Notices

Federal Register
Vol. 62, No. 40
Friday, February 28, 1997

Animal and Plant Health Inspection Service

[Docket No. 96-098-1]

Dupont Agricultural Products: Receipt of Petition for Determination of Nonregulated Status for Genetically Engineered Soybeans

AGENCY: Animal and Plant Health Inspection Service, USDA.

ACTION: Notice.

SUMMARY: We are advising the public that the Animal and Plant Health Inspection Service has received a petition from Dupont Agricultural Products seeking a determination of nonregulated status for soybeans designated as sublines G94-1, G94-19, and G168 derived from transformation event 260-05 that have been genetically engineered to produce high oleic acid oil. The petition has been submitted in accordance with our regulations concerning the introduction of certain genetically engineered organisms and products. In accordance with those regulations, we are soliciting public comments on whether these soybean sublines present a plant pest risk.

DATES: Written comments must be received on or before April 29, 1997.

ADDRESSES: Please send an original and three copies of your comments to

Docket No. 96-098-1. Regulatory Analysis and Development. PPD. APHIS. Suite 3C03. 4700 River Road Unit 118. Riverdale. MD 20737-1238. Please state that your comments refer to Docket No. 96-098-1. A copy of the petition and any comments received may be inspected at USDA. room 1141. South Building. 14th Street and Independence Avenue SW.. Washington. DC. between 8 a.m. and 4:30p.m.. Monday through Friday. except holidays. Persons wishing access to that room to inspect the petition or comments are asked to call in advance of visiting at (202) 690-2817.

FOR FURTHER INFORMATION CONTACT: Dr. Ved Malik. BSS. PPQ. APHIS. Suite 5B05. 4700 River Road Unit 147. Riverdale. MD 20737-1236: (301) 734-7612. To obtain a copy of the petition. contact Ms. Kay Peterson at (301) 734-7612: e-mail: mkpeterson@aphis.usda.gov.

SUPPLEMENTARY INFORMATION: The regulations in 7 CFR part 340. "Introduction of Organisms and Products Altered or Produced Through Genetic Engineering Which Are Plant Pests or Which There Is Reason to Believe Are Plant Pests." regulate. among other things. the introduction (importation. interstate movement. or release into the environment) of organisms and products altered or produced through genetic engineering that are plant pests or that there is reason to believe are plant pests. Such genetically engineered organisms and products are considered "regulated articles."

The regulations in § 340.6(a) provide that any person may submit a petition to the Animal and Plant Health Inspection Service (APHIS) seeking a determination that an article should not be regulated under 7 CFR part 340. Paragraphs (b) and (c) of § 340.6 describe the form that a petition for determination of nonregulated status must take and the information that must be included in the petition.

On January 8. 1997. APHIS received a petition (APHIS Petition No. 97-008-01p) from Dupont Agricultural Products (Dupont) of Wilmington. DE. requesting a determination of nonregulated status under 7 CFR part 340 for high oleic acid soybean sublines G94-1. G94-19. and G168 (sublines G94-1. G94-19. and G168) derived from transformation event 260-05. The Dupont petition states that the subject soybean sublines should not be regulated by APHIS because they do not present a plant pest risk.

As described in the petition. sublines G94-1. G94-19. and G168 have been genetically engineered to contain the GmFad 2-1 gene. which causes a coordinate silencing of itself and the endogenous GmFad 2-1 gene. Suppression of the GmFad 2-1 gene in developing soybeans prevents the addition of a second double bond to oleic acid. resulting in a greatly increased oleic acid content only in the seed. The resulting oil contains an abundance of monosaturated oleic acid (82-85%). a reduced concentration of polysaturated fatty acids. and lower palmitic acid content. While the subject soybean sublines also contain the GUS and Amp marker genes. tests indicate that these genes are not expressed in the transgenic soybean plants. The added genes were introduced into meristems of the elite soybean line A2396 by the particle bombardment method. and their expression is controlled in part by gene sequences derived from the plant pathogens *Agrobacterium tumefaciens* and cauliflower mosaic virus.

Dupont's soybean sublines G94-1. G94-19. and G168 are currently considered regulated articles under the regulations in 7 CFR part 340 because they contain gene sequences derived from plant pathogenic sources. The subject soybean sublines have been evaluated in field trials conducted since 1995 under APHIS notifications. In the process of reviewing these notifications for field trials. APHIS determined that the vectors and other elements were disarmed and that the trials. which were conducted under conditions of reproductive and physical containment or isolation. would not present a risk of plant pest introduction or dissemination.

In the Federal Plant Pest Act. as amended (7 U.S.C. 150aa *et seq.*). "plant pest" is defined as "any living stage of: Any insects. mites. nematodes. slugs. snails. protozoa. or other invertebrate animals. bacteria. fungi. other parasitic plants or reproductive parts thereof. viruses. or any organisms similar to or allied with any of the foregoing. or any infectious substances. which can directly or indirectly injure or cause disease or damage in any plants or parts thereof. or any processed. manufactured or other products of plants." APHIS views this definition very broadly. The definition covers direct or indirect injury. disease. or damage not just to agricultural crops. but also to plants in general. for example. native species. as well as to organisms that may be beneficial to plants. for example. honeybees. rhizobia. etc.

The Food and Drug Administration (FDA) published a statement of policy on foods derived from new plant varieties in the Federal Register on May 29. 1992 (57 FR 22984-23005). The FDA statement of policy includes a discussion of the FDA's authority for ensuring food safety under the Federal Food. Drug and Cosmetic Act (21 U.S.C 201 *et seq.*). and provides guidance to industry on the scientific consideration associated with the development of foods derived from new plant varieties. including those plants developed through the techniques of genetic engineering. Dupont has begun the consultative process with FDA on the subject soybean sublines.

In accordance with § 340.6(d) of the regulations. we are publishing this notice to inform the public that APHIS will accept written comments regarding the Petition for Determination of Nonregulated Status from any interested person for a period of 60 days from the date of this notice. The petition and any comments received are available for public review. and copies of the petition may be ordered (see the **ADDRESSES** section of this notice).

After the comment period closes. APHIS will review the data submitted by the petitioner. all written comments received during the comment period. and any other relevant information. Based on the available information. APHIS will furnish a response to the petitioner. either approving the petition in whole or in part. or denying the petition. APHIS will then publish a notice in the Federal Register announcing the regulatory status of Dupont's high oleic acid soybean sublines G94-1. G94-19. and G168 derived from transformation event 260-05 and the availability of APHIS' written decision.

Authority: 7 U.S.C. 150aa-150jj. 151-167. and 1622n: 31 U.S.C. 9701: 7 CFR 2.22. 2.80 and 371.2(c).

Done in Washington. DC. this 24th day of February 1997.

Terry L. Medley.

Administrator. Animal and Plant Health Inspection Service.

[FR Doc. 97-5023 Filed 2-27-97; 8:45 am]

BILLING CODE 3410-34-P

97-008-01p II

DUPONT

AGRICULTURAL PRODUCTS
Experimental Station
P.O. Box 80402
Wilmington, Delaware 19880-0402

Dr. Vedpal Malik
APHIS
Biotechnology Permits
(301)734-6774

7 February 1997

Ved,

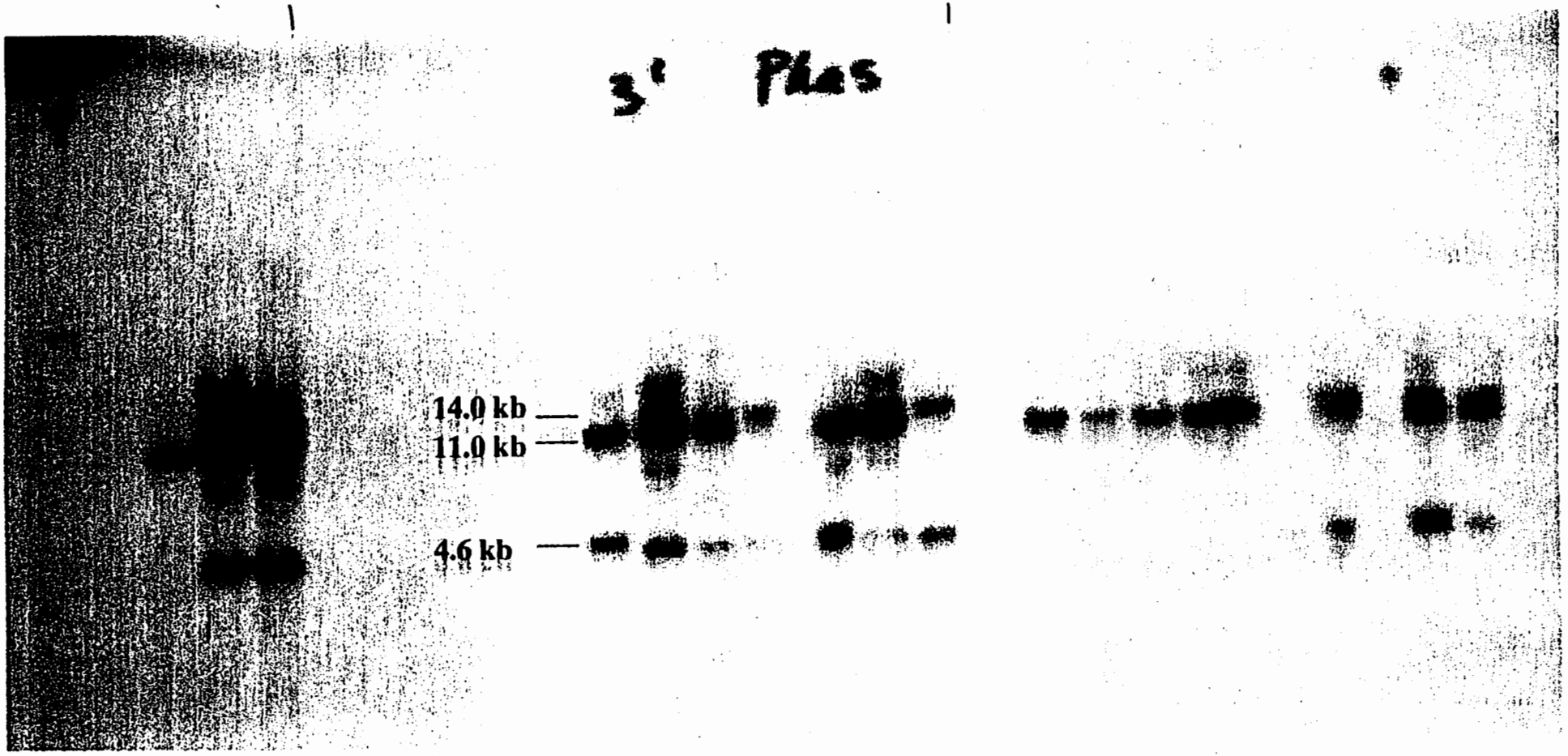
Attached is the requested information to clarify some points in the Petition for Determination of Nonregulated Status for High Oleic Acid Transgenic Soybean. We will fax the information today and send printed copies next week by overnight mail. If you have any questions or require futher clarification please do not hesitate to phone. My number is (302)695 -1241 or you can contact Ed Raleigh (302)992-6158.

Mary Locke

Mary Locke
Group Chemist

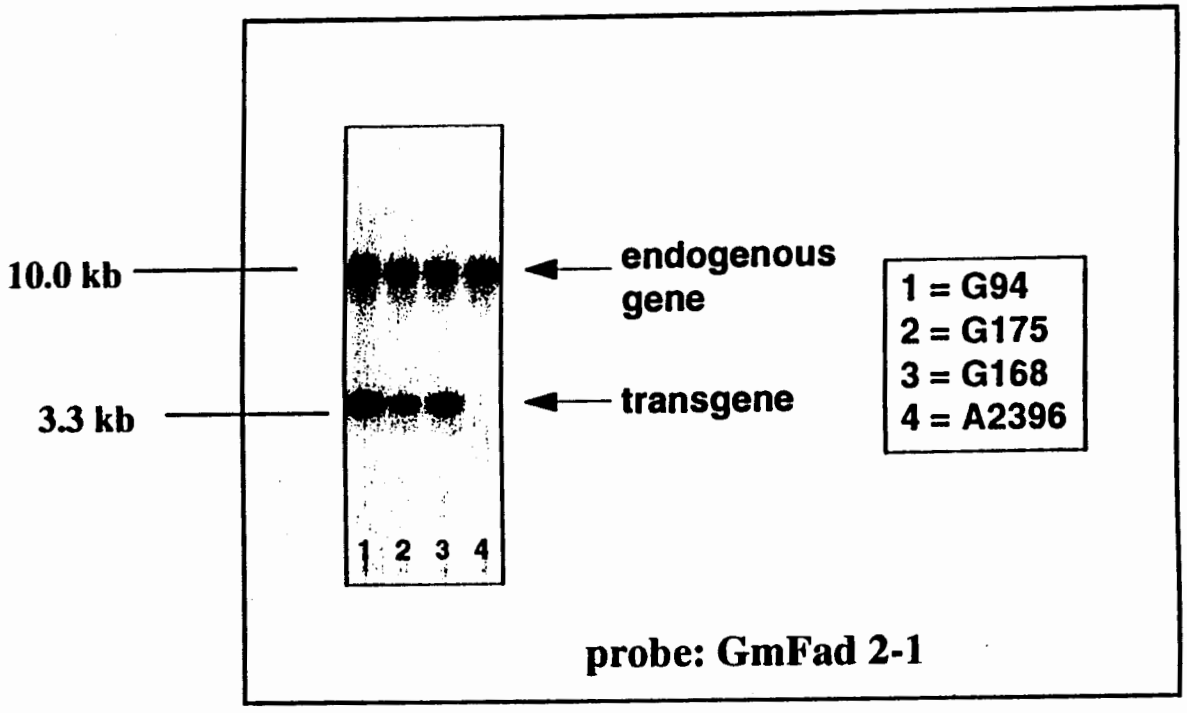
1. On page 20 of the petition, southern blot data in Figure 6 ~~(Fig. 5)~~ should be accompanied by molecular weight markers or some indication of the size of the DNA bands present. This informaiton is needed to determine if gel electrophoresis has provided sufficient resolution in separating DNA bands and determining the number of insertion events.

FIGURE 6 BamHI Digest Phaseolin PROBE



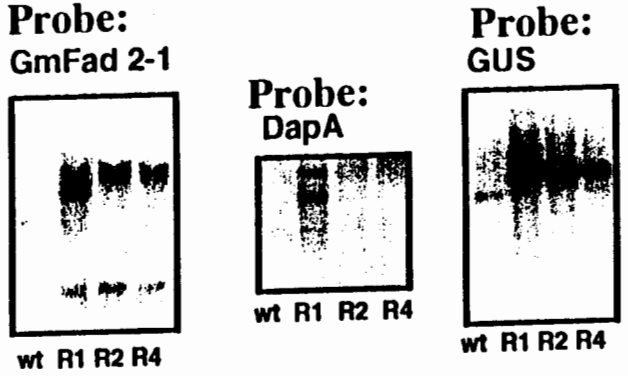
2. On page 21 of the petition, southere blot data in Figure 8 (~~Fig. 6~~) should include an indication of the molecular weitght of the bands present to verify that the transcriptional unit remains intact.

FIGURE 8


10.0 kb
3.3 kb
endogenous gene
transgene
1 = G94
2 = G175
3 = G168
4 = A2396
1 2 3 4
probe: GmFad 2-1

3. On page 22 of the petition, neither the southern blot data in figure 10 (~~fig 8~~) nor the accompanying text indicates what specific probe was used, Please indicate specifi ally what probe was used to generate these data.

FIGURE 10



Probe GmFad 2-1: 1.1 kb Ssp I fragment from pBS43 (fig. 3)

Probe Dap A: 0.87 kb Nco I/Eco RI fragment from pML87

Probe GUS: 1.85 kb Nco I/Asp 718 fragment from pMH40

Additional information on probes on following pages

Gm Fad2-1 probe

E

Ssp I -95

```
CCATATACTAATATTTGCTTGTATTGATAGCCCCTCCGTTCCCAAGAGTATAAAACTGCATCGAATAATA
GGTATATGATTATAAACGAACATAACTATCGGGGAGGCAAGGGTTCTCATATTTTGACGTAGCTTATTAT

 P  Y  T  N  I  C  L  Y  .  .  P  L  R  S  Q  E  Y  K  T  A  S  N  N
  H  I  L  I  F  A  C  I  D  S  P  S  V  P  K  S  I  K  L  H  R  I  I
T  I  Y  .  Y  L  L  V  L  I  A  P  P  F  P  R  V  .  N  C  I  E  .  Y

CAAGCCACTAGGCATGGGTCTAGCAAAGGAAACAACAATGGGAGGTAGAGGTCGTGTGGCCAAAGTGGAA
GTTCGGTGATCCGTACCCAGATCGTTTCCTTTGTTGTTACCCTCCATCTCCAGCACACCGGTTTCACCTT

 T  S  H  .  A  W  V  .  Q  R  K  Q  Q  W  E  V  E  V  V  W  P  K  W  K
  Q  A  T  R  H  G  S  S  K  G  N  N  N  G  R  .  R  S  C  G  Q  S  G
   K  P  L  G  M  G  L  A  K  E  T  T  M  G  G  R  G  R  V  A  K  V  E

GTTCAAGGGAAGAAGCCTCTCTCAAGGGTTCCAAACACAAAGCCACCATTCACTGTTGGCCAACTCAAGA
CAAGTTCCCTTCTTCGGAGAGAGTTCCCAAGGTTTGTGTTTCGGTGGTAAGTGACAACCGGTTGAGTTCT

  F  K  G  R  S  L  S  Q  G  F  Q  T  Q  S  H  H  S  L  L  A  N  S  R
S  S  R  E  E  A  S  L  K  G  S  K  H  K  A  T  I  H  C  W  P  T  Q  E
 V  Q  G  K  K  P  L  S  R  V  P  N  T  K  P  P  F  T  V  G  Q  L  K

AAGCAATTCCACCACACTGCTTTCAGCGCTCCCTCCTCACTTCATTCTCCTATGTTGTTTATGACCTTTC
TTCGTTAAGGTGGTGTGACGAAAGTCGCGAGGGAGGAGTGAAGTAAGAGGATACAACAAATACTGGAAAG

 K  Q  F  H  H  T  A  F  S  A  P  S  S  L  H  S  P  M  L  F  M  T  F
  S  N  S  T  T  L  L  S  A  L  P  P  H  F  I  L  L  C  C  L  .  P  F
K  A  I  P  P  H  C  F  Q  R  S  L  L  T  S  F  S  Y  V  V  Y  D  L  S

ATTTGCCTTCATTTTCTACATTGCCACCACCTACTTCCACCTCCTTCCTCAACCCTTTTCCCTCATTGCA
TAAACGGAAGTAAAAGATGTAACGGTGGTGGATGAAGGTGGAGGAAGGAGTTGGGAAAAGGGAGTAACGT

 H  L  P  S  F  S  T  L  P  P  P  T  S  T  S  F  L  N  P  F  P  S  L  H
  I  C  L  H  F  L  H  C  H  H  L  L  P  P  P  S  S  T  L  F  P  H  C
   F  A  F  I  F  Y  I  A  T  T  Y  F  H  L  L  P  Q  P  F  S  L  I  A
```

```
TGGCCAATCTATTGGGTTCTCCAAGGTTGCCTTCTCACTGGTGTGTGGGTGATTGCTCACGAGTGTGGTC
ACCGGTTAGATAACCCAAGAGGTTCCAACGGAAGAGTGACCACACACCCACTAACGAGTGCTCACACCAG

  G  Q  S  I  G  F  S  K  V  A  F  S  L  V  C  G  .  L  L  T  S  V  V
M  A  N  L  L  G  S  P  R  L  P  S  H  W  C  V  G  D  C  S  R  V  W  S
 W  P  I  Y  W  V  L  Q  G  C  L  L  T  G  V  W  V  I  A  H  E  C  G

ACCATGCCTTCAGCAAGTACCAATGGGTTGATGATGTTGTGGGTTTGACCCTTCACTCAACACTTTTAGT
TGGTACGGAAGTCGTTCATGGTTACCCAACTACTACAACACCCAAACTGGGAAGTGAGTTGTGAAAATCA

 T  M  P  S  A  S  T  N  G  L  M  M  L  W  V  .  P  F  T  Q  H  F  .
  P  C  L  Q  Q  V  P  M  G  .  .  C  C  G  F  D  P  S  L  N  T  F  S
H  H  A  F  S  K  Y  Q  W  V  D  D  V  V  G  L  T  L  H  S  T  L  L  V

CCCTTATTTCTCATGGAAAATAAGCCATCGCCGCCATCACTCCAACACAGGTTCCCTTGACCGTGATGAA
GGGAATAAAGAGTACCTTTTATTCGGTAGCGGCGGTAGTGAGGTTGTGTCCAAGGGAACTGGCACTACTT

S  L  I  S  H  G  K  .  A  I  A  A  I  T  P  T  Q  V  P  L  T  V  M  K
 P  L  F  L  M  E  N  K  P  S  P  P  S  L  Q  H  R  F  P  .  P  .  .
  P  Y  F  S  W  K  I  S  H  R  R  H  H  S  N  T  G  S  L  D  R  D  E

GTGTTTGTCCCAAAACCAAAATCCAAAGTTGCATGGTTTTCCAAGTACTTAAACAACCCTCTAGGAAGGG
CACAAACAGGGTTTTGGTTTTAGGTTTCAACGTACCAAAAGGTTCATGAATTTGTTGGGAGATCCTTCCC

  C  L  S  Q  N  Q  N  P  K  L  H  G  F  P  S  T  .  T  T  L  .  E  G
S  V  C  P  K  T  K  I  Q  S  C  M  V  F  Q  V  L  K  Q  P  S  R  K  G
 V  F  V  P  K  P  K  S  K  V  A  W  F  S  K  Y  L  N  N  P  L  G  R

CTGTTTCTCTTCTCGTCACACTCACAATAGGGTGGCCTATGTATTTAGCCTTCAATGTCTCTGGTAGACC
GACAAAGAGAAGAGCAGTGTGAGTGTTATCCCACCGGATACATAAATCGGAAGTTACAGAGACCATCTGG

 L  F  L  F  S  S  H  S  Q  .  G  G  L  C  I  .  P  S  M  S  L  V  D
  C  F  S  S  R  H  T  H  N  R  V  A  Y  V  F  S  L  Q  C  L  W  .  T
A  V  S  L  L  V  T  L  T  I  G  W  P  M  Y  L  A  F  N  V  S  G  R  P
```

420 490 560 630 700

```
CTATGATAGTTTTGCAAGCCACTACCACCCTTATGCTCCCATATATTCTAACCGTGAGAGGCTTCTGATC
GATACTATCAAAACGTTCGGTGATGGTGGGAATACGAGGGTATATAAGATTGGCACTCTCCGAAGACTAG  770

P M I V L Q A T T T L M L P Y I L T V R G F . S
 L . . F C K P L P P L C S H I F . P . E A S D
  Y D S F A S H Y H P Y A P I Y S N R E R L L I

TATGTCTCTGATGTTGCTTTGTTTTCTGTGACTTACTCTCTCTACCGTGTTGCAACCCTGAAAGGGTTGG
ATACAGAGACTACAACGAAACAAAAGACACTGAATGAGAGAGATGGCACAACGTTGGGACTTTCCCAACC  840

  M S L M L L C F L . L T L S T V L Q P . K G W
L C L . C C F V F C D L L S L P C C N P E R V G
 Y V S D V A L F S V T Y S L Y R V A T L K G L

TTTGGCTGCTATGTGTTTATGGGGTGCCTTTGCTCATTGTGAACGGTTTTCTTGTGACTATCACATATTT
AAACCGACGATACACAAATACCCCACGGAAACGAGTAACACTTGCCAAAAGAACACTGATAGTGTATAAA  910

 F G C Y V F M G C L C S L . T V F L . L S H I
  L A A M C L W G A F A H C E R F S C D Y H I F
V W L L C V Y G V P L L I V N G F L V T I T Y L

GCAGCACACACACTTTGCCTTGCCTCATTACGATTCATCAGAATGGGACTGGCTGAAGGGAGCTTTGGCA
CGTCGTGTGTGTGAAACGGAACGGAGTAATGCTAAGTAGTCTTACCCTGACCGACTTCCCTCGAAACCGT  980

C S T H T L P C L I T I H Q N G T G . R E L W Q
 A A H T L C L A S L R F I R M G L A E G S F G
  Q H T H F A L P H Y D S S E W D W L K G A L A

ACTATGGACAGAGATTATGGGATTCTGAACAAGGTGTTTCATCACATAACTGATACTCATGTGGCTCACC
TGATACCTGTCTCTAATACCCTAAGACTTGTTCCACAAAGTAGTGTATTGACTATGAGTACACCGAGTGG  1050

  L W T E I M G F . T R C F I T . L I L M W L T
N Y G Q R L W D S E Q G V S S H N . Y S C G S P
 T M D R D Y G I L N K V F H H I T D T H V A H
```

11

+1000

Ssp I

```
ATCTCTTCTCTACAATGCCACATTACCATGCAATGGAGGCAACCAATGCAATCAAGCCAATATTGGGTGA
TAGAGAAGAGATGTTACGGTGTAATGGTACGTTACCTCCGTTGGTTACGTTAGTTCGGTTATAACCCACT

 I  S  S  L  Q  C  H  I  T  M  Q  W  R  Q  P  M  Q  S  S  Q  Y  W  V
  S  L  L  Y  N  A  T  L  P  C  N  G  G  N  Q  C  N  Q  A  N  I  G  .
H  L  F  S  T  M  P  H  Y  H  A  M  E  A  T  N  A  I  K  P  I  L  G  E

GTACTACCAATTTTGATGACACACCATTTTACAAGGCACTGTGGAGAGAAGCGAGAGAGTGCCTCTATGTG
CATGATGGTTAAACTACTGTGTGGTAAAATGTTCCGTGACACCTCTCTTCGCTCTCTCACGGAGATACAC

 S  T  T  N  L  M  T  H  H  F  T  R  H  C  G  E  K  R  E  S  A  S  M  W
  V  L  P  I  .  .  H  T  I  L  Q  G  T  V  E  R  S  E  R  V  P  L  C
   Y  Y  Q  F  D  D  T  P  F  Y  K  A  L  W  R  E  A  R  E  C  L  Y  V

GAGCCAGATGAAGGAACATCCGAGAAGGGCGTGTATTGGTACAGGAACAAGTATTGATGGAGCAACCAAT
CTCGGTCTACTTCCTTGTAGGCTCTTCCCGCACATAACCATGTCCTTGTTCATAACTACCTCGTTGGTTA

  S  Q  M  K  E  H  P  R  R  A  C  I  G  T  G  T  S  I  D  G  A  T  N
 G  A  R  .  R  N  I  R  E  G  R  V  L  V  Q  E  Q  V  L  M  E  Q  P  M
  E  P  D  E  G  T  S  E  K  G  V  Y  W  Y  R  N  K  Y  .  W  S  N  Q

GGGCCATAGTGGGAGTTATGGAAGTTTTGTCATGTATTAGTACATAATTAGTAGAATGTTATAAATAAGT
CCCGGTATCACCCTCAATACCTTCAAAACAGTACATAATCATGTATTAATCATCTTACAATATTTATTCA

 G  P  .  W  E  L  W  K  F  C  H  V  L  V  H  N  .  .  N  V  I  N  K
  G  H  S  G  S  Y  G  S  F  V  M  Y  .  Y  I  I  S  R  M  L  .  I  S
 W  A  I  V  G  V  M  E  V  L  S  C  I  S  T  .  L  V  E  C  Y  K  .  V

GGATTTGCCGCGTAATGACTTTGTGTGTATTGTGAAACAGCTTGTTGCGATCATGGTTATAATGTAAAAA
CCTAAACGGCGCATTACTGAAACACACATAACACTTTGTCGAACAACGCTAGTACCAATATTACATTTTT

 W  I  C  R  V  M  T  L  C  V  L  .  N  S  L  L  R  S  W  L  .  C  K  N
  G  F  A  A  .  .  L  C  V  Y  C  E  T  A  C  C  D  H  G  Y  N  V  K
   D  L  P  R  N  D  F  V  C  I  V  K  Q  L  V  A  I  M  V  I  M  .  K
```

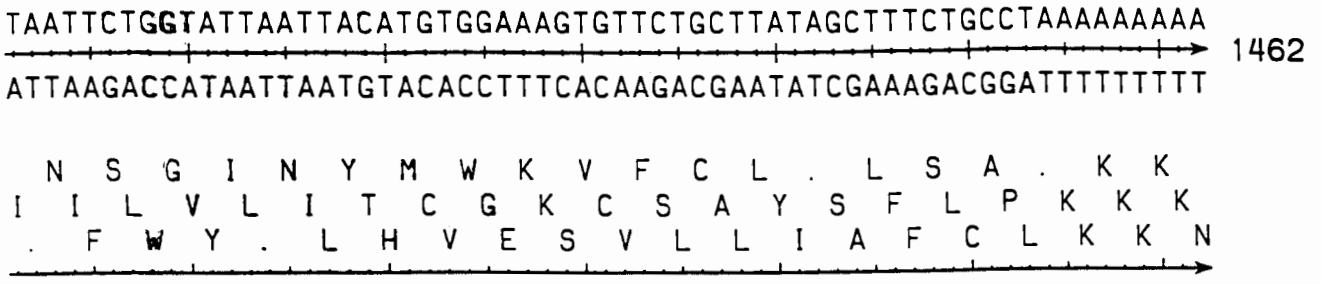
1120

1190

1260

1330

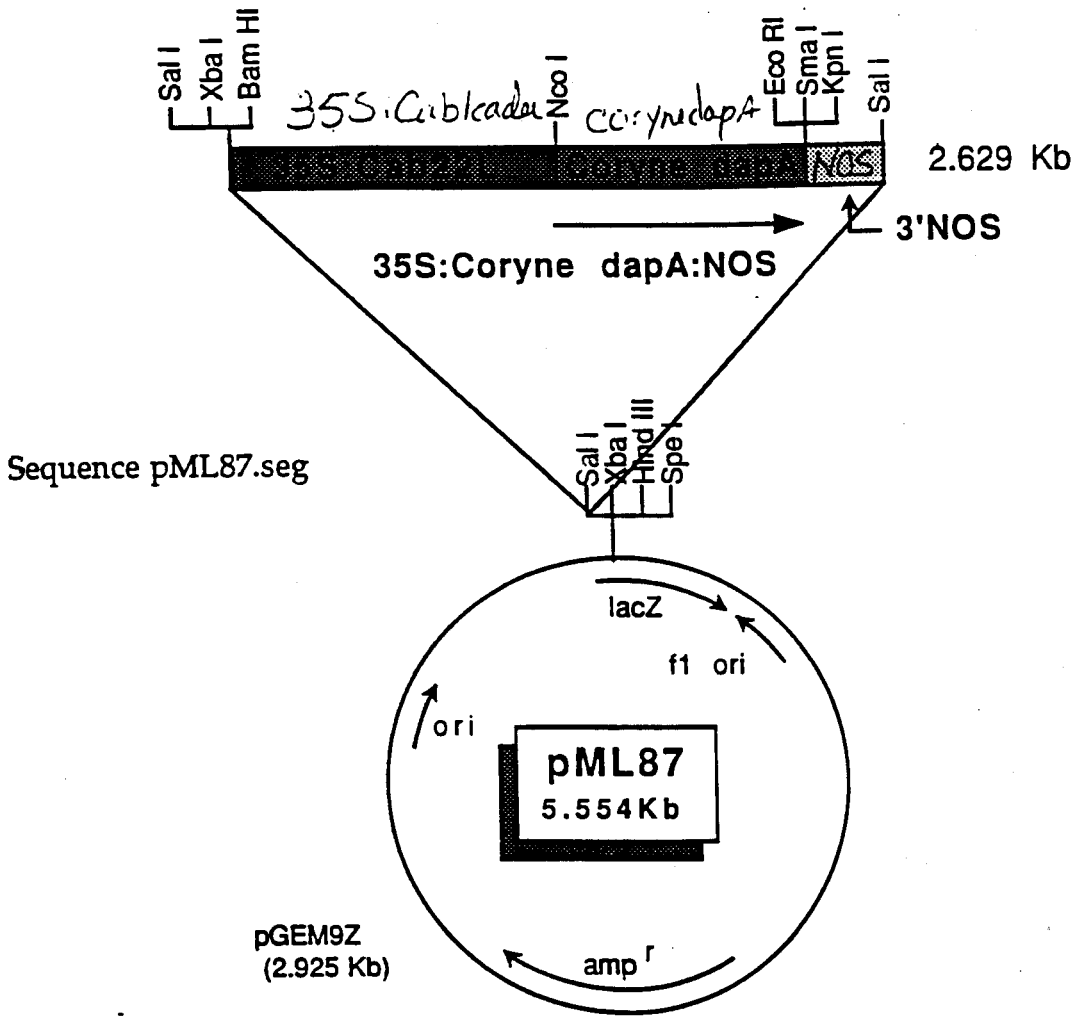
1400


TAATTCTGGTATTAATTACATGTGGAAAGTGTTCTGCTTATAGCTTTCTGCCTAAAAAAAAA
1462
ATTAAGACCATAATTAATGTACACCTTTCACAAGACGAATATCGAAAGACGGATTTTTTTTT
N S G I N Y M W K V F C L . L S A . K K
I I L V L I T C G K C S A Y S F L P K K K
. F W Y . L H V E S V L L I A F C L K K N

DapA Probe: The dapA probe was isolated as a 0.91 kb Nco I to EcoR I fragment from plasmid pML87. The probe contains the entire coding region of the Corynebacterium dapA gene. It does not contain coding suquence for the chloroplast transit peptide.

NcoI (1)

```
   1 CCATGGCTACAGGTTTAACAGCTAAGACCGGAGTAGAGCACTTCGGCACCGTTGGAGTAGCAATGGTTACTCCAT
   1▶MetAlaThrGlyLeuThrAlaLysThrGlyValGluHisPheGlyThrValGlyValAlaMetValThrProP

  76 TCACGGAATCCGGAGACATCGATATCGCTGCTGGCCGCGAAGTCGCGGCTTATTTGGTTGATAAGGGCTTGGATT
  25▶heThrGluSerGlyAspIleAspIleAlaAlaGlyArgGluValAlaAlaTyrLeuValAspLysGlyLeuAspS

 151 CTTTGGTTCTCGCGGGCACCACTGGTGAATCCCCAACGACAACCGCCGCTGAAAAACTAGAACTGCTCAAGGCCG
  50▶erLeuValLeuAlaGlyThrThrGlyGluSerProThrThrThrAlaAlaGluLysLeuGluLeuLeuLysAlaV

 226 TTCGTGAGGAAGTTGGGGATCGGGCGAAGCTCATCGCCGGTGTCGGAACCAACAACACGCGGACATCTGTGGAAC
  75▶alArgGluGluValGlyAspArgAlaLysLeuIleAlaGlyValGlyThrAsnAsnThrArgThrSerValGluL

 301 TTGCGGAAGCTGCTGCTTCTGCTGGCGCAGACGGCCTTTTAGTTGTAACTCCTTATTACTCCAAGCCGAGCCAAG
 100▶euAlaGluAlaAlaAlaSerAlaGlyAlaAspGlyLeuLeuValValThrProTyrTyrSerLysProSerGlnG

 376 AGGGATTGCTGGCGCACTTCGGTGCAATTGCTGCAGCAACAGAGGTTCCAATTTGTCTCTATGACATTCCTGGTC
 125▶luGlyLeuLeuAlaHisPheGlyAlaIleAlaAlaAlaThrGluValProIleCysLeuTyrAspIleProGlyA

 451 GGTCAGGTATTCCAATTGAGTCTGATACCATGAGACGCCTGAGTGAATTACCTACGATTTTGGCGGTCAAGGACG
 150▶rgSerGlyIleProIleGluSerAspThrMetArgArgLeuSerGluLeuProThrIleLeuAlaValLysAspA

 526 CCAAGGGTGACCTCGTTGCAGCCACGTCATTGATCAAAGAAACGGGACTTGCCTGGTATTCAGGCGATGACCCAC
 175▶IaLysGlyAspLeuValAlaAlaThrSerLeuIleLysGluThrGlyLeuAlaTrpTyrSerGlyAspAspProL

 601 TAAACCTTGTTTGGCTTGCTTTGGGCGGATCAGGTTTCATTTCCGTAATTGGACATGCAGCCCCCACAGCATTAC
 200▶euAsnLeuValTrpLeuAlaLeuGlyGlySerGlyPheIleSerValIleGlyHisAlaAlaProThrAlaLeuA

 676 GTGAGTTGTACACAAGCTTCGAGGAAGGCGACCTCGTCCGTGCGCGGGAAATCAACGCCAAACTATCACCGCTGG
 225▶rgGluLeuTyrThrSerPheGluGluGlyAspLeuValArgAlaArgGluIleAsnAlaLysLeuSerProLeuV

 751 TAGCTGCCCAAGGTCGCTTGGGTGGAGTCAGCTTGGCAAAAGCTGCTTCGCGTCTGCAGGGCATCAACGTAGGAG
 250▶alAlaAlaGlnGlyArgLeuGlyGlyValSerLeuAlaLysAlaAlaSerArgLeuGlnGlyIleAsnValGlyA

 826 ATCCTCGACTTCCAATTATGGCTCCAAATGAGCAGGAACTTGAGGCTCTCCGAGAAGACATGAAAAAAGCTGGAG
 275▶spProArgLeuProIleMetAlaProAsnGluGlnGluLeuGluAlaLeuArgGluAspMetLysLysAlaGlyV

                    EcoRI (912)
 901 TTCTATAATGAGAATTC
 300▶alLeu•••
```



Sequence pML87.seg

GUS Probe: The GUS probe was isolated as a 1.87 kb Nco I to Asp718 I fragment from plasmid pMH40. The probe contains the entire coding region of the GUS gene.

NcoI (1)

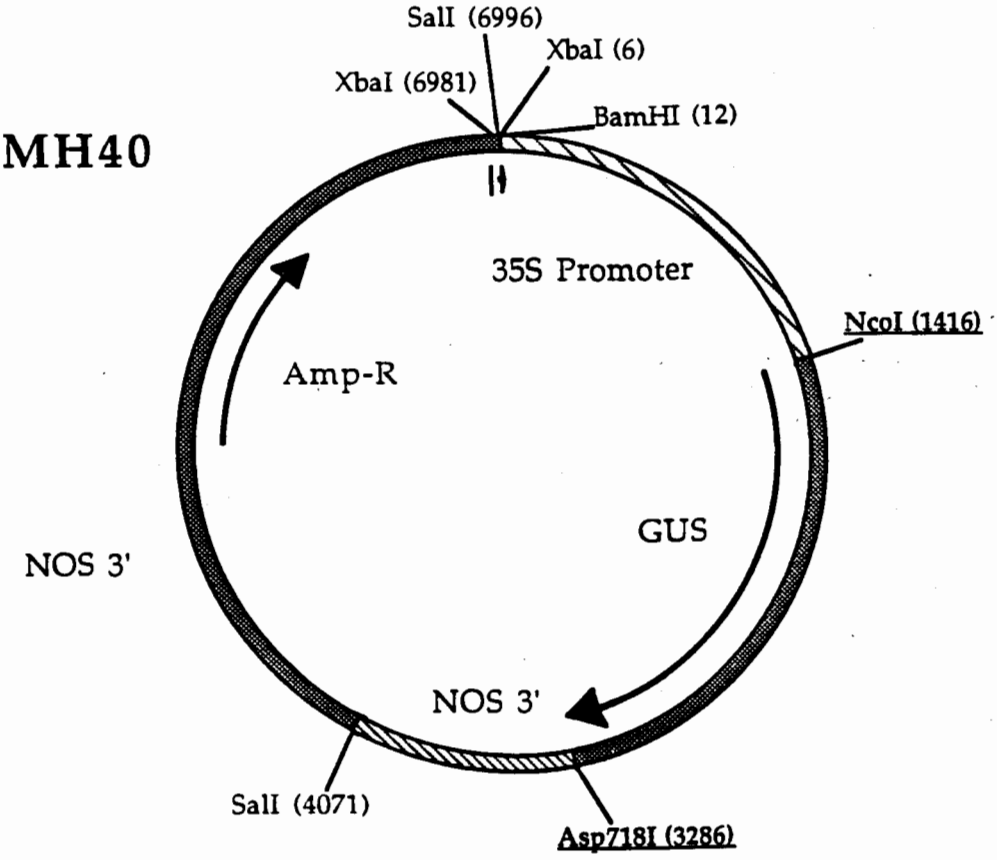
```
   1 CCATGGTACGTCCTGTAGAAACCCCAACCCGTGAAATCAAAAAACTCGACGGCCTGTGGGCATTCAGTCTGGATCGC
   1▶MetValArgProValGluThrProThrArgGluIleLysLysLeuAspGlyLeuTrpAlaPheSerLeuAspArg

  78 GAAAACTGTGGAATTGATCAGCGTTGGTGGGAAAGCGCGTTACAAGAAAGCCGGGCAATTGCTGTGCCAGGCAGTTT
  26▶GluAsnCysGlyIleAspGlnArgTrpTrpGluSerAlaLeuGlnGluSerArgAlaIleAlaValProGlySerPh

 155 TAACGATCAGTTCGCCGATGCAGATATTCGTAATTATGCGGGCAACGTCTGGTATCAGCGCGAAGTCTTTATACCGA
  51▶eAsnAspGlnPheAlaAspAlaAspIleArgAsnTyrAlaGlyAsnValTrpTyrGlnArgGluValPheIleProL

 232 AAGGTTGGGCAGGCCAGCGTATCGTGCTGCGTTTCGATGCGGTCACTCATTACGGCAAAGTGTGGGTCAATAATCAG
  77▶ysGlyTrpAlaGlyGlnArgIleValLeuArgPheAspAlaValThrHisTyrGlyLysValTrpValAsnAsnGln

 309 GAAGTGATGGAGCATCAGGGCGGCTATACGCCATTTGAAGCCGATGTCACGCCGTATGTTATTGCCGGGAAAAGTGT
 103▶GluValMetGluHisGlnGlyGlyTyrThrProPheGluAlaAspValThrProTyrValIleAlaGlyLysSerVa

 386 ACGTATCACCGTTTGTGTGAACAACGAACTGAACTGGCAGACTATCCCGCCGGGAATGGTGATTACCGACGAAAACG
 128▶lArgIleThrValCysValAsnAsnGluLeuAsnTrpGlnThrIleProProGlyMetValIleThrAspGluAsnG

 463 GCAAGAAAAAGCAGTCTTACTTCCATGATTTCTTTAACTATGCCGGAATCCATCGCAGCGTAATGCTCTACACCACG
 154▶lyLysLysLysGlnSerTyrPheHisAspPhePheAsnTyrAlaGlyIleHisArgSerValMetLeuTyrThrThr

 540 CCGAACACCTGGGTGGACGATATCACCGTGGTGACGCATGTCGCGCAAGACTGTAACCACGCGTCTGTTGACTGGCA
 180▶ProAsnThrTrpValAspAspIleThrValValThrHisValAlaGlnAspCysAsnHisAlaSerValAspTrpGl

 617 GGTGGTGGCCAATGGTGATGTCAGCGTTGAACTGCGTGATGCGGATCAACAGGTGGTTGCAACTGGACAAGGCACTA
 205▶nValValAlaAsnGlyAspValSerValGluLeuArgAspAlaAspGlnGlnValValAlaThrGlyGlnGlyThrS

 694 GCGGGACTTTGCAAGTGGTGAATCCGCACCTCTGGCAACCGGGTGAAGGTTATCTCTATGAACTGTGCGTCACAGCC
 231▶erGlyThrLeuGlnValValAsnProHisLeuTrpGlnProGlyGluGlyTyrLeuTyrGluLeuCysValThrAla

 771 AAAAGCCAGACAGAGTGTGATATCTACCCGCTTCGCGTCGGCATCCGGTCAGTGGCAGTGAAGGGCCAACAGTTCCT
 257▶LysSerGlnThrGluCysAspIleTyrProLeuArgValGlyIleArgSerValAlaValLysGlyGlnGlnPheLe

 848 GATTAACCACAAACCGTTCTACTTTACTGGCTTTGGTCGTCATGAAGATGCGGACTTACGTGGCAAAGGATTCGATA
 282▶uIleAsnHisLysProPheTyrPheThrGlyPheGlyArgHisGluAspAlaAspLeuArgGlyLysGlyPheAspA

 925 ACGTGCTGATGGTGCACGACCACGCATTAATGGACTGGATTGGGGCCAACTCCTACCGTACCTCGCATTACCCTTAC
 308▶snValLeuMetValHisAspHisAlaLeuMetAspTrpIleGlyAlaAsnSerTyrArgThrSerHisTyrProTyr

1002 GCTGAAGAGATGCTCGACTGGGCAGATGAACATGGCATCGTGGTGATTGATGAAACTGCTGCTGTCGGCTTTAACCT
 334▶AlaGluGluMetLeuAspTrpAlaAspGluHisGlyIleValValIleAspGluThrAlaAlaValGlyPheAsnLe

1079 CTCTTTAGGCATTGGTTTCGAAGCGGGCAACAAGCCGAAAGAACTGTACAGCGAAGAGGCAGTCAACGGGGAAACTC
 359▶uSerLeuGlyIleGlyPheGluAlaGlyAsnLysProLysGluLeuTyrSerGluGluAlaValAsnGlyGluThrG

1156 AGCAAGCGCACTTACAGGCGATTAAAGAGCTGATAGCGCGTGACAAAAACCACCCAAGCGTGGTGATGTGGAGTATT
 385▶InGlnAlaHisLeuGlnAlaIleLysGluLeuIleAlaArgAspLysAsnHisProSerValValMetTrpSerIle

1233 GCCAACGAACCGGATACCCGTCCGCAAGTGCACGGGAATATTTCGCCACTGGCGGAAGCAACGCGTAAACTCGACCC
 411▶AlaAsnGluProAspThrArgProGlnValHisGlyAsnIleSerProLeuAlaGluAlaThrArgLysLeuAspPr

1310 GACGCGTCCGATCACCTGCGTCAATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCTCTTTGATGTGC
 436▶oThrArgProIleThrCysValAsnValMetPheCysAspAlaHisThrAspThrIleSerAspLeuPheAspVal

1387 TGTGCCTGAACCGTTATTACGGATGGTATGTCCAAAGCGGCGATTTGGAAACGGCAGAGAAGGTACTGGAAAAAGAA
 462▶euCysLeuAsnArgTyrTyrGlyTrpTyrValGlnSerGlyAspLeuGluThrAlaGluLysValLeuGluLysGlu

1464 CTTCTGGCCTGGCAGGAGAAACTGCATCAGCCGATTATCATCACCGAATACGGCGTGGATACGTTAGCCGGGCTGCA
 488▶LeuLeuAlaTrpGlnGluLysLeuHisGlnProIleIleIleThrGluTyrGlyValAspThrLeuAlaGlyLeuHi

1541 CTCAATGTACACCGACATGTGGAGTGAAGAGTATCAGTGTGCATGGCTGGATATGTATCACCGCGTCTTTGATCGCG
 513▶sSerMetTyrThrAspMetTrpSerGluGluTyrGlnCysAlaTrpLeuAspMetTyrHisArgValPheAspArgV
```

```
1618 TCAGCGCCGTCGTCGGTGAACAGGTATGGAATTTCGCCGATTTTGCGACCTCGCAAGGCATATTGCGCGTTGGCGGT
 539▶ alSerAlaValValGlyGluGlnValTrpAsnPheAlaAspPheAlaThrSerGlnGlyIleLeuArgValGlyGly

1695 AACAAGAAAGGGATCTTCACTCGCGACCGCAAACCGAAGTCGGCGGCTTTTCTGCTGCAAAAACGCTGGACTGGCAT
 565▶AsnLysLysGlyIlePheThrArgAspArgLysProLysSerAlaAlaPheLeuLeuGlnLysArgTrpThrGlyMe

1772 GAACTTCGGTGAAAAACCGCAGCAGGGAGGCAAACAATGAATCAACAACTCTCCTGGCGCACCATCGTCGGCTACAG
 590▶tAsnPheGlyGluLysProGlnGlnGlyGlyLysGln•••
```

Asp718I (1871)

```
1849 CCTCGGTGGGGAATTCCCCGGGGGTACC
```

pMH40


SalI (6996)
XbaI (6)
XbaI (6981)
BamHI (12)
35S Promoter
NcoI (1416)
Amp-R
GUS
NOS 3'
NOS 3'
SalI (4071)
Asp718I (3286)

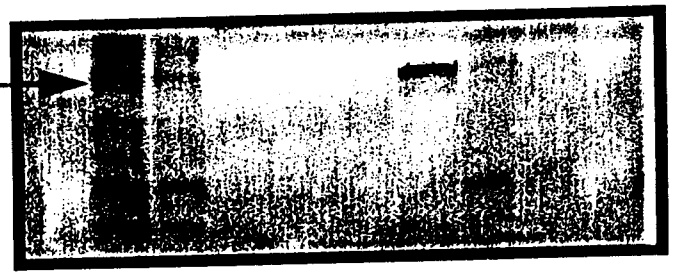
4. On page 23 of the petition, Figue 12 (fig. 10) and the accompanying text assert that the beta-glucuronidase gene is not expressed, What is the lower limit of detection of the GUS protein in the western blot protrayed in Figure 12 (10)? The northern blot in Figure 13 (fig. 11) indicates that the RNA is degraded to a fair degree in the control lane 4 (GUS positive plant). An additional positive control would confirm that undegreded RNA is indeed present in lanes 1-3. This could be demonstrated by probing with a gene that is known to be expressed in the seed. Alternatively, data should be provided that substantiates the claim that "GUS activity was absent" in the transgenic soybean lines (page 23).

Sample Loading Pattern for GUS Colorimetric Assay													
		A	B	C	D	E	F	G	H	I	J	K	L
Seed	1	A2396	A2396	G94-1	G94-1	G94-19	G94-19	A2396	A2396	G94-1	G94-1	G94-19	G94-19
Seed	2	A2396	A2396	GUS +	GUS +	GUS +	GUS +	GUS +	GUS +	GUS +	GUS +	A2396	A2396
Leaf	3	A2396	A2396	G94-1	G94-1	A5403	A5403	GUS +	GUS +	GUS +	GUS +	GUS +	GUS +
GUS Colorimetric Assay Results													
		A	B	C	D	E	F	G	H	I	J	K	L
Seed	1	-	-	-	-	-	-	-	-	-	-	-	-
Seed	2	-	-	+	+	+	+	+	+	+	+	-	-
Leaf	3	-	-	-	-	-	-	+	+	+	+	+	+

GUS Colorimetric Assay:
The top panel indicates the sample loading pattern in a 96-well plate. The top two rows contain seed samples from an elite cultivar (A2396) or transgenic plants that contain the 35S:GUS:NOS gene expression cassette. The third row contains leaf samples from either elite cultivars (A2396, A5403) or transgenic plants that contain the 35S:GUS:NOS gene expression cassette. (G94-1 and G94-19 are the transgenic high oleic acid soybeans. GUS + represents several different transgenic plants from other experiments that were previously shown to express the GUS gene.)

The bottom panel gives the results of the GUS colorimetric assay. A negative indicates samples that did not express the GUS enzyme activity and a positive indicates GUS enzyme acitivity as visualized by the presence of blue colored substrate in the wells.

GUS Western (figure 12, p23)



Lanes 1,2,7,8: partially purified GUS protein from E. coli (Sigma)

Lane 1: 25 ng
Lane 2: 10 ng
Lane 7: 5 ng
Lane 8: 2.5 ng

Minimum detectable amount of GUS: at least 0.75 ng of partially purified E.coli protein

Lanes 3-6: 50 ug of seed protein extract
Lane 3: G168
Lane 4: G94-1
Lane 5: G94-19
Lane 6: GUS positive plant

GUS protein is below level of detection (less than 0.001% of total protein) in lanes 3-5

8

97-008-01p

DuPont Agricultural Products
Walker's Mill, Barley Mill Plaza
P. O. Box 80038
Wilmington, DE 19880-0038

DUPONT

January 7, 1997

DuPont Agricultural Products

Mr. Michael A. Lidsky, J.D., LL.M
Deputy Director, Biotechnology
Coordination and Technical Assistance
BBEP, APHIS, USDA
4700 River Road
Riverdale, MD 20737

JAN - 8 1997

Dear Mr. Lidsky:

Petition for Determination of Nonregulated Status:

High Oleic Acid Transgenic Soybean Sublines G94-1, G-94-19 and G-168 derived from Transformation Event 260-05

DuPont Agricultural Products submits this petition under 7 CFR 340.6 to request that sublines G94-1, G-94-19 and G-168 derived from transformation event 260-05 wherein the only introduced trait gene is the GmFad2-1 gene which is causing a coordinate silencing of itself and endogenous GmFad2-1 gene thereby producing a soybean oil with an oleic acid level exceeding 80%, and all progenies derived from crosses between these sublines and traditional soybean varieties, and any progeny derived from crosses of these sublines with transgenic soybean varieties that have also received a determination of nonregulated status, no longer be considered regulated articles under 7 CFR Part 340. These sublines will be referred to in this petition as high oleic acid transgenic soybeans.

We are submitting ten copies of the petition, which is a revised and modified version of the petition submitted last November. References were sent with the earlier version. Included also is a diskette with the text of the petition in Word Perfect 6.1. If you need additional copies or any other information, please call me at (302) 992-6158.

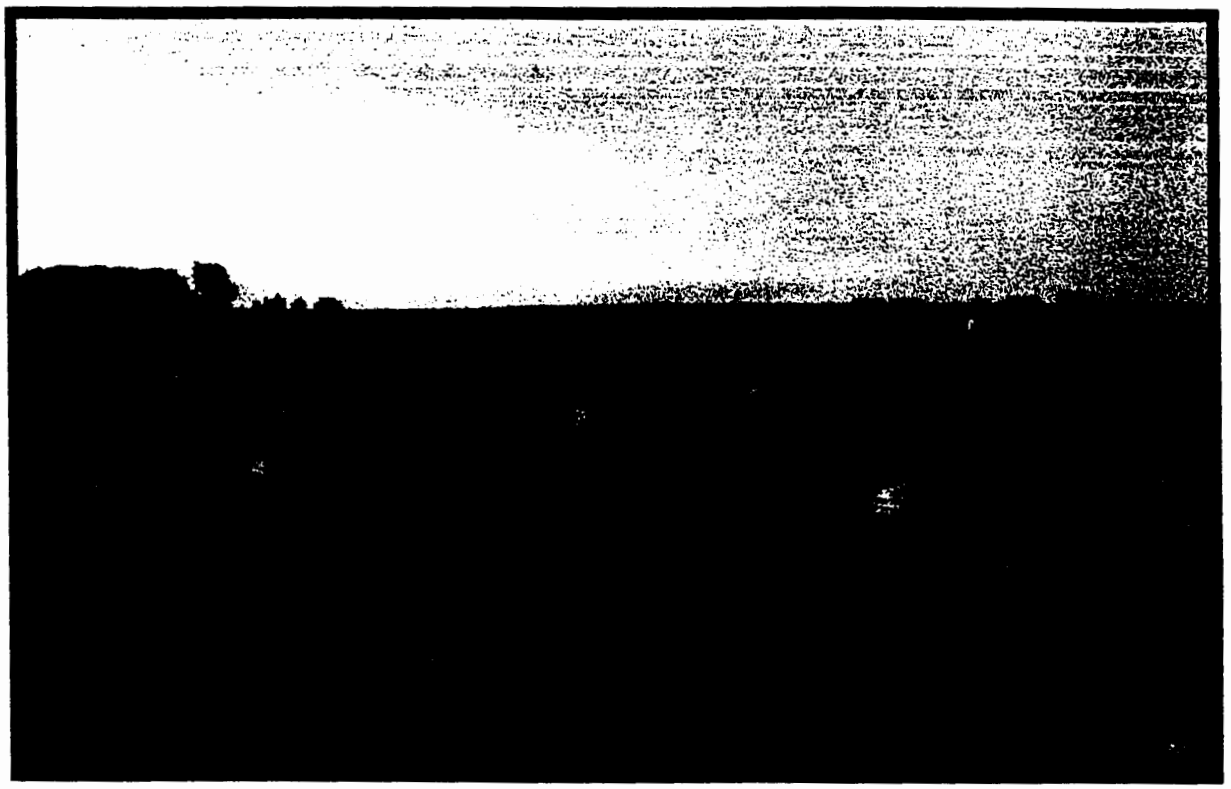
Sincerely,

Edward W. Raleigh, Ph.D.
Manager
Biotechnology Regulatory Affairs

mlc?

PETITION FOR DETERMINATION OF NONREGULATED STATUS

HIGH OLEIC ACID TRANSGENIC SOYBEAN SUBLINES DERIVED FROM TRANSFORMATION EVENT 260-05



Prepared for the Animal and Plant Health Inspection Service
by

DUPONT

DuPont Agricultural Products
January, 1997

Title Page

Petition for Determination of Nonregulated Status:

High Oleic Acid Transgenic Soybeans Sublines G94-1, G94-19 and G-168 Derived from Transformation Event 260-05

The undersigned submits this petition under 7CFR 340.6 to request that the Director, BBEP make a determination that the article should not be regulated under 7CFR 340.

Edward W. Raleigh
Manager, Biotechnology Regulatory Affairs

Jan 7 1997
Date

DuPont Agricultural Products
Barley Mill Plaza
Walker's Mill 2-172
Wilmington, DE 19880-0038
Telephone: (302) 992-6158
Fax: (302) 892-1581

Contains No Confidential Business Information

Certification

The undersigned certifies, that to the best knowledge and belief of the undersigned, this petition includes all information and views on which to base a determination, and that it includes relevant data and information known to the petitioner which are unfavorable to the petition.

Edward W. Raleigh

Jan 7 1997
Date

Edward W. Raleigh
Manager, Biotechnology Regulatory Affairs

DuPont Agricultural Products
Barley Mill Plaza
Walker's Mill 2-172
Wilmington, DE 19880-0038
Telephone: (302) 992-6158
Fax: (302) 892-1581

Table of Contents

Table of Contents (continued)

List of Figures

List of Figures (continued)

I. Introduction

DuPont Agricultural Products is submitting a Petition for Determination of Nonregulated Status to the Animal and Plant Health Inspection Service (APHIS) for high oleic acid transgenic soybeans derived from transformation event 260-05 which are homozygous for a GmFad 2-1 cDNA in the sense orientation and devoid of a high lysine gene which was concomitantly introduced into meristems of Asgrow Seed Company elite soybean line A2396 by particle bombardment, but was lost via segregation prior to field testing of the sublines which are designated as high oleic acid transgenic soybeans. All of these sublines are isogenic, have an identical phenotype, and consistently produce an oil with a relative abundance of oleic acid exceeding 80%, versus approximately 24% found in conventional soybean oil.

DuPont Agricultural Products requests a determination from APHIS that sublines G94-1, G94-19 and G168, derived from transformation event 260-05, wherein the only introduced trait gene is the GmFad 2-1 gene which is causing a coordinate silencing of itself and endogenous GmFad 2-1 thereby producing a soybean oil with an oleic acid level exceeding 80%, and any progeny derived from crosses of these sublines with traditional soybean varieties or crosses with transgenic soybean varieties that also have received a determination of nonregulated status, no longer be considered regulated articles under 7 CFR Part 340. Sublines G94-1, G94-19 and G168 are considered regulated articles because they contain sequences from the plant pests *Agrobacterium tumefaciens* and cauliflower mosaic virus (CaMV).

The sublines G94-1, G94-19 and G168 are homozygous for a GmFad 2-1 cDNA in the sense orientation under the control of a seed-specific promoter. In developing soybeans, the second double bond is added to oleic acid in the δ-12 (n-6) position by a δ-12 desaturase, encoded by the GmFad 2-1 gene. A second Fad2 gene (GmFad 2-2) is expressed in all tissues of the soybean plant. Suppression of the GmFad 2-1 gene in developing soybeans prevents the addition of a second double bond to oleic acid, resulting in greatly increased oleic acid content only in the seed.

To produce the R0 original transformant, two constructs were introduced into meristems of an elite soybean line by particle bombardment. One of the constructs contained the GmFad 2-1 sense cDNA, the other a dapA gene, expression of which leads to an increase in the free lysine content (Falco et al., 1995). The GmFad 2-1 construct contained the β-glucuronidase (GUS) gene with a 35S CaMV promoter, for plant selection, and both constructs contained the β-lactamase (Amp) gene with a bacterial promoter, used for selection when the construct is in bacterial cells.

The original R0 transformant contained inserts at two loci. At one locus (locus A) the GmFad 2-1 construct was causing a suppression of the endogenous GmFad 2-1 gene,

resulting in an oleic content of greater than 80%. At locus B, GmFad 2-1 was over expressing, resulting in a decrease of oleic content to approximately 4%. The dapA gene was only integrated at locus B. In subsequent generations, segregation led to the isolation of isogenic lines homozygous for locus A, lacking locus B, and consistently producing a soybean with a relative abundance of oleic acid above 80% of the total fatty acids. Tests indicated that the β-lactamase gene was not expressed, as expected. The GUS gene at locus A is also silent, and therefore no β-glucuronidase is present in high oleic acid transgenic soybeans.

Other changes in the oil from high oleic acid transgenic soybeans versus the parent elite line include very low abundance of linoleic acid, approximately 1%, and significantly lower linolenic and palmitic acid contents when compared with the parent elite. Trace amounts (<1%) of a linoleic acid isomer were also detected. While absent from non-hydrogenated soybean oil, this positional isomer is present at ranges from 0.02% to 5.4% of the total fatty acids of butterfat, beef and mutton tallow, partially hydrogenated vegetable oils, human milk and mango pulp.

In the high oleic soybeans the concentration of β-conglycinin α and α' subunits has been reduced and replaced with glycinin subunits. This was a result of silencing of the α and α' subunit genes mediated by the α' promoter sequence used in the GmFad 2-1 vector. It is anticipated that increasing the content of glycinin (11S) subunits and decreasing the content of β-conglycinin (7S) subunits will improve the functionality of soy proteins in various foods (Kitamura, 1995).

