MONSANTO

Food - Health - Hope

59364

AG New Products Div.
Plant Science Technology

FINAL REPORT

Report No.: MSL-12738

Job/Project No.: 09-023-760.39-393773

Date: 10/9/93

Title: Purification, Cloning, and Characterization of a Highly Glyphosate-tolerant 5-Enolpyruvylshikimate-3-phosphate Synthase from *Agrobacterium sp.* strain CP4

Authors:
Stephen R. Padgette, Gerard F. Barry, Diane B. Re, David A. Eichholtz, Marcia Weldon, Kathryn Kolacz, and Ganesh M. Kishore

Abstract:
5-Enolpyruvyl-shikimate-3-phosphate synthase (EPSPS), an enzyme of the shikimate pathway for aromatic amino acid biosynthesis in plants and microorganisms, is the biological target enzyme of glyphosate, the active ingredient of Roundup® herbicide. Expression *in planta* of glyphosate-tolerant EPSPSs has proven to be an effective mechanism for conferring glyphosate-tolerance to crop plants. We now wish to report the purification, cloning, and expression in *E. coli* of EPSPS from *Agrobacterium* sp. strain CP4 (CP4 EPSPS). Based on steady-state kinetic analysis, CP4 EPSPS exhibits very high-level glyphosate tolerance (appK_i[glyphosate] = 2.7 mM), while retaining a very low appK_m(PEP) (12 μM), comparable to that of wild-type plant EPSPSs. CP4 EPSPS has the highest appK_i(glyphosate) / appK_m(PEP) ratio, 227, of any EPSPS described to date, while the appK_m(S3P) is approximately 1 μM. CP4 EPSPS has approximately 50-60% similarity to previously described EPSPSs, and numerous active site residues are conserved relative to other EPSPSs. The kinetic data collected supports the use of the CP4 EPSPS gene obtained herein for the development of glyphosate-tolerant crops.

[redacted]

[redacted]

DISTRIBUTION LIST

Copy Number		Abstract Only
1-3	Reports Library/R2C	G. M. Kishore/BB3B
4	B. G. Hammond/C2SE	R. J. Stonard/GG3A
5	S. Padgette/AA3I	R. B. Horsch/GG4J
6	[redacted]	E. Debus/BB4D
7	Reg. Aff. Archives/D. Re/BB3A	F. Serdy/AA2E
8	D. Nida/AA3I	S. G. Rogers/AA2E
9	N. Biest/AA3I	D. Bertrand, C3SH
10	T. Nickson/GG4C	B. Wells, C3NK
11	S. Waters/Bxl 5040	K. Marshall/C2SB
12	M. Kasai/MJL 5265	X. Delannay/GG6A
		C. Morley/BB4D
13	G. Barry/AA3E	M. D. Stowers/BB4D
14	D. Eichholtz/AA3I	G. Keyes/BB4D
15	M. Weldon/AA3E	B. Ziegler/BB4D
16	K. Kolacz/AA3I	R. Fraley/O3E
17	S. Brown/AA2G	S. Berberich/GG4F
18	P. Sanders/AA2G	J. Ream/GG4K
		P. Keck/GG4K
		P. Lavrik/GG4C
		A. Reed/GG4F
		D. Hoerner/BB4D

[redacted]

[redacted]

[redacted]

[redacted]

Table of Contents

List of Tables

List of Figures

List of Appendices

I. Introduction

5-Enolpyruvylshikimate-3-phosphate synthase (EPSP[1] synthase, EPSPS) catalyzes the reversible reaction of phosphoenolpyruvate (PEP) and shikimate-3-phosphate (S3P) to form EPSP and inorganic phosphate (1). EPSPS is found in bacteria, plants, and fungi (but not mammals) as a component of the shikimate pathway of aromatic amino acid biosynthesis. Identification of EPSPS as the *in vivo* enzyme target of glyphosate (N-phosphonomethylglycine) (2), the active ingredient of Roundup® herbicide, has elicited extensive investigations into the enzymology and molecular biology of EPSPS (3-5). One practical use for EPSPS genes is in the development of glyphosate-tolerant crops (5-10). The availability of crops tolerant to environmentally compatible non-selective herbicides, such as Roundup®, will offer farmers new weed control solutions in an era where it is increasingly difficult to discover and develop selective herbicides which meet current use criteria.

There are two main mechanisms to confer herbicide tolerance to a crop using genetic engineering techniques (9). First, in the target-site modification approach, a gene is expressed *in planta* to produce a herbicide-insensitive target site, or, alternatively, a herbicide-sensitive target may be overexpressed (7). In the second mechanism, the metabolic inactivation approach, a gene is expressed in the crop which encodes a protein which converts the herbicide to an inactive metabolite. To confer glyphosate-tolerance to a crop plant using the target-site modification approach, the expression of glyphosate-tolerant EPSPSs *in planta* has proven to indeed yield crops with a glyphosate-tolerant phenotype (5-10).

The interaction of an inhibitor-tolerant enzyme with its natural substrates is an important aspect of that enzyme's ability to maintain a high catalytic activity in the presence of the inhibitor. For EPSPS, glyphosate is a competitive inhibitor versus phosphoenolpyruvate (PEP) (11). Extensive mutagenesis studies at the active site of petunia EPSPS demonstrated that the binding sites for PEP and glyphosate overlap (12). Therefore, one key aspect to consider when optimizing EPSPS genes for use to confer glyphosate tolerance to crops is whether a highly glyphosate-tolerant EPSPS maintains an interaction with PEP comparable to that of the wild-type plant EPSPS. This is because, if, in a glyphosate-tolerant crop which has been treated with glyphosate, PEP levels are low intracellularly (near the K_m[PEP] of wild-type

---

[1] The abbreviations used are EPSPS, 5-enolpyruvylshikimate-3-phosphate synthase; CP4 EPSPS, EPSPS from *Agrobacterium sp.* strain CP4; PEP, phosphoenolpyruvate; S3P, shikimate-3-phosphate; RT, Roundup® tolerant; DTT, dithiothreitol; BAM, benzamidine HCl; SDS-PAGE, sodium dodecyl sulfate - polyacrylamide gel electrophoresis; MES, 2-[N-morpholino]ethanesulfonic acid; MOPS, 3-[N-morpholino]propanesulfonic acid; IPTG, isopropylthiogalactoside; TBST, tris-buffer saline tween; BSA, bovine serum albumin.

petunia EPSPS), then a glyphosate-tolerant EPSPS with a severely elevated $appK_m$(PEP) couldn't maintain an EPSPS reaction rate comparable to the wild-type crop with no glyphosate treatment (assuming comparable expression levels). In this case, one would predict that less than complete tolerance (plant damage) would be exhibited upon glyphosate treatment.

As shown in Table 1, the $appK_m$(PEP) for several glyphosate-tolerant EPSPSs is elevated relative to wild-type EPSPS, indicating perturbation of PEP interactions accompanying reduced glyphosate binding. The $appK_m$(PEP) for the glyphosate-tolerant petunia EPSPS G101A variant is 210 μM, which is elevated significantly relative to the $appK_m$(PEP) of 5 μM for the wild-type petunia EPSPS enzyme. Plants expressing the petunia EPSPS G101A variant were tolerant to glyphosate, but commercial-level tolerance could not be obtained, presumably due to the elevated $appK_m$(PEP).

One measure of the selectivity of the PEP/glyphosate site of an EPSPS is the $appK_i$(glyphosate) / $appK_m$(PEP) ratio. The higher the value of this constant, the better the EPSPS will bind PEP relative to glyphosate. For instance, the wild-type petunia EPSPS, which is inhibited by glyphosate, has a $appK_i$(glyphosate) / $appK_m$(PEP) ratio of 0.08, while the glyphosate-tolerant G101A petunia EPSPS variant has a ratio of 12 (Table 1). While extensive mutagenesis of the petunia EPSPS gene, both random and site-directed, has yielded glyphosate-tolerant EPSPSs with $appK_m$(PEP)'s in the 40-60 μM range (Table 1) (13), no variants have been identified which have had $appK_m$(PEP)'s in the 5-15 μM range while also exhibiting millimolar $appK_i$(glyphosate) constants. Identification of such an EPSPS, with an $appK_i$(glyphosate) / $appK_m$(PEP) ratio in the range of 100 and a low $appK_m$(PEP), has been a goal of our laboratory for use in developing glyphosate-tolerant crops.

Previous studies have shown that EPSPS enzymes from different sources vary widely with respect to their degree of sensitivity to inhibition by glyphosate (14,15). The degree of sensitivity showed no correlation with any genus or species tested. An insensitivity to glyphosate inhibition of the activity of the EPSPS from the *Pseudomonas sp.* PG2982 has also been reported, but with no details of the studies (16,17). In general, while such natural tolerance has been reported, there is no report suggesting the kinetic superiority of the naturally occurring bacterial glyphosate-tolerant EPSPS enzymes over those of mutated EPSPS enzymes, nor have any of the genes heretofore been characterized. Similarly, there are no reports on the expression of naturally glyphosate-tolerant EPSPS enzymes *in planta* to confer glyphosate tolerance. Recently, an EPSPS from the cyanobacterium *Anabaena variabilis* was reported to have an $appK_m$(PEP) of 34 μM and an $appK_i$(glyphosate) of 350 μM, giving an $appK_i$(glyphosate) / $appK_m$(PEP) ratio of 10.3 (18,19), which is comparable to ratios of the petunia EPSPS variants

listed in Table 1.

We now wish to report the purification, cloning, and characterization of the EPSPS from *Agrobacterium* sp. strain CP4 (CP4 EPSPS) (20). CP4 EPSPS is the most highly glyphosate-tolerant EPSPS reported to date, with a appK_i(glyphosate) of 2.7 mM, and the enzyme exhibits an appK_m(PEP) of 12 μM, which is very similar to the value for the appK_m(PEP) for the wild-type petunia EPSPS, 5 μM (Table 1). EPSPS from *Agrobacterium* sp. strain CP4 thus represents a unique opportunity to access an EPSPS gene with near-optimal characteristics for glyphosate-tolerant crop production, with an additional opportunity to gain further information on the PEP/glyphosate selectivity of the EPSPS active site.

II. Experimental Procedures

General methods - Microbial cultures containing EPSPS genes were grown, extracted, desalted, and analyzed for protein content (21) as previously described (12). EPSPS was assayed by HPLC radioassay or phosphate release assay as previously described at 25°C and pH 7.0, in the presence of 1 mM PEP, 2 mM S3P, and 50 mM HEPES (22). EPSPS specific activities are reported as μmol EPSP/min mg protein (U/mg). PEP was from Boehringer, and [^{14}C]-PEP was from Amersham. S3P was synthesized as previously described (23). All protein purification procedures were carried out at 3-5°C. N-terminal amino acid sequencing was performed by loading samples onto a Polybrene precycled filter in aliquots while drying. Automated Edman degradation chemistry was used to determine the N-terminal protein sequence, using an Applied Biosystems Model 470A gas phase sequencer (24) with an Applied Biosystems 120A PTH analyzer. Unless otherwise indicated, SDS-PAGE was performed using the Pharmacia Phast-gel System according to manufacturer's instructions. Cloning and genetic techniques, unless otherwise indicated, are generally as previously described (25). DNA sequences were determined using both the Sequenase kit from IBI (International Biotechnologies Inc.) and the T7 sequencing / Deaza Kit from Pharmacia.

Buffers - Extraction buffer, 100 mM TrisCl, 1 mM EDTA, 1 mM BAM, 5 mM DTT, 10% glycerol, pH 7.5); Q Sepharose buffer, 10 mM TrisCl, 25 mM KCl, 5 mM DTT, 10% glycerol, pH 7.8; Phenyl Superose buffer, 100 mM TrisCl, 5 mM DTT, 1 M ammonium sulfate, 10% glycerol, pH 7.5; EPSPS storage buffer, 50 mM TrisCl, 50 mM KCl, 2 mM DTT, 50% glycerol, pH 7.5; Solution I, 50 mM Glucose, 10 mM EDTA, 25 mM TrisCl pH 8.0; TE, 10 mM Tris pH 8.0; 1.0 mM EDTA; 10X ligation buffer, 250 mM Tris-HCl, pH 8.0, 100 mM $MgCl_2$, 100 mM dithiothreitol, 2 mM spermidine; cracking buffer, 60 mM Tris-HCl 6.8, 1% SDS, 1% 2-mercaptoethanol, 10% glycerol, 0.01% bromophenol blue. For all peptide purifications, 0.1% trifluoroacetic acid (TFA, Pierce) was designated buffer "RP-A" and 0.1% TFA in acetonitrile was buffer "RP-B".

Isolation of Agrobacterium *sp.strain CP4* - The *Agrobacterium* sp. strain CP4 was initially identified by its ability to grow on glyphosate as a carbon source (10 mM) in the presence of 1 mM phosphate. The strain CP4 was identified from a collection obtained from a fixed-bed immobilized cell column that employed Mannville R-635 diatomaceous earth beads. The column had been run for three months on a waste-water feed from a glyphosate production plant. The column contained 50 mg/ml glyphosate and NH_3 as NH_4Cl. Total organic carbon was 300 mg/ml and BOD's (Biological Oxygen Demand - a measure of "soft" carbon availability) were less than 30 mg/ml. This treatment column has been described (26). Dworkin-Foster minimal salts medium containing glyphosate at 10 mM and with phosphate at 1 mM was used to select for microbes from a wash of this column that were capable of growing on glyphosate as sole carbon source. Dworkin-Foster minimal medium was made up by combining 1 ml each of A, B and C and 10 ml of D (see below) and thiamine HCl (5 mg), and diluting to 1 liter with autoclaved water:

A. D-F Salts (1000X stock; per 100 ml; autoclaved):	
H_3BO_3	1 mg
$MnSO_4.7H_2O$	1 mg
$ZnSO_4.7H_2O$	12.5 mg
$CuSO_4.5H_2O$	8 mg
$NaMoO_3.3H_2O$	1.7 mg
B. $FeSO_4.7H_20$ (1000X stock; per 100 ml; autoclaved)	0.1 g
C. $MgSO_4.7H_2O$ (1000X stock; per 100 ml; autoclaved)	20 g
D. $(NH_4)_2SO_4$ (100X stock; per 100 ml; autoclaved)	20 g

Yeast Extract (YE; Difco) was added to a final concentration of 0.01 or 0.001%. The strain CP4 was also grown on media composed of D-F salts, amended as described above, containing glucose, gluconate and citrate (each at 0.1 %) as carbon sources and with inorganic phosphate (0.2 - 1.0 mM) as the phosphorous source.

Purification of EPSPS from Agrobacterium sp. *strain CP4* - Five 10-litre fermentations were carried out on a spontaneous "smooth" isolate of strain CP4 that displayed less clumping when grown in liquid culture (CP4-S1). The cells from the three batches showing the highest EPSPS specific activities (0.006 - 0.010 U/mg) were pooled. Cell paste of *Agrobacterium* sp. strain CP4 (300 g) was washed with 0.5 L of 0.9% saline (0.5 L, twice) and collected by centrifugation (30 minutes, 8000 rpm in a GS3 Sorvall rotor). The cell pellet was suspended in 0.9 L extraction buffer and lysed by 2 passes through a Manton Gaulin cell. The resulting solution was centrifuged (30 minutes, 8000 rpm) and the supernatant was treated with 1.5% protamine sulfate (in 100 mM TrisCl, pH 7.5, 0.2% w/v final protamine sulfate concentration). After stirring (1 h), the mixture was centrifuged (50 minutes, 8000 rpm) and the

resulting supernatant treated with solid ammonium sulfate to 40% saturation and stirred for 1 h. After centrifugation (50 minutes, 8000 rpm), the resulting supernatant was treated with solid ammonium sulfate to 70% saturation, stirred for (50 min), and the insoluble protein was collected by centrifugation (1 h, 8000 rpm). This 40-70% ammonium sulfate fraction was then dissolved in extraction buffer to give a final volume of 0.2 L, and dialyzed twice (Spectrum 10,000 MW cutoff dialysis tubing) against 2 L of extraction buffer (12 h total).

To the resulting dialyzed 40-70% ammonium sulfate fraction (0.29 L) was added solid ammonium sulfate to give a final concentration of 1 M ammonium sulfate. This material was loaded (2 ml/min) onto a column (5 cm x 15 cm, 295 ml) packed with phenyl Sepharose CL-4B (Pharmacia) resin equilibrated with extraction buffer containing 1 M ammonium sulfate, and washed with the same buffer (1.5 L, 2 ml/min). EPSPS was eluted with a linear gradient of extraction buffer going from 1 M to 0.00 M ammonium sulfate (total volume of 1.5 L, 2 ml/min). Fractions were collected (20 ml), assayed for EPSPS activity, and the fractions with the highest EPSPS activity (fractions 36-50) were pooled and dialyzed against 3 x 2 L (18 h) of 10 mM TrisCl, 25 mM KCl, 1 mM EDTA, 5 mM DTT, 10% glycerol, pH 7.8.

The dialyzed EPSPS extract (350 ml) was loaded (5 ml/min) onto a column (2.4 cm x 30 cm, 136 ml) packed with Q-Sepharose Fast Flow (Pharmacia) resin equilibrated with Q Sepharose buffer, and washed with 1 L of the same buffer. EPSPS was eluted with a linear gradient of Q Sepharose buffer going from 0.025 M to 0.40 M KCl (total volume of 1.4 L, 5 ml/min). Fractions were collected (15 ml), assayed for EPSPS activity, and the fractions with the highest EPSPS activity (fractions 47-60) were pooled and the protein was precipitated by adding solid ammonium sulfate to 80% saturation and stirring (1 h). The precipitated protein was collected by centrifugation (20 minutes, 12000 rpm in a GSA Sorvall rotor), dissolved in Q Sepharose buffer (total volume of 14 ml), and dialyzed against the same buffer (2 x 1 L, 18 h).

The resulting dialyzed partially purified EPSPS extract (19 ml) was loaded (1.7 ml/min) onto a Mono Q 10/10 column (Pharmacia) equilibrated with Q Sepharose buffer, and washed with the same buffer (35 ml). EPSPS was eluted with a linear gradient of 0.025 M to 0.35 M KCl in Q Sepharose buffer (total volume of 119 ml, 1.7 ml/min). Fractions were collected (1.7 ml), assayed for EPSPS activity, and the fractions with the highest EPSPS activity (fractions 30-37) were pooled (6 ml).

The Mono Q pool was made 1 M in ammonium sulfate by the addition of solid ammonium sulfate and 2 ml aliquots were chromatographed in three batches on a Phenyl Superose 5/5 column (Pharmacia) equilibrated with Phenyl Superose buffer. Samples were loaded (1 ml/min), washed with Phenyl Superose buffer (10 ml), and eluted with a linear gradient of Phenyl Superose

buffer going from 1 M to 0.00 M ammonium sulfate (total volume of 60 ml, 1 ml/min). Fractions were collected (1 ml), assayed for EPSPS activity, and the fractions from the three runs with the highest EPSPS activity (Run 1: F37-40; Run 2: F37-40; Run 3, F36-39) were pooled together (10 ml, 2.5 mg protein). A portion (300 µl) of F39 from Run 1 was taken before pooling, dialyzed against 50 mM $NaHCO_3$ (2 x 1 L), and the resulting pure EPSPS sample (0.9 ml, 77 µg protein) was analyzed by N-terminal amino acid sequencing.

The remaining Phenyl Superose EPSPS pool was dialyzed against 50 mM TrisCl, 2 mM DTT, 10 mM KCl, 10% glycerol, pH 7.5 (2 x 1 L). An aliquot (0.55 ml, 0.61 mg protein) was loaded (1 ml/min) onto a Mono Q 5/5 column (Pharmacia) equilibrated with Q Sepharose buffer, washed with the same buffer (5 ml), and eluted with a linear gradient of Q Sepharose buffer going from 0-0.14 M KCl in 10 minutes, then holding at 0.14 M KCl (1 ml/min). Fractions were collected (1 ml) and assayed for EPSPS activity by the phosphate release assay and were subjected to SDS-PAGE (10-15%, Phast System, Pharmacia, with silver staining) to determine protein purity. Several fractions exhibiting a single band of protein by SDS-PAGE (fractions 22-25, 222 µg) were pooled and dialyzed against 100 mM ammonium bicarbonate, pH 8.1 (2 x 1 L, 9 h). Table 2 gives the purification table for CP4 EPSPS.

