primary structures. (3) All that is required for functional rescue of a mutant strain is substrate to product conversion; what kind of machine catalyzes the conversion is irrelevant. (4) The catalysis does not need to be efficient; even a small amount of product can permit survival and growth, albeit slow, of the mutant cell. If these premises are correct, it should be possible to identify and isolate specific eukaryotic

cDNAs on the basis of their ability to rescue *E. coli* mutant cells when plated on restrictive medium. In particular, this approach should be effective for cDNAs which encode proteins that function as monomers or homomultimers and without obligatory interaction with any component of the host cell. Glutamine synthetase (GS) cDNAs were chosen as the focus of our initial attempt to isolate eukaryotic cDNAs by direct genetic selection in *E. coli* because of (1) the key role that GS plays in nitrogen metabolism in plants (reviewed by MIFLIN and LEA 1980; STEWART, MANN and FENTEM 1980) and (2) DASSARMA, TISCHER and GOODMAN'S (1986) demonstration of alfalfa GS activity in *E. coli* encoded by a clone constructed *in vitro* by splicing the 5′ segment of the coding region from a genomic clone to a partial 3′ cDNA clone.

In the present study we have characterized maize GS cDNAs isolated from a cDNA library by screening for functional rescue of an *E. coli* Δ*glnA* mutant. Our approach was to use the plasmid vector pUC13 (MESSING 1983) as an expression vector in a manner similar to that described by HELFMAN *et al.* (1983). Two maize GS cDNAs with different 5′ and 3′ termini, but identical ORFs, have been sequenced. Both cDNAs contain a 5′ mini-ORF of 29 codons separated from the major GS-encoding ORF by 37 nucleotides. We have no evidence that the mini-ORF is translated; however, the AUG is in strong context for initiation (KOZAK 1986a) with G's at both the −3 and +4 positions. The predicted amino acid sequence of the maize GS exhibits striking similarity to the GSs of dicotyledonous plants and Chinese hamster. Western blot analyses of proteins in Δ*glnA E. coli* cells transformed with maize GS cDNA clones demonstrated the presence of anti-tobacco GS cross-reacting proteins of the expected sizes.

MATERIALS AND METHODS

Plasmid vectors and bacterial strains: Plasmid vectors pUC13 and pUC119 were our own stock. MV1190 or MV1184 was used as the host for pUC119 and derivatives in all cloning and sequencing experiments [see VIEIRA and MESSING (1987) for genotypes and methodologies]. *E. coli* strains FDB213, LE392, and DH1 were obtained from F. AUSUBEL, L. ENQUIST, and D. HANAHAN, respectively. The genotype of FDB213 is *endoA thi-1 hsdR17 supE44 pro* Δ*glnAG* (DE BRUIJN and AUSUBEL 1981; DASSARMA, TISCHER and GOODMAN 1986). Derivatives of LE392 (r^-m^-) and DH1 (r^-m^+) harboring F′ *lacI*^{q}*Z::Tn5* Y^+A^+ were used as host cells for the cDNA library; see MANIATIS, FRITSCH and SAMBROOK (1982) for the complete genotypes of LE392 and DH1.

Plant materials: The Black Mexican Sweet (BMS) cell culture lines (SHERIDAN 1975, 1982; GREEN 1977) were obtained from C. E. GREEN. Seeds of maize lines BSSS-53, W23, W64A, B37, I11-12E, A188 and W22 were provided by R. L. PHILLIPS. BMS culture cells were grown in MS medium (MURASHIGE and SKOOG 1962) supplemented with 2,4-D (2 mg/liter), asparagine (200 mg/liter), and sucrose (2%); cultures were transferred weekly.

cDNA library construction: The cDNA synthesis was performed by a vector-primed method designed for use with advanced pUC plasmids (J. P. HUNSPERGER and I. RUBENSTEIN, unpublished data) using poly(A$^+$) RNA prepared as previously described (LIZARDI and ENGELBERG, 1979; KIRIHARA *et al.* 1988). Vector DNA was digested with *Sac*I and T-tailed. The DNA was then digested with *Bam*HI to provide a single priming site for reverse transcriptase. Ten micrograms of methyl mercury-denatured poly(A$^+$) RNA was annealed to 2 μg of vector-primer in a first strand synthesis reaction polymerized by AMV reverse transcriptase. After second strand synthesis (OKAYAMA and BERG 1982; HEIDECKER and MESSING 1983), duplex cDNA-vector was methylated with *Eco*RI methylase, ligated to *Eco*RI octameric linkers, and digested with *Eco*RI. The population of linearized cDNA-vector molecules was size-fractionated (10 fractions) on an agarose gel, ligated, and transformed into LE392 F′ *lacI*^{q} and later into DH1 F′ *lacI*^{q} by the procedure of HANAHAN (1983). This library (designated BMS-3) was stored frozen in 20% glycerol at −80°.

Isolation of GS cDNA clones: Library fractions 4 and 5 contained plasmids with cDNA inserts of average size 1.9 kb and 1.4 kb, respectively. These fractions should contain cDNAs large enough to encode proteins the size of GS monomers (~40 kD) plus 5′ and 3′ noncoding sequences and poly(A) tails of 100–300 bp. Thus, only library fractions 4 and 5 were screened.

FDB213 cells were transformed by either the simple $CaCl_2$ procedure or the high frequency transformation protocol of HANAHAN (1983). Maize GS cDNA clones were identified using cells transformed by both procedures. After the heat-shock step, the cells were allowed to cool to room temperature before adding 1.8 ml of 2 × YT medium or SOC medium supplemented with L-glutamine (5 mM) and incubating on a roller drum at 37° for 1 hr. In the initial experiments, the cells were then collected by centrifugation for four min at 3000 rpm in a clinical centrifuge (4°), washed once by resuspension in M9 medium (MILLER 1972) and recentrifugation, and then gently spread on the surface of M9 minimal agar medium containing 0.2% glucose and 2 mM proline. The plates were wrapped in plastic food wrap to prevent dehydration and incubated at 37° in an inverted position. Colonies first became visible after 2 1/2 days and continued to appear for several days. In later experiments, ampicillin (50 or 100 μg/ml) was added to the M9 minimal medium used for selection. The presence of ampicillin decreases the chance of contaminants forming colonies and does not seem to have any deleterious effect on the plating efficiency of rescued cells.

Reconstruction experiments carried out with the initial clones indicated that plating efficiency was quite variable and sometimes quite low. We have observed that this can be largely overcome by allowing the transformed cells to undergo the physiological changes coincident with step-down growth (transfer from rich to minimal medium) in liquid M9 medium supplemented with L-glutamine (10 mM) prior to spreading them on the selection plates. Specifically, there is a lag of about 2 hr before cells subjected to step-down conditions of the type used here start to grow and divide again. Thus, in more recent experiments, the transformed FDB213 cells have been incubated for 2 hr in M9 minimal medium supplemented with glucose (0.2%), L-proline (2 mM), L-glutamine (10 mM), and ampicillin (50 μg/ml) at 37° with aeration by shaking before washing with M9 medium without supplements and spreading on selection plates. Colonies that formed on the selection plates were streaked out on fresh M9 selection plates, and single colonies

were used to grow cultures for plasmid isolations and for storage in glycerol at −80°.

Colony hybridizations were carried out using nick-translated, 2× agarose gel-purified MGS1 cDNA as probe (WEINSTOCK *et al.* 1978). Hybridizations were done at 55° without formamide.

Growth of *E. coli* FDB213 derivatives: Overnight cultures of *E. coli* FDB213 derivatives were grown at 37° with aeration by shaking in M9 medium supplemented with 0.2% glucose, 2 mM L-proline, 10 mM L-glutamine, and 50 μg/ml ampicillin after inoculation from stocks frozen in glycerol at −80°. The cells were collected by centrifugation, washed once with unsupplemented M9 medium, resuspended in unsupplemented M9 medium, and used to inoculate the desired medium. Growth was monitored by measuring the OD_{550} of the culture with a Beckman DB-G spectrophotometer.

Subcloning, template preparation, and DNA sequencing: Recombinant plasmid DNA was digested with restriction enzymes and ligated to appropriately digested pUC119 vector DNA (VIEIRA and MESSING 1987). Packaged single-stranded plasmid DNA was prepared and used for template preparation as described by MCMULLEN *et al.* (1986). Overlapping deletion subclones of the recombinant plasmids were generated by the procedure of DALE, MCCLURE and HOUCHINS (1985). Subclones were sequenced by the dideoxynucleotide chain termination method of SANGER, NICKLEN and COULSON (1977) using a synthetic primer (M13 sequencing primer [17 mer]; Amersham Corp.). Standard sequencing reactions using the Klenow fragment of *E. coli* DNA polymerase I were performed as described by WALKER and GAY (1983) except that [α-^{35}S]dATP (800 Ci/mM; Amersham Corp.) was used as the labeled nucleotide. Sequencing reactions with Klenow and 7-deaza-2′-deoxyguanosine-5′-triphosphate mixes were carried out as described by BARR *et al.* (1986). Sequencing reactions catalyzed by Sequenase were performed according to the protocol provided by the supplier (U.S. Biochemical Corp.). DNA fragments were fractionated on 6% and 7% polyacrylamide sequencing gels as described by CARLSON and MESSING (1984). In a few cases, gels that contained 40% formamide were used to verify the absence of compressions (MARTIN 1987). Sequence data were analyzed by using the IntelliGenetics software on a Sun Microsystems 2/120 computer or the UWGCG programs on a Vax 785 (DEVEREUX, HAEBERLI and SMITHIES 1984).