High oleic acid transgenic soybeans were also analyzed for total protein, oil, and carbohydrate content, crude fiber, ash, individual amino acids, phytic acid, trypsin inhibitor, stachyose, raffinose, and the isoflavones daidzein, genistein and glycitein. In addition the allergenicity potential of transgenic soybeans was evaluated. The abundance of the various components and the allergenicity potential of transgenic soybeans were the same as for the elite parent soybean.

High oleic acid transgenic soybeans will provide soybean farmers, vegetable oil producers, food companies, food frying operations, and ultimately the individual consumer with a soybean oil containing a higher percentage of oleic acid and a lower percentage of saturated fatty acids than olive oil, and without the instability problems normally encountered with typical soybean oil. This superior more heat stable soybean oil may be used in food frying and baking operations without the need for an additional processing step, chemical hydrogenation. In many food applications, the high oleic soybean oil will likely be preferred over conventional soybean oil or partially hydrogenated soybean oil because of its natural stability and favorable fatty acid profile.

Sublines G94-1, G94-19 and G168 have been field tested since 1995 under APHIS notifications, by DuPont Agricultural Products and Asgrow Seed Company, over four generations, in the United States and Puerto Rico. These field tests have been carried out at approximately 25 sites under APHIS notifications 95-088-08N, 95-107-08N, 95-257-10N,

96-071-18N and 96-115-02N (see Appendix 2). A field test was also carried out in Chile in the winter of 95/96.

Data collected from these field tests, from laboratory analyses of soybeans grown in contained facilities and in field tests, and from literature references, demonstrate that sublines G94-1, G94-19 and G168 exhibit no plant pathogenic properties; 2) are no more likely to become a weed than non-modified soybean; 3) are unlikely to increase the weediness potential of any other cultivated plant or native wild species; 4) do not cause damage to processed agricultural commodities; and 5) are unlikely to harm other organisms that are beneficial to agriculture. DuPont Agricultural Products requests a determination from APHIS that sublines G94-1, G94-19 and G168, and any progeny derived from crosses of such sublines with traditional soybean varieties or with transgenic soybean varieties that have also received a determination of nonregulated status, no longer be considered regulated articles under 7CFR Part 340.

II. Rationale for the Development of High Oleic Acid Transgenic Soybeans

Soybean oil is currently the predominant plant oil in the world, and is used in a wide variety of food applications. Untreated, commodity soybean oil is rich in polyunsaturated fatty acids, which are oxidatively unstable, making it unusable for most of these applications (Frankel, 1982). Thus it is necessary to reduce the concentration of the fatty acids causing the stability problems (Smith, 1981; Kinney, 1994).

A chemical solution to enhance stability is selective hydrogenation, a process which leads to a soybean oil more suitable for food frying operations or for use in baked products (Woerfel, 1995). Selective hydrogenation of soybean oil greatly decreases the content of polyunsaturated fatty acids, mainly linoleic (18:2) and linolenic (18:3), increases the relative abundance of monounsaturated oleic acid (18:1) while also producing substantial quantities of the trans isomer of oleic acid and other trans isomers (Koritala & Dutton, 1969; Popescu et. at., 1969). These isomers can constitute up to 40% of the total fatty acid content of commercially prepared foods such as prefried, frozen french fries (Sebedio et. al., 1994; Fernanez San Juan, 1995). Selective hydrogenation also increases the relative abundance of stearic acid (Popescu et. at., 1969; see also Section X, Figure 67).

Epidemiological and metabolic studies during the past few decades have indicated that the consumption of different types of fatty acids can affect the ratios of low-density lipoproteins (LDL) to high density lipoproteins (HDL) in the blood plasma (Spady et. al. 1993). The concentration of cholesterol in the LDL fraction of plasma is considered to be one of the major risk factors for coronary heart disease (Wollett & Dietschy, 1994).

Extensive studies have linked the consumption of saturated fats by animals and humans to coronary heart disease (McNamara, 1992; Allison et. al., 1995). This is presumably since they increase the plasma concentration of total and LDL-cholesterol and/or reduce the concentration of beneficial HDL-cholesterol (McNamara, 1992; Spady et. al. 1993; Katan et. al, 1995b). Not all saturated fatty acids have a similar effect on plasma lipoproteins however. For example, most studies report that stearic acid (18:0) has a neutral effect on the plasma concentration of LDL-cholesterol whereas palmitic acid (16:0), myristic acid (14:0) and lauric acid (12:0) appear to increase plasma LDL-cholesterol (Mensink et. al., 1994; Woollett & Dietschy, 1994; Ascherio & Welt, 1995). There are some reports, however, which correlate stearic acid with heart disease, possibly mediated by factors other than plasma LDL-cholesterol (Raloff, 1996).

Consumption of trans fatty acids by humans has also been linked to unfavorable plasma lipoprotein profiles and coronary heart disease (Mensink & Katan, 1990; Mensink et. al., 1994; Ascherio & Welt, 1995, Katan et. al., 1995a). Unfortunately, the data for trans fatty acids are even more equivocal than those for stearic acid, possibly because of the difficulties involved in quantifying trans fatty acid intake (Allison et. al., 1995) or because of the effects of interactions of trans fatty acids with dietary cholesterol or polyunsaturated fatty acids (Dictenberg et. al., 1995; Bolton-Smith et. al., 1996).

There are numerous studies which demonstrate that dietary unsaturated fatty acids are at the very least benign and most report that both mono- and polyunsaturates have the beneficial effect of reducing plasma LDL-cholesterol concentration (McNamara, 1992; Woollett & Dietschy, 1994; Gardner & Kraemer, 1995; Katan et. al., 1995b).

Thus there is now wide consensus that substituting harmful saturated fats with unsaturated oils is desirable (American Society of Clinical Nutrition, 1996). If this can be done without introducing trans fatty acids or increased amounts stearic acid into the diet, then the unresolved issues surrounding their consumption become less of a concern.

Utilizing the GmFad 2-1 gene, which encodes the enzyme responsible for adding another double bond onto oleic acid in developing soybean seeds (Heppard et. al., 1995), and gene silencing technology (Napoli et. al. 1990), DuPont has developed a soybean with a more stable oil containing 82-85% oleic acid, a greatly reduced concentration of polyunsaturated fatty acids, and a lower palmitic acid content. The stearic acid content of this oil is less than 4%. The increased stability obviates the need for an additional chemical hydrogenation processing step.

Consumption of foods (such as chicken, fish, french fries, potato chips, corn chips, donuts or bakery products) cooked in high oleic soybean oil instead of hydrogenated vegetable oil would significantly decrease a person's daily intake of trans fatty acids. Most likely the high oleic soybean oil also would be preferred over animal-derived cooking fats, such as lard, which contain greater amounts of saturated fatty acids as well as cholesterol.

Conventional breeding techniques (mutational breeding) have been used to produce soybean plants with an elevated oleic acid content (Kinney, 1994). However the environmental instability of the high oleic phenotype is a major drawback of these plants. A typical high oleic soybean produced by conventional techniques, such as DuPont's HO7-9, yields an oil with widely variable oleic acid content (35-55% of total fatty acids). The high oleic transgenic soybeans yield an oil with oleic acid consistently in the 82-85% range, irrespective of the locality or weather conditions. Of all high oleic soybean lines, the transgenic soybean is clearly superior.

III. The Soybean Family

A. Brief History of the Soybean

Soybeans were first domesticated in China about three thousand years ago and were first introduced into the United States in 1765 (Hymowitz and Harlan, 1982). The principal use of soybeans in the United States until well into the twentieth century was as a forage crop or as hay or silage. In the early days of the twentieth century, the use of soybean as an oilseed increased interest in soybeans in the United States. In 1915, oil from U.S. grown soybeans was first produced in U.S. oil mills, but even in the 1920s and 1930s, most planted soybean acres were used for forage. In the 1920s, soybean meal became an accepted ingredient for use in animal feeds. The increase in demand for cooking and salad oils and for red meat during World War II and immediately thereafter lead to a rapid expansion in soybean production (Smith and Huyser, 1987). By 1940, the United States had become a net exporter of soybeans and soybean products. Today soybeans are grown in over 29 states, and soybeans are the second largest crop in terms of cash sales, with half of the grain going into the export market. The European Union collectively represents the largest market for U.S. soybeans with Japan being the number one import country (American Soybean Association, 1995).

Soybean oil is the dominant edible vegetable oil in the world, and soybean meal is the dominant supply of high-protein livestock feed supplements (Lackey).

B. Taxonomy of the Genus Glycine

The genus Glycine is divided into two subgenera, Glycine and Soja. The first comprises perennial species mainly from Australia, the second comprises three annual species from Asia, namely Glycine max, which is the cultivated soybean, Glycine soja Sieb. & Zucc., the wild form of soybean, and Glycine gracilis Skvortz, the weedy form of soybeans. The latter two, i.e. the wild and weedy forms of soybeans grow naturally only in Asia and Australia and associated regions; in the United States they are only found at certain universities and specialized research stations. (Lackey)

C. Pollination of Soybeans

Soybean anthers shed their pollen directly onto the stigma of the same flower and soybeans are almost completely self pollinated. Natural crossing ranges from less than 0.5 to about 1% (Carlson and Lersten, 1987). Certified Seed Regulations (7CFR 201.67 - 291.78), reflecting this low potential for cross-pollination, allow foundation seed to be grown adjacent to other soybean cultivars, as long as the distance is adequate to prevent mechanical mixing of the harvested seeds.

D. Potential for Outcrossing

In the continental United States there are no relatives of cultivated soybean. The only wild species that are sexually compatible with Glycine max are members of the genus Glycine. Intersubgeneric hybrids between Glycine max and the subgenus Glycine species have been obtained only via in vitro seed culture techniques and the hybrids obtained have generally been sterile (USDA-APHIS, 1994).

Interspecific crosses between the annuals G. max and G. soja in subgenus Soja can be easily made. The latter species is found in China, Korea, Japan, Taiwan, and the former USSR. However even in those situations where some G. soya were to be found growing near cultivated soybeans, the flower development and pollination characteristics of G. max would lead to a high percentage of self-pollination (USDA-APHIS, 1994).

E. Weediness Potential

The previously cited APHIS environmental assessment notes that G. max does not show any especially weedy characteristics, the genus Glycine shows no particular aggressive weedy tendencies, and that there are no indications in standard tests and lists of weeds that cultivated soybeans are regarded as weeds. Glycine soya is listed as a common weed in Japan by Holm et al (1979), but is not considered as a harmful weed on cultivated lands, pastures and meadows (Kasahara, 1982; Nemoto, 1982)

F. Description of the Nonmodified Recipient Plant

A2396 is an Asgrow Seed Company early Group II maturity soybean variety that has high yield potential. A2396 is well adapted to areas of the upper midwest including Nebraska, Iowa, Wisconsin, Illinois, and Indiana. Protein and oil characteristics are similar to other soybeans at 40% protein and 22% oil on a dry weight basis (recorded for 1996). It has an indeterminate plant type with ovate shaped leaves, gray pubescence, purple flowers, and a brown pod wall color. The seed have imperfect black hila and dull seed coats.

IV. Biochemistry of High Oleic Acid Transgenic Soybeans

A. The GmFad 2-1 Gene

The synthesis of polyunsaturated fatty acids in developing oilseeds is catalyzed by two membrane associated desaturases which sequentially add a second and third double bond to oleic acid, a monounsaturated fatty acid (Kinney, 1994). The second double bond is added at the δ-12 (n-6) position by a δ-12 desaturase, encoded by the GmFad 2-1 gene (Okuley et al 1994, Heppard et al 1996) (see Figure 1). The third double bond is added at the n-3 (δ-15) position by a n-3 desaturase, encoded by the GmFad 3 gene (Yadav et al 1993):

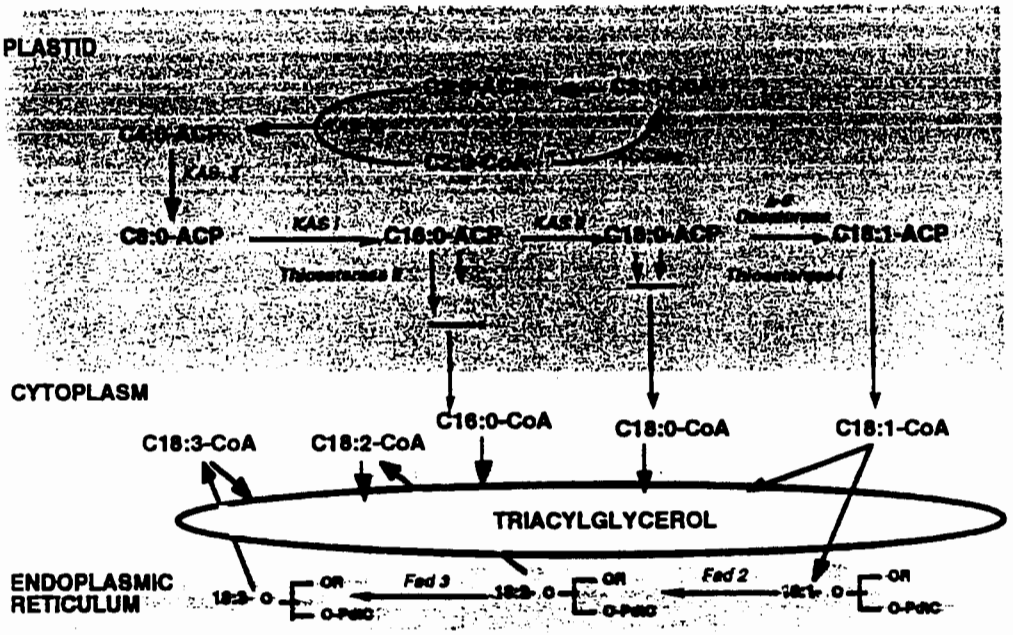


Figure 1: Fatty Acid Biosynthesis in Developing Soybeans

In soybean (*Glycine max*) there are two Fad 2 genes, one of which (GmFad 2-1) is expressed only in the developing seed (Heppard et al 1996). The expression of this gene increases during the period of oil deposition, starting around 19 days after flowering (DAF), and its gene product is responsible for the synthesis of the polyunsaturated fatty acids found in soybean oil. The other gene (GmFad 2-2) is expressed in the seed, leaf (L), root (R) and stem (St) of the soy plant at a constant level and is the "housekeeping" δ-12 desaturase gene. The Fad 2-2 gene product is responsible for the synthesis of polyunsaturated fatty acids for cell membranes. Figure 2 demonstrates that the Fad 2-2 gene is expressed, ie. produces mRNA, in various parts of the plant, while Fad 2-1 is only expressed in the seed.

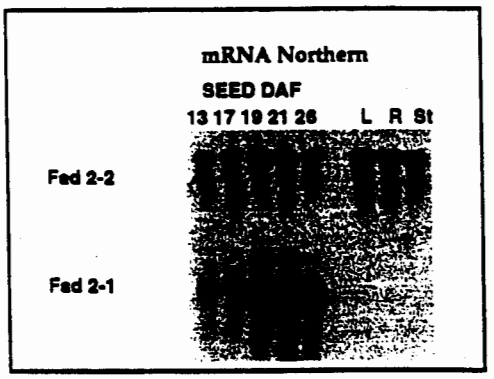


Figure 2: Northern blots of endogenous GmFad 2-1 and GmFad 2-2 genes in soybean tissues

We cloned cDNAs corresponding to both the GmFad 2-1 and the GmFad 2-2 genes using a cDNA library made from developing soybeans. A 1.46 kb fragment of cDNA, corresponding to 100% of the open reading frame of the GmFad 2-1 cDNA, was used in an attempt to suppress the endogenous GmFad 2-1 gene.

V. The Transformation and Vector System

A. Plasmid pB543 Containing the GmFad 2-1 Gene

The plasmid used for soybean transformation, pBS43, (see Figure 3), contains 3 gene expression cassettes, each of which in turn contains one or more modules. Together each cassette thus contains a regulatory segment (commonly called a promoter) responsible for directing expression in the desired cell type at the desired developmental stage, the coding sequence for the protein involved, and a termination region to ensure correct completion of the transcription process:

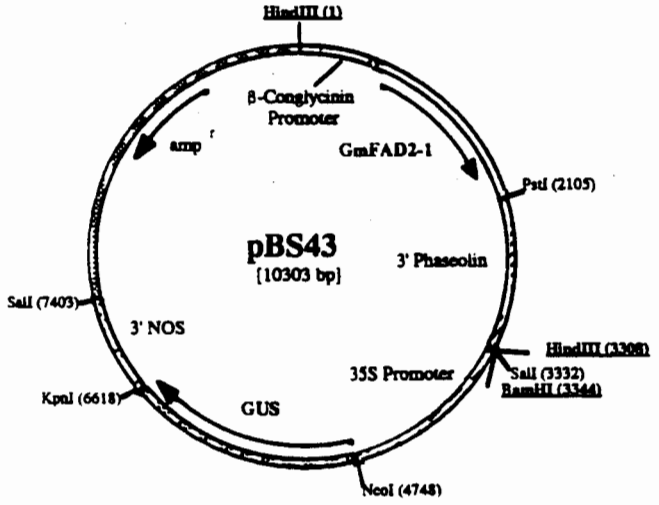


Figure 3: Soybean Transformation vector pBS43

The GmFad 2-1 gene was placed under the control of a strong, seed-specific promoter derived from the α'-subunit of the *Glycine max* β-conglycinin gene. This promoter allows high level, seed specific expression of the trait gene. It spans the 606 bp upstream of the start codon of the α' subunit of the soybean (*Glycine max*) β-congylcinin storage protein. The β-conglycinin promoter sequence represents an allele of the published β-conglycinin gene (Doyle et al., 1986) having differences at 27 nucleotide positions. It has been shown to maintain seed specific expression patterns in transgenic plants (Barker et al., 1988; Beachy et al., 1985).

The reading frame was terminated with a 3' fragment from the phaseolin gene of green bean (*Phaseolus vulgaris*) This is a 1174 bp stretch of sequences 3' of the *Phaseolus vulgaris* phaseolin gene stop codon (originated from clone described in Doyle et al., 1986). Such sequences ensure appropriate termination of transcription. Note that the restriction enzyme *Hind* III will cut either side of the GmFad 2-1 complete transcriptional unit,

whereas the restriction enzyme *Bam* HI will cut only once, outside of the complete transcriptional unit.

The GmFad 2-1 open reading frame (ORF) was in a sense orientation with respect to the promoter. The intention was to create a coordinate gene silencing of the sense GmFad 2-1 cDNA and the endogenous GmFad 2-1 gene. This phenomenon, known as "sense suppression" is reported to be an effective method for deliberately turning off genes in plants and is described in US patent 5034323. The method has been successfully used by the DNAP Company to produce tomatoes with increased ripening time. These tomatoes were described in a data package submitted to the Animal and Plant Health Inspection Service in August, 1994, and were exempted from APHIS oversight in January, 1995.

For maintenance and replication of the plasmid in *E. coli* the GmFad 2-1 transcriptional unit described above was cloned into plasmid pGEM-9z(-) (Promega Biotech, Madison WI, USA). This is a basic ampicillin resistance plasmid derived from pBR322. This plasmid contains the complete β -lactamase gene (Sutcliffe, 1979) for bacterial selection on the antibiotic ampicillin. This prokaryotic gene allows selection of transformed *E. coli* during laboratory recombinant DNA steps and is not expressed in the transformed plants. It contains its own *E. coli* regulatory sequences.

For identification of transformed soybean plants the β-glucuronidase gene (GUS) from *E. coli.* was used. The cassette used consisted of the three modules; the Cauliflower Mosaic Virus 35S promoter, the β-glucuronidase gene (GUS) from *E. coli.* and a 0.77 kb DNA fragment containing the gene terminator from the nopaline synthase (NOS) gene of the Ti-plasmid of *Agrobacterium tumefaciens*. The 35S promoter is a 1.4 kb promoter region from CaMV for constitutive gene expression in most plant tissues (Odell et al., 1985), the GUS gene a 1.85 kb fragment encoding the enzyme β-glucuronidase (Jefferson et al., 1986) and the NOS terminator a portion of the 3' end of the nopaline synthase coding region (Fraley et al., 1983). The NOS terminator does not encode the entire protein nor any start signals for transcription or translation of the gene; it is only used to terminate transcription of the selectable marker gene.

The CaMV sequence and the NOS sequence, as used in soybeans derived from transformation event 260-05, do not cause the soybeans to become a plant pests.

B. Co-bombarded Construct pML102

The GmFad 2-1 construct (pBS43) was transformed into meristems of the elite soybean line A2396, by the method of particle bombardment (Christou et al., 1990). The meristems were co-bombarded with a construct, pML102, containing a *Corynebacterium* gene encoding the enzyme dihydrodipicolinic acid synthase or dapA gene. This *Corynebacterium* gene was not present in plants with the high oleic acid phenotype (82-85% oleic) which were selected for field testing, as described in Section VI. The plasmid pML102 contains two gene expression cassettes (see Figure 4). The first cassette encodes the β-lactamase gene for bacterial transformation selection on ampicillin, as described above for pBS43. Expression of the high lysine phenotype is encoded by a seed specific gene expression cassette. The soybean

Kunitz trypsin inhibitor 3 (KTi3) promoter allows high level seed expression of the dapA gene.

The promoter spans the 2.0 kb up-stream of the start codon of the KTi3 open reading frame (Jofuku & Goldberg, 1989). No portions of the coding sequence of Kunitz trypsin inhibitor are present. Expression of KTi3 has been shown to occur at high levels in soybean seed and to turn on early during embryogenesis (Perez-Grau & Goldberg, 1989).

A chloroplast transit peptide sequence from the small subunit of ribulose bisphosphate carboxylase of soybean was cloned onto the 5' end of the dapA gene (Berry-Lowe et al. 1982). This 170 bp fragment encodes a plant chloroplast transit peptide that will direct the protein into the chloroplast where lysine biosynthesis is carried out. In pML102 the transit peptide sequence was cloned onto the dapA gene to make a translational fusion.

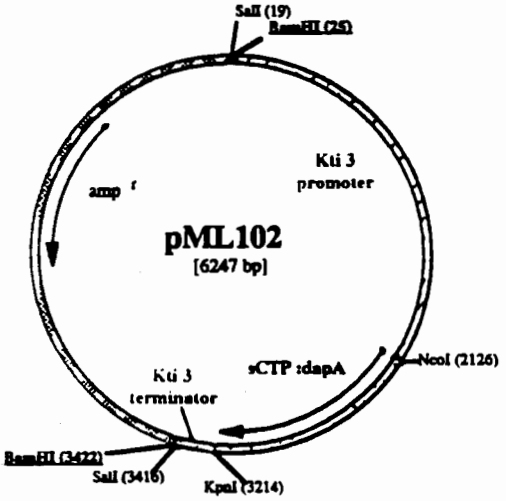


Figure 4: Plant Expression Vector pML102

The *Corynebacterium* dapA gene encodes the enzyme dihydrodipicolinic acid synthase (DHDPS) (Bonnassie et al., 1990; Yeh et al., 1988). This enzyme catalyzes the condensation of aspartyl β-semialdehyde with pyruvate and is the first reaction committed to lysine biosynthesis. In plants DHDPS is feedback inhibited by lysine and serves as the major regulator of the lysine branch of the aspartate family biosynthetic pathway. The *Corynebacterium* DHDPS is insensitive to feedback inhibition by lysine and expression of this enzyme in soybean seeds de-regulates the pathway (Falco et al., 1995) resulting in accumulation of free lysine (non protein bound lysine).

The soybean KTi3 transcriptional terminator, a 250 bp stretch of sequences 3' of the KTi3 gene stop codon (Jofuku & Goldberg, 1989) was used to ensure appropriate termination of transcription.

Expression of the dapA gene in soybeans and canola results in an increase in the free lysine content of the meal by up to five times (Falco et al., 1995). The intention was to produce

transgenic soybeans with increased lysine in their meal fraction and reduced polyunsaturates in their oil fraction (see Figure 5):

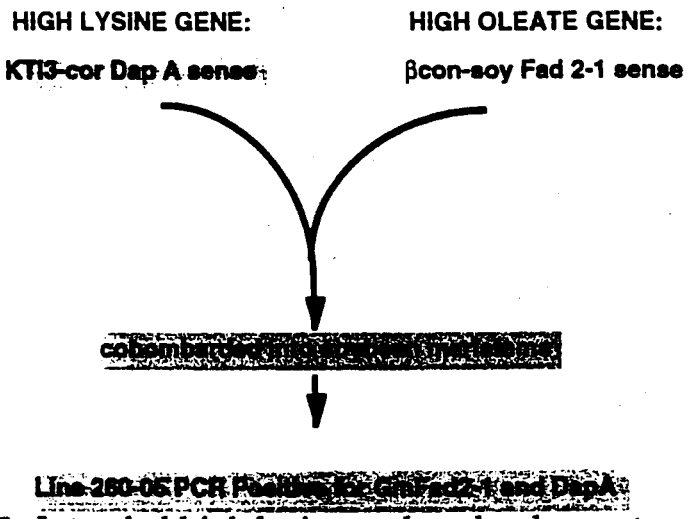


Figure 5: Intended high lysine, reduced polyunsaturates soybean

Although both constructs contained the bacterial β-lactamase gene for expression in bacteria, only the GmFad 2-1 construct contained the 35S-GUS gene for plant selection. The transformation was done for DuPont by Agracetus, Inc. using their published procedure (Christou et al., 1990). The sublines for exemption contain only the GmFad 2-1 construct and not the dapA construct, for reasons that are described below.

VI. Genetic, Molecular and Biochemical Analysis of 260-05 and Sublines

A. Selection Process for Sublines Homozygous for Silencing Locus A

1. Southern and Northern Analyses

From the initial population of transformed plants, a plant was selected which was expressing GUS activity and which was also PCR positive for the GmFad 2-1 gene (Event 260-05). Small chips were taken from a number of R1 seeds of plant 260-05 and screened for fatty acid composition. The chipped seed was then planted and germinated. Genomic DNA was extracted from the leaves of the resulting plants and cut with the restriction enzyme *Bam* HI. The blots were probed with a phaseolin probe. A single band will be seen for each locus of insertion (since the enzyme will cut once in the transgene and once randomly in the soy genome). Furthermore, only the transgene GmFad 2-1 (which contains the phaseolin 3' region) will hybridize.

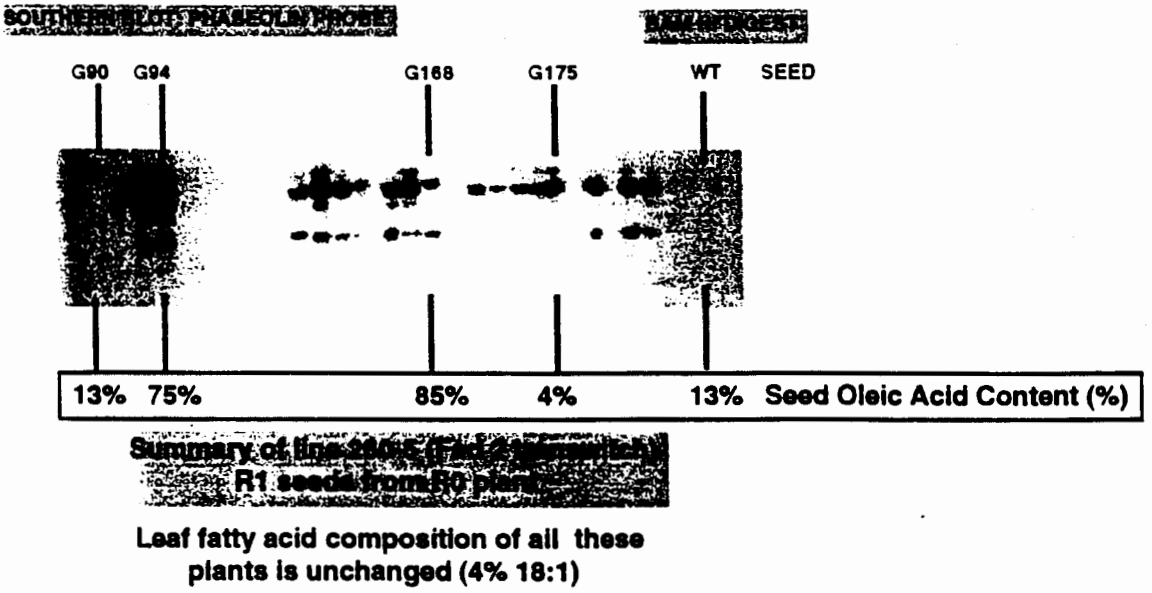


Figure 6: Southern blot of genomic DNA, digested with *Bam* HI, from leaves of plants grown from R1 seeds of Line 260-05.

From the DNA hybridization pattern in Figure 6 it was clear that in the original transformation event the GmFad 2-1 construct had become integrated at two different loci in the soybean genome. At one locus (Locus A) the GmFad 2-1 construct was causing a silencing of the endogenous GmFad 2-1 gene, resulting in a relative oleic acid content of about 85% (compared with about 20% in elite soybean varieties). At locus A (e.g. G168) there were two copies of pBS43. On the DNA hybridization blot this is seen as two cosegregating bands. One of these bands (the higher of the two) corresponds to a *Bam* HI site in the plant genome and the *Bam* HI site of pBS43. The other, lower, band corresponds

to a piece of DNA containing the phaseolin region between the *Bam* HI sites of the first and second inserts at locus A.

At the other integration locus (Locus B) (see Figure 6) the GmFad 2-1 was over-expressing, thus decreasing the oleic acid content to about 4%. At locus B (e.g. G175), there was a single copy of pBS43. The construct containing the *Corynebacterium* dapA gene was only integrated at locus B. Figure 7 contrasts the southerns of seeds containing the dapA locus and those where the dapA locus is absent.

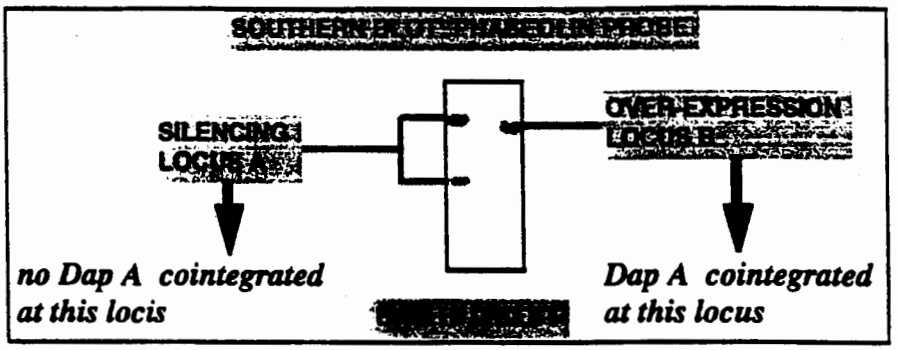


Figure 7: Southerns of seeds with or without dapA

The genomic DNA of plants with either or both loci was cut with *Hind* III (which excises the transcriptional unit from the plasmid DNA) and probed with GmFad 2-1 probe. At both loci the transcriptional unit was intact and thus integration into the soybean DNA must have occurred at some other point in the plasmid (see Figure 8).

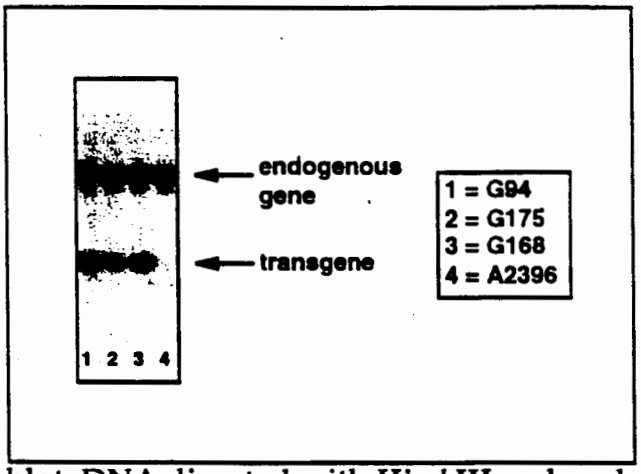


Figure 8: Southern blot: DNA digested with *Hind* III and probed with GmFad 2-1

The expression pattern of dapA in seeds containing neither locus (1), both loci (2), only the gene suppression locus A (3) and only the overexpression locus B (4) was confirmed by Northern analysis as shown in Figure 9.

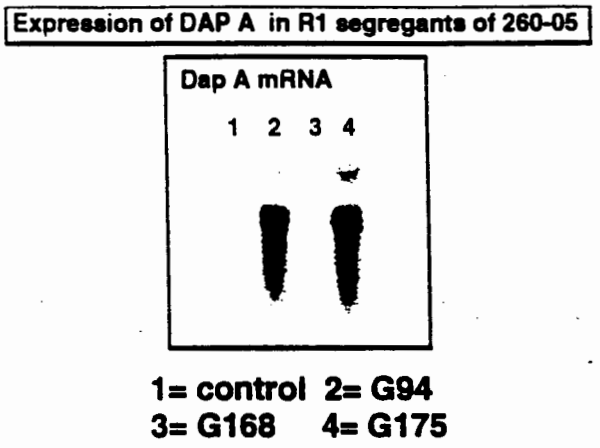


Figure 9: Northern blot probed with dapA showing expression of dapA mRNA in R1 seeds G94 and G175 but not G168

Thus R1 seed G94 contained both loci (A and B). It expressed the dapA gene, as did seeds which contained only locus B (e.g. G175). Seeds containing only locus A (e.g. G168) did not contain the dapA gene and therefore no dapA mRNA could be detected (see Figure 9). Thus subline G168 only contains locus A and no locus B.

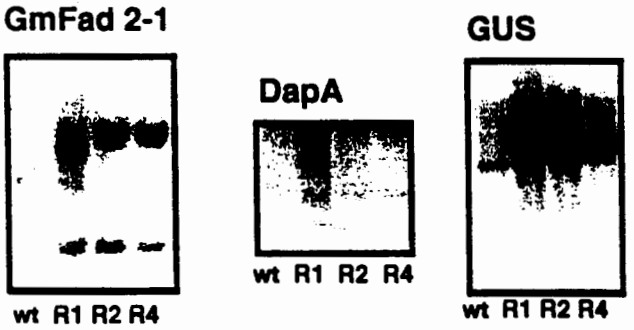


Figure 10: Southern blots of genomic DNA digested with *Bam* HI from R1, R2 & R4 G94 plants. Elite soybean line A2396 is marked as "wt" in this figure.

R2 segregant seeds of G94 were selected which contained only locus A, thus they did not contain the dapA gene. Both loci contained copies of the GUS gene, one copy segregating with the silencing locus (Figure 10).

In contrast to developing A2396 seeds (20 days after flowering), R4 seeds at a similar stage of development which contained only the silencing locus A (e.g. G94-1) did not contain any detectable GmFad 2-1 mRNA, as measured by Northern blotting (see Figure 11). GmFad 2-2 mRNA, although reduced somewhat compared with controls, was not suppressed. Thus the GmFad2-1 sense construct had the desired effect of preventing the expression of the GmFad2-1 gene and thus increasing the oleic acid content of the seed.

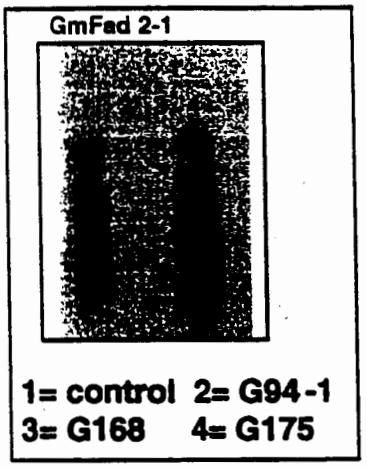


Figure 11: Northern analysis of GmFad 2-1 RNA in control (A2396) and transgenic lines

B. β-Glucuronidase Gene not Expressed

Although the GUS (β-Glucuronidase) gene was present in plants which only contained the silencing locus, GUS activity was absent in the leaves and seed of these plants. This coincided with a total absence of GUS protein on Western blots of leaf and seeds proteins from these plants (see Figure 12) and with the absence of GUS mRNA on Northern blots (see Figure 13). We concluded that the GUS gene was not expressed at this locus.

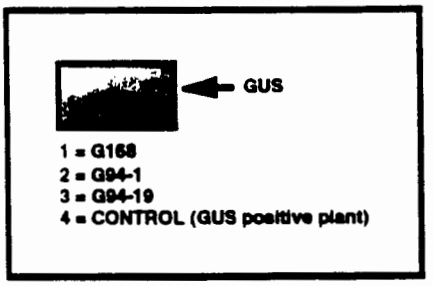


Figure 12: Western blot of GUS protein in seeds of G168, G94-1, G94-19 and a positive control plant.

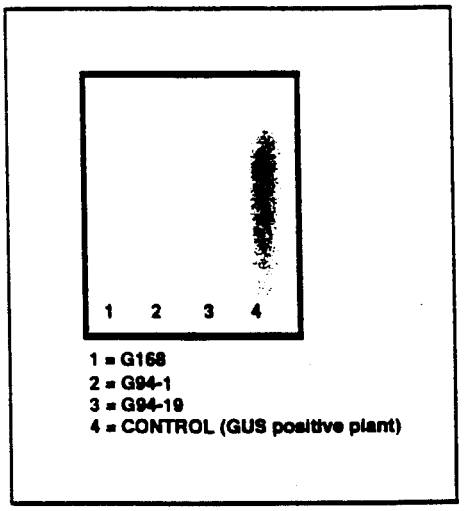


Figure 13: Northern analysis of RNA from seeds of G168, G94-1, G94-19 and a positive control plant.

C. β-Lactamase Gene not Expressed

All of the lines derived from event 260-05 which contained only the GmFad 2-1 silencing locus also contained the gene for β-lactamase. This gene, however, is under the control of a bacterial promoter and is not expressed in the plant (see Figure 14).

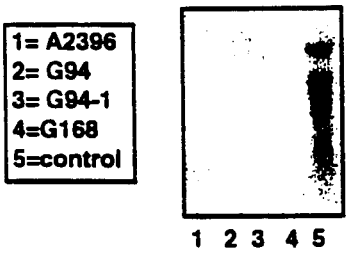
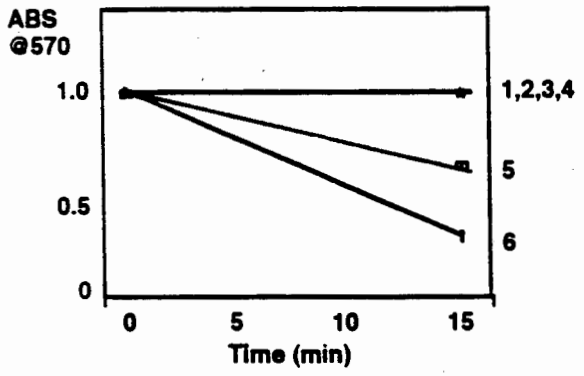


Figure 14: Northern analysis of β-lactamase expression in *E. coli* cells containing plasmid pBS43 (Control), elite soybeans (A2396) and transgenic (G94, G94-1, G168) soybeans

The absence of β-lactamase RNA in the transgenic soybeans coincided with the lack of β-lactamase activity in cell-free extracts of these plants. Enzyme activity was measured in leaf extracts by the method of DeBoer et al (1991). Fresh tissue was ground in liquid nitrogen with mortar and pestle and then homogenized with three volumes enzyme extraction buffer (DeBoer et al 1991) per gram of tissue. The supernatant was used for the β-lactamase enzyme assay. The positive control for both the RNA blots and the enzyme assays was *E. coli* cells transformed with the GmFad 2-1 expression vector (pBS43). For the enzyme assays, *E. coli* cells were washed in 50mM Tris-acetate, 2mM EDTA; pH 7.5 and the

sonicated supernatant used as a positve β–lactamase enzyme control. The negative control was 1 mg/ml Bovine serum albumin.

In the assay, 1ml 50uM pyridium-2-azo-p-dimethylaniline chromophore in 30mM potassium phosphate; pH 7.5 was used as a substrate for fifteen minute kinetic time course studying the change in absorbance at wavelengh 570nm, with readings taken every 30 seconds. The maximum absorbance was standardized to 1 unit (see Figure 15).



1= 50 ug BSA, 2= 500 ug A2396, 3= 500 ug G94-1, 4= 2500 ug G94-1, 5=50 ug *E.coli*, 6= 100 ug *E. coli*)

Figure 15: β-lactamase activity in elite and transgenic soybeans and in *E. Coli* transformed with pBS43

VII. Sublines of 260-05 Selected for Field Testing

Seeds selected for advancement were of two types:

R1 seeds homozygous for the silencing locus A (e.g. G168) that did not contain the GmFad 2-1 overexpression/dapA locus B.

R2 segregants of G94 which were homozygous for the silencing locus A and did not contain the GmFad2-1/dapA locus B (e.g. G94-1, G94-19).

These lines are therefore isogenic and can be considered equivalent. All plants homozygous for the GmFad 2-1 silencing locus had an identical Southern blot profile (digested with *Bam* HI and probed with phaseolin) over a number of generations. This indicates that the insert was stable and at the same position in the genome over at least four generations.

VIII. Description of Phenotype

A. Fatty Acid Profile

Sublines G94-1, G94-19 and G-168 differ from parent line A2396 in that the fatty acid profile is changed to produce an oil containing approximately 82-85% oleic acid, low levels of linoleic and linolenic acids, and approximately 6-7% palmitic acid and 3.5% stearic acid. Details on the fatty acid analyses and comparisons of the fatty acid profile with other oils are given in Sections IX and X, respectively.

Other constituents in sublines G94-1, G94-19 and G-168 are the same as in A2396 as outlined in Section IX. There are no expressed gene products in sublines G94-1, G94-19 and G-168 and hence no new enzymes. Except for the changes in the fatty acid metabolism, there are no changes in plant metabolism. Changes in the oil in the seed only would not be expected to have any impact on nontarget organisms.

The phenotype of sublines G94-1, G94-19 and G-168 is very similar to that of high oleic soybeans producted by mutational breeding, except that sublines G94-1, G94-19 and G-168 are superior, have higher levels of oleic acid, are more environmentally stable, and do not have any changes in the oleic levels in any plant part with the exception of the seeds.

B. Stability of the Fatty Acid Phenotype

The fatty acid phenotype is also stable. The oleic acid content of R1 seeds homozygous for the silencing locus alone was greater than 80%. The most recent analysis was done on R6 seeds from plants grown in Isabella, Puerto Rico. The oleic acid content of these seeds was greater than 80% (84-86%), indicating that the trait is still stable after six generations. We also planted three different generations of seeds (R1, R2 and R3) seeds side by side in fields plots at Stine Research Farm, Md., during the summer of 1995. Again there was no difference in the bulk oleic acid content of seeds harvested from these plants (see Figure 16).

**High oleic acid phenotype is stable in the field:
Multiple generations grown at Stine in summer 1995**

Plant ID	Generation Planted	Seed Analyzed	Bulk Oleic Acid (%)
G253	R0:1	R1:2	84.1%
G276	R0:1	R1:2	84.2%
G296	R0:1	R1:2	84.1%
G313	R0:1	R1:2	83.8%
G328	R0:1	R1:2	84.0%
G168-187	R1:2	R2:3	84.4%
G168-171	R1:2	R2:3	85.2%
G168-59-4	R2:3	R3:4	84.0%
G168-72-1	R2:3	R3:4	84.1%
G168-72-2	R2:3	R3:4	84.5%
G168-72-3	R2:3	R3:4	84.3%
G168-72-4	R2:3	R3:4	83.3%

Figure 16: Stability of R1 (R0:1) through R4 (R3:4) phenotypes in the field

The trait is also environmentally stable. R3 seeds were grown at a number of US locations during the summer of 1995 and R4 seeds in Isabella, Puerto Rico during the fall of 1995 (see Figure 17).

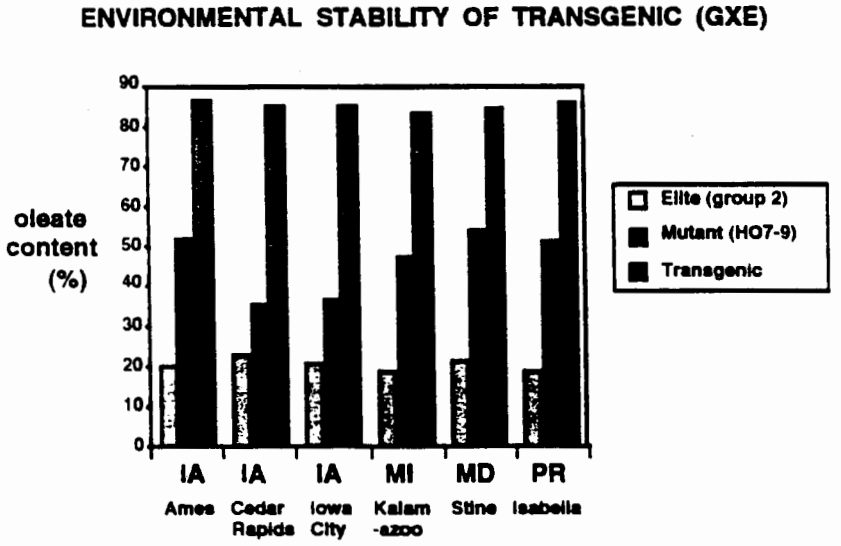


Figure 17: Oleic acid content of high oleic transgenic soybeans grown at several locations in 1995

At each of these sites the average oleic acid content of plants was 83-85% indicating that the trait is stable over a number of different growing environments. This is in contrast to high oleic soybeans produced by germplasm selection or mutation breeding, which have different oleic acid contents in different environments. A typical mutant (DuPont's HO7-9, our most environmentally stable mutant) is shown below and contrasted with high oleic transgenic and Elite soybeans from the same genetic background and maturity group (A2396).

IX. Compositional Analyses of High Oleic Acid Transgenic Soybeans

Soybean seeds from high oleic lines were compared with the parent variety A2396 in order to determine that there were no unexpected changes in composition. The soybeans were grown in two locations: Slator, Iowa, and Isabela, Puerto Rico during the summer of 1995 and the winter of 1995/1996. Two lines were analyzed designated G94-1 and G94-19. These are isogenic lines derived from the same transformation event as described previously. The seeds represent the R4 and R5 generation of the high oleic soybean lines. Compositional analysis included determination of proximate, fatty acid, amino acid, isoflavone, raffinose, stachyose, phytic acid, and trypsin inhibitor content. Reported values on each chart provide reference data obtained from the literature or determined experimentally. All soybean samples were coarsely ground in a coffee mill prior to analysis. Proximate, amino acid, phytic acid, and trypsin inhibitor assays were performed by either of two commercial testing labs[1]. Isoflavone analysis was performed at a university research lab expert in the field[2]. Fatty acid composition, raffinose and stachyose assays were performed by DuPont personnel.

A. Proximate Analyses

Soybean seeds consist of proteins, lipids, carbohydrates and minerals. Proteins and lipids are the two most important fractions commercially, accounting for more than 60% of the seed. Soybeans can vary in content due to both varietal differences and environmental factors which affect the overall seed composition. Proximate analysis was done on high oleic and the parental control soybeans to confirm that there was no change in the primary constituents of the seed. Proximate analysis includes the measurement of protein, oil, crude fiber, carbohydrate (by calculation), and ash content. The results are shown below in Figures 18-22. High oleic soybeans were indistinguishable from commodity soybeans in proximate composition. Literature values showing normal ranges in soybean composition were those reported by Orthoefer (1978), Pryde (1980), and Mounts (1987).

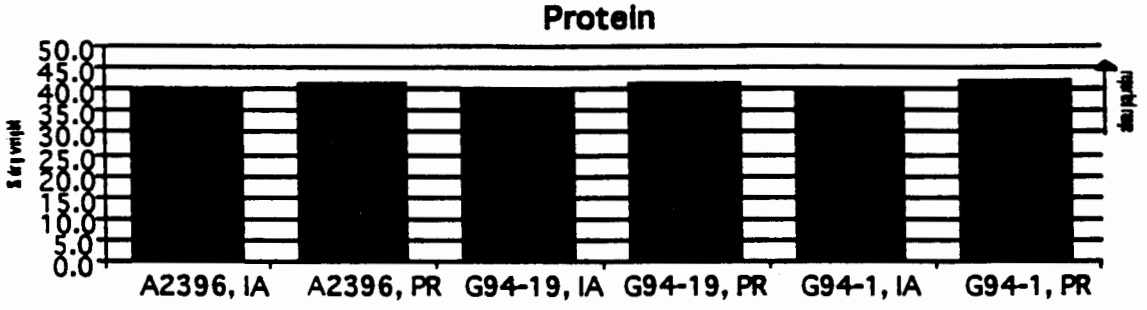


Figure 18. Protein content of high oleic and control (A2396 parent) soybean.

[1] Woodson-Tenent Laboratories, Inc. Des Moines, Iowa
CN Laboratories. Courtland, Minnesota

[2] Dr. Patricia A. Murphy, Iowa State University, Ames, Iowa

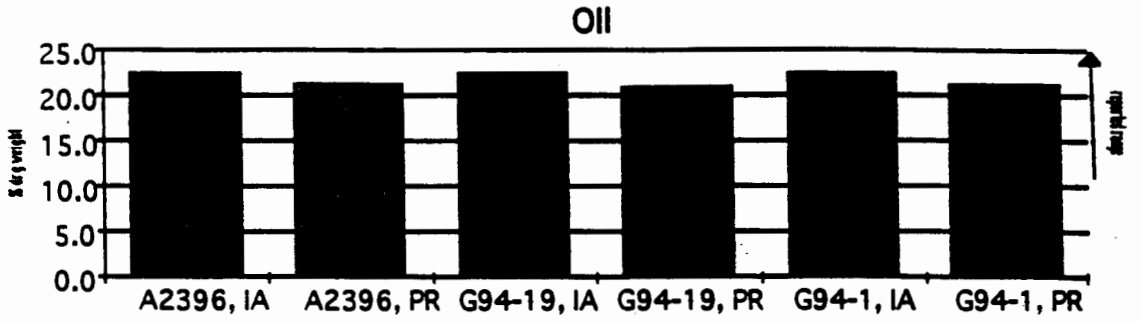


Figure 19. Crude Oil content of high oleic and control (A2396 parent) soybean.

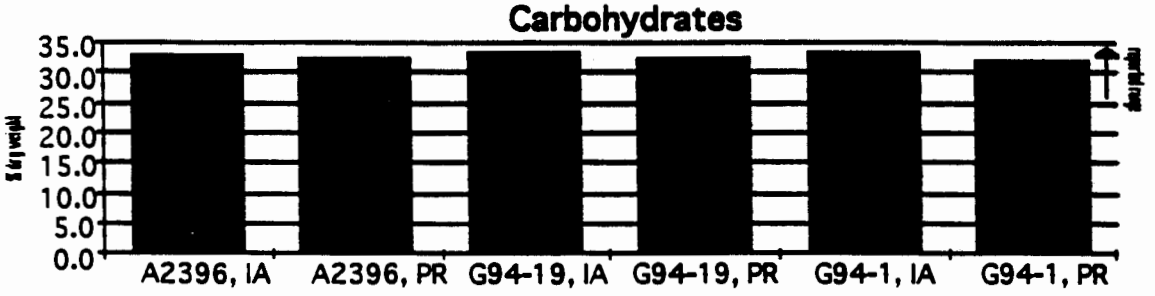


Figure 20. Carbohydrate content of high oleic and control (A2396 parent) soybean.

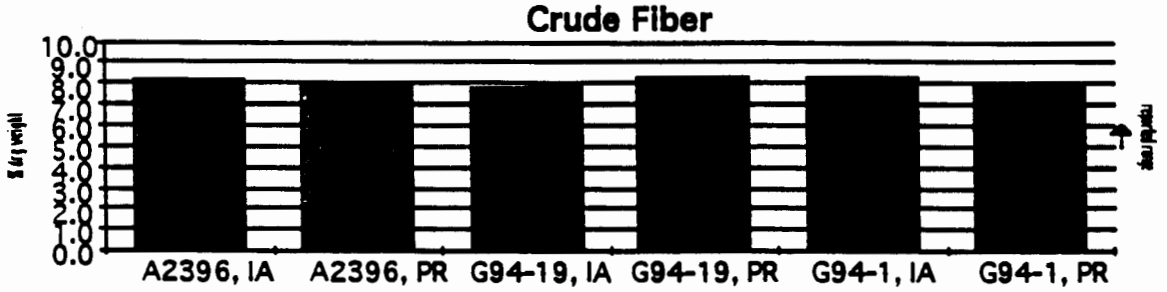


Figure 21. Crude Fiber content of high oleic and control (A2396 parent) soybean.

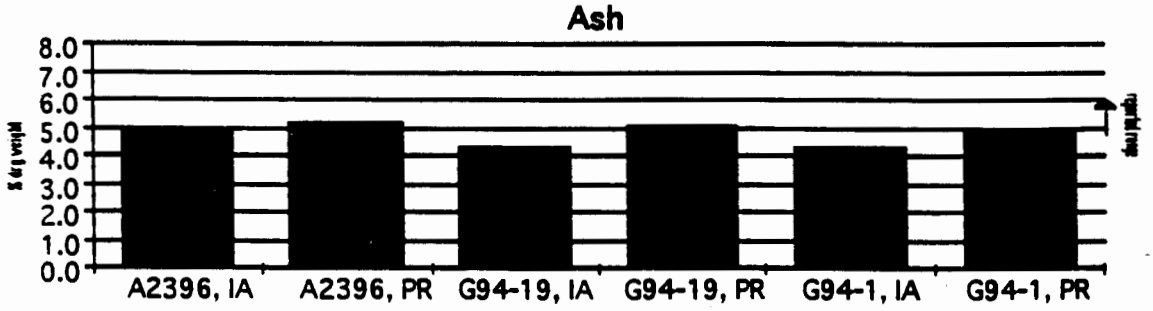


Figure 22. Ash content of high oleic and control (A2396 parent) soybean.

B. Amino Acid Composition

Soybeans are valued as a major source of a high protein animal feed supplement and have increasingly important utility in human food applications. We therefore wanted to establish that there was no change in the amino acid composition of high oleic lines. High oleic and A2396 control soybeans were analyzed for amino acid composition and the data is presented below (Figures 23-39). High oleic soybeans did not differ substantially from control in any of the 17 amino acids measured. Literature values for commodity soybean for each amino acid were as reported by Han et al. (1991).

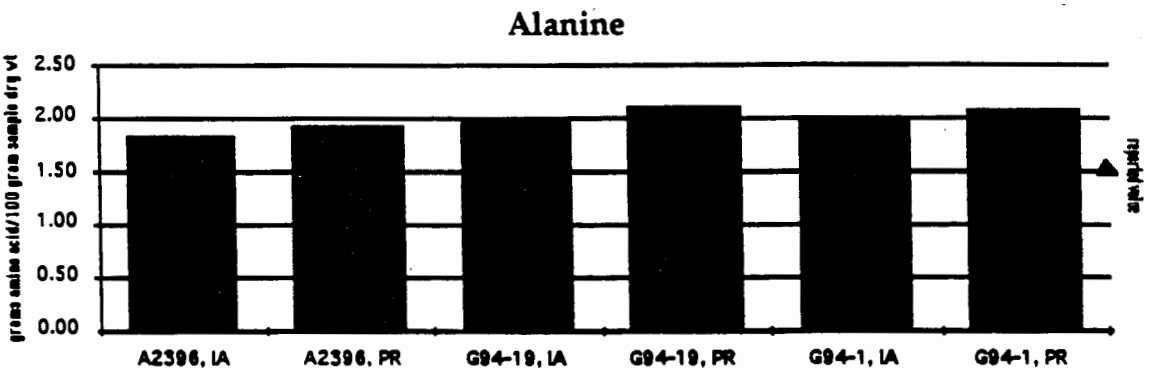


Figure 23. Alanine analysis of high oleic and control (A2396 parent) soybean.

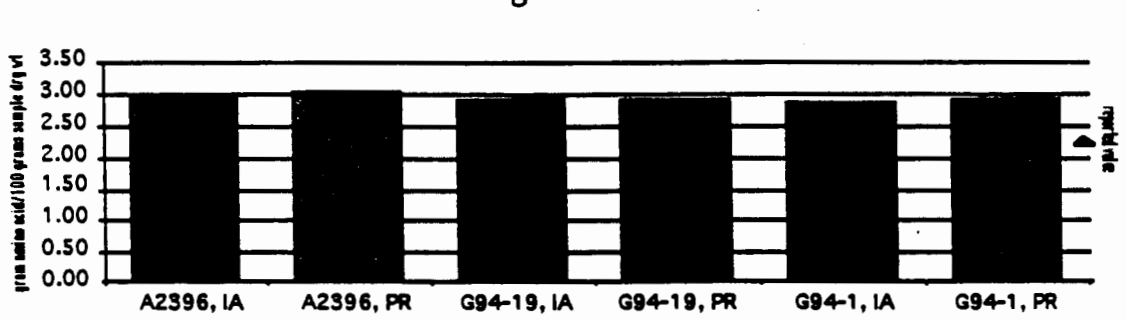


Figure 24. Arginine analysis of high oleic and control (A2396 parent) soybean.

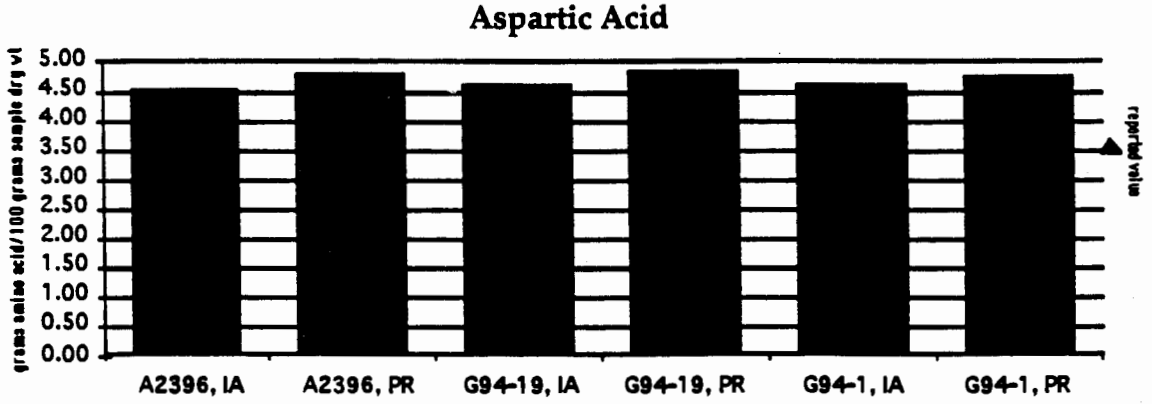


Figure 25. Aspartic Acid analysis of high oleic and control (A2396 parent) soybean.

Cystine

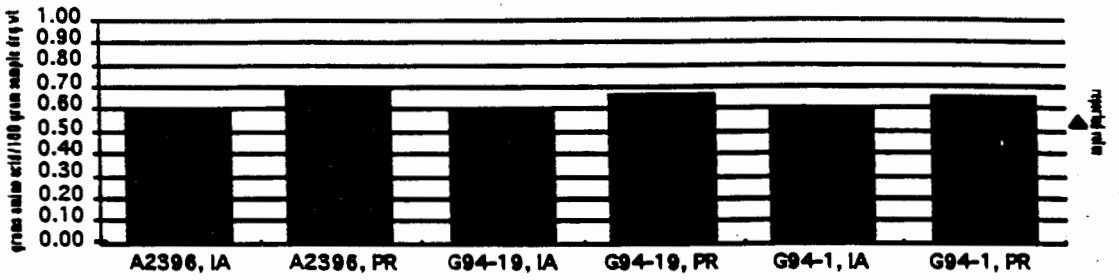


Figure 26. Cystine analysis of high oleic and control (A2396 parent) soybean.

Glutamic Acid

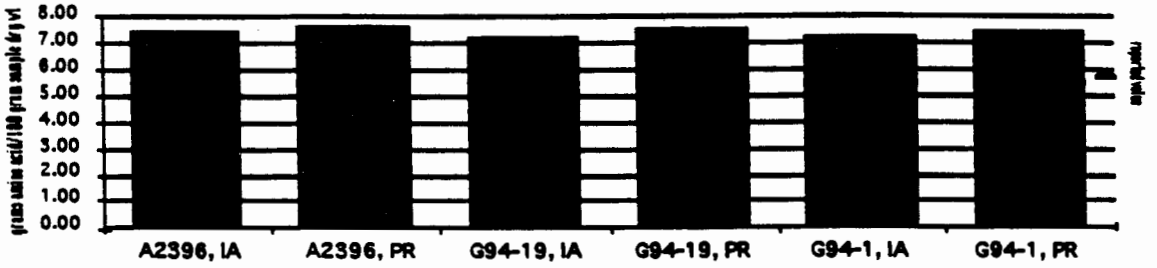


Figure 27. Glutamic Acid analysis of high oleic and control (A2396 parent) soybean.

Glycine

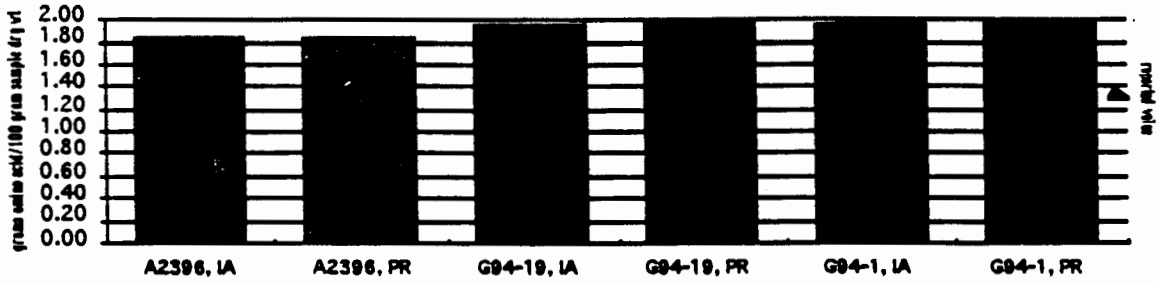


Figure 28. Glycine analysis of high oleic and control (A2396 parent) soybean.