Trypsinolysis of Agrobacterium *sp. strain CP4 EPSPS* - To the resulting pure *Agrobacterium sp.* strain CP4 EPSPS in 100 mM ammonium bicarbonate (111 µg) was added 3 µg of trypsin (Calbiochem), and the trypsinolysis reaction was allowed to proceed for 16 h at 37°C. The tryptic digest was then chromatographed (1 ml/min) on a C18 reverse phase HPLC column (Vydac) as previously described (22). The gradient used for elution of the *Agrobacterium sp.* CP4 EPSPS tryptic peptides was: 0-8 minutes, 0% RP-B; 8-28 minutes, 0-15% RP-B; 28-40 minutes, 15-21% RP-B; 40-68 minutes, 21-49% RP-B; 68-72 minutes, 49-75% RP-B; 72-74 minutes, 75-100% RP-B. Fractions were collected (1 ml) and, based on the elution profile at 210 nm, at least 70 distinct peptides were produced from the trypsinized EPSPS. Fractions 40-70 were evaporated to dryness and redissolved in 150 µl each of 10% acetonitrile, 0.1% trifluoroacetic acid.

The CP4 EPSPS fraction 61 peptide was further purified on the C18 column by the gradient: 0-5 minutes, 0% RP-B; 5-10 minutes, 0-38% RP-B; 10-30 minutes, 38-45% B. Fractions were collected based on the UV signal at 210 nm. A large peptide peak in fraction 24 eluted at 42% RP-B and was dried down, resuspended as described above, and rechromatographed on the C18 column with the gradient: 0-5 minutes, 0% RP-B; 5-12 min, 0-38% RP-B; 12-15 min, 38-39% RP-B; 15-18 minutes, 39% RP-B; 18-20 minutes, 39-41% RP-B; 20-24 minutes, 41% RP-B; 24-28 minutes, 42% RP-B. The peptide in fraction 25, eluting at 41% RP-B and designated peptide 61-24-25, was subjected to N-terminal amino acid sequencing.

The CP4 EPSPS fraction 53 tryptic peptide was further purified by C18 HPLC by the gradient 0% RP-B (5 minutes), 0-30% RP-B (5-17 minutes), 30-40% RP-B (17-37 minutes). The peptide in fraction 28, eluting at 34% RP-B and designated peptide 53-28, was subjected to N-terminal amino acid sequencing.

The CP4 EPSPS fraction 41 tryptic peptide was further purified by C18 HPLC by the gradient 0% RP-B (3 minutes), 0-20% RP-B (3-10 minutes), 20-30% RP-B (10-30 minutes). The peptide in fraction 25, eluting at 35% RP-B was dried down, resuspended as described above, and rechromatographed on the C18 column with the gradient 0% RP-B (3 minutes), 0-20 % RP-B (7 minutes), 20-30 % RP-B (20 minutes). The peptide in fraction 27, eluting at 35% RP-B and designated peptide 41-25-27, was subjected to N-terminal amino acid sequencing.

Cloning of the Agrobacterium *sp. strain CP4 EPSPS gene in* E. coli - Chromosomal DNA was prepared from strain *Agrobacterium sp.* strain CP4 by resuspending the cell pellet from a 200 ml L-Broth (27), late log phase culture of *Agrobacterium sp.* strain CP4 in 10 ml of Solution I (see "Buffers" above") (28). SDS was added to a final concentration of 1% and the suspension was subjected to three freeze-thaw cycles, each consisting of immersion in dry ice for 15 minutes and in water at 70°C for 10 minutes. The lysate was then extracted four times with equal volumes of phenol:chloroform (1:1; phenol saturated with TE) and the phases separated by centrifugation (15000 g; 10 minutes). The ethanol-precipitable material was pelleted from the supernatant by brief centrifugation (8000 g; 5 minutes) following addition of two volumes of ethanol. The pellet was resuspended in 5 ml TE and dialyzed for 16 h at 4°C against 2 liters TE. This preparation yielded a 5 ml DNA solution of 552 µg/ml. Partially-restricted DNA was prepared by treating three 100 µg aliquot samples of CP4 DNA for 1 h at 37°C with restriction endonuclease HindIII at rates of 4, 2 and 1 enzyme unit/µg DNA, respectively. The DNA samples were pooled, made 0.25 mM with EDTA and extracted with an equal volume of phenol:chloroform. Following the addition of sodium acetate and ethanol, the DNA was precipitated with two volumes of ethanol and pelleted by centrifugation (12000 g; 10 minutes). The dried DNA pellet was resuspended in 500 µl TE and layered on a 10-40% Sucrose gradient (in 5% increments of 5.5 ml each) in 0.5 M NaCl, 50 mM Tris pH 8.0, 5 mM EDTA. Following centrifugation for 20 h at 26,000 rpm in a SW28 rotor, the tubes were punctured and ~1.5 ml fractions collected. Samples (20 µl) of each second fraction were run on 0.7% agarose gel and the size of the DNA determined by comparison with linearized lambda DNA and HindIII-digested lambda DNA standards. Fractions containing DNA of 25-35 kb fragments were pooled, desalted on Amicon10 columns (7000 rpm; 20°C; 45 minutes) and concentrated by precipitation. This procedure yielded 15 µg of CP4 DNA of the required size. A cosmid bank was then constructed using the vector

pMON17020, which is based on the pBR327 replicon and contains the spectinomycin/ streptomycin (Spr;spc) resistance gene from Tn7 (29), the chloramphenicol resistance gene (Cmr;cat) from Tn9 (30), the gene10 promoter region from phage T7 (31), and the 1.6 kb BglII phage lambda cos fragment from pHC79 (32). A number of cloning sites are located downstream of the cat gene.

The vector was then cut with HindIII and treated with calf alkaline phosphatase (CAP) in preparation for cloning. Vector and target sequences were ligated by combining vector DNA (HindIII/CAP) (3 μg), size fractionated CP4 HindIII fragments (1.5 μg), 10X ligation buffer (2.2 μl), T4 DNA ligase (New England Biolabs) (400 U/μl) (1.0 μl), and diluted with H_2O to a volume of 22.0 μl. This mixture was incubated for 18 h at 16°C. The ligated DNA (5 μl) was packaged into lambda phage particles (Stratagene, Gigapack Gold) using the manufacturer's procedure.

A sample (200 μl) of *E. coli* HB101 (33) containing the T7 polymerase expression plasmid pGP1-2 (34) and grown overnight in L-Broth (with maltose at 0.2% and kanamycin at 50 μg/ml) was infected with the packaged DNA (50 μl). Transformants were selected at 30°C on M9 (27) agar containing kanamycin (50 μg/ml), chloramphenicol (25 μg/ml), L-proline (50 μg/ml), L-leucine (50 μg/ml) and B1 (5 μg/ml), and containing glyphosate at a concentration of 3.0 mM. Aliquot samples were also plated on the same media lacking glyphosate to titer the packaged cosmids. Cosmid transformants were isolated on this latter medium at a rate of ~5 x 10^5 per μg CP4 HindIII DNA after 3 days at 30°C. Colonies arose on the glyphosate agar from day 3 until day 15 with a final rate of ~1 per 200 cosmids. DNA was prepared from 14 glyphosate tolerant clones and, following verification of this phenotype, was transformed into *E. coli* GB100/pGP1-2 (*E. coli* GB100 is an aroA derivative of MM294 (35)) and tested for complementation for growth in the absence of added aromatic amino acids and aminobenzoic acids. Other aroA strains such as SR481 (36,37) can also be used for this experiment. This aroA strain usually requires that growth media be supplemented with L-phenylalanine, L-tyrosine and L-tryptophan each at 100 μg/ml and with para–hydroxybenzoic acid, 2,3–dihydroxybenzoic acid and para-aminobenzoic acid each at 5 μg/ml for growth in minimal media. Of the fourteen cosmids tested only one showed complementation of the aroA- phenotype. Transformants of this cosmid, pMON17076, showed weak but uniform growth on the unsupplemented minimal media after 10 days.

The proteins encoded by the cosmids were determined *in vivo* using a T7 expression system (34). Cultures of *E. coli* containing pGP1-2 (34) and test and control cosmids were grown at 30°C in L-broth (2 ml) with chloramphenicol and kanamycin (25 and 50 μg/ml, respectively) to a Klett reading of ~ 50. An aliquot was removed and the cells collected by centrifugation, washed with M9

salts (27) and resuspended in 1 ml M9 medium containing glucose at 0.2%, thiamine at 20 µg/ml and containing the 18 amino acids at 0.01% (minus cysteine and methionine). Following incubation at 30°C for 90 minutes, the cultures were transferred to a 42°C water bath and held there for 15 minutes. Rifampicin (Sigma) was added to 200 µg/ml and the cultures held at 42°C for 10 additional minutes and then transferred to 30°C for 20 minutes. Samples were pulsed with 10 µCi of ^{35}S-methionine for 5 minutes at 30°C. The cells were collected by centrifugation and suspended in 60-120 µl cracking buffer. Aliquot samples were electrophoresed on 12.5% SDS-PAGE and following soaking for 60 minutes in 10 volumes of acetic acid-methanol–water (10:30:60), the gel was soaked in ENLIGHTNING™ (DuPont) following manufacturer's directions, dried, and exposed at -70°C to X-ray film. Proteins of about 45 kd in size, labeled with 35S-methionine, were detected in number of the cosmids, including pMON17076.

Oligonucleotide probes - In order to verify the CP4 EPSPS cosmid clone, a number of oligonucleotide probes were designed on the basis of the sequence of two of the tryptic sequences from the CP4 enzyme (Table 3). The probe identified as MID was very low degeneracy and was used for initial screening. The probes identified as EDV-C and EDV-T were based on the same amino acid sequences and differ in one position (underlined in Table 3) and were used as confirmatory probes, with a positive to be expected only from one of these two probes. For the oligonucleotides in Table 3, alternate acceptable nucleotides at a particular position are designated by a "/" such as A/C/T. Probes were labeled using gamma-^{32}P-ATP and polynucleotide kinase, and DNA from fourteen of the cosmids described above was restricted with EcoRI, transferred to membrane and probed with the oligonucleotide probes. The conditions used were as follows: prehybridization was carried out in 6X SSC, 10X Denhardt's for 2-18 h periods at 60°C, and hybridization was for 48-72 h in 6X SSC, 10X Denhardt's, 100 µg/ml tRNA at 10°C below the Td for the probe. The Td of the probe was approximated by the formula 2°C x (A+T) + 4°C x (G+C). The filters were then washed three times with 6X SSC for ten minutes each at room temperature, dried and autoradiographed. Using the MID probe, an ~9.9 kb fragment in the pMON17076 cosmid gave the only positive signal. This cosmid DNA was then probed with the EDV-C and EDV-T probes separately; again, the ~9.9 kb band gave a signal, but only with the EDV-T probe.

Localization and subcloning of the CP4 EPSPS gene - The CP4 EPSPS gene was further localized by performing additional Southern analyses on different restriction digests of pMON17076 using the MID and EDV-T probes separately. Based on these analyses and on subsequent detailed restriction mapping of the pBlueScript (Stratagene) subclones of the ~9.9 kb fragment from pMON17076, a 3.8 kb EcoRI-SalI fragment was identified to which both probes hybridized. This analysis also showed that the MID and EDV-T probes

hybridized to different sides of BamHI, ClaI, and SacII sites. This 3.8 kb fragment was cloned in both orientations in pBlueScript to form pMON17081 and pMON17082. The phenotypes imparted to *E. coli* by these clones were then determined. Glyphosate tolerance was determined following transformation into *E. coli* MM294 containing pGP1-2 (pBlueScript also contains a T7 promoter) on M9 agar media containing glyphosate at 3 mM. Both pMON17081 and pMON17082 showed glyphosate tolerant colonies at three days at 30°C at about half the size of the controls on the same media lacking glyphosate. This result suggested that the 3.8 kb fragment contained an intact EPSPS gene. The apparent lack of orientation-dependence of this phenotype could be explained by the presence of the T7 promoter at one side of the cloning sites and the lac promoter at the other. The aroA phenotype was determined in transformants of *E. coli* GB100 on M9 agar media lacking aromatic supplements. In this experiment, carried out with and without the Plac inducer IPTG, pMON17082 showed much greater growth than pMON17081, suggesting that the EPSPS gene was expressed from the SalI site towards the EcoRI site.

Nucleotide sequencing - Nucleotide sequencing was begun from a number of restriction site ends, including the BamHI site discussed above. Sequences encoding protein sequences that closely matched the N-terminus protein sequence and that for the tryptic fragment 53-28 (the basis of the EDV-T probe) were localized to the SalI side of this BamHI site. These data provided conclusive evidence for the cloning of the CP4 EPSPS gene and for the direction of transcription of this gene. These data coupled with the restriction mapping data also indicated that the complete gene was located on an ~2.3 kb XhoI fragment and this fragment was subcloned into pBlueScript. The nucleotide sequence of almost 2 kb of this fragment was determined by a combination of sequencing from cloned restriction fragments and by the use of specific primers to extend the sequence.

E. coli *expression of CP4 EPSPS in pMON17101* - By a series of site directed mutageneses, carried out by the procedures of Kunkel et al. (38), essentially as described in Sambrook *et al.* (39), BglII and NcoI sites were placed at the N-terminus with the fMet contained within the NcoI recognition sequence, the first internal NcoI site was removed (the second internal NcoI site was removed later), and a SacI site was placed after the stop codons. At a later stage the internal NotI site was also removed by site-directed mutagenesis. The CP4 EPSPS gene was then cloned as a NcoI-BamHI N-terminal fragment plus a BamHI-SacI C-terminal fragment into a PrecA-gene10L expression vector similar to those described (40,41) to form pMON17101. The nucleotide sequences of the primers used for the site-directed mutageneses are given in Table 4. Note that this cloning scheme changed the second amino acid of CP4 EPSPS from serine to alanine.

E. coli *expression of CP4 EPSPS in pMON21104* - It was also of interest to express in *E. coli* the CP4 EPSPS gene with an N-terminus identical to that of the CP4 EPSPS gene cloned into glyphosate-tolerant crops. In the plant expression vectors, the NcoI site, containing the initiating ATG, was mutagenized to an SphI site (42). This resulted in a change of the second amino acid of the CP4 EPSPS sequence from an serine to a leucine. In order to clone the CP4 sequence with the second amino acid being leucine into an *E. coli* expression vector, it was necessary to mutagenize the SphI site to an NdeI site by site directed mutagenesis; the CP4 EPSPS gene chosen as the starting material was a CP4 EPSPS gene with plant-preferred codons, which is described elsewhere[2]. The primer used to convert the second amino acid of the plant-preferred CP4 EPSPS gene to leucine from alanine is given in Table 4.

Western blots - Antibodies were raised in goat to CP4 EPSPS purified from pMON17101 and petunia EPSPS purified from pMON342 (37). SDS-PAGE was run using Daiichi 10 cm x 10 cm 4-20% gradient gels, run at 25 mA/gel in Enprotech mini gel rigs (approx. 3 hour running time). For Western blots, proteins were transferred to Immobilon PVDF membrane overnight (16.5 hours) at 0.1 Amps (1.6 Amp hours total) in a Hoeffer TE Series Transphor unit. Blots were blocked with 1% BSA in TBST for 30 min. To detect EPSPS, blots were incubated with either anti-CP4 EPSPS antibody (1:1000 dilution, goat DR2 Bleed 4) or anti-petunia EPSPS antibody (1:1000 goat DR3 Bleed 7) for 1 hour and washed 3 x 5 min with TBST. The blots were then probed with ^{125}I Protein G in 0.5% BSA in TBST for 45 min and subsequently washed with TBST with 0.1% Triton (4 x 10 min). The blots were allowed to dry and exposed to film overnight. For Coomassie blue staining, the gel was stained overnight and destained in 30% MeOH, 10% acetic acid.

Sequence comparisons - The sequence comparisons were performed using the software from UWCGC, Version 7 (43). Using the program "FastA", the most homologous protein sequences to the CP4 EPSPS protein sequence were found from the SwissProt data base. The twelve most homologous EPSPS sequences found (*B. subtilis*, accession number P20691; *Y. enterocolitica*, P19688; *E. coli*, P07638; *S. typhimurium*, P19786; *S. gallinarum*, P22299; *S. typhimurium*, P07637; *P. hybrida*, P11043; *M. tuberculosis*, P22487; *B. napus*, P17688; *L. esculentum*, P10748; *A. thaliana*, P05466; *B. pertussis*, P12421) were compared using the program "Pileup". The resulting consensus EPSPS sequence was then compared to CP4 EPSPS. Paired comparisons of EPSPS sequences were performed with the program "Gap" (43).

III. Results and Discussion

Purification of EPSPS from Agrobacterium *sp. strain CP4* - EPSPS from *Agrobacterium sp.* strain CP4 was purified using ammonium sulfate

---

[2] G. Barry et *al.*, Monsanto, manuscript in progress.

fractionation and Phenyl Sepharose, Q-Sepharose, MonoQ 10/10, Phenyl Superose 5/5, and MonoQ 5/5 chromatography steps. As shown in Table 2, the overall yield was 6.9%, the final specific activity was 11.8 U/mg, and the purification achieved was 2360-fold. Upon N-terminal sequencing of the purified EPSPS in fraction 39 of the first Phenyl Superose run (shown in Figure 1A), a single N-terminal amino acid sequence was obtained, with the sequence: XH(G)ASSRPATARKS-S(G)LX(G)(T)V(R)IPG(D)(K)(M). Amino acids in parentheses represent probable but not certain identification. Upon pooling the most pure fractions from the three Phenyl Superose runs (see Experimental Procedures), it was clear by SDS-PAGE with silver staining that the Phenyl Superose pool was not homogeneous (Figure 1B). After a subsequent MonoQ 5/5 column chromatography step, several fractions were obtained which appeared to be homogeneous CP4 EPSPS (Figure 1C). As shown in Figure 1, the purified CP4 EPSPS migrated on SDS-PAGE with a molecular weight of approximately 45-46 kD kilodaltons, which is typical of EPSPS enzymes from *E. coli* (44), as well as petunia and tomato (37,45).

Trypsinolysis of purified CP4 EPSPS - In order to obtain peptide sequences of potential use for synthesis of DNA probes for gene cloning, the purified CP4 EPSPS from the MonoQ 5/5 pool was trypsinized, and several peptides were purified and subjected to N-terminal sequencing analysis. Amino acid sequences were successfully obtained for three tryptic peptides, designated peptide 61-24-25, peptide 53-28, and peptide 41-25-27. The amino sequences obtained from the tryptic peptides are shown in Table 3.

Cloning and sequencing of the CP4 EPSPS gene - Two parallel approaches were undertaken to clone the EPSPS gene from *Agrobacterium sp.* strain CP4: cloning based on the expected phenotype for a glyphosate tolerant EPSPS, and purification of the enzyme to provide material to raise antibodies and to obtain amino acid sequences from the protein to facilitate the verification of clones. The cloning strategy selected was introduction of a cosmid DNA bank of *Agrobacterium sp.* strain CP4 into *E. coli* and selection for the EPSPS gene by selection for growth on inhibitory concentrations of glyphosate.

The Agrobacterium sp. strain CP4 cosmid bank was constructed using DNA fragments of 25-35 kb and the vector pMON17020 (see Experimental Procedures). The packaged cosmid DNA transformants were cloned into *E. coli* HB101, and selected on 3.0 mM glyphosate. Fourteen glyphosate-tolerant clones were selected, and the DNA from these clones were transformed into an AroA- strain of *E. coli*, GB100. Selection of functional CP4 EPSPS clones on media lacking aromatic amino acids yielded one glyphosate-tolerant clone, designated pMON17076. Cultures of *E. coli* containing pGP1-2 (34) and pMON17076, when pulse-labelled with ^{35}S-methionine and examined by SDS-PAGE with autoradiography, were positive for the expression of proteins in the 45 kD range. Using the MID probe synthesized based on the amino acid

sequence of peptide 61-24-25 (Table 3), a ~9.9 kb fragment in the pMON17076 cosmid gave the only positive signal. This cosmid DNA was then probed with the EDV-C and EDV-T probes separately (Table 3); again, the ~9.9 kb band gave a signal, but only with the EDV-T probe. Thus, based on the glyphosate-tolerant phenotype, the complementation of the *E. coli* aroA- phenotype, the expression of a ~45 Kd protein, and the hybridization to two probes derived from the CP4 EPSPS amino acid sequence, the data strongly suggested that the pMON17076 cosmid contained the EPSPS gene.