Deletion of the mini-ORF of cDNA clone pMGS1: Removal of the mini-ORF of pMGS1 was facilitated by a unique *Nar*I recognition sequence at position −31 to −26 (cleavage site 5′ to nucleotide −27) (see Figures 3 and 4). The complete MGS1 cDNA was excised from pMGS1 with *Eco*RI, separated from vector DNA by agarose gel electrophoresis, extracted from the agarose gel slice with GENECLEAN, and then cut with *Nar*I. The large *Nar*I-*Eco*RI restriction fragment containing the major ORF was again gel isolated and extracted from the excised gel slice with GENECLEAN. *Nar*I cleavage produces 5′ CG overhangs that are identical to those produced by cleavage of the pUC119 polylinker with *Acc*I. The *Nar*I-*Eco*RI fragment of MGS1 was then ligated into *Acc*I, *Eco*RI-double cut pUC119 to force clone the fragment into pUC in the sense orientation. The resulting plasmid, designated pMGS1-S ("S" for *S*hort), carries the major ORF out of frame (in the +1 reading frame) relative to the *lac* initiation codon. Therefore, an insertion of one or four bp or a deletion of two bp is required to place the major ORF of pMGS1-S in frame with the *lac* AUG. Although the insertion of four bp at the *Hin*dIII site of the polylinker is the simplest, it generates a distal in-frame TAG termination codon. We thus deleted two bp from the polylinker region as follows. Plasmid pMGS1-S was cleaved with *Hin*dIII which yields 5′ AGCT overhangs. Klenow and dATP were used to add a single A to each 3′ terminus, a brief treatment on ice with S1 nuclease was used to cleave off the remaining AGC 5′ overhangs for some of the molecules, and the nascent ends were joined by unimolecular blunt-end ligation. The desired in-frame derivatives, designated pMGS1-SR ("SR" for *S*hort, *R*escues), were identified by transforming *E. coli* FDB213 cells with the ligation reaction products and plating the transformed cells on selective plates (M9 medium, + proline, − glutamine).

Western blot analysis: Overnight cultures of *E. coli* strains FDB213(pUC119), FDB213(pMGS1), and FDB213 (pMGS3) were grown in LB medium supplemented with 1% glycerol and 150 μg/ml ampicillin. The culture medium for FDB213(pUC119) also contained 10 mM L-glutamine. The cells were collected by centrifugation and resuspended directly in sample loading buffer (13% glycerol, 65 mM Tris-HCl (pH 6.8), 2.5% SDS, and 1.3% 2-mercaptoethanol). The cells were disrupted by sonication.

Maize leaf extracts were prepared from leaves of 10-day-old seedlings of inbred line W23. Tissue was ground in a Polytron for 30 sec at high power in 3 ml/gm tissue homogenization buffer (18% sucrose, 10 mM $MgCl_2$, 0.1 M Tris-HCl (pH 8), 40 mM 2-mercaptoethanol). The brei was filtered through miracloth and microfuged for 10 min. The supernatant solution was adjusted to 2% SDS and 6% sucrose for gel loading.

Proteins were separated by electrophoresis using a 9% SDS polyacrylamide gel and were transferred to nitrocellulose with a Hoefer TE42 transphor cell by the protocol provided by the manufacturer. After blocking nonspecific protein binding sites with 5% Carnation non-fat dry milk, the membrane was incubated first with 0.1% (v/v) antiserum to tobacco glutamine synthetase (generously provided by B. HIREL) and then with 10 μCi ^{125}I-labeled protein A (ICN 68038) as described by TOWBIN, STAEHELIN and GORDON (1979).

Genomic blot hybridizations: Genomic DNA was isolated from leaf tissue of 3-week-old maize seedlings as described by SHURE, WESSLER and FEDOROFF (1983). Genomic DNA was isolated from BMS culture cells by an analogous protocol except that Whatman CF-11 cellulose was used to remove carbohydrates as described by MOZER (1980) prior to precipitation. DNAs were digested with the appropriate restriction enzyme and subjected to electrophoresis through 0.8% agarose gels. Transfers of DNA fragments to Zeta Probe (Bio-Rad) nylon filters were done as described by SOUTHERN (1975) with the modifications of REED and MANN (1985). The filters were prewashed in 0.1 × SSC, 0.1% SDS for 1 hr at 65° with agitation, and prehybridized overnight at 42° in a solution that contained either 40% or 50% formamide (MCB Manufacturing Chemists, Inc.), 5 × SSCP, 50 mM Tris-HCl (pH 7.5), 1 × Denhardt's solution (0.02% each of bovine serum albumin, Ficoll, and polyvinylpyrolidone), 1% SDS, 100 μg/ml of sheared, denatured salmon sperm DNA, and 5% dextran sulfate. Hybridization was carried out in prehybridization solution which contained 1–2 × 10^6 cpm/ml of ^{32}P-labeled probe at 37° or 42° for 24–36 hr. The probe used was 2 × agarose gel-purified MGS1 cDNA labeled by nick-translation (WEINSTOCK *et al.* 1978). The final two filter washes were for 30 min each in 0.1 × SSC, 0.2% SDS at 60°.

RESULTS

Isolation of maize GS cDNA clones: When *E. coli* FDB213 cells, which contain a deletion of the gluta-

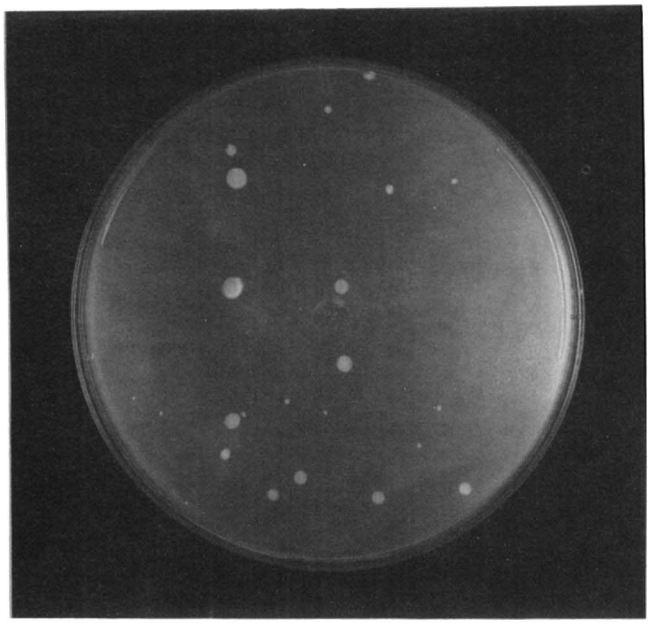

FIGURE 1.—Functional rescue of Δ*glnA* mutant FDB213 *E. coli* cells after transformation with plasmids in fraction 5 (average insert size = 1.4 kb) of the maize BMS-3 cDNA library.

mine synthetase structural gene, were transformed with fractions 4 and 5 (average cDNA insert size = 1.9 and 1.4 kb, respectively) of the pUC13-derived maize BMS culture cell cDNA expression library and were plated on selective medium lacking glutamine, about one of 10^5 transformants was functionally rescued and capable of growth without exogenous glutamine. These fractions of the library were selected for screening based on the estimated sizes of plant GS monomers (reviewed by STEWART, MANN and FENTEM 1980) and the estimated sizes of GS mRNAs in *Phaseolus vulgaris* (CULLIMORE *et al.* 1984) and *Medicago sativa* (DONN *et al.* 1984). A major advantage of the genetic selection procedure used here over colony lift protocols is that large numbers of transformants can be screened, and thus many independent cDNA clones identified, on a single selective plate. For example, Figure 1 shows a plate from which 25 maize GS cDNA clones were isolated. Since strain FDB213 is also a proline auxotroph, the possibility that the viable colonies were produced by contaminating cells was examined by sterile toothpick transfers of cells from the selected colonies to minimal medium lacking proline. All rescued cells were still Pro^-. These rescued cell lines were then grown overnight in rich medium containing ampicillin (150 μg/ml), and plasmid preparations were made and used in a second cycle of transformation of FDB213 with the result that all transformants were now capable of growth without exogenous glutamine.

When the sizes of the cDNA inserts in the selected recombinant plasmids were examined by agarose electrophoresis, they were found to range from ~1.5 kb to ~1.7 kb. Two clones, one (designated MGS1) with

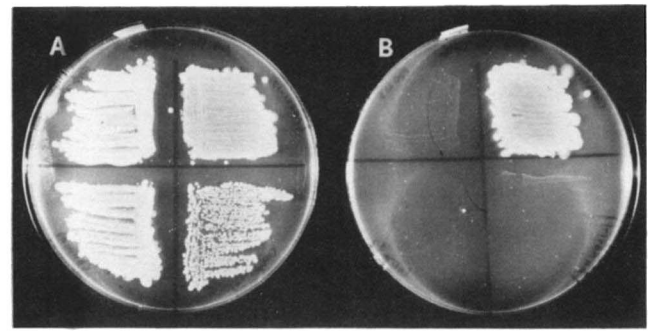

FIGURE 2.—Functional rescue of the Δ*glnA E. coli* cells requires a correctly oriented maize GS cDNA insert. M9 minimal agar medium containing 0.2% glucose was supplemented with (A) 2 mM L-proline plus 10 mM L-glutamine and (B) only 2 mM L-proline. Bacteria were spread on quadrants of the plates as follow: (top left) the *E. coli* Δ*glnA* mutant strain FDB213; (bottom left) FDB213 (pUC119); (top right) FDB213 (pMGS1), pMGS1 is pUC119 carrying the maize GS1 cDNA in the sense orientation; (bottom right) FDB213(pnMGS1), pnMGS1 is pUC119 carrying the maize GS1 cDNA in the antisense orientation.

insert size ~1.5 kb from library fraction 5 and one (designated MGS2) with insert size ~1.7 kb from fraction 4, were chosen for initial characterization. Both cDNAs were transferred into pUC119 (VIEIRA and MESSING 1987) for sequence analysis. In order to verify that the functional rescue of the Δ*glnA* mutation in FDB213 cells was due to the maize cDNA inserts and not some unknown aspect of the vector *per se*, we examined the growth of FDB213 cells transformed with pUC119 alone (no insert) and with pUC119 containing the MGS1 cDNA insert in both orientations. The results (Figure 2) clearly demonstrated that functional rescue of the FDB213 *glnA* defect occurred only when the MGS1 cDNA insert was present in one orientation (subsequently shown to be the sense orientation by sequence analysis, see below) and not in the other (antisense) orientation.