Histidine

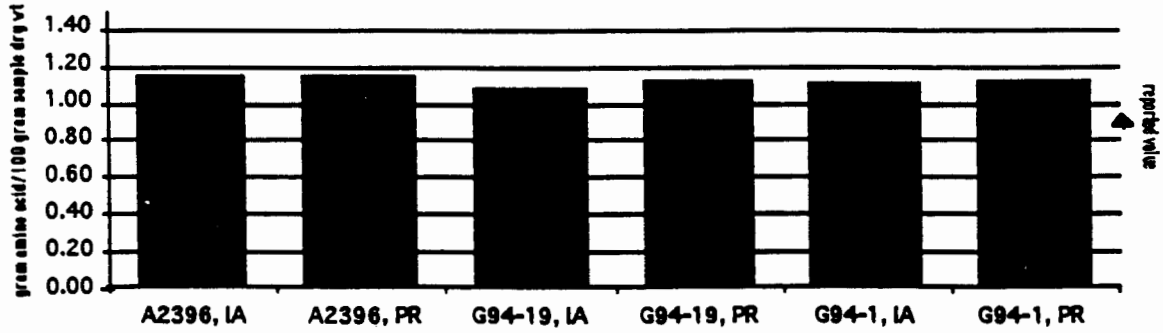


Figure 29. Histidine analysis of high oleic and control (A2396 parent) soybean.

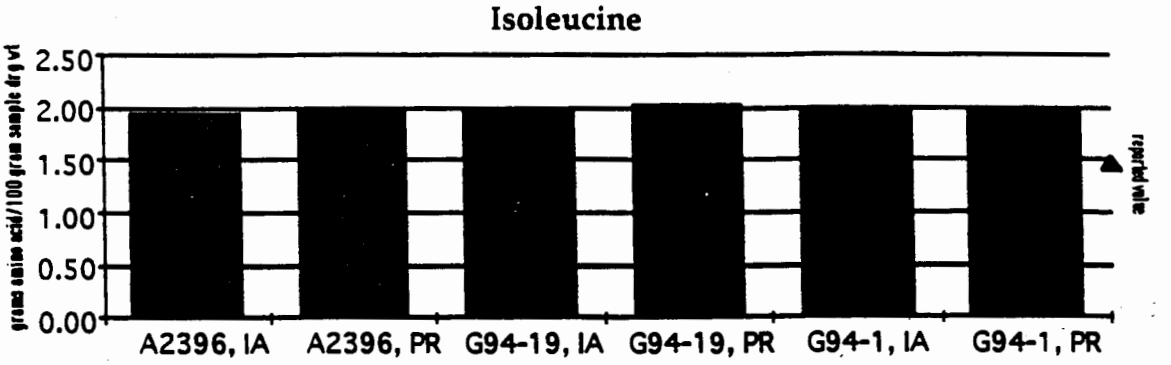


Figure 30. Isoleucine analysis of high oleic and control (A2396 parent) soybean.

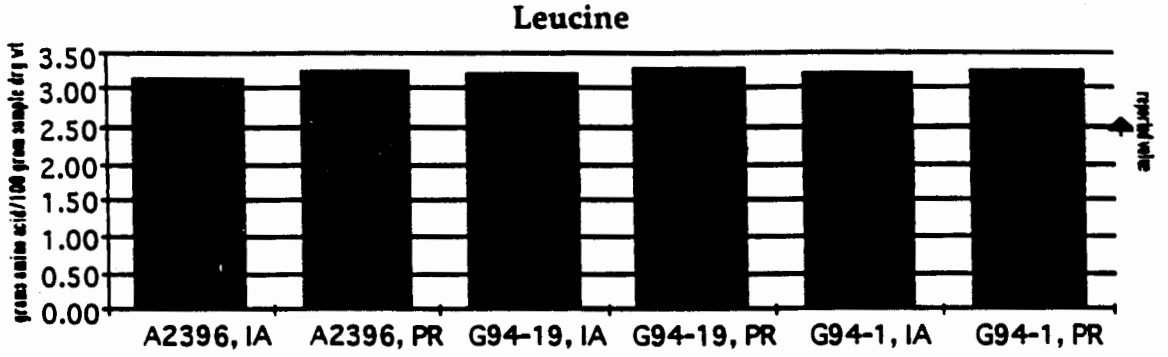


Figure 31. Leucine analysis of high oleic and control (A2396 parent) soybean.

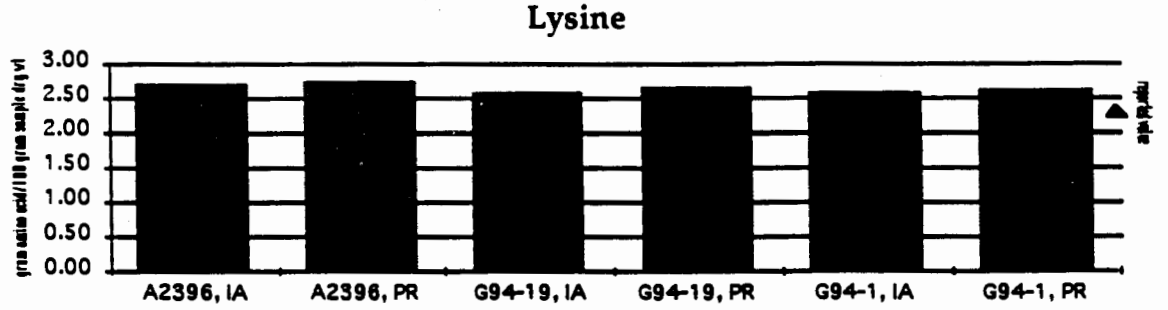


Figure 32. Lysine analysis of high oleic and control (A2396 parent) soybean.

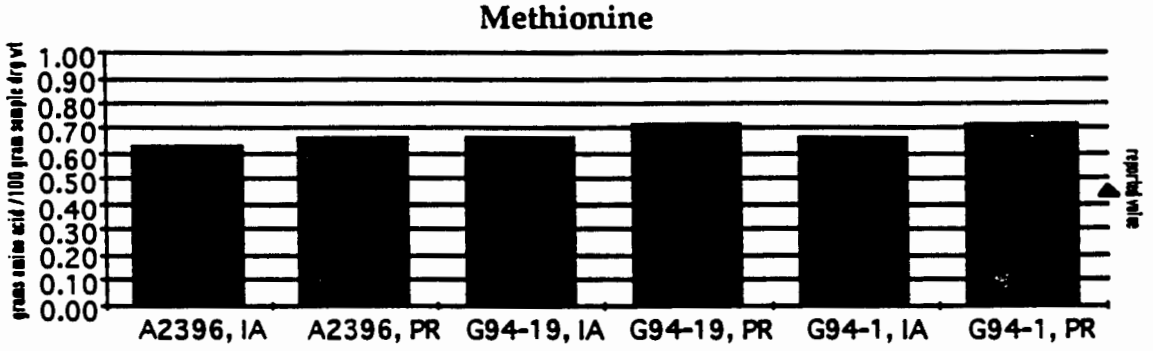


Figure 33. Methionine analysis of high oleic and control (A2396 parent) soybean.

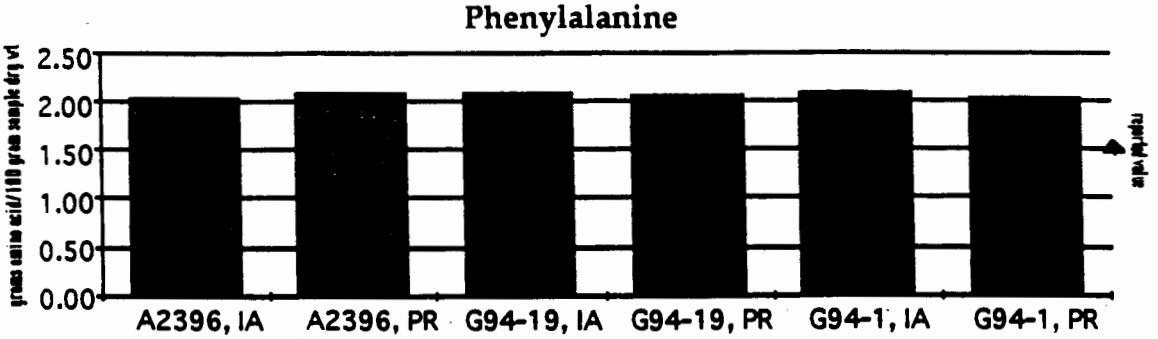


Figure 34. Phenylalanine analysis of high oleic and control (A2396 parent) soybean.

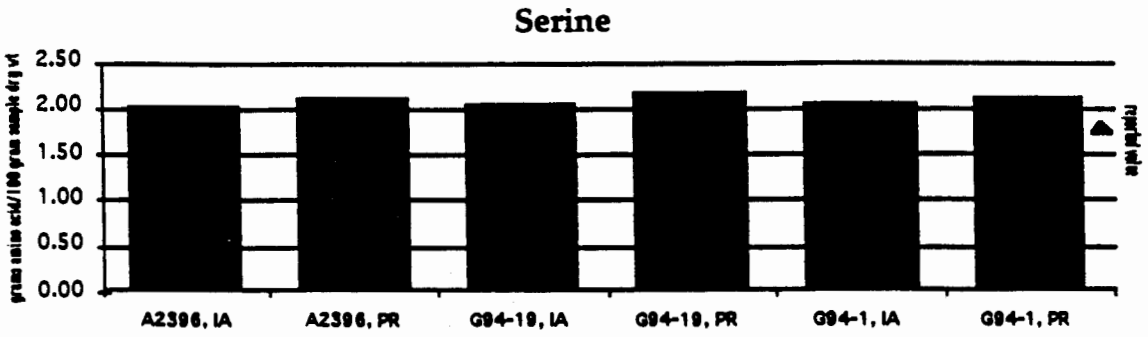


Figure 35. Serine analysis of high oleic and control (A2396 parent) soybean.

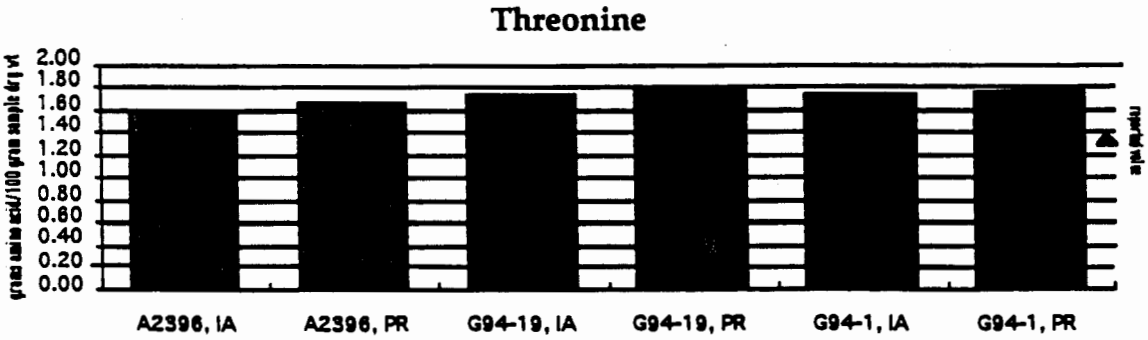


Figure 36. Threonine analysis of high oleic and control (A2396 parent) soybean.

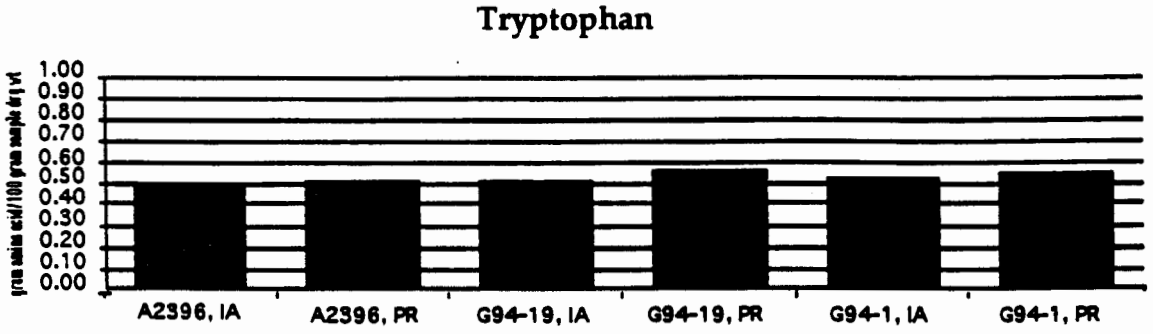


Figure 37. Tryptophan analysis of high oleic and control (A2396 parent) soybean.

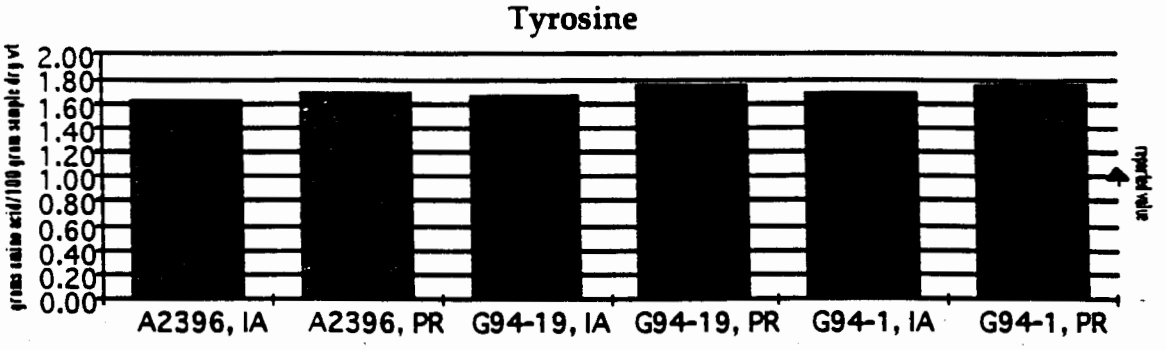


Figure 38. Tyrosine analysis of high oleic and control (A2396 parent) soybean.

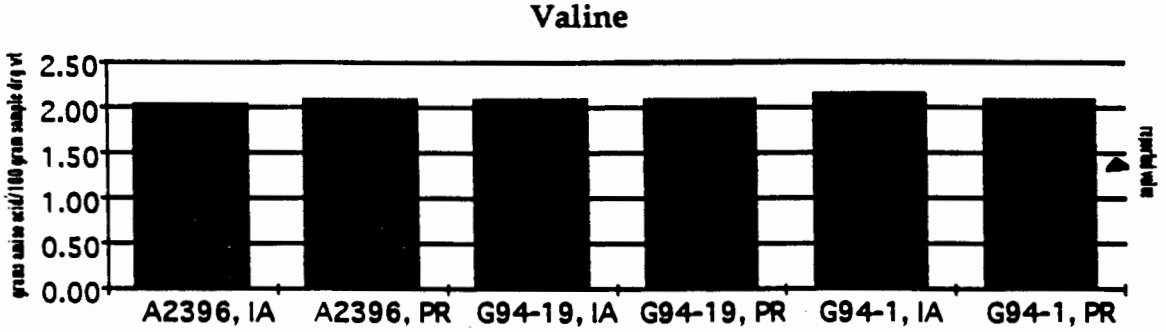


Figure 39. Valine analysis of high oleic and control (A2396 parent) soybean.

C. Fatty Acid Composition

As described previously, high oleic soybeans differ substantially from commodity soybean in the level of oleic, linoleic, linolenic, and to a lesser extent, palmitic acid present in the oil. Also, there is a trace amount of the 9,15 isomer of linoleic acid normally found only in hydrogenated soybean oils. These changes in fatty acid composition are shown in Figures 40-45. Literature values are those reported by Pryde (1980). With the exception of the 9,15 isomer, the relative abundance of minor fatty acids was similar in elite and transgenic beans and was within Codex ranges.

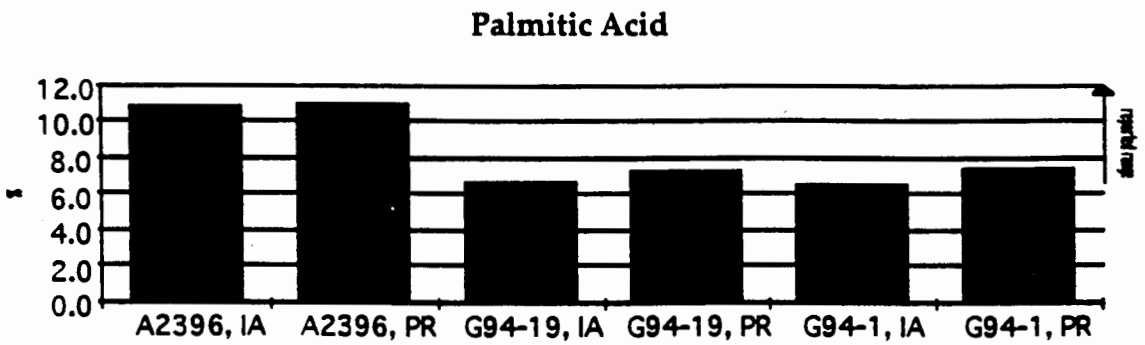


Figure 40. Palmitic Acid content of high oleic and control (A2396 parent) soybean.

Stearic Acid

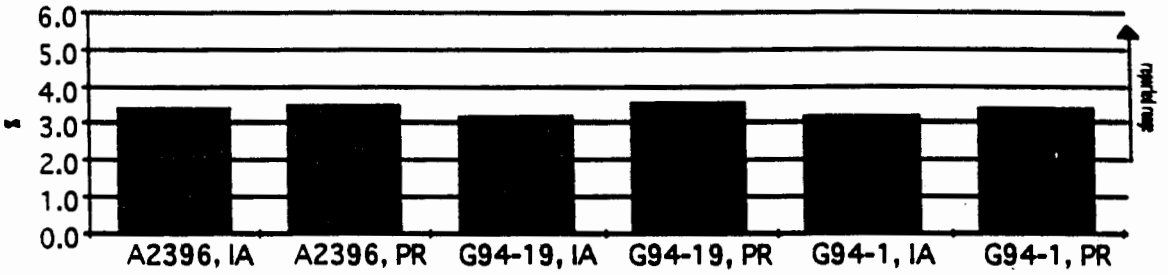


Figure 41. Stearic Acid content of high oleic and control (A2396 parent) soybean.

Oleic Acid

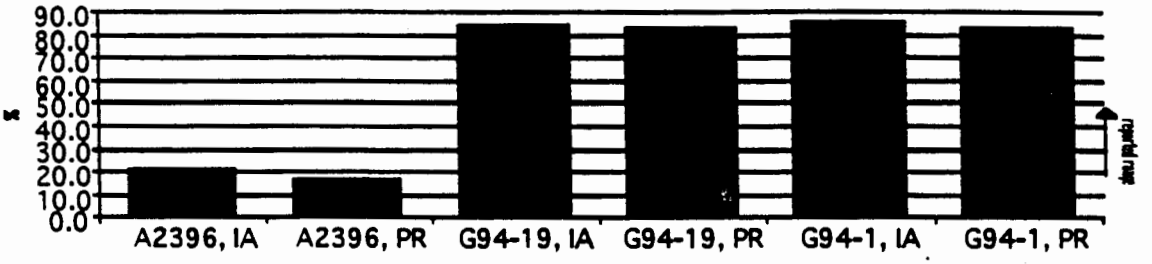


Figure 42. Oleic Acid content of high oleic and control (A2396 parent) soybean.

Linoleic Acid

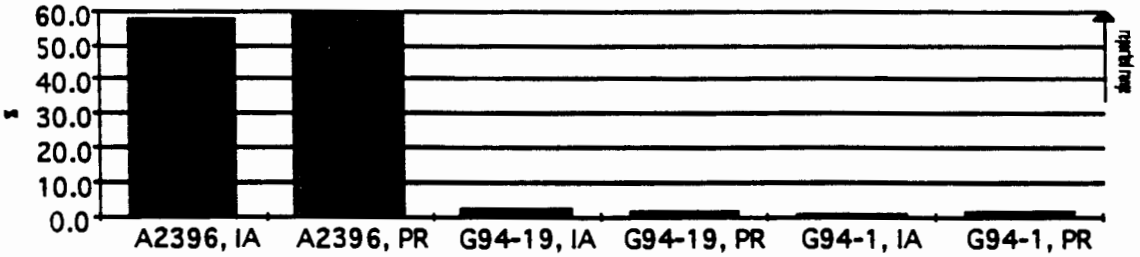


Figure 43. Linoleic Acid content of high oleic and control (A2396 parent) soybean.

Linolenic Acid

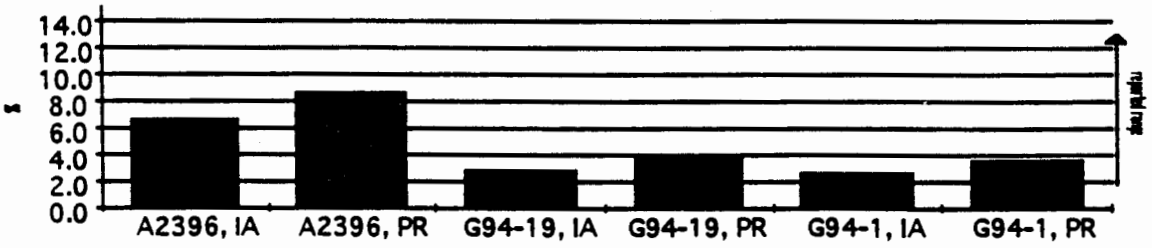


Figure 44. Linolenic Acid content of high oleic and control (A2396 parent) soybean.

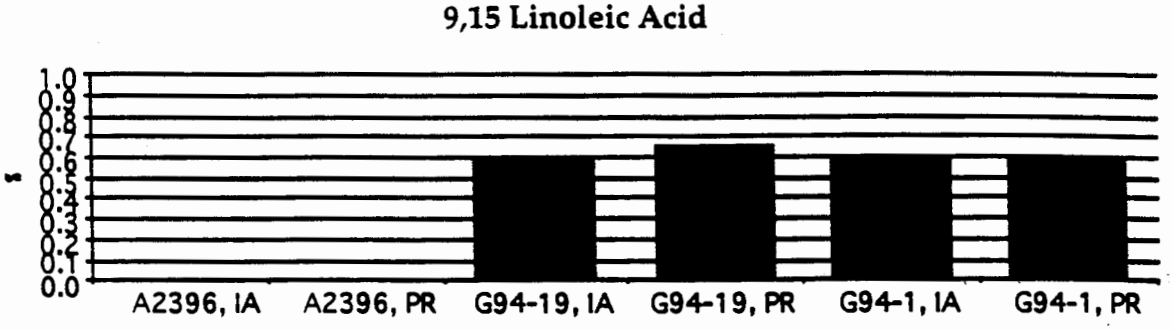


Figure 45. 9,15 Linoleic Acid content of high oleic and control (A2396 parent) soybean.

D. Identification of A Non-Methylene Interrupted Diene

A fatty acid was observed in the high oleic soybean oil which, although common in many edible oils (see Figure 46), is not present in the oil of non-transgenic soybeans:

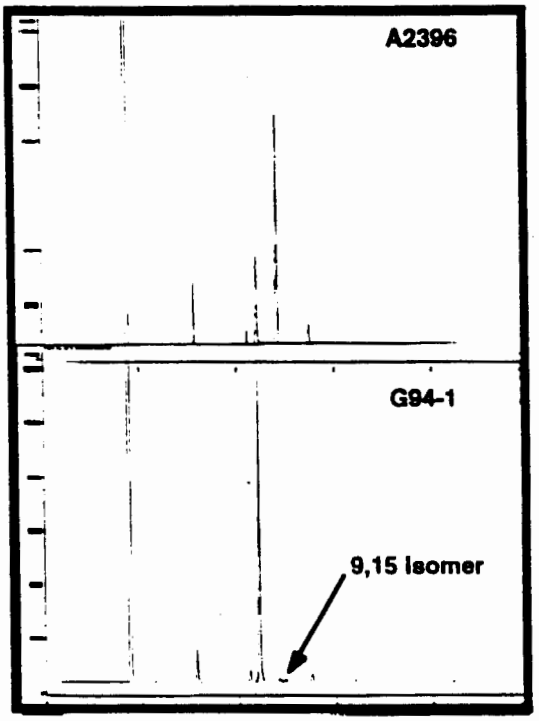


Figure 46: GC of Fatty Acid Methyl esters from elite (A2396) and transgenic (G94-1) soybeans

This fatty acid was identified as cis-9,cis-15-octadecadienoic acid which is the 9,15 isomer of linoleic acid. The linoleic acid fraction was first isolated by reverse-phase HPLC which yielded two fatty acids, one which had the retention time on the GC of 9,12-linoleic acid

and a second, unknown peak which had coeluted with the linoleic acid on the HPLC (see Figure 47).

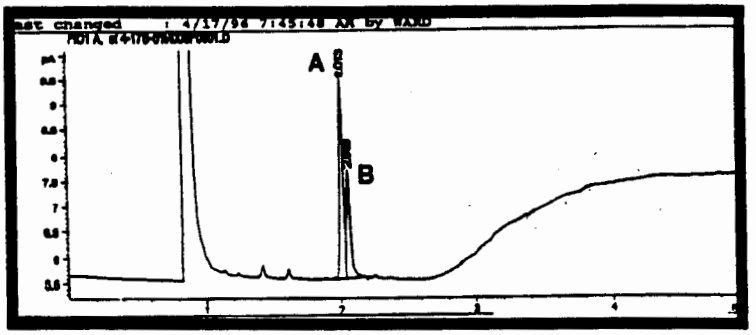


Figure 47: GC of fatty acids methyl esters derived from a fatty acid fraction purified by RP-HPLC

The second of the two peaks was identified by its mass spectra (GC-MS) as 9,15-linoleic acid (see Figure 48).

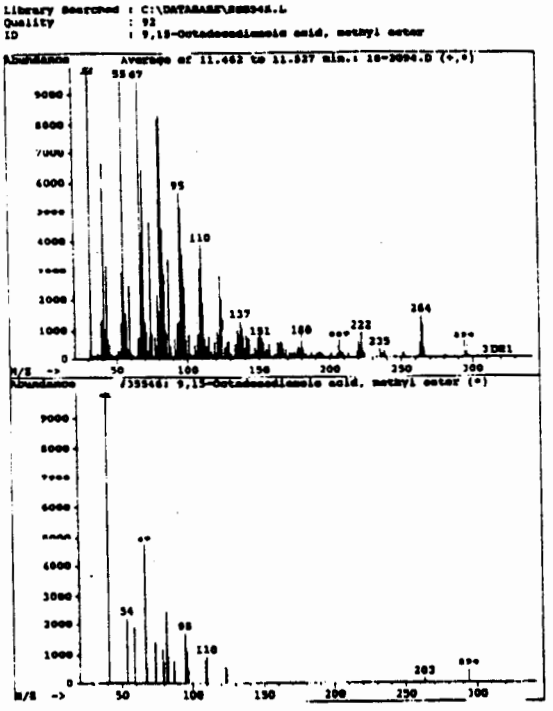


Figure 48: GCMS of 18:2-Peak B shown in Figure 2

We deduce that the isomer is a result of the activity of a δ-15 (n-3) desaturase (GmFad3) which normally inserts a δ-15 double bond into 9,12-linoleic acid (see Figure 49).

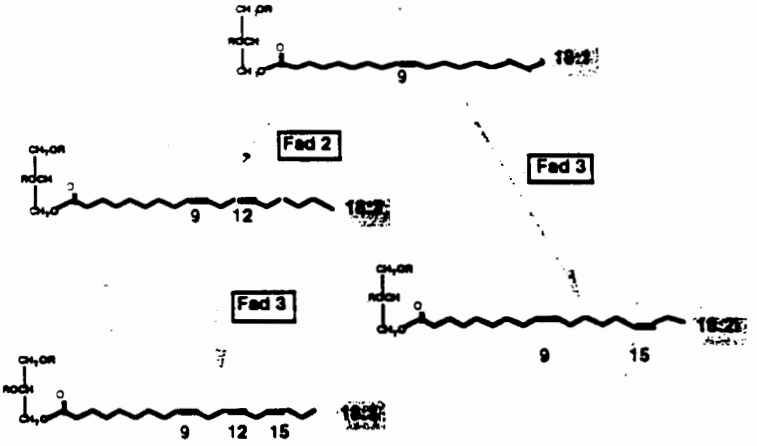


Figure 49: Addition of Double Bonds to Oleic Acid

In the transgenic plants the linoleic acid content is reduced from over 50% of the total fatty acids to less than 2% and so the GmFad3 enzyme probably creates a small amount of the isomer by putting a δ-15 double bond into 9-oleic acid. This view is supported by the results of crossing the high oleic soybeans with soybean containing a suppressed GmFad3 gene. In the resulting progeny the isomer is either reduced or eliminated as shown in Figure 50.

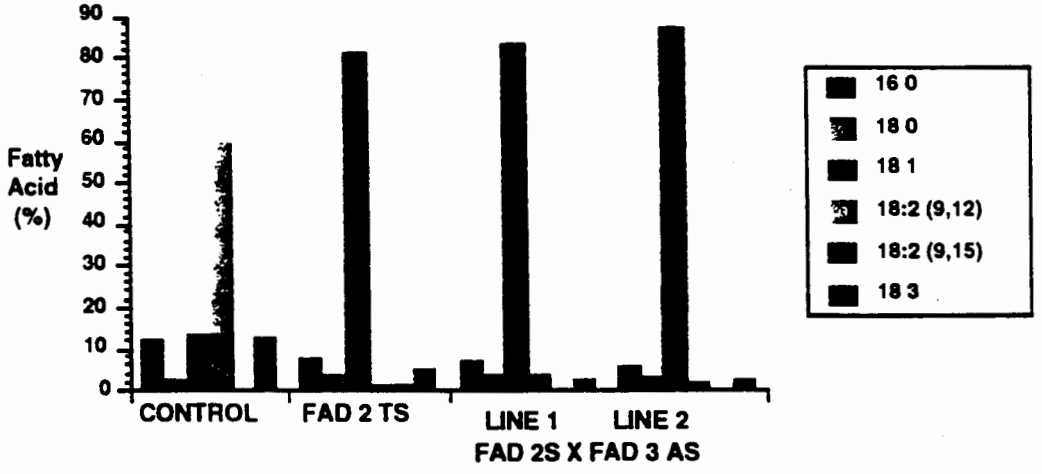


Figure 50: Fatty acid composition of A2396 (*CONTROL*), G94-1 (*FAD 2S*) and G94-1 crossed with GmFad3 antisense (*FAD 2S x FAD 3AS*) plants

The 9,15 isomer is present in the transgenic oil at less than 1% of the total fatty acid content. This isomer is also found, at concentrations ranging from 0.02% to 5.4% of the total fatty acids, in many edible sources of fat including butterfat, cheese, beef and mutton tallow, partially hydrogenated vegetable oils, human milk and mango pulp (De Jong & van der Wel, 1964; Keppler et al., 1964; Stroink & Sparreboom, 1967; Hoffman & Meijboom, 1969; Murawski et al., 1971; Mallet et al., 1985; Ratnayake & Pelletier, 1992; Werner et al., 1992; Shibahara et al., 1993).

One of the intended uses of high oleic soybean oil is to replace heavy duty shortenings produced by hydrogenation of regular soy oil. In our own studies we observed that commercially available hydrogenated soybean oil contains about twice the mol % of 9,15-isomer than high oleic oil from transgenic soybeans. The heavy duty shortening also contains other positional isomers of unsaturated fatty acids not found in high oleic soybean oil (see Figure 51).

	16:0	18:0	18:1 (1)	18:2	18:2i (2)	others (3)	18:3
A2396	10.1	3.3	14.8	62.1	0	0	9.7
G94-19(4)	6.7	3.2	86.1	0.9	0.6	0	2.6
HDS (5)	11.3	10.4	70.7	2.4	1.6	3.6	0

Notes:
(1) Includes *cis* and *trans* isomers of 18:1
(2) 9,15-isomer of linoleic acid
(3) Positional isomers of 18:1 and 18:2
(4) Oil from seeds grown at Stine, Md., 1995
(5) HDS = Commercial heavy duty soy shortening.

Figure 51: Relative fatty acid content of three soy oils determined by GC of fatty acid methyl esters

All of the other minor fatty acids detected in the high oleic soybean oil were present at a similar relative abundance to traditional elite soybeans. A complete fatty acid analysis of oil extracted from high oleic and elite (A2396) soybeans grown in Puerto Rico in the winter of 1995/96 is shown in Figure 52.

	A2396	G94-1	G94-19
14:0	trace	trace	trace
16:0	10.1	6.3	6.6
16:1	0.1	0.12	0.2
16:2	trace	trace	trace
16:3	trace	trace	trace
18:0	3.2	3.7	3.6
18:1	14.7	84.6	84.9
18:2 (9,12)	61.6	0.9	0.6
18:2 (9,15)	nd	0.8	0.7
18:3	9.5	2.4	1.9
20:0	0.2	0.4	0.5
20:1	0.2	0.4	0.4
20:2	nd	nd	nd
22:0	0.3	0.4	0.5
22:1	trace	trace	trace
24:0	0.1	0.1	0.2

Figure 52: Complete fatty acid analysis of soybean oil from elite (A2396) and high oleic transgenic soybeans (trace = less than 0.1%, nd = not detected).

E. Antinutritional Factors

Several heat labile and heat stabile antinutritional factors are known to exist in soybean including trypsin inhibitors, phytic acid, and the oligosaccharides raffinose and stachyose (Mounts et al., 1987). Trypsin inhibitors are heat labile and are destroyed during the processing of soy protein products by heat treatment. They are associated with hypertrophy and lesions of the pancreas and the loss of S-containing amino acids which limits animal growth. Phytic acid is heat stable and has been implicated in interfering with the bioavailability of minerals such as Ca, Mg and Zn. Phytic acid remains stable through most soybean processing steps. Raffinose and stachyose are associated with the flatulence resulting from ingestion of soybean flours. Further processing of soybean flours into concentrates and isolates removes these oligosaccharides. High oleic and control soybeans were analyzed for each of these constituents and the data is shown in Figures 53-56. No differences were observed between control and high oleic soybeans in any of these anti nutritional components. The reported values for phytic acid and trypsin inhibitor are those described by Han et al. (1991) and Mounts et al. (1987). The reported range for raffinose and stachyose was obtained from Dr. Ken Leto, DuPont Quality Grains (personal communication) who surveyed soluble carbohydrate content in a wide range of commodity soybean lines grown in the same locations and seasons as high oleic and A2396 soybeans.

Phytic Acid

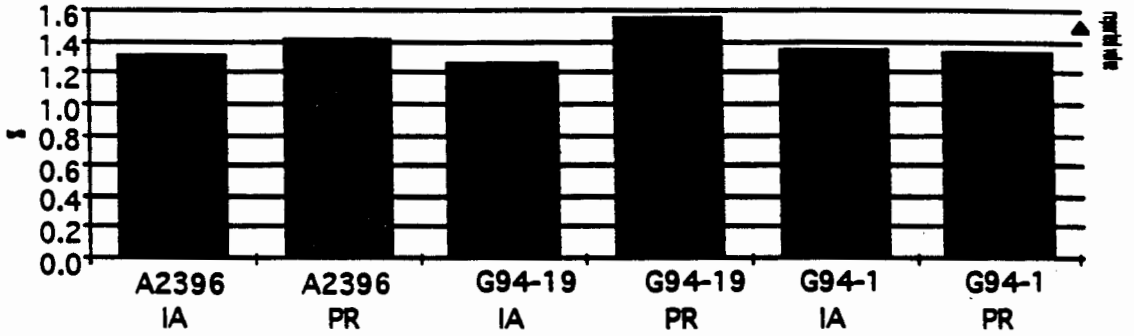


Figure 53. % Phytic Acid (dry weight basis) in high oleic and control (A2396 parent) soybean.

Trypsin Inhibitor

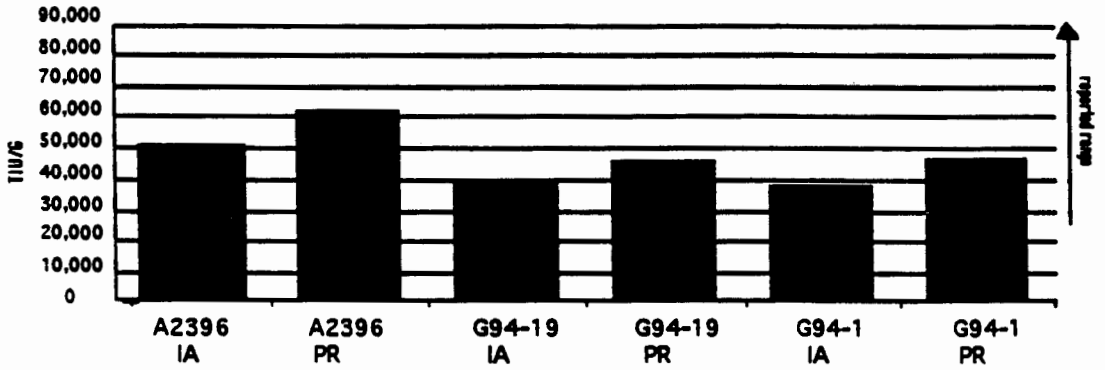


Figure 54. Trypsin Inhibitor Activity in high oleic and control (A2396 parent) soybean.

Stachyose

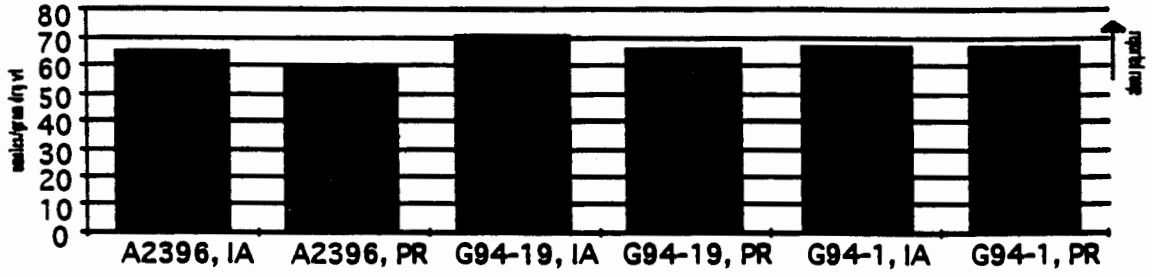


Figure 55. Stachyose content of high oleic and control (A2396 parent) soybeans.

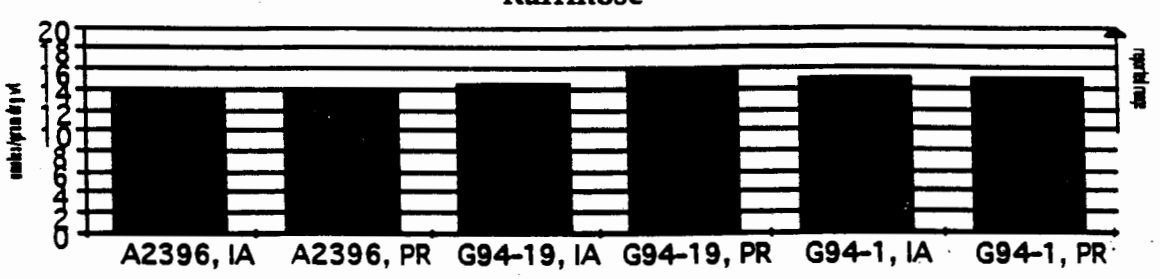


Figure 56. Raffinose content of high oleic and control (A2396 parent) soybeans.

F. Isoflavones

Soybeans contain several biologically active phytoestrogens known as isoflavones which are present in various concentrations in soy protein products. Isoflavone content varies between soybean varieties and is affected by environmental factors. Minimally processed soy products contain higher levels of isoflavones than do products that undergo an alcohol wash as part of their purification. Recent studies suggest that the isoflavones may act as cancer-protective agents since they have been linked to a variety of anticarcinogenic activities including antioxidant, radical scavenging, serum cholesterol lowering and antiestrogenic and antiproliferative properties (Wang et al., 1994, Messina et al, 1991). At high doses they have been associated with reproductive problems in animals feeding on clover silage (Wang et al, 1990). The major isoflavones in soybeans and soybean products include daidzin, genistin, and their corresponding aglycons, daidzein, genistein. Glycitin and glycitein also occur in trace amounts. We have analyzed high oleic and control soybeans for isoflavone content and the data is shown in Figures 57-59. The concentration of total genistein in high oleic soybeans fell below the reported range however since the control soybeans also fell below this range, there is indeed no difference when compared with commodity soybean. No differences were observed between control and high oleic soybean in either total daidzein or glycitien content. Reported ranges were those provided by Dr. P. Murphy, Iowa State University (personal communication).

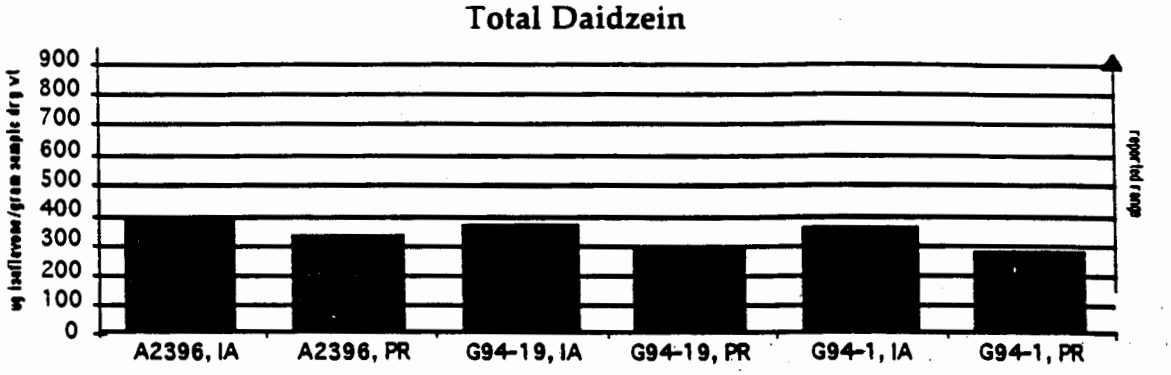


Figure 57. Total Daidzein in high oleic and control (A2396 parent) soybean. Total daidzein represents the sum of individual isomers normalized for molecular weight differences. Isomers included daidzin, malonyl daidzin, acetyl daidzin, and daidzein.

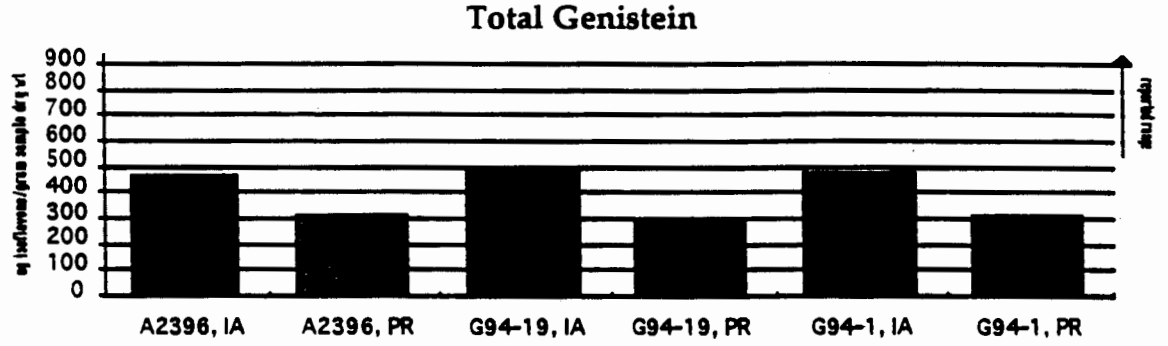


Figure 58. Total Genistein in high oleic and control (A2396 parent) soybean. Total genistein represents the sum of individual isomers normalized for molecular weight differences. Isomers included genistin, malonyl genistin, acetyl genistin, and genistein.

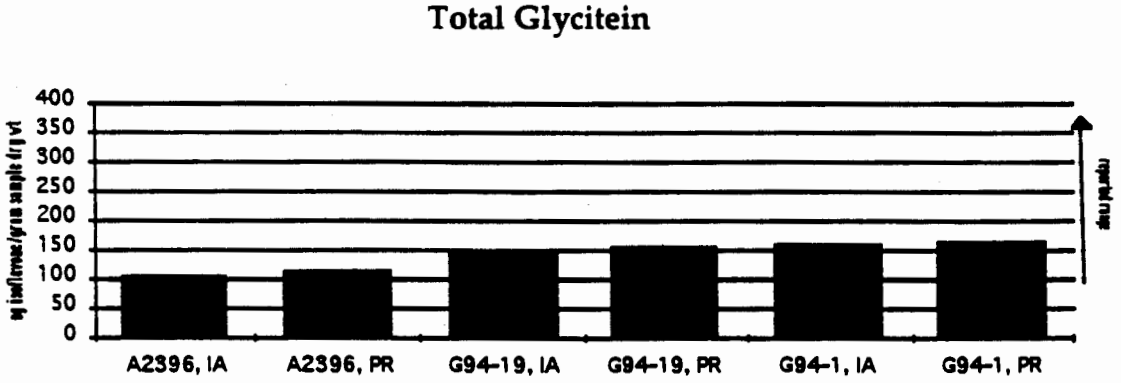


Figure 59. Total Glycitein in high oleic and control (A2396 parent) soybean. Total glycitein represents the sum of individual isomers normalized for molecular weight differences. Isomers included glycitin, malonyl glycitin, acetyl glycitin, and glycitein.

G. Storage Proteins - Changes in the Ratio of β-Conglycinin to Glycinin

There was a reduction in the concentration of the α and α′ subunits of β-conglycinin in the high oleic acid transgenic soybeans when compared with A2396 beans. This was coincident with an increase in the concentration of the acidic and basic subunits of glycinin in the high oleic acid transgenic soybeans, in addition to an increase in the concentration of the A2B1A glycinin precursor. Based on SDS-gel electrophoresis of total protein extracts, the profile of other storage proteins appeared to be identical in A2396 and high oleic acid transgenic beans.

Soybean 7S and 11S globulins are two major storage proteins, acounting for about 70% of total meal protein (Kinsella, 1979). The 7S fraction is made up of the α, α′ and β subunits of β-conglycinin. The 11S fraction is made up of the acidic (A) and basic subunits (B) of glycinin. The two globulins show considerable differing effects on the functional aspects of the soy proteins such as thermal stability, gel-making ability and emulsifying capacity (Kinsella, 1979, Utsumi & Kinsella, 1985; Kitamura, 1995). It is anticipated that increasing the content of glycinin (11S) subunits and decreasing the content of β-conglycinin (7S) subunits will improve the functionality of soy proteins in various foods (Kitamura, 1995).

A reduction of α-type subunits of β-conglycinin has previously been observed in naturally occuring soybean varieties (Ladin et al., 1984) and in soybeans with induced mutations (Takahashi et al 1994, Kitamura, 1995). A similar effect has been achieved in the high oleic soybeans described as a result of promoter cosuppression. The phenomenon of promoter cosuppression has been previously documented for other genes and plants (Brusslan & Tobin, 1995). In the high oleic soybeans the concentration of β-conglycinin α and α′ subunits have been reduced and replaced with glycinin subunits. This was a result of cosuppression of the α and α′ subunits mediated by the α′ promoter sequence used in the GmFad 2-1 vector (pBS43). This observation is described in detail below.

Proteins were extracted from the meal of field-grown transgenic beans (G94-1, G94-19) and from non-transgenic A2396 beans (wt) and analyzed by SDS-Gel electrophoresis as shown in Figure 60.

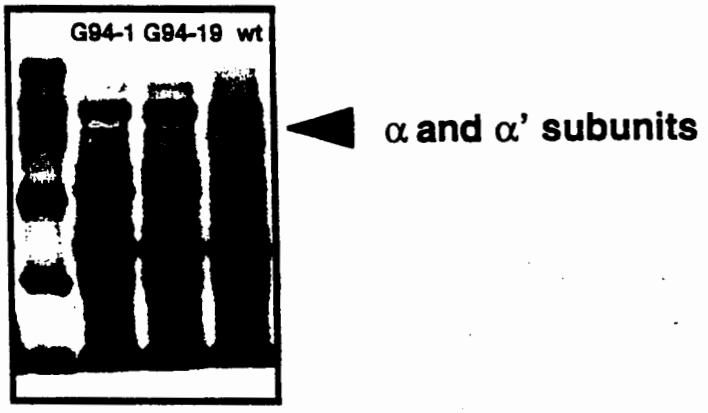


Figure 60: SDS Gel electrophoresis of total proteins from transgenic (G94) and elite (wt) soybean lines

Two major protein bands were absent in the transgenic samples when compared with controls. These bands represent the α and α' subunits of β-conglycinin. Western blots using antibodies against β-conglycinin confirmed that the two alpha-type subunits were greatly reduced in concentration although not completely suppressed as shown in Figure 61.

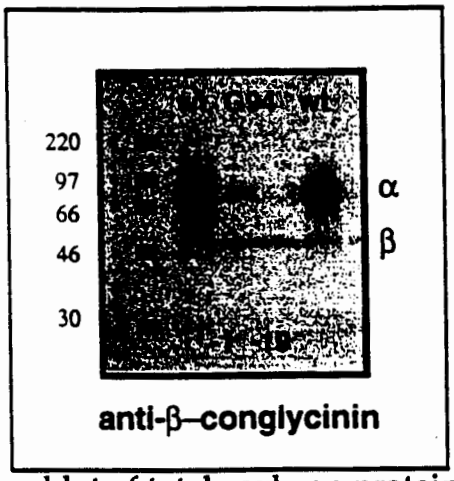


Figure 61: Western blot of total soybean proteins from transgenic (G94) and elite (wt) soybean lines using anti-β-conglycinin antibodies

Three bands were observed to increase in response to the decreased β-conglycinin subunit. Two of these correspond to acidic (A) and basic (B) subunits of glycinin. The third was a polypeptide with an estimated molecular weight of about 55 kDa. The polypeptide was isolated and was determined to be the precursor of the A2 and B1A subunits of glycinin by N-terminal amino acid sequencing. A Western blot confirmed that the newly identified band cross-reacted with antibodies raised against glycinin (see Figure 62).

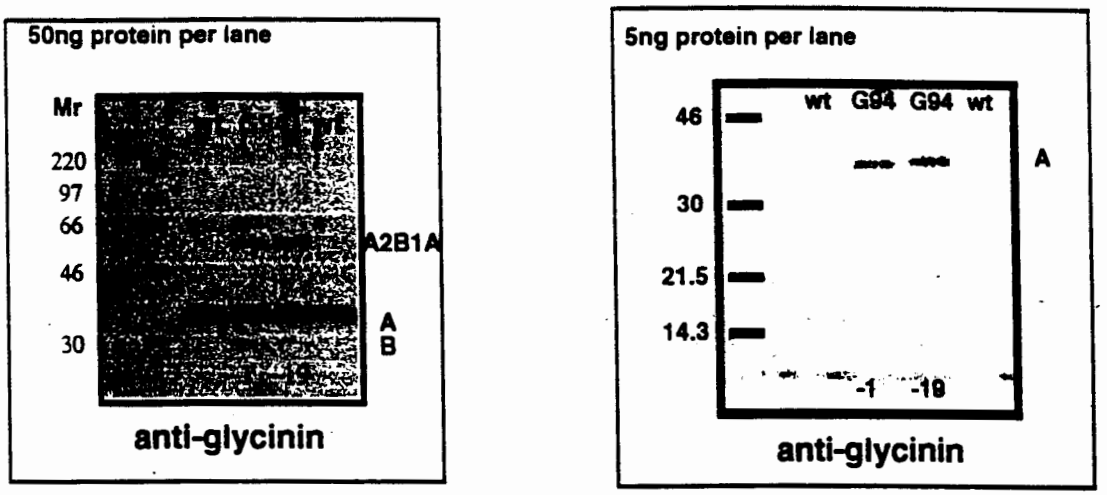


Figure 62: Western blot of total soybean proteins from transgenic (G94) and elite (wt) soybean lines using anti-glycinin antibodies

Western blotting also confirmed, after prolonged exposure, that the precursor is also present in A2396 soybeans, but at a lower concentration (see Figure 63).

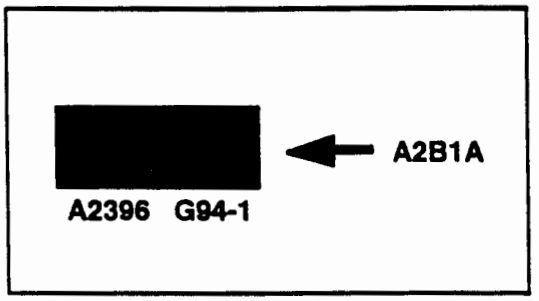


Figure 63: Western blot of glycinin A2B1A precursor from transgenic (G94) and elite (A2396) soybean lines using anti-glycinin antibodies

The total protein content and the amino acid composition of the transgenic meal was unchanged by these changes in the ratio of glycinin to β-conglycinin. The allergenicity of high oleic acid soybeans was evaluated by Dr. Samuel Lehrer of Tulane University, as reported in Appendix 1. Results demonstrate that there is no significant quantitative or qualitative difference between the transgenic and elite soybeans with regard to their allergen content.

X. Improved Oxidative Stability of High Oleic Acid Transgenic Soybeans

Modifying the fatty acid composition of oilseeds to improve the functional performance of vegetable oils has been the goal for many plant breeders over the years. One goal has been to improve the oxidative stability of oils by reducing the amount of polyunsaturated fatty acids. Using traditional crop breeding techniques, commercial varieties with modified fatty acid composition have been produced from sunflower, canola, peanut, soybean, and safflower. High oleic sunflower oil, high oleic/low linolenic canola oil, and low linolenic soybean oil are sold commercially as premium edible oil products.

Soybean oil has poor oxidative stability due to naturally high levels of polyunsaturated fatty acids. This limits its use to those applications that do not require a high degree of stability. Processors have responded by hydrogenating soybean oil: reducing the levels of linoleic and linolenic acid thereby extending the range of applications in which it can be used. However, hydrogenation has undesirable consequences including the formation of trans fatty acid isomers and a characteristic "hydrogenated flavor". Spurred by the initial studies which indicated that trans fatty acids may affect blood cholesterol levels similar to that of saturated fats, questions regarding the health implications of dietary trans fatty acids continue to be hotly debated in the industry.

A natural soybean oil with improved oxidative stability characteristics similar to that of partially hydrogenated soybean oil would be of considerable importance to the food industry. We therefore wanted to determine if the oil derived from high oleic soybeans was more stable than commodity soybean oil and how it compared to other oils currently sold in the high stability market. For this study, refined, bleached, and deodorized (RBD) high oleic soybean oil, control soybean oil, and high oleic corn oil were produced under contract by POS Pilot Plant Corporation (Canada). All other oils were commercial samples obtained from the manufacturer.

There are many methods used to evaluate an oil's stability but one of the most common is a test known as the AOM (Active Oxygen Method). This is an accelerated oxidation test in which an oil is aerated under a constant, elevated temperature (97.8 C) and degradation is monitored by measuring peroxide accumulation. The end point, or induction time, is determined by the number of hours required to reach a peroxide value of 100 meq/kg. Thus the longer the induction time the more stable the oil. Almost all commercial oil samples specify an AOM induction time as a component of the technical data sheet. Shown below is the AOM induction time for high oleic soybean oil compared to control soybean oil (see Figure 64):

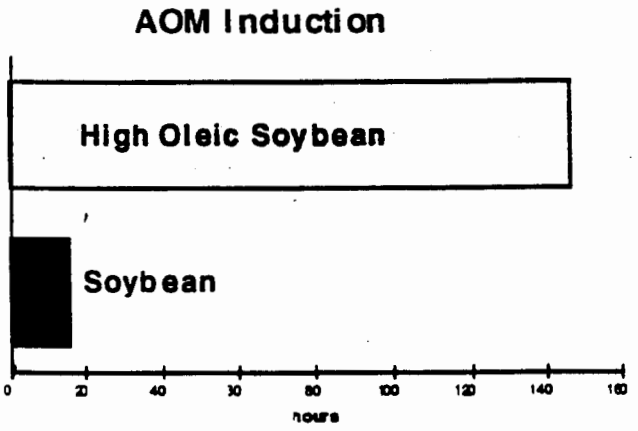


Figure 64. AOM induction time of high oleic and control soybean oil.

We also compared (Figure 65) the high oleic soybean oil to other natural high stability oils (high oleic canola, high oleic sunflower, high oleic corn, and low linolenic soybean):

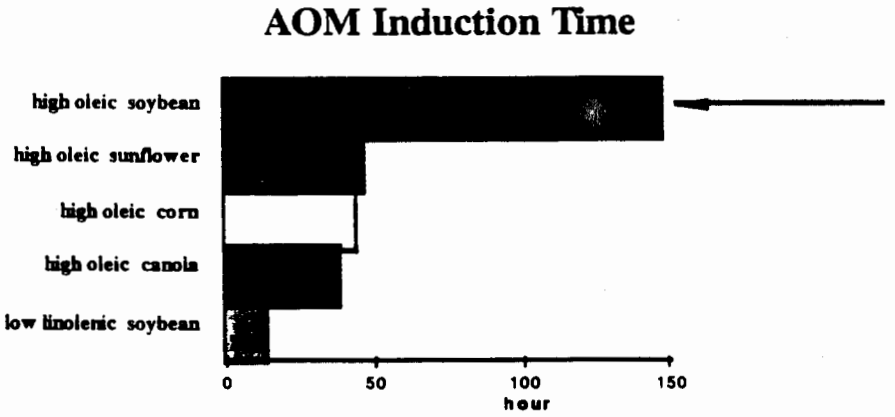


Figure 65. AOM induction time for high oleic soybean and other natural high stability oils.

The oil was also compared with frying and hydrogenated soybean oils containing TBHQ (an antioxidant) and silicone (antifoam agent), see Figure 66.

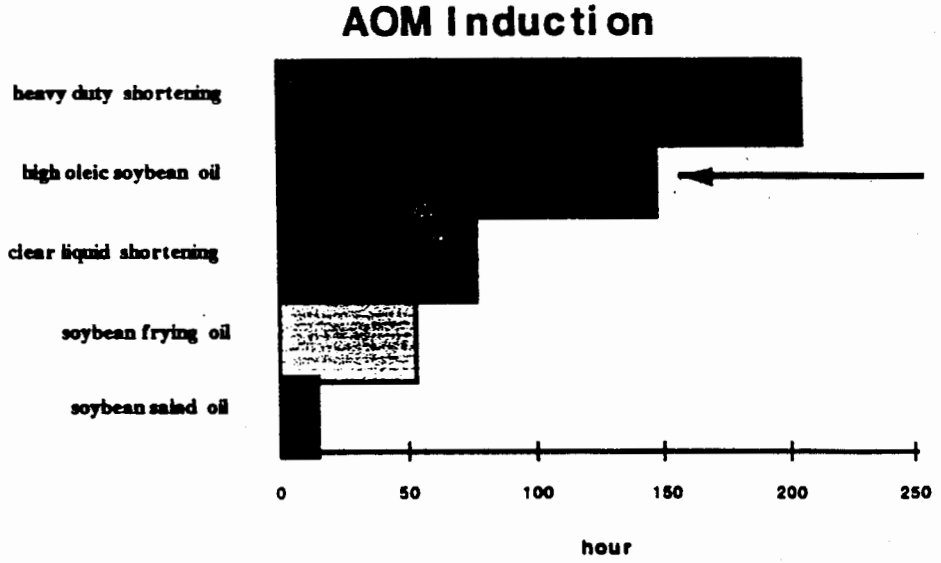


Figure 66. AOM induction times for high oleic soybean oil and hydrogenated or additive-enhanced soybean oils.

Fatty acid compositions for each of the oils is shown below in Figure 67 and analytical data indicating oil quality and content is shown in Figure 68.

Fatty Acid Composition

	16:0	18:0	18:1 (cis)	18:1 (trans)	18:2	18:3
high oleic soybean	6.4	3.3	85.6		1.6	2.2
control soybean	10.4	4.1	22.9		52.9	7.5
heavy duty shortening	10.5	10.2	42.5	31.7	3.8	0.0
clear liquid shortening	11.3	4.7	40.4	8.8	34.7	0.0
soybean frying oil	11.4	4.5	25.1		53.0	6.0
high oleic sunflower	3.0	3.9	84.1		7.9	0.5
high oleic corn	8.8	2.3	64.3		22.2	1.1
low linolenic soybean	10.6	5.1	28.3		53.6	2.3
high oleic canola	4.0	2.9	76.9		10.4	4.1

Figure 67. Fatty Acid composition of test oils.

Analytical Data- RBD Oil

	PV	FFA	Color	Additives
	meq/kg	%	Lov 5 1/4	
high oleic soybean	0.01	0.01	4.5Y.5R	none
control soybean	0.01	0.03	2Y.4R	none
heavy duty shortening	0.06	0.06	2Y.4R	TBHQ, silicone
clear liquid shortening	0.09	0.04	2Y.3R	TBHQ, silicone
soybean frying oil	0.08	0.04	3Y.5R	TBHQ, silicone
high oleic sunflower	0.37	0.01	4Y0.7R	none
high oleic corn	0.00	0.01	2Y0.4R	none
low linolenic soybean	0.42	0.02	10Y1R	none
high oleic canola	0.55	0.03	10Y2R	none

Figure 68. Analytical and content data of test oils.
PV=peroxide value, FFA=free fatty acids

As indicated by the AOM data above, high oleic soybean oil has dramatically improved oxidative stability over control soybean oil as well as any of the other natural high stability oils currently on the market. When compared with hydrogenated oils, it is superior to a liquid partially hydrogenated product containing antioxidants and can likely replace solid shortening products in many applications. With its dramatically improved fatty acid spectra, high oleic soybean oil promises to greatly extend the usage of soybean oil in a wide variety of food applications, without the need for chemical hydrogenation.

XI. Agronomic Performance of High Oleic Acid Transgenic Soybeans

A. Field Tests of Subline G94-1, G94-19 and G168

Field tests were carried out with high oleic acid transgenic soybean sublines G94-1, G94-19 and G168 in the summer of 1995, in Puerto Rico during winter 1995/1996, and in summer of 1996 at approximately 25 sites under APHIS notifications 95-088-08N, 95-107-08N, 95-257-10N, 96-071-18N and 96-115-02N. Some other isogenic sublines were also tested in earlier field studies, but G94-1, G94-19 and G168 were the sublines chosen for advancement and testing in the larger studies carried out in Puerto Rico and in the summer of 1996. The purpose of the tests varied, with early tests to select sublines with superior agronomic characteristics, testing in different environments, yield studies versus commercial soybean varieties, etc. Observations were made on agronomic characteristics and disease and pest characteristics.