As described in Experimental Procedures, the CP4 EPSPS gene was localized to a 3.8 kb EcoRI-SalI fragment, which was cloned in both orientations in pBlueScript to form pMON17081 and pMON17082. Nucleotide sequencing was begun from a number of restriction site ends, and the data indicated that the complete gene was located on an ~2.3 kb XhoI fragment and this fragment was subcloned into pBlueScript. The nucleotide sequence of almost 2 kb of this fragment was determined by a combination of sequencing from cloned restriction fragments and by the use of specific primers to extend the sequence. The nucleotide sequence of the CP4 EPSPS gene is shown in Figure 2, and the deduced amino acid sequence is shown in Figure 3. The calculated molecular mass of CP4 EPSPS, based on the gene cloned into pMON17101 (Appendix A) is 47,572, and consists of 455 amino acid residues. The amino acid sequences corresponding to peptides 61-24-25, 53-28, and 41-25-27, as well as the N-terminus, were located and are underlined in Figure 3. The identification in the gene of the four amino acid sequences obtained directly from the purified enzymatically active CP4 EPSPS conclusively demonstrates that the gene cloned was CP4 EPSPS.

E. coli expression of CP4 EPSPS in pMON17101 - In order to express the CP4 EPSPS gene in *E. coli*, the gene was cloned as a NcoI-BamHI N-terminal fragment plus a BamHI-SacI C-terminal fragment into a PrecA-gene10L expression vector similar to those described (40,41) to form pMON17101. The nucleotide sequences of the primers used for the site-directed mutageneses are given in Table 4 (see Experimental Procedures for details). The plasmid map of pMON17101 is shown in Figure 4. We note that the EPSPS N-terminal amino acid sequence was altered by the introduction of the NcoI site, from Met-ser-his-gly... in the gene as cloned from *Agrobacterium* (shown in Figure 3) to Met-ala-his-gly... in the gene as cloned into pMON17101. pMON17101 was transformed in *E. coli* JM101, cultures were induced with nalidixic acid (40), and extracted and assayed for EPSPS activity.

Extracts of *E. coli* JM101: pMON17101 yielded EPSPS specific activities of 0.96 U/mg with no glyphosate present in the assay mixture, and a specific activity of 0.97 U/mg in the presence of 0.5 mM glyphosate. In contrast, *E. coli* JM101 (with no CP4 EPSPS vector present) extracts exhibited a specific activity of 0.006 U/mg, and no activity remained in the presence of 0.5 mM

glyphosate. These results clearly indicate that the CP4 EPSPS gene in pMON17101 encodes the functional glyphosate-tolerant EPSPS from *Agrobacterium* sp. strain CP4. CP4 EPSPS expressed in pMON17101 was purified to homogeneity in an analogous method to that described below for the pMON21104 CP4 EPSPS, and antibodies were raised against the enzyme in goat. Based on isoelectric focusing gels, the pI of CP4 EPSPS purified from pMON17101 was 4.1-4.3 (data not shown), compared to a calculated pI of 4.99 (Appendix A).

E. coli expression of CP4 EPSPS in pMON21104 - It was also of interest to express in *E. coli* the CP4 EPSPS gene with an N-terminus identical to that of the CP4 EPSPS gene cloned into glyphosate-tolerant crops. In the plant expression vectors, the NcoI site, containing the initiating ATG, was mutagenized to an SphI site (42). This resulted in a change of the second amino acid of the "mature" CP4 EPSPS sequence from an serine in the native gene to a leucine. In order to clone the CP4 sequence with the second amino acid being leucine into an *E. coli* expression vector, it was necessary to mutagenize the SphI site to an NdeI site by site directed mutagenesis; the CP4 EPSPS gene chosen as the starting material was the "plant-preferred" CP4 EPSPS gene. The oligonucleotide primer used to insert the NdeI site is shown in Table 4. The resulting *E. coli* expression vector, pMON21104, is shown in Figure 5, and was transformed into *E. coli* GB100. Shown in Figure 6 is the nucleic acid comparison of the CP4 EPSPS genes in pMON17101 and pMON21104. As noted previously, the only amino acid sequence difference between the two genes is at the second position, which is alanine and leucine in pMON17101 and 21104, respectively; the DNA sequences are different due to the pMON21104 vector containing the "plant-preferred" CP4 EPSPS gene. The specific activity of CP4 EPSPS expressed from pMON21104 was found to be 2.3 U/mg in small-scale shake flasks. A purification scheme was developed to purify CP4 EPSPS from the pMON21104 *E. coli* culture, and the large-scale fermentation and purification of CP4 EPSPS from *E. coli* GB100:pMON21104 is described elsewhere (46,47).

Relationship to other EPSPSs - The deduced amino acid sequences from CP4 EPSPS and a number of other EPSPS enzymes were compared using the "Gap" computer program provided in the UWGCG package (43), and the results are shown in Table 5. Comparing the deduced amino acid sequences of CP4 EPSPS with EPSPS from soybean, corn, petunia, *E. coli, B. subtilis,* and *S. cerevisiae* (Baker's yeast) yields similarities of 51.2%, 48.5%, 50.1%, 52.2%, 59.3%, and 53.5%, respectively, and identities of 26.0%, 24.1%, 23.3%, 26.0%, 41.1%, and 29.9%, respectively. The listings of the sequences and alignments are given as Appendices A-Q of this report. As shown in Table 5, the amino acid sequence homology between CP4 EPSPS and EPSPSs typically present in plant and bacterial food sources (48.5-59.3% similar, 23.3-41.1% identical) is comparable to the homology between the EPSPSs from soybean and *B.*

subtilis (55.6% similar, 30.1% identical). *B. subtilis* is used in food production and is considered GRAS (generally regarded as safe) (48)), and soybean protein (including soybean EPSPS) is widely consumed by animals and humans. This indicates that there is considerable divergence in the EPSPSs which are typically present in foods, and that the divergence of the CP4 EPSPS sequence from typical food EPSPS sequences is of the same order. As also shown in Table 5, there is a higher degree of relatedness between soybean and corn EPSPS (90.3% similar, 83.1% identical), and soybean and *E. coli* EPSPS (71.8% similar, and 54.8% identical). In any case, CP4 EPSPS is no more different from food EPSPSs than some food EPSPSs are different from each other.

Another important result which supports the high degree of relatedness between CP4 EPSPS and other EPSPSs is the recently solved 3-dimensional X-ray crystal structure of CP4 EPSPS[3]. This data shows that, on a molecular level, the CP4 EPSPS has the same overall folding patterns as the *E. coli* EPSPS. Since the *E.coli* EPSPS is more similar in amino acid sequence to plant EPSPSs than is CP4 EPSPS, the structural similarity of CP4 EPSPS to the *E. coli* EPSPS allows the inference that the CP4 EPSPS is structurally similar to plant EPSPSs.

Antibody reactivity - CP4 EPSPS was also compared to a plant EPSPS (petunia EPSPS) by Western blot analysis. Two separate Western blots containing both CP4 EPSPS and petunia EPSPS were probed with anti-CP4 EPSPS and anti-petunia EPSPS antibodies, respectively. As shown in Figure 7, at similar protein loadings the goat antibody to CP4 EPSPS recognized the CP4 EPSPS standard, but not the petunia EPSPS (Figure 7A), and the goat antibody to petunia EPSPS recognized petunia EPSPS but not CP4 EPSPS (Figure 7B). Shown in Figure 7C is a Coomassie blue-stained SDS-PAGE gel indicating that both samples used in Figures 7A and B were pure and migrate with similar molecular weights. These results indicate that the antibodies developed distinguish between the plant EPSPS and CP4 EPSPS under the conditions utilized, likely due to the differences in amino acid sequence between the two proteins.

Active site homology - Although the overall amino acid similarity between CP4 EPSPS and other published EPSPSs may be lower than the similarity between plant EPSPSs, the CP4 EPSPS still retains an active site similar to other well-studied EPSPS enzymes. This was determined by aligning the amino acid sequence of CP4 EPSPS against a consensus sequence of previously-identified EPSPSs (Figure 8). Previous analyses of EPSPS sequences have noted the high degree of conservation of sequences of the enzymes and the almost invariance of sequences in two regions - the "20-35" and "95-107" regions (45)

---

[3] D. Niedhart, W. Stallings, Monsanto unpublished results.

(numbered according to the petunia EPSPS sequence) - and these regions are less conserved in the case of CP4 when compared to other bacterial and plant EPSPS sequences. However, several residues which have been identified as important for EPSPS function are conserved in CP4 EPSPS, as shown in Table 6. The residues include Lys^{28}, which corresponds to Lys^{22} in *E. coli* EPSPS and likely interacts with PEP (49); Arg^{33}, which corresponds to Arg^{27} of *E. coli* EPSPS and is involved in S3P binding (50); and Arg^{128}, which corresponds to Arg^{124} of *E. coli* EPSPS, and is believed to be involved in PEP binding (50). One quite interesting feature of the CP4 EPSPS primary structure is the presence of an alanine at position 100 corresponding to Gly^{96} of *E. coli* EPSPS (and Gly^{101} of petunia EPSPS). This glycine to alanine substitution has been well-characterized as imparting glyphosate-tolerance to glyphosate-sensitive EPSPSs (12). It appears that CP4 EPSPS has naturally compensated for the increase in appK_m(PEP) which in all other known examples accompanies the glycine to alanine substitution. The CP4 EPSPS must have subtle changes in the active site residues which sets it apart from the other known EPSPS, with respect to PEP and glyphosate interactions.

Comparison of microbial and plant-expressed CP4 EPSPS N-terminal sequences- Shown in Table 7 is a summary of the N-terminal amino acid sequences of CP4 EPSPS as deduced from the gene sequence and originally purified from 1) *Agrobacterium sp.* strain CP4, 2) pMON17101, and 3) pMON21104, as well as the from glyphosate-tolerant soybean and glyphosate-tolerant canola seeds, versus the deduced "mature" CP4 EPSPS sequence from the plant expression vector. It can be seen that the N-terminal sequences from purified pMON17101 and pMON21104 proteins verify the changes made at the second position by site-directed mutagenesis. It is also clear from the results obtained that the initiating methionine is absent from the original *Agrobacterium*-expressed protein (ser at the second position) and the *E. coli*-expressed pMON17101 protein (ala at the second position), but present in the *E. coli*-expressed pMON21104 protein (leu at the second position). This result indicates that the endogenous *E. coli* N-terminal aminopeptidase is not able to cleave off the N-terminal methionine when leu is at the second position. Table 7 also shows that the N-terminal sequence obtained from the plant-expressed CP4 EPSPS matches that of the pMON21104-expressed protein, with the exception of the initiating methionine on pMON21104 EPSPS.

Steady-state kinetic characterization of purified CP4 EPSPS - Initial measurement of the appK_m(PEP) and appK_i(glyphosate) versus PEP for CP4 EPSPS were performed in desalted crude extracts of *Agrobacterium sp.* strain CP4, using the HPLC radioassay method. Those initial experiments (data not shown) indicated that the appK_m(PEP) was approximately 9 μM ([PEP]=1.5-40 μM range), and the appK_i(glyphosate) was approximately 1.8 mM ([glyphosate] = 0, 0.5, 1, and 3 mM). Upon purification of the CP4 EPSPS from *Agrobacterium sp.* strain CP4 (see Experimental Procedures), several kinetic

constants were redetermined using the purified enzyme. As shown in Figure 9, the inhibition of CP4 EPSPs by glyphosate was competitive versus PEP, as expected, and the appK_i(glyphosate) was found to be 2.7 mM. As summarized in Table 8, the appK_m(PEP) of the purified CP4 EPSPS was found to be 12 μM, giving an appK_i(glyphosate)/appK_m(PEP) ratio of 227, which is the highest ratio reported to date. As expected, based on previous studies with glyphosate-tolerant EPSPSs (12), there was no increase in appK_m(S3P) accompanying the very high appK_i(glyphosate); in fact, the appK_m(S3P) of 1 μM is very low for an EPSPS enzyme, which typically exhibit values around 7 μM (12). As discussed previously, the specific activity of purified CP4 EPSPS from *Agrobacterium* was observed to be 11.8 U/mg. Preparations of CP4 EPSPS expressed in *E. coli* (from pMON17101 or 21104) often yielded specific activities in the range of 2-7 U/mg. Further investigation revealed that the CP4 EPSPS activity was stimulated by salt, such as 50 mM KCl, which gave a 2-2.6 fold activation of the enzyme (data not shown). Using purified pMON21104 CP4 EPSPS, dialyzed into storage buffer and assayed in the presence of 1 mM KCl (carried over from storage buffer), the specific activity was determined to be 5.5 U/mg. We have not ruled out some partial inactivation taking place in the large-scale *E. coli*-expressed CP4 EPSPS purifications. Comparing CP4 EPSPS to the petunia G101A variant (Table 8), it is clear that based on the k_{cat}/K_m(PEP) ratio (0.78 sec^{-1} μM^{-1} vs. 0.1 sec^{-1} μM^{-1} for CP4 EPSPS and G101A petunia EPSPS, respectively), and the k_{cat} • appK_i(glyphosate)/appK_m(PEP)ratio (2106 sec^{-1} vs. 200 sec^{-1} for CP4 EPSPS and G101A petunia EPSPS, respectively), the CP4 EPSPS is 8-10 times more efficient than the G101A variant, either in the presence or absence of glyphosate.

pH dependence of CP4 EPSPS activity and stability - The pH dependence of CP4 EPSPS was measured in a MES / Tricine / MOPS / KCl buffer system, from pH 4 to pH 11. As can be seen in Figure 10, the maximal activity of CP4 EPSPS under these conditions is observed at pH 9-9.5. Virtually no activity remains at less than pH 5.5, and the activity at the standard assay pH, pH 7, is approximately 60% of the maximal activity. To address whether the loss of activity in the low pH range is due to irreversible denaturation, a stability experiment was performed, whereby the enzyme was preincubated at various pH values for 15 minutes, then assayed at pH 7. These results (Figure 10) indicate that although there is a trend towards slightly lower activity at the low end of the pH range, the majority of enzymatic activity is not irreversibly lost at either the low or high pH range.

Temperature dependence of CP4 EPSPS activity - One aspect of CP4 EPSPS which was of interest was the temperature dependence of the stability of the enzyme. One reason for this interest was to attempt to project whether glyphosate-tolerant raw agricultural commodities expressing CP4 EPSPS would be likely to be inactivated upon heat processing. As shown in Figure 11,

CP4 EPSPS was incubated for either 15 or 30 minutes at the indicated temperature, then an aliquot was diluted in a standard assay mixture at 25°C and the activity was measured. Upon incubation of CP4 EPSPS at 55°C for 15 minutes, less than half of the 25°C incubation activity was present; enzymatic activity was completely abolished after 15 minutes incubation at 65°C. These results indicate that it is very likely that the enzymatic activity of CP4 EPSPS will be lost or significantly decreased upon processing crops using heat treatment.

IV. Conclusion

EPSPS from *Agrobacterium sp.* strain CP4 has been purified, cloned, and overexpressed in *E. coli*. Kinetic analyses confirm that this EPSPS has the highest K_i(glyphosate)/K_m(PEP) ratio, 227, of any EPSPS described to date. CP4 EPSPS has approximately 50% similarity to previously described EPSPSs, and several key active site residues are conserved, indicating that the active site is largely conserved between CP4 EPSPS and other EPSPSs. CP4 EPSPS is rapidly inactivated by heat treatment, and is not active at pH values less than pH 5.5. This information supports the use of CP4 EPSPS as a highly efficient glyphosate-tolerant EPSPS for use in the development of glyphosate-tolerant crops.

Acknowledgements - The authors wish to acknowledge Dr. Robert T. Fraley for his support.

V. References

1. Levin, J. G. and D. B. Sprinson. 1964. The Enzymatic Formation and Isolation of 3-Enolpyruvylshikimate 5-phosphate. *J. Biol. Chem.* 239:1142-1150.

2. Steinrucken, H. C. and N. Amrhein. 1980. The Herbicide Glyphosate is a Potent Inhibitor of 5-Enolpyruvyl Shikimic Acid-3-Phosphate Synthase. *Biochem. Biophys. Res. Comm.* 94:1207-1212.

3. Alberg, D. G., c. T. Lauhon, R. Nyfeler, A. Fassler, and P. A. Bartlett. 1992. Inhibition of EPSP Synthase by Analogues of the Tetrahedral Intermediate and of EPSP. *J. A. C. S.* 114:3535-3546.

4. Anderson, K. S. and K. A. Johnson. 1990. Kinetic and Structural Analysis of Enzyme Intermediates: Lessons from EPSP Synthase. *Chem. Rev.* 90:1131-1149.

5. Kishore, G. and D. Shah. 1988. Amino Acid Biosynthesis Inhibitors as Herbicides. *Ann. Rev. Biochem.* 57:627-663.

6. Comai, L., D. Facciotti, W. R. Hiatt, G. Thompson, R. E. Rose, and D. M. Stalker. 1985. Expression in Plants of a Mutant *aroA* Gene from *Salmonella typhimurium* Confers Tolerance to Glyphosate. *Nature* 317:741-744.

7. Shah, D., R. Horsch, H. Klee, G. Kishore, J. Winter, N. Tumer, C. Hironaka, P. Sanders, C. Gasser, S. Aykent, N. Siegal, S. Rogers, and R. Fraley. 1986. Engineering Herbicide Tolerance in Transgenic Plants. *Science* 233:478-481.

8. Padgette, S. R., G. della-Cioppa, D. M. Shah, R. T. Fraley, and G. M. Kishore. 1989. Selective Herbicide Tolerance through Protein Engineering. In Cell Culture and Somatic Cell Genetics of Plants, Vol. 6. J. Schell and I. Vasil, editors. Academic Press, New York. 441-476.

9. Hinchee, M. A. W., SR. Padgette, G. M. Kishore, X. Delannay, and R. T. Fraley. 1993. Herbicide Tolerant Crops. In Transgenic Plants Vol. 1. S- Kung and Wu Ray, editors. Academic Press, Inc., Orlando. 243-263.

10. Kishore, G., D. Shah, S. Padgette, G. della-Cioppa, C. Gasser, D. Re, C. Hironaka, M. Taylor, J. Wibbenmeyer, D. Eichholtz, M. Hayford, N. Hoffmann, X. Delannay, R. Horsch, H. Klee, S. Rogers, D. Rochester, L. Brundage, P. Sanders, and R. T. Fraley. 1988. 5-Enolpyruvylshikimate-3-Phosphate Synthase: From Biochemistry to Genetic Engineering of Glyphosate Tolerance. In Biotechnology for Crop Protection, ACS Symposium Series No. 379. P.A. Hedin, J.J. Menn, and R.M. Hollingsworth, editors. Amercan Chemical Society, Washington. 37-48.

11. Boocock, M. R. and J. R. Coggins. 1983. Kinetics of 5-Enolpyruvylshikimate-3-phosphate Synthase Inhibition by Glyphosate. *FEBS Letters* 154:127-133.

12. Padgette, S. R., D. B. Re, C. S. Gasser, D. A. Eichholtz, R. B. Frazier, C. M. Hironaka, E. B. Levine, D. M. Shah, R. T. Fraley, and G. M. Kishore. 1991. Site-directed Mutagenesis of a Conserved Region of the 5-enolpyruvylshikimate-3-phosphate synthase Active Site. *J. Biol. Chem.* 266:22364-22369.

13. Ruff, T., D. Eichholtz, D. Re, S. Padgette, and G. Kishore. 1991. Effects of Amino Acid Substitutions on Glyphosate Tolerance and Activity of EPSPS. *Plant Physiol. Suppl* 96:94.

14. Schulz, A., D. Sost, and N. Amrhein. 1984. Insensitivity of 5-Enolpyruvylshikimic Acid-3-Phosphate Synthase to Glyphosate Confers Resistance to this Herbicide in a Strain of *Aerobacter aerogenes. Arch. Microbiol.* 137:121-123.

15. Schulz, A., A. Kruper, and N. Amrhein. 1985. Differential Sensitivity of Bacterial 5-Enolpyruvylshikimate-3-Phosphate Synthases to the Herbicide Glyphosate. *FEMS Microbiology Letters* 28:297-301.

16. Fitzgibbon, J. E. and H. D. Braymer. 1990. Cloning of a Gene from *Pseudomonas* sp. Strain PG2982 Conferring Incresed Glyphosate Resistance. *App. Environ. Microbiol.* 56:3382-3388.

17. Fitzgibbon, J.E. 1988. *Psuedomonas* sp. strain PG2982: Uptake of Glyphosate and Cloning of a Gene which Confers Increased Resistance to Glyphosate. Louisiana State University Ph.D Dissertation, Baton Rouge.