Sequence analysis: The MGS1 and MGS2 cDNAs were subcloned into pUC119 in both orientations, and a set of overlapping subclones was generated for each orientation of each cDNA by the unidirectional deletion procedure of DALE, MCCLURE and HOUCHINS (1985). The strategies employed for dideoxy sequencing (SANGER, NICKLEN and COULSON 1977) of the two cDNAs are shown in Figure 3; the sequences are presented in Figure 4. The two cDNAs differed only at their termini. They were almost certainly synthesized from transcripts of the same gene produced by the utilization of different polyadenylation signals. When MGS1 and MGS2 were compared, MGS2 was found to be 17 bp longer at the 5′ terminus and to contain an extra 93 bp of genome-encoded sequence at the 3′ end as well as an approximately 100 bp-longer poly(A) tail.

The surprising result was that both cDNAs contained a small (29 codons) ORF located upstream from the major ORF (Figure 4). The first AUG is located

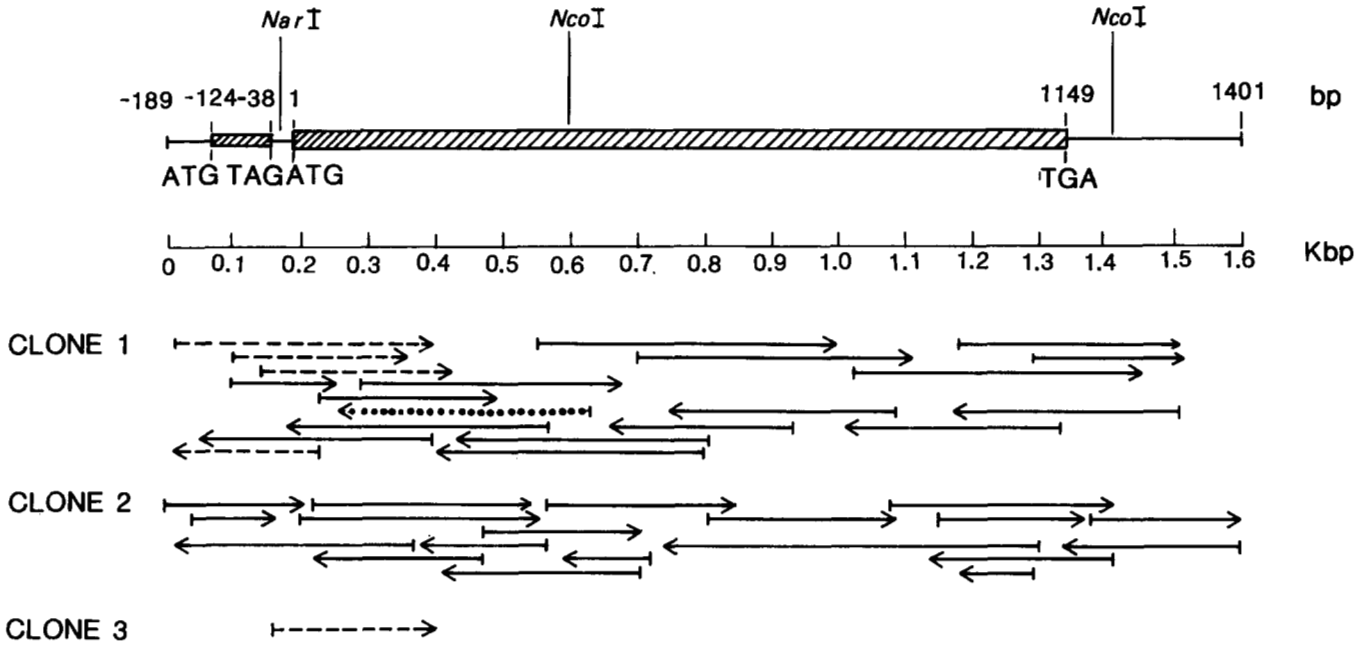

FIGURE 3.—Physical map of the maize MGS1 and MGS2 cDNA clones (top) and diagram of the strategies used in sequencing MGS1, MGS2, and the 5′ terminus of MGS3 (bottom). ORFs are shown as boxes. The A of the AUG initiation codon of the major GS-encoding ORF is identified as nucleotide 1; upstream (5′) sequences are numbered −1 through −189 (the 5′ end of the longest GS clone, MGS2). Clones 1, 2, and 3 are MGS1, MGS2, and MGS3, respectively. The arrows indicate the directions and extents of DNA sequence determined for each subclone. All subclones were sequenced at least twice; the critical subclones of MGS1 spanning the −38 to −124 mini-ORF were sequenced at least five times. MGS1 and MGS3 subclones were sequenced by D. P. SNUSTAD at Minnesota; MGS2 subclones were sequenced by B. M. CHERESKIN at Rutgers. All subclones were sequenced using standard dG mixes and either Klenow or Sequenase enzyme or both. Subclones represented by dashed arrows were also sequenced using both (1) 7-deaza-dG mixes and Klenow and (2) dI mixes and Sequenase. The dotted arrow indicates a subclone that was analyzed using a polyacrylamide gel containing 7 M urea and 40% formamide. *Nar*I and *Nco*I identify restriction enzyme cleavage sites utilized in subcloning or as diagnostic markers.

48 bp and 65 bp from the 5′ termini of the MGS1 and MGS2 cDNAs, respectively. This AUG is in strong context for translational initiation (KOZAK 1984, 1986a) with a G at −3 and a G at +4, just like the AUG of the major ORF. This first 5′ AUG is followed by 27 codons (one an AUG in weak context) and then a TAG termination codon. The second AUG that is in strong context for initiation is located 37 bp away (3′) from the TAG mini-ORF termination codon. This second AUG begins the major ORF which predicts a protein of 382 amino acids with striking similarity to the proteins predicted by other sequenced eukaryotic GS genes (see Figure 8).

Since the observed sequence is not consistent with the rule of monocistronic mRNAs in eukaryotes and therefore with KOZAK's (1978) scanning mechanism, we carefully repeated the sequence analysis in this region by isolating and sequencing additional subclones with staggered termini within the 189 bp at the 5′ ends of the cDNAs (Figure 3). These analyses yielded maximal band separations on short sequences of 50 bp or less from the bottom of the wedge gel (CARLSON and MESSING 1984). Sequencing gels were run at 2500 V (constant voltage) to minimize compressions, as well as at the usual 1600 V. In addition, the key templates spanning the mini-ORF were sequenced with (1) 7-deaza-2′-deoxyguanosine-5′-triphosphate mixes with "Klenow" enzyme (BARR *et al.* 1986) and (2) 2′-deoxyinosine-5′-triphosphate mixes (GOUGH and MURRAY 1983) and Sequenase (TABOR and RICHARDSON 1987). Finally, we used sequencing gels containing 7 M urea and 40% formamide (MARTIN 1987) to separate the reaction products for some templates. The results obtained by using these approaches clearly confirmed the sequences presented in Figure 4.