B. Agronomic Characteristics

During the course of the field studies reported in Appendix 2, soybean breeders made visual observations to evaluate seed emergence, plant appearance, seed set, flowering, maturity. Sublines chosen in 1995 for advancement were visually identical to the variety A2396. In comparative yield studies in 1996, yields were slightly below those of variety A2396 and very similar to some other commercial varieties. The yields in some of the large field plots carried out under notification 96-15-02N, were somewhat low due in large part to the late planting and the adverse weather conditions during and shortly after planting. Off-types in the range of 0.05 - 0.06% range were found in these large grow-outs, which is not considered unusual.

The high oleic levels in high oleic acid transgenic soybean sublines G94-1, G94-19 and G168 consistently was at the 82-85% level. Plots were evaluated for volunteers in the Spring of 1996 and none were found.

C. Disease and Pest Characteristics

In addition to evaluation of the plants for agronomic characteristics, plant breeders and other investigators made visual observations to determine the prevalence of plant diseases and insects. In none of the field trials was any difference noted in disease susceptibility and incidence of insect attack between the transgenic sublines and the non-transgenic counterpart A2396. In the large grow-outs carried out under notification 96-15-02N disease there were no unusual occurrences of insect attack or disease pressure.

XII. Environmental Consequences to the Noncontained Use of High Oleic Acid Transgenic Soybean Sublines G94-1, G94-19 and G168

High oleic acid transgenic soybeans will be grown just as any other soybean variety, except they will be grown identity preserved so that they can be stored and processed separate from regular soybeans. This is to allow for the capturing of the improved high oleic acid transgenic soybean oil. The soybeans will be grown as any other soybean variety and would not be expected to have any greater potential to become a weed than the large number of soybean varieties, including non-transgenic high oleic soybeans, that currently are grown under non-contained conditions. Since the high oleic acid trait in the transgenic soybeans is produced by sense suppression and no new protein is produced in the soybean plant or seed, there is no potential for exposing humans, animals, or other organisms to some new protein component, from consumption of high oleic acid transgenic soybeans.

As discussed in the earlier section on the soybean family, there exists only a minimal potential for outcrossing of soybeans with other soybeans, and there likewise would be only a minimal potential for outcrossing of the high oleic acid transgenic soybeans with any other soybeans, because of the self-pollination characteristics of the soybean plant. In the United States there are no weedy relatives of cultivated soybeans with which the transgenic soybeans could have the possibility of crossing. Even if there were some minimal cross-pollination of high oleic acid transgenic soybeans with a neighboring non-transgenic soybean, the consequences to the environment would be nil.

The possibility of transfer of genetic information to organisms with which soybeans cannot interbreed, i.e. the movement of transgenes from transgenic plants to microorganisms has been reviewed in many APHIS environmental assessments. As pointed out in the USDA Environmental Assessment of a field test of a transgenic cotton plant, APHIS noted that "Horizontal movement of the introduced genes is not known to be possible. No mechanism that can transfer an inserted gene from a chromosome of a transformed plant to a chromosome of another organism has been shown to exist in nature." (APHIS, 1991).

XIII. Statement of Grounds Unfavorable

No unfavorable information or data has been demonstrated for high oleic acid transgenic soybean sublines G94-1, G94-19 and G-168 derived from transformation event 260-05.

XIV. References

Allison, D.B., Denke, M.A., Dietschy, J.M., Emken, E.A. Nicolosi, R.J. (1995) Am. J. Clin. Nutr. 62:655-707

American Society of Clinical Nutrition: Task Force on Trans Fatty Acids. Position Paper (1996) Am. J. Clin. Nutr. 63:663-670

American Soybean Association (1995) Soy Stats Reference Guide.

Ascherio, A., Willett, W.C. (1995) J. Nutr. 125: 647-655.

Bolton-Smith, C., Woodward, M., Fenton, S., Brown, C.A. (1996) Euro. Heart. J. 17:837-845

Carlson, J.B. and Lersten, N. R. (1987) in Soybeans: Improvement, Production and Uses (Wilcox, J.R. ed.) pp110-113. 2nd. Ed.

Christou, P., McCabe, D.E., Martinell, B.J., Swain, W.F. (1990) Trends Biotechnol. 8:145-151.

Ceska, M., Lundkvist, U. (1972) Immunochemistry 9:1021-1030

Barker, S.J., Harada, J.J., Goldberg, R.B. (1988) Proc. Natl. Acad. Sci. (USA) 85:458-462.

Beachy, R.N., Chen, Z.-L., Horsch, R.B., Rogers, S.G., Hoffmann, N.J., Fraley, R.T. (1985) EMBO J. 4:3047-3053.

Berry-Lowe, S.L., McKnight, T.D., Shah, D.M., Meagher, R.B. (1982) J.Mol. Genet. 1:483-498.

Bonnassie, S., Oreglia, J., Sicard, A.M. (1990) Nucleic Acid Res. 18:6421.

Brusslan, J.A., Tobin, E.M. (1995) Plant Molec. Biol. 27:809-813.

de Boer, D., Bakker, H., Lever, A., Bouma, T., Salentijn, E., Weisbeek, P. (1991) EMBO J. 10:2765-2772.

de Jong, K., van der Wel, H. (1964) Nature 202:553-555.

Dictenberg, J.B., Proczuk, A., Hayes, K.C. (1995) J. Nutr. Biochem. 6:353-361

Doyle, J.J., Schuler, M.A., Godette, W.D., Zenger, V., Beachy, R.N., Slightom, J.L. (1986) J. Biol. Chem. 261:9228-9238.

Environmental Assessment and Finding of No Significant Impact, APHIS Permit No. 91-025-02, April 29, 1991

Falco, S.C., Guida, T., Locke, M. Mauvais, J., Sanders, C., Ward, R.T., Weber, P. (1995) Bio/Technology 13:577-582.

Fenandez San Juan, P.M. (1995) Alimentaria 33:93-98

Fraley, R.T., Rogers, S.G., Horsch, R.B., Sanders, P.R., Flick, J.S., Adams, A.P., Bittner, M.L., Brand, L.A., Fink, C.L., Fry, J.S., Galluppi, G.R., Goldberg, S.B., Hoffman, N.L., Woo, S.C. (1983) Proc. Natl. Acad. Sci. (USA) 80:4803-4807.

Frankel, E.N. (1980) Prog. Lipid Res. 19, 1-22

Gardner, C.D., Kraemer, H.C. (1995) Arterio. Thromb. Vasc. Biol. 15:1917-1927

Gillatt, P. (1994) Inform 5 : 981-986.

Han, Y., Parsons, C. M. Hymowitz T. (1991) Poultry Sci. 70:896-906.

Heppard, E.P., Kinney, A. J., Stecca, K.L., Miao, G-H (1996) Plant Physiol. 110:311-319.

Hoffman, G., Meijboom, P.W. (1969) JAOCS 46:620-622.

Holm, L, Pancho, J.V.; Herbarger, J.P., Plucknett, D.L. (1979) A Geographical Atlas of World Weeds. John Wiley and Sons, New York

Hymowitz, T. and Harlan, J.R. (1982) Introduction of soybean to North America by Samuel Bowen in 1765. Econ. Bot. 37:371-379

Jefferson, R.A., Burgess, S.M. Hirsh D. (1986) Proc. Natl. Acad. Sci. (USA) 83:8447-8451.

Jofuku, D., Goldberg, R.B. (1989) The Plant Cell 1:1079-1093.

Kasahara, Y. (1982) Japan. In: Geobotany 2; Biology and Ecology of Weeds. (Holzner, W., Numata, M. eds.) pp 285-297. W. Junk, Netherlands.

Katan, M.B., Mensink, R.P., Zock, P.L (1995a) Ann. Rev. Nutr. 15:473-496

Katan, M.B., Zock, P.L., Mensink, R.P (1995b) Am. J. Clin. Nutr. 6:1368-1373

Keppler, J.G., Schols, J.A., Feenstra, W.H., Meijboom, P.W. (1965) JAOCS 42:246-249.

Kinney, A.J. (1994) Curr. Opin. Biotechnol. 5:144-151.

Kinsella, J.E. (1979) JAOCS 56:242-258.

Kitamura, K. (1995) JARQ 29:1-8.

Koritalia, S., Dutton, H.J., (1969) JAOCS 46:245-248

Ladin, B.F., Doyle, J.J., Beachy, R. N. (1984) J. Mol. Appl. Genet. 2:372-380

Lackey, J., APHIS Internet Server http://www.aphis.usda.gov/bbep/bp/soybean.html

Mallet, G., Dimitriades, C., Ucciani, E., Morin, O. (1985) Rev. Fr. Corps Gras 32:387-395.

McNamara, D. (1992) Adv. Food Nutr. Res. 36:254-351

Mensink, R.P., Temme, E.H.M., Hornstra, G. (1994) Annal. Med. 26:461-464

Mensink, R. P., Katan, M.B. (1990) New England Journal of Medicine 323 (7) 439-445

Messina, M., Barnes, S. (1991) J. Natl. Cancer Inst. 83:541-546.

Mounts, T.L., Wolf, W.J., Martinez, W.H. (1987) In: "Soybeans: Improvement, Production and Uses", pp 820-866.

Murawski, U., Egge, P., Gyorgy, P., Zilliken, F. (1971) FEBS Lett. 18:290-292.

Napoli, C., Lemieux, C., Jorgensen, R. (1990) Plant Cell 2:279-289

Nemoto, N. (1982) Weeds of Pastures and Meadows in Japan. In: Geobotany 2; Biology and ecology of weeds. (Holzner, W., Numata, M. eds.) pp. 395-401 W Junk, Netherlands

Odell, J.T., Nagy, F., Chua, N.-H. (1985) Nature 303:810-812.

Okuley, J., Lightner, J., Feldman, K., Yadav, N., Lark, E., Browse, J. (1994) Plant Cell 6:147-158.

Orthoefer, F.T. (1978) In: "Soybean: Physiology, Agronomy, and Utilization", pp219-246.

Popescu, O., Koritala, S., Dutton, H.J., (1969) JAOCS 46:97-99

Pryde, E.H. (1980) In: Handbook of Soy Oil Processing and Utilization, 13-31.

Raloff, J. (1996) Science News 150:87

Ratnayake, W.M.N., Pelletier, G. (1992) JAOCS 69:95-105.

Sebedio, J.L., Bonpunt, A., Prevost, J., Grandgirard, A. (1994) Fett Wissenschaft Technologie 96:235-239

Shibahara, A., Yamamoto, K., Shinkai, K., Nakayama, T., Kajimoto, G. (1993) Biochim. Biophys. Acta 1170:245-252.

Smith, K.J. (1981). JAOCS 135-139.

Smith, K.J., Huyser, W. (1987) in Soybeans: Improvement, Production, and Uses, (Wilcox, J.R. ed.) pp1-4. 2nd. Ed

Spady, D.K., Woollett, L.A., Dietschy, J.M. (1993) Ann. Rev. Nutr. 13:355-381

Stroink, J.B.A., Sparreboom, S. (1967) JAOCS 44:531-533.

Sutcliffe, J. (1979) Cold Spring Harbor Symposium. Quantitative Biology 43:77-90.

Takahashi, K., Banba, H., Kikuchi, A., Ito, M., Nakamura, S. (1994) Breeding Science 44:65-66.

USDA-APHIS Petition 93-258-01 for Determination of Nonregulated Status for Glyphosate-Tolerant Soybean Line 40-3-2; Environmental Assessment and Finding of No Significant Impact, May 1994

Utsumi, S., Kinsella, J.E. (1985) J. Agric. Food Chem. 33:297-303.

Wang G., S.S. Kuan, O.J. Francis, G.M.Ware, A.S. Carman (1990) J. Agric. Food Chem. 38:185-190.

Wang, H, Murphy, P.A. (1994) J. Agric. Food Chem. 42:1666-1673.

Werner, S.A., Luedecke, L.O., Shultz, T.D. (1992) J. Agric. Food Chem. 40:1817-1821.

Woerfel, J.B. (1995) Grasas y Acietas 46:357-365

Wollett, L.A., Dietschy, J.M (1994) Amer. J. Clin. Nutr. 60:991-996

Yadav, N.S., Wierzibicki, A., Aegerter, M., Caster, C.S., Perez-Grau, L., Kinney, A.J., Hitz, W.D. et. al. (1993) Plant Physiol. 103:467-476.

Yeh, P., Sicard, A.M., Sinskey A.J. (1988) Mol. Gen. Genet. 212:105-111.

XV. Appendices

A. Appendix 1: Allergenicity of High Oleic Acid Transgenic Soybeans

The protein profile of high oleic acid transgenic soybeans is somewhat different from that of A2396, the elite parent line. To assess any changes in the allergenic activity of the transgenic in comparison to elite soybeans, allergenicity testing was carried out by Dr. Samuel Lehrer at Tulane University School of Medicine, Section of Allergy and Clinical Immunology, New Orleans, La.

Elite soybeans (150g of A2396) and high oleic acid transgenic soybeans (11.95g of G94-1), were extracted by homogenization (1:20 w/v) in phosphate buffered saline (PBS) extraction buffer (100 µmol/L sodium phosphate, 0.9% NaCl, pH 7.2) for one hour at 4°C under constant stirring. The soybean extracts were centrifuged (80,000 xg), the aqueous layer removed, concentrated (Amicon YM-10 MWCO 1,000D) and recentrifuged (105,000 xg). The supernatant was aliquoted and stored at -20°C. Wild-type soybean yielded 18.785g (12.52%) of extract and transgenic soybean 1.165 g (9.75%) of extract.

The sera used were from 31 subjects who had a positive history of documented soybeans or food allergy, plus a positive skin test to soybean extract, and/or a positive IgE antibody response to soybean extract. Nine/31 sera (#6-14) were from a commercial laboratory (Plasma Lab., Seattle, Wa.); another 9/31 (#15-23) were obtained from workers with occupational soybean allergy; one sample was a serum pool of soybean allergic individuals, a gift from Steve Taylor and Julie Nordlee of the University of Nebraska, and the remaining 12 sera were obtained locally from soybean allergic subjects. Control sera were obtained from soybean tolerant individuals with negative skin test and/or RAST to soy extract with total IgE levels similar to those sera of soybean-sensitive subjects. Individuals participating in this study gave informed consent and completed a questionnaire requesting information concerning their allergy history, reactivity to soy and skin test reactivity to common inhalant and food allergens.

All sera were assayed for soybean-specific IgE antibody by radioallergosorbent test (RAST). Wild type soy extracts were coupled to cyanogen bromide activated paper discs (Whatman #50) at 1.0 mg dry weight (Ceska & Lundkvist, 1972). One hundred µl of serum were added to duplicate RAST discs and incubated overnight on a rotator at room temperature. Discs were washed three times with 0.9% saline to remove unbound serum. I^{125} labelled equine anti-human IgE (5,000 cpm/disc; Sanofi Diagnostics Pasteur, Inc.) was added to each disc and incubated overnight on a rotator at room temperature. Discs were washed again three times with 0.9% saline and bound I^{125} counted in a Beckman 5500 gamma counter. The mean of duplicate counts were obtained and expressed as percent bound of the total I^{125} added. RAST values equal to or greater than 3% binding were considered positive.

The results of the RAST assay for the 31 sera described above are summarized in Figure 1. Many of the sera demonstrated significant IgE antibody reactivity to soybean extracts. Twenty-one of the 31 sera tested had IgE antibody percent binding greater than or equal to

4%. Eleven of the 21 positive sera had IgE antibody binding in excess of 20%. The sera with the most significant RAST reactivity (#5,6,7,8,10,12 and 31) were pooled for RAST inhibition studies.

Subject	Source	% Bound
1	U. of Neb. Pool	30.3
2	Baton Rouge	1.8
3	Baton Rouge	9.0
4	Lafayette	4.3
5	New Orleans	32.9
6	Commercial	38.3
7	Commercial	39.4
8	Commercial	34.0
9	Commercial	45.4
10	Commercial	43.1
11	Commercial	20.0
12	Commercial	30.6
13	Commercial	14.2
14	Commercial	12.1
15	Occupational Soybean Allergy	3.7
16	Occupational Soybean Allergy	3.6
17	Occupational Soybean Allergy	2.6
18	Occupational Soybean Allergy	6.7
19	Occupational Soybean Allergy	4.0
20	Occupational Soybean Allergy	2.5
21	Occupational Soybean Allergy	15.6
22	Occupational Soybean Allergy	8.1
23	Occupational Soybean Allergy	1.6
24	New Orleans	3.1
25	New Orleans	3.8
26	New Orleans	0.8
27	New Orleans	5.4
28	New Orleans	1.7
29	New Orleans	50.4
30	New Orleans	4.6
31	New Orleans	21.5

Figure 1: RAST reactivity to soybean extracts.

Prior to RAST inhibition studies, the sera was tested undiluted and diluted 1:2.5, 1:5, 1:10, 1:20, 1:40 and 1:80 against discs coated with elite soybean extract. The serum pool diluted 1:5 was the highest dilution that still gave significant reactivity and was selected for subsequent RAST inhibition analysis.

The RAST inhibition studies were performed using discs coated with the elite soybean extract testing increasing amounts (0.1, 1.0, 10, 100, 1000 µg) of either elite or transgenic soybean extracts as inhibitors. The results, summarized in Figure 2, indicated that both the elite and the transgenic soybean extracts yielded identical RAST inhibition of the elite soybean RAST. Both inhibition curves were analyzed by logit-log transformation and linear regression analysis. The two regression lines were compared for statistically significant differences of their slopes and y-axis intercepts. The correlation coefficients for both elite and transgenic soybean were high (r_{wt} = 0.9991, r_t = 0.9978) and the 50% inhibition concentrations were very similar ($I_{50,wt}$ = 16.7 µg/ml, $I_{50,t}$ = 11.7 µg/ml). Analysis showed that both slopes and y-axis interceps are statistically identical. Thus, from a quantitiative viewpoint, the allergen content of both soybean extracts is identical.

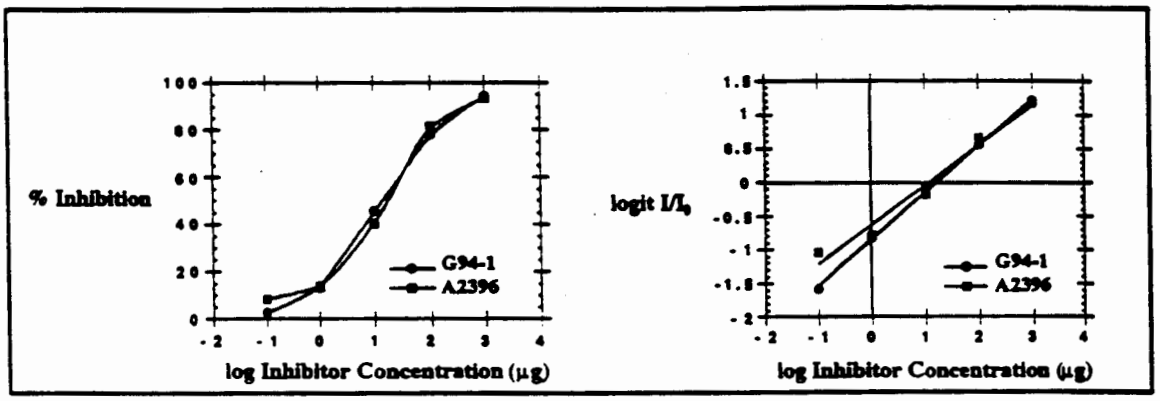


Figure 2: RAST inhibition plots of elite (A2396) and transgenic (G94-1) soybean extracts

The 21 most potent RAST positive sera were selected for immunoblot analyses of soybean allergens. For the detection of IgE antibody reactivities, blots were incubated overnight with patients' serum pool, diluted with TBS-Tween (100 mM Tris/HCl; 100 mM NaCl; 2.5 mM $MgCl_2$; pH 7.4, 0.05% v/v Tween 20). TBS-Tween was used for all incubation and washing steps throughout the experiments. For the detection of specific IgE antibodies, the strips were incubated with 1:1,000 diluted alkaline phosphatase-conjugated monoclonal anti-human-IgE (Southern Biotechnology Associates, Birmingham AL, USA), washed with freshly prepared assay buffer (100 mM diethanolamine/HCl, 1.0 mM $MgCl_2$, pH 10.0), incubated in 1:20 diluted Nitroblock® chemiluminescence enhancer (Tropix, Bedford, MA) for 5 min, washed with assay buffer and incubated in 250 mM CSPD (disodium 3-(4-methoxyspiro{dioxetane-3,2'-(5'chloro)tricyclo[3.3.1.1.^{3,7}]decan}-4-yl)phenyl phosphate; Tropix) for 5 min. Excessive liquid was drained, the strips sealed between transparencies and exposed to autoradiography film for 15, 30, 60 and 120 sec.

Additionally, the antibody binding was made visible by colorimetric detection. After washing with TBS-Tween and TBS-AP (100 mM Tris/HCl; 100 mM NaCl; 5 mM $MgCl_2$; pH 9.5), the strips were incubated in substrate/chromogen mixture for alkaline phosphatase at 37°C containing 450 µM 5-bromo-4-chloro-indolyl-phosphate disodium salt

(BCIP; Sigma, St. Louis, Mo.) and 400 µM nitroblue tetrazolium chloride (NBT; Sigma) solubilized in TBS-AP.

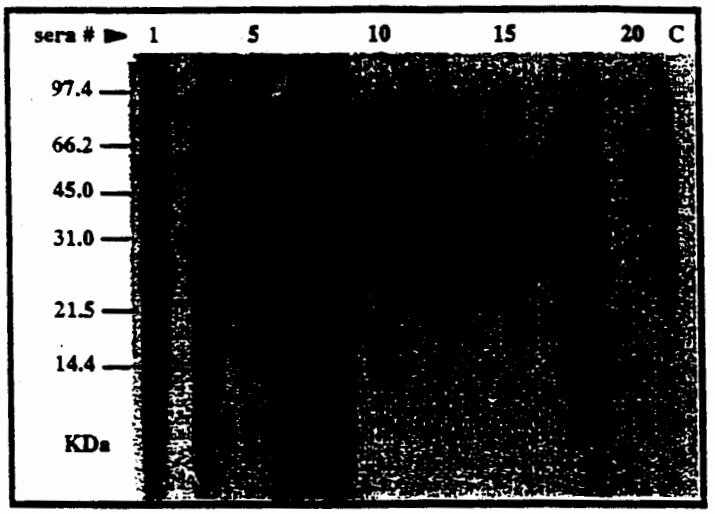


Figure 3: Immunoblot analysis of the 21 most potent RAST positive sera (Sera 1-21)

Results, shown in Figure 3, indicate that there are a number of protein bands that bind IgE antibodies from soybean allergic sera. Some sera appear to be more reactive with soybean extracts (sera #1, 3, 6, 8, 15, 18, and 21) whereas others show little or no reactivity. Based on these results the six most reactive sera (#1, 3, 6, 8, 18 and 21) were selected and pooled for further study of allergens present in elite and transgenic soybeans.

The sera pool were serially diluted at 1:2, 1:4, 1:8, and 1:16 and tested for IgE antibody reactivity to transgenic and elite soybean extracts. The results, summarized in Figure 4, show that no significant differences in the number of bands to which the sera react or the intensity of the IgEreactivity. The IgE antibody binding to the transgenic soybean extract was compared to the elite soybean extract. The inhibition patterns were remarkably similar with blots stained by colormetric (Figure 4a) or chemiluminescence/autoradiography (Figure 4b) methods.

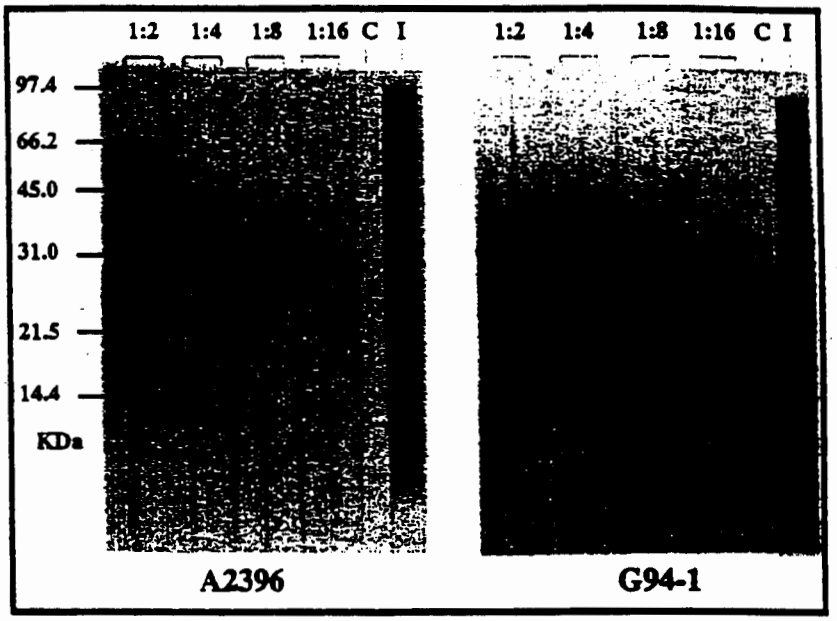


Figure 4a: Antibody reactivity to elite (A2396) and transgenic (G94) soybean extracts (colormetric detection method). C = negative control, I = positive stain (India ink).

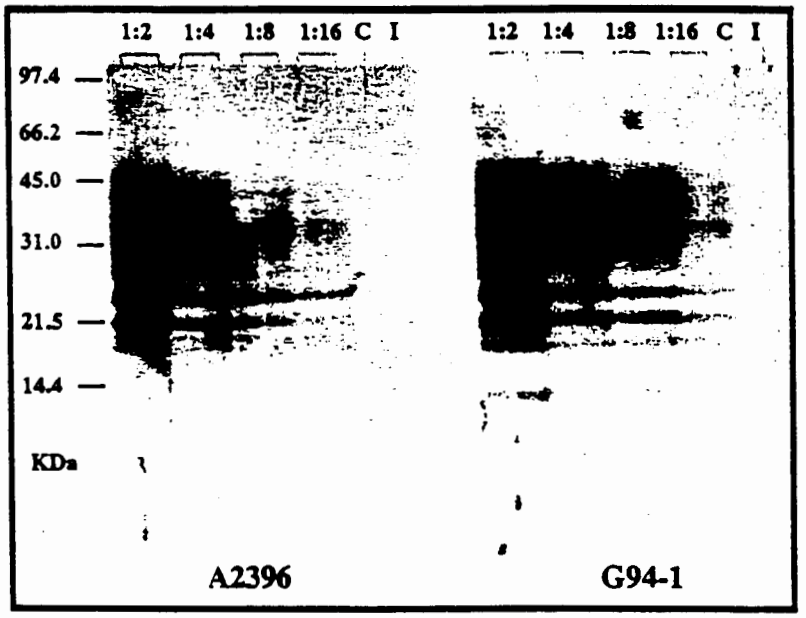


Figure 4b: Antibody reactivity to elite (A2396) and transgenic (G94) soybean extracts (chemiluminescence detection method). C = negative control, I = positive stain (India ink).

Control sera from soybean tolerant subjects did not react to either soybean extract (results not shown).

Taken together, these results demonstrate that there is no significant quantitative or qualitative difference between the transgenic and elite soybeans with regard to their allergen content.

B. Appendix 2: Field Test Reports

Field Test Report for High Oleic Acid Transgenic Soybeans

APHIS Notification Number: 95-088-08N

Reference Number: 95SRN23

Phenotype: improved oil quality

Three sublines (G94-1, G94-19, and G168-12) derived from transformation event 260-05 and each of which contained a single locus of the transgene in the homozygous condition, were tested at both the Maryland and Iowa sites. Additional isogenic lines containing a single locus of the transgene in the homozygous condition, derived from G94 and G168, were field tested at the Iowa site in Palestine Township (Slater), namely:

G94-1, G94-8, G94-11, G94-17, G94-19, G94-21, G94-33, G94-34, G168-17,
G168-18, G168-26, G168-32, G168-37, G168-38, G168-40, G168-42 , G168-44,
G168-50, G168-51, G168-53, G168-55, G168-59, G168-61, G168-65, G168-67,
G168-72, G168-76, G168-78, G168-79, G168-80, G168-81, G168-84, G168-87
G168-88, G168-90, G168-95, G168-98, G168-100, G168-101, G168-105,
G168-112, G168-115, G168-117, G168-119, G168-121, G168-122, G168-125
G168-127, G168-129, G168-131, G168-133, G168-137, G168-143, G168-144
G168-149, G168-150, G168-154, G168-155, G168-157

In addition, in a small experimental plot, three different generations of seeds (R1, R2 and R3) were planted side by side and all the seeds from each plant were collected for oleic acid analyses, to check the stability of the fatty acid phenotype over generations. Individual seeds planted were G253, G276, G296, G313, G328 (R1 seeds); G168-187, G168-171 (R2 seeds); and G168-59-4, G168-72-1 through G168-72-4 (R3 seeds).

Researcher's Name: Scott A. Sebastian

Field Trial Locations:

a. DuPont Stine Research Facilities
Maryland Side (Cecil County)

b. NW 1/4 of Section 31 of Palestine Township
Story County, IA (Slater)

Date Planted: 5/22/95 (Stine) and 5/25 (Slater)

Date of seedling emergence: ca. 5 days after planting for each site

Estimated percent germination: Some plots were started in the greenhouse and transplanted into the field while others were planted directly into the field (as noted below).

	% Germination Stine	% Germination Slater
G94-1	73% field	90% field
G94-19	75% greenhouse	80% field
G168-12	92% greenhouse	not planted

Agronomic observations:

The three transgenic sublines were indistinguishable from the original parent line A2396 in appearance, seed set, flowering, and maturity based on visual observations. G94-1 and G94-19 were ca. 2" shorter than A2396 at the Slater location but this did not appear to affect pod set or other agronomic properties. Plant-to-plant and plot-to-plot variation in these traits was not any greater than variation within the original parent line A2396. More highly replicated data will be collected during the summer of 1996 to see if more subtle differences exist.

No differences were noted between the three transgenic sublines and the parent line A2396 as regards levels of disease and insects. Although the entire Stine field was sprayed three times during June 1996 (twice with Asana and once with Sevin) to control thrips, no difference in insect susceptibility were observed between transgenic and original parent line.

Date trial terminated: 9/21/95(Stine) 10/17/95(Slater)

Method of destruction: Plants were harvested by hand by pulling them from the ground. Seeds were threshed and stored in controlled facilities.

Field monitoring for transgenics volunteers was done by Scott A. Sebastian on May 8, 1996 in fields H6 and H8 at the Stine Farm plots (Maryland side). No volunteer soybeans were observed.

Comments on the sublines grown at Slater, IA. After harvesting and checking the notes taken by the soybean breeder, several lines were removed from consideration for further advancement. The following lines were selected on the basis of superior agronomic characteristics making them visually identical to the variety A2396. These selections made up the majority of the lines grown in Puerto Rico in the winter of 1995/96. G94-1, G94-19, G168-44, G168-61, G168-79, G168-84, G168-95, G168-105, G168-115, G168-117, G168-119, G168-143, G168-150, G168-155, G168-157

Method of Destruction: Plants were harvested by hand by pulling them from the ground or breaking them off at the ground level. Seeds were threshed and stored in controlled facilities.

Dates of Monitoring for volunteer plants: Fields were checked for volunteers in June 1996. No volunteers were found.

Field Test Report for High Oleic Acid Transgenic Soybeans

Field Test Report for subline G168-12, derived from transformation event 260-05 which contains a single locus of the transgene in the homozygous condition, grown at the 1995 GXE locations (1) Schoolcraft, MI (Kalamazoo) (2) Ames, IA (3) Norway, IA (Cedar Rapids), (4) Ainsworth, IA (Iowa City) in 1995:

APHIS Notification Number: 95-088-08N, 95-107-08N

Reference Number: 95SRN23, 95SRN23A

Phenotype: High Oleic Subline G168-12

Researcher's Name: Allen R. LeRoy

Locations:

	Date Planted	Date Terminated
Schoolcraft, MI	5-31-95	10-5-95
Ames, Ia	5-31-95	10-15-95
Norway, IA	6-12-95	10-17-95
Ainsworth, IA	6-13-95	10-20-95

Date of Seedling Emergence: About 5 days after planting for each site.

Estimated Percent Germination: 90%

Agronomic Observations: The transgenic subline was not different from the original parent line A2396 in appearance, seed set, flowering, and maturity based on visual observations. In addition, there was no difference observed for insect or disease susceptibility.

Method of Destruction: Plants were harvested by hand by pulling them from the ground or breaking them off at ground level. Seeds were threshed and stored in controlled facilites.

Dates Of Monitoring For Volunteer Plants: Fields were checked for volunteers in June 1996. No volunteers were observed.

Field Test Report for High Oleic Acid Transgenic Soybeans

Field Test Report for sublines derived from transformation event 260-05 which contain a single locus of the transgene in the homozygous condition, grown at Asgrow Winter Nursery at Isabela, PR in winter 1995/1996.

APHIS Notification Number: 95-257-10N

Reference Number: 95SRN44

Phenotype: High Oleic Sublines derived from transformation event 260-05.

LINE	Area for 11-22-95 Planting (Acres)	Area for 2-15-96 Planting (Acres)
5GH-94-1	0.3	6.5
5GH-94-19	0.25	4.5
5GH-G168-12	0.02	0.1
5GH-G168-44	0.02	0.1
5GH-G168-61	0.02	0.1
5GH-G168-79	0.02	0.1
5GH-G168-84	0.02	0.1
5GH-G168-95	0.02	0.1
5GH-G168-105	0.02	0.1
5GH-G168-115	0.02	0.1
5GH-G168-117	0.02	0.1
5GH-G168-119	0.02	0.1
5GH-G168-143	0.02	0.1
5GH-G168-150	0.02	0.1
5GH-G168-155	0.02	0.1
5GH-G168-157	0.02	0.1

Researcher's Name: Allen R. LeRoy

Locations:

	Date Planted	Date Terminated
Asgrow Station Field G1	11-22-95	2-12-96
Asgrow Station Field EEA3	2-15-96	5-7-96

Date of Seedling Emergence: About 5 days after planting for each site.

Estimated Percent Germination: 90%

Agronomic Observations: The transgenic sublines were not different from the original parent line A2396 in appearance, seed set, flowering, and maturity based on visual observations. In addition, there was no difference observed for insect or disease susceptibility.

Method of Destruction: Plants were harvested by hand by pulling them from the ground or breaking them off at ground level. Seeds were threshed and stored in controlled facilites.

Dates Of Monitoring For Volunteer Plants: Volunteers were destroyed by cultivation at monthly intervals following harvest.

Field Test Report for High Oleic Acid Transgenic Soybeans

APHIS Notification Number 96-071-18n

Reference Number: 96SRN3

Phenotype: improved oil quality

Researcher's Name: Scott A. Sebastian

Field Trial Locations:

DuPont Stine Research Facilities
Maryland Side (Cecil County)
Elkton, MD 21921

Date Planted: 5/17/96

Date of seedling emergence: ca. 5 days after planting

Agronomic observations:

Control rows of the original transformant 260-05 (sublines G168-12, G94-1, and G94-19), three rows of the elite line A2396 (the line used as parent for transformation), and 568 F2 lines from crosses involving 260-05 were planted within the same field nursery. The F2 lines (derived from various crosses with elite and pre-commercial lines) were chosen as candidates for further development since they confirmed as containing an improved oil quality phenotype.

The three transgenic sublines were indistinguishable from the original parent line A2396 in appearance, seed set, flowering, and maturity based on visual observations. Plant-to-plant and plot-to-plot variation in these traits was not any greater than variation within the original parent line A2396. No differences were noted between the three transgenic sublines and the parent line A2396 as regards levels of disease and insects.

Because of the variation introduced by the various parents crossed with 260-05, the F2-derived lines were segregating for a variety of agronomic traits including height and maturity. No significant disease or insect pressure was noted in the 1996 nursery. Since the transformant checks were similar in agronomics to the "isogenic" elite line A2396, said segregation was not be attributed to the transgene imparting the improved oil quality. The lines with improved oil quality did not have any obvious"average" effects on height, disease resistance, or insect resistance from wild type check varieties within the same nursery.

Date trial terminated: All plots containing the gene for improved oil quality were harvested on October 21, 1996. Plants were harvested by hand by pulling them from the ground. Seeds were threshed and stored in controlled facilities.

Dates of monitoring for volunteer plants: Fields will be checked for volunteers in May of 1997.

Field Test Report for High Oleic Acid Transgenic Soybeans

Field test report for subline G168-12 soybeans grown at the 1996 GXE locations (1) Stine Farm, MD (2) Schoolcraft, MI (3) Ames, IA, (4) Norway, IA (Cedar Rapids), (5) Ainsworth, IA.

APHIS Notification Number: 96-071-18N

Reference Number: 96SRN3

Phenotype: High Oleic Subline G168-12 and lines derived from crosses between this line and transgenic and nontransgenic (mutant) sources of altered fatty acid germplasm.

Researcher's Name: Allen R. LeRoy

Locations:

	Date Planted	Date Terminated
DuPont Stine Research	5-17-96	10-21-96
Schoolcraft, MI	6-10-96	10-30-96
Ames, IA	6-4-96	11-8-96
Norway, IA	6-13-96	10-19-96
Ainsworth, IA	6-15-96	10-26-96

Date of Seedling Emergence: About 5 days after planting for each site.

Estimated Percent Germination: 90%

Agronomic Observations: The transgenic subline G168-12 was not different from the original parent line A2396 in appearance, seed set, flowering, and maturity based on visual observations. In addition,there was no difference observed for insect or disease susceptibility.

Method of Destruction: Plants were harvested by hand by pulling them from the ground or breaking them off at ground level. Seeds were threshed and stored in controlled facilites.

Dates Of Monitoring For Volunteer Plants: Fields will be checked for volunteers in June 1997.

Field Test Report for High Oleic Acid Transgenic Soybeans

Field Test Report for high oleic rDNA Soybeans Grown at Mid Western Yield Test Sites in 1996

APHIS Notification Number: 96-071-18N

Reference Number: 96SRN3

Phenotype: High Oleic Sublines derived from transformation event 260-05.

LINE	Area planted at each location
5GH-94-1	0.01 Acres
5GH-94-19	0.01 Acres
5GH-G168-12	0.01 Acres
5GH-G168-44	0.01 Acres
5GH-G168-61	0.01 Acres
5GH-G168-79	0.01 Acres
5GH-G168-84	0.01 Acres
5GH-G168-95	0.01 Acres
5GH-G168-105	0.01 Acres
5GH-G168-115	0.01 Acres
5GH-G168-117	0.01 Acres
5GH-G168-119	0.01 Acres
5GH-G168-143	0.01 Acres
5GH-G168-150	0.01 Acres
5GH-G168-155	0.01 Acres
5GH-G168-157	0.01 Acres

Researcher's Name: Allen R. LeRoy

Locations:

	Date Planted	Date Terminated
Asgrow Station, Ames, IA	6-5-96	10-12-96
Norway, IA	6-4-96	10-10-96
Prairie City, IA	6-12-96	10-5-96
Asgrow Station, Redwood Falls, MN	5-20-96	10-8-96
Amboy, MN	5-25-96	10-10-96
Asgrow Station, Janesville, WI	5-20-96	10-4-96
Dixon, IL	5-24-96	10-10-96
Asgrow Station, Oxford, IN	5-20-96	9-28-96
Wanatah, IN	5-28-96	10-10-96
Asgrow Station, Schoolcraft, MI	5-20-96	10-12-96

Date of Seedling Emergence: About 5 days after planting for each site.

Estimated Percent Germination: 90%

Agronomic Observations: The transgenic sublines were not different from the original parent line A2396 in appearance, seed set, flowering, and maturity based on visual observations. In addition, there was no difference observed for insect or disease susceptibility.

Method of Destruction: Plants were mechanically harvested by self propelled combine. Seeds were threshed and stored in controlled facilites.

Dates of Monitoring For Volunteer Plants: Fields will be checked for volunteers in June of 1997.

Yield Data - 1996 Field Tests

6IAVV023---COMBINED SITE MEANS, ALL TRAITS--NATURAL DATA--BU/A--SORTED
SITES: Prarie City, IA; Norway, IA; Dixon, IL; Oxford, IN; Schoolcraft, MI; Amboy, MN; Janesville, WI

STATISTIC			YIELD BU/A	MOIS	MAT	HGT	LDG	GR2
OVERALL MEAN			50.09	12.11	23.9	29.1	1.1	2.2
NUMBER OF LOCATIONS			7	7	7	7	6	7
STANDARD ERROR (FIXED MODEL)			.922	.165	.241	.314	.048	.090
ENT#	ENTRY CODE	PEDIGREE						
4	PION 9281	RkS	55.89	12.31	27.3	27.2	1.2	1.6
3	STIN 2250	S S A3127*11702	53.74	12.67	26.9	27.9	1.2	1.9
2	ASGR A2396	Ra CM214/2943	53.68	11.87	22.9	30.0	1.0	2.2
23	PUB BURLISON	Rb K7411376486*CENT	52.49	12.99	27.4	30.2	1.4	1.9
1	ASGR A1900	Rk 86PO58-14*A2234	51.09	11.93	21.4	28.2	1.2	1.6
8	5GH-G168-117	2396D2S	50.26	11.66	23.3	27.6	1.0	2.2
18	5GH-G168-79	2396D2S	50.15	12.06	23.4	28.5	1.0	2.2
10	5GH-G168-44	2396D2S	49.98	12.06	23.6	28.0	1.0	2.3
11	5GH-G168-95	2396D2S	49.92	12.24	23.5	28.8	1.0	2.2
17	5GH-G168-119	2396D2S	49.90	11.96	23.6	28.0	1.0	2.0
20	5GH-G168-155	2396D2S	49.79	11.53	23.2	27.9	1.0	2.4
24	VINTON 81	L60-347-4-4G-2-B*VIN	49.69	12.44	22.8	36.0	1.9	2.8
5	5GH-G168-12	2396D2S	49.60	12.91	25.2	28.6	1.0	2.5
25	HP204	VINTON81/2/HARDIN	49.41	12.37	24.0	35.0	1.9	2.8
12	5GH-G168-150	2396D2S	49.26	11.84	23.3	28.1	1.0	2.1
6	5GH-G94-19	2396D2S	48.88	11.56	23.6	27.8	1.0	2.3
9	5GH-G168-143	2396D2S	48.52	11.73	23.2	27.8	1.0	2.4
14	5GH-G168-61	2396D2S	48.12	12.03	23.4	28.3	1.0	2.2
16	5GH-G168-115	2396D2S	47.99	11.89	23.5	28.0	1.0	2.4
15	5GH-G168-84	2396D2S	47.84	11.74	23.8	28.3	1.1	2.4
13	5GH-G168-157	2396D2S	46.88	12.01	23.9	27.7	1.0	2.3
19	5GH-G168-105	2396D2S	46.80	11.97	23.6	27.5	1.0	2.4
7	5GH-G94-1	2396D2S	45.57	11.66	23.0	27.4	1.0	2.4

mois - moisture; mat - maturity, days after 9/31; hgt - height; ldg - lodging (1-5); GR2 - general rating (1-5) (1 = best); 2396D2S = high oleic acid transgenic soybean

Field Test Report for High Oleic Acid Transgenic Soybeans

1. Reference Number: 96-115-02N

2. Applicant Reference Number: 96SRN4

Indiana Tests

Variety G94-19
26 Acres
6508 West 500 North
West Lafayette, IN 47906-9228
Tippecanoe county, Shelby Township, section #30

Variety G94-19
39 Acres
6722 West 500 North
West Lafayette, IN 47906-9228
Tippecanoe County, Shelby Township, section #25

Variety G94-1
73 Acres
6722 West 500 North
West Lafayette, IN 47906-9228
Tippecanoe County, Shelby Township, Section #26

After Harvest Production Estimates for Indiana Field Tests:

Variety	Bushels	Acres	Yield /Acre
G94-1	2400	75.2	31.9
G94-19	1827	83.8	21.8

Iowa Tests

Variety G94-19
40 Acres Approx. 1580 bushels or 39.5bu/ac
Greene County
South Washington Township
Section #11

Variety G94-19
20 Acres Approx. 820 bushels or 41.0bu/ac
Greene County
South Washington Township
Section #13

Variety G94-1
52.1 Acres Approx. 2495 bushels or 47.9bu/ac
Dallas County
Spring Valley Township
Section #7

Variety G94-1
51.2 Acres Approx. 2500 bushels or 48.8bu/ac
Dallas County
Spring Valley Township
Section #7

After Harvest Production Estimates for Iowa Field Tests:

Variety	Bushels	Acres	Yield /Acre
G94-1	4995	103.3	48.35
G94-19	2400	60	40

Summary

The two varieties G94-1 and G94-19 did not appear to be unusual in comparison to normal soybeans of a similar parental background when planted late and under stress conditions. No unusual occurrences of insect attack or disease pressure were noted. Fields of G94-1 and G94-19 were planted in Indiana and Iowa in early/mid June of 1996, i.e. later than the preferred planting time. Plantings were delayed because of the very wet conditions. Poor emergence scores were noted because of the cool damp soils; the poor emergence rating does not differ from non-transgenic soybeans planted in the area. Yields at the Indiana sites were lower than expected due to stress, cool, damp conditions and late planting. In contrast, yields at the Iowa sites were much higher, 48.35 bushels/acre for G94-1, and 40 bushels/acre for G94-19.

Details

Indiana Sites

Two days after planting additional rains fell causing puddles in the fields. The G94-1 soybeans planted on June 6th had relatively good emergence scores with a final stand count of 5 plants per foot. The planting of G94-1 on June 8th had a very poor emergence rating due to the rain and puddling of the soil. The initial weed control practices for these two plantings failed. A follow-up program to control the weeds set the plants back and may have reduced the potential yield. The June 6th planting estimated average is 33 bushels per acre, while the June 8th planting dropped to 20 bushels per acre. The G94-19 planted on June 14th ended up with an established stand of 6 plants per foot. Pod height was very low due to the late planting and poor conditions. Two to five pods were so low to the ground they were not harvestable. The total estimated bushels of G94-1 is 2,400 bushels. The G94-19 total estimated bushels is 1,827. The after harvest report indicates that nearly 10 acres of production had less the 4 plants per foot due to flooding. There were a few off-types in the final stand.

The off-types were comprised of pubescence color and late maturity.

Iowa Sites

The G94-1 fields were planted on June 10, 1996 at a rate of 5 seeds per foot to stretch the seedstock and enhance the total yield. The fields were checked for off-types late in the growing season. The flower color off-type was at a rate of 2/4000 plants (0.05%), tall, later maturing off-types were found at a rate of approximately 6/10,000 or 0,06%. The off-types were rogued to a level of 1/4000 or less plants; some were kept for future analysis. The off-types tended to have a tawny pubescence. The normal plants have gray pubescence, tan pod walls and imperfect black hilums. Fatty acid analyses of some of the tall off-type plant seeds indicated that they were a mixture of plants, some with normal fatty acid profiles, some with high oleic profiles and the rest displayed a range in fatty acid profile between the two extremes. The G94-19 fields did not show any off-type plants for flower color or pubescence color. The grain samples have beans with imperfect-black hilums. The 0.05 - 0.06% range of off-types is not unusual and could be caused by a very small amount of outcrossing.

The two varieties G94-1 and G94-19 did not appear to be unusual in comparison to normal soybeans of a similar parental background growing under similar conditions. There was very little insect or disease pressure noted in these tests and both of the transformed lines, G94-1 and G94-19, seemed to perform similarly to the parental line A2396 with respect to insects and diseases. Yields of G94-1 were 48.35 bushels per acre while yields of G94-19 were 40 bushels per acre.

97-008-01p

VOLUME 2 - REFERENCES

Petition for Determination of Nonregulated Status:

High Oleic Acid Transgenic Soybean Sublines G94-1, G-94-19 and G-168 derived from Transformation Event 260-05

97-008-01p

VOLUME 2 - REFERENCES

Petition for Determination of Nonregulated Status:

High Oleic Acid Transgenic Soybean Sublines G94-1, G-94-19 and G-168 derived from Transformation Event 260-05

Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 458–462, January 1988
Botany

Cellular localization of soybean storage protein mRNA in transformed tobacco seeds

(transformation/*in situ* hybridization/gene regulation)

SUSAN J. BARKER, JOHN J. HARADA*, AND ROBERT B. GOLDBERG†

Department of Biology, University of California, Los Angeles, CA 90024

Communicated by Charles S. Levings III, September 28, 1987

ABSTRACT **We transformed tobacco plants with a soybean β-conglycinin gene that encodes the 1.7-kilobase β-subunit mRNA. We showed that the β-conglycinin mRNA accumulates and decays during tobacco seed development and that β-conglycinin mRNA is undetectable in the tobacco leaf. We utilized *in situ* hybridization to localize β-conglycinin mRNA within the tobacco seed. β-Conglycinin mRNA is not detectable within the endosperm but is localized within specific embryonic cell types. The highest concentration of β-conglycinin mRNA is found in cotyledon storage parenchyma cells. We conclude that sequences required for embryo expression, temporal control, and cell specificity are linked to the β-conglycinin gene, and that factors regulating β-conglycinin gene expression are compartmentalized within analogous soybean and tobacco seed regions.**

β-Conglycinin is a storage protein that accumulates to high levels in the soybean seed (1, 2). The major β-conglycinin polypeptides, designated α′, α, and β (3, 4), are encoded by 2.5-kilobase (kb) (α′/α) and 1.7-kb (β) mRNAs, respectively (5, 6). β-Conglycinin mRNAs are highly homologous, differ primarily by the presence or absence of a 0.56-kb insertion sequence, and are transcribed from different genes (refs. 7 and 8; also unpublished data). The β-conglycinin gene family contains at least 15 members that are clustered in small domains on at least two different chromosomes (ref. 9; J.J.H. and R.B.G., unpublished results). Most β-conglycinin genes either encode or are preferentially homologous with the 1.7-kb β-subunit mRNA (J.J.H. and R.B.G., unpublished results).

β-Conglycinin gene expression is highly regulated during the soybean life cycle. β-Conglycinin mRNA accumulates at precise embryonic stages, is less prevalent in the axis than the cotyledons, and is undetectable in mature plant organ systems (5, 8, 10, 11). In addition, the 2.5-kb α′/α mRNA accumulates prior to the 1.7-kb β mRNA during embryogenesis (refs. 8 and 10; J.J.H. and R.B.G., unpublished results). Transcriptional control processes are responsible for restricting β-conglycinin gene expression to the embryonic phase of the life cycle (5, 11). By contrast, posttranscriptional events regulate β-conglycinin mRNA differential accumulation (unpublished data). Neither selective gene amplification nor detectable differences in methylation patterns accompany the developmental changes in β-conglycinin gene expression (5, 11).

The DNA sequences and cellular factors responsible for controlling β-conglycinin gene expression have not yet been identified. To begin to identify cis control regions, Beachy *et al.* (12) and Chen *et al.* (13) demonstrated that an α′-subunit gene is expressed during petunia seed development and that a 0.09-kb region located within 0.26 kb of the 5′ gene end is required for high seed expression levels. It has not yet been established, however, whether β-conglycinin mRNA or any other seed protein mRNA is spatially distributed to the correct cell types in transformed seeds. Nor has it been established whether or not β-conglycinin mRNA accumulates and decays during seed development in transformed plants as occurs in soybean.

Here, we introduced a β-conglycinin β-subunit gene into tobacco plants. Using *in situ* hybridization, we demonstrate that β-conglycinin mRNA is detectable only within the tobacco embryo and is localized preferentially within cotyledon and upper axis cells. In addition, we show that β-conglycinin mRNA is not detectable in the tobacco leaf, and that β-conglycinin mRNA appears and disappears at precise stages of tobacco seed development. We conclude that sequences required for both temporal control and cell specificity are linked to the β-subunit gene and that regulatory molecules that control β-conglycinin gene expression are spatially distributed within analogous tobacco and soybean seed regions.

MATERIALS AND METHODS

Isolation of a β-Conglycinin Plasmid. The characteristics of the β-conglycinin cDNA plasmid A-16 have been described (5).

DNA Isolation and Labeling. DNAs were isolated as described by Fischer and Goldberg (14) and were labeled by nick-translation (15).

Polysomal mRNA Isolation. Soybean and tobacco polysomal mRNAs were isolated as described (16, 17).

Filter Hybridization Experiments. DNA gel-blot and RNA dot-blot experiments were carried out according to published procedures (18–20). RNA gel-blot studies were performed as described (5) except that nitrocellulose filters were used.

***In Situ* Hybridization Experiments.** *In situ* hybridization studies were carried out as described by Cox *et al.* (21) with modification. In brief, seeds 19 days after pollination (DAP) were fixed in 10% formalin/5% acetic acid/50% ethanol (22), embedded in paraffin, and sliced into 10-μm sections. Single-stranded β-conglycinin ^{35}S-labeled RNA probes were synthesized by using the pGEM transcription system (Promega Biotec). The ^{35}S-labeled RNA probes were hydrolyzed to approximately 0.2 kb in size and then were hybridized with slides containing fixed seed sections for 14 hr at selection criteria of 42°C, 0.3 M Na^+, and 50% formamide (21). A ^{35}S-labeled anti-mRNA probe was used to localize β-conglycinin mRNA within the tobacco seed while a ^{35}S-labeled mRNA probe monitored nonspecific hybridization within the fixed seed section. Following hybridization, the fixed seed sections were incubated with RNase A at 50

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: DAP, days after pollination.

*Present address: Department of Botany, University of California, Davis, CA 95616.
†To whom reprint requests should be addressed.

μg/ml, washed at criteria of 57°C and 0.02 M Na^+ and overlaid with film emulsion (21). Seed sections were stained with 0.5% toluidine blue, and both bright-field and dark-field microscopy were used to visualize cell types and autoradiographic grains.

R-Loop Formation. R loops were formed as described by Fischer and Goldberg (14).

Nuclease S1 Mapping. Nuclease S1 protection experiments were carried out by published procedures (23, 24).

Transformation of Tobacco Plants. Tobacco (*Nicotiana tabacum* cv. Wisconsin 38) leaf disks (25) were transformed with the β-conglycinin gene as described by Okamuro *et al.* (17).

RESULTS

Characterization of a β-Conglycinin Gene Region. We utilized the β-conglycinin cDNA clone A-16 (5) to select β-conglycinin genomic clones from a soybean genomic DNA library. A-16 represents the 2.5-kb α/α'-subunit mRNA but hybridizes with the 1.7-kb β-subunit mRNA as well (5). Fig. 1 shows the β-conglycinin genomic clone, designated λA16-B3, that contains the gene region that was utilized in the transformation studies. Two β-conglycinin genes, designated *CG-4* and *CG-8*, and a nonseed protein gene (NSP) are localized within this 17.3-kb genomic segment (Fig. 1*A*). Both β-conglycinin gene family members hybridized with the 2.5-kb and 1.7-kb β-conglycinin mRNAs but were preferentially homologous with the 1.7-kb β-subunit mRNA (J.J.H. and R.B.G., unpublished results). By contrast, the nonseed protein gene was transcriptionally active during embryogenesis and was represented in leaf and root mRNAs (ref. 11; J. K. Okamuro and R.B.G., unpublished results).

We characterized the *CG-4* β-conglycinin gene by R-loop analysis to visualize its structure in the electron microscope. Fig. 1*B* shows that the *CG-4* β-conglycinin gene contains at least six exons and five introns. To verify this structure directly, we sequenced the *CG-4* β-conglycinin gene and its contiguous regions. A schematic representation of the *CG-4* gene structure is presented in Fig. 1*C*. We determined that the DNA sequence produces an open reading frame with N-terminal amino acids similar to those present in the β-conglycinin β-subunit (4), that the *CG-4* gene contains five introns as predicted by the R-loop analysis, and that it differs from the *CG-1* α'-subunit gene (7) by the absence of a 0.56-kb insertion within the first exon.

We hybridized soybean midmaturation-stage embryo mRNA with a *CG-4* β-conglycinin gene 5' probe (Fig. 1*C*, probe I) to determine whether this gene was expressed during embryogenesis. Several fragments differing by one nucleotide in length were resistant to nuclease S1 digestion (Fig. 2, lane Soy +), indicating that the *CG-4* gene was represented in embryo mRNA. The fragment producing the strongest signal (0.24 kb) suggests that the transcription start site is 0.03 kb 3' to a TATA consensus sequence (Fig. 1*C*). Together, these data show that *CG-4* is a functional β-conglycinin gene and that it is clustered within a genomic domain containing a related β-conglycinin gene and a differentially regulated nonseed protein gene.

Representation of the β-Conglycinin Gene Region in Transformed Tobacco Plants. We introduced a 12.3-kb DNA fragment containing the *CG-4* β-conglycinin gene (Fig. 1*A*) into tobacco plants by transforming leaf disk cells with a Ti-plasmid vector (17, 25, 27–29). Two independent transformants, designated TOB-T1 and TOB-T2, were selected for analysis. We hybridized a *CG-4* β-conglycinin probe (Fig. 1*C*, probe II) with TOB-T1 and TOB-T2 DNA gel blots to determine the β-conglycinin gene copy number within the transformed plants. As shown in Fig. 3*A* and schematically presented in Fig. 3*B*, no detectable signals were obtained

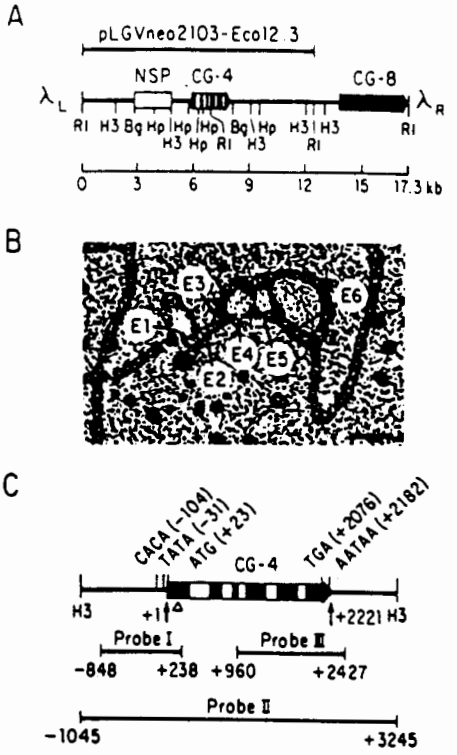


FIG. 1. The soybean *CG-4* β-conglycinin gene. (*A*) Organization of the *CG-4* β-conglycinin gene region. RI, H3, Hp, and Bg refer to *Eco*RI, *Hin*dIII, *Hpa* I, and *Bgl* II restriction endonucleases, respectively. Boxes/arrows represent gene locations/transcriptional orientations, respectively. *CG-4* exons and introns are displayed in black and white, respectively. NSP box refers to a nonseed protein gene and represents the maximum gene length. NSP and *CG-8* gene structures have not yet been characterized. pLGVneo2103-Eco12.3 is the recombinant plasmid used for the gene transfer studies. This plasmid contains the bracketed 12.3-kb *Eco*RI fragment that was obtained from a partial *Eco*RI digestion of λA16-3 phage DNA. (*B*) R-loop analysis of the *CG-4* β-conglycinin gene. λA16-3 DNA was hybridized with excess midmaturation-stage soybean embryo mRNA under conditions that form R loops (14). E1 through E6 designate exons oriented relative to the 5' and 3' gene ends. The bar represents approximately 0.6 kb. (*C*) A schematic representation of the *CG-4* β-conglycinin gene. Nucleotides +1 and +2221 designate the 5' and 3' ends, respectively. These were obtained from nuclease S1 protection studies (Fig. 2). CACA represents a consensus sequence (5' CAACACACG 3') that is present at least once in the 5' flanking region of all soybean seed protein genes sequenced to date (26). Black and white areas represent exons and introns, respectively. The bracketed lines below the β-conglycinin gene represent DNA probes used in the experiments reported here. Δ refers to the 0.56-kb insertion within the *CG-1* α'-subunit β-conglycinin gene (7) that is absent from *CG-4* at nucleotide +91.

with untransformed tobacco DNA (Fig. 3*A*, lane TOB-C). By contrast, low-copy soybean DNA fragments of the expected size were present in each transformant (Fig. 3*B*, lanes TOB-T1 and TOB-T2). We determined that each tobacco plant contained a single-copy β-conglycinin gene region by using a restriction endonuclease that does not digest within vector DNA, enabling two distinct-sized border fragments to be visualized (data not shown). Together, these findings demonstrate that each tobacco transformant contains one unrearranged 12.3-kb β-conglycinin gene region.

β-Conglycinin Gene Is Expressed Correctly in Transformed Tobacco Plants. We hybridized 14- to 16-DAP tobacco seed mRNA with a *CG-4* 5' probe (probe I, Fig. 1*C*) to determine whether the β-conglycinin gene was expressed in tobacco. The nuclease S1 assay (Fig. 2, lanes TOB-T1 and -T2) indicates that both TOB-T1 and TOB-T2 seed mRNAs

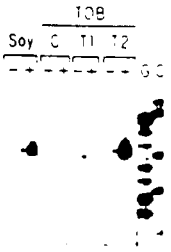


FIG. 2. Nuclease S1 analysis of β-conglycinin gene expression. Probe I (Fig. 1*C*) was hybridized with either soybean midmaturation-stage embryo mRNA or tobacco 14- to 16-DAP seed mRNA under conditions that favor DNA·RNA hybrid formation (24). Lanes: Soy, TOB C, TOB T1, and TOB T2, soybean, untransformed tobacco, transformant-1, and transformant-2 mRNAs, respectively (film exposures of lanes TOB-C, TOB-T1, and TOB-T2 differ from that of the Soy lanes); G and C, sequencing ladders from the same *CG-4* region; − and +, hybridization reactions that were not treated (−) or treated (+) with nuclease S1, respectively. The dots designate the major protected fragment in each experiment. This fragment was taken to be the 5′ terminal nucleotide of the mRNA (Fig. 1*C*).

protected DNA fragments similar in size to those produced with soybean mRNA (Fig. 2, lane Soy$^+$). This result shows that the β-conglycinin gene is expressed in each transformed plant and that the transcription start site is similar to that observed in soybean plants.