18. Powell, H. A., N. W. Kerby, and P. Powell. 1991. Natural Tolerance of Cyanobacteria to the Herbicide Glyphosate. *New Phytol.* 119:421-426.

19. Powell, H. A., N. W. Kerby, P. Rowell, D. M. Mousdale, and J. R. Coggins. 1992. Purification and Properties of a Glyphosate-tolerant 5-Enolpyruvylshikimate-3-Phosphate Synthase from the Cyanobacterium *Anabaena variabiliis. Planta* 188:484-490.

20. Barry, G., G. Kishore, S. Padgette, M. Taylor, K. Kolacz, M. Weldon, D. Re, D. Eichholtz, K. Fincher, and L. Hallas. 1992. Inhibitors of Amino Acid Biosynthesis: Strategies for Imparting Glyphosate Tolerance to Crop Plants. In Biosynthesis and Molecular Regulation of Amino Acids in Plants. B.K. Singh, H.E. Flores, and J.C. Shannon, editors. American Society of Plant

Physiologists, 139-145.

21. Bradford, M. 1976. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* 72:248-254.

22. Padgette, S. R., Q. K Huynh, S. Aykent, R. D. Sammons, J. A. Sikorski, and G. M. Kishore. 1988. Identification of the Reactive Cysteines of *Escherichia coli* 5-Enolypyruvylshikimate-3-phosphate Synthase and Their Nonessentiality for Enzymatic Catalysis. *J. Biol. Chem.* 263:1798-1802.

23. Wibbenmeyer, J., L. Brundage, S. R. Padgette, J. J. Likos, and G. M. Kishore. 1988. Mechanism of the EPSP Synthase Catalyzed Reaction: Evidence for the Lack of a Covalent Carboxyvinyl Intermediate in Catalysis. *Biochem. Biophys. Res. Comm.* 153:760-766.

24. Hunkapiller, M. W., R. M. Hewick, R. J. Dreyer, and L. E. Hood. 1983. High-sensitivity Sequencing with a Gas-phase Sequenator. In Methods in Enzymology, Vol. 91. Academic Press, San Diego. 399-413.

25. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

26. Heitkamp, M., W. J. Adams, and L. E. Hallas. 1992. Glyphosate Degradation by Immobilized Bacteria: Laboratory Studies Showing Feasibility for Glyphosate Removal from Waste Water. *Can. J. Microbiol.* 38:921-928.

27. Miller, J. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y..

28. Birnboim, H. C. and J. Doly. 1979. A Rapid Alkaline Extraction Procedure for Screening Recombinant Plasmid DNA. *Nuc. Acids. Res.* 7:1513-1523.

29. Fling, M. E., J. Kopt, and C. Richards. 1985. Nucleotide Sequence of the Transposon Tn7 Gene Encoding an Aminoglycoside-modifying Enzyme, 3"(9)-O-Nucleotidyltransferase. *Nuc. Acids. Res.* 13:7095-7106.

30. Alton, N. K. and D. Vapnek. 1979. Nucleotide Sequence Analysis of the Chloramphenicol Resistance Transposon Tn9. *Nature* 282:864-869.

31. Dunn, J. J. and F. W. Studier. 1983. Complete Nucleotide Sequence of Bacteriophage T7 DNA and the Locations of T7 Genetic Elements. *J. Mol. Biol.* 166:477-535.

32. Hohn, B. and J. Collins. 1980. A Small Cosmid for Efficient Cloning of Large DNA Fragments. *Gene* 11:291-298.

33. Boyer, H. W. and D. Roulland-Dussoix. 1969. A Complementation Analysis of the Restriction and Modification of DNA in *Escherichia coli*. *J. Mol. Biol.* 41:459-472.

34. Tabor, S. and C. C. Richardson. 1985. A Bacteriophage T7 RNA Polymerase/Promoter System for Controlled Exclusive Expression of Specific Genes. *Proc. Natl. Acad. Sci. USA* 82:1074-1078.

35. Talmadge, K. and W. Gilbert. 1980. Construction of Plasmid Vectors with Unique *PstI* Cloning Sites in a Signal Sequence Coding Region. *Gene* 12:235-241.

36. Bachmann, B. J. and K. B. Low. 1980. Linkage Map of *Escherichia coli* K-12, Edition 6. *Microbiol. Rev.* 44:1-56.

37. Padgette, S. R., Q. K. Huynh, J. Borgmeyer, D. M. Shah, L. A. Brand, D. B. Re, B. F. Bishop, S. G. Rogers, R. T. Fraley, and G. M. Kishore. 1987. Bacterial Expression and Isolation of *Petunia hybrida* 5-Enolpyruvyl-shikimate-3-phosphate Synthase. *Arch. Biochem. Biophys.* 258:564-573.

38. Kunkel, T. A., J. D. Roberts, and R. A. Zakour. 1987. Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Expression. *Meth. Enzymol.* 154:367-382.

39. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

40. Wong, E. Y., R. Seetharam, C. E. Kotts, R. A. Heeren, B. K. Klein, S. R. Braford, J. J. Mathis, B. F. Bishop, N. R. Siegel, C. E. Smith, and W. C. Tacon. 1988. Expression of Secreted Insulin-like Growth Factor-1 in *Escherichia coli*. *Gene* 68:193-203.

41. Olins, P. O., C. S. Devine, S. H. Rangwala, and K. S. Kavka. 1988. The T7 Phage Gene 10 leader RNA, a Ribosome-binding Site that Dramatically Enhances Expression of Foreign Genes in *Escherichia coli*. *Gene* 73:227-235.

42. Eichholtz, D.A., G.F. Barry, M.L. Taylor, and S.R. Padgette. 1993. Construction of DNA Vectors for the Production of Glyphosate-tolerant Soybeans. Monsanto Technical Report MSL-12905, St. Louis.

43. Manual, University of Wisconsin Genetics Computer, V.7, April, 1991. University of Wisconsin Genetics Computer Group, Madison, WI.

44. Duncan, K., A. Lewendon, and J. R. Coggins. 1984. The Purification of 5-Enolpyruvylshikimate-3-Phosphate Synthase from an Overproducing Strain of *Escherichia coli*. *FEBS* 165:121-127.

45. Gasser, C. S., J. A. Winter, C. M. Hironaka, and D. M. Shah. 1988. Structure, Expression, and Evolution of the 5-Enolpyruvyl-shikimate-3-Phosphate Synthase Genes of Petunia and Tomato. *J. Biol. Chem.* 263:4280-4289.

46. Bishop, B. 1992. Production of CP4 EPSP Synthase in a 100 liter Recombinant *Escherichia coli* Fermentation. Monsanto Technical Report MSL-12389, St. Louis.

47. Heeren, R.A., S.R. Padgette, and M.E. Gustafson. 1993. Purification of Recombinant *Escherichia coli* CP4 5-Enol Pyruvyl Shikimate 3-Phosphate Synthase for Equivalence Studies. Monsanto Report MSL-12574, St. Louis.

48. Mountain, A. 1989. Gene Expression Systems for *Bacillus subtilis*. In *Bacillus*. C.R. Harwood, editor. Plenum Press, New York. 73-114.

49. Huynh, Q. K., G. M. Kishore, and G. S. Bild. 1988. 5-Enolpyruvyl Shikimate-3-Phosphate Synthase from *Escherichia coli*: Identification of Lys-22 as a Potential Active Site Residue. *J. Biol. Chem.* 263:735-739.

50. Padgette, S. R., C. E. Smith, Q. K. Huynh, and G. M. Kishore. 1988. Arginine Chemical Modification of *Petunia hybrida* 5-Enolpyruvylshikimate-3-Phosphate Synthase. *Arch. Biochem. Biophys.* 266:254-262.

51. Duncan, K., A. Lewendon, and J. R. Coggins. 1984. The Complete Amino Acid Sequence of *Escherichia coli* 5-Enolpyruvylshikimate-3-phosphate Synthase. *FEBS Letters* 170:59-63.

52. Henner, D. J., L. Band, G. Flaggs, and E. Chen. 1986. The Organization and Nucleotide Sequence of the *Bacillus subtilis his*H, *tyr*A, and *aro*E Genes. *Gene* 49:147-152.

53. Duncan, K., R. M. Edwards, and J. R. Coggins. 1987. The Pentafunctional *arom* Enzyme of *Saccharomyces cerevisiae* is a Mosaic of Monofunctional Domains. *Biochem. J.* 246:375-386.

54. Harrison, L A., M.R. Bailey, R.M. Leimgruber, C.E. Smith, D.L. Nida, M.L. Taylor, and S.R. Padgette. 1993. Characterization of Microbially-Expressed Protein: CP4 EPSPS. Monsanto Study 92-01-30-14, Technical Report MSL-12901, St. Louis.

55. Padgette, S.R. 1993. Purification of *Agrobacterium* sp. strain CP4 5-Enolpyruvyl-shikimate-3-phosphate Synthase Expressed in Glyphosate-tolerant Soybean Seeds. Monsanto Technical Report MSL-12739, St. Louis.

56. Harrison, L.A., M. Bailey, and T. Nickson. 1993. Equivalence of Plant and Microbially-Expressed Proteins: CP4 EPSPS in Roundup tolerant Canola and *E. coli*. Monsanto Study 92-02-30-03, Technical Report MSL-12899, St. Louis.

Table 1. Kinetic characteristics of several EPSPSs

EPSPS source	appK_m(PEP) μM	appK_i(glyphosate) μM	appK_i(glyphosate)/ appK_m(PEP)	Reference
E. coli	3.5	1.1	0.31	(11)
Petunia	5	0.4	0.08	(37)
petunia G101A	210	2000	9.5	(12)
petunia G101A,G144D	47	330	7.0	(13)
petunia G101A,P167S	35	238	6.8	(13)
petunia P106S	44	3	0.07	(12)
Anabaena variablis	34	350	10.3	(19)

Table 2. Purification table for *Agrobacterium sp.* strain CP4 EPSPS

Step	Volume, ml	Act., U/ml	Act., U	Protein, mg/ml	Protein, mg	Spec. Act., U/mg	Yield	Purif., fold
crude	1400	0.0618	86.5	13.4	18800	0.005	100%	1
protamine sulfate	1600	0.0585	93.6	10.5	16700	0.006	108%	1.2
ammonium sulfate	290	0.180	52.1	35.4	10300	0.005	60.2%	1
Phenyl Sepharose	300	0.108	32.4	3.03	909	0.036	37.5%	7.2
Q Sepharose	200	0.277	55.4	0.816	163	0.339	64.0%	68
Mono Q 10/10	6	6.94	41.6	2.76	16.5	2.51	48.1%	502
Phenyl Superose	20.5	0.40	8.29	0.143	2.93	2.83	9.6%	566
Mono Q 5/5 load	0.55	3.1	1.71	1.1	0.605	2.83	-	-
Mono Q 5/5 product	0.65	2.0	1.24	0.17	0.11	11.8	72% (step) 6.9%(overall)	2360

Table 3. Peptide sequences and oligonucleotide probes for CP4 EPSPS

Peptide	Peptide Sequence	Probe #	Probe Sequence
61-24-25	APSM(I)(D)EYPILAV mixed probe 24-fold degenerate	MID; 17-mer;	ATGATA/C/TGAC/TGAG/ATAC/TCC
53-28	ITGLLEGEDVINTGK 17-mer; mixed probe, 48-fold degenerate	EDV-C;	GAA/GGAC/TGTA/C/G/TATA/C/TAACAC
	17-mer; mixed probe, 48-fold degenerate	EDV-T;	GAA/GGAC/TGTA/C/G/TATA/C/TAATAC
41-25-27	NVLMNPTR	NA	(Sequence not used for probe)

Table 4. Primers used for site-directed mutagenesis of CP4 EPSPS

Primer name (explanation)	Primer oligonucleotide sequence
PRIMER BgNc: (addition of BglII and NcoI sites to N-terminus)	CGTGGATAGATCTAGGAAGACAACCATGGCTCACGGTC
PRIMER Sph2: (addition of SphI site to N-terminus)	GGATAGATTAAGGAAGACGCGCATGCTTCACGGTGCAAGCAGCC
PRIMER S1: (addition of SacI site immediately after stop codons)	GGCTGCCTGATGAGCTCCACAATCGCCATCGATGG
PRIMER N1: (removal of internal NotI recognition site)	CGTCGCTCGTCGTGCGTGGCCGCCCTGACGGC
PRIMER Nco1: (removal of first internal NcoI recognition site)	CGGGCAAGGCCATGCAGGCTATGGGCGCC
PRIMER Nco2 (removal of second internal NcoI recognition site)	CGGGCTGCCGCCTGACTATGGGCCTCGTCGG
PRIMER sph to nde (introduction of leu at second position of plant-preferred codon CP4 EPSPS gene to make pMON21104)	CACGGCGCATATGCTTCACGG

Table 5. Comparison of relatedness[a] of EPSPS protein sequences

Comparison	Similarity	Identity	Appendix	Reference for non-CP4 EPSPS sequence
CP4 EPSPS vs. soybean EPSPS	51.2	26.0	C	Footnote b
CP4 EPSPS vs. corn EPSPS	48.5	24.1	E	Footnote b
CP4 EPSPS vs. petunia EPSPS	50.1	23.3	G	(45)
CP4 EPSPS vs. *E. coli* EPSPS	52.2	26.0	I –	(51)
CP4 EPSPS vs. *B. subtilis* EPSPS	59.3	41.1	K	(52)
CP4 EPSPS vs. yeast EPSPS	53.5	29.9	M	(53)
soybean EPSPS vs. *B. subtilis* EPSPS	55.6	30.1	N	(52), footnote b
soybean EPSPS vs. corn EPSPS	90.3	83.1	O	Footnote b
soybean EPSPS vs. *E. coli* EPSPS	71.8	54.8	P	(51), footnote b
B. subtilis EPSPS vs. *E. coli* EPSPS	55.5	29.1	Q	(51,52)

[a] The deduced amino acid sequences were compared using the "Gap" computer program provided in the UWGCG package (43).
[b] Unpublished data, Monsanto

Table 6. EPSPS active site residues in CP4 EPSPS homologous to other EPSPS enzymes

Residue number in CP4 EPSPS	Residue in *E. coli* EPSPS	Residue in *p. hybrida* EPSPS	Postulated interaction with...
Arginine 33	Arginine 27	Arginine 28	S3P
Lysine 28	Lysine 22	Lysine 23	PEP/glyphosate
Arginine 128	Arginine 124	Arginine 131	PEP/glyphosate
Alanine 100	Glycine 96	Glycine 101	glyphosate > PEP

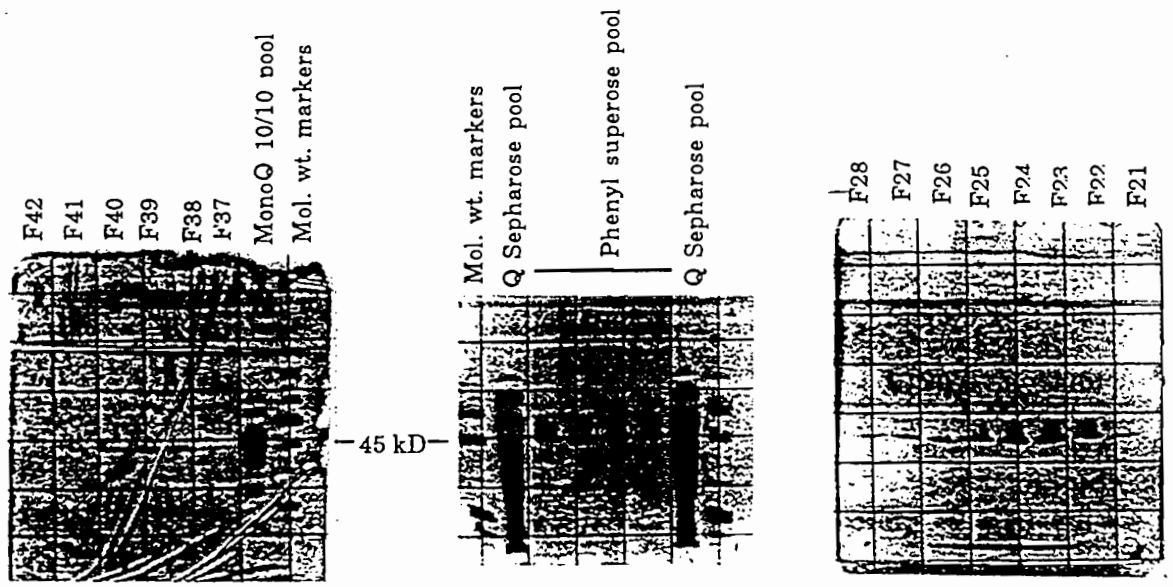
Table 7. N-terminal sequence comparison between several CP4 EPSPS proteins

N-terminal amino acid sequence	Source
M S H G ASS R PATA R K SS GL	Deduced N-terminus from original CP4EPSPS gene from *Agrobacterium sp.* strain CP4 pMON17081
X H(G)ASS R PATA R K SS(G)L	N-terminal sequence from purified CP4 EPSPS from *Agrobacterium sp.* strain CP4
M A H G ASS R PATA R K SS GL	Deduced N-terminus from the CP4 EPSPS gene overexpressed in *E. coli* from pMON17101
A H G ASS R PATA R K SS	N-terminal sequence from purified pMON17101 CP4 EPSPS
M L H G ASS R PATA R K SS GL	Deduced N-terminus from the CP4 EPSPS gene overexpressed in *E. coli* from pMON21104
M L H G ASS R PA	N-terminal sequence from purified pMON21104 CP4 EPSPS (54)
L H G ASS R PATA R K SS	Deduced sequence of N-terminus of mature CP4 EPSPS from plant expression vectors using the sphI EPSPS chloroplast transit peptide fusion
(L)H G ASS R PATA R K SS	Major N-terminal sequence from purified RT soybean seed-expressed CP4 EPSPS (55)
(L)H G ASS R PATA R K SS	N-terminal sequence from purified RT canola seed-expressed CP4 EPSPS (56)

Table 8. Steady-state kinetic characterization of CP4 EPSPS

Kinetic constant	CP4 EPSPS	Petunia G101A EPSPS (12)
appKm(PEP), μM	12 μM	210 μM
$appK_m$(S3P), μM	1 μM	11 μM
$appK_i$(glyphosate), μM	2.7 mM	2000 μM
$appK_i$(glyphosate)/$appK_m$(PEP)	227	9.5
Specific activity	11.8 U/mg	27 U/mg
k_{cat}	9.4 sec^{-1}	21 sec^{-1}
k_{cat}/K_m(PEP)	0.78 sec^{-1} μM^{-1}	0.1 sec^{-1} μM^{-1}
k_{cat} • $appK_i$(glyphosate) / $appK_m$(PEP)	2106 sec^{-1}	200 sec^{-1}

Figure 1. SDS-PAGE of purified EPSPS from *Agrobacterium sp.* strain CP4. Samples were run on a 10-15% Phast gel (Pharmacia) and developed with silver stain according to the manufacturer's instructions.