Deletion of the mini-ORF of pMGS1 results in increased GS activity in *E. coli*: In order to obtain independent evidence that the TAG termination codon of the mini-ORF was in-frame and functional, we tested whether the deletion of the mini-ORF from the MGS1 cDNA might eliminate or decrease its ability to rescue the FDB213 Δ*glnA* mutant. The strategy that we used to delete the mini-ORF from pMGS1 is described in MATERIALS AND METHODS and is diagrammed in Figure 5. The resulting plasmid, designated pMGS1-SR, contains the major ORF of pMGS1 in frame with the *lac* AUG. When the growth rates of FDB213(pMGS1) and FDB213(pMGS1-SR) in proline-supplemented M9 lactose medium were compared, FDB213(pMGS1-SR) was found to grow about four times as fast as FDB213(pMGS1) (Figure 6). Clearly, deletion of the mini-ORF did not eliminate or decrease GS activity; to the contrary, it resulted in increased GS activity. This result thus provides addi-

```
-189 taggt ctcggcggga gaggggcggc ggccggtccg tgtccgtgtc cgtcgacggt tggttcggga -125

ATG GCG CAG GCG GTG GTG CCG GCG ATG CAA GTG CCG GGT CGG AGT GAA GGC GGC      -71
MET Ala Gln Ala Val Val Pro Ala MET Gln Val Pro Gly Arg Ser Glu Gly Gly

GGC GGG GAG GGT GTG GAG CGC CGG CAG GAC TAG gaccggccgc ggcggcgcct cgccggggtt  -8
Gly Gly Glu Gly Val Glu Arg Arg Gln Asp

caaggtc ATG GCC GTC AGC ACG GGC AGC ACC GGG GTG GTG CCG CGC CTC GAG CAG      48
       MET Ala Val Ser Thr Gly Ser Thr Gly Val Val Pro Arg Leu Glu Gln

CTG CTC AAC ATG GAC ACC ACG CCC TAC ACC GAC AAG GTC ATC GCC GAG TAC ATC     102
Leu Leu Asn MET Asp Thr Thr Pro Tyr Thr Asp Lys Val Ile Ala Glu Tyr Ile

TGG GTC GGA GGA TCT GGA ATC GAC ATC AGA AGC AAA TCA AGG ACG ATT TCG AAA     156
Trp Val Gly Gly Ser Gly Ile Asp Ile Arg Ser Lys Ser Arg Thr Ile Ser Lys

CCC GTG GAG GAT CCC TCA GAA CTA CCA AAA TGG AAC TAC GAT GGA TCT AGC ACA     210
Pro Val Glu Asp Pro Ser Glu Leu Pro Lys Trp Asn Tyr Asp Gly Ser Ser Thr

GGA CAA GCC CCG GGA GAA GAC AGT GAA GTC ATT CTA TAC CCC CAG GCT ATC TTC     264
Gly Gln Ala Pro Gly Glu Asp Ser Glu Val Ile Leu Tyr Pro Gln Ala Ile Phe

AAG GAC CCA TTC CGA GGT GGC AAC AAC GTT TTG GTT ATC TGT GAC ACC TAC ACG     318
Lys Asp Pro Phe Arg Gly Gly Asn Asn Val Leu Val Ile Cys Asp Thr Tyr Thr

CCA CAG GGG GAA CCC CTT CCA ACT AAC AAA CGC CAC AGG GCT GCG CAA ATT TTC     372
Pro Gln Gly Glu Pro Leu Pro Thr Asn Lys Arg His Arg Ala Ala Gln Ile Phe

AGC GAC CCA AAG GTC GGT GAA CAA GTG CCA TGG TTT GGC ATA GAG CAA GAG TAC     426
Ser Asp Pro Lys Val Gly Glu Gln Val Pro Trp Phe Gly Ile Glu Gln Glu Tyr

ACT TTG CTC CAG AAA GAT GTA AAT TGG CCT CTT GGT TGG CCT GTT GGA GGC TTC     480
Thr Leu Leu Gln Lys Asp Val Asn Trp Pro Leu Gly Trp Pro Val Gly Gly Phe

CCT GGT CCC CAG GGT CCA TAC TAC TGT GCC GTA GGA GCC GAC AAA TCA TTT GGC     534
Pro Gly Pro Gln Gly Pro Tyr Tyr Cys Ala Val Gly Ala Asp Lys Ser Phe Gly

CGT GAC ATA TCA GAT GCT CAC TAC AAG GCA TGC CTC TAC GCT GGA ATC AAC ATT     588
Arg Asp Ile Ser Asp Ala His Tyr Lys Ala Cys Leu Tyr Ala Gly Ile Asn Ile

AGT GGA ACA AAC GGG GAG GTC ATG CCT GGT CAG TGG GAG TAC CAA GTT GGA CCT     642
Ser Gly Thr Asn Gly Glu Val MET Pro Gly Gln Trp Glu Tyr Gln Val Gly Pro

AGT GTT GGT ATT GAA GCA GGA GAT CAC ATA TGG ATT TCG AGA TAC ATT CTC GAG     696
Ser Val Gly Ile Glu Ala Gly Asp His Ile Trp Ile Ser Arg Tyr Ile Leu Glu

AGA ATC ACA GAG CAA GCT GGG GTT GTC CTT ACC CTT GAT CCA AAA CCA ATT CAG     750
Arg Ile Thr Glu Gln Ala Gly Val Val Leu Thr Leu Asp Pro Lys Pro Ile Gln

GGT GAC TGG AAC GGA GCT GGC TGC CAC ACA AAT TAC AGC ACA AAG ACC ATG CGC     804
Gly Asp Trp Asn Gly Ala Gly Cys His Thr Asn Tyr Ser Thr Lys Thr MET Arg

GAA GAC GGC GGG TTT GAA GAG ATC AAG AGA GCA ATC CTG AAC CTT TCT CTG CGC     858
Glu Asp Gly Gly Phe Glu Glu Ile Lys Arg Ala Ile Leu Asn Leu Ser Leu Arg

CAT GAT CTG CAT ATT AGT GCA TAC GGA GAA GGA AAT GAA AGA AGA TTG ACT GGG     912
His Asp Leu His Ile Ser Ala Tyr Gly Glu Gly Asn Glu Arg Arg Leu Thr Gly

AAA CAT GAG ACT GCG AGC ATC GGA ACC TTC TCA TGG GGT GTG GCA AAC CGC GGC     966
Lys His Glu Thr Ala Ser Ile Gly Thr Phe Ser Trp Gly Val Ala Asn Arg Gly

TGC TCT ATC CGT GTG GGG CGG GAT ACC GAG GCA AAA GGG AAA GGT TAC CTG GAA    1020
Cys Ser Ile Arg Val Gly Arg Asp Thr Glu Ala Lys Gly Lys Gly Tyr Leu Glu

GAC CGT CGG CCG GCA TCA AAC ATG GAC CCG TAC ATT GTG ACG GGG CTA CTG GCC    1074
Asp Arg Arg Pro Ala Ser Asn MET Asp Pro Tyr Ile Val Thr Gly Leu Leu Ala

GAG ACC ACG ATC CTC TGG CAG CCA TCC CTC GAG GCG GAG GCT CTT GCC GCC AAG    1128
Glu Thr Thr Ile Leu Trp Gln Pro Ser Leu Glu Ala Glu Ala Leu Ala Ala Lys

AAG CTG GCG CTG AAG GTG TGA agcagctgaa ggatggttca ggcaccaata taaacggtc     1189
Lys Leu Ala Leu Lys Val

cgcgacaaga ttgatctttg tgtccatggc cgttgggtct tgcgactctc tgctcggcgg tgccactctg 1259

tacaaaatca cggctgtctt tgattcatcg gatattcgga tacgtttgtt tgttactttt tgcttggaca 1329

cccacatgtt tggaactttt ttgggctccg tttgggggct gaacgatggt cagtggaaat tttaagaatt 1399

cgtcgtc (A)n                                                               1406
```

FIGURE 4.—Nucleotide sequences of the MGS1 and MGS2 cDNAs, and the predicted translation products of the two ORFs present on these cDNAs. The A of the initiator AUG of the major ORF is defined as nucleotide position 1; upstream nucleotides are numbered with minus designators. The termini of the MGS1 cDNA are indicated by the solid arrows. The 5′ terminus of the MGS3 cDNA is indicated by the open arrow.

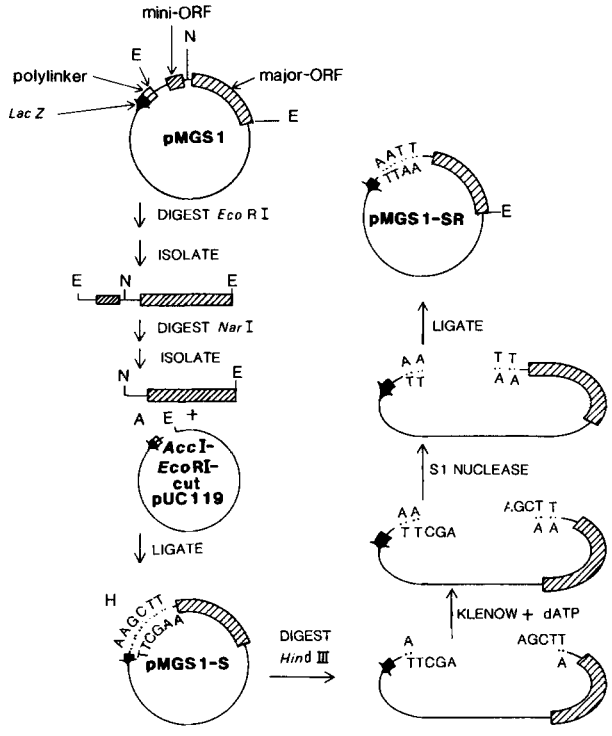

FIGURE 5.—Schematic diagram (not to scale) of the procedures used to (1) delete the mini-ORF from pMGS1 (pUC119 containing the entire MGS1 cDNA, see Figures 3 and 4) to produce pMGS1-S and (2) to position the major ORF in frame with the *lacZ* AUG to produce pMGS1-SR. Details of the procedures are given in the text.

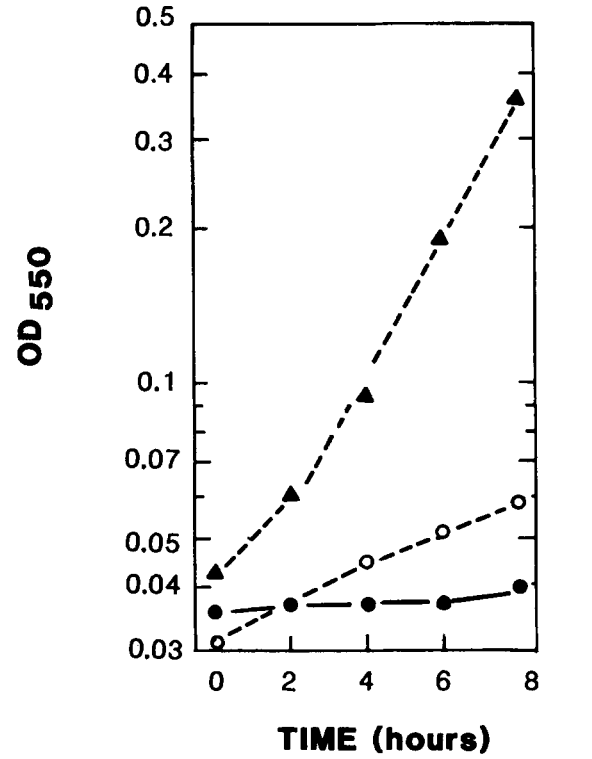

FIGURE 6.—Growth rates of FDB213 (pMGS1) (○), FDB213 (pMGS1-S) (●), and FDB213(pMGS1-SR) (▲) in M9 minimal medium containing 0.2% lactose and 2 mM proline. Fresh overnight cells grown in M9 medium with glucose, proline, and glutamine were washed and diluted 40-fold into M9 medium with lactose and proline at time 0. Cells were incubated at 37° with aeration by shaking.

tional evidence for the accuracy of the cDNA structure shown in Figure 4; in particular, it demonstrates that the mini-ORF is not essential for GS activity.