We hybridized a *CG-4* probe (Fig. 1*C*, probe II) with gel blots containing tobacco seed and leaf mRNAs to determine whether the β-conglycinin gene was regulated correctly

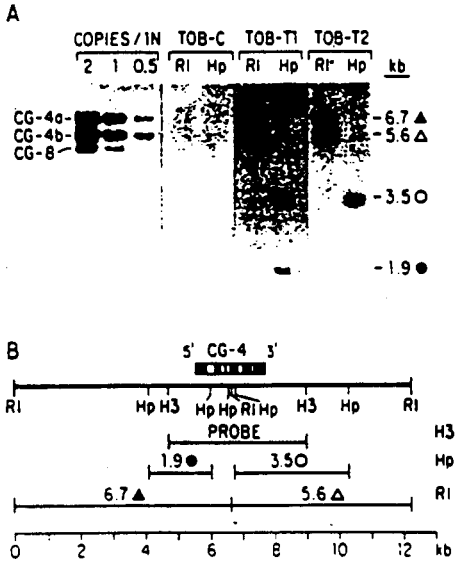


FIG. 3. Representation of the β-conglycinin gene region in transformed tobacco plants. (*A*) DNA gel blot of untransformed and transformed tobacco DNAs. Leaf nuclear DNA from individual plants was digested with restriction endonuclease, fractionated by electrophoresis, transferred to nitrocellulose, and hybridized with *CG-4* gene probe II (Fig. 1*C*). Lanes: Copies/1N, reconstruction lanes containing 0.5-, 1-, and 2-copy equivalents of *Eco*RI-digested λA16-3 DNA [DNA copy equivalents were calculated using a tobacco genome size of 2.4×10^6 kb (30)]; TOB-C, untransformed tobacco; TOB-T1, transformant 1 mRNA; TOB-T2, transformant 2 mRNA. CG-4a and CG-4b refer to the 6.7-kb and 5.6-kb *CG-4 Eco*RI fragments, while *CG-8* refers to the 5-kb *CG-8 Eco*RI fragment (Fig. 1*A*). The reduced *CG-8* hybridization signal indicates that this gene has diverged from the *CG-4* β-conglycinin gene. ▲ △, *Eco*RI DNA fragments; ● ○, *Hpa* I DNA fragments. (*B*) Map of relevant restriction endonuclease sites within the *CG-4* β-conglycinin gene region. RI, H3, and Hp refer to *Eco*RI, *Hin*dIII, and *Hpa* I restriction endonucleases, respectively. CG-4 exons and introns are represented by black and white boxes, respectively.

during the tobacco life cycle. Fig. 4 shows that the *CG-4* probe produced a 1.7-kb signal with both TOB-T1 and TOB-T2 seed mRNAs. Densitometric analysis indicated that these signals were 1/20th (TOB-T2) to 1/100th (TOB-T1) of that obtained with soybean mid-maturation stage embryo mRNA. We presume that this prevalence difference is due to position effect because TOB-T1 and TOB-T2 plants contain a single-copy β-conglycinin gene insert at distinct genomic locations (Fig. 3). By contrast, no detectable signals were obtained with untransformed tobacco seed and leaf mRNAs (Fig. 4, lanes TOB-C), or with transformed tobacco leaf mRNA (Fig. 4, lanes L of TOB-T1 and TOB-T2). Together, these results show that the 1.7-kb soybean β-conglycinin mRNA is produced in the correct developmental state within each transformed tobacco plant.

β-Conglycinin Gene Expression Is Regulated Temporally During Seed Development. We isolated transformed tobacco seed mRNA at different times after pollination and hybridized each mRNA with a β-conglycinin probe (Fig. 1*C*, probe II) to determine whether the β-conglycinin gene was regulated temporally during seed development. We constructed a timetable for tobacco seed development to specify stages from which mRNAs were isolated (data not shown). Fig. 5*A* shows that β-conglycinin mRNA accumulates and decays during tobacco seed development. Transformed tobacco leaf and untransformed tobacco seed mRNAs did not produce hybridization signals (Fig. 5*B*). At its peak prevalence (23 DAP), β-conglycinin mRNA represented approximately 0.5% of the TOB-T2 tobacco seed mRNA mass or approximately 1/20th of that observed during soybean embryogenesis (Fig. 5*C*). We conclude that the soybean β-conglycinin gene displays a precise temporal expression pattern during tobacco seed development and that this pattern is analogous to that which occurs in soybean (5, 11).

Soybean β-Conglycinin mRNA Is Concentrated Within Specific Tobacco Seed Regions. We utilized *in situ* hybridization (21) to localize β-conglycinin mRNA within transformed tobacco seed sections (22). Fig. 6*A* shows a bright-field micrograph of a 19-DAP tobacco seed longitudinal section that was stained with toluidine blue. A prominent embryo (E) with well-developed cotyledons (C) and a hypocotyl-radicle axis (A) is embedded within several layers of nonembryonic

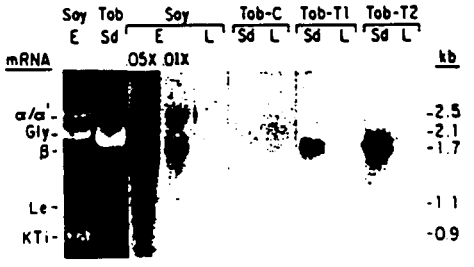


FIG. 4. β-conglycinin gene expression in transformed tobacco plants. Soybean and tobacco mRNAs were fractionated by electrophoresis on a methylmercury(II) hydroxide-agarose gel, stained with ethidium bromide to visualize mRNA bands, and transferred to nitrocellulose paper. The RNA gel blots were hybridized with *CG-4* gene probe II (Fig. 1*C*). Lanes: Soy, TOB-C, TOB-T1, and TOB-T2, soybean, untransformed tobacco, and transformed tobacco mRNAs 1 and 2, respectively; E, Sd, and L, embryo, seed, and leaf mRNAs, respectively, at 5 μg of mRNA; soybean embryo 0.05× and 0.01×, 0.25 μg and 0.05 μg of midmaturation-stage mRNA representing ≈1000 and 250 molecules per cell equivalents of β-conglycinin mRNA (5). α/α′, Gly, β, Le, and KTi refer to mRNAs encoding the β-conglycinin α/α′-subunit, glycinin, β-conglycinin β subunit, lectin, and Kunitz trypsin inhibitor seed proteins, respectively. Seed protein mRNA sizes have been described previously (5, 14). The prevalent 1.9-kb tobacco seed mRNA (lane Tob Sd) hybridized weakly with a soybean β-conglycinin gene probe (data not shown), suggesting that it encodes an endogenous tobacco storage protein

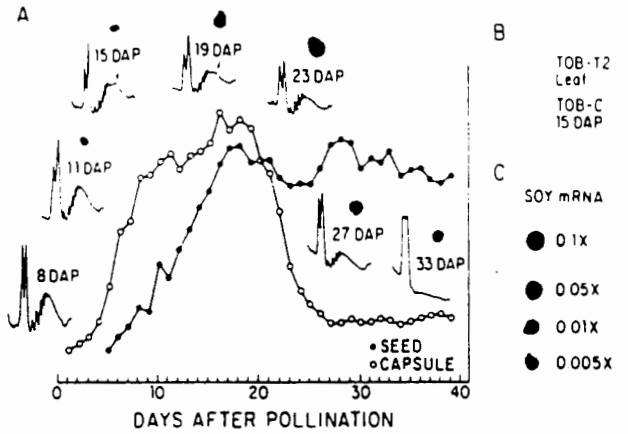


FIG. 5. β-conglycinin gene expression during tobacco seed development. (*A*) Representation of β-conglycinin transcripts in tobacco seed mRNA. One microgram of TOB-T2 mRNA from different developmental stages was spotted onto nitrocellulose and hybridized with a *CG-4* gene probe (Fig. 1C, probe II). ●, Seed dry weight in milligrams per 100 seeds; ○, fruit capsule fresh weight in grams. Fruit capsule fresh weight and seed dry weight at 19 DAP were 1.4 g and 7.7 mg, respectively. Polysome profiles for seeds at different DAP are shown below the dot-blot signals. The arrow indicates a 15-mer polysome that probably contains the 1.9-kb tobacco storage protein mRNA visualized in Fig. 4 (lane Tob Sd). (*B*) Hybridization of the β-conglycinin gene probe with transformed tobacco leaf mRNA (TOB-T2) and untransformed tobacco seed mRNA (TOB-C). Each spot contained one microgram of mRNA. (*C*) Soybean mRNA standards used to calibrate β-conglycinin mRNA prevalences. The 0.1×, 0.05×, 0.01×, and 0.005× spots contained 0.1 μg, 0.05 μg, 0.01 μg, and 0.005 μg of soybean midmaturation-stage mRNA and correspond to β-conglycinin mRNA prevalences of 1%, 0.5%, 0.1%, and 0.05%, respectively (5).

endosperm cells (En). In addition, a provascular cylinder (V) is apparent within both the cotyledons and the axis.

The dark-field micrograph (Fig. 6*B*) indicates that β-conglycinin ^{35}S-labeled anti-mRNA (Fig. 1*C*, probe III) hybridized intensely with RNA present in embryo cotyledon cells and upper axis cells. By comparison with the ^{35}S-labeled mRNA control (Fig. 6*C*), no grains above background were observed in the endosperm or in the lower axis embryonic region. Shorter exposures of a seed section analogous to that presented in Fig. 6*B* (data not shown) and hybridization with sections sliced in different planes (L. Perez-Grau and R.B.G., unpublished results) indicated that β-conglycinin mRNA was present in embryo storage parenchyma cells but undetectable in cells of the provascular cylinder. These findings indicate that the β-conglycinin mRNA concentration differs significantly within the seed and within different embryonic regions and cell types. Assuming that endosperm and embryo cells have similar mRNA contents, we estimate that the actual β-conglycinin mRNA prevalence in the embryo is at least 2% of the mRNA mass or only one-fifth that measured in soybean for the family as a whole (5). We conclude that soybean β-conglycinin mRNA is spatially regulated within the tobacco seed and that the β-conglycinin gene has an embryo-specific gene expression program similar to that observed in soybean.

DISCUSSION

Soybean *CG-4* β-Conglycinin Gene Region. We utilized a 12.3-kb genomic segment containing a β-subunit β-conglycinin gene for the gene transfer studies reported here (Fig. 1). Closely linked to the *CG-4* gene is a β-subunit gene relative (*CG-8*) and an unidentified gene that is expressed during embryogenesis and in the mature plant. DNA walking experiments showed that within 45 kb no other β-conglycinin genes are contiguous to *CG-4* and *CG-8* (unpublished data). By contrast, other regions exist that contain up to five interspersed α′/α-subunit and β-subunit β-conglycinin genes (unpublished data). Close linkage of differentially expressed genes also occurs in the glycinin, lectin, and leghemoglobin gene regions (14, 17, 31) and appears to be a general characteristic of the soybean genome (17). We infer that each gene within the *CG-4* β-conglycinin region has a unique regulatory domain that programs its expression during the soybean life cycle, at least at the transcriptional level.

***CG-4* β-Conglycinin Gene Is Expressed Correctly in Transformed Tobacco Plants.** Both genetic and molecular analyses indicated that a single-copy soybean gene is present in each transformed tobacco plant (Fig. 3). Nuclease S1 (Fig. 2), RNA gel-blot (Fig. 4), and RNA dot-blot (Fig. 5) studies demonstrated that the 1.7-kb β-subunit mRNA is present at prevalent levels in developing tobacco seeds, that this mRNA

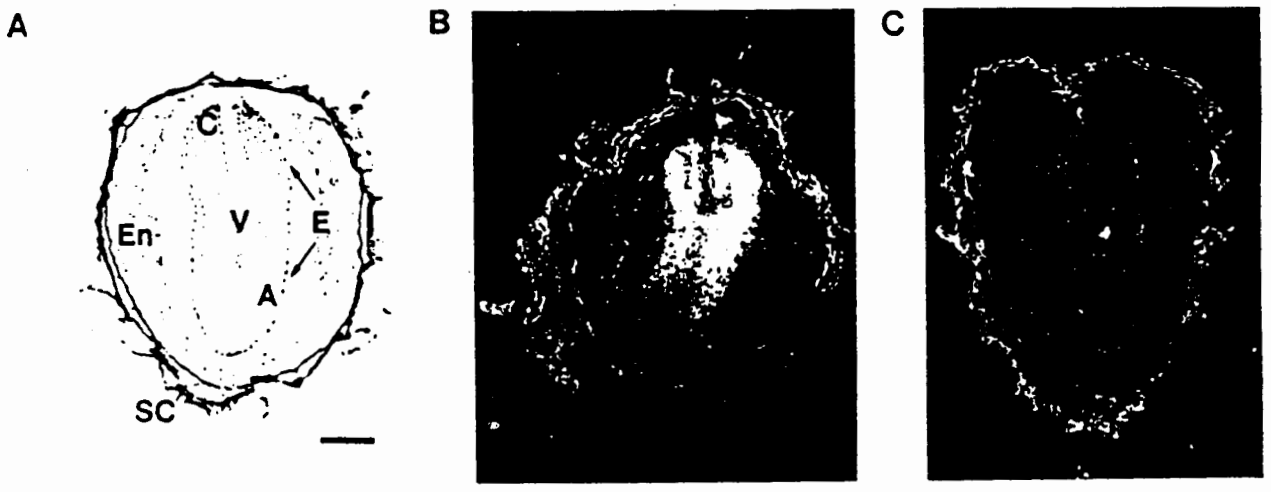


FIG. 6. Localization of β-conglycinin mRNA in transformed tobacco seeds. TOB-T2 19-DAP tobacco seeds were fixed, embedded in paraffin, sectioned, and hybridized *in situ* with single-stranded ^{35}S-labeled RNA probes (Fig. 1*C*, probe III) as outlined. (*A*) Longitudinal section of a tobacco seed. SC, E, En, A, C, and V refer to seed coat, embryo, endosperm, hypocotyl-radicle axis, cotyledons, and provascular cylinder, respectively (22). The photograph was taken using bright-field microscopy. (Bar = 100 μm.) (*B*) Hybridization with a ^{35}S-labeled anti-mRNA probe. White grains represent regions containing RNA·RNA hybrids. The photograph was taken with dark-field microscopy. (*C*) Hybridization with a ^{35}S-labeled mRNA probe. White grains represent background hybridization levels. The photograph was taken with dark-field microscopy.

is undetectable in the tobacco leaf, and that β-conglycinin mRNA accumulates and decays during tobacco seed development. This regulated β-conglycinin gene expression program is qualitatively indistinguishable from that observed in the soybean plant (5, 6, 11).

We have not yet carried out transcription experiments with the transformed tobacco plants analogous to those reported for soybean (11). However, we infer from our results that the transcriptional controls responsible for activating and repressing β-conglycinin gene expression during seed development and the posttranscriptional processes that operate to degrade β-conglycinin mRNA prior to seed dormancy are highly conserved in tobacco and soybean. This inference probably applies to legume seed protein genes in general because several different gene classes (e.g., storage protein, trypsin inhibitor, and lectin) have been shown to be expressed correctly in solanaceous plants (refs. 12, 13, 17, and 32; K. D. Jofuku and R.B.G., unpublished results). We presume that molecules regulating soybean seed protein gene expression have counterparts within the tobacco seed. Whether these molecules are identical or highly related to those in soybean remains to be determined; however, they must be sufficiently homologous in order to recognize soybean control sequences that participate in transcriptional and posttranscriptional regulatory processes.

β-Conglycinin mRNA Is Present Within Specific Tobacco Embryo Cell Types. A clue to the distribution of molecules that control β-conglycinin gene expression within the tobacco seed was obtained from the *in situ* hybridization experiments presented in Fig. 6. No detectable hybridization above background was observed within the endosperm (Fig. 6 *B* and *C*). By contrast, β-conglycinin mRNA was detected within the embryo and was preferentially localized within the cotyledons and upper axis cells (Fig. 6*B*). Greenwood and Chrispeels (33) and Sengupta-Gopalan *et al.* (32) showed that the bean storage protein phaseolin is present primarily in the embryo of transformed tobacco seeds, although a small quantity of protein is found in the endosperm. Our mRNA localization studies indicate that soybean β-conglycinin mRNA is only detectable in the tobacco embryo. Thus, β-conglycinin mRNA is either absent entirely from the tobacco endosperm or is present at a concentration too low to be detected by our *in situ* hybridization procedure.

By comparison with tobacco, soybean seeds are composed almost entirely of embryonic cells and possess a vestigial endosperm (34). During early seed development, however, a significant mass of endosperm tissue surrounds the developing embryo (34). Recently, we localized β-conglycinin mRNA sequences within developing soybean seeds, using *in situ* hybridization (L. Perez-Grau and R.B.G., unpublished data). β-Conglycinin transcripts are only detectable during the maturation phase of embryogenesis and are confined to the embryo. Thus, the distribution of β-conglycinin mRNA within the tobacco seed approximates that which is observed within soybean. We conclude that sequences required for embryonic expression, temporal control, and cell specificity are contained within the 12.3-kb β-conglycinin DNA segment and are probably less than 1 kb 5′ and/or 4 kb 3′ to the *CG-4* β-subunit gene (Fig. 1*A*). Whether the temporal and spatial β-conglycinin gene expression programs are controlled by a single regulatory element or several different elements, and the relationship, if any, between α′-subunit (12, 13) and β-subunit control sequences is not yet known.

We showed recently that a soybean embryo DNA binding protein interacts with a specific soybean lectin gene 5′ region and that its activity parallels lectin gene transcriptional levels during the soybean life cycle (35). The β-conglycinin gene competes with the lectin gene for this DNA binding protein (C. Reeves and R.B.G., unpublished results). Although we have no information regarding the physiological relevance of this DNA binding protein, it may perform some role in regulating seed protein gene transcription. A major conclusion of the present study is that factors regulating seed protein gene expression are compartmentalized within specific regions of soybean and tobacco seeds. What these factors are and how they become spatially distributed during embryogenesis remain to be determined.

We express our gratitude to Dr. Kathleen Cox for teaching us the *in situ* hybridization procedure and to Drs. Diane Jofuku and Jack Okamuro for developing the gene transfer technology in our laboratory. S.J.B. was supported by McKnight Foundation and UCLA Graduate Research Predoctoral Fellowships. This research was supported by a U.S. Department of Agriculture Grant to R.B.G.

1. Hill, J. E. & Breidenbach, R. W. (1974) *Plant Physiol.* **53**, 742–746.
2. Hill, J. E. & Breidenbach, R. W. (1974) *Plant Physiol.* **53**, 747–751.
3. Thanh, V. H. & Shibasaki, K. (1978) *J. Agric. Food Chem.* **26**, 692–695.
4. Coates, J. B., Medeiros, J. S., Thanh, V. H. & Nielsen, N. C. (1985) *Dev. Biol.* **83**, 218–231.
5. Goldberg, R. B., Hoschek, G., Ditta, G. S. & Breidenbach, R. W. (1981) *Dev. Biol.* **83**, 218–231.
6. Beachy, R. N., Jarvis, N. P. & Barton, K. A. (1981) *J. Mol. Appl. Genet.* **1**, 19–27.
7. Doyle, J. J., Schuler, M. A., Gudette, W. D., Zenger, V., Beachy, R. N. & Slightom, J. L. (1986) *J. Biol. Chem.* **261**, 9228–9238.
8. Ladin, B. F., Tierney, M. L., Meinke, P., Hosangadi, P., Veith, M. & Beachy, R. N. (1987) *Plant Physiol.* **84**, 35–41.
9. Davies, C. S., Coates, J. B. & Nielsen, N. C. (1985) *Theor. Appl. Genet.* **71**, 351–358.
10. Meinke, D. W., Chen, J. & Beachy, R. N. (1981) *Planta* **153**, 130–139.
11. Walling, L., Drews, G. N. & Goldberg, R. B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2123–2127.
12. Beachy, R. N., Chen, Z. L., Horsch, R. B., Rogers, S. G., Hoffman, N. J. & Fraley, R. T. (1985) *EMBO J.* **4**, 3047–3053.
13. Chen, Z. L., Schuler, M. A. & Beachy, R. N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8560–8564.
14. Fischer, R. L. & Goldberg, R. B. (1982) *Cell* **29**, 651–660.
15. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
16. Goldberg, R. B., Hoschek, G., Tam, S. H., Ditta, G. S. & Breidenbach, R. W. (1981) *Dev. Biol.* **83**, 218–231.
17. Okamuro, J. K., Jofuku, K. D. & Goldberg, R. B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8240–8244.
18. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
19. Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3693–3697.
20. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
21. Cox, K. H., DeLeon, D. V., Angerer, L. M. & Angerer, R. C. (1984) *Dev. Biol.* **101**, 485–502.
22. Avery, G. S. (1933) *Am. J. Bot.* **20**, 309–327.
23. Berk, A. J. & Sharp, P. A. (1977) *Cell* **12**, 721–722.
24. Rosbash, M., Blank, D., Fahrner, K., Hereford, L., Ricciardi, R., Roberts, B., Ruby, S. & Woolford, J. (1979) *Methods Enzymol.* **68**, 454–469.
25. Horsch, R. B., Fry, J. E., Hoffman, N. L., Eichholtz, D., Rogers, S. E. & Fraley, R. T. (1985) *Science* **227**, 1229–1231.
26. Goldberg, R. B. (1986) *Philos. Trans. R. Soc. London Ser. B* **314**, 343–353.
27. DeBlock, M., Herrera-Estrella, L., Van Montagu, M., Schell, J. & Zambryski, P. (1984) *EMBO J.* **3**, 1681–1689.
28. Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M. & Schell, J. (1983) *EMBO J.* **2**, 2143–2150.
29. Hain, R., Czernilofsky, A. P., Steinbiss, H. N., Herrera-Estrella, L. & Schell, J. (1985) *Mol. Gen. Genet.* **199**, 161–168.
30. Okamuro, J. K. & Goldberg, R. B. (1985) *Mol. Gen. Genet.* **198**, 290–298.
31. Lee, J. S., Brown, G. G. & Verma, D. P. S. (1983) *Nucleic Acids Res.* **11**, 5541–5553.
32. Sengupta-Gopalan, C., Reichert, N. A., Barker, R. F., Hall, T. L. & Kemp, J. D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3320–3324.
33. Greenwood, J. S. & Chrispeels, M. J. (1985) *Plant Physiol.* **79**, 65–71.
34. Carlson, J. B. (1973) in *Soybeans: Improvement, Production, and Use*, ed. Caldwell, B. E. (Am. Soc. Agronomy, Madison, WI), pp. 17–95.
35. Jofuku, K. D., Okamuro, J. K. & Goldberg, R. B. (1987) *Nature (London)* **328**, 734–737.

Accumulation and assembly of soybean β-conglycinin in seeds of transformed petunia plants

R.N.Beachy, Z.-L.Chen, R.B.Horsch[1], S.G.Rogers[1], N.J.Hoffmann[1] and R.T.Fraley[1]

Department of Biology, Washington University, St Louis, MO 63130, and [1]Monsanto Company, 700 Chesterfield Village Parkway, St Louis, MO 63198, USA

Communicated by G.Kreil

A gene encoding the α′-subunit of β-conglycinin, a seed storage protein of soybean (*Glycine max*), was transformed into petunia cells on a disarmed Ti-plasmid of *Agrobacterium tumefaciens*, and plants were regenerated. Transcripts of the introduced gene accumulated in immature embryos but not in leaves of the transformed plants. Soybean protein was first detected immunologically in proteins extracted from embryos at 10 days post pollination (d.p.p.), concurrent with the accumulation of subunits of the major petunia seed proteins. Between 10 and 16 d.p.p. the primary soybean protein detected had an apparent mol. wt. of 55 kd. The 76-kd α′-subunit and several smaller polypeptides accumulated between 16 and 24 d.p.p., when seeds had matured. Polypeptides <76 kd probably resulted from specific proteolytic cleavage of the α′-subunit. The α′-subunit and the smaller polypeptides assembled into multimeric proteins with sedimentation coefficients of 7 – 9S, similar to the sedimentation coefficients of β-conglycinins isolated from soybean seeds. This transformation and expression system should be ideally suited for testing gene mutations to alter the amino acid composition of these seed storage proteins.

Key words: β-conglycinin/soybean/seed storage protein/petunia

Introduction

Storage proteins accumulate to high levels in many types of seeds as a means of storing nitrogen for use during germination and early seedling growth. Because the genes encoding seed storage proteins are expressed at a specific period in the life cycle of the plant, seed proteins represent excellent model systems for studies of the regulated expression of genes in plants.

Legume seeds contain proportionately larger amounts of proteins than most cereal grains, with soybean seeds containing as much as 50% protein by weight. Soybeans contain two major storage proteins, the glycinins or 11S proteins, and the β-conglycinins, or 7S proteins. The latter are trimeric molecules of approximate mol. wt. 200 000, and are produced by assembly of three subunits, α′ (76 kd), α (72 kd) and β (53 kd) subunits. The expression of the genes encoding the subunits of β-conglycinin and the accumulation of their respective mRNAs has been studied in several laboratories (Goldberg *et al.*, 1981; Meinke *et al.*, 1981). More recent studies of the β-conglycinin genes have included *in vitro* run-off transcription experiments (Walling, Harada and Goldberg, in preparation), micrococcal nuclease and DNase sensitivity of gene sequences in nuclei isolated from embryos (Kerney and Beachy, unpublished), and studies of the role of abscisic acid in regulating the expression of genes in embryos (Bray and Beachy, 1985). Studies of the biosynthesis and assembly of the subunits to form β-conglycinin (Beachy *et al.*, 1981; Shattuck-Eidens and Beachy, 1985), and degradation of β-conglycinin during seed germination (Bryant, Hosangadi and Beachy, in preparation) have also been completed. These and other experiments have resulted in a good understanding of the biology of the β-conglycinins, and the types of regulation that control their accumulation during seed development.

Soybean seed proteins, in particular the β-conglycinin, contain low amounts of sulfur-amino acids (Holowach, 1984; Koshiyama, 1983), preventing the use of soy protein as a fully balanced source of protein. It may be possible to alter the amino acid composition of the storage protein genes to improve the nutritional quality of the encoded protein. Each mutation made, however, must be tested for effects of the mutation on the structure and function of the protein product. In this paper we report the results of experiments to introduce a gene encoding the α′-subunit β-conglycinin into petunia cells *via* a disarmed Ti-plasmid in *Agrobacterium tumefaciens*. Seeds produced on plants regenerated from transformed cells produced high levels of the α′-subunit and a 55-kd breakdown product of the α′-subunit. Soybean proteins isolated from petunia seeds have sedimentation coefficients of 7 – 9S, consistent with the hypothesis that they assemble into a multimeric form similar to the assembly of β-conglycinin in soybeans. This system should provide a convenient assay for testing *in vivo* the effect of mutations that alter the amino acid composition of a soybean protein on its structure and function.

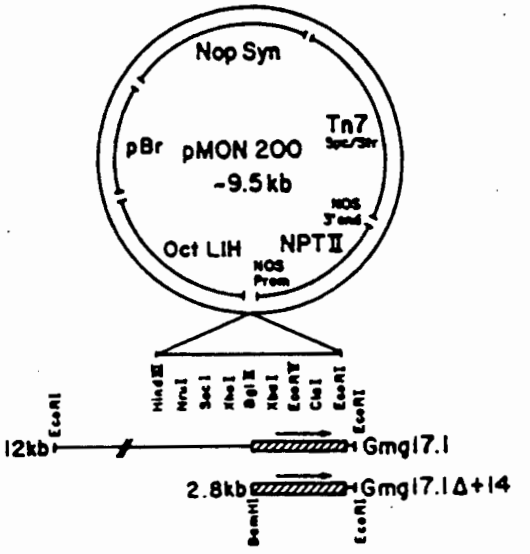
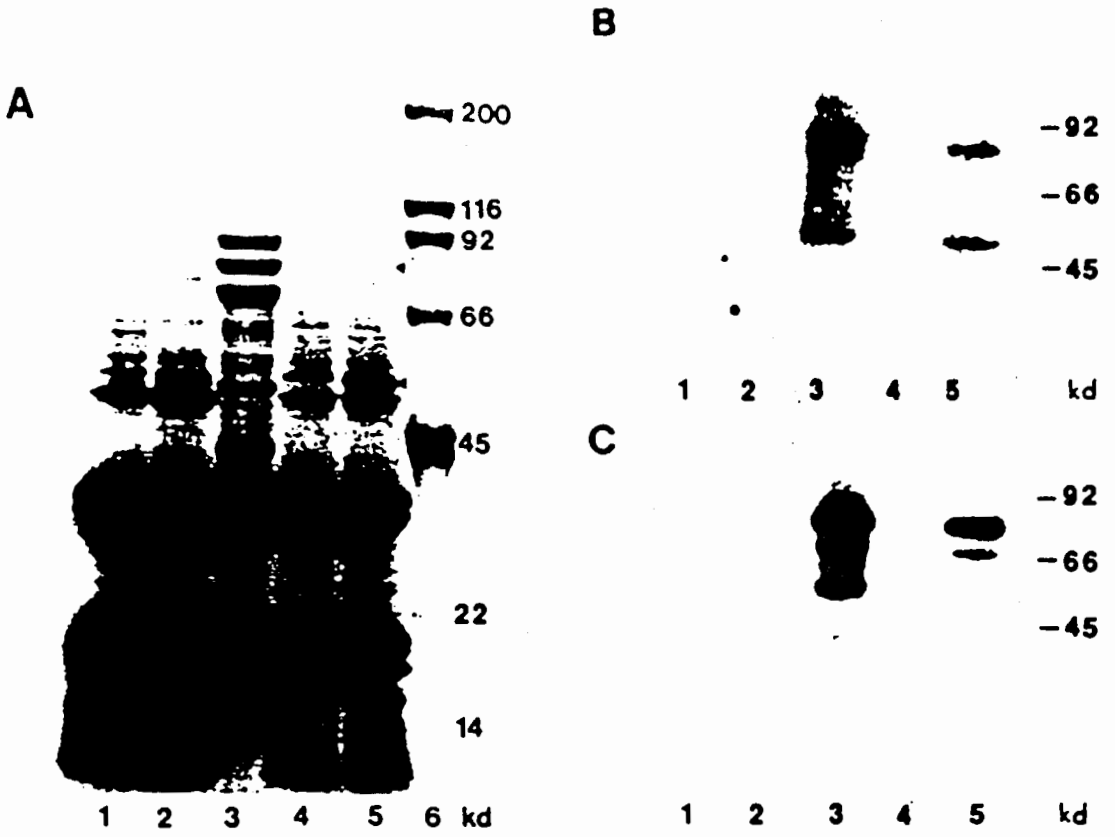


Fig. 1. Intermediate plasmid used to deliver target genes to *A. tumefaciens* Ti-plasmid. pMON200 (Fraley *et al.*, 1985) was linearized with *Eco*RI, or *Eco*RI and *Bgl*II, and ligated with Gmg 17.1 or Gmg 17.1 Δ + 14. Gmg 17.1 contains the α′-subunit gene (boxed area; arrows indicate direction of transcription) flanked by 11.5 kb of DNA 5′ of the gene. Gmg 17.1 Δ + 14 lacks 5′ flanking sequences, and begins at nucleotide +14.



2. Accumulation of the soybean α'-subunit in petunia seeds. 50 μg of protein extracted from petunia seeds or 10 μg of soybean protein were subjected to SDS-PAGE and stained with Coomassie blue (A) or immunoblot analysis (B,C). Lane 1: protein from seeds of normal petunia plants; lane 2: protein from seeds of petunia plants transformed to contain vector pMON200; lane 3: protein from soybean seeds; lane 4: protein from seeds of petunia plants transformed to contain Gmg 17.1 on pMON200; lane 5: protein from seeds of petunia plants transformed to contain Gmg 17.1 Δ + 14 on pMON200. (A) is a 5–25% gradient gel. In B and C, proteins were separated on 10% acrylamide gels and then electroblotted to nitrocellulose. Blotted proteins were reacted with a rabbit polyclonal antibody raised against soybean β-conglycinin followed by ^{125}I-donkey anti-rabbit antibodies (B) or a mouse monoclonal antibody that reacts to the α'-subunit followed by ^{125}I-rabbit anti-mouse antibodies (C). The positions of mol. wt. markers are indicated.

Results

Accumulation of soybean proteins and mRNA in seeds of transformed petunia plants

)nts were regenerated from transformed cells selected on medium containing kanamycin as described by Horsch *et al.* (1985). To confirm that regenerated plants were transformed, leaf segments were removed periodically and assayed for the presence of nopaline by paper electrophoresis using purified nopaline as standard. Kan^r nop$^+$ plants were grown in the greenhouse, the flowers were self-pollinated and mature seeds were collected. 50 μg of seed protein was extracted and subjected to SDS-PAGE on gels containing a gradient of 5–25% polyacrylamide. Proteins that stained with Coomassie blue are shown in Figure 2A. Included as controls in these experiments were proteins extracted from non-transformed petunia seed (lane 1), transformed petunias carrying pMON200 (lane 2), and transformed petunias carrying Gmg 17.1 Δ + 14 (lane 4). These proteins are compared with proteins extracted from transformed petunias carrying Gmg 17.1 (lane 5) and with proteins extracted from soybean seeds (lane 3). The arrows to the left indicate the position of new protein bands found in plants containing Gmg 17.1 that are not present in the controls. The intensely stained proteins in lanes 1, 2, 4 and 5 with apparent mol. wts. of 30–35 kd and 17–20 kd are subunits of the major petunia seed protein, a legumin-like protein with a sedimentation coefficient of ~12S.

Seed proteins were also subjected to analysis by immunoblot

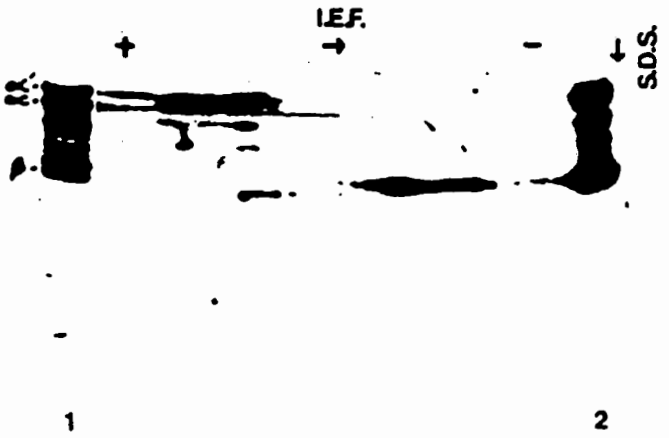


Fig. 3. Immunoblot analysis of petunia seed proteins separated on a two-dimensional gel. Petunia seed proteins were separated in the first dimension by isoelectric focussing and in the second dimension by SDS-PAGE using the methods of O'Farrell (1975). The direction of the first dimension is indicated at the top. Lane 1 contains soybean seed proteins; lane 2 contains petunia seed proteins. Electroblotted proteins were reacted with the polyclonal antibody as described in Figure 2.

assays using polyclonal antibodies against total soybean β-conglycinin raised in rabbits, or to a monoclonal antibody AG3.4 that is specific for the α'-subunit. As shown in Figure 2B and C, only transformed petunia plants carrying Gmg 17.1 contain pro-

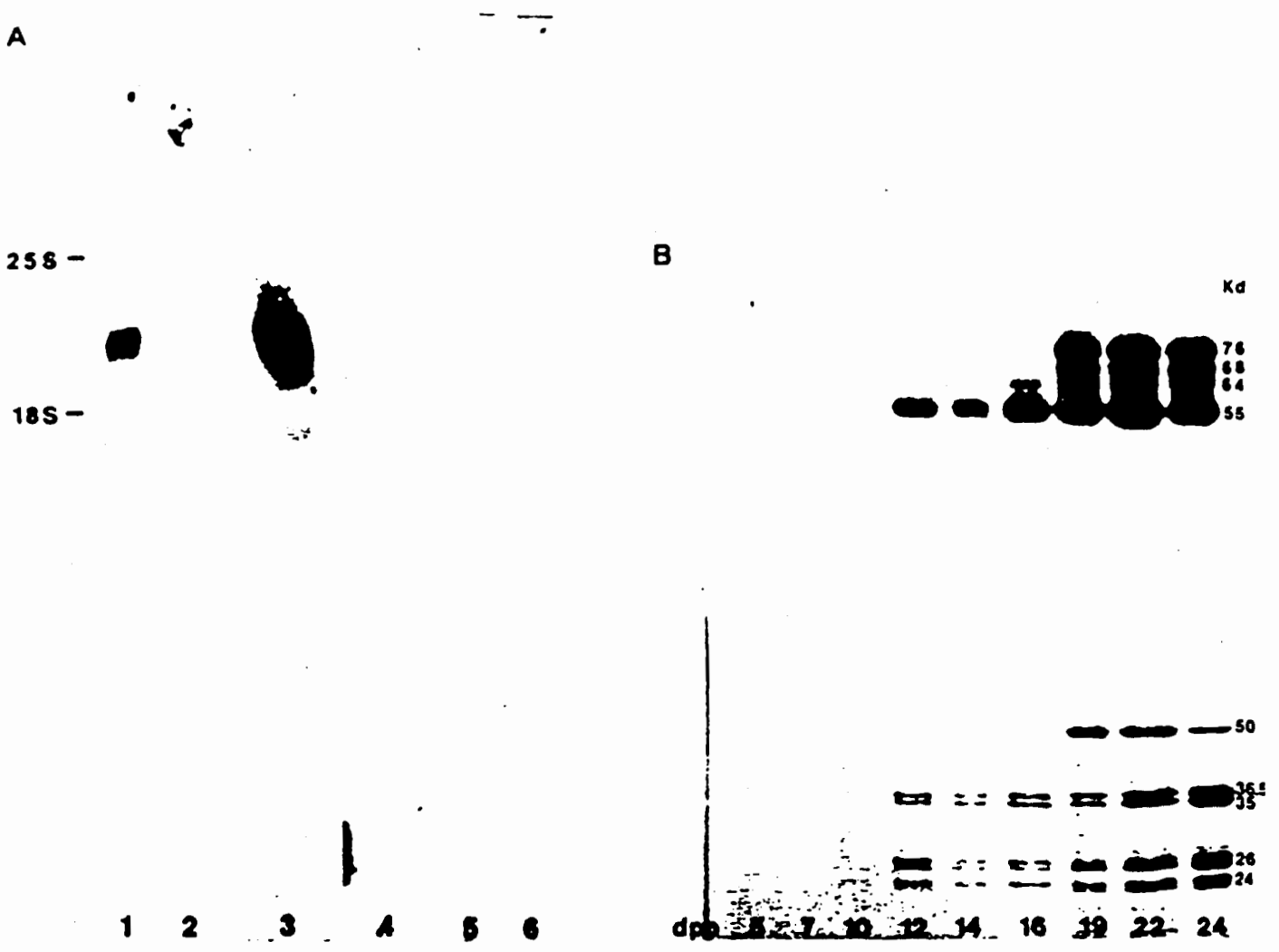


Fig. 4. Regulated expression of the α′-subunit genes in transformed petunia plants. (A) Northern blot analysis of 10 μg of total RNA extracted from immature embryos of plants transformed to contain Gmg 17.1 (lane 1) or Gmg 17.1 Δ + 14 (lane 2). Lane 3 contains RNA from immature soybean seeds; lane 4, poly(A)$^+$ RNA from leaves of normal petunia; lane 5, poly(A)$^+$ RNA from leaves transformed to contain Gmg 17.1, lane 6, poly(A)$^+$ RNA from leaves of plants transformed to contain Gmg 17.1 Δ + 14. (B) Analysis of proteins extracted from embryos collected from transformed plants containing Gmg 17.1 at increasing d.p.p. Each lane contained 20 μg of protein. Upper: electroblotted proteins were reacted with the rabbit polyclonal antibody against soybean β-conglycinin, followed by ^{125}I-donkey anti-rabbit antiserum. Lower: proteins were stained with Coomassie blue. The apparent mol. wt. of the most abundant seed proteins is indicated.

proteins which reacted with the antibodies used in these experiments. The 76 kd protein has electrophoretic mobility similar to the α′-subunit of soybean, and is immunoreactive with the α′-specific monoclonal antibody. The major immunoreactive protein has an apparent mol. wt. of 55 kd, slightly greater than the soybean β-subunit (53 kd), and was not immunoreactive with Ag3.4. In addition, other polypeptides, with apparent mol. wts. of 68, 64 and 20 kd, were also detected in the protein samples extracted from the plants transformed with Gmg 17.1 (Figure 4B).

To characterize further the soybean proteins that accumulated in petunia seeds, total seed proteins were separated in two-dimensional gels and analyzed by immunoblots. As shown in Figure 3, the α′-subunit from petunia seeds is made up of at least two charged isomeric forms, while the 55 kd polypeptide exhibits a wide range of charge heterogeneity. Each of the other polypeptides that are antigenically related to the α′-subunit are also made up of multiple charged isomers. Heterogeneity in charge of α′-subunits was also observed in proteins extracted from soybean seeds (Ladin *et al.*, 1983).

Since multiple polypeptides related to β-conglycinin accumulated in petunia seeds it was important to determine whether single or multiple soybean mRNA(s) was produced in transformed petunias. Poly(A)$^+$ RNA from leaves and total RNA from seeds were hybridized on Northern blots with pMON200 containing Gmg 17.1 Δ + 14 as the probe. As shown in Figure 4A only seeds of petunia plants carrying Gmg 17.1 (lane 1) contained an RNA with mobility equivalent to the α′-mRNA extracted from the soybean seeds (lane 3). This mRNA is ~2400 nucleotides in length, the size of the α′-mRNA as determined by Beachy *et al.* (1981), and by DNA sequence analysis of the α′-gene on Gmg 17.1 (Doyle *et al.*, in preparation) and a cloned cDNA corresponding to the α′-mRNA (Schuler *et al.*, 1982). Leaves of transformed plants did not contain the 2.4-kb α′-mRNA.

The results of these experiments indicated that the α′-subunit gene on Gmg 17.1 is expressed in a tissue-specific manner in transformed petunias, producing the expected 2.4-kb mRNA. The product of the translation of this mRNA is the α′-subunit. The other antigenically related polypeptides found in seeds of plants

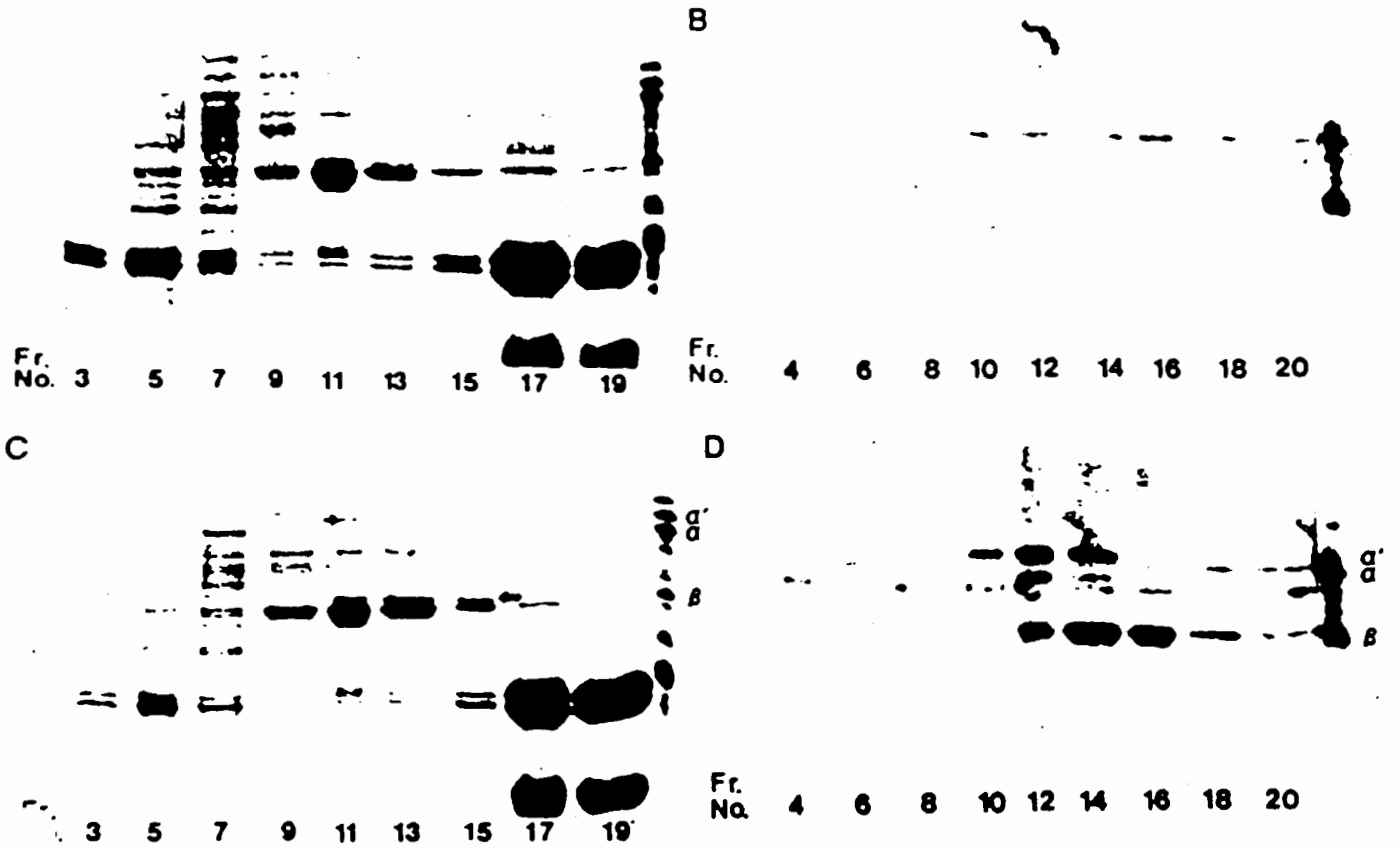


Fig. 5. Sedimentation of petunia seed proteins in sucrose density gradients. 5 mg of proteins extracted from petunia seeds was sedimented in 5–20% linear sucrose gradients. Proteins were acetone precipitated from 20 fractions of the 14 ml gradients, subjected to SDS-PAGE, and stained with Coomassie blue (A,C) or analyzed by immunoblotting with polyclonal antibodies (B,D) as described in Figure 2. Panels A and B show the separation of proteins from transformed plants containing Gmg 17.1 Δ + 14; (C) and (D) show the separation of proteins from transformed plants containing Gmg 17.1. Arrows in (C) indicate proteins in seeds of plants carrying Gmg 17.1 that are not present in seeds of plants carrying Gmg 17.1 Δ + 14. The right-most lanes in each panel contain soybean seed proteins. β-Conglycinin from soybean seed sediments in fractions 10–14 (not shown).

containing the α'-gene are presumed to result from limited proteolysis of the α'-subunit, but further characterization of the related polypeptides remains to be done.

Accumulation of soybean protein during development of petunia seeds

The β-conglycinins are temporally regulated throughout soybean embryogenesis (Meinke *et al.*, 1981; Goldberg *et al.*, 1981; Gayler and Sykes, 1981). To determine if the α'-subunit gene present on Gmg 17.1 was temporally regulated in petunia embryos, proteins were extracted from petunia seeds taken at various days post pollination (d.p.p.) and subjected to SDS-PAGE and immunoblot analyses. In these experiments the seed capsule was mature 22 d.p.p. The 55 kd polypeptide related to the α'-subunit was first detected at 10 d.p.p. and the 76 kd α'-subunit and the other smaller polypeptides were detected at later times in seed development (Figure 4B, upper panel). The petunia seed storage proteins were also first detected at 10 d.p.p. (Figure 4B, lower panel) indicating that accumulation of the soybean protein is temporally regulated during petunia seed development in a manner similar to that of the major petunia storage proteins.

Assembly of the soybean polypeptides into multimeric proteins in petunia seeds

In soybean seeds α'-subunits are assembled with α- and β-subunits to produce multimeric proteins with sedimentation coefficients of 7–9S (Thanh and Shibasaki, 1978; reviewed by Koshiyama, 1983). To determine whether the α'-subunit in petunia seeds was assembled into multimers, 5 mg of protein from mature petunia seeds collected from transformed plants containing Gmg 17.1 or Gmg 17.1 Δ + 14 were subjected to sedimentation in sucrose gradients. Proteins present in 0.6-ml fractions were collected and resolved by SDS-PAGE and were either stained with Coomassie blue or transferred to nitrocellulose for immunoblot analyses. As shown in Figure 5A and C the stained protein patterns are similar in the two samples except for the presence of the 76 and 55 kd polypeptides (indicated by arrows) in fractions 10–16. These proteins are immunologically related to β-conglycinin (Figure 5D). The 76 kd protein was present in fractions 10–13 and has sedimentation characteristics similar to those of β-conglycinin isolated from soybeans (data not shown; see Beachy *et al.*, 1979). The 55 kd polypeptide was, however, present in protein complexes that sedimented between 7S and ~9S (fractions 10–18) indicating that under these experimental conditions the multimeric forms of the protein containing primarily the 55 kd protein are physically different from proteins that contain primarily the α'-subunit.

Quantitation of soybean proteins in petunia seeds and plant-to-plant variability

We examined the seed proteins of four independently transformed

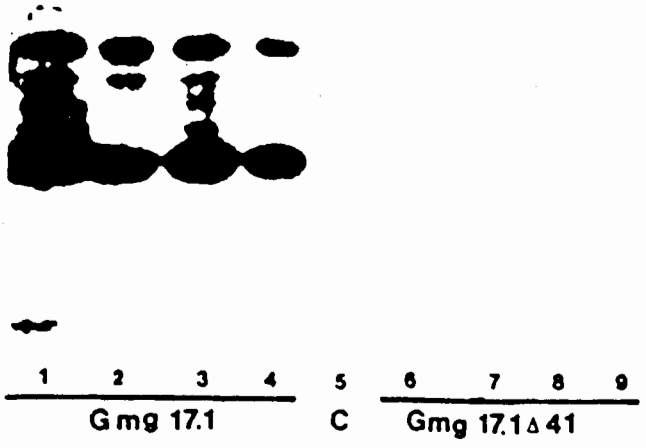


Fig. 6. Immunoblot analysis of proteins from four different transformants containing Gmg 17.1 or Gmg 17.1 Δ + 14. Each lane contains equivalent amounts of protein. Electroblotted proteins were allowed to react with the polyclonal antibody as described in Figure 2. Position of the 76-kd α′-subunit and the 55-kd polypeptide are indicated.

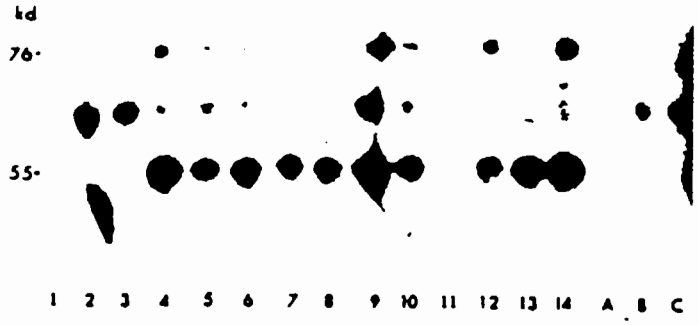


Fig. 7. Immunoblot analysis of the segregation of Gmg 17.1 in single seeds of transformed petunia plants. Proteins extracted from single seeds of a transformed petunia plant containing Gmg 17.1 were reacted with the polyclonal antibody as described in Figure 2 (lanes 1 – 14). Lanes A, B and C contain proteins of seeds from untransformed petunia, petunia transformed with pMON200, and petunia transformed to contain Gmg 17.1 Δ + 14, respectively.

plants carrying Gmg 17.1 to determine whether or not there was marked variability in the amount of accumulated soybean protein in different plants. In seeds of each of the four plants there was a nearly identical amount of soybean protein (Figure 6). The amount of soybean protein in the seeds was estimated to be between 0.1 and 1% of the total seed proteins extracted in an SDS-containing buffer. We also examined the amount of protein produced in plants that contained Gmg 17.1 in opposite orientation (relative to the NPTII gene; Figure 1) to that in these four plants. The level of protein in these plants was identical to that shown in Figure 6 (data not shown). Lanes 6 – 9 in Figure 6 contain proteins extracted from seeds of four different plants containing Gmg 17.1 Δ + 14, none of which produced a protein that was antigenically related to the α′-subunit.

To study the inheritance and segregation of Gmg 17.1 Western blot analyses were carried out on protein extracts prepared from single seeds from one transformed plant. In these experiments 8 of 77 seeds (75.3%) examined contained soybean protein, closely approximating the expected segregation ratios of 3:1. Figure 7 presents the results of one experiment in which 14 such seeds were examined. Lanes 1, 3 and 11 were scored as negative since they do not contain either the 76 kd or the 55 kd polypeptides. Lanes A, B and C contain extracts of seeds from various control plants (as indicated in the figure legend). We were unable to distinguish unequivocally homozygous seeds from heterozygous seeds on the basis of these experiments, probably due to the unequal extraction of proteins from individual seeds.

Discussion

The gene encoding the α′-subunit of the soybean seed storage protein β-conglycinin was introduced into petunia cells *via* a modified Ti-plasmid. The intermediate plasmid pMON200, which contains several unique restriction sites to facilitate the introduction of genes, and a strain of *A. tumefaciens* carrying a disarmed Ti-plasmid (Fraley *et al.*, 1985) were used throughout. The soybean proteins that accumulated in petunia seeds have physical characteristics similar to β-conglycinins isolated from soybean seeds. First, there are multiple isomeric forms of the 76-kd α′-subunit both in petunia seeds (Figure 3) and in soybean seeds (Ladin *et al.*, 1983, Figure 2). This is an important observation since it indicates that the number of charged isomers of the subunits does not necessarily reflect the number of different genes or alleles that encode these subunits.

Second, the 76-kd α′-subunit and the antigenically related 68, 64, 55, 20 and 18 kd polypeptides produced in petunia seeds assemble into multimeric proteins with sedimentation coefficients similar to those of the β-conglycinins from soybean seeds. The majority of the β-conglycinins from soybean and petunia seeds have sedimentation coefficients of 7 – 8S under the conditions used in these experiments (Figure 5). However, other β-conglycinins have sedimentation coefficients of ~9 – 11S. This is especially apparent in the case of the 55 kd polypeptide that accumulates in petunia seeds (fractions 14 – 18, Figure 5C,D). It is well known that the sedimentation properties of the β-conglycinins are variable in solutions containing different molarities of the NaCl (Koshiyama, 1983). However, it is unknown whether sedimentation of the 55 kd polypeptide at ~9S reflects an altered assembly of the protein, perhaps into a hexameric molecule (Koshiyama, 1983), or assembly with other petunia proteins to produce novel multimeric proteins.

Thanh and Shibasaki (1978) reported that β-conglycinin oligomers containing three α′-subunits were not present in soybean seeds. In petunia seeds which contain only the α′-subunit, oligomers comprised of the α′-subunit and polypeptides derived therefrom were found. These results clearly demonstrated that the product of the α′-subunit gene can assemble into a multimeric protein, and may indicate that the conclusions of Thanh and Shibaski (1978) should be re-examined. On the other hand, our conclusion that the α′-subunit self-reassmbles in petunia seeds is tentative until it is demonstrated that β-conglycinin produced in petunia seeds does not contain petunia polypeptides.

A third similarity between soybean and petunia β-conglycinins is that each contains numerous polypeptides that are antigenically related to, and probably derived from, the α′-subunit. These polypeptides in soybean β-conglycinins are collectively designated the γ-subunits (Ladin *et al.*, in preparation), and generally have mol. wts. between 68 and 55 kd. Several of these polypeptides have electrophoretic mobilities similar to the 68, 60 and 55 kd polypeptides found in petunia seeds (Figure 2). Whether the polypeptides found in petunia seeds are in fact identical to those found in soybean seed remains to be determined.

The 55-kd polypeptide is the first soybean protein that ac-

cumulates in petunia seeds during embryogenesis, followed by the 76-kd α'-subunit (Figure 3) and the 68 and 64 kd polypep- . This suggests that the 76 kd protein is initially unstable in the immature embryos, but becomes increasingly stable as the seed matures. Recently we reported that the α'- and α-subunits are also unstable during early stages of soybean seed development (Shattuck-Eidens and Beachy, 1985). Taken together, these results lend support to the hypothesis that accumulation of β-conglycinin is, to some degree, dependent upon the stability of the protein in the protein body. Protein bodies in soybean seeds are derived from the cell vacuole (Beachy *et al.*, 1979; Yoo and Chrispeels, 1979), and it appears that storage proteins are inherently resistant to the activity of hydrolases in vacuolar protein bodies, or that the hydrolases are less active in mature seeds than in immature seeds. Either or both conditions would result in accumulation of seed storage proteins as the seed matures. Sengupta *et al.* (1985) recently reported the expression of a gene encoding the β-subunit of phaseolin in seeds of transformed tobacco plants. In this study the 48 kd phaseolin β-subunit was the first protein that accumulated while smaller polypeptides derived f the β-subunit accumulated late in seed development. These authors suggested that the phaseolin β-subunit was less protected from proteolysis as the seed matures. Thus, the results of the experiments reported by Sengupta *et al.* (1985) differ significantly from those reported here.

Plant transformation experiments have previously been used to study the expression of genes under control of promoters taken from the T-DNA region of the Ti-plasmid (Fraley *et al.*, 1983; Bevan *et al.*, 1983; Herrera-Estrella *et al.*, 1983), from cauliflo mosaic virus (Koziel *et al.*, 1984), and from the small su t of ribulose bisphosphate carboxylase (RuBisCo; Broglie *et al.*, 1984; Herrera-Estrella *et al.*, 1984). In the latter case, transcription was induced by exposing transformed tissues to light. We have studied the expression of a gene which is under stringent developmental control and is expressed primarily in developing embryos. We have found that Gmg 17.1 is under tight control in plants regenerated from transformed petunia cells, and is expressed in much higher levels in developing seeds than in other parts of the plant (to be reported elsewhere). Similar results have recently been reported for the phaseolin gene (Sengupta *et al.* 985). We are currently identifying the DNA sequences wh are required for developmental regulation of the Gmg 17.1 gene.

Since the protein product of the α'-gene Gmg 17.1 assembles into a multimeric form in petunia, this system represents an excellent model system for testing proteins with modified amino acid composition. Each modified gene can be tested in petunia plants where the expression and stability of the modified protein can be determined. Such studies may make it possible to improve the nutritional quality of these proteins with the anticipation of eventually returning the modified genes into agronomically valuable seed grains.

Materials and methods

Construction of the intermediate plasmid and conjugation into A. tumefaciens

The isolation and partial characterization of a 12-kb *Eco*RI fragment of soybean genomic DNA containing an α'-subunit gene has been previously reported (Schuler *et al.* 2). On this fragment, Gmg 17.1, the α'-gene is flanked on the 5' end by ~ of DNA and by 0.4 kb of DNA on the 3' end. No other storage protein genes are located on this DNA fragment. A second gene fragment used in these experiments, referred to as Gmg 17.1 Δ + 14, was produced by restriction digestion and *Bal*31 exonuclease digestion to remove all sequences 5' of the gene, and ending at nucleotide +14 (+1 is the first nucleotide transcribed). Gmg 17.1 Δ + 14 contains a *Bam*HI site at its 5' end. The methods for creating this deletion will be described in a subsequent paper.

The intermediate plasmid in these experiments was pMON200 (Figure 1). This plasmid is derived from pMON128 (Fraley *et al.*, 1983) and contains genes that permit selection in bacteria as well as in transformed plant cells. A full description of pMON200 will appear elsewhere (Rogers *et al.*, in preparation). Gmg 17.1 or Gmg 17.1 Δ + 14 was ligated to pMON200 restricted with *Eco*RI, or pMON200 restricted with *Eco*RI and *Bgl*II. pMON200 and derivatives were conjugated into *A. tumefaciens* by the triple mating procedure previously described. For these experiments *A. tumefaciens* strain GV3111SE, carrying a disarmed Ti-plasmid (pTiB6S3-SE) in which the T-DNA phytohormone biosynthetic genes, the T_L DNA right border and all of T_R DNA have been deleted and replaced with a bacterial kanamycin resistance marker was utilized as the recipient. Full details of the pTiB6S3-SE construction are described elsewhere (Fraley *et al.*, 1985). Transformation of leaf discs or strips and plant regeneration of a diploid hybrid petunia was carried out as previously described (Horsch *et al.*, 1985). Regenerated plants were tested for the production of nopaline as described by Otten and Schilperoort (1978).

Extraction and analysis of proteins from petunia seeds

Proteins were extracted from immature seeds and from single mature seeds by grinding in Laemmli (1970) sample buffer (2% SDS, 0.285 M Tris-HCl, pH 6.8, 20% glycerol, 4% β-mercaptoethanol and 0.0025% bromophenol blue). Proteins for two-dimensional gel analysis and sucrose gradient centrifugation were extracted by grinding seeds in the high salt buffer used to extract globulin seed proteins, i.e., 0.4 M NaCl, 0.035 M $NaPO_4^-$, 0.01 M β-mercaptoethanol, pH 7.2. Extracted proteins were quantitated by the Coomassie blue assay (Bradford, 1976) using bovine serum albumin as standard.

SDS-PAGE was carried out essentially as described by Laemmli (1970) in gels containing 10% polyacrylamide in the resolving gel and 5% polyacrylamide in the stacking gel. Gels were either 23 cm × 15 cm × 1.5 mm or 7.7 cm × 13 cm × 0.7 mm. Proteins subjected to electrophoresis in one or two dimensions (O'Farrell, 1975) were electrophoretically transferred to nitrocellulose and reacted with antibodies essentially as described by Symington *et al.* (1981). Polyclonal antibodies to β-conglycinins were produced in rabbits as previously described (Beachy *et al.*, 1981). In some experiments a monoclonal antibody Ag3.4 was used. Ag3.4 is immunoreactive with the α'-subunit and closely related breakdown products, but not to α- or β-subunits of β-conglycinin or their breakdown products. Characterization of Ag3.4 will be reported elsewhere. Antibody:antigen reactions were detected with ^{125}I-donkey anti-rabbit antibody for the polyclonal antibodies, or ^{125}I-rabbit anti-mouse antibodies in the case of AG3.4.