A. Run 1, Phenyl Superose column. Fraction 39 was submitted for N-terminal sequencing

B. Final Pool, Phenyl Superose column. The phenyl superose pool was comprised of the fractions from three separate runs as described in Experimental Procedures

C. Mono Q 5/5 column Fractions. Fractions 22-25 were pooled and trypsinized

Figure 2. Nucleotide sequence of the *Agrobacterium sp.* strain CP4 EPSPS gene as cloned into pMON17081

```
        10                  30                  50
ATGTCGCACGGTGCAAGCAGCCGGCCCGCAACCGCCCGCAAATCCTCTGGCCTTTCCGGA
        70                  90                  110
ACCGTCCGCATTCCCGGCGACAAGTCGATCTCCCACCGGTCCTTCATGTTCGGCGGTCTC
       130                 150                 170
GCGAGCGGTGAAACGCGCATCACCGGCCTTCTGGAAGGCGAGGACGTCATCAATACGGGC
       190                 210                 230
AAGGCCATGCAGGCCATGGGCGCCAGGATCCGTAAGGAAGGCGACACCTGGATCATCGAT
       250                 270                 290
GGCGTCGGCAATGGCGGCCTCCTGGCGCCTGAGGCGCCGCTCGATTTCGGCAATGCCGCC
       310                 330                 350
ACGGGCTGCCGCCTGACCATGGGCCTCGTCGGGGTCTACGATTTCGACAGCACCTTCATC
       370                 390                 410
GGCGACGCCTCGCTCACAAAGCGCCCGATGGGCCGCGTGTTGAACCCGCTGCGCGAAATG
       430                 450                 470
GGCGTGCAGGTGAAATCGGAAGACGGTGACCGTCTTCCCGTTACCTTGCGCGGGCCGAAG
       490                 510                 530
ACGCCGACGCCGATCACCTACCGCGTGCCGATGGCCTCCGCACAGGTGAAGTCCGCCGTG
       550                 570                 590
CTGCTCGCCGGCCTCAACACGCCCGGCATCACGACGGTCATCGAGCCGATCATGACGCGC
       610                 630                 650
GATCATACGGAAAAGATGCTGCAGGGCTTTGGCGCCAACCTTACCGTCGAGACGGATGCG
       670                 690                 710
GACGGCGTGCGCACCATCCGCCTGGAAGGCCGCGGCAAGCTCACCGGCCAAGTCATCGAC
       730                 750                 770
GTGCCGGGCGACCCGTCCTCGACGGCCTTCCCGCTGGTTGCGGCCCTGCTTGTTCCGGGC
       790                 810                 830
TCCGACGTCACCATCCTCAACGTGCTGATGAACCCCACCCGCACCGGCCTCATCCTGACG
       850                 870                 890
CTGCAGGAAATGGGCGCCGACATCGAAGTCATCAACCCGCGCCTTGCCGGCGGCGAAGAC
       910                 930                 950
GTGGCGGACCTGCGCGTTCGCTCCTCCACGCTGAAGGGCGTCACGGTGCCGGAAGACCGC
       970                 990                1010
GCGCCTTCGATGATCGACGAATATCCGATTCTCGCTGTCGCCGCCGCCTTCGCGGAAGGG
      1030                1050                1070
GCGACCGTGATGAACGGTCTGGAAGAACTCCGCGTCAAGGAAAGCGACCGCCTCTCGGCC
      1090                1110                1130
GTCGCCAATGGCCTCAAGCTCAATGGCGTGGATTGCGATGAGGGCGAGACGTCGCTCGTC
      1150                1170                1190
GTGCGCGGCCGCCCTGACGGCAAGGGGCTCGGCAACGCCTCGGGCGCCGCCGTCGCCACC
      1210                1230                1250
CATCTCGATCACCGCATCGCCATGAGCTTCCTCGTCATGGGCCTCGTGTCGGAAAACCCT
      1270                1290                1310
GTCACGGTGGACGATGCCACGATGATCGCCACGAGCTTCCCGGAGTTCATGGACCTGATG
      1330                1350                1370
GCCGGGCTGGGCGCGAAGATCGAACTCTCCGATACGAAGGCTGCCTGATGA
```

Figure 3. Deduced amino acid of the *Agrobacterium sp.* strain CP4 EPSPS gene from pMON17081

```
Pmon17081.pep  Length: 457  April 9, 1993  15:20  Type: P  Check: 9740  ..

      1  MSHGASSRPA TARKSSGLSG TVRIPGDKSI SHRSFMFGGL ASGETRITGL

     51  LEGEDVINTG KAMQAMGARI RKEGDTWIID GVGNGGLLAP EAPLDFGNAA

    101  TGCRLTMGLV GVYDFDSTFI GDASLTKRPM GRVLNPLREM GVQVKSEDGD

    151  RLPVTLRGPK TPTPITYRVP MASAQVKSAV LLAGLNTPGI TTVIEPIMTR

    201  DHTEKMLQGF GANLTVETDA DGVRTIRLEG RGKLTGQVID VPGDPSSTAF

    251  PLVAALLVPG SDVTILNVLM NPTRTGLILT LQEMGADIEV INPRLAGGED

    301  VADLRVRSST LKGVTVPEDR APSMIDEYPI LAVAAAFAEG ATVMNGLEEL

    351  RVKESDRLSA VANGLKLNGV DCDEGETSLV VRGRPDGKGL GNASGAAVAT

    401  HLDHRIAMSF LVMGLVSENP VTVDDATMIA TSFPEFMDLM AGLGAKIELS

    451  DTKAA**
```

Figure 4. Plasmid map of pMON17101, an *E. coli* expression vector for *Agrobacterium sp.* strain CP4 EPSPS

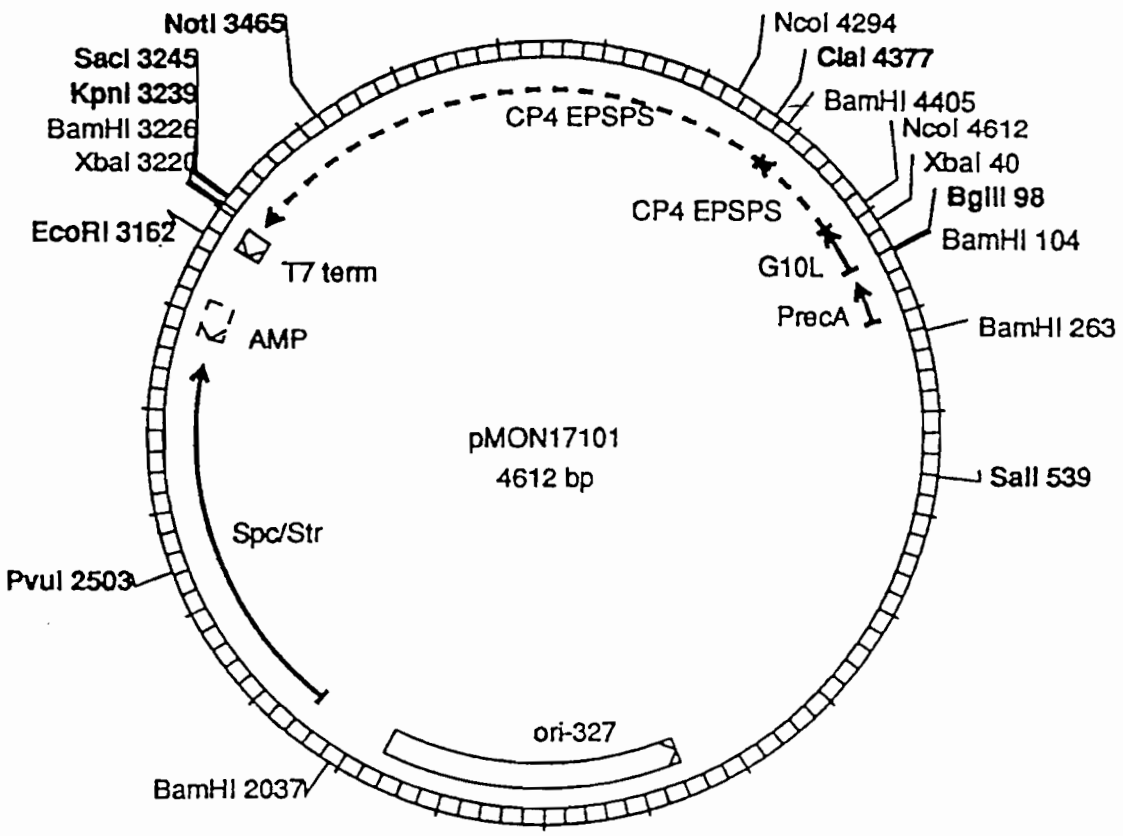


Figure 5. Plasmid map of pMON21104, an *E. coli* expression vector for *Agrobacterium sp.* strain CP4 EPSPS

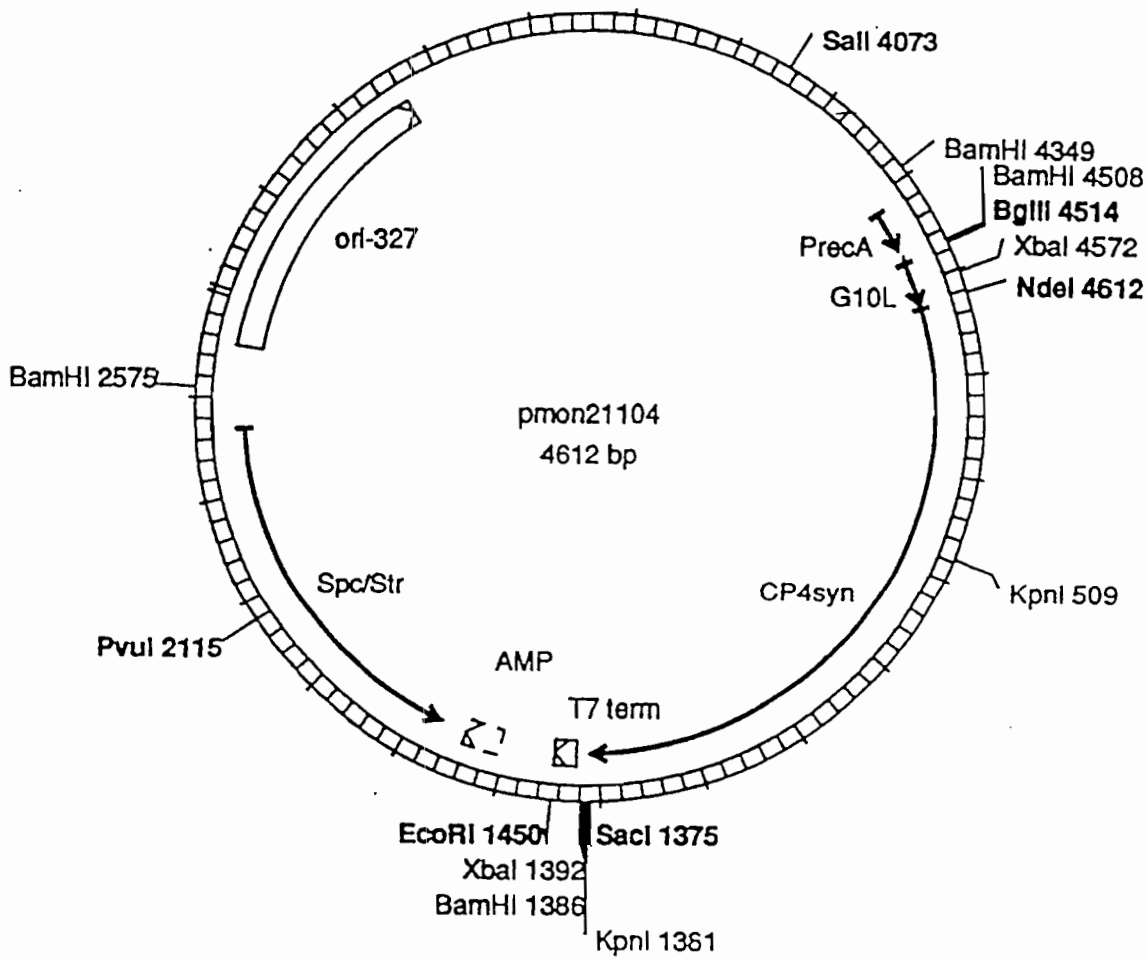


Figure 6. Comparison of DNA and protein sequences of CP4 EPSPS expressed in pMON17101 (top) and pMON21104 (bottom). The CP4 EPSPS amino acid sequences in both vectors are identical, with the exception of the asterisked second residue, which is leu in pMON21104.

```
          *
          MetAlaHisGlyAlaSerSerArgProAlaThrAlaArgLysSerSerGlyLeuSerGly
          ATGGCTCACGGTGCAAGCAGCCGGCCCGCAACCGCCCGCAAATCCTCTGGCCTTTCCGGA
1         |||  ||||||||||||||||||| || ||||| || || || |||||||| ||||| |||   60
          ATGCTTCACGGTGCAAGCAGCCGTCCAGCAACTGCTCGTAAGTCCTCTGGTCTTTCTGGA

          ThrValArgIleProGlyAspLysSerIleSerHisArgSerPheMetPheGlyGlyLeu
          ACCGTCCGCATTCCCGGCGACAAGTCGATCTCCCACCGGTCCTTCATGTTCGGCGGTCTC
61        |||||||| ||||| || |||||||| |||||||| |||||||||||| || ||||||   120
          ACCGTCCGTATTCCAGGTGACAAGTCTATCTCCCACAGGTCCTTCATGTTTGGAGGTCTC

          AlaSerGlyGluThrArgIleThrGlyLeuLeuGluGlyGluAspValIleAsnThrGly
          GCGAGCGGTGAAACGCGCATCACCGGCCTTCTGGAAGGCGAGGACGTCATCAATACGGGC
121       || |||||||||| || |||||||| ||| ||||||| || || || ||||| || ||   180
          GCTAGCGGTGAAACTCGTATCACCGGTCTTTTGGAAGGTGAAGATGTTATCAACACTGGT

          LysAlaMetGlnAlaMetGlyAlaArgIleArgLysGluGlyAspThrTrpIleIleAsp
          AAGGCCATGCAGGCTATGGGCGCCAGGATCCGTAAGGAAGGCGACACCTGGATCATCGAT
181       ||||| ||||| |||||||| ||||| |||||||||||||| || || |||||||| |||   240
          AAGGCTATGCAAGCTATGGGTGCCAGAATCCGTAAGGAAGGTGATACTTGGATCATTGAT

          GlyValGlyAsnGlyGlyLeuLeuAlaProGluAlaProLeuAspPheGlyAsnAlaAla
          GGCGTCGGCAATGGCGGCCTCCTGGCGCCTGAGGCGCCGCTCGATTTCGGCAATGCCGCC
241       || || || || || || ||||| || |||||||| || |||||||||| || || ||   300
          GGTGTTGGTAACGGTGGACTCCTTGCTCCTGAGGCTCCTCTCGATTTCGGTAACGCTGCA

          ThrGlyCysArgLeuThrMetGlyLeuValGlyValTyrAspPheAspSerThrPheIle
          ACGGGCTGCCGCCTGACCATGGGCCTCGTCGGGGTCTACGATTTCGACAGCACCTTCATC
301       || || |||||  |||| ||||| || || || || |||||||||| ||||| |||||   360
          ACTGGTTGCCGTTTGACTATGGGTCTTGTTGGTGTTTACGATTTCGATAGCACTTTCATT

          GlyAspAlaSerLeuThrLysArgProMetGlyArgValLeuAsnProLeuArgGluMet
          GGCGACGCCTCGCTCACAAAGCGCCCGATGGGCCGCGTGTTGAACCCGCTGCGCGAAATG
361       || ||||| || ||||| ||||| || ||||| || |||||||||| || |||||||||   420
          GGTGACGCTTCTCTCACTAAGCGTCCAATGGGTCGTGTGTTGAACCCACTTCGCGAAATG
```

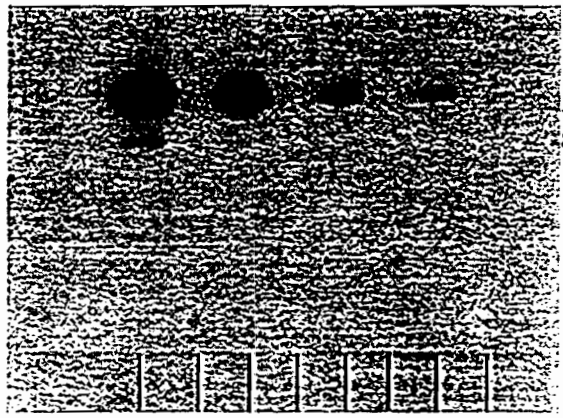
Figure 6. Comparison of DNA and protein sequences of CP4 EPSPS expressed in pMON17101 (top) and pMON21104 (bottom) (continued).

```
          GlyValGlnValLysSerGluAspGlyAspArgLeuProValThrLeuArgGlyProLys
          GGCGTGCAGGTGAAATCGGAAGACGGTGACCGTCTTCCCGTTACCTTGCGCGGGCCGAAG
421       || |||||||||||| || |||||||||| |||||||| |||||||||| || || |||   480
          GGTGTGCAGGTGAAGTCTGAAGACGGTGATCGTCTTCCAGTTACCTTGCGTGGACCAAAG

          ThrProThrProIleThrTyrArgValProMetAlaSerAlaGlnValLysSerAlaVal
          ACGCCGACGCCGATCACCTACCGCGTGCCGATGGCCTCCGCACAGGTGAAGTCCGCCGTG
481       || || ||||| ||||||||| | || || ||||| ||||| || ||||||||||| ||   540
          ACTCCAACGCCAATCACCTACAGGGTACCTATGGCTTCCGCTCAAGTGAAGTCCGCTGTT

          LeuLeuAlaGlyLeuAsnThrProGlyIleThrThrValIleGluProIleMetThrArg
          CTGCTCGCCGGCCTCAACACGCCCGGCATCACGACGGTCATCGAGCCGATCATGACGCGC
541       ||||| || || |||||||| || || ||||| || || |||||||| |||||||| ||   600
          CTGCTTGCTGGTCTCAACACCCCAGGTATCACCACTGTTATCGAGCCAATCATGACTCGT

          AspHisThrGluLysMetLeuGlnGlyPheGlyAlaAsnLeuThrValGluThrAspAla
          GATCATACGGAAAAGATGCTGCAGGGCTTTGGCGCCAACCTTACCGTCGAGACGGATGCG
601       || || || |||||||||||| || || ||||| || |||||||||||| ||||| |||||   660
          GACCACACTGAAAAGATGCTTCAAGGTTTTGGTGCTAACCTTACCGTTGAGACTGATGCT

          AspGlyValArgThrIleArgLeuGluGlyArgGlyLysLeuThrGlyGlnValIleAsp
          GACGGCGTGCGCACCATCCGCCTGGAAGGCCGCGGCAAGCTCACCGGCCAAGTCATCGAC
661       ||||| ||||| |||||||| || ||||| || || |||||||||||| ||||| || ||   720
          GACGGTGTGCGTACCATCCGTCTTGAAGGTCGTGGTAAGCTCACCGGTCAAGTGATTGAT

          ValProGlyAspProSerSerThrAlaPheProLeuValAlaAlaLeuLeuValProGly
          GTGCCGGGCGACCCGTCCTCGACGGCCTTCCCGCTGGTTGCGGCCCTGCTTGTTCCGGGC
721       || || || || || ||||| || || |||||  ||||||| ||| |||||||||| ||   780
          GTTCCAGGTGATCCATCCTCTACTGCTTTCCCATTGGTTGCTGCCTTGCTTGTTCCAGGT

          SerAspValThrIleLeuAsnValLeuMetAsnProThrArgThrGlyLeuIleLeuThr
          TCCGACGTCACCATCCTCAACGTGCTGATGAACCCCACCCGCACCGGCCTCATCCTGACG
781       |||||||||||||||| |||||  |||||||||| ||||| || || |||||| ||||   840
          TCCGACGTCACCATCCTTAACGTTTTGATGAACCCAACCCGTACTGGTCTCATCTTGACT

          LeuGlnGluMetGlyAlaAspIleGluValIleAsnProArgLeuAlaGlyGlyGluAsp
          CTGCAGGAAATGGGCGCCGACATCGAAGTCATCAACCCGCGCCTTGCCGGCGGCGAAGAC
841       |||||||||||||| |||||||||||||| |||||||| || ||||| || || ||||||   900
          CTGCAGGAAATGGGTGCCGACATCGAAGTGATCAACCCACGTCTTGCTGGTGGAGAAGAC

          ValAlaAspLeuArgValArgSerSerThrLeuLysGlyValThrValProGluAspArg
          GTGGCGGACCTGCGCGTTCGCTCCTCCACGCTGAAGGGCGTCACGGTGCCGGAAGACCGC
901       ||||| ||| |||| ||||| || || ||  ||||||| || || || || ||||||||   960
          GTGGCTGACTTGCGTGTTCGTTCTTCTACTTTGAAGGGTGTTACTGTTCCAGAAGACCGT
```