Preliminary characterization of additional GS clones: The sequenced MGS1 and MGS2 cDNAs contain two *Nco*I cleavage sites: one between nucleotides 400 and 401 and the second between nucleotides 1213 and 1214 (Figure 4). Cleavage of plasmids pMGS1 and pMGS2 thus cuts out an internal 813 bp restriction fragment. The first *Nco*I recognition sequence spans triplets encoding a Pro-Trp dipeptide sequence that is conserved in all eukaryotic GSs for which predicted sequences are available (see Figure 8). However, the second *Nco*I recognition sequence occurs in the 3′ noncoding region and thus would not be expected to be conserved in different GS genes. We have used this internal *Nco*I restriction fragment as a diagnostic marker in a preliminary screen of maize GS cDNA clones for different GS cDNAs. Thirteen cDNA clones identified by genetic selection and four cDNA clones identified by medium stringency (aqueous, 55°) colony hybridization have been examined to date. All yielded a restriction fragment indistinguishable from the 813 bp *Nco*I fragment of MGS1 by agarose gel electrophoresis upon digestion with *Nco*I (data not shown). The four clones identified by colony hybridization were isolated from the original BMS-3 cDNA library in LE392 F′ *lacI*^{q}. This demonstrates the stability of the GS cDNA over several generations and after transfer to different *E. coli* strains.

Synthesis of maize GS in *E. coli*: Western blot analyses carried out with antiserum prepared against purified tobacco GS and generously provided by B. HIREL have been used to demonstrate the synthesis of maize GS in *E. coli* cells harboring maize cDNAs (Figure 7). Two different maize GS cDNAs have been examined for synthesis of GS in *E. coli* strain FDB213 (Δ*glnA*). One, MGS1, contains both the 5′ mini-ORF and the major GS ORF. In MGS1, the major ORF is located 37 bp distal to the termination codon of the mini-ORF and is thus in a different reading frame (Figure 4). The second, MGS3, contains only the major GS ORF; its 5′ terminus is five bp 3′ from the TAG termination codon of the mini-ORF (nucleotide position −33; Figure 4). The pMGS3 clone was singled out for study because FDB213(pMGS3) transformants grew faster and produced larger colonies on M9 minimal proline ampicillin plates than did any of the other 54 maize GS cDNA clones analyzed up to that time. When MGS3 was subcloned in pUC119 in the antisense orientation so that single-stranded template could be prepared and the 5′ terminus sequenced (Figure 3), it became clear that the rapid growth of

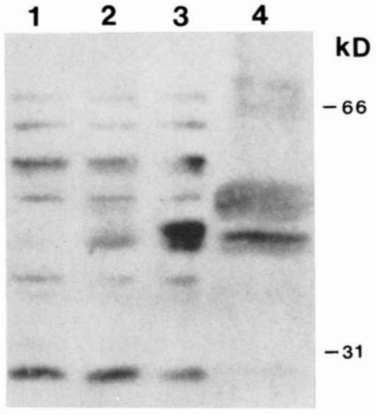

FIGURE 7.—Western blot analysis of the proteins present in extracts of FDB213(pUC119) (lane 1), FDB213(pMGS1) (lane 2), FDB213(pMGS3) (lane 3), and maize leaves (lane 4) using antiserum prepared against purified tobacco GS as probe. Size markers are bovine serum albumin, 66.2 kD, and bovine carbonic anhydrase, 31 kD. The two bands in lane 4 are presumably the result of binding of GS antibodies to the cytosolic and chloroplastic GSs present in maize leaves.

FDB213(pMGS3) and the large amount of GS produced in these cells (at least a large amount of anti-GS cross-reacting material) is the result of the absence of the mini-ORF and its TAG termination codon. The sequence analysis showed that the GS ORF in pMGS3 is joined in frame to the *lac* translation initiation signals with no intervening termination codons.

Although the antiserum used for the western blot analyses contains antibodies that react with several *E. coli* proteins, these can be identified in extracts of FDB213(pUC119) cells (Figure 7, lane 1). Western blots performed on extracts of FDB213(pMGS1) and FDB213(pMGS3) (Figure 7, lanes 2 and 3, respectively) revealed one or two additional proteins that reacted with the anti-GS serum. There was a light band (possibly two) for pMGS1 and two much darker bands for pMGS3, indicating that pMGS3 directs the synthesis of larger amounts of GS than does pMGS1. These results correlate well with the relative growth rates of FDB213(pMGS1) and FDB213(pMGS3) on M9 minimal plates containing proline but lacking glutamine.

The larger of the two new proteins detected in FDB213(pMGS3) with the anti-GS serum is intermediate in size between the two major cross-reacting species in maize leaves (Figure 7, lane 4), which presumably are the cytosolic and chloroplastic forms of maize GS (MCNALLY *et al.* 1983). The origin of the smaller protein detected in extracts of FDB213(pMGS3) is not clear. It may result from an unexpected secondary translation initiation event, a secondary termination event, or proteolytic cleavage of the *lac*-GS fusion protein.

Plasmid pMGS1-directed GS activity in *E. coli* is dependent on translation initiated at the *lacZ* AUG: Is the MGS1-directed maize GS activity responsible for the functional rescue of the Δ*glnA* defect in FDB213 dependent on translation initiated at the *lacZ* AUG or does the G-rich region of this cDNA 5′ (relative to the direction of transcription) to the GS ORF contain fortuitous ribosome-binding activity that permits translational initiation at the AUG of the GS ORF? This question was answered by using a reading-frame change to introduce a termination codon in the *lacZ*-proximal region of the poly-linker in pMGS1. Specifically, pMGS1 was cleaved with *Pst*I, the TGCA 3′ overhangs were removed with the 3′→5′ exonuclease activity of T4 DNA polymerase, and the blunt ends were rejoined with T4 DNA ligase. This procedure deletes four bp and produces an in-frame TAG termination codon near the *Xba*I recognition sequence of the poly-linker. FDB213 cells were then transformed with the ligation products, and transformants were identified by plating the cells on L medium containing ampicillin. Glutamine is growth-limiting for Gln^- *E. coli* in L medium. Thus on L plates, Gln^+ cells produce large colonies, and Gln^- cells produce small colonies. The vast majority of the colonies produced by FDB213 cells transformed with pMGS1 plasmids that had been subjected to *Pst*I cleavage, detailing, and ligation were small. When cells from 36 such colonies were replicated by sterile toothpick transfers sequentially onto M9 + proline medium and M9 + proline + glutamine medium, all were unable to grow in the absence of glutamine (data not shown). Thus, the pMGS1-directed GS activity in *E. coli* is dependent on translation initiated at the *lac* AUG. We believe that translation is initiated at the *lacZ* AUG and continues through the poly-linker, the 5′ leader region, and the mini-ORF. The AUG's of *lacZ* and the mini-ORF are in frame with no intervening terminators. Termination should then occur at the TAG of the mini-ORF. Since the AUG of the major GS ORF is close to the terminator (37 bp away), some ribosomes would be expected to reinitiate translation of the major ORF. Translational reinitiation of this type has been shown to occur at distances up to 102 bp in the *lacI* gene of *E. coli* (STEEGE 1977).

The introduction of a termination codon near the *Xba*I recognition site of the poly-linker places the first TAG termination codon at a position 200 bp away from the initiator AUG of the major ORF. Presumably, this would be too large a distance to permit enough translational reinitiation by ribosome shuttling to provide sufficient GS activity for functional rescue of FDB213. Some translational reinitiation might occur at the AUG of the mini-ORF 77 bp distal to the introduced terminator, but translation of GS from the major ORF would now require two such low-frequency translation reinitiation events. The latter would presumably not yield sufficient GS activity for functional rescue of FDB213.

Similarity of the predicted maize GS to the predicted products of other sequenced GS genes: The nucleotide sequence of the major ORF of MGS1 and MGS2 predicts that it encodes a primary translation product containing 382 amino acids and with a molecular weight of 42,896 (Figure 4). The predicted amino acid sequence of this maize GS exhibits striking similarity to the amino acid sequences of the predicted products of other eukaryotic GS genes (Figure 8). When compared with the predicted amino acid sequences of GS polypeptides that have been published to date, the maize GS sequence reported here is the most closely related to the chloroplastic GS of pea. Although the chloroplastic GS cDNA sequence of pea published by TINGEY, WALKER and CORUZZI (1987) is incomplete at the 5′ end, it predicts the sequence of 373 amino acids. The pea sequence shares 85% (317/373) amino acid identity with the corresponding segment of the predicted maize GS. Moreover, both are 15–16 amino acids longer at the carboxyl terminus than the other plant GS polypeptides, and 13 of these additional 15–16 amino acids are identical in the pea chloroplastic GS and the maize GS (Figure 8). In contrast, the maize GS has only 76% amino acid identity with the published cytosolic GS sequence of pea (TINGEY, WALKER and CORUZZI 1987) and only 76–77% amino acid identity with the published GS sequences of *Phaseolus vulgaris* (GEBHARDT *et al.* 1986), *Medicago sativa* (TISCHER, DASSARMA and GOODMAN 1986), and *Nicotiana plumbaginifolia* (TINGEY and CORUZZI 1987). Based on the alignment shown in Figure 8, the maize GS and the Chinese hamster GS sequence published by HAYWARD *et al.* (1986) have only 50% amino acid identity; however, there are several regions where the plant and the Chinese hamster GSs are highly conserved. When the plant GSs were compared with GSs from *E. coli* (COLOMBO and VILLAFRANCA, 1986) and *Anabaena* (TUMER, ROBINSON and HASELKORN 1983), only three short regions of similarity were detected, all within the carboxyl-terminal half of the molecule and all consisting of sequences of ten or less amino acids (Figure 8).