Isolation and analysis of RNA from transformed tissues

RNA was isolated from leaves by multiple phenol extraction essentially as described by Lane and Tumaitis-Kennedy (1981). RNA was isolated from immature seeds removed from seed capsules collected between 15 and 20 d.p.p. Immature fresh seeds were immersed in phenol and homogenized with an equal volume of 1 M Tris-HCl, pH 9.0, with a ground glass pestle in a microfuge tube. The phenol phase was re-extracted with 0.1 M Tris, pH 9.0, and combined aqueous phases were re-extracted with phenol. RNAs that were precipitated with 2.0 M NaCl were collected and subjected to oligo(dT) cellulose chromatography. Poly(A)$^+$ RNA or total RNA [not selected on oligo(dT) cellulose] was resolved by electrophoresis in 1.2% agarose gels containing formaldehyde, and blotted to nitrocellulose as described in Maniatis *et al.* (1982). The hybridization probe was pMON200 containing Gmg 17.1 Δ + 14 labeled with ^{32}P by nick-translation to a sp. act. of $1-3 \times 10^8$ c.p.m./μg DNA. The filters were hybridized in a solution containing 4 × SSC and 50% formamide at 42°C. After hybridization filters were washed at 65°C in 1 × SSC containing 0.1% SDS prior to exposure to X-ray film for 1–4 days.

Sucrose density gradient centrifugation

Linear sucrose gradients containing 5–20% sucrose in grinding buffer were overlaid with 5 mg of proteins extracted from petunia seeds. After centrifugation at 20°C for 14–16 h at 35 000 r.p.m. in a Beckman SW40 rotor, gradients were fractionated on an ISCO gradient fractionator and monitored at 280 nm. The collected fractions were dialyzed to remove sucrose, and proteins were precipitated by addition of five volumes of acetone. Proteins were suspended in Laemmli (1970) sample buffer and subjected to SDS-PAGE.

Acknowledgements

The authors thank Matt Harline, P.Hosangadi, P.Dubé and M.Dyer for technical assistance and maintenance of transformed plants, and Dr M.L.Tierney for helpful discussion during the course of the work. This research was supported by a grant from the United States Department of Energy, Biological Energy Research. Z.-L.Chen was supported by a graduate fellowship from the Peoples Republic of China.

References

Beachy,R.N., Jarvis,N.P. and Barton,K.A. (1981) *J. Mol. Appl. Gen.*, 1, 19-27.

Beachy,R.N., Thompson,J.F. and Madison,J.T. (1979) in Rubenstein,I., Phillips, R.L., Green,C.E. and Gengenbach,B.G. (eds.), *The Plant Seed. Development, Preservation and Germination*. Academic Press, NY, pp. 67-84.

Bevan,M.W., Flavell,R.B. and Chilton,M.-D. (1983) *Nature*, 304, 184-187.

Bradford,M.M. (1976) *Anal. Biochem.*, 72, 248-254.

Bray,E. and Beachy,R.N. (1985) *Plant Physiol.*, in press.

Broglie,R., Coruzzi,G., Fraley,R.T., Rogers,S.G., Horsch,R.B., Niedermeyer, J.G., Fink,C.L., Flick,J.S. and Chua,N.-H. (1984) *Nature*, 224, 838-843.

Fraley,R.T., Rogers,S.G., Horsch,R.B., Sanders,P.S., Flick,J.S., Adams,S.P., Bittner,M.L., Brand,L.A., Fink,C.L., Fry,J.S., Galluppi,G.R., Goldberg,S.B., Hoffmann,N.L. and Woo,S.C. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 4803-4807.

Fraley,R.T., Rogers,S.G., Horsch,R.B., Eichholtz,D.A., Flick,J.S., Fink,C.L., Hoffmann,N.A. and Sanders,P.R. (1985) *Biotechnology*, in press.

Gayler,K.R. and Sykes,G.E. (1981) *Plant Physiol.*, 67, 958-961.

Goldberg,R.B., Hoschek,G., Ditta,G.S. and Breidenbach,R.W. (1981) *Dev. Biol.*, 83, 218-231.

Herrera-Estrella,L., Depicker,A., Van Montagu,M. and Schell,J. (1983) *Nature*, 303, 209-213.

Herrera-Estrella,L., Van den Broeck,G., Maenhaut,R., Van Montagu,M., Schell, J., Timko,M. and Cashmere,A. (1984) *Nature*, 310, 115-120.

Holowach,L.P. (1981) Ph.D. Thesis, Cornell University, Ithaca, NY.

Horsch,R.B., Fry,J.F., Hoffmann,N.L., Eichholtz,D., Rogers,S.G. and Fraley, R.T. (1985) *Science (Wash.)*, 227, 1229-1231.

Koshiyama,I. (1983) in Gottschalk,W. and Müller,H.P. (eds.), *Seed Proteins Biochemistry, Genetics, Nutritive Value*, M.Nijhoff and W.Junk, Boston, pp. 427-450.

Koziel,M.G., Adams,T.L., Hozlet,M.A., Damm,D., Miller,J., Dahlbeck,D., Jayne,S. and Staskawicz,B.J. (1984) *J. Mol. Appl. Gen.*, 2, 549-562.

Ladin,B.F., Doyle,J.J. and Beachy,R.N. (1983) *J. Mol. Appl. Gen.*, 2, 372-380.

Laemmli,U.K. (1970) *Nature*, 227, 680-685.

Lane,B.G. and Tumaitis-Kennedy,T.D. (1981) *Eur. J. Biochem.*, 114, 457-463.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.

Meinke,D.W., Chen,J. and Beachy,R.N. (1981) *Planta*, 153, 130-139.

O'Farrell,P.H. (1975) *J. Biol. Chem.*, 250, 4007-4021.

Otten,L. and Schilperoort,R. (1978) *Biochim. Biophys. Acta*, 527, 497-500.

Schuler,M.A., Schmitt,E.S. and Beachy,R.N. (1982) *Nucleic Acids Res.*, 10, 8225-8244.

Sengupta-Gopalan,C., Reichert,N.A., Barker,R.F., Hall,T.C. and Kemp,J.D. (1985) *Proc. Natl. Acad. Sci. USA*, 82, 3320-3324.

Shatuck-Eidens,D.M. and Beachy,R.N. (1985) *Plant Physiol.*, 78, 895-898.

Symington,J., Green,M. and Brackman,K. (1981) *Proc. Natl. Acad. Sci. USA*, 78, 177-181.

Thanh,V.H. and Shibasaki,K. (1978) *J. Agric. Food Chem.*, 26, 692-695.

Yoo,B.Y. and Chrispeels,M.J. (1981) *Protoplasma*, 103, 201-204.

Received on 22 August 1985

LOCKE

Nucleotide sequence of the *dapA* gene from *Corynebacterium glutamicum*

S.Bonnassie, J.Oreglia and A.M.Sicard
Laboratoire de Génétique, Centre de Recherche de Biochimie et Génétique Cellulaires du CNRS, 31062 Toulouse Cedex, France

Submitted October 3, 1990 EMBL accession no. X53993

The *dapA* gene from *Corynebacterium glutamicum* encoding the L-2,3-dihydrodipicolinate synthetase was cloned by complementation of *Escherichia coli* auxotrophs with genomic DNA from *C. glutamicum* AS019 (1). We present here the nucleotide sequence of the *dapA* gene, which was obtained from the 1.4 kb *Bam*HI, *Sal*I, fragment subcloned into the corresponding sites of the plasmid pBR322. Our results indicate that the *dapA* gene consist of 903 nucleotides encoding a protein of 301 amino-acid residues.

REFERENCE

1. Yeh,P., Sicard,A.M. and Sinskey,A.J. (1988) *Mol. Gen. Genet.* **212**, 105–111.

```
TCTTATCGGAATGTGGCTTGGGCGATTGTTATGCAAAAGTTGTTAGGTTTTTTGCGGGGTTGTTTAACCCCCAAA  75
TGAGGGAAGAAGGTAACCTTGAACTCTATGAGCACAGGTTTAACAGCTAAGACCGGAGTAGAGCACTTCGGCACC 150
                          MetSerThrGlyLeuThrAlaLysThrGlyValGluHisPheGlyThr
                                                ClaI
GTTGGAGTAGCAATGGTTACTCCATTCACGGAATCCGGAGACATCGATATCGCTGCTGGCCGCGAAGTCGCGGCT 225
ValGlyValAlaMetValThrProPheThrGluSerGlyAspIleAspIleAlaAlaGlyArgGluValAlaAla

TATTTGGTTGATAAGGGCTTGGATTCTTTGGTTCTCGCGGGCACCACTGGTGAATCCCCAACGACAACCGCCGCT 300
TyrLeuValAspLysGlyLeuAspSerLeuValLeuAlaGlyThrThrGlyGluSerProThrThrThrAlaAla

GAAAAACTAGAACTGCTCAAGGCCGTTCGTGAGGAAGTTGGGGATCGGGCGAAGCTCATCGCCGGTGTCGGAACC 375
GluLysLeuGluLeuLeuLysAlaValArgGluGluValGlyAspArgAlaLysLeuIleAlaGlyValGlyThr

AACAACACGCGGACATCTGTGGAACTTGCGGAAGCTGCTGCTTCTGCTGGCGCAGACGGCCTTTTAGTTGTAACT 450
AsnAsnThrArgThrSerValGluLeuAlaGluAlaAlaAlaSerAlaGlyAlaAspGlyLeuLeuValValThr
                                                      PstI
CCTTATTACTCCAAGCCGAGCCAAGAGGGATTGCTGGCGCACTTCGGTGCAATTGCTGCAGCAACAGAGGTTCCA 525
ProTyrTyrSerLysProSerGlnGluGlyLeuLeuAlaHisPheGlyAlaIleAlaAlaAlaThrGluValPro

ATTTGTCTCTATGACATTCCTGGTCGGTCAGGTATTCCAATTGAGTCTGATACCATGAGACGCCTGAGTGAATTA 600
IleCysLeuTyrAspIleProGlyArgSerGlyIleProIleGluSerAspThrMetArgArgLeuSerGluLeu
                                                   BclI
CCTACGATTTTGGCGGTCAAGGACGCCAAGGGTGACCTCGTTGCAGCCACGTCATTGATCAAAGAAACGGGACTT 675
ProThrIleLeuAlaValLysAspAlaLysGlyAspLeuValAlaAlaThrSerLeuIleLysGluThrGlyLeu

GCCTGGTATTCAGGCGATGACCCACTAAACCTTGTTTGGCTTGCTTTGGGCGGATCAGGTTTCATTTCCGTAATT 750
AlaTrpTyrSerGlyAspAspProLeuAsnLeuValTrpLeuAlaLeuGlyGlySerGlyPheIleSerValIle
                                          HindIII
GGACATGCAGCCCCCACAGCATTACGTGAGTTGTACACAAGCTTCGAGGAAGGCGACCTCGTCCGTGCGCGGGAA 825
GlyHisAlaAlaProThrAlaLeuArgGluLeuTyrThrSerPheGluGluGlyAspLeuValArgAlaArgGlu

ATCAACGCCAAACTATCACCGCTGGTAGCTGCCCAAGGTCGCTTGGGTGGAGTCAGCTTGGCAAAAGCTGCTTCG 900
IleAsnAlaLysLeuSerProLeuValAlaAlaGlnGlyArgLeuGlyGlyValSerLeuAlaLysAlaAlaSer
  PstI
CGTCTGCAGGGCATCAACGTAGGAGATCCTCGACTTCCAATTATGGCTCCAAATGAGCAGGAACTTGAGGCTCTC 975
ArgLeuGlnGlyIleAsnValGlyAspProArgLeuProIleMetAlaProAsnGluGlnGluLeuGluAlaLeu

CGAGAAGACATGAAAAAAGCTGGAGTTCTATAAATATGAATGATTCCCGAAATCGCGGCCGGAAGGTTACCCGCA 1050
ArgGluAspMetLysLysAlaGlyValLeu---
AGGCGGCCCACCAGAAGCTGGTCAGGAAAACCATCTGGATACCCCTGTCTTTCAGGCACCAGATGCTTCCTCTAA 1125
```

pea (Polhill, 1981). Thus, nucleotide or amino acid sequences shared between pea and either one of the other two species presumably represent the ancestral state, although it is possible that such shared features could reflect the existence of functionally diverged classes of genes within the multigene families of each species (Shah *et al.*, 1983). However, this appears unlikely for the 7 S storage protein gene families because subunit sequences are, in all cases, more similar within a species than between species (Hall *et al.*, 1983; Lycett *et al.*, 1983; Schuler *et al.*, 1982; Slightom *et al.*, 1985).[2]

Our results extend previous observations of the similarities among the phaseolin and β-conglycinin genes (Schuler *et al.*, 1983). This comparison shows that the major difference in size between the α′- and α-subunits of β-conglycinin and the smaller subunits of other 7 S storage proteins from soybean, common bean, and pea is attributable to a large insertion in the first exon of the genes that encode the larger subunits. Comparative analyses of both coding and noncoding DNA regions have been useful in distinguishing nucleotide and amino acid sequences that have diverged from those which have remained conserved throughout the evolution of these species. The identification of conserved sequence elements in the 5′-flanking DNA of the β-type phaseolin and α′-subunit of β-conglycinin genes should be helpful in locating DNA elements that can be tested for their ability to control organ specificity and developmental expression of these genes. Also, the identification of conserved and diverged regions of protein-encoding DNA may be useful in targeting regions of the proteins that will accept amino acid modifications.

EXPERIMENTAL PROCEDURES

Gene Isolation and DNA Sequencing—Isolation of the Charon 24A clone 177.4 containing a complete *P. vulgaris* β-type phaseolin storage protein gene has been described (Sun *et al.*, 1981). Most of the phaseolin gene sequences used for comparisons were obtained using phaseolin gene subclones AG-pPvPh3.0 (*Eco*RI-*Bam*HI fragment) and AG-pPvPh3.8 (*Bgl*II-*Bam*HI fragment) which have been reported previously by Slightom *et al.* (1983). For the purpose of the comparison, we obtained an additional 790 bp of nucleotide sequence, extending 5′ from previously published data, using the subclone pPvPh 177.4-E1.6 containing the 5′ adjacent 1.6-kDa *Eco*RI fragment (Sun *et al.*, 1981). DNA sequences were determined using the chemical method (described by Maxam and Gilbert (1980)) procedure on 5′ end-labeled DNA fragments using restriction endonucleases *Bgl*II, *Nco*I, and *Eco*RI (Slightom *et al.*, 1983).[3] The isolation and partial nucleotide sequences of the gene encoding the α′-subunit of β-conglycinin from *G. max*, Gmg 17.1, have been reported by Schuler *et al.* (1982). The sequence of 850 bp 5′ of the α′-subunit gene was determined by first deleting nucleotides from the 5′ end with the exonuclease *Bal*31 followed by DNA sequence analysis according to the methods of Smith and Calvo (1980). Preparation and analysis of the *Bal*31 deletions will be described elsewhere.

Nucleotide and Amino Acid Sequence Comparisons—Initial alignments of nucleotide sequences were done using the computer program NUCALN modified from the original Wilbur and Lipman (1983) algorithm by M. Boguski (Washington University Computer Center). Coding and noncoding regions were aligned separately, with higher gap penalties (3 or 5 *versus* 1) used in coding regions so as to maintain codon integrity. Nucleotide sequence alignments were also performed using the computer program GAP supplied by the University of Wisconsin Genetics Computer Group (Devereux *et al.*, 1984). Alignments subsequently were modified manually in regions where duplications or other discrete evolutionary events could be inferred. Other computer programs were obtained from Compugene, Inc. (Creve Couer, MO) and included translation of nucleic acid sequences, codon usage, dot matrix analysis, and nucleic acid and protein secondary structure predictions. A program for computing corrected levels of silent *versus* replacement nucleotide substitutions following the method of Perler *et al.* (1980) was developed by V. Zenger. The program used to construct hydrophobicity index plots is based on the program developed by Kyte and Doolittle (1982).

RESULTS[4]

Nucleotide Sequence Alignment and Gene Organizations—The nucleotide sequence alignment of the β-type phaseolin gene (Pvu β; 2880 bp) against the gene encoding the α′-subunit of β-conglycinin (Gma α′; 3636 bp) is shown in Fig. 1. The aligned sequences begin approximately 900 bp 5′ of their common cap sites and extend through the structural gene regions to their respective polyadenylated nucleotide. Alignment of these sequences including the placement of gaps (shown as *asterisks* in Fig. 1) was relatively straightforward due to the similar structures of these two genes, both of which have six exons and five introns and share nucleotide sequence elements in their 5′- and 3′-untranslated regions. The nucleotide sequence numbering system is positive starting at the cap sites and extending in the 3′ direction, while sequences extending 5′ of the cap sites are numbered negatively.

The comparison presented in Fig. 1 reveals considerable homology at both the nucleotide and amino acid sequence levels. It is apparent that homologous sequence regions are interspersed with diverged regions, both in coding and noncoding DNAs. Using both natural (introns, exons, and untranslated regions) and arbitrary (in the 5′-flanking DNA) boundaries, we calculated the degree of apparent nucleotide divergence for 18 regions (Table 1). The overall corrected divergence (Hayashida and Miyata, 1983) between these genes is about 41%, with exons and introns showing about the same degree of divergence, 43 and 37%, respectively. The lack of divergence between intron sequences (which is not expected) is due, in part, to our method of analysis, since gaps in intron sequences are not heavily weighted (see "Experimental Procedures"). In addition, the small size of these introns (the largest is 203 bp) may reduce their degree of divergence; this has been observed for the small (about 120 bp) orthologous introns of mammalian globin genes which appear to diverge less than the larger (about 850 bp) introns contained in the same genes (Chang and Slightom, 1984; Efstratiadis *et al.*, 1980; Goodman *et al.*, 1984; Slightom *et al.*, 1985). Small orthologous related intron sequences may diverge less because of biological constraints imposed by the minimal size required by the splicing mechanism (Rautmann *et al.*, 1984; Wieringa *et al.*, 1984).

5′-Flanking DNA: A Search for Possible cis-Acting Regulatory Elements—Several cis-acting regulatory elements have been identified in the 5′-flanking DNA of eukaryotic genes. The CCAAT and TATA elements are located −77 and −31 bp, respectively, 5′ from the mRNA cap site of many eukaryotic genes (Efstratiadis *et al.*, 1980). Functional experiments have shown that the TATA element is important for RNA polymerase II binding (McKnight and Kingsbury, 1982). In addition, functional investigations of mammalian virus genes have led to the identification of another cis-acting regulatory element known as the enhancer element which has a consensus sequence of $\overset{\text{TTT}}{\underset{\text{GTGGAAAG}}{}}$ (Gruss, 1984). Identical CCAAT and TATA elements are shared in the 5′-flanking DNAs of

[2] J. Harada, personal communication.

[3] J. L. Slightom, *et al.*, manuscript in preparation.

[4] Tables 3–5 and Fig. 4 are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 8M5-3318, cite the authors, and include a check or money order for $2.00 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

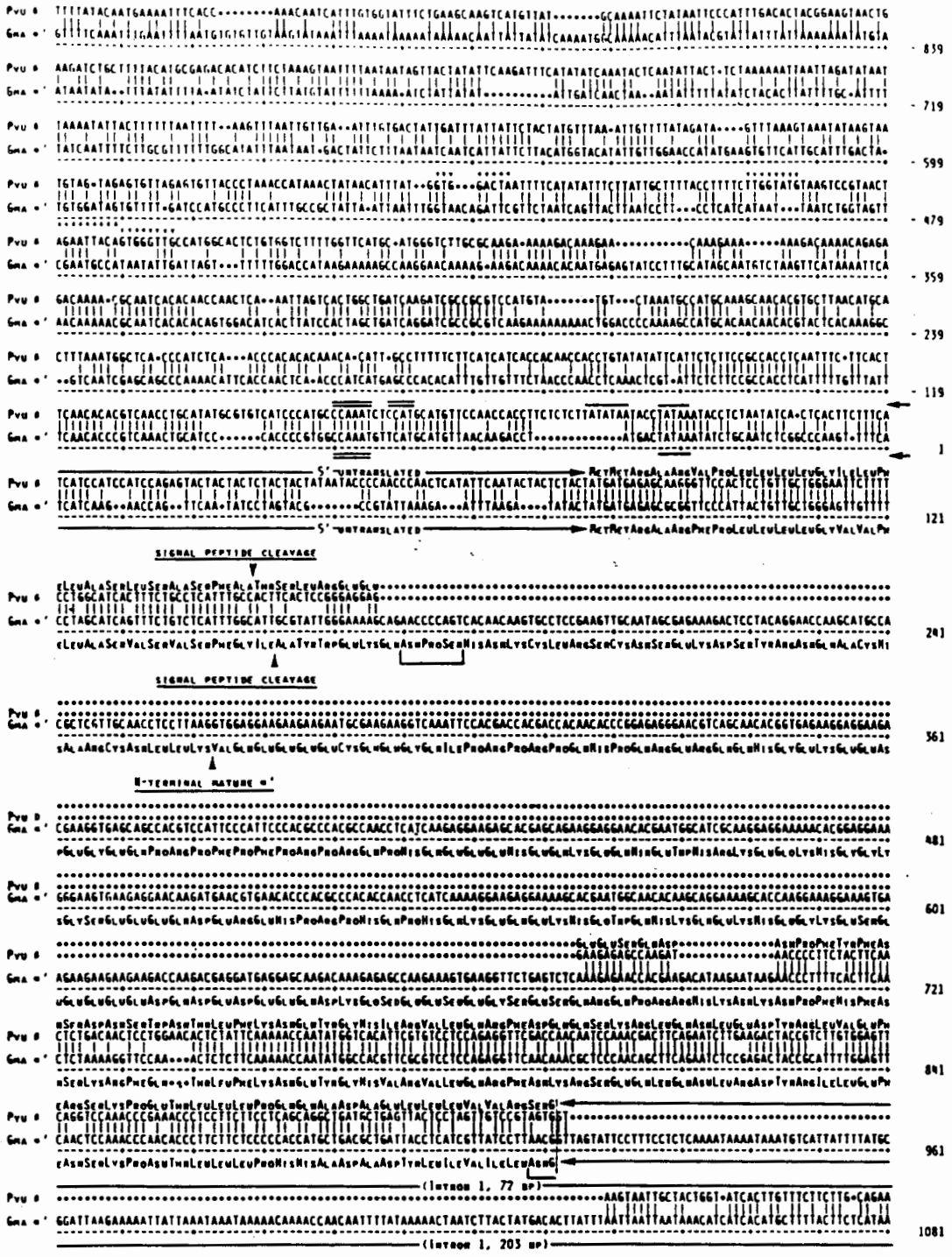
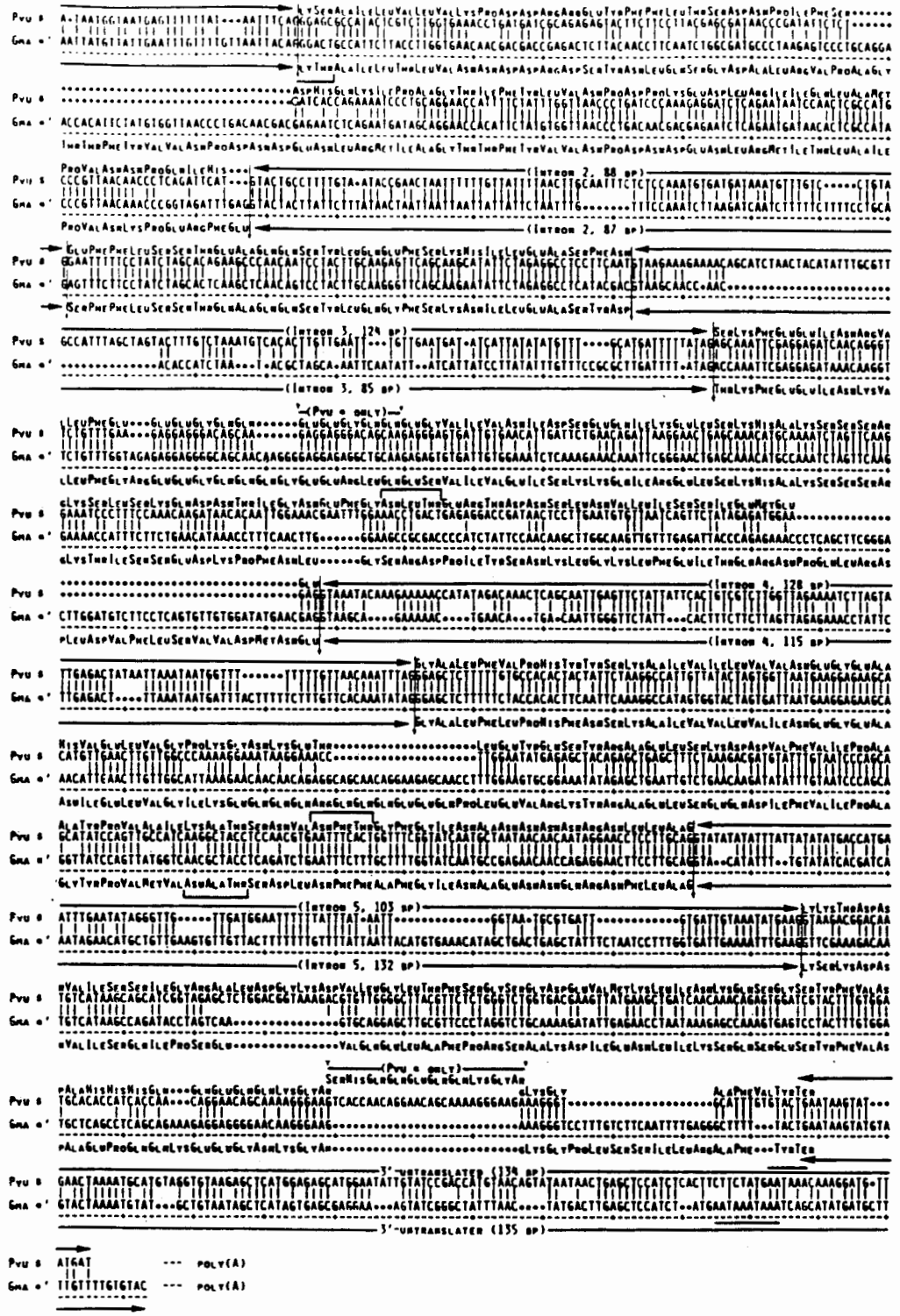


FIG. 1. **Nucleotide sequence comparison of 7 S storage protein gene isolated from *G. max* and *P. vulgaris*.** Nucleotide sequences determined for the *P. vulgaris* 7 S storage protein gene (gene contained in clone λ177.4 (Sun *et al.*, 1981) encodes a β-type phaseolin polypeptide (Slightom *et al.*, 1981)) is presented on the top line (Pvu β), and it is compared to the sequence of the *G. max* 7 S gene contained in λ clone Gma α′ 17.1 (Schuler *et al.*, 1982) which encodes an α′-type β-conglycinin polypeptide. The nucleotide sequence numbering system used was set by the overall alignment between these genes. Complete sequences for both genes are presented starting 889 bp (for Pvu β) and 905 bp (for Gma α′) 5′ from their respective mRNA-capped nucleotides and extend 3′ to the nucleotide used for poly(A) attachment, 1990 bp (for Pvu β) and 2732 bp (for Gma α′). The presence of *asterisks* in nucleotide and amino acid sequences indicate gaps which were used to maximize identities. Different gap weights were used depending upon whether gaps were needed in coding regions (gap = 5 substitutions) or noncoding regions (gap = 1.1 substitutions). Nucleotide sequence alignments were done using computer programs NUCALN and GAP. *Vertical lines* between nucleotides indicate base matches. Nucleotide sequences which may have biological importance to the expression of these genes are indicated; TATAA and poly(A) addition elements are *once underlined* or *overlined*, and CCAAT elements are *twice underlined* or *overlined*. Also nucleotide sequences which match the consensus sequence for mammalian enhancer elements (Gruss, 1984) are indicated by *arrows* pointing to nucleotides above (Pvu β) or below (Gma α′) the respective sequence lines. *Vertical arrows* between sequence lines show the position of exon-intron boundaries. Translation initiation can occur at either of two Met residues at the beginning of each gene, and the termination codons are designated *TER*. Protein signal peptide and



secondary processing (Gma α' only) cleavage positions are indicated, and possible *N*-glycoside attachment sites are shown as *boxed* amino acids in the respective amino acids sequence lines. Nucleotide and amino acid sequences that correspond to the duplicated regions responsible for the *P. vulgaris* α-type polypeptides (Hall *et al.*, 1983; Slightom *et al.*, 1985) are shown.

Gma α' and Pvu β (Fig. 1). The multiple CCAAT and TATA elements found in Pvu β are believed to be responsible for the numerous cap sites found in phaseolin mRNAs (Hall *et al.*, 1983; Slightom *et al.*, 1985b). The sequences surrounding the TATA elements of both genes are nearly identical to the "plant consensus," while sequences surrounding the CCAAT elements are more similar to the mammalian consensus sequence (Table 2).

Whether plant genes possess elements that are similar in function to mammalian enhancers is currently unknown. However, we searched the 5'-flanking DNAs of these two seed protein genes for sequences matching the consensus mammalian enhancer element and have located an identical match in the Gma α' sequence 553 bp upstream from the mRNA cap site (position −590 in Fig. 1). The Pvu β sequence has several closely matching sequences located between position

TABLE 1

Divergence in and around 7 S storage protein genes from G. max and P. vulgaris

Assignments of gene regions and nucleotide position are from Fig. 1. Corrected percentage sequence divergences were calculated as described by Hayashida and Miyata (1983). Gaps, regardless of length and location, were counted as one substitution. Calculations for Exon 4 and 6 use the nucleotide sequence from the Pvu β gene and do not include the Pvu α gene sequences shown in Fig. 1.

Structural region compared	Nucleotide positions	Number of gaps	Match/ compared	Corrected % divergence
			bp	
1. 5′ Flanking	−958 to −358	28	304/549	67.8
2. 5′ Flanking	−357 to −248	4	75/101	31.5
3. 5′ Flanking	−247 to −150	7	46/96	87.3
4. 5′ Flanking	−149 to −52	2	73/94	26.5
5. 5′ Flanking	−51 to −1	3	26/39	44.1
6. 5′ Untranslated	1 to 77	6	39/64	55.2
7. Exon 1	78 to 911	3	237/311	28.6
8. Intron 1	912 to 1115	5	52/77	42.5
9. Exon 2	1116 to 1348	2	123/193	49.6
10. Intron 2	1349 to 1442	3	60/84	36.0
11. Exon 3	1443 to 1523		67/81	19.6
12. Intron 3	1524 to 1655	10	61/87	38.1
13. Exon 4	1656 to 1938	4	143/229	52.1
14. Intron 4	1939 to 2082	7	91/116	25.4
15. Exon 5	2083 to 2372	1	198/260	28.7
16. Intron 5	2373 to 2508	7	74/106	38.6
17. Exon 6	2509 to 2747	4	115/176	46.5
18. 3′ Untranslated	2748 to 2886	6	92/129	36.1

TABLE 2

Consensus sequences elements and their comparison with similar elements from Gma α′ and Pvu β genes

TATA sequences:	
	C T C
Plant consensus (34):	TGTATAAA$_{1-3}$TA
Gma α′:	ACTATAAA TA
Pvu β:	CCTATAAA TA
2. CCAAT sequences:	
	C G CC
Plant consensus (34):	TA$_{2-3}$TNGA$_{2-4}$TT
Gma α′:	GGCCAA TCTGTTCAT
Pvu β:	GGCCAA TCTCTCCAT
Animal consensus (15):	GGCCAA TCT
3. Enhancer sequences:	T
	TTT
Mammalian consensus (55):	GTGGAAAG
Gma α′:	GTGGATAG
Pvu β:	GTGGACTA
	TTGGTATG
	GTGGGTTG
4. Translation initiation:	
	C
Plant consensus (34):	GAANNATGG
Gma α′ (first AUG):	ATACTATGA
Gma α′ (second AUG):	CTATGATGA
Pvu β (first AUG):	CTACTATGA
Pvu β (second AUG):	CTATGATGA

−550 and −470 (Fig. 1; Table 2). These putative plant enhancer elements are located much farther 5′ from the cap site of these genes than are functional mammalian enhancers, which generally are found within 100 bp 5′ of the mRNA cap sites (Gruss, 1984). However, because mammalian enhancers can function in both orientations (Moreau *et al.*, 1981; Weiher *et* 1983), we scanned the opposite strands and found enhan--like elements at positions −54 and −206 of Pvu β (shown as CCAACCAC and CAACCCAC, respectively, Fig. 1) and at position −87 of Gma α′ (CATCCCAC, Fig. 1). The ability of these sequences to regulate expression of the Gma α′ gene is being tested using deletion mutants.

Because the accumulation of the seed storage proteins derived from these genes is, in both cases, subject to similar tissue and developmental regulation, it is conceivable that these genes could share conserved cis-acting regulatory elements in their 5′-flanking DNAs. Experiments involving the transfer of the Pvu β gene and Gma genes into genomes of tobacco and petunia (respectively) via *A. tumefaciens* vector systems have been completed. These experiments support the presence of regulatory sequences within 800 bp of the cap site of Pvu β (Sengupta-Gopalan *et al.*, 1985) and 200 bp of the Gma α′ cap site (Beachy *et al.*, 1985).[5] Fig. 1 shows that there are several regions of homology in the 5′-flanking DNAs, between positions −357 to −248 and −149 to −1. These two homologous regions are separated by sequences that share little homology (Fig. 1, Table 1). Within positions −357 to −248 there are three sequence clusters that show a very high degree of homology, positions −357 to −338 (19/20 matches), −330 to −293 (24/26 matches), and −273 to −248 (22/26 matches). In the second region, four sequence clusters show a high degree of homology, positions −149 to −127 (22/23 matches), −119 to −98 (20/22 matches), −89 to −61 (24/29 matches), and −33 to −13 (14/17 matches). Any or all of these conserved sequence regions may provide information responsible for controlling the tissue and developmental expression of these genes. Recently, a light-inducible control sequence element for the pea ribulose 1,5-bisphosphate carboxylase gene (Morelli *et al.*, 1985) was found near the cap site of this gene.

We have extended this comparison of 5′-flanking DNAs to include the sequences from another gene family known to be subject to similar tissue and developmental controls, the *P. vulgaris* seed lectin genes. Our comparison (data not shown) of 300 bp of flanking DNA sequences from two seed lectin genes (Hoffman and Donaldson, 1985) failed to reveal any extensive sequence homologies. One of two possibilities may account for this observation. First, shared seed-specific regulatory elements for these genes might be located further upstream and thus were not identified. Second, *P. vulgaris* lectin and phaseolin genes, even though appearing to be controlled by a similar mechanism, are not and thus do not share common cis-acting regulatory sequences. Deletion and site-specific mutant constructions of the Gma α′ gene are being used to test the functionality of the putative regulatory sequences identified above.

5′- and 3′-Untranslated Regions—The 5′-untranslated regions of Gma α′ and Pvu β are the most diverged (55%) structural regions of these genes (Table 1), sharing homology only around the cap site and site of translation initiation. Assuming that the first methionine codon is used for translation initiation, the sizes of the 5′-untranslated regions differ significantly, 61 bp for Gma α′ and 80 bp for Pvu β. Thus, it appears that the 5′-untranslated regions of these storage protein genes can tolerate a considerable amount of change without affecting function. In contrast, the 3′-untranslated regions appear to be subject to more stringent evolutionary constraints. Both are 135 bp in length and appear to have diverged to about the same amount (36%) as have most of the coding sequences. These 3′-untranslated sequences share a high degree of homology around the translation termination codon but have lower homology near the poly(A) addition signals. A cluster of 12 identical nucleotides, located just 5′ of the poly(A) addition signals, is the last stretch of homology between these 3′-untranslated regions. It is not known whether these 12 shared nucleotides play a role in the poly(A) addition mechanism.

[5] Z.-L. Chen, M. A. Schuler, and R. N. Beachy, manuscript in preparation.

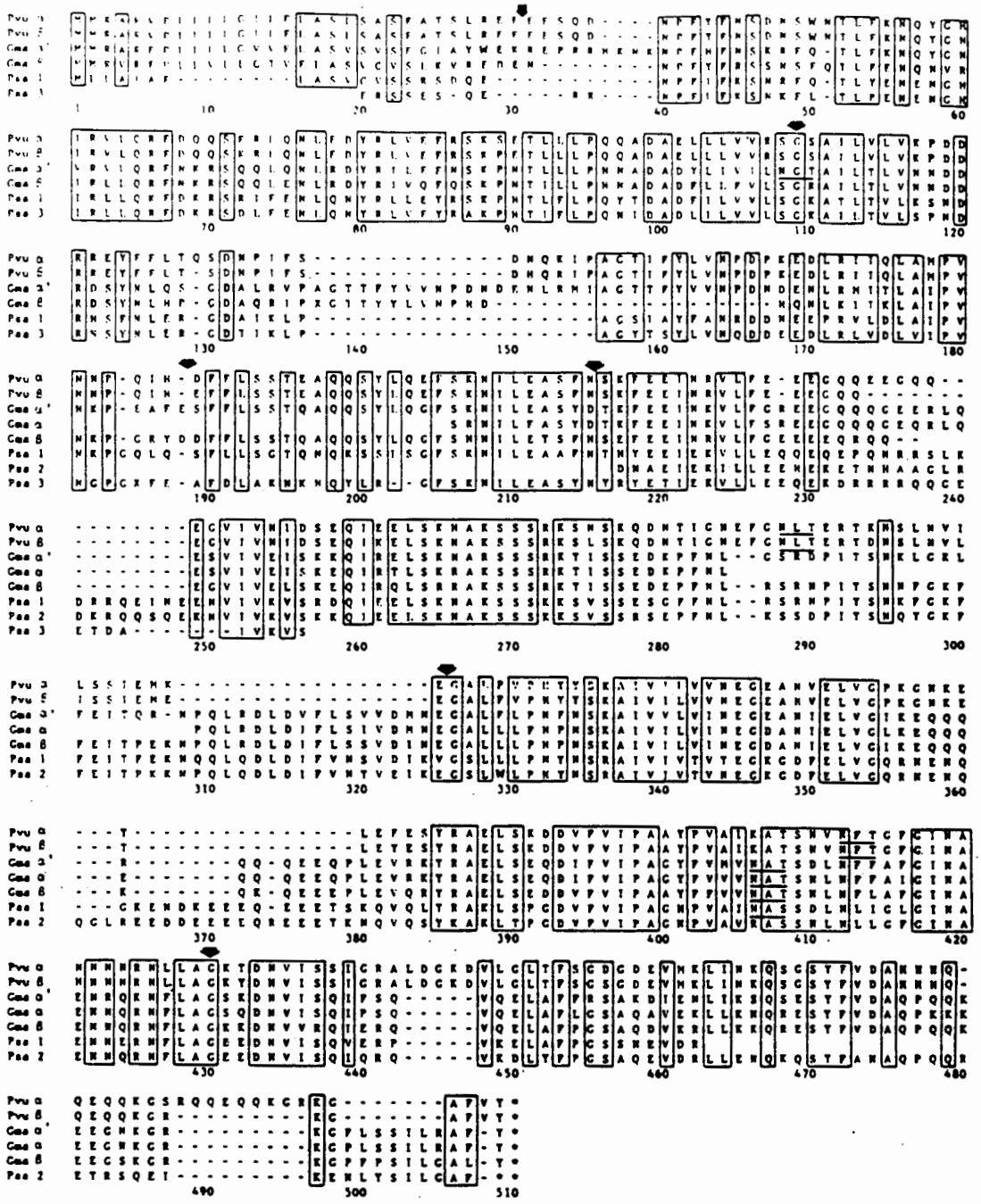


FIG. 2. **Comparison of 7 S storage protein amino acid sequences from three legumes.** Pvu α, from *P. vulgaris*, is predicted from DNA sequence data (Hall *et al.*, 1983; Slightom *et al.*, 1983); Pvu β and Gma α' are predicted from the sequence given in Fig. 1; Gma α is from Schuler *et al.* (1982). Psa 1 and Psa 2 are predicted from the sequences of *P. sativum* cDNA clones C47 and C50 of Lycett *et al.* (1983); Psa 3 is predicted by convicilin cDNA clone pCD59 (Casey *et al.*, 1984). The sequences have been aligned with gaps (represented by *dashes*) included where required to give maximal homologies. The large insertion in exon 1 of Gma α' is not shown; its position is represented by an *arrow*. Intron positions in Pvu β and Gma α' are marked by *numbered arrows*; the position of introns in the other sequences is unknown. Potential glycosylation sites are *underlined*. Amino acid residues shared by at least one storage protein in each of the three genera are *boxed*; conservative substitutions (V = I = L; E = D; Y = F; S = T) are considered shared residues.

Alignment of Exon Regions—Despite weighting against gaps in coding regions, gaps were needed in five of the six exons to maximize the identity of these two genes. The degree of homology found between these exons varies considerably; even within individual exons we find regions of high homology (at both nucleotide and amino acid sequence levels) separated by regions which show complete divergence or loss of genetic information (due to deletions or insertions). These findings suggest that evolutionary conservation of these storage protein genes may be operating on units smaller than the exons themselves, perhaps due to the presence of important nucleotide sequence elements or importance of amino acid structures (see below). The analysis of both nucleotide and amino acid sequence divergence and of the gaps contained in these exons is more meaningful when associated with additional sequence information provided by evolutionary outgroups. Published nucleotide and derived amino acid sequences of the 7 S storage protein subunits from pea vicilin cDNA clones c47 and c50

cett *et al.*, 1983), referred to subsequently as Psa 1 and 2, respectively, a convicilin cDNA clone pCD59 (Casey *et al.*, 1984), referred to as Psa 3, provide information which is useful in determining the nature and direction of evolutionary events. These pea gene nucleotide sequences have been aligned with those from Gma α' and Pvu β; however, because these comparisons are confined mainly to coding sequence, only the resulting amino acid alignment is presented in Fig. 2. This amino acid alignment also includes, from soybean, the partial sequence of the α-subunit, Gma α, (Schuler *et al.*, 1982) and the recently sequenced β-subunit, Gma β^6 of β-conglycinin, as well as from common bean, the sequence of the α-type phaseolin subunit, Pvu α, which corresponds to cDNA clone 6 (Hall *et al.*, 1983; Slightom *et al.*, 1985).

Exon 1; Gma α' Contains a Unique Insertion—The alignment of exon 1 shown in Fig. 1 is unique because, while it shows relatively low divergence between Gma α' and Pvu β (29%, Table 1), 522 bp of the Gma α' exon 1 nucleotide sequence is not present in Pvu β. The difference between these exon 1 sequences is almost twice as large as exon 1 of β. As a result of our alignment method, we have represented the difference as two gaps (Fig. 1). The first gap of 508 bp introduces a shift in reading frame for five amino acids which is then corrected by a second gap of 14 bp. This large insertion could also be represented as a single gap at the minimal expense of 6-bp matches. The 522-bp insert is unique as it does not appear either completely or partially in any other location of the other 7 S storage protein genes examined (except the α-subunit, see below); thus, it may well have been derived from a source outside of these storage protein genes.

nucleotide sequence at the core of this large insert is composed of numerous short, perfect, or near-perfect repeats as shown by dot matrix analysis (Fig. 3), and it contains several open-reading frames of unknown functionality on both strands (data not shown). These two findings are perhaps not unrelated because multiple open-reading frames may be a common and important evolutionary property of repeated sequences (Ohno, 1984).

The large insertion adds 174 amino acid residues or about 20,000 to the molecular weight of Gma α' polypeptides, which accounts for most of the molecular weight difference observed between the Gma α' and Pvu β polypeptides. This insert, or diverged counterpart, is also present in the 72-kDa α-subunit of β-conglycinin.[7] Because this insert is not found in the pea vicilin gene, Psa 1, or convicilin gene, Psa 3, nor in the soybean β-conglycinin subunit, Gma β, it almost certainly represents an insertion event into an ancestral α'-subunit gene of the soybean β-conglycinin gene family. This event may not be unique to soybean. An immunological survey of a large number of legume genera has shown that 7 S polypeptides in excess of 65 kDa are common only to the group of genera closely related to soybean (Doyle *et al.*, 1985). This suggests that the large insertion in the first exon occurred in a progenitor of this entire group of legumes, a hypothesis that can be tested by isolating storage protein genes from representative members of those genera. The ancestral state of Gma α' may indeed be represented by the smaller Gma β subunit of β-conglycinin. The amino acid alignment presented in Fig. 2 supports this possibility because Gma β encodes a p peptide almost identical in size to that encoded by pea n and convicilin genes, and they all share amino acid homologies. These observations also suggest that when the

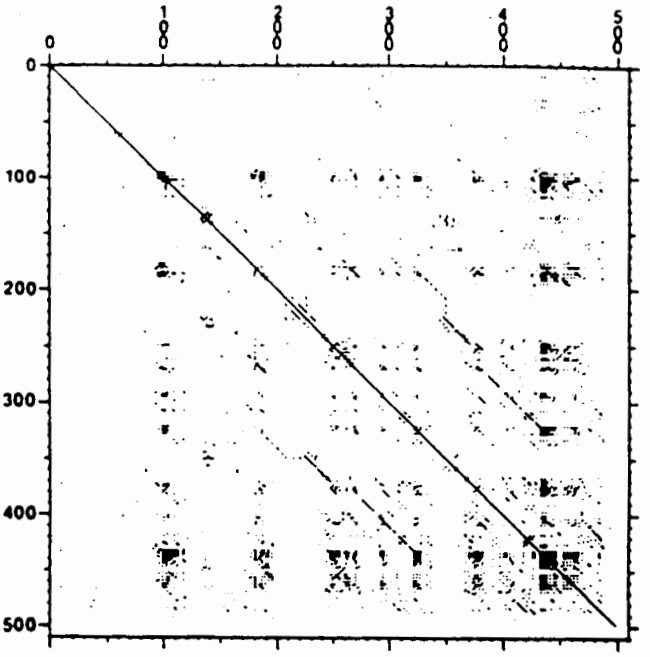


FIG. 3. **Dot matrix comparison of the 509-nucleotide insertion in the first exon of the Gma α'-subunit gene.** This insertion is responsible for the increase in size of the α' and α subunits of *G. max* relative to the Pvu α and β proteins of *P. vulgaris*. In this comparison the first nucleotide was marked if 10 of 12 nucleotides compared were identical.

insertion event occurred, the genes encoding the soybean 7 S proteins were already a multigene family, which through postinsertion duplications and sequence divergence have formed the present multigene subfamilies of β-conglycinin. We found no evidence of transposon-like sequence elements (flanking direct or indirect repeats) near the boundaries of this large insert, which suggests that if a transposon-like mechanism is responsible for the insertion event, the transposon-like sequences have been lost due to evolutionary mutations. Thus, if this large insert once was an active mobile genetic element, it is probably no longer active.

Comparison of portions of the insert region from Gma α and α' genes shows a relatively high degree of nucleotide and amino acid sequence homology which suggests that these inserts have achieved both functional and evolutionary stability. Surprisingly, this stability has been achieved even though the amino acid compositions of these inserts are quite different from that present in the subunits of other 7 S proteins, being particularly rich in glutamate residues (Table 3). The stability of inserts that encode acidic amino acid residues appears to be quite common in legume storage proteins (Argos *et al.*, 1985). If the amino acids encoded by the insert are excluded, the composition of Gma α' closely resembles that of Pvu β (Table 3). The similarity of the mature Gma α' and Pvu β polypeptides is also evident from their similar codon usage (Table 4). Where any preferential codon use is apparent, the same codons are preferred by both genes, such as Asn: AAC >AAU, or Val: GUU, GUG >GUC, GUA.

The insert in Gma α' is also of interest because it does contain the NH_2 terminus of the mature Gma α' polypeptide, as deduced by amino acid analysis (Coates *et al.*, 1985). The NH_2 terminus of the Gma α' polypeptide does not immediately follow the signal peptide but instead is located 27 amino acid residues downstream (Fig. 1), within the large insert. Thus, the Gma α' polypeptide is subject to a second cleavage process that is not found for Pvu β polypeptide, because the NH_2 terminus of the latter polypeptide is located at the signal

[6] J. J. Harada, S. J. Barker, and R. B. Goldberg, manuscript in preparation.

[7] M. A. Schuler, F. Sebastiani, and R. N. Beachy, manuscript in preparation.

peptide cleavage site.[8] A biological explanation for the necessity for this second cleavage step in Gma α' is not apparent, but it could be due to constraints imposed by the presence of the additional 174 amino acids. The 27 amino acids involved in this second cleavage may, for example, need to be removed for proper packaging in protein storage bodies. Alternatively, the additional amino acids may be folded in such a way as to expose a proteolytically labile site. The significance of this cleavage process could be addressed by site-specific (amino acid-changing) substitutions designed to prevent cleavage. Preventing this second cleavage step in Gma α' would provide a simple method for increasing the number of sulfur-containing amino acids residues in this sulfur-deficient polypeptide because four of the 27 residues removed are cysteines (Fig. 1). However, blocking of this cleavage could dramatically change the secondary and tertiary structures of the protein. The resultant Gma α' polypeptide would also have a quite different NH_2-terminal amino acid composition than its present mature NH_2 terminus, which is rich in glutamine and similar to the NH_2 terminus of Pvu β.

One of the most conserved features of the Gma α', Gma β, and Pvu β (also Pvu α (Slightom *et al.*, 1985)) proteins is the signal peptide region (Figs. 1 and 2). Gma α' and Pvu β share 83 and 72% of their nucleotide and amino acid sequences, respectively. The signal peptide cleavage positions for both Gma α' and Pvu β appear to occur at similar locations; however, the cleavage for Pvu β is heterogeneous.[9] All of these signal peptides share extensive homology at their NH_2 termini, but homology is lacking around the expected cleavage sites. None of these signal peptides nor the very different signal peptide of vicilin, Psa 1, conform to the amino acid sequence pattern for signal peptide processing described by Von Heijne (1983). However, all of these signal peptides have hydrophobic cores (see hydropathic index, Fig. 4), and all but Psa 1 have the same charged amino acid residues near their NH_2 termini. The NH_2 termini encoded by all of these soybean and common bean storage protein genes contain tandem methionine codons. The nucleotide sequences surrounding both methionine codons do not match the plant consensus translation initiator sequence (Messing *et al.*, 1983), but they do share homology with the consensus sequence derived from the comparison of animal genes (Kozak, 1983) (Table 2).

Evolutionary Relationships as Determined by Other Exon Gaps—As stated above, gaps occur in all exons except exon 3, and from our alignment of amino acid sequences in Fig. 2, these gaps provide a significant amount of information concerning the evolution of these genes. In exon 2, Gma α' has a 57-bp direct repeat which results in the duplication of 19 amino acids. The alignment of this region with the other polypeptides (Fig. 2, positions 137–175) shows that this duplication has not occurred in the pea polypeptides and that much of it is not represented in either Gma β or Pvu β polypeptides. The partial alignments between Gma α' and Gma β and Pvu β are probably artifacts due to separate small insertions occurring in the latter two genes. This 57-bp duplication event appears, therefore, to have occurred only in the soybean, and after the separation of proto-Gma α' from the Gma β lineages. If this duplication occurred after divergence of the common ancestor of *Phaseolus* and *Glycine* from *Pisum*, but before the former two genera diverged from each other, this would require a second event in two species, *i.e.* the loss of this duplication from proto-Pvu β and proto-Gma β. Further analysis of the 7 S storage protein genes from other species closely related to *Glycine* and *Phaseolus* may be useful in resolving these two possibilities.

Exon 4 shows several regions of length variations, between positions 226–250 and 308–325 (Fig. 2). In the first region Gma α' and Pvu β differ considerably from Psa 1, 2, and 3, which themselves show considerable divergence. Reasons for divergence among the pea sequences are not apparent, but duplications appear responsible for variations in soybean and common bean. This region of the soybean Gma α' gene has an 11-bp indirect repeat (Fig. 1, positions 1693–1723) which is not present in Pvu β, while a 15-bp direct repeat is present only in Pvu α. The presence of repeated nucleotide sequence elements appears to correlate with evolutionary instability in these storage protein genes (see below). A second region in exon 4 (positions 308–325) is presumably a deletion in both Pvu α and β, since the sequences absent in these common bean genes are present in all the other 7 S storage proteins (Fig. 2). An additional small insertion (positions 285–286) encodes two amino acids that distinguish the Pvu α and β sequences from soybean and pea. Exon 5 contains a hot spot for length variation (positions 357–380) which is bounded on both sides by regions which share a high degree of homology. Instability in this region may also be due to the presence of several direct and indirect repeat sequences (see Fig. 1, positions 2186–2221). Exon 6 includes four regions of length variability (positions 443–447; 480; 488–496; and 499–505) all which appear to be due to insertion or deletion events that have occurred in *Phaseolus* (Fig. 2). The insertion at position 488–496 in Pvu α is due to a 27-bp duplication of a region from Pvu β which, along with the duplication at position 234–238 (in exon 4) characterizes the Pvu α gene family (Slightom *et al.*, 1985).

Only exon 3, the shortest exon (81 bp) shared by these 7 S storage protein genes, did not require gaps for alignment. At the nucleotide and amino acid sequences levels these exon 3 alignments share the following per cent homologies: 83 and 78% for Gma α' *versus* Pvu β; 79 and 67% for Gma α' *versus* Psa 1; and 65 and 59% for Pvu β *versus* Psa 1, respectively. These results are congruent with accepted phylogenies and taxonomic treatments of the legumes (Polhill, 1981) as they show that *Glycine* and *Phaseolus* are more closely related to each other than either is to *Pisum*.

Evolution of Gma α' and Pvu β Introns—Another of the interesting features deduced from the comparison of Gma α' and Pvu β nucleotide sequences is that both genes have 5 introns which interrupt the coding region at identical positions. The divergences calculated in Table 1 are not indicative of the considerable length divergence of the introns (see above). Intron 1 shows the largest length variation, 203 bp (Gma α') *versus* 72 bp (Pvu β), while intron 2 lengths are almost the same (87 *versus* 88 bp, respectively). All intron/exon splice junctions conform to the consensus plant sequences (Shah *et al.*, 1983; Slightom *et al.*, 1983), and all have DNA compositions which are A+T rich. Introns from Gma α' and Pvu β have average A+T percentages of 80.0 and 72.4%, respectively. Aside from the splice junctions and nucleotide compositions, these introns do not appear to share any common "core" elements which might play a functional role in the splicing mechanism.

Glycosylation of 7 S Storage Proteins—The 7 S seed storage proteins are glycoproteins, and all of the amino acid sequences compared contain at least one *N*-glycosyl recognition site (Asn-*X*-Thr or Ser) with the exception of Psa 2 (Fig. 2). The fact that the amino acid sequence of Psa 2 does not contain a *N*-glycosyl recognition site is puzzling; it could be due to an error in the data or Psa 2 could represent a pea vicilin variant that has lost the site. The number of *N*-glycosyl sites varies.

[8] H. Paaren, personal communication.

[9] Paaren, H., Slightom, J., Hall, T., Inglis, A., and Blagrove, R. (1986) *Phytochemistry*, in press.

ith three in Gma α′, two in Pvu β, and only one in Psa 1.

ne first glycosylation site in Gma α′ has the sequence Asn-Pro-Ser, a sequence that is rarely glycosylated (Mononen and Karjalainen, 1984), and is located in the 27-amino acid polypeptide that is removed by secondary processing (see above). Because this region of Gma α is removed, the glycosylation status of this site is unknown.

The second *N*-glycosyl site of Gma α′ is located between exons 1 and 2, while the third site is located in exon 5; the two sites for Pvu β occur in exons 4 and 5 (Figs. 1 and 2). The locations of the *N*-glycosyl sites in exon 5 of Gma α′ and Pvu β are not identical, but they are located in a hydrophobic domain (Fig. 4) at about the same distance from the COOH terminus of each polypeptide. The *N*-glycosyl site found in Psa 1 is located at exactly the same position in exon 5 as that of Gma α′ but differs from that of Gma α′ in having Ser in the third position instead of Thr. Gma α′ and Pvu β would share identical *N*-glycosyl sites in exon 5 except for a single base mutation C→G (see position 2302, Fig. 1). This would also be true for Psa 2 except for the apparent amino acid eplacement of Asn by Arg. Conservation of this *N*-glycosyl recognition site in exon 5 suggests that glycosylation near the COOH terminus of these polypeptide backbones may be important to the biological processing and/or storage of these polypeptides. The nonconserved *N*-glycosyl recognition sites found in Gma α′ and Pvu β are not only located in different positions, but also in quite different hydropathic environments (Fig. 4). The site in exon 4 of Pvu β is located in a hydrophilic domain while the site in between exons 1 and 2 f Gma α′ is located in a hydrophobic domain. The type of dropathic environment surrounding these *N*-glycosyl sites may influence the degree to which such sites are glycosylated; this remains to be demonstrated experimentally.

Properties of 7 S Proteins: Structure and Domains—Many of the changes that occur in the two genes involve the substitution of amino acids with similar physical properties, for example, isoleucine to valine. For this reason, two approaches to inferring physical properties of the proteins were adopted. Although some conservation of protein secondary structure is apparent from Chou and Fasman (1974) predictions (data not shown) and has also been reported by Argos *et al.* (1985), more striking similarities are apparent from comparisons of predicted hydropathies. Regions of very similar properties occur in both Pvu β and Gma α′, many of which are also shared with pea vicilin polypeptides (Fig. 4). The large insertion in conglycinin α′ is hydrophilic (Fig. 4) and represents a major departure from the patterns observed in the smaller subunits of the other genera. The overall pattern of hydropathy appears to transcend exon boundaries and is, in this respect, consistent with the patterns of amino acid conservation since both suggest that there is little relationship between exons and evolutionary domains (Fig. 2). These observations together with the patterns of predicted secondary structure for both vicilin and legumin proteins (see Argos *et al.*, 1985) suggest that these legume storage proteins have not been constructed by splicing together exons representing structural domains as has been suggested for other proteins (Gilbert, 1978).

Evolutionary Tempo and Mode—Divergence values presented in Table 1 were calculated by the method of Hayashida and Miyata (1983) to compensate for multiple substitutions at single positions in the nucleic acid sequence and thus are higher than those derived by simply comparing shared nucleotides. Homology values for the coding region have been further subdivided (Table 5) using the method of Perler *et al.* (1980) to approximate the actual degree of silent and replacement substitutions. A modification of this method takes into account the evolutionary preponderance of transition mutations over transversions, a phenomenon that is obscured over long evolutionary times (Brown *et al.*, 1982). Although this latter method is presumably a better approximation of actual divergences, we lack sufficient data on the actual frequencies of transition and transversion mutations. The observed rates of transition mutations in the different comparisons are, however, of interest. The comparison that should reflect the smallest evolutionary time frame would involve the inserted portion of the first exons of conglycinin α and α′, since the large insertion appears to be the most recent event in these gene families. If this assumption is valid, the finding that the frequency of transitions in this comparison, 72%, is the highest figure encountered for these genes is consistent with the prediction of transition excess. The high figure in this region may, however, be an artifact of its unusual amino acid composition. The frequencies of transition mutations in the intergeneric comparisons are considerably lower: Pvu β/Gma α′ = 44%; Psa (1+2)/Gma α′ = 48%. When the sequences are compared within a species, the transition frequency is still low with levels similar to those observed in intergeneric comparisons for *Pisum* and *Phaseolus*: Pvu β/Pvu α = 45%; Psa 1/Psa 2 = 51%. In contrast, a comparison of Gma α′ with Gma α yields a higher transition value, 64%, despite the fact that these two sequences appear to be less homologous than, for example, the two phaseolin sequences (Schuler *et al.*, 1982).[6] Assuming that transition excess has occurred, failure to observe high transition levels could reflect either long divergence times for gene copies within a multigene family or the presence of variation only at hot spots that rapidly become saturated with mutations. The latter is perhaps more likely, given what appears to be concerted evolution of the multigene families encoding the 7 S proteins in each species (Slightom *et al.*, 1985).

The values given in Table 5 indicate that divergence is much more apparent at silent codon sites where in most cases greater than 50% substitution levels are observed. This is consistent with selection pressures at the level of the amino acid sequence. It is interesting to note that silent sites have not been saturated by mutations in any of these exons (corrected divergences of over 1.0). The ratio of replacement to silent substitutions, *Rcs*, has been used to assess the degree of evolutionary conservation of globin gene sequences (Czelusniak *et al.*, 1982). In the absence of selection, a value of *Rcs* = 3 is expected, while for many conserved regions of coding sequences, *Rcs* values lower than 1 have been observed. The *Rcs* values observed for the Pvu β/Gma α′ comparison are quite high, with some exons approaching *Rcs* = 3 (Table 5). While this suggests a low degree of evolutionary conservation, it is also possible that rapid divergence is a function of strong positive selection of regions of these proteins, as has been suggested for globin genes (Czelusniak *et al.*, 1982).

DISCUSSION

It is clear that the 7 S seed storage proteins of legumes are not merely random aggregations of amino acids, as one might predict for proteins whose only known function is nutritional. Indeed, as has been pointed out elsewhere (Blagrove *et al.*, 1984; Schuler *et al.*, 1983), a number of potential evolutionary constraints exist for these proteins, even in the fulfillment of this nutritional role. These include subunit interactions necessary for proper holoprotein assembly, packaging of holoproteins in storage bodies, membrane transit during synthesis, and general stability and solubility constraints. The fact that many of these functions are also common to the 11 S storage proteins which, in pea, at least, are found within the same cotyledonary storage bodies as the 7 S proteins (Craig *et al.*,

1980a, 1980b) could be responsible for the similarities observed between the two classes of proteins (Argos *et al.*, 1985). In addition to these potential constraints, conglycinin and phaseolin subunits are not randomly degraded during germination but rather show evidence of discrete proteolytic processing sites (Nielsen and Liener, 1984).[10] In pea vicilin, proteolytic cleavage takes place prior to germination, and the cleavage sites are known. It does not appear that the sequences around these cleavage sites are strongly conserved between vicilin and β-conglycinin or phaseolin (Lycett *et al.*, 1983), and it will be interesting to locate the cleavage sites of conglycinin and phaseolin proteins to see whether or not they lie in the same regions as those of vicilin.