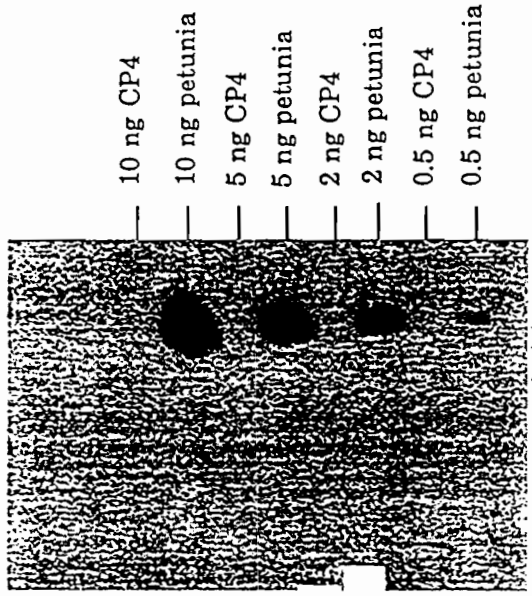
Figure 6. Comparison of DNA and protein sequences of CP4 EPSPS expressed in pMON17101 (top) and pMON21104 (bottom) (continued).

```
        AlaProSerMetIleAspGluTyrProIleLeuAlaValAlaAlaAlaPheAlaGluGly
        GCGCCTTCGATGATCGACGAATATCCGATTCTCGCTGTCGCCGCCGCCTTCGCGGAAGGG
961     || ||||| |||||||||||| ||||| |||||||||||| || || || ||||| |||||   1020
        GCTCCTTCTATGATCGACGAGTATCCAATTCTCGCTGTTGCAGCTGCATTCGCTGAAGGT

        AlaThrValMetAsnGlyLeuGluGluLeuArgValLysGluSerAspArgLeuSerAla
        GCGACCGTGATGAACGGTCTGGAAGAACTCCGCGTCAAGGAAAGCGACCGCCTCTCGGCC
1021    || ||||| ||||||||| |||||||||||||| || ||||||||||||||| || || ||   1080
        GCTACCGTTATGAACGGTTTGGAAGAACTCCGTGTTAAGGAAAGCGACCGTCTTTCTGCT

        ValAlaAsnGlyLeuLysLeuAsnGlyValAspCysAspGluGlyGluThrSerLeuVal
        GTCGCCAATGGCCTCAAGCTCAATGGCGTGGATTGCGATGAGGGCGAGACGTCGCTCGTC
1081    ||||| || || |||||||||||| || || |||||||||||| || ||||| || ||||||   1140
        GTCGCAAACGGTCTCAAGCTCAACGGTGTTGATTGCGATGAAGGTGAGACTTCTCTCGTC

        ValArgGlyArgProAspGlyLysGlyLeuGlyAsnAlaSerGlyAlaAlaValAlaThr
        GTGCGCGGCCGCCCTGACGGCAAGGGGCTCGGCAACGCCTCGGGCGCCGCCGTCGCCACC
1141    ||||| || || |||||||| ||||| ||||| ||||| || || || || ||||| |||   1200
        GTGCGTGGTCGTCCTGACGGTAAGGGTCTCGGTAACGCTTCTGGAGCAGCTGTCGCTACC

        HisLeuAspHisArgIleAlaMetSerPheLeuValMetGlyLeuValSerGluAsnPro
        CATCTCGATCACCGCATCGCCATGAGCTTCCTCGTCATGGGCCTCGTGTCGGAAAACCCT
1121    || |||||||||| ||||| |||||||||||||| ||||| ||||| || |||||||||   1260
        CACCTCGATCACCGTATCGCTATGAGCTTCCTCGTTATGGGTCTCGTTTCTGAAAACCCT

        ValThrValAspAspAlaThrMetIleAlaThrSerPheProGluPheMetAspLeuMet
        GTCACGGTGGACGATGCCACGATGATCGCCACGAGCTTCCCGGAGTTCATGGACCTGATG
1261    || || || || ||||| || |||||||| || |||||||| |||||||||| |||||   1320
        GTTACTGTTGATGATGCTACTATGATCGCTACTAGCTTCCCAGAGTTCATGGATTTGATG

        AlaGlyLeuGlyAlaLysIleGluLeuSerAspThrLysAlaAlaEnd
        GCCGGGCTGGGCGCGAAGATCGAACTCTCCGATACGAAGGCTGCCTGA
1321    || || || || || |||||||||||||||| !! |||||||| |||
        GCTGGTCTTGGAGCTAAGATCGAACTCTCCGACACTAAGGCTGCTTGA
```

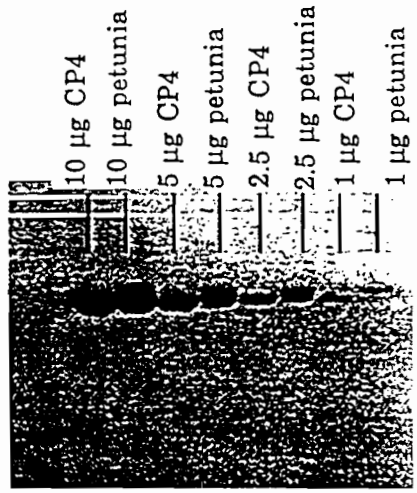
Figure 7. Comparison of specificity of CP4 and petunia EPSPS antibodies.



A. Western blot probed with anti-CP4 EPSPS antibody



B. Western blot probed with anti-petunia EPSPS antibody



C. Coomassie-blue stained SDS-PAGE

Figure 8. Comparison of CP4 EPSPS with a consensus EPSPS sequence. The EPSPS consensus sequence was obtained using published EPSPS sequences as described in "Experimental Procedures," with absolutely conserved residues in capital letters. The consensus EPSPS sequence thus obtained was compared to the CP4 EPSPS sequence described herein, resulting in a new consensus sequence.

```
                51                                                  100
CP4             .......... .......... ..MSHGASSR PATARKSSGL SGTVRIPGDK
EPSPS           .......... .......... ........me sltlqpiar. .gtinlPGsK
Consensus       ---------- ---------- ---------- --T------- -GT---PG-K

                101                                                 150
CP4             SISHRSFMFG GLAS....GE TRITGLLEGE DVINTGKAMQ AMGARIRKEG
EPSPS           svsnRallla ALae....Gt tv.tnlLdsd Dirhmlnalk alGln.elsa
Consensus       S-S-R----- -LA-----G- T--T-LL--- D------A-- A-G-------

                151                                                 200
CP4             DTWIIDGVGN GGLLAPE... .APLD..FGN AATGCR.LTM GLVGVYDFD.
EPSPS           drtrcdvegc gg.lpas... .galelflGn aGTamRpl.a alcl..g.n.
Consensus       D----D--G- GG-L------ ---L----GN A-T--R-L-- -L--------

                201                                                 250
CP4             STFIGDASLT KRPMGRVLNP LREMGVQV.. .KSEDGDRLP VTLRGPKTPT
EPSPS           .vl.G.prm. eRPig.lvd. Lrq.Ganidy leqenypplr ....gg.pgG
Consensus       ----G----- -RP-G----- LR--G----- ---E----L- ----G-----

                251                                                 300
CP4             PITYRVPMAS AQVKSAVLLA GLNTPGITTV IEPIMTRDHT EKMLQGFGAN
EPSPS           kvkvdgsvSs QfltalLm.a pl....a.gd teieikgel. SkpyidiTln
Consensus       ---------S ---------A -L-------- -E-------- -K-------N

                301                                                 350
CP4             L......TVE TDADGVRTIR LEGRGKLTGQ VIDVPGDPSS TAFPLVAALL
EPSPS           lmkrfGv..e h..h.qrfvv kggqkykspg rylvegDass AsyFLaaaa.
Consensus       L--------E ------R--- --G------- ---V-GD-SS ----L-AA--

                351                                                 400
CP4             VPGSDVTILN VLMNPTRTGL ILT..LQEMG ADIEVINPRL AGGEDVADLR
EPSPS           ikggtv.vtg ig..s.qgd. kfadvLekmg a.vtwgddsi ...gp....a
Consensus       --G--V---- ---------- -----L--MG A--------- ----------

                401                                                 450
CP4             VRSSTLKGVT VPEDRAPSMI DEYPILAVAA AFAEGATVMN GLEELRVKES
EPSPS           cgrgelraid .dmnhipdaa mtiat.AlfA ...dgptt.r nia.wrvkET
Consensus       -----L---- ------P--- ------A--A ----G-T--- -----RVKE-
```

Figure 8. Comparison of CP4 EPSPS with a consensus EPSPS sequence (continued).

```
                  451                                                 500
CP4               DRLSAVANGL KLNGVDCDEG ETSLVVRGRP DGKGLGNASG AAVATHLDHR
EPSPS             dRl.amatEl rklGaeveeg hDyiri.tpp ak..l.h....aeigty.DHR
Consensus         DRL-A-A--L ---G----EG ---------P ----L----- A---T--DHR

                  501                                                 550
CP4               IAMSFLVMGL VSENPVTVDD ATMIATSFPE FMDLMAGLGA KIELSDTKAA
EPSPS             maMcfslaa. lsdtpvtild p.ctaktfPd yfe.larmsk .a........
Consensus         -AM-F----- -S--PVT--D ----A--FP- -----A---- ----------

                  551           566
CP4               .......... ......
EPSPS             .......... ......
Consensus         ---------- ------
```

Figure 9. Determination of the appK$_i$(glyphosate) for CP4 EPSPS. EPSPS assays were performed using the HPLC radioassay as described in Experimental Procedures. Inset: replot of the slope values obtained by linear regression; the X-intercept, 2.7 mM, represents the appK$_i$(glyphosate) for CP4 EPSPS.

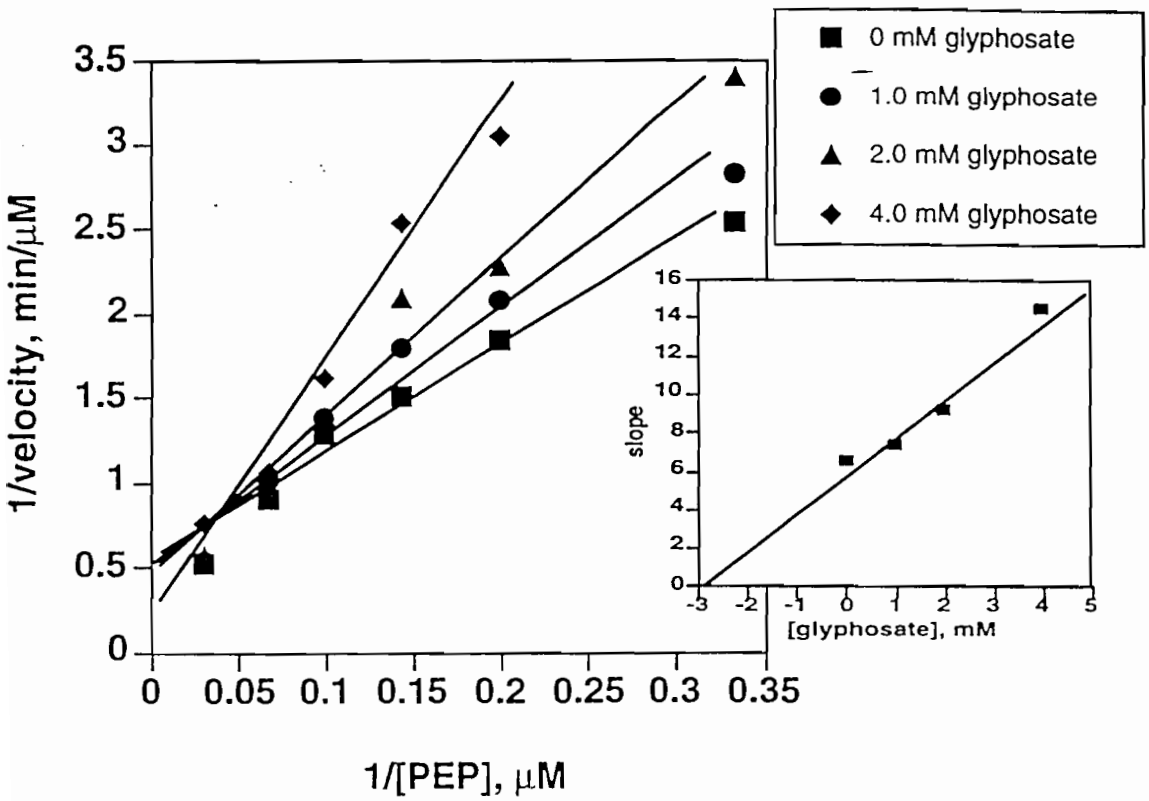


Figure 10. pH dependence of CP4 EPSPS activity; filled circles: The pH dependence of CP4 EPSPS activity was determined in an assay buffer solution comprised of 75 mM MES, 75 mM Tricine, 75 mM MOPS, and 50 mM KCl, using the standard PEP and S3P concentrations of 1 mM and 2 mM, respectively, at 25°C. The enzyme used was purified from *E. coli* expressing pMON21104. **Open circles: The pH stability profile of CP4 EPSPS.** CP4 EPSPS (0.06 mg/ml) was incubated for 15 minutes at the indicated pH in the MES/Tricine/MOPS/KCl buffer system, at 25°C, then an aliquot (5 µl) was diluted into the pH 7.0 MES/Tricine/MOPS/KCl buffer system (100 µl total volume), and 10 µl of this diluted enzyme was then assayed at pH 7 in the same buffer system as the pH activity measurements.

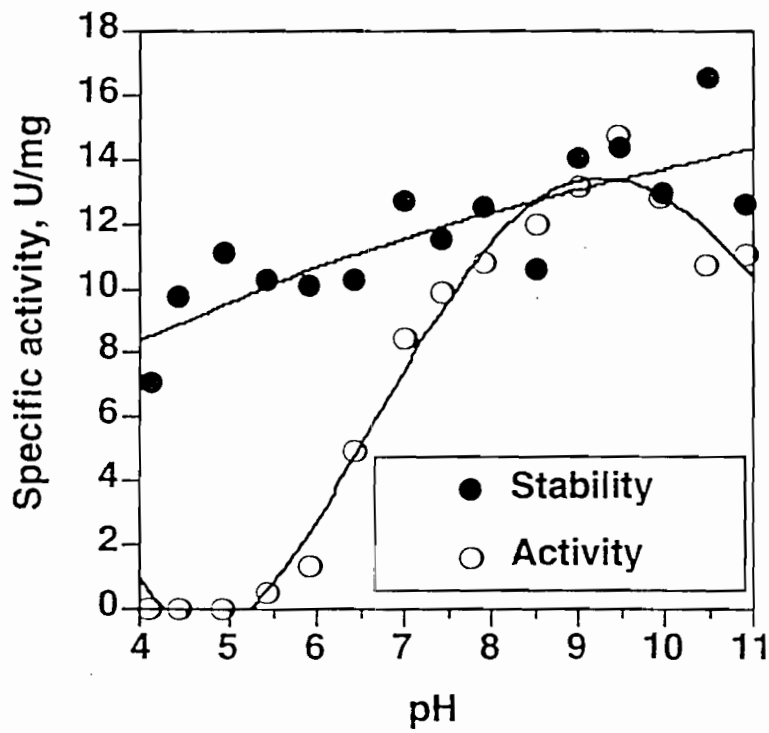
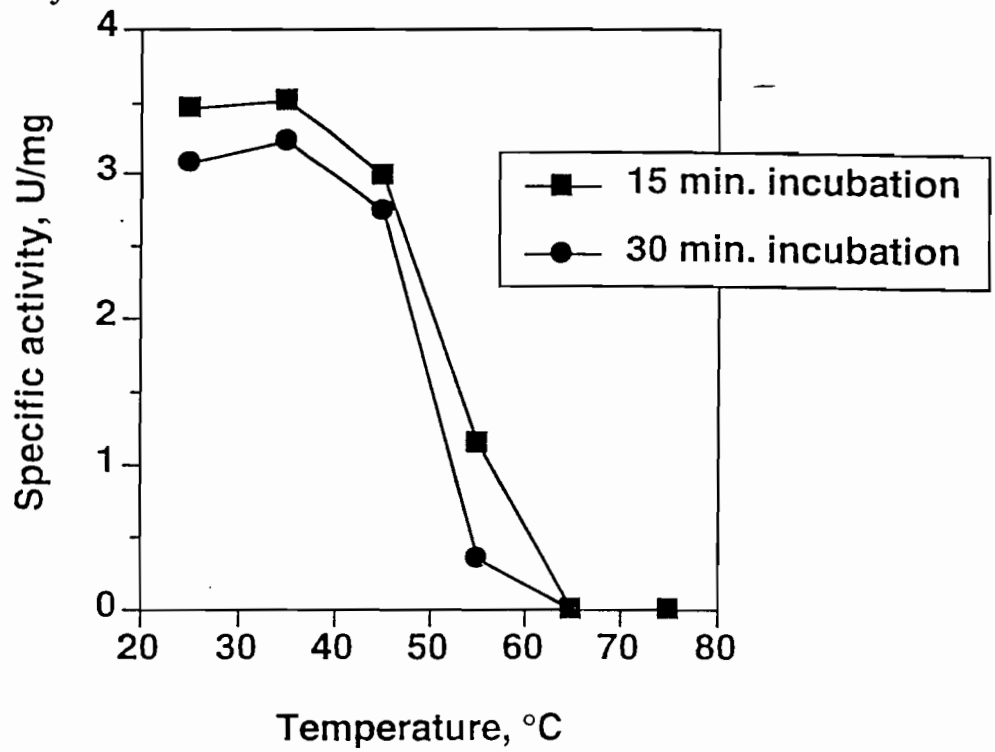


Figure 11. Temperature stability of CP4 EPSPS. Purified CP4 EPSPS from pMON17101 in storage buffer was diluted (1:1) with 50 mM HEPES, pH 7.0 (dilution buffer) to a concentration of 1.0 mg/ml and incubated at the indicated temperature. Aliquots (5 µl) of the incubation mixture were withdrawn at 15 min and 30 min after incubation initation, diluted with 27.5 µl of dilution buffer, and 10 µl was assayed in duplicate by the standard HPLC radioassay.



Appendix A. Amino acid sequence of CP4 EPSPS utilized for homology lineups (pMON17101)

Metalacp4.Pep Length: 455 November 22, 1993 09:18 Type: P Check: 7268

```
   1  MAHGASSRPA TARKSSGLSG TVRIPGDKSI SHRSFMFGGL ASGETRITGL
  51  LEGEDVINTG KAMQAMGARI RKEGDTWIID GVGNGGLLAP EAPLDFGNAA
 101  TGCRLTMGLV GVYDFDSTFI GDASLTKRPM GRVLNPLREM GVQVKSEDGD
 151  RLPVTLRGPK TPTPITYRVP MASAQVKSAV LLAGLNTPGI TTVIEPIMTR
 201  DHTEKMLQGF GANLTVETDA DGVRTIRLEG RGKLTGQVID VPGDPSSTAF
 251  PLVAALLVPG SDVTILNVLM NPTRTGLILT LQEMGADIEV-INPRLAGGED
 301  VADLRVRSST LKGVTVPEDR APSMIDEYPI LAVAAAFAEG ATVMNGLEEL
 351  RVKESDRLSA VANGLKLNGV DCDEGETSLV VRGRPDGKGL GNASGAAVAT
 401  HLDHRIAMSF LVMGLVSENP VTVDDATMIA TSFPEFMDLM AGLGAKIELS
 451  DTKAA
```

Summary for whole sequence:

Molecular weight = 47572.05 Residues = 455
Average Residue Weight = 104.554 Charged = -11
Isoelectric point = 4.99
Extinction coefficient = 9650

Residue	Number	Mole Percent
A = Ala	45	9.890
B = Asx	0	0.000
C = Cys	2	0.440
D = Asp	30	6.593
E = Glu	25	5.495
F = Phe	11	2.418
G = Gly	53	11.648
H = His	5	1.099
I = Ile	23	5.055
K = Lys	16	3.516
L = Leu	46	10.110
M = Met	19	4.176
N = Asn	14	3.077
P = Pro	25	5.495
Q = Gln	6	1.319
R = Arg	28	6.154
S = Ser	28	6.154
T = Thr	37	8.132
V = Val	38	8.352
W = Trp	1	0.220
Y = Tyr	3	0.659
Z = Glx	0	0.000
A + G	98	21.538
S + T	65	14.286
D + E	55	12.088
D + E + N + Q	75	16.484
H + K + R	49	10.769
D + E + H + K + R	104	22.857
I + L + M + V	126	27.692
F + W + Y	15	3.297

Appendix B. Amino acid sequence of mature soybean EPSPS utilized for homology lineup with CP4 EPSPS

Mature soybean epsps

Msepsp.Pep Length: 447 June 27, 1991 15:06 Check: 559 ..

```
  1  KPSTSPEIVL EPIKDFSGTI TLPGSKSLSN RILLLAALSE GTTVVDNLLY
 51  SEDIHYMLGA LRTLGLRVED DKTTKQAIVE GCGGLFPTSK ESKDEINLFL
101  GNAGTAMRPL TAAVVAAGGN ASYVLDGVPR MRERPIGDLV AGLKQLGADV
151  DCFLGTNCPP VRVNGKGGLP GGKVKLSGSV SSQYLTALLM AAPLALGDVE
201  IEIVDKLISV PYVEMTLKLM ERFGVSVEHS GNWDRFLVHG GQKYKSPGNA
251  FVEGDASSAS YLLAGAAITG GTITVNGCGT SSLQGDVKFA EVLEKMGAKV
301  TWSENSVTVS GPPRDFSGRK VLRGIDVNMN KMPDVAMTLA VVALFANGPT
351  AIRDVASWRV KETERMIAIC TELRKLGATV EEGPDYCVIT PPEKLNVTAI
401  DTYDDHRMAM AFSLAACGDV PVTIKDPCTR KTFPDYFEVL ERLTKH*
```