Genomic blot hybridizations: The results of blot hybridization of the MGS1 cDNA to *Eco*RI- or *Hin*dIII-digested Black Mexican Sweet genomic DNA showed that the entire MGS1 genomic sequence resides on one *Eco*RI restriction fragment and on two *Hin*dIII restriction fragments (Figure 9). Since the cDNA does not contain a *Hin*dIII recognition sequence, this sequence probably resides in one of several introns in the maize GS1 gene (B. M. CHERESKIN AND J. MESSING, unpublished results). It is also possible that the BMS cells studied contain a restriction site polymorphism in copies of the GS1 gene. Genomic blot hybridizations have also been performed on *Bam*HI-, *Bgl*II-, and *Eco*RI-digested genomic DNA isolated from leaf tissue of seedlings of maize lines BSSS-53, W23, W64A, B37, Ill-12E, A188 and W22, with the MGS1 cDNA as probe. Restriction fragment sizes on these blots were carefully calibrated by using the BRL 1-kb ladder size-marker system. Although restriction fragment length polymorphisms were evident, genomic DNAs from five lines were found to contain all of the sequences homologous to MGS1 on a single *Eco*RI restriction fragment (data not shown). The size of the fragment was ~5.5 kb in three lines and ~6.5 kb in two lines. The other two lines exhibited MGS1 homology on two *Eco*RI restriction fragments. Similar results were observed for *Bam*HI- and *Bgl*II-digested genomic DNAs. These results are most easily explained if MGS1 is specified by a single-copy gene.

DISCUSSION

The results of this study clearly demonstrate the feasibility of identifying eukaryotic cDNAs encoding proteins of interest by directly screening cDNA libraries for functional rescue of known mutational defects in *E. coli*. Such direct genetic selection exhibits several orders of magnitude greater resolving power than standard colony or plaque hybridization approaches. We thus predict that direct genetic selection will prove useful in isolating rare cDNAs from libraries. Given the vast repertoire of extensively characterized mutations in *E. coli* (BACHMAN 1987) and the large number of selectable phenotypes that they define (VINOPAL 1987), the approach outlined here might prove to have considerable applicability. WENKO, TREICK and WILSON (1985) screened a phage λ Charon 4a soybean genomic library for functional rescue of *E. coli* lysine auxotrophs using the "lytic complementation" procedure of DAVIS, BOTSTEIN and ROTH (1980). However, this approach would not be expected to work for plant genes containing introns. In contrast, the cDNA library screening procedure utilized in the present study should work irrespective of the presence of introns in the encoding genes. Clearly, this procedure would not be applicable to cDNAs encoding proteins that require post-translational processing events unique to eukaryotes. A major advantage of our approach is that it selects cDNA clones that carry complete or, at least, nearly complete coding sequences.

The two maize GS cDNAs sequenced in this study were found to carry a major GS-encoding ORF and a 5′ mini-ORF of 29 codons that is out-of-frame from the major ORF. The AUG of the mini-ORF has a G in the −3 position and a G in the +4 position, just like the AUG of the major ORF. Thus, the mini-ORF AUG is in strong context for translational initiation

Z. mays GS1	1	MAVSTGSTGV	VPRLEQLLNM	DTTPYTDKVI	AEYIWVGGSG	IDIRSKSRTI	SKPVEDPSEL	60
P. vulgaris pR-1	1	----------	MSL SD I L	NLSDT E	I	L L A L	PG KN	50
P. vulgaris pR-2	1	----------	MSL SD I L	NLSES E I		M L A L	PG D AK	50
P. sativum pGS341	1	----------	MSS SD I F	NLSDS E I		A L	PG S AK	50
P. sativum pGS197	1*	*********T	IN V D L	I F SI	G I T	V	SH V	51*
M. sativa GS	1	----------	MSL SD I L	DLSET E I	I	L L A L	PG T Q	50
N. plumbaginifolia pGS1	1	----------	MSL SD I L	NLSDS E I	I	M L A L	G T AK	50
C. griseus GS	1	M T AS HLN	KNIKQMY --	-CL QGE Q	M D T	EGL C T L	DCEPKCVE	57

Z. mays GS1	61	PKWNYDGSST	GQAPGEDSEV	ILYPQAIFKD	PFRGGNNVLV	ICDTYTPQGE	PLPTNKRHRA	120
P. vulgaris pR-1	51		Q	I	R I	A A	I N	110
P. vulgaris pR-2	51		D D		R I	V A	YD	110
P. sativum pGS341	51		N K		R I	V A	YN	110
P. sativum pGS197	52*				I	V A A	I	111*
M. sativa GS	51			I	R I	M A A	I A	110
N. plumbaginifolia pGS1	51				R I	M A A	I A	110
C. griseus GS	58	E F	F SE SN DM	Y S V M R	RDP K	F EVFKYNRK	AE L SC	117

Z. mays GS1	121	AQIFSDPKVG	EQVPWFGIEQ	EYTLLQKDVN	WPLGWPVGGF	PGPQGPYYCA	VGADKSFGRD	180
P. vulgaris pR-1	111	K N D V	AEE Y	E	V	G	A	170
P. vulgaris pR-2	111	K H D V	AE Y		L Y	G	AY	170
P. sativum pGS341	111	K H D A	AE Y	I	I Y	K G	I AY	170
P. sativum pGS197	112*	E N E	AEI Y	TN K	Y		A	171*
M. sativa GS	111	K H D V	AE Y	I		G	A A	170
N. plumbaginifolia pGS1	111	K N D V	AEE Y	R I	I	G	T A	170
C. griseus GS	118	KR -M M- S	N H M	MGT -G	H F SN	G	AY	174

Z. mays GS1	181	ISDAHYKACL	YAGINISGTN	GEVMPGQWEY	QVGPSVGIEA	GDHIWISRYI	LERITEQAGV	240
P. vulgaris pR-1	171	V V	I	F	A S	EL VA	V	230
P. vulgaris pR-2	171	V V	I	F	S	EV AA	L A	230
P. sativum pGS341	171	V	F I	F	S	E AA	I	230
P. sativum pGS197	172*	I		W		A		231*
M. sativa GS	171	V S	I	F	S	E VA	V	230
N. plumbaginifolia pGS1	171	V S Y	I	F	S	EV VA	A I	230
C. griseus GS	175	VE R	VK T	A A F	I CE RM	L VA F	H VC DF	234

Z. mays GS1	241	VLTLDPKPIQ	GDWNGAGCHT	NYSTKTMRED	GGFEEIKRAI	LNLSLRHDLH	ISAY----GE	296
P. vulgaris pR-1	231	SF K	A	N	Y S	QK GK KE	A ----	286
P. vulgaris pR-2	231	VSF P	A S	E	Y V K	EK G KE	A ---- K	286
P. sativum pGS341	231	VSF P	A A	F S N	Y V K	EK G KE	A ----	286
P. sativum pGS197	232*	E		S	V K	KI	E ----	287*
M. sativa GS	231	SF K	A	S	Y V LK	EK GKK KE	A ----	286
N. plumbaginifolia pGS1	231	VSF P	A	S	Y V LK	EK G K KE	A ----	286
C. griseus GS	235	IA F P	N	F A E	N LKH EE	EK K RY	R DPKG L	294
E. coli GS	253	TA FM MF	- S M C	HM LSKNGVN	LFAGDKYAGL	SEQA YYIGG	VIKHAKAINA	311
Anabaena 7120 GS	256	TV FM F	- S M C	HQ IWKDGKP	LFAGDQYAGL	SEMG YYIGG	LLKHAPALLA	314

Z. mays GS1	297	GNERRLTGKH	ETASIGTFSW	GVANRGCSIR	VGRDTEAKGK	GYLEDRRPAS	NMDPYIVTGL	356
P. vulgaris pR-1	287	R	D N L	A	KA	F	V SM	346
P. vulgaris pR-2	287	R	D N	S V	KQ	F	V SM	346
P. sativum pGS341	287		D NV	S	KD	F	V SM	346
P. sativum pGS197	288*		ND		KN		V A	347*
M. sativa GS	287	R	D N L	A	KA	F S	V SM	346
N. plumbaginifolia pGS1	287		N S K	A V	KA	F	V AM	346
C. griseus GS	295	D A G F	SN ND A	SA	IP TVGQEK	F SA	C FA EA	354
E. coli GS	318	NSYK VPGY	APVMLAY -	-AR SA	IPVVSSP -A	RRI V F DP	AAN LCFAA	374
Anabaena 7120 GS	321	NSYK VPGY	APVNLAY Q	-- SA	IPLSGTNPKA	KR F C DA	TSN LAFAA	378

Z. mays GS1	357	LAETTILWQP	SLEAEALAAK	KLALKV.				382
P. vulgaris pR-1	347	I D K	.					356
P. vulgaris pR-2	347	I K	.					356
P. sativum pGS341	347	I KK	P					357
P. sativum pGS197	348*	S L E	T Q	I				373*
M. sativa GS	347	I D K	.					356
N. plumbaginifolia pGS1	347	I D IGKS	.					356
C. griseus GS	355	IVR CL NET	GD PFQYKN.					373
E. coli GS	404	EAKEIPQVAG	- NEL	DLDREFLKAG	GVFTDEAIDA	YIALRREEDD	RVRMT-PHPV	461
Anabaena 7120 GS	408	ELAKVPSTPG	-L E L	ENDHAFLTDT	GVFTEDFIQN	WIDYKLANEV	KQMQLRPHPY	466