In marked contrast to the zein storage proteins of maize, which are encoded by a large multigene family of over 100 members (Hagen and Rubenstein; 1981; Viotti *et al.*, 1979), the 7 S storage protein subunits of pea, soybean, and common bean are all encoded by smaller multigene families. In these legumes, gene copy number analyses have found fewer than 10 members/haploid genome in *Phaseolus* and *Pisum* (Domoney and Casey, 1985; Gatehouse *et al.*, 1983; Talbot *et al.*, 1984) and fewer than 20 members in the polyploid *Glycine* (Ladin *et al.*, 1984). The level of homology within the individual families is quite high and of a comparable level between 85 and 95% in all three genera. This suggests either the operation of strong evolutionary selection for species-specific characteristics or the maintenance of homogeneity within a species by concerted evolution (Slightom *et al.*, 1980; Zimmer *et al.*, 1980). The two alternatives have very different consequences for those interested in genetic modifications to alter amino acid compositions. Regions conserved among all three genera are presumably poor choices for modification, while areas of low intergeneric homology almost certainly represent regions not required for the general functionality of 7 S proteins and, thus, are attractive sites for modification. However, if the high levels of homology observed within a species represent selectional constraints at the species level, several of these areas may also be essential and cannot be changed. In support of this latter alternative, the proteolytic processing sites of pea vicilin, though almost certainly important in the proper functioning of the vicilin protein, lie in nonconserved areas (Lycett *et al.*, 1983). Areas not highly conserved in amino acid sequence may also be under the influence of selection for physical properties, as is suggested from our hydropathy predictions. Nonetheless, it does appear that several regions of the coding sequences of these genes have been prone to evolutionary change, especially in the accumulation of deletions and insertions, even within a particular species. In the case of soybean, an additional level of structural complexity in the proteins has been imposed by the insertion of a large nucleotide sequence element into one class of ancestral protein, such that both the ancestral type gene (represented by the β-subunit) and genes encoding much larger polypeptides (α- and α′-subunits) are present. This is also presumably true of the near relatives of soybean and may, in fact, be a useful evolutionary marker in establishing relationships in this taxonomically complex group of plants. In soybean Gma α′ and α genes the presence of this large insert offers an attractive target for protein modification.

Acknowledgments—We are grateful to Drs. John Harada and Robert Goldberg for sharing with us the sequence of the β-subunit of β-conglycinin prior to publication and to Dr. Niels Nelson for sharing amino acid sequencing information for the NH_2 terminus of the α′ subunit of β-conglycinin prior to publication.

[10] J. Bryant and R. N. Beachy, manuscript in publication.

REFERENCES

Argos, P., Narayana, S. V. L., and Nielson, N. C. (1985) *EMBO J.* **4**, 1111-1117
Beachy, R. N., Chen, Z.-L., Horsch, R. B., Rogers, S. G., Hoffmann, S. G., and Fraley, R. T. (1985) *EMBO J.* **4**, 3047-3053
Blagrove, R. J., Lilley, G. G., Van Donkelaar, A., Sun, S. M., and Hall, T. C. (1984) *Int. J. Biol. Macromol.* **6**, 137-141
Brown, J. W. S., Bliss, F. A., and Hall, T. C. (1981) *Theor. Appl. Genet.* **60**, 251-258
Brown, W. M., Prager, E. M., Wang, A., and Wilson, A. C. (1982) *J. Mol. Evol.* **18**, 225-239
Casey, R., Domoney, C., and Stanley, J. (1984) *Biochem. J.* **224**, 661-666
Chang, L.-Y. E., and Slightom, J. L. (1984) *J. Mol. Biol.* **180**, 767-784
Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry* **13**, 222-245
Coates, J. B., Medeiros, J. S., Thanh, V. H., and Nielsen, N. C. (1985) *Arch. Biochem. Biophys.* **243**, 184-198
Craig, S., Goodchild, D. J., and Miller, C. (1980a) *Aust. J. Plant Physiol.* **7**, 329-338
Craig, S., Millerd, A., and Goodchild, D. J. (1980b) *Aust. J. Plant Physiol.* **7**, 339-346
Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M. L., Venta, P. J., and Tashian, R. E. (1982) *Nature* **298**, 297-300
Derbyshire, E., Wright, D. J., and Boulter, D. (1976) *Phytochemistry (Oxf.)* **15**, 3-24
Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387-395
Domoney, C., and Casey, R. (1985) *Nucleic Acids Res.* **13**, 687-699
Doyle, J. J., Ladin, B. F., and Beachy, R. N. (1985) *Biochem. Syst. Ecol.* **13**, 123-132
Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980) *Cell* **21**, 653-668
Gatehouse, J. A., Croy, R. R. D., Morton, H., Tyler, M., and Boulter, D. (1981) *Eur. J. Biochem. (Tokyo)* **118**, 627-633
Gatehouse, J. A., Lycett, G. W., Croy, R. R. D., and Boulter, D. (1982) *Biochem. J.* **207**, 629-632
Gatehouse, J. A., Lycett, G. W., Delauney, A. J., Croy, R. R. D., and Boulter, D. (1983) *Biochem. J.* **212**, 427-432
Gilbert, W. (1978) *Nature* **271**, 501
Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. L., and Slightom, J. L. (1984) *J. Mol. Biol.* **180**, 803-823
Gruss, P. (1984) *DNA* **3**, 1-5
Hagen, G., and Rubenstein, I. (1981) *Gene (Amst.)* **13**, 239-249
Hall, T. C., Slightom, J. L., Ersland, D. R., Murray, M. G., Hoffman, L. M., Adang, M. J., Brown, J. W. S., Ma, Y., Matthews, J. A., Cramer, J. H., Barker, R. F., Sutton, D. W., and Kemp, J. D. (1983) in *Structure and Function of Plant Genomes* (Ciferri, O., and Dure, L., III, eds) pp. 123-142, Plenum Publishing Corp., New York
Hayashida, H., and Miyata, T. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 2671-2675
Hoffman, L. M., and Donaldson, D. D. (1985) *EMBO J.* **4**, 883-889
Kloz, J., and Klozova, E. (1974) *Biol. Plant (Prague)* **16**, 290
Kozak, M. (1983) *Microbiol. Rev.* **47**, 1-45
Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105-132
Ladin, B. F., Doyle, J. J., and Beachy, R. N. (1984) *J. Mol. Appl. Genet.* **2**, 372-380
Lycett, G. W., Delauney, A. J., Gatehouse, J. A., Gilroy, J., Croy, R. R. D., and Boulter, D. (1983) *Nucleic Acids Res.* **11**, 2367-2380
Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560
McKnight, S. L., and Kingsbury, R. (1982) *Science* **217**, 316-324
Meinke, D. W., Chen, J., and Beachy, R. N. (1981) *Planta (Berl.)* **153**, 130-139
Messing, J., Geraghty, D., Heidecker, G., Hu, N. T., Kridl, J., and Rubenstein, I. (1983) in *Genetic Engineering of Plants* (Kosuge, T., Meredith, C. P., and Hollaender, A., eds) pp. 211-227, Plenum Publishing Corp., New York
Mononen, I., and Karjalainen, E. (1984) *Biochim. Biophys. Acta* **788**, 364-367
Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P., and Chambon, P. (1981) *Nucleic Acids Res.* **9**, 6047-6068
Morelli, G., Nagy, F., Fraley, R. T. Rogers, S. G., and Chua, N. H. (1985) *Nature* **315**, 200-204
Nielsen, S. S., and Liener, I. E. (1984) *Plant Physiol. (Bethesda)* **74**,

494–498
Ohno, S. (1984) *J. Mol. Evol.* **20**, 313-321
Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980) *Cell* **20**, 555-566
Polhill, R. M. (1981) in *Advances in Legume Systematics* (Polhill, R. M., and Raven, R. H., eds) pp. 191-208, Royal Botanic Gardens, Edinburgh, Scotland
Rautmann, G., Matthes, H. W. D., Gait, M. J., and Breathnach, R. (1984) *EMBO J.* **3**, 2021-2028
Schuler, M. A., Schmitt, E. S., and Beachy, R. N. (1982) *Nucleic Acids Res.* **10**, 8225–8244
Schuler, M. A., Doyle, J. J., and Beachy, R. N. (1983) *Plant Mol. Biol.* **2**, 119-127
Sengupta-Gopalan, C., Reichert, N. A., Barker, R. F., Hall, T. C., and Kemp, J. D. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 3320–3324
Shah, D. M., Hightower, R. C., and Meagher, R. B. (1983) *J. Mol. Appl. Genet.* **2**, 111-126
Slightom, J. L., Blechl, A. E., and Smithies, O. (1980) *Cell* **21**, 627–638
Slightom, J. L., Sun, S. M., and Hall, T. C. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1897-1901
Slightom, J. L., Chang, L. Y. E., Koop, B. F., and Goodman, M. (1985a) *Mol. Biol. Evol.* **2**, 370-389
Slightom, J. L., Drong, R. F., Klassy, R. C., and Hoffman, L. M. (1985b) *Nucleic Acids Res.* 6483-6498
Smith, D. R., and Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255-2274
Sun, S. M., Slightom, J. L., and Hall, T. C. (1981) *Nature* **289**, 37-41
Talbot, D. R., Adang, M. J., Slightom, J. L., and Hall, T. C. (1984) *Mol. Gen. Genet.* **198**, 42–49
Viotti, A., Sala, E., Marotta, R., Alberi, P., Balducci, C., and Soave, C. (1979) *Eur. J. Biochem.* **102**, 211–222
Von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17-21
Weiher, H., König, M., and Gruss, P. (1983) *Science* **219**, 626–631
Wieringa, B., Hofer, E., and Weissmann, C. (1984) *Cell* **37**, 915-925
Wilbur, W. J., and Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 726-730
Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W., and Wilson, A. C. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 2158–2162

Supplementary Material to: THE GLYCOSYLATED SEED STORAGE PROTEINS OF GLYCINE MAX AND PHASEOLUS VULGARIS: structural homologies of genes and proteins

Table 3. Effect of Gma a' exon 1 major insertion on the amino acid composition of Gma a' polypeptide.

	Percent Composition		
Amino Acid	Phaseolus	Glycine without Insertion	Glycine Insertion
Ala	5.9	5.5	1.2
Arg	4.5	6.4	7.7
Asn	7.1	7.9	3.0
Asp	5.0	4.7	4.1
Cys	0.0	0.0	3.0
Gln	5.5	7.4	11.2
Glu	8.6	7.2	27.8
Gly	5.5	4.9	4.7
His	2.4	1.5	8.9
Ile	5.9	6.0	0.6
Leu	10.9	9.8	1.8
Lys	5.7	6.0	8.3
Met	1.2	1.3	0.0
Phe	5.9	6.4	1.2
Pro	3.6	4.5	8.3
Ser	9.3	7.9	5.9
Thr	3.3	3.0	0.0
Trp	0.2	0.2	1.2
Tyr	3.1	3.0	0.6
Val	6.4	6.6	0.6

Table 4. Codon usage of mature conglycinin and phaseolin proteins.

Phe	UUU	10	7	Ser	UCU	12	10	Tyr	UAU	6	5	Cys	UGU	0	0
Phe	UUC	19	16	Ser	UCC	5	9	Tyr	UAC	7	8	Cys	UGC	1	0
Leu	UUA	0	2	Ser	UCA	6	1	*oc	UAA	0	0	*op	UGA	1	1
Leu	UGG	10	6	Ser	UCG	1	1	*am	UAG	0	0	Trp	UGG	2	1
Leu	CUU	12	11	Pro	CCU	13	5	His	CAU	6	5	Arg	CGU	3	3
Leu	CUC	9	7	Pro	CCC	4	4	His	CAC	14	5	Arg	CGC	8	1
Leu	CUA	6	5	Pro	CCA	14	4	Gln	CAA	32	12	Arg	CGA	5	1
Leu	CUG	4	5	Pro	CCG	1	1	Gln	CAG	20	11	Arg	CGG	3	0
Ile	AUU	13	9	Thr	ACU	3	3	Asn	AAU	10	10	Ser	AGU	9	4
Ile	AUC	6	9	Thr	ACC	8	5	Asn	AAC	27	20	Ser	AGC	7	10
Ile	AUA	9	6	Thr	ACA	3	2	Lys	AAA	18	15	Arg	AGA	10	5
Met	AUG	9	6	Thr	ACG	0	3	Lys	AAG	20	9	Arg	AGG	9	6
Val	GUU	9	9	Ala	GCU	9	7	Asp	GAU	11	13	Gly	GGU	10	8
Val	GUC	4	4	Ala	GCC	7	6	Asp	GAC	17	8	Gly	GGC	4	1
Val	GUA	2	2	Ala	GCA	6	8	Glu	GAA	41	17	Gly	GGA	10	10
Val	GUG	12	11	Ala	GCG	1	0	Glu	GAG	38	19	Gly	GGG	5	3

Numbers of codons in each protein are given to the right of each codon class, with Gma a' listed first.

Table 5. Divergence between phaseolin 8 and conglycinin a' at silent and replacement sites in coding regions.

	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6
observed silent %	0.21	0.27	0.20	0.27	0.19	0.25
corrected silent %	0.43	0.51	0.50	0.64	0.42	0.52
observed replac. %	0.19	0.31	0.11	0.32	0.19	0.28
corrected replac. %	0.24	0.43	0.14	0.47	0.26	0.44
Rcs	1.93	2.73	1.30	2.87	2.43	2.74

Corrected divergences reflect multiple mutations at a site (Perler et al., 1980).

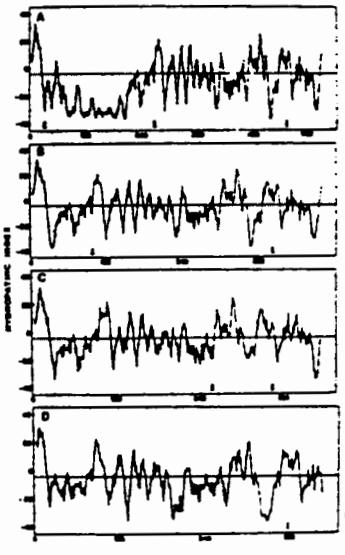


Fig. 4. Hydropathic profiles for 7S storage protein genes from three legumes. Hydropathic profiles have been constructed using the 7S storage protein amino acid sequences shown in Figs. 1 and 2. Frame A shows the profile for the complete Gmaa' protein (including the 174 amino acid residue insert), while frame B shows the profile of Gma a' without the insert (the amino acid sequence numbering coordinate has been adjusted between frames A and B). Frames C and D show the profile for Pvu 8 and Psa 1 and Psa 2 amino sequences, respectively. The hydropathic profiles in frames B,C, and D have similar amino acid sequence coordinates so that amino acid hydropathic domains can be aligned vertically between the different frames. Hydrophobic domains are indicated by a positive index and hydrophilic domains have a negative index. The arrows indicate the location of N-glycosyl recognition sites present in each polypeptide.

Transgenic Canola and Soybean Seeds with Increased Lysine

S. C. Falco[1,*], T. Guida[1], M. Locke[1], J. Mauvais[1], C. Sanders[1,3], R. T. Ward[1] and P. Webber[2]

[1]Agricultural Products and [2]Central Research & Development, E. I. DuPont de Nemours & Co., Wilmington, DE 19880-0402. [3]Department of Plant and Soil Sciences, University of Delaware, Newark, DE 19717-1303. *Corresponding author (email: scfalco@esvax.dnet.dupont.com).

We have increased the lysine content in the seeds of canola and soybean plants by circumventing the normal feedback regulation of two enzymes of the biosynthetic pathway, aspartokinase (AK) and dihydrodipicolinic acid synthase (DHDPS). Lysine-feedback-insensitive bacterial DHDPS and AK enzymes encoded by the *Corynebacterium dapA* gene and a mutant *E. coli lysC* gene, respectively, were linked to a chloroplast transit peptide and expressed from a seed-specific promoter in transgenic canola and soybean seeds. Expression of *Corynebacterium* DHDPS resulted in more than a 100-fold increase in the accumulation of free lysine in the seeds of canola; total seed lysine content approximately doubled. Expression of *Corynebacterium* DHDPS plus lysine-insensitive *E. coli* AK in soybean transformants similarly caused several hundred-fold increases in free lysine and increased total seed lysine content by as much as 5-fold. Accumulation of α-amino adipic acid (AA) in canola and saccharopine in soybean, which are intermediates in lysine catabolism, was also observed.

Received 22 December 1994; accepted 27 March 1995.

Human food and animal feed derived from many grains are deficient in some of the ten essential amino acids which are required in an animal diet. Corn, which is a preferred animal feed because it is a low cost energy source, is relatively poor in amino acid content, with lysine being the most limiting amino acid for the dietary requirements of many animals. Soybean meal, which is rich in lysine and other essential amino acids, is used to supplement corn-based animal feeds. In addition, about 200,000 tons of lysine are produced annually via fermentation, mostly for use as an animal feed additive. An increase in the lysine content of corn, soybean, or other animal feed sources would reduce the need to supplement the seeds with crystalline lysine.

One approach to raising the lysine content of seeds is to increase its production by deregulating the biosynthetic pathway. Lysine, along with threonine, methionine, and isoleucine, are amino acids derived from aspartate (Fig. 1). The first step in the pathway is the phosphorylation of aspartate by the enzyme aspartokinase (AK), and this enzyme has been found to be an important target for regulation, usually via end-product inhibition, in plants. The condensation of aspartyl β-semialdehyde with pyruvate catalyzed by dihydrodipicolinic acid synthase (DHDPS) is the first reaction committed to lysine biosynthesis. In plants DHDPS is feedback inhibited by lysine and serves as the major regulator of the lysine branch of the pathway. Many attempts have been made to obtain lysine overproducing mutants of various plants by selecting for resistance to the lysine analog S(2-aminoethyl)-cysteine (AEC), but only one tobacco mutant with a significant increase in free lysine, in leaves only, has been reported[1].

Recently, genetic engineering technology has been used to increase free lysine production in the leaves of plants. The *E. coli dapA* gene, which encodes a DHDPS enzyme that is about 20-fold less sensitive to inhibition by lysine than than a typical plant DHDPS, was linked to the 35S promoter of Cauliflower Mosaic Virus and a plant chloroplast transit sequence to direct the protein to the chloroplast, wherein most amino acid biosynthetic pathway enzymes are localized. The chimeric gene was introduced into tobacco cells via transformation and shown to cause a substantial increase in free lysine levels in leaves[2-5]. However, the lysine content in the seeds of the transformed plants was not increased in these studies.

A mutant *E. coli lysC* gene that encodes a lysine-insensitive AK enzyme was linked to the 35S promoter of Cauliflower Mosaic Virus and a plant chloroplast transit sequence and expressed in tobacco cells in concert with the *E. coli dapA* gene[4-6]. More free lysine accumulated in the leaves of plants expressing lysine-insensitive AK plus *E. coli* DHDPS than in plants expressing DHDPS alone, but again there was no increase in the level of free lysine in seeds. Strong seed-specific promoters were used to achieve high-level expression of both the *E. coli dapA* and mutant *E. coli lysC* genes in tobacco seed, but still no increase in the accumulation of free lysine was observed[5,7]. However, accumulation of an intermediate of lysine catabolism was seen in one study[5], and an increased level of a lysine catabolic enzyme in the other[4], suggesting that lysine breakdown was preventing accumulation of excess free lysine in seeds.

In this paper we report on genetic engineering of the lysine biosynthetic pathway in soybean and canola seeds. In these crops we have been able to achieve a large increase in seed lysine content.

Results

Genes and constructs. We attempted to bypass feedback regulation of the lysine biosynthetic pathway (see Fig. 1) by

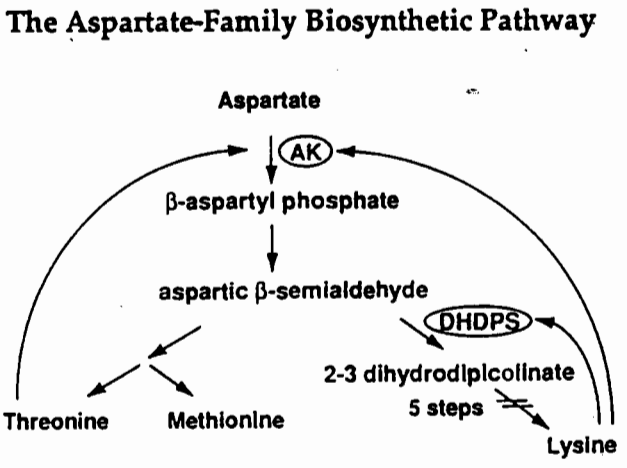


FIGURE 1. Biosynthetic pathway for amino acids derived from aspartate illustrating end-product feedback inhibition of dihydrodipicolinic acid synthase (DHDPS) and aspartokinase (AK).

FIGURE 2. Maps of canola transformation vectors pFS926 and pBT597, and soybean vector pBT614. *Kan* is kanamycin resistance gene for selection in bacteria; *ori* is origin of DNA replication in *E. coli*; *rep pVSI* and *sta* allow stable DNA replication in *Agrobacterium*; *35S 5'* is 35S promoter from Cauliflower Mosaic Virus; *NPTII* is neomycin phosphotransferase gene for selection of kanamycin resistant transformed plants; *NOS 3'* is transcription termination region from the nopaline synthase gene; *Pv 5'* is the promoter from the bean β-phaseolin gene; *cts/dapA* is the coding region for the chloroplast transit peptide from the small subunit of ribulose-bis-phosphate carboxylase from soybean fused to the *Corynebacterium glutamicum dapA* coding region; *Pv 3'* is the transcription termination region from the bean β-phaseolin gene; *cts/lysC-M4* is the coding region for the chloroplast transit peptide from the small subunit of ribulose-bis-phosphate carboxylase from soybean fused to the mutant *E. coli lysC-M4* gene; *Amp* is the ampicillin resistance gene for selection in bacteria; *GUS* is the coding region of the *E. coli uidA* gene for expression of β-glucuronidase, the scorable marker for soybean transformation.

expressing lysine insensitive forms of the two key enzymes, aspartokinase (AK) and dihydrodipicolinic acid synthase (DHDPS) in the seeds of plants. The *Corynebacterium dapA* gene from ATCC strain 13032 (ref. 8), which encodes a lysine insensitive DHDPS, was amplified from genomic DNA using the polymerase chain reaction (PCR). The *E. coli lysC* gene, which encodes a lysine sensitive AK[9], was isolated from the ordered *E. coli* genomic library of Kohara, Akiyama, and Isono[10]. Then mutations that rendered the enzyme insensitive to lysine were selected in the cloned gene[3]. A mutation designated *lysC*-M4, which results in an isoleucine for threonine substitution at amino acid 352, was used in this work. Both genes were modified to have an Nco I restriction enzyme site at the translation start codon. Then, the chloroplast transit sequence (cts) of the small subunit of ribulose 1,5-bisphosphate carboxylase from soybean[11] was fused to the *dapA* and *lysC*-M4 coding sequences. The cts/*dapA* and cts/*lysC*-M4 genes were inserted into a seed-specific gene expression cassette composed of the promoter and transcription terminator from the gene encoding the β-subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris*[12]. Details on the isolation and construction of these chimeric genes are in the Experimental Protocol.

Transgenic canola seeds with high levels of lysine. The binary vectors pFS926 (Fig. 2A), carrying the chimeric gene cassette phaseolin promoter/cts/cor*dapA*/phaseolin 3', and pBT597 (Fig. 2B), carrying the phaseolin promoter/cts/cor*dapA*/phaseolin 3' plus phaseolin promoter/cts/*lysC*-M4/phaseolin 3', were used for *Agrobacterium*-mediated transformation of canola. In these binary vectors the 35S promoter from Cauliflower Mosaic Virus drives expression of the NPT II gene permitting selection of kanamycin-resistant transformants. Transformed canola plants were self-pollinated and mature seeds were harvested and analyzed.

Western blots of protein extracts from mature seeds (not shown) revealed that eight of eight pFS926 transformants

TABLE 1. Canola transformants: pFS926: phaseolin 5'/cts/*dapA*/phaseolin 3', pBT597: phaseolin 5'/cts/*dapA*/phaseolin 3' + phaseolin 5'/cts/*lysC*-M4/phaseolin 3'.

Plasmid-Line	Free Amino Acids[1] K/L	AA/L	Western *Coryne.* DHDPS	Western *E. coli* AKIII-M4	Percent Total Amino Acids[2] K	AA
None-Westar	0.8	0	–	–	6.5	0
Vector-1	1.3	0	–	–	6.3	0
pFS926-3	140	16	++++	–	12	1.0
pFS926-9	110	12	++++	–	11	0.8
pFS926-29	38	4.7	++++	–	12	1.6
pFS926-6	14	4.6	+++	–	8.2	0
pFS926-68	4.2	0.9	++	–	8.3	0
pFS926-11	7.9	5.2	++	–	7.7	0
pFS926-27	4.2	1.1	++	–	7.1	0
pFS926-22	3.1	0.3	+	–	6.9	0
pBT597-4	38	4.5	++++	++++	13	1.6
pBT597-100	9.1	1.7	+++	++	6.6	0
pBT597-148	7.6	0.9	+++	+	7.3	0
pBT597-169	5.6	1.7	+++	+++	6.6	0
pBT597-14	6.0	4.3	++	+/–	7.0	0
pBT597-68	4.7	1.5	++	+	6.9	0

[1]Free amino acids are shown as ratios of K (lysine) to L (leucine) or AA (α-aminoadipic acid) to L (leucine). [2]Total amino acids are shown as percentages of K (lysine) or AA (α-aminoadipic acid) of the total seed amino acid content (excluding tryptophan and cysteine).

and seven of seven pBT597 transformants expressed the DHDPS protein. Five of the seven pBT597 transformants also expressed the AKIII-M4 protein. From the size of the proteins on the western blots it was apparent that most of the expressed protein had the chloroplast transit peptide removed, suggesting that most of the protein had been imported into plastids.

Expression of *Corynebacterium* DHDPS lead to a large increase in accumulation of free lysine in canola transformants (Table 1). The highest expressing lines showed a greater than 100-fold increase in free lysine level in the seeds. The free lysine increase was great enough to cause a significant increase in the total seed lysine content. In the transformants with the highest levels, lysine makes up about 12% of the total seed amino acids, nearly double the lysine content of standard canola.

Control of lysine accumulation in canola seeds. There was a good correlation between expression level of *Corynebacterium* DHDPS protein, as estimated from western blots, and the level of lysine accumulation in the seeds of transformants (Table 1). The expression level of DHDPS protein appears to be primarily determined by the transgene copy number. For example, southern blots (not shown) indicate that lines pFS926-3 and 9 have at least three sites of insertion of the transgene, while lines pFS926-11, 27 and 22 have one.

Free amino acids were extracted and quantitated from 16 single seeds in the segregating population of each of the primary transformants. The results from three lines that contain a single site of insertion are shown in Figure 3A; approximately one quarter of these seeds are expected to be wild type, one half heterozygous, and one quarter homozygous for the transgene. None of the lines show a three step pattern of free lysine levels, but instead only a few seeds (three to five) from each of the transformants show a four-fold or greater level of free lysine compared to the control seeds. It seems reasonable to propose that the high lysine seeds are homozygous for the transgene. If so, this would suggest that at least two copies of the transgene are needed to provide sufficient expression of the *Corynebacterium* DHDPS enzyme to accumulate this level of excess free lysine.

The free lysine levels from single seeds of three transformants that contain two insertion sites, are shown in Figure 3B. For two independently segregating insertions eleven of sixteen seeds (on average) would be expected to contain two or more copies of the transgene. In this case nine to twelve seeds from each of the transformants show a four-fold or greater increase in the accumulation of free lysine, consistent with the proposal that two copies of the transgene are needed to accumulate this level of excess free lysine. Twenty to 100-fold increases in free lysine were seen in many of the seeds; some of these seeds may contain three or four transgene copies.

Accumulation of lysine in the canola seeds is controlled not only by the rate of its synthesis, but also by the rate of breakdown. This is shown by the build-up of α-aminoadipic acid in the transformed lines. This compound has been found to be an intermediate in the catabolism of lysine in wheat[13], but is normally detected only via radioactive tracer experiments due to its low level of accumulation. The observation of levels of α-aminoadipic acid higher than most free amino acid pools indicates that a large amount of lysine is entering the catabolic pathway and that the breakdown of α-aminoadipic acid is a rate-limiting step in lysine catabolism in canola.

Expression of lysine-insensitive *E. coli* AKIII in canola seeds results in an increase in free threonine as it does in tobacco seeds[4], indicating that the intracellular lysine concentration controls flux down the aspartate-family pathway. However, expression of *E. coli* AKIII along with *Corynebacterium* DHDPS does not increase lysine accumulation in canola seeds over the increase obtained by expressing *Corynebacterium*

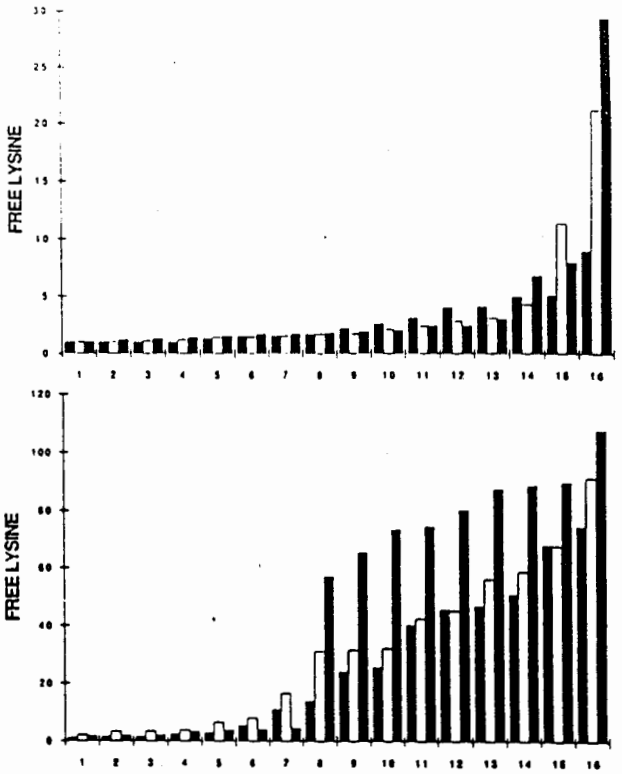


FIGURE 3. Free lysine in single canola seeds. Free amino acids were extracted from sixteen single seeds from each line. Amino acids were separated and quantitated as described in the Experimental Protocol. (A) lines pFS926-22 (black), pFS926-11 (white), pFS926-27 (gray) contained a single transgene insertion. (B) lines pFS926-68 (black), pFS926-6 (white), pFS926-29 (gray) contained two transgene insertions.

DHDPS alone (Table 1). On the other hand, expression of *Corynebacterium* DHDPS prevents the accumulation of excess threonine that is seen in canola seeds expressing lysine-insensitive *E. coli* AKIII alone; this effect was also previously observed in tobacco seeds[3,7] and leaves[4]. Thus, *Corynebacterium* DHDPS apparently directs the flow of pathway intermediates down the lysine branch and away

TABLE 2. Soybean transformants: pBT614: phaseolin 5′/cts/*dapA*/phaseolin 3′ + phaseolin 5′/cts/*lysC*-M4/phaseolin 3′.

Cultivar-Line-Seed	Western *Coryne.* DHDPS	Western *E. Coli* AKIII-M4	Percent Total Amino Acids[1] K	Sac
A2396	–	–	5.8	0
A2396-145-2	+	–	7.2	0.3
A2396-233-B1	+	+	25	n.d.
A2396-234-B28	+	+	15	0.1
A2396-240-E7	+	+	11	0
A2396-248-B33	+	–	7.1	n.d.
A2396-267-E14	+	+	6.4	0
A2242	–	–	6.0	0
A2242-273-D28	+	+	15	1.0
A2242-315-E25	+	+	17	0.4
A2242-316-E37	+	+	12	0.5
A5403	–	–	5.9	0
A5403-175-8	+	+	7.7	0.7
A5403-183-2	+	+	7.3	0.4
A5403-196-4	+	+	7.6	0.4
A5403-214-E51	+	+	34	0.1
A5403-222-E65	+	+	15	0.5
A5403-225-E75	+	+	13	0.5

[1]Total amino acids are shown as percentages of K (lysine) or Sac (saccharopine) of the total seed amino acid content (excluding tryptophan and cysteine).

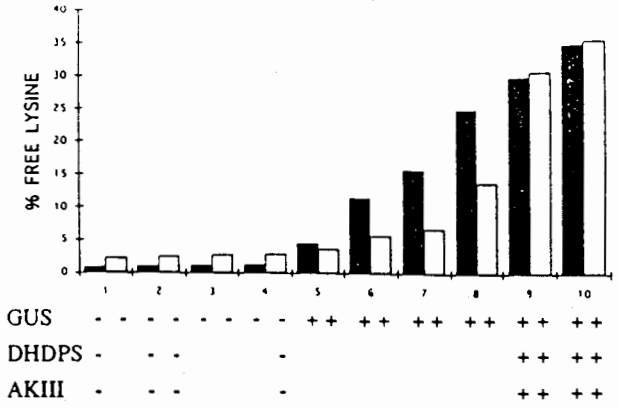


FIGURE 4. Free lysine in single soybean seeds. Free amino acids were extracted from ten single seeds from each line. Amino acids were separated and quantitated as described in the Experimental Protocol. Line A2396-145 (black), A2396-183 (white). Expression of GUS, DHDPS and AK is indicated where tested.

from threonine.

Inheritance of the high lysine trait. R1 seeds from line pFS926-27, which carried the transgenes at a single insertion site, were morphologically indistinguishable from standard canola seeds and germinated well. Three R1 plants predicted to be homozygous for the transgene, based upon quantitative PCR analysis of leaf DNA, were grown to maturity and R2 seed was analyzed for lysine content. Seeds from all three plants exhibited higher levels of free lysine than the pooled R1 (segregating) seed. Total lysine content in the R2 seeds averaged 9.3% (a 45% increase over standard canola), compared to 7.1% lysine (or an 11% lysine increase) seen in the R1 seed.

Seeds from canola transformants accumulating high levels of free lysine were also morphologically indistinguishable from standard canola seeds, but germination of the R1 seeds from the highest lysine lines, e.g. pFS926-3 and pFS926-9, was poor. Because both of these lines carried multiple insertions, it was not clear whether the high lysine content of the seeds was the cause of the poor germination. If only low lysine containing seeds were able to germinate, R2 seeds would be expected to have lower lysine content than the R1 seeds. However, pools of (still segregating) R2 seed from pFS926-3 and pFS926-9 maintained high levels of free lysine and the total lysine level of these seeds remained about 12% of the total seed amino acid content. Thus, association of the high lysine phenotype with poor germination of the R1 seeds is uncertain.

Transgenic soybean seeds with high levels of lysine. The vector pBT614 (Fig. 2C), carrying the chimeric gene cassettes phaseolin promoter/cts/cor*dapA*/phaseolin 3′ and phaseolin promoter/cts/*lysC*-M4/phaseolin 3′, was used for DNA-coated particle bombardment mediated transformation of soybean. As a scorable marker gene for transformation, the vector also carried a 35S promoter/*uidA*/Nos 3′ gene for expression of *E. coli*, β-glucuronidase (GUS)[14]. GUS-positive transformants were also screened for the presence of the *dapA* coding region using PCR. Transformed soybean plants were self-pollinated and mature seeds were harvested.

Single seeds from transformed lines were analyzed for GUS expression via enzyme assay and DHDPS and AK protein via western blot. All transformants showed linkage of GUS, DHDPS and AK expression in individual R1 seeds. Twenty-one of twenty-three transformants expressed the DHDPS protein and eighteen of the twenty-one also expressed the AKIII-M4 protein. As was seen in canola, the size of the DHDPS protein on western blots indicated that most of the protein had the chloroplast transit peptide removed, suggesting that most of the DHDPS had been imported into the plastids.

Soybean transformants expressing *Corynebacteria* DHDPS alone or in concert with *E. coli* AKIII-M4 accumulated high levels of free lysine in their seeds. Analysis of individual seeds in the segregating R1 population showed that increased free lysine was observed in seeds expressing the transgenes (Fig. 4). It seemed likely that the seeds exhibiting the highest levels of free lysine would be homozygous for the transgene insertion, including the linked GUS marker gene. It was possible to test this directly since only a fraction of the R1 seed was used for analysis and the remainder of the seed could be planted. In several cases where this was done, analysis of the R2 seed confirmed the prediction. For example, the two seeds from line A2396-145 in Figure 4 containing the highest free lysine were planted and the R2 seeds showed no segregation of the GUS marker gene.

The level of free lysine accumulation varied from ten-fold increases in some lines to several hundred-fold increases in others. It has not been possible to correlate this with gene copy number, as in canola, because the DNA insertion patterns resulting from gene gun mediated transformation were usually more complex and more difficult to interpret. It was generally true in transformants where insertion was at a single locus that seeds wherein the transgene was homozygous accumulated more lysine than in the heterozygous siblings.

Accumulation of free lysine in soybean seeds is also controlled in part by the catabolic rate, as was seen in canola. In contrast to canola however, it is saccharopine, which is the first intermediate of lysine catabolism, rather than α-aminoadipic acid, that accumulates. Levels of saccharopine higher than that of most of the other free amino acids demonstrate that a large amount of lysine is entering the catabolic pathway. The build-up of this intermediate indicates that the breakdown of saccharopine is a rate-limiting step for lysine catabolism in soybean.

The free lysine increases obtained in these transformants resulted in significant increases in total seed lysine. In seeds from transformed soybean lines that expressed *Corynebacteria* DHDPS alone (see for examples Table 2, lines A2396-145-2 and A2396-248-B33) lysine makes up as much as 7.2% of the total seed amino acids, about a 25% increase over untransformed soybean. Soybean seeds expressing *Corynebacteria* DHDPS in concert with *E. coli* AKIII-M4 showed much greater accumulation of total seed lysine than those expressing *Corynebacteria* DHDPS alone, in contrast to the results in canola. Seeds with lysine making up 11–17% of the total amino acid content, a two to three-fold increase over standard soybean, were obtained from seven different transgenic lines in three different elite parents (Table 2). Even more striking are two transgenic lines with seeds showing a four to five-fold increase in total lysine content (Table 2, lines A2396-233-B1 and A5403-214-E51). In the latter seeds lysine makes up 25–33% of the total seed amino acids.

Seed morphology and inheritance of the high lysine trait. Seeds with a lysine content up to about 12% of the total amino acids, double the content of standard soybeans, were normal in appearance and germinated well. The lysine content

in R2 and R3 seeds from the lines so-far analyzed has remained at least as high as that observed in the R1 seed, demonstrating heritability of the trait. Seeds from soybean transformants accumulating the highest levels of lysine were wrinkled in appearance, and germination of these R1 seeds was poor. Correlation of wrinkled seed with poor germination was also seen in the R2 generation. This effect was evident in seeds from several different lines where lysine was 15% or more of the total seed amino acids, and thus appears to be a consequence of the increased lysine.

Discussion

We have been able to substantially increase the lysine content in the seeds of canola and soybean. This was accomplished by bypassing the normal cellular control of the biosynthetic pathway resulting in the accumulation of excess free lysine. To be successful, it was necessary to achieve a rate of synthesis that exceeded the rate of lysine catabolism. In the future it may be possible to make the process more efficient by decreasing the rate of catabolism through mutation or genetic engineering.

It is possible to estimate the economic value of the high lysine seeds, although this depends upon the prices of the various animal feed ingredients and the specific uses. The cost of the crystalline lysine is currently about $1.20/1b. Transgenic lines that double the seed lysine content provide about 3 lbs of additional lysine per 100 lbs of soybean meal. Thus, a soybean meal with double the normal lysine content is worth an additional $3.60/100 lbs over commodity soybean meal, which represents a 30–35% increase in value.

It must be pointed out that there are a number of uncertainties that will need to be addressed before these nutritionally improved crops can be commercialized. Agronomic effects such as seed germination, pest susceptibility, and yield will have to be carefully evaluated. Because these seeds have a different composition from the standard varieties, seed handling, storage and processing characteristics will also have to be studied. The extensive use of crystalline lysine as a supplement in animal diets gives us reason to expect that the excess free lysine in these seeds will be available to animals, but this will have to be demonstrated through animal feeding studies. The effects of the catabolic by-products, saccharopine or α-amino adipic acid, on animals will also have to be thoroughly investigated. Animals have the capability of breaking down lysine via the same pathway that operates in plants. Thus it is unlikely that these non-protein amino acids will pose a problem.

Recently we have applied this technology to corn and achieved a two to three-fold increase in corn seed lysine content (manuscript in preparation). We are currently working on extending this metabolic engineering approach to increase the accumulation of the other amino acids derived from aspartate, threonine and methionine.

Experimental Protocol

Gene construction. The *Corynebacterium dapA* gene was isolated from genomic DNA from ATCC strain 13032 using polymerase chain reaction (PCR). The nucleotide sequence of the *Corynebacterium dapA* gene has been published[4]. The oligonucleotide primers used for PCR were: 5′-CCCGGGCCAT GGCTACAGGT TTAACAGCTA AGACCGGAGT AGAGCACT and 5′-GATATCGAAT TCTCATTATA GAACTCCAGC TTTTTTC. The primers added unique restriction sites (underlined) at the start codon (Nco I) and just past the stop codon (EcoR I) of the gene. In addition to introducing an Nco I site at the translation start codon, the PCR primers also resulted in a change of the second codon from AGC coding for serine to GCT coding for alanine. Several clones that expressed active, lysine-insensitive DHDPS were isolated, indicating that the second codon amino acid substitution did not affect activity. The *E. coli lysC*, which has been cloned and sequenced previously[5], was obtained on bacteriophage lambda clones 4E5 and 7A4 from the ordered *E. coli* genomic library of Kohara, Akiyama and Isono[10]. An EcoR I-Nhe I fragment was isolated and subcloned in plasmid pBR322 yielding pBT436. To establish that the cloned *lysC* gene was functional it was shown to relieve the nutritional requirements of *E. coli* strain Gif106M1 (*E. coli* Genetic Stock Center strain CGSC-5074) which has mutations in each of the three *E. coli* AK genes[11]. Addition of 0.2mM lysine to minimal growth medium inhibits the growth of Gif106M1 transformed with pBT436 because AKIII is inhibited by exogenously added lysine leading to starvation for threonine and methionine. This property of pBT436-transformed Gif106M1 was used to select for mutations in *lysC* that encoded lysine-insensitive AKIII. Three lysine tolerant mutants expressed AKIII that was uninhibited by 15mM lysine, whereas wild type AKIII is 50% inhibited by 0.3–0.4mM lysine. The sequences of the three mutant *lysC* genes each differed from the wild type sequence by a single nucleotide. Mutant M2 is an A to G transition resulting in an isoleucine for methionine substitution at amino acid 318 and mutants M3 and M4 have identical T for C transitions resulting in an isoleucine for threonine substitution at amino acid 352. An Nco I site was inserted at the translation initiation codon of the wild type and mutant *lysC* genes using oligonucleotide adaptors. The *chloroplast transit sequence* (cts) of the small subunit of ribulose 1,5-bisphosphate carboxylase from soybean[12] was fused to the *dapA* and *lysC*-M4 coding sequences. The cts was added in two steps using 75 and 90 base pair oligonucleotides for the *lysC* gene, and then PCR was used to generate a full-length cts for addition to the *dapA* gene. The cts/dapA and cts/lysC genes were inserted into a seed-specific gene expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris*[13]. The phaseolin cassette includes about 500 nucleotides upstream (5′) from the translation initiation codon and about 1650 nucleotides downstream (3′) from the translation stop codon of phaseolin[14]. Between the 5′ and 3′ regions the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I were added using an oligonucleotide adaptor. The entire cassette is flanked by Hind III sites. The chimeric gene cassettes phaseolin promoter/cts/cor*dapA*/phaseolin 3′ plus phaseolin promoter/cts/*lysC*-M4/phaseolin 3′ were inserted into the binary vector pZS199. To insert the phaseolin promoter/cts/cor*dapA*/phaseolin 3′, oligonucleotide adaptors were used to add BamH I sites to the ends of the gene cassette, which was then isolated as a 2.7 kb BamH I fragment and inserted into BamH I digested pZS199, yielding plasmid pFS926 (Fig. 2A) and into a previously constructed pZS199 derivative containing the phaseolin promoter/cts/*lysC*-M4/phaseolin 3′, yielding pBT597 (Fig. 2B). The chimeric gene cassettes were inserted into a soybean transformation vector consisting of the 35S promoter from Cauliflower Mosaic Virus driving expression of the *E. coli* β-glucuronidase gene[14] with the Nos 3′ region in a modified pGEM9Z plasmid. To insert the phaseolin promoter/cts/*lysC*-M4/phaseolin 3′, the gene cassette was isolated as a 3.3 kb Hind III fragment and inserted into Hind III digested vector yielding plasmid pBT609. To insert the phaseolin promoter/cts/cor*dapA*/phaseolin 3′, the gene cassette was isolated as a 2.7 kb BamH I fragment and inserted into BamH I digested pBT609, yielding plasmid pBT614 (Fig. 2C).

Transformation of canola and soybeans. The binary vectors pFS926 or pBT597 were transferred by tri-parental matings[17] to the disarmed *Agrobacterium tumefaciens* strain LBA4404/pAL4404[18]. *Brassica napus* cultivar "Westar" was transformed by co-cultivation of seedling pieces with the two *Agrobacterium* strains (for details see ref. 5). Transformed canola plants were grown under a 16:8-hr photoperiod, with a daytime temperature of 23°C and a nightime temperature of 17°C. When the primary flowering stem began to elongate, it was covered with a mesh pollen-containment bag to prevent outcrossing. Self-pollination was facilitated by shaking the plants several times each day. Mature seeds derived from self-pollinations were harvested about three months after planting. Soybean was transformed with plasmid pBT614 by Agracetus Company (Middleton, WI) according to the procedure described in Christou et al.[19].

Evaluation of transgenic canola and soybean seeds. A partially defatted canola seed meal was prepared as follows: 40 milligrams of mature dry seed was ground with a mortar and pestle under liquid nitrogen to a fine powder. One ml of hexane was added and the mixture was shaken at room temperature for 15 min. The meal was pelleted in an eppendorf centrifuge, the hexane was removed and the hexane extraction was repeated. Then the meal was dried at 65° for 10 min. To extract protein, approximately 30–40 mg of meal was put into a 1.5 ml disposable plastic microfuge tube and ground in 0.25 ml of 50 mM Tris-HCl pH 6.8, 2 mM EDTA, 1% SDS, 1% β-mercaptoethanol. The resultant suspensions were centrifuged for 5 min at room temperature in a microfuge to remove particulates. Three volumes of extract was mixed with 1 volume of 4 X SDS-gel sample buffer (0.17M Tris-HCl pH6.8, 6.7% SDS, 16.7% β-mercaptoethanol, 33% glycerol) and 5 μl from each extract were run per lane on an SDS polyacrylamide gel, with bacterially produced DHDPS or AKIII serving as a size standard and protein extracted from untransformed seeds serving as a negative control. The proteins were then electrophoretically blotted onto a nitrocellulose membrane. The membranes were exposed to the DHDPS or AKIII antibodies at a 1:5000 dilution of the rabbit serum using standard protocol provided by BioRad with their Immun-Blot Kit. Following rinsing to remove unbound primary antibody the membranes were exposed to the secondary antibody, donkey anti-rabbit Ig conjugated to horseradish peroxidase (Amersham) at a 1:3000 dilution. Following rinsing to remove unbound secondary antibody, the membranes were exposed to Amersham chemiluminescence reagent and X-ray film. To measure free amino acid composition of the seeds, they were extracted

from 40 mg of the defatted meal according to the method of Bieleski and Turner[20]. Samples were hydrolyzed in 6N hydrochloric acid, 0.4% β-mercaptoethanol under nitrogen for 24 h at 110–120°C, and run on a Beckman Model 6300 amino acid analyzer using post-column ninhydrin detection. Soybean meal was prepared by grinding a fragment from individual seeds into a fine powder. Total proteins were extracted from the meal by adding 1 mg to 0.1 ml of 43 mM Tris-HCl pH 6.8, 1.7% SDS, 4.2% β-mercaptoethanol, 8% glycerol, vortexing the suspension, boiling for 2–3 minutes and vortexing again. The resultant suspensions were centrifuged for 5 min at room temperature in a microfuge to remove particulates and 10 µl from each extract were run per lane on an SDS polyacrylamide gel, and analyzed as described above. Soybean free amino acids were extracted from 8–10 mg of the meal as described above; 1/10 of the sample was used for analysis. To identify seeds that expressed β-glucuronidase (GUS) the transformation marker, a small chip of the seed was cut off with a razor and put into a well in a disposable plastic microtiter plate. A GUS assay mix consisting of 100 mM NaH_2PO_4, 10 mM EDTA, 0.5 mM $K_4Fe(CN)_6$, 0.1% Triton X-100, 0.5 mg/ml 5-bromo-4-chloro-3-indolyl β-D-glucuronic acid was prepared and 0.15 ml was added to each microtiter well. The microtiter plate was incubated at 37° for 45 minutes. The development of blue color indicated the expression of GUS. To measure the total amino acid composition of mature seeds, 1–2 mg of the defatted meal (canola) or non-defatted meal (soybean) were hydrolyzed in 6N hydrochloric acid, 0.4% β-mercaptoethanol under nitrogen for 24 h at 110–120°C; 1/50–1/100 of the sample was run on a Beckman Model 6300 amino acid analyzer using post-column ninhydrin detection.

Acknowledgments

We thank Jerry Slightom and Roger Drong, Upjohn Company, for providing the seed expression cassette prior to publication. We thank Enno Krebbers for help on the manuscript. C.S. thanks Dr. Sherry Kitto, University of Delaware, for guidance and suggestions.

References

1. Negrutiu, I., Cattoir-Reynearts, A., Verbruggen, I. and Jacobs, M. 1984. Lysine overproducer mutants with altered dihydrodipicolinate synthase from protoplast culture of *Nicotiana sylvestris* (Spegazzini and Comes). Theor. Appl. Genet 68:11–20.
2. Glassman, K. F., Barnes, L. and Pilacinski, W. P. 1988. Method of inducing lysine overproduction in plants. PCT Patent Appl., Int. Pub. No. W0 89/11789.
3. Shaul, O. and Galili, G. 1992. Increased lysine synthesis in tobacco plants that express high levels of bacterial dihydrodipicolinate synthase in their chloroplasts. Plant Jour. 2:203–209.
4. Galili, G. and Perl, A. 1992. Transgenic plants overproducing threonine and lysine. EPO Patent Appl., Pub. No. 0 485 970 A2.
5. Falco, S. C., Keeler, S. J. and Rice, J. A. 1993. Chimeric genes and methods for increasing the lysine and threonine content of the seeds of plants. PCT Patent Appl., Int. Pub. Number WO 93/19190.
6. Shaul, O. and Galili, G. 1993. Concerted regulation of lysine and threonine synthesis in tobacco plants expressing bacterial feedback-insensitive aspartate kinase and dihydrodipicolinate synthase. Plant Mol. Biol. 23:759–768.
7. Karchi, H., Shaul, O. and Galili, G. 1994. Lysine synthesis and catabolism are coordinately regulated during tobacco seed development. Proc. Natl. Acad. Sci. USA 91:2577–2581.
8. Bonnassie, S., Oreglia, J. and Sicard, A. M. 1990. Nucleotide sequence of the *dapA* gene from *Corynebacterium glutamicum*. Nucleic Acids Res. 18:6421.
9. Cassan, M., Parsot, C., Cohen, G. N. and Patte, J.-C. 1986. Nucleotide sequence of *lysC* gene encoding the lysine-sensitive aspartokinase III of *Escherichia coli* K12. J. Biol. Chem. 261:1052–1057.
10. Kohara, Y., Akiyama, K. and Isono, K. 1987. The physical map of the whole *E. coli* chromosome: application of a strategy for rapid analysis and sorting of a large genomic library. Cell 50:595–508.
11. Berry-Lowe, S. L., McKnight, T. D., Shah, D. M. and Meagher, R. B. 1982. The nucleotide sequence, expression, and evolution of one member of a multigene family encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean. J. Mol. Appl. Genet. 1:483–498.
12. Slightom, J. L., Sun, S. M. and Hall, T. C. 1983. Complete nucleotide sequence of a French bean storage protein gene: phaseolin. Proc. Natl. Acad. Sci. USA 80:1897–1901.
13. Nigan, S. N. and McConnell, W. B. 1963. Studies on wheat plants using carbon^{14} compounds. XIX. Observations on the metabolism of lysine C^{14}. Can. J. Biochem. Physiol. 41:1367–1371.
14. Jefferson, R. A. 1987. Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol. Biol. Rep. 5:387–405.
15. Theze, J., Margarita, D., Cohen, G. N., Borne, F. and Patte, J. C. 1974. Mapping of the structural genes of the three aspartokinases and of the two homoserine dehydrogenases of *Escherichia coli*. J. Bacteriol. 117:133–143.
16. Slightom, J. L., Drong, R. F., Sieu, L. C. and Chee, P. C. 1991. Custom polymerase chain reaction engineering plant expression vectors and genes for plant expression. Plant Mol. Biol. Man. B16:1–55.
17. Ruvkun, G. B. and Ausubel, F. M. 1981. A general method for site directed mutagenesis in prokaryotes. Nature 289:85–88.
18. Hockema, A., Hirsch, P. R., Hooykaas, P. J. and Schilperoot, R. A. 1983. A binary plant vector strategy based on separation of *vir*- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature 303:179–180.
19. Christou, P., Swain, W. F., Yang, N.-S. and McCabe, D. E. 1989. Inheritance and expression of foreign genes in transgenic soybean plants. Proc. Natl. Acad. Sci. USA 86:7500–7504.
20. Bieleski, R. L. and Turner, N. A. 1966. Separation and estimation of amino acids in crude plant extracts ny thin-layer electrophoresis and chromatography.

Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 4803–4807, August 1983
Genetics

Expression of bacterial genes in plant cells

(plant protoplasts/transformation/foreign DNA/antibiotic resistance/selectable markers)

ROBERT T. FRALEY, STEPHEN G. ROGERS, ROBERT B. HORSCH, PATRICIA R. SANDERS, JEFFERY S. FLICK, STEVEN P. ADAMS, MICHAEL L. BITTNER, LESLIE A. BRAND, CYNTHIA L. FINK, JOYCE S. FRY, GERALD R. GALLUPPI, SARAH B. GOLDBERG, NANCY L. HOFFMANN, AND SHERRY C. WOO

Monsanto Company, 800 North Lindbergh Boulevard, St. Louis, Missouri 63167

Communicated by Howard A. Schneiderman, April 25, 1983

ABSTRACT Chimeric bacterial genes conferring resistance to aminoglycoside antibiotics have been inserted into the *Agrobacterium tumefaciens* tumor-inducing (Ti) plasmid and introduced into plant cells by *in vitro* transformation techniques. The chimeric genes contain the nopaline synthase 5′ and 3′ regulatory regions joined to the genes for neomycin phosphotransferase type I or type II. The chimeric genes were cloned into an intermediate vector, pMON120, and inserted into pTiB6S3 by recombination and then introduced into petunia and tobacco cells by cocultivating *A. tumefaciens* cells with protoplast-derived cells. Southern hybridization was used to confirm the presence of the chimeric genes in the transformed plant tissues. Expression of the chimeric genes was determined by the ability of the transformed cells to proliferate on medium containing normally inhibitory levels of kanamycin (50 μg/ml) or other aminoglycoside antibiotics. Plant cells transformed by wild-type pTiB6S3 or derivatives carrying the bacterial neomycin phosphotransferase genes with their own promoters failed to grow under these conditions. The significance of these results for plant genetic engineering is discussed.

The transformation of plant cells by virulent strains of *Agrobacterium tumefaciens* has been studied extensively by several laboratories (1–4). A small fragment of the tumor-inducing (Ti) plasmid, called transferred DNA (T-DNA), is known to be transferred to and stably incorporated in the nuclear DNA of transformed plant cells (5–7). The T-DNA is actively transcribed in plant cells (8–10) and specific gene products have been shown to be responsible for the observed phytohormone-independent growth characteristics (11, 12) and novel metabolic capacities (13) exhibited by crown gall tumor cells. The transfer and insertion of T-DNA into plant DNA is thought to involve repeated nucleotide sequences present near the T-DNA "borders" (14, 15) as well as other genes of unknown function located in specific virulence regions outside of T-DNA (16, 17).

In spite of our considerable understanding of the *A. tumefaciens*-Ti plasmid system, several problems remain which limit its use as a vector for genetically modifying higher plants. Because of the high levels of phytohormones produced by crown gall tumor cells (18) they have generally proven recalcitrant to attempts to induce regeneration into whole plants (19, 20). Exceptions to this are cases in which, as a result of aberrant integration or spontaneous deletion events, transformed cells have lost all or part of the Ti plasmid tumor genes and can now be regenerated (21, 22). In addition, transformation of cells by weakly virulent, mutant Ti plasmids (23) and transformation by root-inducing (Ri) plasmids (24, 25) have been shown to produce callus that can be regenerated into whole plants. However, these plants often display morphological aberrations and may retain certain tumorous properties (26). Another obstacle has been the failure to obtain expression from a variety of foreign genes that have been introduced into plants (23, 27). Reasons for this include the fact that, up to now, most studies have utilized either heterologous genes from bacteria, fungi, and mammalian cells whose regulatory regions may not be recognized by the plant RNA polymerases or highly regulated plant genes which are normally expressed in specialized tissues and which may not be transcribed in undifferentiated crown gall tumor tissue.

To bypass the dependence on tumor genes for identifying transformed plant cells and to overcome the barriers to gene expression in plants, chimeric genes that function as dominant selectable markers have been assembled. These contain the neomycin phosphotransferase (NPTase) coding sequences from the bacterial transposons Tn*5* (type II) or Tn*601* (type I) joined to the 5′ and 3′ regulatory regions of the nopaline synthase gene from the Ti plasmid. This paper describes the construction of these chimeric genes and their introduction and expression in plant cells.

MATERIALS AND METHODS

DNA Preparation. Plasmid pBR322 and its derivatives or M13 replicative form DNAs were purified by using either a Triton-X-100/CsCl procedure (28) or a large-scale alkaline lysis procedure (29), followed by purification on hydroxylapatite (30).

DNA fragments were isolated by electroelution into dialysis bags after polyacrylamide gel electrophoresis and band excision or by adsorption onto NA-45 DEAE membrane (Schleicher & Schuell) after agarose gel electrophoresis (31).

The *Bam*HI synthetic DNA linkers (5′ C-C-G-G-A-T-C-C-G-G) were purchased from Collaborative Research (Waltham, MA). Other synthetic DNAs were synthesized by using a modification of the phosphite procedure (32).

Enzymes. All restriction endonucleases and the large Klenow fragment of DNA polymerase I were obtained from New England Biolabs or Bethesda Research Laboratories and were used according to the instructions of the supplier. Phage T4 DNA ligase was prepared as in ref. 33. DNA fragment assembly was carried out as described (31).

***Transformation of Escherichia coli* Cells.** Plasmid DNAs were introduced into *E. coli* cells by using $CaCl_2$-treated or $RuCl_3$-treated cells (31). The recipient *E. coli* K-12 strains were SR200 = C600 *thr pro recA56 hsdR*(r^-m^+) (34); LE392 = ED8554 *hsdR* (r^-m^+) (31); SR20 = GM42 = *his dam*-3 (35); and the M13

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s); NPTase I and NPTase II, neomycin phosphotransferase, types I and II, respectively; Ti plasmid, tumor-inducing plasmid; T-DNA, transferred DNA; Ri plasmid, root-inducing plasmid.

phage host, JM101 (36). Cells carrying recombinant plasmids were selected or grown (or both) on Luria medium plates or broth at 37°C containing appropriate antibiotics (ampicillin, 200 μg/ml; spectinomycin, 50 μg/ml; and kanamycin, 40 μg/ml).

Introduction of pMON120 Derivatives into *A. tumefaciens*. Plasmid pMON120 or its derivatives were transferred to a chloramhpenicol-resistant *A. tumefaciens* strain GV3111 = C58C1 Cm^R carrying pTiB6S3tra^c (37) by using a triparental plate mating procedure (ref. 38; R. Riedel, personal communication). Briefly, 0.2 ml of a fresh overnight culture of LE392 carrying pMON120 or its derivative was mixed with 0.2 ml of an overnight culture of HB101 (31) carrying the pRK2013 (38) plasmid and 0.2 ml of an overnight culture of GV3111 cells. The mixture of cells was spread on an LB plate and incubated for 16–24 hr at 30°C to allow plasmid transfer and recombination. The cells were resuspended in 3 ml of 10 mM $MgSO_4$ and a 0.2-ml aliquot was then spread on an LB plate containing 25 μg of chloramphenicol per ml and 100 μg each of spectinomycin and streptomycin per ml to select *A. tumefaciens* carrying the pMON120 derivatives. After incubation for 48 hr at 30°C, ≈10 colonies per plate were obtained. Control matings between HB101/pRK2013 cells and GV3111 cells never gave rise to colonies after this selection. Typically, one colony was chosen and grown at 30°C in LB medium containing chloramphenicol, spectinomycin, and streptomycin at the same concentrations given above.

Protoplast Isolation and Culture. "Mitchell" petunia plants were grown in environmental chambers under fluorescent and incandescent illumination (≈5,000 lux, 12 hr/day) at 21°C in a 50:50 mixture of vermiculite and Pro-mix BX (Premier Brands, Quebec, PQ, Canada). Leaves were surface sterilized, cut into 2-mm strips, and enzymatically digested as described (39). The resulting protoplasts were purified by passage through stainless steel meshes and by density floatation as described (39). The protoplasts were plated in tissue culture flasks (T75, Falcon; 6 ml per flask) at a cell density of 10^5 cells per ml in culture medium [MS salts (GIBCO), B-5 vitamins, 3% (wt/vol) sucrose, 9% (wt/vol) mannitol, 1 μg of 2,4-D per ml, and 0.5 μg of benzyl adenine per ml, pH 5.7].