Appendix C. GAP of CP4 EPSPS versus mature soybean EPSPS

```
        Gap Weight:  3.000      Average Match:  0.540
    Length Weight:  0.100   Average Mismatch: -0.396

           Quality:  228.6             Length:    472
             Ratio:  0.511               Gaps:     13
Percent Similarity: 51.163   Percent Identity: 26.047

Msepsp2.Pep x Metalacp4.Pep November 22, 1993  12:28  ..

      1 ..KPSTSPEIVLEPIKDFSGTITLPGSKSLSNRILLLAALSEGTTVVDNL 48
        ....|.. .  . .::|||: :||.||:|:| :::::|..|.| :..|
      1 MAHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGL 50

     49 LYSEDIHYMLGALRTLGLRVEDDKTTKQAIVEGCGGLFPTSKESKDEINL 98
        | :||:     |:..:| |: .: .|.   . | |||::.      |  |
     51 LEGEDVINTGKAMQAMGARIRKEGDTWIIDGVGNGGLLAP......EAPL 94

     99 FLGNAGTAMRPLTAAVVAAGGNASYVLDGVPRMRERPIGDLVAGLKQLGA 148
         :|||:|: | || ::|:. : .| .: | :.: .||:| ::..|:::|.
     95 DFGNAATGCR.LTMGLVGVYDFDS.TFIGDASLTKRPMGRVLNPLREMGV 142

    149 DVDCFLGTNCPPVRVNGKGGLPGGKVKLSGSVSSQYLTALLMAAPLALGD 198
        :|.:   .:...:.|. :|. ....:....:.|. :.. :: |.|. .:
    143 QVKS...EDGDRLPVTLRGPKTPTPITYRVPMASAQVKSAVLLAGLNTPG 189

    199 VEIEIVDKLISVPYVEMTLKLMERFG..VSVEHSGNWDRFLVHGGQKYKS 246
        :.. | . :..    : | |::: ||  :.|| .::  | :  :|.   .
    190 ITTVIEPIMTR....DHTEKMLQGFGANLTVETDADGVRTIRLEGRGKLT 235

    247 PGNAFVEGDASSASY.LLAGAAITGGTITVNGCGTSSLQGDVKFAEVLEK 295
        ..   |.||:||..: |:|:  :.|:.:|: ..  .. ..:: :.  |:.
    236 GQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGLILT..LQE 283

    296 MGAKVTWSENSVTVSGPPRDFSGR.KVLRGIDVNMNKMP...DVAMTLAV 341
        |||.:.  : .:.:.::.. |:..| ..|:|:.|  :: |   |   .|||
    284 MGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAV 333

    342 VALFANGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCVITP 391
        .| ||:|:|.:.:::..:||||.:|: |:..:|:  |...:||..  |: .
    334 AAAFAEGATVMNGLEELRVKESDRLSAVANGLKLNGVDCDEGETSLVVRG 383

    392 P......EKLNVTAIDTYDDHRMAMAF.SLAACGDVPVTIKDPCTRKT.F 433
        .      :. ...|:.|. |||:||.|  :: .:: |||:.|:.   | |
    384 RPDGKGLGNASGAAVATHLDHRIAMSFLVMGLVSENPVTVDDATMIATSF 433

    434 PDYFEVLERLTKH*........ 447
        |:::::::. |. .
    434 PEFMDLMAGLGAKIELSDTKAA 455
```

Appendix D. Amino acid sequence of mature corn EPSPS utilized for homology lineup with CP4 EPSPS

Mzmepsp.Pep Length: 445 June 27, 1991 15:08 Check: 9577 ..

```
   1  AGAEEIVLQP IKEISGTVKL PGSKSLSNRI LLLAALSEGT TVVDNLLNSE
  51  DVHYMLGALR TLGLSVEADK AAKRAVVVGC GGKFPVEDAK EEVQLFLGNA
 101  GTAMRPLTAA VTAAGGNATY VLDGVPRMRE RPIGDLVVGL KQLGADVDCF
 151  LGTDCPPVRV NGIGGLPGGK VKLSGSISSQ YLSALLMAAP LPLGDVEIEI
 201  IDKLISIPYV EMTLRLMERF GVKAEHSDSW DRFYIKGGQK YKSPKNAYVE
 251  GDASSASYFL AGAAITGGTV TVEGCGTTSL QGDVKFAEVL EMMGAKVTWT
 301  ETSVTVTGPP REPFGRKHLK AIDVNMNKMP DVAMTLAVVA LFADGPTAIR
 351  DVASWRVKET ERMVAIRTEL TKLGASVEEG PDYCIITPPE KLNVTAIDTY
 401  DDHRMAMAFS LAACAEVPVT IRDPGCTRKT FPDYFDVLST FVKN*
```

Appendix E. GAP of CP4 EPSPS versus mature corn EPSPS

```
      Gap Weight:  3.000      Average Match:  0.540
   Length Weight:  0.100   Average Mismatch: -0.396

         Quality:  226.2            Length:    469
           Ratio:  0.508              Gaps:     11
Percent Similarity: 48.492   Percent Identity: 24.130

Mzmepsp.Pep x Metalacp4.Pep November 22, 1993  11:16  ..

     1 ....AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNL 46
         |:. . . .. .::||||::||.||:|:| :::::|..|.| :..|
     1 MAHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGL 50

    47 LNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLF 96
       |::|||     |:..:| .:  : ...   .|| || ::.|.: :
    51 LEGEDVINTGKAMQAMGARIRKEGDTWIIDGVGNGGLLAPEAPLD..... 95

    97 LGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGAD 146
       :|||:|: |   : |.. : :.|::  | :.: .||:| :: .|:::|.:
    96 FGNAATGCRLTMGLVGVYDFDSTFI..GDASLTKRPMGRVLNPLREMGVQ 143

   147 VDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYLSALLMAAPLPLGDV 196
       |.:   .|.-..:.|.  |. ....:.....:.|. :.. :: |.|  .::
   144 VKS...EDGDRLPVTLRGPKTPTPITYRVPMASAQVKSAVLLAGLNTPGI 190

   197 EIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQKYKSPKN 246
       .. | . :.. .  .|   : ..:.|..: .: :. : :.| .|..:.
   191 TTVIEPIMTRDHTEKMLQGFGANLTVETDADGVRT.IRLEGRGKLTGQVI 239

   247 AYVEGDASSASY.FLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGA 295
       . |.||:||..: ::|:  :.|:.||: ..  .. ..:: :.  |: |||
   240 D.VPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGLILT..LQEMGA 286

   296 KVIWTETSVTVTGPPREPFGR.KHLKAIDVNMNKMP...DVAMTLAVVAL 341
       .:. .:..:...:.. :  .| . ||::.|  :: |   |   .|||.|
   287 DIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAA 336

   342 FADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP.. 389
       ||:|:|.:.::..:||||.:|: |: .:|.  |...:||..  :: ..
   337 FAEGATVMNGLEELRVKESDRLSAVANGLKLNGVDCDEGETSLVVRGRPD 386

   390 ....EKLNVTAIDTYDDHRMAMAF.SLAACAEVPVTIRDPGCTRKTFPDY 434
           :. ...|:.|. |||:||.|  :: ..| |||: |:. . ..||::
   387 GKGLGNASGAAVATHLDHRIAMSFLVMGLVSENPVTVDDATMIATSFPEF 436

   435 FDVLSTFVKN*........ 445
       :|::..:. .
   437 MDLMAGLGAKIELSDTKAA 455
```

Appendix F. Amino acid sequence of mature petunia EPSPS utilized for homology lineup with CP4 EPSPS

Mpepsp.Pep Length: 445 February 28, 1989 15:34 Check: 2981 ..

```
   1  KPSEIVLQPI KEISGTVKLP GSKSLSNRIL LLAALSEGTT VVDNLLSSDD
  51  IHYMLGALKT LGLHVEEDSA NQRAVVEGCG GLFPVGKESK EEIQLFLGNA
 101  GTAMRPLTAA VTVAGGNSRY VLDGVPRMRE RPISDLVDGL KQLGAEVDCF
 151  LGTKCPPVRI VSKGGLPGGK VKLSGSISSQ YLTALLMAAP LALGDVEIEI
 201  IDKLISVPYV EMTLKLMERF GISVEHSSSW DRFFVRGGQK YKSPGKAFVE
 251  GDASSASYFL AGAAVTGGTI TVEGCGTNSL QGDVKFAEVL EKMGAEVTWT
 301  ENSVTVKGPP RSSSGRKHLR AIDVNMNKMP DVAMTLAVVA LYADGPTAIR
 351  DVASWRVKET ERMIAICTEL RKLGATVEEG PDYCIITPPE KLNVTDIDTY
 401  DDHRMAMAFS LAACADVPVT INDPGCTRKT FPNYFDVLQQ YSKH*
```

Appendix G. GAP of CP4 EPSPS versus mature petunia EPSPS

```
       Gap Weight:  3.000      Average Match:  0.540
    Length Weight:  0.100   Average Mismatch: -0.396

          Quality:  227.3             Length:    471
            Ratio:  0.511               Gaps:     11
Percent Similarity: 50.117   Percent Identity: 23.310

Mpepsp.Pep x Metalacp4.Pep December 1, 1993  08:04  ..

                                                    —
                .         .         .         .         .
      1 .....KPSEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNL 45
            .... . .. .::||||::||.||:|:| :::::|..|.| :..|
      1 MAHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGL 50
          .         .         .         .         .
     46 LSSDDIHYMLGALKTLGLHVEEDSANQRAVVEGCGGLFPVGKESKEEIQL 95
        |.::|:     |:..:| :: .::..    . | |||::      .| .|
     51 LEGEDVINTGKAMQAMGARIRKEGDTWIIDGVGNGGLLA......PEAPL 94
                .         .         .         .         .
     96 FLGNAGTAMRPLTAAVTVAGGNSRYVLDGVPRMRERPISDLVDGLKQLGA 145
         :|||:|: |   : |.| : :| ::  | :.: .||:: :::.|:::|.
     95 DFGNAATGCRLTMGLVGVYDFDSTFI..GDASLTKRPMGRVLNPLREMGV 142
                .         .         .         .         .
    146 EVDCFLGTKCPPVRIVSKGGLPGGKVKLSGSISSQYLTALLMAAPLALGD 195
        :|.:  |.:   :.:. :|. ....:.....:.|. :.. :: |.|. .:
    143 QVKSEDGDR...LPVTLRGPKTPTPITYRVPMASAQVKSAVLLAGLNTPG 189
                .         .         .         .         .
    196 VEIEIIDKLISVPYVEMTLKLMERFGISVEHSSSWDRFFVRGGQKYKSPG 245
        :.. | . :.. .  .|   : ..:.:..: .: :.   :| :.: | .|
    190 ITTVIEPIMTRDHTEKMLQGFGANLTVETDADGVRT...IRLEGRGKLTG 236
                .         .         .         .         .
    246 KAF.VEGDASSASY.FLAGAAVTGGTITVEGCGTNSLQGDVKFAEVLEKM 293
        ..: |.||:||..: ::|:  |.|:.:|: ..  |. ..:: :.  |:.|
    237 QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGLILT..LQEM 284
                .         .         .         .         .
    294 GAEV....TWTENSVTVKGPPRSSSGRKHLRAIDVNMNKMPDVAMTLAVV 339
        ||::    .: ..: .| : . .||. | : . : .  .| |   .|||.
    285 GADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVA 334
                .         .         .         .         .
    340 ALYADGPTAIRDVASWRVKETERMIAICTELRKLGATVEEG.PDYCIITP 388
        | :|:|:|.:.:::..:||||.:|: |:..:|:  |...:|| .....: ..
    335 AAFAEGATVMNGLEELRVKESDRLSAVANGLKLNGVDCDEGETSLVVRGR 384
                .         .         .         .         .
    389 PEKLNVTD.....IDTYDDHRMAMAF.SLAACADVPVTINDPGCTRKTFP 432
        |:  .:.:     :.|. |||:||.|  :: ..: |||::|:. . ..||
    385 PDGKGLGNASGAAVATHLDHRIAMSFLVMGLVSENPVTVDDATMIATSFP 434
                .         .
    433 NYFDVLQQYSKH*........ 445
        :::|::...: .
    435 EFMDLMAGLGAKIELSDTKAA 455
```

Appendix H. Amino acid sequence of *E. coli* EPSPS utilized for homology lineup with CP4 EPSPS

TRANSLATE of: ecoaroa.ba check: 8829 from: 1 to: 1284
generated symbols 1 to: 428.

E.coli aroA gene for 5-enolpyruvylshikimate 3-phosphate synthase (EPSP synthase), complete cds. 9/89

ec.pep Length: 428 October 19, 1989 15:41 Check: 6927 ..

```
  1  MESLTLQPIA RVDGTINLPG SKTVSNRALL LAALAHGKTV LTNLLDSDDV

 51  RHMLNALTAL GVSYTLSADR TRCEIIGNGG PLHAEGALEL FLGNAGTAMR

101  PLAAALCLGS NDIVLTGEPR MKERPIGHLV DALRLGGAKI TYLEQENYPP

151  LRLQGGFTGG NVDVDGSVSS QFLTALLMTA PLAPEDTVIR IKGDLVSKPY

201  IDITLNLMKT FGVEIENQHY QQFVVKGGQS YQSPGTYLVE GDASSASYFL

251  AAAAIKGGTV KVTGIGRNSM QGDIRFADVL EKMGATICWG DDYISCTRGE

301  LNAIDMDMNH IPDAAMTIAT AALFAKGTTR LRNIYNWRVK ETDRLFAMAT

351  ELRKVGAEVE EGHDYIRITP PEKLNFAEIA TYNDHRMAMC FSLVALSDTP

401  VTILDPKCTA KTFPDYFEQL ARISQAA*
```

Appendix I. GAP of CP4 EPSPS versus *E. coli* EPSPS

```
        Gap Weight:  3.000      Average Match:  0.540
    Length Weight:  0.100   Average Mismatch: -0.396

           Quality:  235.7             Length:    460
             Ratio:  0.551              Gaps:      9
Percent Similarity: 52.246   Percent Identity: 26.005

E coli.Pep x Metalacp4.Pep    November 22, 1993  10:18  ..

                                                  —
                .         .         .         .         .
      1 ......MESLTLQPIARVDGTINLPGSKTVSNRALLLAALAHGKTVLTNL 44
                . | .. . :.||:.:||.|.:|:|.:::::|| |.| :|.|
      1 MAHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGL 50
                .         .         .         .         .
     45 LDSDDVRHMLNALTALGVSYTLSADRTRCEIIGNGGPLHAEGALELFLGN 94
        |:::|| :  .|: |:|...  .:|   .: :|||| | :|::|:  :||
     51 LEGEDVINTGKAMQAMGARIRKEGDTWIIDGVGNGGLLAPEAPLD..FGN 98
                .         .         .         .         .
     95 AGTAMRPLAAALCLGSNDIVLTGEPRMKERPIGHLVDALRLGGAKITYLE 144
        |:|: |   : :.: . | .:.|::.:..||:|:::::||  |..:.  |
     99 AATGCRLTMGLVGVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVK.SE 147
                .         .         .         .         .
    145 QENYPPLRLQGGFTGGNVDVDGSVSSQFLTALLMTAPLAPEDTVIRIKGD 194
        :::  |: |.|. |.. :.  ..:.|. :.. :: |.|...:... |..
    148 DGDRLPVTLRGPKTPTPITYRVPMASAQVKSAVLLAGLNTPGITTVIEPI 197
                .         .         .         .         .
    195 LVSKPYIDITLNLMKTFGVEIENQHYQQFVVKGGQSYQSPGTYLVEGDAS 244
        :....  .:  .:  .:.||.:.:  . : :.| ... :... :.::..|
    198 MTRDHTEKMLQGFGANLTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSS 247
                .         .         .         .         .
    245 SASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATI....... 287
        .|  ::||  :.|:.|.: .: .|. ..:: :.  |:.|||.|
    248 TAFPLVAALLVPGSDVTILNVLMNPTRTGLILT..LQEMGADIEVINPRL 295
                .         .         .         .         .
    288 CWGDDY..ISCTRGELNAIDMDMNHIP...DAAMTIATAALFAKGTTRLR 332
        . |:|   :.. .:.|.::.:. :: |   |.  .:|.|| ||.|.| :.
    296 AGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAEGATVMN 345
                .         .         .         .         .
    333 NIYNWRVKETDRLFAMATELRKVGAEVEEGHDYIRI.TPPEKLNF..... 376
        .: ::|||| .|| |:|.:|:  |.:.:||.. : : ..|:  .:
    346 GLEELRVKESDRLSAVANGLKLNGVDCDEGETSLVVRGRPDGKGLGNASG 395
                .         .         .         .         .
    377 AEIATYNDHRMAMCFSLVAL.SDTPVTILDPKCTAKTFPDYFEQLARISQ 425
        |.:||. |||:||:| :::| |:.|||: |:. .|..||:::: :| ::.
    396 AAVATHLDHRIAMSFLVMGLVSENPVTVDDATMIATSFPEFMDLMAGLGA 445
                .
    426 AA*..... 428

    446 KIELSDTKAA 455
```

Appendix J. Amino acid sequence of *B. subtilis* EPSPS utilized for homology lineup with CP4 EPSPS

Bsubepsps.pep

P20691 Length: 428 May 21, 1992 08:49 Type: P Check: 3714 ..

```
   1  MKRDKVQTLH GEIHIPGDKS ISHRSVMFGA LAAGTTTVKN FLPGADCLST
  51  IDCFRKMGVH IEQSSSDVVI HGKGIDALKE PESLLDVGNS GTTIRLMLGI
 101  LAGRPFYSAV AGDESIAKRP MKRVTEPLKK MGAKIDGRAG GEFTPLSVSG
 151  ASLKGIDYVS PVASAQIKSA VLLAGLQAEG TTTVTEPHKS RDHTERMLSA
 201  FGVKLSEDQT SVSIAGGQKL TAADIFVPGD ISSAAFFLAA GAMVPNSRIV
 251  LKNVGLNPTR TGIIDVLQNM GAKLEIKPSA DSGAEPYGDL IIETSSLKAV
 301  EIGGDIIPRL IDEIPIIALL ATQAEGTTVI KDAAELKVKE TNRIDTVVSE
 351  LRKLGAEIEP TADGMKVYGK QTLKGGAAVS SHGDHRIGMM LGIASCITEE
 401  PIEIEHTDAI HVSYPTFFEH LNKLSKKS
```