FIGURE 8.—Comparison of predicted amino acid sequences of glutamine synthetases encoded by sequenced cDNAs and genes. The complete amino acid sequence (single-letter code) is shown only for the predicted product of the maize glutamine synthetase cDNA GS1. The predicted amino acid sequences of the other glutamine synthetases are given on the lines below the maize GS1 sequence, showing only those amino acids that differ from the corresponding amino acids of the GS1 sequence except that all initiator methionine residues are shown. A dash indicates the absence of an amino acid; the translation termination codons are represented by periods. The *P. sativum* pGS197 cDNA is incomplete; the asterisks are used to represent sequence that is not available and as a reminder that this sequence is numbered from the most amino-terminal residue available, not from the methionine initiator residue. The *E. coli* and *Anabaena* sequences are shown only for those lines which contain segments of homology with the eukaryotic sequences. Sources of the predicted sequences are as follow: *Phaseolus vulgaris* (GEBHARDT *et al.* 1986), *Pisum sativum* (TINGEY, WALKER and CORUZZI 1987), *Medicago sativa* (TISCHER, DASSARMA and GOODMAN 1986), *Nicotiana plumbaginifolia* (TINGEY and CORUZZI 1987), *Cricetulus griseus* (HAYWARD *et al.* 1986), *Escherichia coli* (COLOMBO and VILLAFRANCA 1986), and *Anabaena* 7120 (TUMER, ROBINSON and HASELKORN 1983).

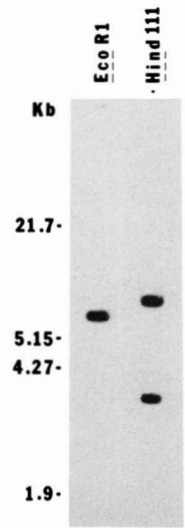

FIGURE 9.—Southern blot analysis of *Eco*RI and *Hin*dIII-digested genomic DNA of BMS culture cells hybridized to labeled MGS1 cDNA. The size markers shown are the positions of selected restriction fragments of an *Eco*RI, *Hin*dIII-double digest of phage λ DNA.

and should not be bypassed during ribosome scanning (KOZAK 1978, 1984, 1986a).

A possible explanation of the apparently anomalous structures of MGS1 and MGS2 is that the 5′ mini-ORF is an artifact of the cloning process. However, this is unlikely because our results require that the same artifact would have had to occur twice. MGS1 and MGS2 have different 3′ sequences subterminal to the poly(A) tails. Since cDNA synthesis was primed from T-tailed vector, MGS1 and MGS2 had to result from different mRNAs produced by the use of two distinct polyadenylation signals. Since both MGS1 and MGS2 carry the mini-ORF, duplicate identical events would have had to occur, one in each cloning sequence, for the mini-ORF to be a cloning artifact. This seems unlikely. Since neither MGS1 nor MGS2 contains an internal *Eco*RI cleavage sequence, it is difficult to envision a cloning event that might have generated artificial mini-ORFs of the type observed here.

Another possible explanation for the apparent violation of the maize GS cDNA sequence of the rule that eukaryotic mRNAs are monocistronic is that the mini-ORF has resulted from some genome rearrangement or other alteration that occurred during the growth of the BMS cells in culture. However, a genomic clone isolated from a lambda library constructed from maize inbred line A188 has now been partially sequenced, and has been shown to contain a sequence identical to that of the mini-ORF, the inter-ORF region, and the first part of the major ORF reported here for MGS1 and MGS2 (B. M. CHERESKIN and J. MESSING, unpublished results).

The results reported by MCNALLY *et al.* (1983) indicate that the GS activity in maize leaves is approximately 45% cytosolic GS_1 and 55% chloroplastic GS_2. The maize MGS1 and MGS2 cDNAs sequenced in this study predict a GS polypeptide which exhibits 75–76% amino acid identity to the putative cytosolic GSs of dicotyledonous plants and 85% amino acid identity to the predicted sequence of the pea chloroplastic GS (see legend to Figure 8 for references). This suggests that the maize cDNAs that we have studied encode the maize chloroplastic GS. Interestingly, when the amino acid sequences predicted by translating the mini-ORF through the inter-ORF sequence in all three reading frames (to allow the detection of possible effects of frameshift mutations that might have occurred during growth in culture) were examined for possible similarities to the conserved sequences in chloroplast transit peptides reported by KARLIN-NEUMANN and TOBIN (1986), no similarities were evident. However, the amino-terminal sequence predicted by the major ORF (**Met Ala** Val **Ser Thr Gly Ser Thr**) is similar to the conserved "box I" sequence identified by KARLIN-NEUMANN and TOBIN (*e.g.*, "box I" for ferredoxin = **Met Ala Ser Thr Leu Ser Thr**). Whether this is significant or coincidental remains to be determined.

All 19 maize cDNAs analyzed to date appear to be derived from transcripts of the same gene. Perhaps the maize genome contains only a single gene, and the cytosolic and chloroplastic forms of GS result from differential processing of the transcript or primary translation product. However, at this point, it must be considered more likely that one or more additional GS genes are present. Our genomic Southern blot analyses have provided no evidence for a small gene family encoding three or four distinct GSs as has been observed in dicotyledonous plants (DONN *et al.* 1984; CULLIMORE *et al.* 1984; GEBHARDT *et al.* 1986; TINGEY, WALKER and CORUZZI 1987). Presumably, one or more additional GS genes are also present in the maize genome. If so, these other GS sequences must be quite divergent and cross-hybridize with the MGS1 probe too weakly to yield convincing bands on intermediate level strigency genomic Southern blots. The characterization of maize GS cDNAs isolated from libraries recently constructed from leaf, stem, and root mRNAs should lead to the identification of additional GS genes in maize if present.

As DASSARMA, TISCHER and GOODMAN (1986) have pointed out, the expression of plant genes in *E. coli* provides obvious advantages for genetic manipulations such as mutational dissections, selection for resistance to inhibitors of the gene-products, and other genetic modifications prior to their reintroduction into plant cells. It will be interesting to see how frequently plant cDNAs produce active proteins in *E. coli.*

The most surprising result of this study was the

discovery of a 29-codon mini-ORF 5′ to and out-of-frame from the major ORF of cDNAs MGS1 and MGS2. The AUG of the mini-ORF is in strong context for translational initiation with G's in the −3 and +4 positions (KOZAK 1984, 1986a) suggesting that the mini-ORF should be translated. If it is, translation of the major ORF encoding GS would require translational reinitiation after movement of the ribosome through an inter-ORF region of 37 nucleotides. Several similar examples are known in animals and animal viruses [see KOZAK (1986b) for a list of examples and references]. The function of such upstream mini-ORFs is not known; however, KOZAK (1986b) has speculated that they exist "for regulatory purposes." Based on studies of GS in prokaryotes (reviewed by REITZER and MAGASANIK 1987), one certainly expects the synthesis of GS in eukaryotes to be highly regulated. However, in prokaryotes this regulation is transcriptional. Might translational regulation be superimposed on or replace transcriptional regulation of GS synthesis in higher plants?

Most of the recent literature refers to the functional rescue of mutants by cDNAs or genes as "complementation." We have resisted using this terminology because in the strict sense genetic complementation occurs when two mutant alleles or otherwise defective genetic elements each provide a component that is missing or defective in the other so as to "complement one another" and yield an active or partially active product (when acting together). The functional rescue of mutants by cDNAs is really a dominance interaction, with the wild-type allele carried by the cDNA being dominant to the resident mutant allele of the host cell or organism. It is analogous to the dominant expression of a wild-type allele when present with a mutant allele in the common protoplasm of any *trans*-heterozygote. Since dominance and complementation have always had distinct meanings in genetics, we suggest that it is not appropriate to call a dominance interaction of this type "complementation."

This research was supported by grant DE-FG05-85ER13367 from the U.S. Department of Energy (to J. M.), by U.S. Public Health Service grant GM-35636 from the National Institutes of Health (to J. M.), and by a Seed Money Grant from the University of Minnesota Plant Molecular Genetics Institute (to D. P. S.). J. P. H. was supported by grants DE-AC02-83ER13071 from the U.S. Department of Energy and U.S. Public Health Service grant GM-24756 from the National Institutes of Health (both to IRWIN RUBENSTEIN). The authors thank J. KIRIHARA, P. DAS, J. VIEIRA, S. LUDWIG, D. OPPENHEIMER, P. HUSSEY, S. KOPCZAK, S. PLOENSE AND S. DOTSON for helpful discussions during the course of this study.

LITERATURE CITED

BACHMANN, B. J., 1987 Linkage map of *Escherichia coli* K-12, edition 7. pp. 807–876. In: *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*, Edited by F. C. NEIDHARDT, J. L. INGRAHAM, K. B. LOW, B. MAGASANIK, M. SCHAECHTER AND H. E. UMBARGER. American Society for Microbiology, Washington, D.C.

BARR, P. J., R. M. THAYER, P. LAYBOURN, R. C. NAJARIAN, F. SEELA and D. R. TOLAN, 1986 7-Deaza-2′-deoxyguanosine-5′-triphosphate: enhanced resolution of M13 dideoxy sequencing. BioTechniques **4:** 428–432.

BAZZICALUPO, M., R. FANI, E. GALLORI, L. TURBANTI and M. POLSINELLI, 1987 Cloning of the histidine, pyrimidine and cysteine genes of *Azospirillum brasilense*: expression of pyrimidine and three clustered histidine genes in *Escherichia coli*. Mol. Gen. Genet. **206:** 76–80.