Appendix K. GAP of CP4 EPSPS versus *B. subtilis* EPSPS

```
        Gap Weight:  3.000      Average Match:  0.540
     Length Weight:  0.100   Average Mismatch: -0.396

           Quality:  333.7             Length:    455
             Ratio:  0.780               Gaps:      2
Percent Similarity: 59.346   Percent Identity: 41.121

Bsubepsp.Pep x Metalacp4.Pep November 22, 1993  11:56  ..
                                                    —

                .         .         .         .         .
   1 .........MKRDKVQTLHGEIHIPGDKSISHRSVMFGALAAGTTTVKNF  41
             .  |  .| |.::||||||||||||.|||:||.|.| :..:
   1 MAHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGL  50
                .         .         .         .         .
  42 LPGADCLSTIDCFRKMGVHIEQSSSDVVIHGKGIDALKEPESLLDVGNSG  91
     |.|.|.:.| ..:. ||.:| ..:.. :|.| | ::| .||. ||.||.:
  51 LEGEDVINTGKAMQAMGARIRKEGDTWIIDGVGNGGLLAPEAPLDFGNAA 100
                .         .         .         .         .
  92 TTIRLMLGILAGRPFYSAVAGDESIAKRPMKRVTEPLKKMGAKIDGRAGG 141
     |..|| :|:::. .| |.. ||.|:.|||| || :||:.||..:.: .|:
 101 TGCRLTMGLVGVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSEDGD 150
                .         .         .         .         .
 142 EFTPLSVSGASLKGIDYVSPVASAQIKSAVLLAGLQAEGTTTVTEPHKSR 191
      :..   :. . ..|.|  |:||||:||||||||...|.|||.|| ..|
 151 RLPVTLRGPKTPTPITYRVPMASAQVKSAVLLAGLNTPGITTVIEPIMTR 200
                .         .         .         .         .
 192 DHTERMLSAFGVKL.....SEDQTSVSIAGGQKLTAADIFVPGDISSAAF 236
     ||||:|| :||..|     .::  .:.:.| .|||:. | |||| ||.||
 201 DHTEKMLQGFGANLTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAF 250
                .         .         .         .         .
 237 FLAAGAMVPNSRIVLKNVGLNPTRTGIIDVLQNMGAKLEIKPSADSGAEP 286
      |.|: :||.| :.: || :|||||:| .||:|||.:|:  .  .|:|.
 251 PLVAALLVPGSDVTILNVLMNPTRTGLILTLQEMGADIEVINPRLAGGED 300
                .         .         .         .         .
 287 YGDLIIETSSLKAVEIGGDIIPRLIDEIPIIALLATQAEGTTVIKDAAEL 336
      :|| : .|.||:|.:.:|  |.:|||.||:|: |. |||.||:.: .||
 301 VADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAEGATVMNGLEEL 350
                .         .         .         .         .
 337 KVKETNRIDTVVSELRKLGAEIEPTADGMKVYGKQTLKG.....GAAVSS 381
     :|||.:|:..|..:|:  |.:.:....:: | |:.. ||     ||||..
 351 RVKESDRLSAVANGLKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVAT 400
                .         .         .         .         .
 382 HGDHRIGMMLGIASCITEEPIEIEHTDAIHVSYPTFFEHLNKLSKKS... 428
     | ||||:| : : : :.|:|:.:::... | .|:|.|:: :. |: |
 401 HLDHRIAMSFLVMGLVSENPVTVDDATMIATSFPEFMDLMAGLGAKIELS 450
                                  .
                                  .
                                  .
```

Appendix L. Amino acid sequence of yeast EPSPS utilized for homology lineup with CP4 EPSPS

Yeast Aro I gene Aro A (EPSP synthase) region. Starts at base 1481 of sequence obtained from Searle. Entered 1/26/85 by C. Gasser

yepsp.pep Length: 469 25-SEP-1987 13:52 Check: 3990 ..

```
  1  TLVYPFKDIP ADQQKVVIPP GSKSISNRAL ILAALGEGQC KIKNLLHSDD
 51  TKHMLTAVHE LKGATISWED NGETVVVEGH GGSTLSACAD PLYLGNAGTA
101  SRFLTSLAAL VNSTSSQKYI VLTGNARMQQ RPIAPLVDSL RANGTKIEYL
151  NNEGSLPIKV YTDSVFKGGR IELAATVSSQ YVSSILMCAP YAEEPVTLAL
201  VGGKPISKLY VDMTIKMMEK FGINVETSTT EPYTYYIPKG HYINPSEYVI
251  ESDASSATYP LAFAAMTGTT VTVPNIGFES LQGDARFARD VLKPMGCKIT
301  QTATSTTVSG PPVGTLKPLK HVDMEPMTDA FLTACVVAAI SHDSDPNSAN
351  TTTIEGIANQ RVKECNRILA MATELAKFGV KTTELPDGIQ VHGLNSIKDL
401  KVPSDSSGPV GVCTYDDHRV AMSFSLLAGM VNSQNERDEV ANPVRILERH
451  CTGKTWPGWW DVLHSELGA
```

Appendix M. GAP of CP4 EPSPS versus yeast EPSPS

```
        Gap Weight:  3.000      Average Match:  0.540
    Length Weight:  0.100   Average Mismatch: -0.396

           Quality:  227.8             Length:    496
             Ratio:  0.501               Gaps:     18
Percent Similarity: 53.505   Percent Identity: 29.907

Yepsp.Pep x Metalacp4.Pep December 1, 1993  08:47  ..

                .         .         .         .         .
  1 TLVYPFKDIPADQQK......VVIPPGSKSISNRALILAALGEGQCKIKN 44
    :... .. ||...|      .|  ||.||||:|.:::::|:.|:.:|..
  1 .MAHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITG 49
          .         .         .         .         .
 45 LLHSDDTKHMLTAVHELKGATISWEDNGETVVVEGHGGSTLSACADPLYL 94
    ||.::|. :  .|::.: || |..|  |:| :::| |.:.| |...|| :
 50 LLEGEDVINTGKAMQAM.GARIRKE..GDTWIIDGVGNGGLLAPEAPLDF 96
                .         .         .         .         .
 95 GNAGTASRFLTSLAALVNSTSSQKYIVLTGNARMQQRPIAPLVDSLRANG 144
    |||:|::|:  :|.:: : .|.      :.|:|.: .||::.:::.||. |
 97 GNAATGCRLTMGLVGVYDFDST.....FIGDASLTKRPMGRVLNPLREMG 141
                .         .         .         .         .
145 TKIEYLNNEGSLPIKVYTDSVFKGGRIELAATVSSQYVSSILMCAPYAEE 194
    ..:   ..|:: .:.|   :. ....|.. ..:.|. |.| :: |.....
142 VQV...KSEDGDRLPVTLRGPKTPTPITYRVPMASAQVKSAVLLAGLNTP 188
                .         .         .         .         .
195 PVTLALVGGKPISKLYVDMTIKMMEKFGIN..VETSTTEPYTYYIPKGHY 242
    .:| .:   .||  :  | | ||:: || |  |||...:. |. :.
189 GITTVI...EPI..MTRDHTEKMLQGFGANLTVETDADGVRTIRLEGRGK 233
                .         .         .         .         .
243 INPSEYVIESDASSATYPLAFAAM.TGTTVTVPNIGFESLQGDARFARDV 291
    :..  . :.:|:||..:||. | : .|..||: |: ::. ..:  :.
234 LTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGLILT... 280
                .         .         .         .         .
292 LKPMGCKI....TQTATSTTVSGPPV..GTLKPLKHVD..MEPMTDAFLT 333
    |..||..|    .. |.:..|.: .|  :|||.:. .:   ..|.|.: .
281 LQEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI 330
                .         .         .         .         .
334 ACVVAAISHDSDPNSANTTTIEGIANQRVKECNRILAMATELAKFGVKTT 383
     .|.||:..:       .|.::|:.: ||||::|: |:|.:|   ||...
331 LAVAAAFAEG.......ATVMNGLEELRVKESDRLSAVANGLKLNGVDCD 373
                .         .         .         .         .
384 ELPDGIQVHGLNSIKDLKVPSDSSGPVGVCTYDDHRVAMSFSLLAGMVNS 433
    | ..:: |:|  . |:|  :|:.    :|.|. |||:|||| |: |:|.
374 EGETSLVVRGRPDGKGLGNASGA....AVATHLDHRIAMSF.LVMGLVS. 417
                .         .         .         .
434 QNERDEVANPVRILERHCTGKTWPGWWDVLHSELGA.......... 469
           .||| : :   .:..:|:: |:: .:|||
418 .......ENPVTVDDATMIATSFPEFMDLM.AGLGAKIELSDTKAA 455
```

Appendix N. GAP of mature soybean EPSPS vs. *B. subtilis* EPSPS

```
          Quality:  243.6              Length:    456
             Ratio:  0.569                Gaps:     16
 Percent Similarity: 55.609   Percent Identity: 30.072

 Msepsp2.Pep x Bsubepsp.Pep December 2, 1993  10:44  ..


      1 KPSTSPEIVLEPIKDFSGTITLPGSKSLSNRILLLAALSEGTTVVDNLLY 50
        :  :.:..: |.| :||.||:|:| ::::||..|||.+.|:|
      1 .......MKRDKVQTLHGEIHIPGDKSISHRSVMFGALAAGTTTVKNFLP 43

     51 SEDIHYMLGALRTLGLRVEDDKTTKQAIVEGCGGLFPTSKESKDEINLF. 99
        :.|.   ::.:|.:|:::|:...    ..|.::|       |:. .|:
     44 GADCLSTIDCFRKMGVHIEQSSSDVVIHGKGIDAL.......KEPESLLD 86

    100 LGNAGTAMRPLTAAVVAAGGNASYVLDGVPRMRERPIGDLVAGLKQLGAD 149
        :||.||.:| |  :::|: .  | | .| ..: .||:  :...||.:||.
     87 VGNSGTTIR.LMLGILAGRPFYSAV.AGDESIAKRPMKRVTEPLKKMGAK 134

    150 VDCFLGTNCPPVRVNGKGGLPGGKVKLSGSVSSQYLTALLMAAPLALGDV 199
        :|.  |.: .|:.|.| ::|.| .  :|. .|.|. .|:|:|:  | |..
    135 IDGRAGGEFTPLSVSG.ASLKGIDY.VSPVASAQIKSAVLLAGLQAEGTT 182

    200 EIEIVDKLISVPYVEMTLKLMERFGVSVEHSGNWDRFLVHGGQKYKSPGN 249
        .:. ..|       : | :::. |||.:  |:: ... : ||||. .:::
    183 TVTEPHK.....SRDHTERMLSAFGVKL..SEDQTSVSIAGGQKL.TAAD 224

    250 AFVEGDASSASYLL.AGAAITGGTITVNGCGTSSLQGDVKFAEVLEKMGA 298
         ||.|| |||.::| ||| :..: |.:...| .. ..::   :||:.|||
    225 IFVPGDISSAAFFLAAGAMVPNSRIVLKNVGLNPTRTGI..IDVLQNMGA 272

    299 KVTWSENSVTVSGPPRDF.SGRKVLRGIDVNMNKMPDVAMTLAVVALF.. 345
        |:. .... ...:|  |:  : . |:::::. : :| :  .::::||:
    273 KLEIKPSADSGAEPYGDLIIETSSLKAVEIGGDIIPRLIDEIPIIALLAT 322

    346 .ANGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCVITPPEK 394
         |:|.|.|:|.|.::||||:|: .:..|||||||.:|..:|   : ..:.
    323 QAEGTTVIKDAAELKVKETNRIDTVVSELRKLGAEIEPTADGMKVYGKQT 372

    395 L.NVTAIDTYDDHRMAMAFSLAAC.GDVPVTIKDP.CTRKTFPDYFEVLE 441
        | ...|:...:|||::| :::|.| .: |:.|... ..: .:|.:|| |:
    373 LKGGAAVSSHGDHRIGMMLGIASCITEEPIEIEHTDAIHVSYPTFFEHLN 422

    442 RLTKH* 447
        :|.|.
    423 KLSKKS 428
```

Appendix O. GAP of mature soybean EPSPS vs. mature corn EPSPS

```
  Gap Weight:  3.000      Average Match:  0.540
     Length Weight:  0.100   Average Mismatch: -0.396

           Quality:  582.3              Length:    448
             Ratio:  1.309                Gaps:      2
Percent Similarity: 90.315   Percent Identity: 83.108

Msepsp2.Pep x Mzmepsp.Pep December 2, 1993  10:47  ..

                .         .         .         .         .
     1 KPSTSPEIVLEPIKDFSGTITLPGSKSLSNRILLLAALSEGTTVVDNLLY 50
       ....|||||:|||::|||:.||||||||||||||||||||||||||||||
     1 ..AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLN 48
                .         .         .         .         .
    51 SEDIHYMLGALRTLGLRVEDDKTTKQAIVEGCGGLFPTSKESKDEINLFL 100
       |||:||||||||||||.||.||..|.|:| |||| ||. .:.|:|:.|||
    49 SEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV.EDAKEEVQLFL 97
                .         .         .         .         .
   101 GNAGTAMRPLTAAVVAAGGNASYVLDGVPRMRERPIGDLVAGLKQLGADV 150
       ||||||||||||||.||||||.||||||||||||||||||.|||||||||
    98 GNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGADV 147
                .         .         .         .         .
   151 DCFLGTNCPPVRVNGKGGLPGGKVKLSGSVSSQYLTALLMAAPLALGDVE 200
       ||||||:||||||||  ||||||||||||||:|||||.||||||||:||||
   148 DCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYLSALLMAAPLPLGDVE 197
                .         .         .         .         .
   201 IEIVDKLISVPYVEMTLKLMERFGVSVEHSGNWDRFLVHGGQKYKSPGNA 250
       |||:|||||:|||||||:|||||||..|||:.||||.:.|||||||| ||
   198 IEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQKYKSPKNA 247
                .         .         .         .         .
   251 FVEGDASSASYLLAGAAITGGTITVNGCGTSSLQGDVKFAEVLEKMGAKV 300
       :||||||||||:||||||||||:||:||||.||||||||||||.|||||
   248 YVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKV 297
                .         .         .         .         .
   301 TWSENSVTVSGPPRDFSGRKVLRGIDVNMNKMPDVAMTLAVVALFANGPT 350
       ||.|.||||.||||:  ||| |::|||||||||||||||||||:|||:|||
   298 TWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPT 347
                .         .         .         .         .
   351 AIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCVITPPEKLNVTAI 400
       ||||||||||||||||:|| ||| ||||.||||||||:||||||||||||
   348 AIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTAI 397
                .         .         .         .
   401 DTYDDHRMAMAFSLAACGDVPVTIKDP.CTRKTFPDYFEVLERLTKH* 447
       ||||||||||||||||::||||||:|| ||||||||||:||. :.|:|
   398 DTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN* 445
```

Appendix P. GAP of mature soybean EPSPS vs. *E. coli* EPSPS

```
     Quality:  403.0              Length:     450
               Ratio:  0.942                Gaps:       7
Percent Similarity: 71.765   Percent Identity: 54.824

Msepsp2.Pep x Ec.Pep      December 2, 1993  10:40  ..

                .         .         .         .         .
   1 KPSTSPEIVLEPIKDFSGTITLPGSKSLSNRILLLAALSEGTTVVDNLLY 50
          ..:.|:||  ..|||.|||||.:||| ||||||..|.||:.|||
   1 ....MESLTLQPIARVDGTINLPGSKTVSNRALLLAALAHGKTVLTNLLD 46
                .         .         .         .         .
  51 SEDIHYMLGALRTLGLRVE.DDKTTKQAIVEGCGGLFPTSKESKDEINLF 99
     |:|::.||.|| .||:. . ... |: .|::..|.|     ...:.::||
  47 SDDVRHMLNALTALGVSYTLSADRTRCEIIGNGGPL.....HAEGALELF 91
                .         .         .         .         .
 100 LGNAGTAMRPLTAAVVAAGGNASYVLDGVPRMRERPIGDLVAGLKQLGAD 149
     ||||||||||||.||:. ::.:  .||.| |||:||||||.||.:|:  ||.
  92 LGNAGTAMRPLAAALCLGSND..IVLTGEPRMKERPIGHLVDALRLGGAK 139
                .         .         .         .         .
 150 VDCFLGTNCPPVRVNGKGGLPGGKVKLSGSVSSQYLTALLMAAPLALGDV 199
     :.:: ..|:||:|:  .||:.||.|.:.||||||:||||||.|||| :|.
 140 ITYLEQENYPPLRL..QGGFTGGNVDVDGSVSSQFLTALLMTAPLAPEDT 187
                .         .         .         .         .
 200 EIEIVDKLISVPYVEMTLKLMERFGVSVEHSGNWDRFLVHGGQKYKSPGN 249
      | | :.|:| ||:::||.||. |||.:|: .::::.|:|.|||.|.|||.
 188 VIRIKGDLVSKPYIDITLNLMKTFGVEIEN.QHYQQFVVKGGQSYQSPGT 236
                .         .         .         .         .
 250 AFVEGDASSASYLLAGAAITGGTITVNGCGTSSLQGDVKFAEVLEKMGAK 299
      :||||||||||:||:|||.|||:.|.|.| .|:|||::||:|||||||.
 237 YLVEGDASSASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGAT 286
                .         .         .         .         .
 300 VTWSENSVTVSGPPRDFSGRKVLRGIDVNMNKMPDVAMTLAVVALFANGP 349
     :.|::: :.         :.|  |.:||::||.:||.|||:|..||||.|.
 287 ICWGDDYIS........CTRGELNAIDMDMNHIPDAAMTIATAALFAKGT 328
                .         .         .         .         .
 350 TAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCVITPPEKLNVTA 399
     | :|:: .||||||:|::|:.|||||:||.||||.||. ||||||||...
 329 TRLRNIYNWRVKETDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAE 378
                .         .         .         .         .
 400 IDTYDDHRMAMAFSLAACGDVPVTIKDP.CTRKTFPDYFEVLERLTKH*. 447
     |.||:||||||.|||.| :|.|||| || || |||||||| |.|:..
 379 IATYNDHRMAMCFSLVALSDTPVTILDPKCTAKTFPDYFEQLARISQAA* 428
```

Appendix Q. GAP of *B. subtilis* EPSPS vs. *E. coli* EPSPS

```
        Gap Weight:  3.000      Average Match:  0.540
     Length Weight:  0.100   Average Mismatch: -0.396

           Quality:  236.8             Length:    447
             Ratio:  0.553               Gaps:     13
Percent Similarity: 55.501   Percent Identity: 29.095

Bsubepsp.Pep x Ec.Pep     December 1, 1993  09:27  ..

                                                   —
                .         .         .         .         .
  1 ...MKRDKVQTLHGEIHIPGDKSISHRSVMFGALAAGTTTVKNFLPGADC 47
     :. :.:. :.|.|::||.|.:|:|.::::||| |.|.:.|:|.:.|.
  1 MESLTLQPIARVDGTINLPGSKTVSNRALLLAALAHGKTVLTNLLDSDDV 50
          .         .         .         .         .
 48 LSTIDCFRKMGVHIEQSS....SDVVIHGKGIDALKEPESLLDVGNSGTT 93
     ::.:  :|| .. |.    :::: :| '.:.|  . | :|  ||.||.
 51 RHMLNALTALGVSYTLSADRTRCEIIGNGGPLHAEGALELFL..GNAGTA 98
                .         .         .         .         .
 94 IRLMLGILAGRPFYSAVAGDESIAKRPMKRVTEPLKKMGAKIDGRAGGEF 143
    :| : : |.  .   .:.|:..: .||: ::.::|:  ||||.  ..:::
 99 MRPLAAAALCLGSNDIVLTGEPRMKERPIGHLVDALRLGGAKITYLEQENY 148
                .         .         .         .         .
144 TPLSVSGASLKGIDYVSPVASAQIKSAVLL.AGLQAEGTTT.VTEPHKSR 191
    .||.: |:   |   |.. .|.|: .|:|: |.|.:|:|.. :.:.  |:
149 PPLRLQGGFTGGNVDVDGSVSSQFLTALLMTAPLAPEDTVIRIKGDLVSK 198
                .         .         .         .         .
192 ...DHTERMLSAFGVKLSED..QTSVSIAGGQKLTAADIFVPGDISSAAF 236
       | | .:.:..|||.:.::  |  |  :|.   .:.:..:|.|| |||.:
199 PYIDITLNLMKTFGVEIENQHYQQFVVKGGQSYQSPGTYLVEGDASSASY 248
                .         .         .         .         .
237 FLAAGAMVPNSRIVLKNVGLNPTRTGI..IDVLQNMGAKLEIKPSADSGA 284
    ||||:| :..: : :..:| |. ..:|   |||:.|||.:
249 FLAAAA.IKGGTVKVTGIGRNSMQGDIRFADVLEKMGATI.......... 287
                .         .         .         .         .
285 EPYGDLIIET..SSLKAVEIGGDIIPRLIDEIPIIALLATQAEGTTVIKD 332
     .:|| .|..  :.|.|:::: : ||    .|:. ||:   |.||| :::
288 .CWGDDYISCTRGELNAIDMDMNHIPDAAMTIATAALF...AKGTTRLRN 333
                .         .         .         .         .
333 AAELKVKETNRIDTVVSELRKLGAEIEPTADGMKVYGKQTLKGGAAVSSH 382
      :::||||:|: .:..||||:|||:|.. | ::: ..:.| . |.:...
334 IYNWRVKETDRLFAMATELRKVGAEVEEGHDYIRITPPEKL.NFAEIATY 382
                .         .         .
383 GDHRIGMMLGIASCITEEPIEIEHTDAIHVSYPTFFEHLNKLSKKS. 428
    .|||:.| ::::.. :.:.|:.| .....  .:|.:|||:|.::|. .
383 NDHRMAMCFSLVA.LSDTPVTILDPKCTAKTFPDYFEQLARISQAA* 428
```