BENTON, W. D., and R. W. DAVIS, 1977 Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196:** 180–182.

BENZER, S., 1957 The elementary units of heredity. pp. 70–93. In: *The Chemical Basis of Heredity*, Edited by W. D. MCELROY AND B. GLASS. The Johns Hopkins Press, Baltimore.

BONDARYK, R. P., and H. PAULUS, 1985 Expression of the gene for *Bacillus subtilis* aspartokinase II in *Escherichia coli*. J. Biol. Chem. **260:** 592–597.

CARLSON, J., and J. MESSING, 1984 Efficiency in cloning and sequencing using the single-stranded bacteriophage M13. J. Biotechnol. **1:** 253–264.

COLOMBO, G., and J. J. VILLAFRANCA, 1986 Amino acid sequence of *Escherichia coli* glutamine synthetase deduced from the DNA nucleotide sequence. J. Biol. Chem. **261:** 10587–10591.

CULLIMORE, J. V., C. GEBHARDT, R. SAARELAINEN, B. J. MIFLIN, K. B. IDLER and R. F. BARKER, 1984 Glutamine synthetase of *Phaseolus vulgaris* L: organ-specific expression of a multigene family. J. Mol. Appl. Genet. **2:** 589–599.

DALE, R. M. K., B. A. MCCLURE and J. P. HOUCHINS, 1985 A rapid single stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13:** 31–40.

DASSARMA, S., E. TISCHER and H. M. GOODMAN, 1986 Plant glutamine synthetase complements a *glnA* mutation in *E. coli*. Science **232:** 1242–1244.

DAVIS, R. W., D. BOTSTEIN and J. R. ROTH, 1980 *Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

DE BRUIJN, F. J., and F. M. AUSUBEL, 1981 The cloning and transposon Tn5 mutagenesis of the *glnA* region of *Klebsiella pneumoniae*: identification of *glnR*, a gene involved in the regulation of the *nif* and *hut* operons. Mol. Gen. Genet. **183:** 289–297.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DONN, G., E. TISCHER, J. A. SMITH and H. M. GOODMAN, 1984 Herbicide-resistant alfalfa cells: an example of gene amplification in plants. J. Mol. Appl. Genet. **2:** 285–300.

GEBHARDT, C., J. E. OLIVER, B. G. FORDE, R. SAARELAINEN and B. J. MIFLIN, 1986 Primary structure and differential expression of glutamine synthetase genes in nodules, roots and leaves of *Phaseolus vulgaris*. EMBO J. **5:** 1429–1435.

GOUGH, J. A., and N. E. MURRAY, 1983 Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. **166:** 1–19.

GREEN, C. E., 1977 Prospects for crop improvement in the field of cell culture. Hortic. Sci. **12:** 131–134.

GRUNSTEIN, M., and D. S. HOGNESS, 1975 Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72:** 3961–3965.

HANAHAN, D., 1983 Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:** 557–580.

HANAHAN, D., and M. MESELSON, 1980 Plasmid screening at high colony density. Gene **10:** 63–67.

HAYWARD, B. E., A. HUSSAIN, R. H. WILSON, A. LYONS, V. WOODCOCK, B. MCINTOSH and T. J. R. HARRIS, 1986 The cloning and nucleotide sequence of cDNA for an amplified glutamine synthetase gene from the Chinese hamster. Nucleic Acids Res. **14:** 999–1008.

HEIDECKER, G., and J. MESSING, 1983 Sequence analysis of zein cDNAs obtained by an efficient mRNA cloning method. Nucleic Acids Res. **11:** 4891–4906.

HELFMAN, D. M., J. R. FERAMISCO, J. C. FIDDES, G. P. THOMAS and S. H. HUGHES, 1983 Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. Proc. Natl. Acad. Sci. USA **80:** 31–35.

KARLIN-NEUMANN, G. A., and E. M. TOBIN, 1986 Transit peptides of nuclear-encoded chloroplast proteins share a common amino acid framework. EMBO J. **5:** 9–13.

KIRIHARA, J. A., J. P. HUNSPERGER, W. C. MAHONEY and J. W. MESSING, 1988 Differential expression of a gene for a methionine-rich storage protein in maize. Mol. Gen. Genet. **211:** 477–484.

KOZAK, M., 1978 How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell **15:** 1109–1123.

KOZAK, M., 1984 Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12:** 857–872.

KOZAK, M., 1986a Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44:** 283–292.

KOZAK, M., 1986b Bifunctional messenger RNAs in eukaryotes. Cell **47:** 481–483.

LIZARDI, P. M., and A. ENGELBERG, 1979 Rapid isolation of RNA using proteinase K and sodium perchlorate. Anal. Biochem. **98:** 116–122.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MARTIN, R., 1987 Overcoming DNA sequencing artifacts: stops and compressions. FOCUS **9:** 8–10.

MCMULLEN, M. D., B. HUNTER, R. L. PHILLIPS and I. RUBENSTEIN, 1986 The structure of the maize ribosomal DNA spacer region. Nucleic Acids Res. **14:** 4953–4968.

MCNALLY, S. F., B. HIREL, P. GADAL, A. F. MANN and G. R. STEWART, 1983 Glutamine synthetases of higher plants, evidence for a specific isoform content related to their possible physiological role and their compartmentation within the leaf. Plant Physiol. **72:** 22–25.

MESSING, J., 1983 New M13 vectors for cloning. Methods Enzymol. **101:** 20–78.

MIFLIN, B. J., and P. J. LEA, 1980 Ammonia assimilation. pp. 169–202. In: *The Biochemistry of Plants,* Vol. 5, Edited by B. J. MIFLIN. Academic Press, New York.

MILLER, J. H., 1972 *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MOZER, T. J., 1980 Partial purification and characterization of the mRNA for a-amylase from barley aleurone layers. Plant Physiol. **65:** 834–837.

MURASHIGE, T., and F. SKOOG, 1962 A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant **15:** 473–497.

OKAYAMA, H., and P. BERG, 1982 High efficiency cloning of full-length cDNA. Mol. Cell. Biol. **2:** 161–170.

REED, K. C., and D. A. MANN, 1985 Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13:** 7207–7221.

REITZER, L. J., and B. MAGASANIK, 1987 Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, L-alanine. and D-alanine, pp. 302–320. In: *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology,* Edited by F. C. NEIDHARDT, J. L. INGRAHAM, K. B. LOW, B. MAGASANIK, M. SCHAECHTER and H. E. UMBARGER. American Society for Microbiology, Washington, D.C.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SHERIDAN, W. J., 1975 Growth of corn cells in culture. J. Cell Biol. **67:** 396a.

SHERIDAN, W. F., 1982 Black Mexican sweet corn: its use for tissue cultures. pp. 385–388. In: *Maize for Biological Research,* Edited by W. F. SHERIDAN. Plant Molecular Biology Association, Charlottesville, Va.

SHURE, M., S. WESSLER and N. FEDOROFF, 1983 Molecular identification and isolation of the *waxy* locus in maize. Cell **35:** 225–233.

SOUTHERN, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

STEEGE, D. A. 1977 5′-Terminal nucleotide sequence of *Escherichia coli* lactose repressor mRNA: features of translational initiation and reinitiation sites. Proc. Natl. Acad. Sci. USA **74:** 4163–4167.

STEWART, G. R., A. F. MANN and P. A. FENTEM, 1980 Enzymes of glutamate formation: glutamate dehydrogenase, glutamine synthetase, and glutamate synthase. pp. 271–327. In: *The Biochemistry of Plants.* Vol. 5, Edited by B. J. MIFLIN. Academic Press, New York.

TABOR, S., and C. C. RICHARDSON, 1987 DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84:** 4767–4771.

TINGEY, S. V., and G. M. CORUZZI, 1987 Glutamine synthetase of *Nicotiana plumbaginifolia,* cloning and *in vivo* expression. Plant Physiol. **84:** 366–373.

TINGEY, S. V., E. L. WALKER and G. M. CORUZZI, 1987 Glutamine synthetase genes of pea encode distinct polypeptides which are differentially expressed in leaves, roots and nodules. EMBO J. **6:** 1–9.

TISCHER, E., S. DASSARMA and H. M. GOODMAN, 1986 Nucleotide sequence of an alfalfa glutamine synthetase gene. Mol. Gen. Genet. **203:** 221–229.

TOWBIN, H., T. STAEHELIN and J. GORDON, 1979 Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76:** 4350–4354.

TUMER, N. E., S. J. ROBINSON and R. HASELKORN, 1983 Different promoters for the *Anabaena* glutamine synthetase gene during growth using molecular or fixed nitrogen. Nature **306:** 337–342.

VIEIRA, J., and J. MESSING, 1987 Production of single-stranded plasmid DNA. Methods Enzymol. **153:** 3–11.

VINOPAL, R. L., 1987 Selectable phenotypes. pp. 990–1015. In: *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology,* Edited by F. C. NEIDHARDT, J. L. INGRAHAM, K. B. LOW, B. MAGASANIK, M. SCHAECHTER and H. E. UMBARGER. American Society for Microbiology, Washington, D.C.

WALKER, J. E., and N. J. GAY, 1983 Analysis of *Escherichia coli* ATP synthase subunits by DNA and protein sequencing. Methods Enzymol. **97:** 195–218.

WEINSTOCK, R., R. SWEET, M. WEISS, H. CEDAR and R. AXEL, 1978 Intragenic DNA spacers interrupt the ovalbumin gene. Proc. Natl. Acad. Sci. USA **75:** 1299–1303.

WENKO, L. K., R. W. TREICK and K. G. WILSON, 1985 Isolation and characterization of a gene encoding meso-diaminopimelate dehydrogenase from *Glycine max.* Plant Mol. Biol. **4:** 197–204.

Communicating editor: W. F. SHERIDAN