Cocultivation of *A. tumefaciens* Cells with Plant Protoplasts. On day 2 after protoplast isolation, aliquots (10–50 μl) of an overnight culture of *A. tumefaciens* cells were added to each flask (final bacterial cell density = 10^8 cells per ml) and cocultivation with plant cells was carried out for 24–30 hr essentially as described (40). On day 3, 6 ml of culture medium (lacking phytohormones and mannitol) containing carbenicillin (1.5 mg/ml) was added to each flask (final concentration = 500 μg/ml) to prevent further bacterial growth. On day 4, an additional 6 ml of the above medium (containing carbenicillin at 500 μg/ml) was added. On day 6, 0.5 ml of the cell mixture was transferred to and spread in a thin layer on the surface of double-filter feeder plates (41). These consisted of agar medium (MS salts, B-5 vitamins, 3% sucrose, 3% mannitol, 0.1 μg of indole acetic acid per ml, and 500 μg of carbenicillin per ml at pH 5.7), a layer of *Nicotinia tabacum* suspension cells, a tight fitting 8.5-cm Whatman filter paper disc (guard disc), and a 7.0-cm Whatman filter paper disc (transfer disc). After 7–10 days, microcolonies (≈0.5 mm) were observable on the feeder plates and the transfer disc was removed and placed on selection medium (MS salts, B-5 vitamins, 3% sucrose, 500 μg of carbenicillin per ml at pH 5.7) lacking phytohormones. Within 2 wk, hormone-independent transformants could be readily distinguished as green colonies against a background of dying, brown nontransformed cells. The transformation frequency in these experiments was ≈10^{-1}. The hormone-independent transformants were then transferred to medium (MS salts, B-5 vitamins, 3% sucrose, 500 μg of carbenicillin per ml at pH 5.7) containing kanamycin (50 μg/ml).

Analysis of Transformants. Octopine and nopaline synthase activities were determined as in ref. 42 with the substitution of [^{14}C]arginine (Amersham, 0.5 μCi/2.5-μl assay; 1 Ci = 3.7 × 10^{10} Bq) for the unlabeled arginine in the assay buffer. The conditions for electrophoresis were as described (42) and the resulting electrophoretograms were exposed to x-ray film (Kodak, XAR-5) for 16–24 hr. The positions of octopine, nopaline, and arginine were established by their comigration with authentic standards.

Callus for NPTase assays were frozen in liquid N_2 and extracted by using a mortar and pestle in a minimal volume of buffer (0.2 M Tris·HCl/2 mM EDTA/7.5% polyvinylpolypyrrolidone). The crude extract was clarified by centrifugation (Eppendorf; Brinkmann) and assays were performed as described (43).

RESULTS

NPTase coding sequences were used in the initial chimeric gene constructions described in this study because plant cells were determined to be sensitive to various aminoglycoside antibiotics (unpublished data), and the expression of NPTase in yeast (43) and mammalian cells (44, 45) has been previously shown to confer resistance to the antiobiotic, G418. The nopaline synthase gene promoter and 3′-nontranslated regions were selected because this gene has been well characterized (9, 46) and it is known to be expressed constitutively in most plant tissues transformed with the *A. tumefaciens* Ti plasmid (47).

Construction of Chimeric Genes. The nopaline synthase promoter region, obtained on a 350-base-pair (bp) *Sau*3A fragment from the *Hin*dIII-23 fragment of pTiT37 (Fig. 1; ref. 46), was engineered to remove the entire nopaline synthase coding sequence. The resulting promoter fragment that extends from base −264 to base 35 of the nopaline synthase sequence (46) was positioned next to the *Bgl* II site located just outside the NPTase II coding sequence (49). In addition, a 260-bp *Mbo* I fragment, extending from base 1,297 to base 1,554 of the published nopaline synthase sequence (46), was isolated from the *Hin*dIII-23 fragment. This *Mbo* I fragment contains the nopaline synthase 3′-nontranslated region and polyadenylylation site. This fragment was ligated together with the *Eco*RI–*Bam*HI fragment that contained the nopaline synthase promoter and NPTase II structural gene to yield the intact chimeric gene on a 1.5-kilobase (kb) *Eco*RI fragment (Fig. 1). A second chimeric gene, containing the nopaline synthase promoter and 3′-nontranslated region joined to the NPTase I coding sequence (Fig. 2), was constructed in a similar fashion. As controls, plasmids were constructed that contained an intact NPTase II promoter and structural sequence with the nopaline synthase 3′-nontranslated region (pMON139 and pMON140; Fig. 2).

Introduction of Chimeric Genes into the Ti Plasmid. The vector pMON120 used for the transfer of the chimeric genes into *A. tumefaciens* cells is shown in Fig. 2. Its essential features include (*i*) a segment of pBR322 DNA for replication in *E. coli*, (*ii*) a segment from pTiT37 that contains a functional nopaline synthase gene to facilitate the rapid identification of transformants, (*iii*) a segment of Tn7 carrying the spectinomycin/streptomycin-resistance determinant for selection in *A. tumefaciens*, (*iv*) a DNA segment obtained from the pTiA6 T-DNA fragment *Hin*dIII-18c (see T-DNA map, ref. 11), which is included to provide homology for recombination with a resident octopine-type Ti plasmid in *A. tumefaciens*, and (*v*) unique restriction sites (*Eco*RI and *Hin*dIII) for insertion of the chimeric genes. The pMON120 plasmid and derivatives were in-

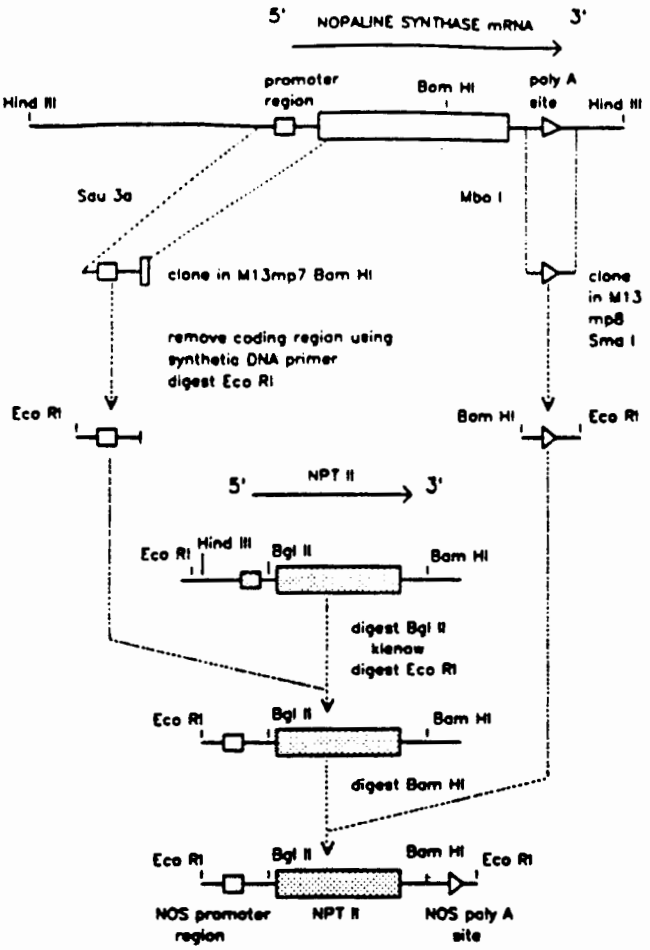


FIG. 1. Isolation and assembly of a chimeric gene containing nopaline synthase (NOS) promoter-NPTase II (NPT II) coding sequence-nopaline synthase 3'-nontranslated region. The nopaline synthase promoter was isolated on a 350-bp *Sau*3A fragment that also contained the first 44 bp of the nopaline synthase coding sequence. The sense strand of this fragment was cloned into the *Bam*HI site in M13 mp7 (36), the 44 bp was removed by using a modification of a published synthetic primer procedure (48) with a primer complementary to bases 22–35 of the published nopaline synthase sequence (46), and a 308-bp promoter fragment was obtained after digestion with *Eco*RI. The flush-end of the promoter fragment was joined to a 1-kb *Bgl* II–*Bam*HI fragment carrying the NPTase II coding sequence (a *Bam*HI linker had been inserted at the *Sma* I site) at the filled-in *Bgl* II site (49). This fusion regenerates the *Bgl* II site. The chimeric gene was completed by the addition of a 260-bp *Mbo* I fragment that contained the nopaline synthase 3'-nontranslated region. This fragment, which contains a polyadenylylation signal (46), was converted to a flush-ended fragment with Klenow polymerase and cloned into the *Sma* I site of a M13 mp8 (50) to introduce *Bam*HI and *Eco*RI sites at the 5' and 3' ends, respectively. The resulting 280-bp fragment was joined to the 1,300-bp *Eco*RI–*Bam*HI nopaline synthase promoter-NPTase II coding sequence fragment to generate the complete chimeric gene.

duced into *A. tumefaciens* as described in *Materials and Methods*.

Selection of Kanamycin-Resistant Petunia Transformants. Several hundred hormone-independent calli (1–2 mm in diameter) obtained from cocultivation experiments with *A. tumefaciens* strains carrying pTiB6S3::pMON120 (or derivatives) recombinant plasmids were pooled and analyzed by DNA blot hybridization for the presence of the chimeric genes (Fig. 3). The results confirm the presence of the expected 1.6-kb *Eco*RI fragment, which carries the chimeric nopaline synthase-NPTase II-nopaline synthase gene in pMON128 and pMON129 transformants, and the control NPTase II-NPTase II-nopaline synthase construct in pMON139 and pMON140 transformants (Fig. 3*a*).

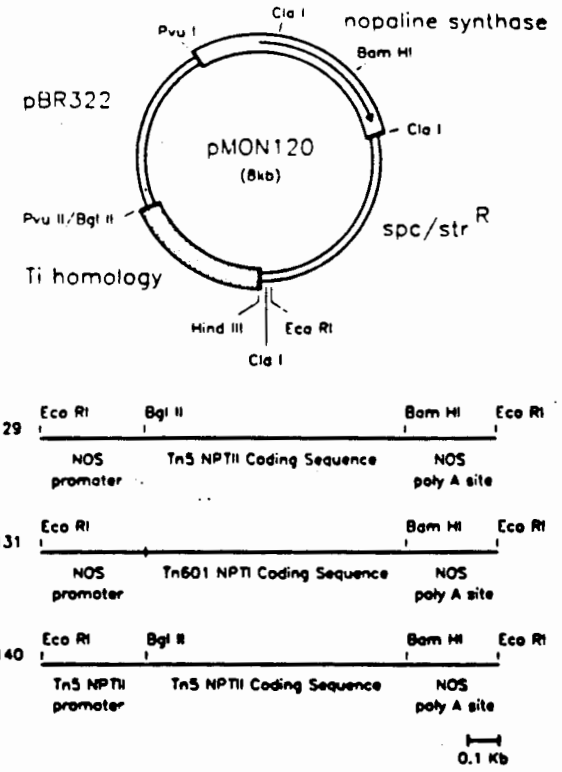


FIG. 2. Structures of the pMON120 intermediate vector and chimeric genes introduced into plant cells. Plasmid pMON120 contains the following segments of DNA: the 1.7-kb pBR322 *Pvu* II to *Pvu* I fragment that carries the origin of replication and *bom* site (51), a 2.2-kb partial *Cla* I to *Pvu* I fragment of pTiT37 DNA that encodes an intact nopaline synthase (NOS) gene, a 2.7-kb *Cla* I-*Eco*RI fragment of Tn7 (37) DNA carrying the determinant for spectinomycin/streptomycin resistance, and the 1.6-kb *Hind*III–*Bgl* II fragment from the *Hind*III-18c fragment of the pTiA6 plasmid. This T-DNA fragment is known to specify two transcripts that are not essential for tumorous growth (8, 12). At the bottom are three chimeric genes inserted at the unique *Eco*RI site of pMON120. The chimeric nopaline synthase-NPTase II-nopaline synthase gene was inserted to give pMON129 and pMON128. In all of these examples, the first plasmid carries the inserted gene as it is drawn in the figure. The second plasmid carries the insert in the opposite orientation to that drawn. Plasmids pMON131 and pMON130 carry a chimeric nopaline synthase-NPTase I-nopaline synthase gene. The final chimeric gene is carried in plasmids pMON140 and pMON139. The bacterial NPTase II promoter and coding sequence have been joined to the nopaline synthase 3'-nontranslated region.

Similar results were obtained for pMON130 and pMON131 transformants, which contain the chimeric nopaline synthase-NPTase I-nopaline synthase gene on a 1.5-kb *Eco*RI fragment (Fig. 3*b*). No hybridization with either the Tn*5*- or Tn*601*-specific probe was detected in transformants containing only the pMON120 vector. Other minor bands of hybridization are present in the pMON129 and pMON140 transformants; these may be attributable to partial digestion or aberrant integration events and their assignment awaits further analysis of clonal tissue. Blot hybridization analysis of DNA from these transformants using T-DNA-specific probes confirmed the presence of the expected internal T-DNA fragments in the transformed tissues and ruled out any possibility that the plant tissue was contaminated by *A. tumefaciens* cells (data not shown).

Other transformed, hormone-independent calli from these experiments were transferred to agar medium containing kana-

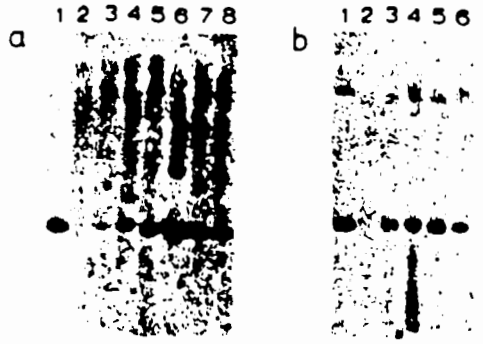


FIG. 3. DNA blot hybridization analysis of *in vitro* transformants. Several hundred hormone-independent *in vitro* transformants from each experiment were pooled and total DNA was extracted (52). The DNAs were digested with *Eco*RI and the fragments were separated by electrophoresis and transferred to nitrocellulose (53). (*a*) Hybridization with NPTase II-specific probe. A gel-purified 3.3-kb *Hin*dIII fragment from Tn*5* (54) was used as probe. Lane 1, pMON128::Ti plasmid marker; lane 2, pMON120 transformants; lane 3, pMON139 transformants; lane 4, pMON140 transformants; lane 5, pMON128 transformants; lane 6, pMON129 transformants; lane 7, pMON128 transformants; and lane 8, pMON129 transformants. Lanes 2–6 represent transformants selected for hormone-independent growth prior to scoring for kanamycin resistance; lanes 7 and 8 represent transformants selected only for kanamycin resistance on medium containing phytohormones. (*b*) Hybridization with NPTase I-specific probe. A gel-purified 1.2-kb *Ava* II fragment from Tn*601* (55) was used as a probe. Lane 1, pMON130::Ti plasmid marker; lane 2, pMON120 transformants; lane 3, pMON130 transformants; lane 4, pMON131 transformants; lane 5, pMON130 transformants; and lane 6, pMON131 transformants. Lanes 2–4 represent transformants selected for hormone-independent growth prior to scoring for kanamycin resistance; lanes 5 and 6 represent transformants selected only for kanamycin resistance on medium containing phytohormones.

mycin (50 μg/ml) and these were scored after 2–3 wk for resistance to the antibiotic. All transformants obtained from experiments utilizing pMON120, pMON139, or pMON140 failed to grow on medium supplemented with kanamycin, whereas all the transformants from experiments utilizing pMON128, pMON129, pMON130, or pMON131 grew on medium containing the antibiotic at rates comparable to growth on normal medium. A quantitative assessment of the level of resistance conferred by the chimeric genes is shown for pMON120, pMON129, and pMON131 (Fig. 4). The results are based on the net growth of independent transformants on medium containing the levels of antibiotic shown in the figure, compared to growth in the absence of antibiotics. It is apparent that transformants containing the chimeric nopaline synthase-NPTase II-nopaline synthase gene (pMON129) require ≈20-fold higher levels of kanamycin to depress net growth by 50% in comparison to transformants lacking the chimeric gene (pMON120). Similar results were obtained for pMON128, which contains the chimeric gene in the opposite orientation in the pMON120 vector (not shown). Transformants containing pMON139 and pMON140 have dose responses identical to pMON120. Transformants containing pMON130 or pMON131 (chimeric nopaline synthase-NPTase I-nopaline synthase gene) are less resistant to kanamycin than those containing pMON128 or pMON129 (results shown for pMON130). However, this level of resistance (≈3-fold greater than control cells) is still quite adequate for selection (see below).

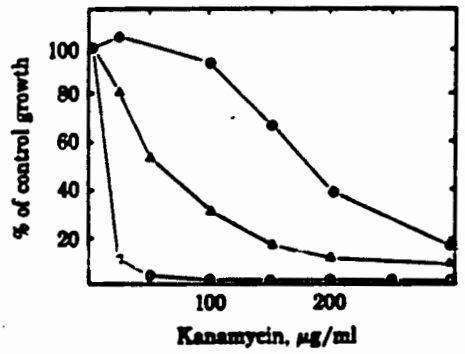


FIG. 4. Growth of transformants at various antibiotic concentrations. *In vitro* transformants were obtained after cocultivation with *A. tumefaciens* strains carrying cointegrate pMON120, pMON129, or pMON131. Hormone-independent calli (1- to 2-mm diameter) from each experiment were transferred to plates (16 calli per plate) containing the antibiotic concentration shown. After 3 wk, the net growth (wet weight) at each antibiotic concentration was determined and the results were expressed as the % of control growth (growth in the absence of antibiotics). ●, pMON129 transformants; △, pMON131 transformants; and ○, pMON120 transformants.

Additional cocultivation experiments were carried out without hormone-independent selection (i.e., medium supplemented with phytohormones which support the growth of nontransformed cells). The resulting microcolonies (≈1 mm) were transferred to phytohormone-supplemented medium containing kanamycin (50 μg/ml) and within 2–3 wk, growing colonies were readily observable on plates containing cells that were transformed with pMON128, pMON129, pMON130, or pMON131. The frequency of transformation obtained by using antibiotic selection was comparable to that obtained by using hormone-independent selection. Opine (data not shown) and Southern hybridization analysis (Fig. 3*a*, lanes 7 and 8; Fig. 3*b*, lanes 5 and 6) of the kanamycin-resistant colonies confirmed that they were indeed transformants. No growing colonies were observable on plates containing cells transformed by pMON120, pMON139, or pMON140 plasmids.

DISCUSSION

The expression of the prokaryotic NPTase I and NPTase II enzymes in plant cells by using the intermediate vector pMON120 probably depends on transcription from the nopaline synthase promoter. Support for this comes from the facts that (*i*) the prokaryotic genes with their own promoters do not confer antibiotic resistance to petunia cells (Figs. 2 and 4) and (*ii*) all of the constructions function identically in either orientation in the pMON120 vector, suggesting that transcription does not initiate elsewhere in the vector. RNA blot hybridization experiments have confirmed the presence of NPTase II-specific mRNA in the transformed tissues and nuclease S1 mapping experiments demonstrate the expected 5′ and 3′ ends for the chimeric NPTase II mRNA (data not shown). In addition, low levels of neomycin-dependent NPTase II activity have been reproducibly observed in crude cell extracts from tissues transformed with pMON128 or pMON129 (no activity has been detected in extracts from control cells or cells transformed with pMON120, pMON139, or pMON140).

The useful range of these chimeric antibiotic resistance genes appears to be quite broad. In addition to the results presented for petunia, successful selection of aminoglycoside-resistant transformants has also been demonstrated for tobacco, sunflower, and carrot (results not shown). It seems likely that most plants within the host range of *A. tumefaciens* could be transformed and identified in this manner. Those plant cells that are not particularly sensitive to kanamycin may be killed by other aminoglycoside antibiotics. In this respect pMON128 (or pMON129) and pMON130 (or pMON131) also function to confer resistance to G418 and neomycin on petunia, carrot, sun-

flower, and tobacco (unpublished data).

The availability of dominant selectable markers on small plasmids such as pMON120 should facilitate the development of alternate, non-*A. tumefaciens*-mediated methods for transforming plant cells such as spheroplast fusion (56) or the use of liposomes (57) or calcium-phosphate (58) techniques. These chimeric genes should also prove useful as markers in somatic hybridization experiments or as sensitive probes for studying promoter function. Finally, two obvious but significant aspects of the results presented in this paper are (*i*) it should now be possible, by using Ti plasmids that have the tumor genes (i.e., tms and tmr loci, 12) deleted, to obtain kanamycin-resistant transformants that can be readily and reproducibly regenerated into phenotypically normal plants, and (*ii*) there is no reason to believe that NPTase I and NPTase II are unique in their ability to be expressed in plant cells and it is quite likely that other bacterial, fungal, or mammalian genes, including those whose products could be expected to modify plant properties in a useful manner, could also be successfully engineered and expressed.

We gratefully acknowledge the contributions and strains provided by Drs. M.-D. Chilton and J. Schell. We also thank Dr. M. Bevan for letting us compare our sequence data on the nopaline synthase gene with his prior to its publication, Dr. S. Gelvin for T-DNA probes, P. Kelly for helpful comments on the manuscript, and Ms. D. Lam and P. Guenther for preparing the manuscript. Finally, we would like to thank Dr. E. Jaworski for his support and encouragement throughout the course of this project.

1. Chilton, M.-D., Drummond, M. H., Merlo, D. J., Sciaky, D., Montoya, A. L., Gordon, M. & Nester, E. (1977) *Cell* 11, 263–271.
2. Van Larebeke, N., Engler, G., Holsters, M., Van der Elsacker, S., Zaenen, I., Schilperoort, R. & Schell, J. (1974) *Nature (London)* 252, 169–170.
3. Kerr, A., Manigault, P. & Tampé, J. (1977) *Nature (London)* 265, 560–561.
4. Braun, A. (1956) *Cancer Res.* 16, 53–56.
5. Chilton, M.-D., Saiki, R., Yadav, N., Gordon, M. & Quétier, F. (1980) *Proc. Natl. Acad. Sci. USA* 77, 4060–4064.
6. Yadav, N., Postle, K., Saiki, R., Thomashow, M. & Chilton, M.-D. (1980) *Nature (London)* 287, 458–461.
7. Willmitzer, L., DeBeuckeleer, M., Lemmers, M., Van Montagu, M. & Schell, J. (1980) *Nature (London)* 287, 359–361.
8. Willmitzer, L., Simons, G. & Schell, J. (1982) *EMBO J.* 1, 139–146.
9. Bevan, M. & Chilton, M.-D. (1982) *J. Mol. Appl. Genet.* 1, 539–546.
10. Gelvin, S., Gordon, M., Nester, E. & Aronson, A. (1981) *Plasmid* 6, 17–29.
11. Leemans, J., Deblaere, R., Willmitzer, L., DeGreve, H., Hernalsteens, J., Van Montagu, M. & Schell, J. (1982) *EMBO J.* 1, 147–152.
12. Garfinkel, D., Simpson, R., Ream, R., White, F., Gordon, M. & Nester, E. (1981) *Cell* 27, 143–153.
13. Holsters, M., Silva, B., Van Vliet, F., Genetello, C., DeBlock, M., Dhaese, P. Depicker, A., Inze, D., Engler, G., Villarael, R., Van Montagu, M. & Schell, J. (1980) *Plasmid* 3, 212–230.
14. Zambryski, P., Depicker, A., Kruger, K. & Goodman, H. (1982) *J. Mol. Appl. Genet.* 1, 361–370.
15. Yadav, N., Vanderleyden, J., Bennet, D., Barnes, W. & Chilton, M.-D. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6322–6326.
16. Hille, J., Klasen, I. & Schilperoort, h. (1982) *Plasmid* 7, 107–118.
17. Klee, H., Gordon, M. & Nester, E. (1982) *J. Bacteriol.* 150, 327–331.
18. Akiyoski, D., Morris, R., Hinz, R., Mischke, B., Kosuge, T., Garfinkel, D., Gordon, M. & Nester, E. (1983) *Proc. Natl. Acad. Sci. USA* 80, 407–411.
19. Braun, A. & Wood, H. (1976) *Proc. Natl. Acad. Sci. USA* 73, 496–500.
20. Yang, F., Montoya, A., Merlo, D., Drummond, H., Chilton, M.-D., Nester, E. & Gordon, M. (1980) *Mol. Gen. Genet.* 177, 707–714.
21. Otten, L., DeGreve, H., Hernalsteens, J., Van Montagu, M., Schieder, O., Straub, J. & Schell, J. (1981) *Mol. Gen. Genet.* 183, 209–213.
22. Wullems, G., Molendijk, L., Ooms, G. & Schilperoort, R. (1981) *Cell* 24, 719–727.
23. Barton, K., Binns, A., Matzke, A. & Chilton, M.-D. (1983) *Cell* 32, 1033–1043.
24. Chilton, M.-D., Tepfer, D., Petit, A., David, C., Casse-Delbart, F. & Tempé, J. (1982) *Nature (London)* 295, 432–434.
25. White, F., Ghidossi, G., Gordon, M. & Nester, E. (1982) *Proc. Natl. Acad. Sci. USA* 79, 3193–3198.
26. Spano, L. & Costantino, P. (1982) Z. *Pflanzenphysiol.* 106, 87–92.
27. Chilton, M.-D., Bevan, M., Yadav, N., Matzke, A., Byrne, M., Grula, M., Barton, K., Vanderleyden, J., DeFramond, A. & Barnes, W. (1981) *Stadler Genet. Symp.* 13, 39–51.
28. Davies, R. W., Botstein, D. & Roth, J. R. (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 116.
29. Ish-Horowicz, D. & Burke, J. F. (1981) *Nucleic Acids Res.* 9, 2989–2998.
30. Colman, A., Bevers, M. J., Primrose, S. B. & Lyons, A. (1978) *Eur. J. Biochem.* 91, 303–310.
31. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 504.
32. Adams, S. P., Holder, S. B., Wykes, E. J., Kavka, K. S. & Galluppi, G. R. (1983) *J. Am. Chem. Soc.* 105, 661–663.
33. Murray, N. E., Bruce, S. A. & Murray, K. (1979) *J. Mol. Biol.* 132, 493–505.
34. Rogers, S. G. & Weiss, B. (1980) *Gene* 11, 187–195.
35. Bale, A., d'Alarcao, M. & Marinus, G. M. (1979) *Mutat. Res.* 59, 157–165.
36. Messing, J., Crea, R. & Seeburg, P. (1981) *Nucleic Acids Res.* 9, 309–321.
37. DeGreve, H., Decraemer, H., Seurinck, J., Van Montagu, M. & Schell, J. (1981) *Plasmid* 6, 235–248.
38. Ditta, G., Stanfield, S., Corbin, D. & Helinski, D. (1980) *Proc. Natl. Acad. Sci. USA* 77, 7347–7351.
39. Ausubel, F., Bahnsen, K., Hanson, M., Mitchell, A. & Smith, H. (1980) *Plant Mol. Biol. Newsl.* 1, 26–32.
40. Wullems, G., Molendijk, L., Ooms, G. & Schilperoort, R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4344–4348.
41. Horsch, R. & Jones, G. (1980) *In Vitro* 16, 103–108.
42. Otten, L. & Schilperoort, R. (1978) *Biochim. Biophys. Acta* 527, 497–500.
43. Jimenez, A. & Davis, J. (1980) *Nature (London)* 287, 869–871.
44. Colbere-Garapin, F., Horodniceanu, F., Kourilsky, P. & Garapin, A.-C. (1981) *J. Mol. Biol.* 150, 1–14.
45. Southern, P. & Berg, P. (1982) *J. Mol. Appl. Genet.* 1, 327–341.
46. Depicker, A., Stachel, S., Dhaese, P., Zambryski, P. & Goodman, H. (1982) *J. Mol. Appl. Genet.* 1, 561–574.
47. Tempé, J. & Goldmann, A. (1982) in *Molecular Biology of Plant Tumors*, eds. Kahl, G. & Schell, J. (Academic, New York), pp. 427–449.
48. Goeddel, D., Shepard, H., Yelverton, E., Leung, D., Crea, R., Sloma, A. & Pestka, S. (1980) *Nucleic Acids Res.* 8, 4057–4074.
49. Beck, E., Ludwig, G., Auerswald, E., Reiss, B. & Schaller, H. (1982) *Gene* 19, 327–336.
50. Messing, J. & Vieira, J. (1982) *Gene* 19, 269–276.
51. Covarrubias, L., Cervantes, L., Covarrubias, A., Soberon, X., Vichido, L., Blanco, A., Kuperztoch-Portnoy, Y. & Bolivar, F. (1981) *Gene* 13, 25–35.
52. Nagao, R., Shah, D., Eckenrode, V. & Meagher, R. (1981) *DNA* 1, 1–9.
53. Southern, E. (1975) *J. Mol. Biol.* 98, 503–517.
54. Berg, D. Davies, J., Allet, B. & Rochaix, J. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3628–3632.
55. Oka, A., Sugisaki, H. & Takanami, M. (1981) *J. Mol. Biol.* 147, 217–226.
56. Hasezawa, S., Nagata, T. & Syono, K. (1981) *Mol. Gen. Genet.* 182, 206–210.
57. Fraley, R. & Papahadjopoulos, D. (1982) in *Current Topics in Microbiology and Immunology*, eds. Hofschneider, P. & Goebel, W. (Springer, New York), pp. 171–192.
58. Krens, F., Molendijk, L., Wullems, G. & Schilperoort, R. (1982) *Nature (London)* 296, 72–74.

Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 8447–8451, November 1986
Biochemistry

β-Glucuronidase from *Escherichia coli* as a gene-fusion marker

(DNA sequence/*uidA* gene/reporter gene/enzyme purification)

RICHARD A. JEFFERSON*, SEAN M. BURGESS, AND DAVID HIRSH†

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309

Communicated by William B. Wood, June 23, 1986

ABSTRACT **We have developed a gene-fusion system based on the *Escherichia coli* β-glucuronidase gene (*uidA*). The *uidA* gene has been cloned from *E. coli* K-12 and its entire nucleotide sequence has been determined. β-Glucuronidase has been purified to homogeneity and characterized. The enzyme has a subunit molecular weight of 68,200, is very stable, and is easily and sensitively assayed using commercially available substrates. We have constructed gene fusions of the *E. coli lacZ* promoter and coding region with the coding region of the *uidA* gene that show β-glucuronidase activity under *lac* control. Plasmid vectors have been constructed to facilitate the transfer of the β-glucuronidase coding region to heterologous control regions, using many different restriction endonuclease cleavage sites. There are several biological systems in which *uidA*-encoded β-glucuronidase may be an attractive alternative or complement to previously described gene-fusion markers such as β-galactosidase or chloramphenicol acetyltransferase.**

The use of fusions between a gene of interest and a reporter gene with an easily detectable product offers several advantages for the study of gene expression. The use of a single set of assays to monitor the expression of diverse gene control regions simplifies analysis and often enhances the sensitivity with which measurements of gene activity can be made. Many genes in higher organisms are members of gene families consisting of several related genes whose expression may be independently controlled (1). It is often desirable to study the expression of one member of such a gene family free from the background of the other members of the family. The use of *in vitro*-generated gene fusions and DNA transformation permits such an analysis.

The most frequently used reporter gene is probably the *Escherichia coli lacZ* gene, which encodes a β-galactosidase (2, 3). β-Galactosidase has many features that make it attractive as a gene fusion marker. The gene and gene product are well characterized genetically and biochemically (3). There are sensitive assays for the enzyme that utilize commercially available substrates, including several that allow visualization of enzyme activity *in situ*. β-Galactosidase is not, however, ideal for all systems. There are several intensively studied biological systems in which endogenous β-galactosidase levels are high enough that it is difficult or impossible to detect chimeric β-galactosidase by enzymatic methods. In addition, the enzyme and gene are very large, sometimes making the *in vitro* construction and analysis of gene fusions unwieldy.

Another gene that has been used recently in the analysis of *in vitro*-generated gene fusions encodes a chloramphenicol acetyltransferase (CAT). There is very little endogenous CAT activity in most eukaryotic systems that have been studied, but quantitative enzyme assays are expensive, laborious, and complicated by the presence of endogenous esterases and there are no histochemical methods for analyzing the spatial distribution of enzyme activity in tissues (4).

Because of some of these limitations, we have developed a gene fusion system that uses the *E. coli* β-glucuronidase gene (*uidA*) as the reporter gene. β-Glucuronidase (β-D-glucuronoside glucuronosohydrolase, EC 3.2.1.31) is an acid hydrolase that catalyzes the cleavage of a wide variety of β-glucuronides. Substrates for β-glucuronidase are generally water-soluble, and due to the extensive analysis of mammalian glucuronidases (6), many substrates are commercially available, including substrates for spectrophotometric, fluorometric, and histochemical analyses. This ability to perform histochemical analysis of gene fusions is an important feature for the study of gene expression in metazoans and plants, where spatial discrimination is often essential for assessing the regulation of genes. Methods have been described that allow subcellular localization of glucuronidase activity (reviewed in ref. 7).

The *uidA* gene has been analyzed genetically and was shown by Novel and coworkers (8–11) to be the β-glucuronidase structural gene. Plasmid clones have been obtained that contain the *uidA* locus, and a partial DNA sequence of the *uidA* regulatory region has been published (12, 13).

MATERIALS AND METHODS

DNA Manipulation. Restriction endonucleases and DNA-modifying enzymes were obtained from New England Biolabs whenever possible and used per the instructions of the supplier. Plasmid DNA preparations were done by the method of Birnboim and Doly (14) as described by Maniatis *et al.* (15). Routine cloning procedures, including ligations and transformation of *E. coli* cells, were performed essentially as described (15). DNA fragments were purified from agarose gels by electrophoresis onto Schleicher & Schuell NA45 DEAE membrane (16) as recommended by the manufacturer. DNA sequences were determined by the dideoxy chain-terminator method of Sanger and Coulson (17), as modified by Biggin *et al.* (18). Oligodeoxynucleotide primers for sequencing and site-directed mutagenesis were synthesized using an Applied Biosystems (Foster City, CA) DNA synthesizer and were purified by preparative polyacrylamide gel electrophoresis. Site-directed mutagenesis was performed on single-stranded DNA obtained from pEMBL-derived plasmids, essentially as described (19). The strain used for routine manipulation of the *uidA* gene was RAJ201, a *recA* derivative of JM83 (20) generated by bacteriophage P1 transduction. Strain PK803 was obtained from P. Kuempel (University of Colorado at Boulder) and contains a deletion of the *manA–uidA* region. Plasmid vectors pUC7, -8, and -9 (20) and pEMBL-9 (21) have been described.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s).

*Present address: Department of Molecular Genetics, Plant Breeding Institute, Maris Lane, Trumpington, Cambridge, England CB2 2LQ.

†Present address: Synergen, Inc., 1885 33rd Street, Boulder, CO 80301.

Protein Sequencing and Amino Acid Analysis. Sequence analysis was performed by A. Smith (Protein Structure Laboratory, University of California, Davis), using a Beckman 890M spinning-cup sequenator. Amino acid composition was determined by analysis of acid hydrolysates of purified β-glucuronidase on a Beckman 6300 amino acid analyzer.

Protein Analysis. Protein concentrations were determined by the dye-binding method of Bradford (22), using a kit supplied by Bio-Rad Laboratories. $NaDodSO_4$/PAGE was performed using the Laemmli system (23).

β-Glucuronidase Assays. Glucuronidase was assayed in a buffer consisting of 50 mM sodium phosphate (pH 7.0), 10 mM 2-mercaptoethanol, 0.1% Triton X-100, and 1 mM *p*-nitrophenyl β-D-glucuronide. Reactions occurred in 1-ml volumes at 37°C and were terminated by the addition of 0.4 ml of 2.5 M 2-amino-2-methylpropanediol. *p*-Nitrophenol absorbance was measured at 415 nm. Routine testing of bacterial colonies for β-glucuronidase activity was done by transferring bacteria with a toothpick into microtiter wells containing the assay buffer. During the preparation of this paper, the histochemical substrate 5-bromo-4-chloro-3-indolyl β-D-glucuronide (analogous to the β-galactosidase substrate "X-Gal") became commercially available (Research Organics, Cleveland, OH). We found it to be an excellent and sensitive indicator of β-glucuronidase activity *in situ* when included in agar plates at a concentration of 50 μg/ml.

Purification of β-Glucuronidase. β-Glucuronidase was purified by conventional methods from the strain RAJ201 containing the plasmid pRAJ210 (see Fig. 1). Details of the method are available upon request (26).

RESULTS

Subcloning and Sequencing of the *uidA* Gene. The starting point for the subcloning and sequencing of the β-glucuronidase gene was the plasmid pBKuidA (Fig. 1). This plasmid has been shown to complement a deletion of the *uidA–manA* region of the *E. coli* chromosome (R. Bitner and P. Kuempel, personal communication), restoring β-glucuronidase activity when used to transform the deleted strain, PK803. The strategy for the localization of the gene on the insert is shown in Fig. 1. A restriction map of the insert was obtained, ar various subclones were generated in the plasmid vect pUC9 and tested for their ability to confer β-glucuronida activity upon transformation of PK803. The intermedia plasmid pRAJ210 conferred high levels of glucuronida activity on the deleted strain and was used for the purificatio of the enzyme. Several overlapping subclones containe within an 800-bp *Eco*RI–*Bam*HI fragment conferred hig levels of constitutive β-glucuronidase production only whe transformed into a *uidA*$^+$ host strain and showed no effec when transformed into PK803. We surmised that the 800-b fragment carried the operator region of the *uidA* locus an was possibly titrating repressor to give a constitutivel expressing chromosomal *uidA*$^+$ gene. With this informatio to indicate a probable direction of transcription and a mini mum gene size estimate obtained from characterization of th purified enzyme (see below), we generated a series o BAL-31 deletions from the *Xho* I site of pRAJ210. Th fragments were gel-purified, ligated into pUC9, and trans formed into PK803. The resulting colonies were then assayec for β-glucuronidase activity. The smallest clone obtained tha still gave constitutive levels of β-glucuronidase was pRAJ-220, which contained a 2.4-kilobase (kb) insert. Subclones of this 2.4-kb fragment were generated in phage vectors M13mp8 and -mp9 and their DNA sequences were determined (Fig. 2).

Manipulation of the *uidA* Gene for Vector Construction. The plasmid pRAJ220 contains the promoter and operator of the *E. coli uidA* locus, as well as additional out-of-frame ATG codons that would reduce the efficiency of proper translational initiation in eukaryotic systems (24). It was necessary to remove this DNA to facilitate using the structural gene as a reporter module in gene-fusion experiments. This was done by cloning and manipulating the 5′ region of the gene separately from the 3′ region and then rejoining the two parts as a *lacZ-uidA* fusion that showed β-glucuronidase activity under *lac* control. The resulting plasmid was further modified by progressive subcloning, linker additions, and site-directed mutagenesis to generate a set of useful gene-module vectors. The details of these manipulations are in the legend to Fig. 3.

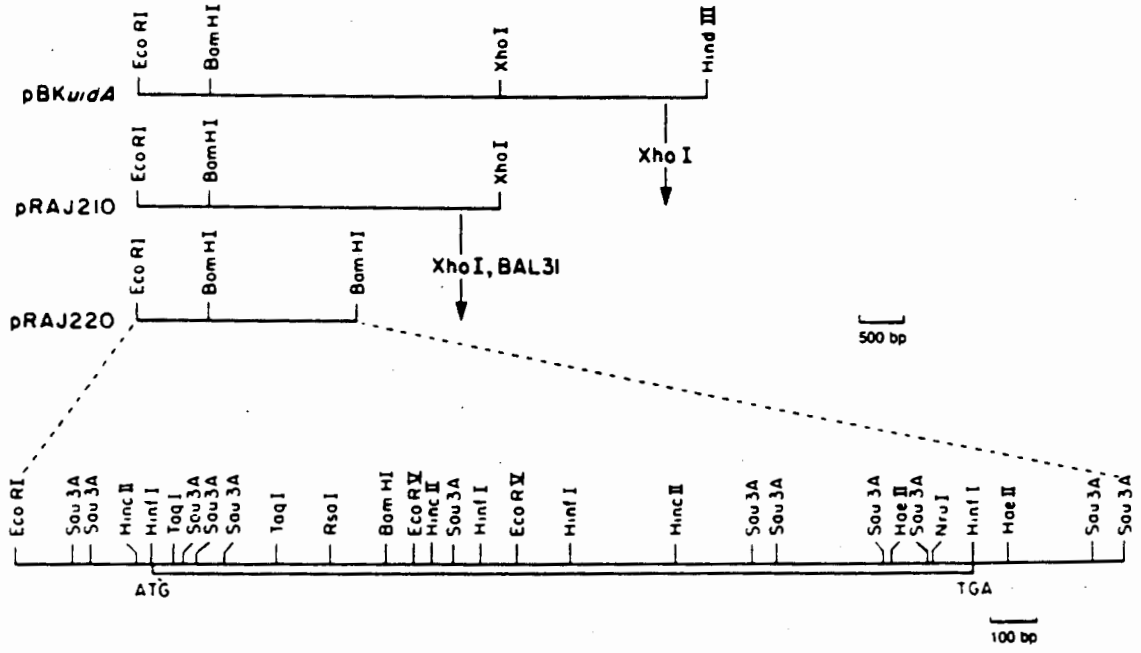


FIG. 1. Subcloning and strategy for determining the nucleotide sequence of the *uidA* gene. pBKuidA was generated by cloning into pBR325. pRAJ210 and pRAJ220 were generated in pUC9, with the orientation of the *uidA* gene opposite to that of the *lacZ* gene in the vector. Sequence was determined from both strands for all of the region indicated except nucleotides 1–125. Orientation of the coding region is from left to right. bp, Base pairs.

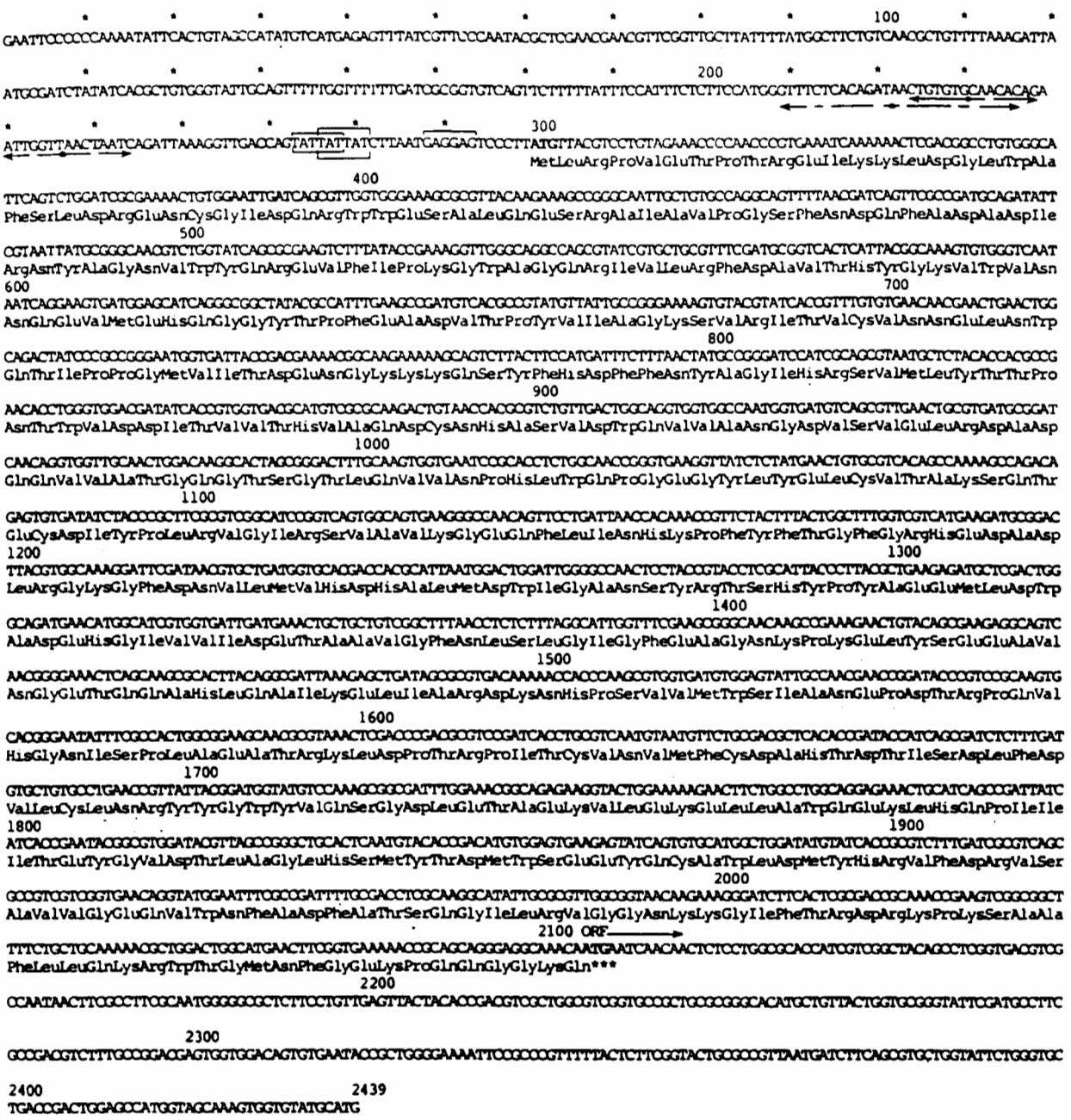


FIG. 2. DNA sequence of the 2439-bp insert of pRAJ220, containing the β-glucuronidase gene. The arrows before the coding sequence indicate regions of dyad symmetry that could be recognition sequences for effector molecules. The overlined region is the putative Shine–Dalgarno (ribosome binding site) sequence for the *uidA* gene; brackets indicate two possible Pribnow boxes. All of the palindromic regions fall within the smallest subcloned region (from the *Sau*3A site at 166 to the *Hin*fI site at 291) that gave constitutive genomic expression of *uidA* when present in high copy in *trans*, consistent with their proposed function as repressor binding sites. The terminator codon at 2106 overlaps with an ATG that may be the initiator codon of a second open reading frame, as indicated (see *Discussion*).

Purification and Properties of β-Glucuronidase. β-Glucuronidase activity in *E. coli* is induced by a variety of β-glucuronides; methyl glucuronide is among the most effective (25). To determine the size and properties of the enzyme and to verify that the enzyme produced by the clone pRAJ210 was in fact the product of the *uidA* locus, we purified the enzyme from the overproducing strain and compared the purified product with the enzyme induced from the single genomic locus by methyl glucuronide.

Aliquots of supernatants from induced and uninduced cultures of *E. coli* C600 were analyzed by $NaDodSO_4$/PAGE and compared with aliquots of the purified β-glucuronidase (Fig. 4). The induced culture of C600 shows only a single band difference relative to the uninduced culture. The new band comigrates with the purified β-glucuronidase, indicating that the enzyme purified from the overproducing plasmid strain has the same subunit molecular weight as the wild-type enzyme.

The purified enzyme was analyzed for amino acid composition and subjected to 11 cycles of Edman degradation to determine the amino-terminal sequence of amino acids. The amino acid composition agrees with the predicted composition derived from the DNA sequence, and the determined amino acid sequence agrees with the predicted sequence, identifying the site of translational initiation and indicating that the mature enzyme is not processed at the amino terminus (26).

E. coli β-glucuronidase is a very stable enzyme, with a broad pH optimum (pH 5.0–7.5); it is half as active at pH 4.3

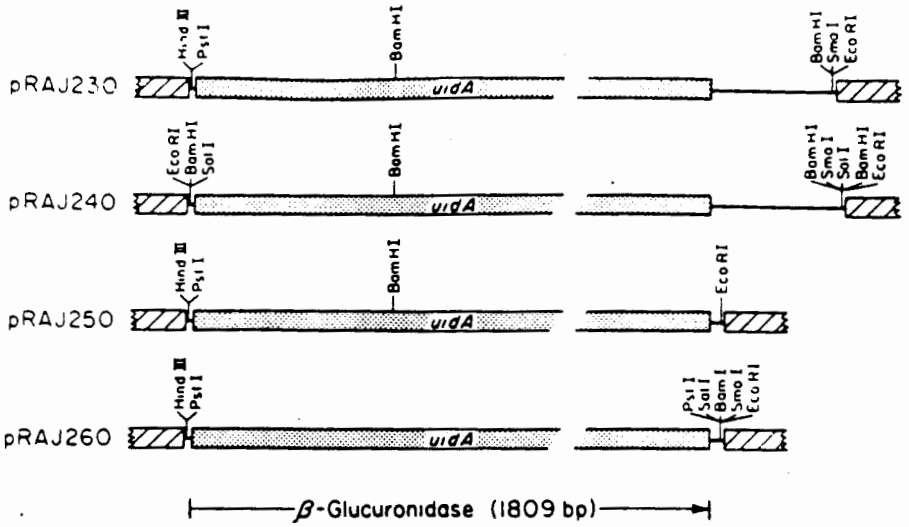


FIG. 3. β-Glucuronidase gene-module vectors. pRAJ220 (see Fig. 1) was digested with *Hin*fI, which cleaves between the Shine–Dalgar sequence and the initiator ATG, the single-stranded tails were filled in and digested with *Bam*HI, and the resulting 515-bp fragment v gel-purified and cloned into pUC9 that had been cut with *Hin*cII and *Bam*HI. This plasmid was digested with *Bam*HI, and the 3′ region of *uidA* gene carried on a 1.6-kb *Bam*HI fragment from pRAJ220 was ligated into it. The resulting plasmid, pRAJ230, showed isopro β-D-thiogalactoside inducible β-glucuronidase activity when transformed into *E. coli* JM103. pRAJ230 was further modified by the addition *Sal* I linkers to generate pRAJ240, an in-frame *lacZ-uidA* fusion in pUC7. pRAJ230 was digested with *Aat* II, which cuts 45 bp 3′ of the *ui* translational terminator, the ends were filled and digested with *Pst* I, and the resulting 1860-bp fragment was gel-purified and cloned into pEMB cut with *Pst* I and *Sma* I. The resulting plasmid, pRAJ250, is an in-frame *lacZ-uidA* fusion. We eliminated the *Bam*HI site that occurs wit the coding region at nucleotide 807 by oligonucleotide-directed mutagenesis of single-stranded DNA prepared from pRAJ250, changing t *Bam*HI site from GGATCC to GAATCC, with no change in the predicted amino acid sequence. The clone resulting from the mutagenes pRAJ255, shows normal β-glucuronidase activity and lacks the *Bam*HI site. This plasmid was further modified by the addition of a *Pst* I link to the 3′ end and cloned into pEMBL9 cut with *Pst* I, to generate pRAJ260.

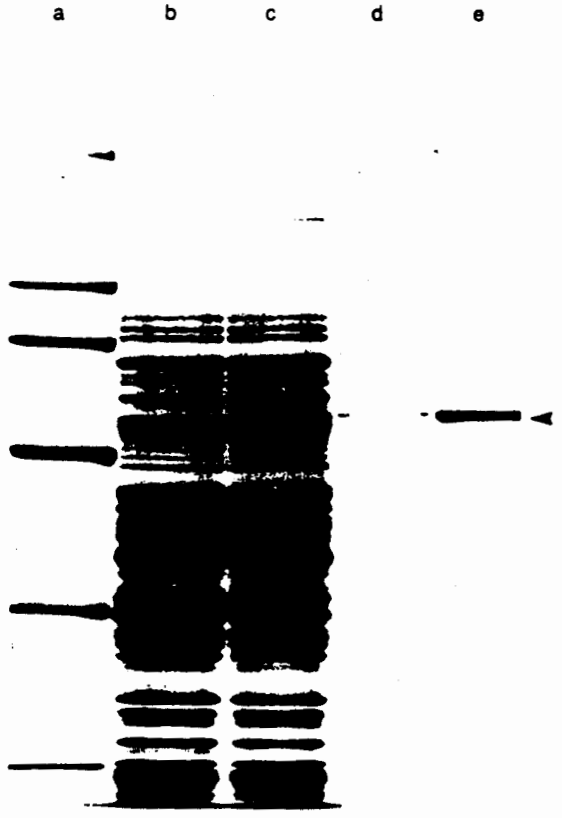


FIG. 4. NaDodSO$_4$/7.5% PAGE analysis of β-glucuronidase. Lanes: a, molecular weight standards; b, extract from uninduced *E. coli* C600; c, extract from C600 induced for β-glucuronidase with methyl glucuronide; d, 0.3 μg of purified β-glucuronidase (calculated to contain the same activity as the induced extract); e, 3.0 μg of purified β-glucuronidase.

and pH 8.5 as at its neutral optimum, and it is resistant thermal inactivation at 50°C (26).

DISCUSSION

Molecular Analysis of the *uid* Locus. We have determine the complete nucleotide sequence of the *E. coli uidA* gen encoding β-glucuronidase. The coding region of the gene 1809 bp long, giving a predicted subunit molecular weight fc the enzyme of 68,200, in agreement with the experimentall determined value of about 73,000. The translational initiatic site was verified by direct amino acid sequence analysis of th purified enzyme.

Genetic analysis of the *uidA* locus has shown three distinc controlling mechanisms, two repressors and a cAMP-depen dent factor, presumably the catabolite activator protein CAI (11). The DNA sequence that we have determined include three striking regions of dyad symmetry that could be th binding sites for the two repressors and CAP. One of th sequences matches well with the consensus sequence fo CAP binding and is located at the same distance from th putative transcriptional initiation point as the CAP bindin site of the *lac* promoter. It is interesting that the putative CAI binding site overlaps one of the other palindromic sequences suggesting a possible antagonistic effect of CAP and one o both repressors.

Our sequence analysis indicates the presence of a second open reading frame of at least 340 bp, whose initiator codon overlaps the translational terminator of the *uidA* gene. This open reading frame is translationally active (26). Although a specific glucuronide permease has been described biochemically (25), the level of genetic analysis performed on the *uid* locus would not have distinguished a mutation that eliminated glucuronidase function from a mutation that eliminated transport of the substrate (8, 9). All mutations that specifically eliminated the ability to grow on a glucuronide mapped to the *uidA* region of the *E. coli* map, indicating that if there is a gene responsible for the transport of glucuronides, it is

tightly linked to *uidA*. By analogy to the *lac* operon, we propose that the coupled open reading frame may encode a release that facilitates the uptake of β-glucuronides.

***uidA* as a Gene-Fusion Marker.** We have constructed plasmid vectors in which the *uidA* structural gene has been separated from its promoter/operator and Shine–Dalgarno (ribosome binding site) region and placed within a variety of convenient restriction sites. The gene on these restriction fragments contains all of the β-glucuronidase coding information, including the initiator codon; there are no ATGs upstream of the initiator. These vectors allow the routine transfer of the β-glucuronidase structural gene to the control of heterologous sequences, thereby facilitating the study of chimeric gene expression in other systems.

The *uidA*-encoded β-glucuronidase is functional with several combinations of up to 20 amino acids derived from the *lacZ* gene and/or polylinker sequences. Translational fusions to β-glucuronidase have also been used successfully in transformation experiments in the nematode *Caenorhabditis elegans*, and in *Nicotiana tabacum*, giving enzyme activity with many different combinations of amino-terminal structures (refs. 5 and 26; T. Kavanagh, R.A.J., and M. Bevan, unpublished data).

There are several systems currently amenable to DNA transformation in which the study of gene fusions using β-glucuronidase as the reporter enzyme may be advantageous. Very little, if any, β-glucuronidase activity has been detected in most higher plants, including tobacco (*Nicotiana tabacum*) and potato (*Solanum tuberosum*) (unpublished data). Fusions of β-glucuronidase to several plant genes have recently been used to monitor tissue-specific gene activity in transformed tobacco plants (R.A.J., T. Kavanagh and M. Bevan, unpublished data). There is no detectable β-glucuronidase activity in the slime mold *Dictyostelium discoidum* (R. Firtel, personal communication) or the yeast *Saccharomyces cerevisiae* (unpublished data). Extracts from *Drosophila melanogaster* have shown no β-glucuronidase activity under conditions that show β-galactosidase levels several hundredfold over background (26).

We thank Peter Kuempel and Rex Bitner for the gift of pBKuidA; Chris Link for enthusiasm; and Fran Storfer, Chris Fields, and Mike Krause for reading the manuscript. This work was supported by Grant GM26515 from the National Institutes of Health.

1. Cox, G. N. & Hirsh, D. (1985) *Mol. Cell. Biol.* **5**, 363–372.
2. Lis, J. T., Simon, J. A. & Sutton, C. A. (1984) *Cell* **35**, 403–410.
3. Beckwith, J. R. & Zipser, D., eds. (1970) *The Lactose Operon* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
4. Gorman, C. M., Moffatt, L. F. & Howard, B. H. (1982) *Mol. Cell. Biol.* **2**, 1044–1051.
5. Jefferson, R. A., Klass, M. & Hirsh, D. (1986) *J. Mol. Biol.*, in press.
6. Paigen, K. (1979) *Annu. Rev. Genet.* **13**, 417–466.
7. Pearse, A. G. E. (1972) *Histochemistry: Theoretical and Applied* (Churchill Livingstone, Edinburgh), 3rd Ed., Vol. 2, pp. 808–840.
8. Novel, G. & Novel, M. (1973) *Mol. Gen. Genet.* **120**, 319–335.
9. Novel, G., Didier-Fichet, M. L. & Stoeber, F. (1974) *J. Bacteriol.* **120**, 89–95.
10. Novel, M. & Novel, G. (1976) *J. Bacteriol.* **127**, 406–417.
11. Novel, M. & Novel, G. (1976) *J. Bacteriol.* **127**, 418–432.
12. Blanco, C., Ritzenthaler, P. & Mata-Gilsinger, M. (1982) *J. Bacteriol.* **149**, 587–594.
13. Blanco, C., Ritzenthaler, P. & Mata-Gilsinger, M. (1985) *Mol. Gen. Genet.* **199**, 101–105.
14. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
16. Dretzen, G., Bellare, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112**, 295–298.
17. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94**, 441–448.
18. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
19. Zoller, M. J. & Smith, M. (1982) *Nucleic Acids Res.* **10**, 6487–6500.
20. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
21. Dente, L., Cesarini, G. & Cortese, R. (1983) *Nucleic Acids Res.* **11**, 1645–1655.
22. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254.
23. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
24. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
25. Stoeber, F. (1961) These de Docteur des Sciences (University of Paris, Paris, France).
26. Jefferson, R. A. (1985) Dissertation (University of Colorado, Boulder, CO).

The Plant Cell, Vol. 1, 1079–1093, November 1989

Kunitz Trypsin Inhibitor Genes Are Differentially Expressed during the Soybean Life Cycle and in Transformed Tobacco Plants

K. Diane Jofuku[1] and Robert B. Goldberg[2]

Department of Biology, University of California, Los Angeles, California 90024-1606

We investigated the structure, organization, and developmental regulation of soybean Kunitz trypsin inhibitor genes. The Kunitz trypsin inhibitor gene family contains at least 10 members, many of which are closely linked in tandem pairs. Three Kunitz trypsin inhibitor genes, designated as KTi1, KTi2, and KTi3, do not contain intervening sequences, and are expressed during embryogenesis and in the mature plant. The KTi1 and KTi2 genes have nearly identical nucleotide sequences, are expressed at different levels during embryogenesis, are represented in leaf, root, and stem mRNAs, and probably do not encode proteins with trypsin inhibitor activity. By contrast, the KTi3 gene has diverged 20% from the KTi1 and KTi2 genes, and encodes the prominent Kunitz trypsin inhibitor found in soybean seeds. The KTi3 gene has the highest expression level during embryogenesis, and is also represented in leaf mRNA. All three Kunitz trypsin inhibitor genes are regulated correctly in transformed tobacco plants. Our results suggest that Kunitz trypsin inhibitor genes contain different combinations of *cis*-control elements that program distinct qualitative and quantitative expression patterns during the soybean life cycle.

INTRODUCTION

Soybean seeds contain two major proteinase inhibitor classes—the Kunitz trypsin inhibitor and the Bowman-Birk proteinase inhibitor (Laskowski and Kato, 1980; Ryan, 1981, 1988). These inhibitors are found in other legumes (Norioka et al., 1988), are localized within protein bodies (Horisberger and Tacchini-Vonlanthen, 1983; Vodkin and Raikhel, 1986), and are specific for serine proteases (Laskowski and Kato, 1980; Ryan, 1981). The Kunitz trypsin inhibitor class is the most prevalent soybean protease inhibitor, is represented by a 21-kD protein, and is specific for trypsin (Ryan, 1981; Kim et al., 1985; Tan-Wilson, 1988). By contrast, the Bowman-Birk class consists of several related 8-kD proteins, and inhibits trypsin, chymotrypsin, and elastase (Laskowski and Kato, 1980; Ryan, 1981, 1988; Tan-Wilson, 1988).

Proteinase inhibitors have been proposed to function as storage proteins, regulators of endogenous proteinases, or factors that protect plants from insect attack (Ryan, 1981). Transfer of an alfalfa Bowman-Birk inhibitor gene to tobacco plants showed directly that this proteinase inhibitor can function as an endogenous insecticide (Hilder et al., 1986). Because soybean null lines exist that lack either the Kunitz trypsin inhibitor (Orf and Hymowitz, 1979) or the Bowman-Birk protease inhibitor (Stahlhut and Hymowitz, 1983), these proteins are not essential for normal growth and development.

We showed previously that the Kunitz trypsin inhibitor gene family is highly regulated during the soybean life cycle (Goldberg et al., 1981; Walling, Drews, and Goldberg, 1986; Jofuku, Schipper, and Goldberg, 1989). These genes are expressed at precise times during embryogenesis, have different expression levels in the embryo cotyledon and axis, and are either inactive or expressed at low levels in mature plant organ systems (Goldberg et al., 1981; Walling et al., 1986). Both transcriptional and post-transcriptional processes regulate Kunitz trypsin inhibitor gene expression (Walling et al., 1986). Jofuku et al. (1989) showed that a null line that lacks Kunitz trypsin inhibitor activity has a frameshift mutation in the major Kunitz trypsin inhibitor gene (KTi3). This mutation prevents KTi3 mRNA from being translated correctly in embryo cells (Jofuku et al., 1989). Other Kunitz trypsin inhibitor mRNAs (e.g., KTi1 and KTi2) are unaffected by the frameshift mutation (Jofuku et al., 1989).

To begin to identify the DNA sequences that regulate Kunitz trypsin inhibitor gene expression, we investigated the organization and developmental regulation of specific Kunitz trypsin inhibitor gene family members. Our studies indicate that the Kunitz trypsin inhibitor family contains at least 10 members, and that some of these genes are

[1] Current address: Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium.

[2] To whom correspondence should be addressed.

organized into tandem pairs. We further show that three distinct Kunitz trypsin inhibitor genes (KTi1, KTi2, and KTi3) have different expression patterns during the soybean life cycle and in transformed tobacco plants. Our results suggest that each Kunitz trypsin inhibitor gene contains a unique combination of *cis*-control elements that programs specific expression patterns during embryogenesis and in the mature plant.

RESULTS

Kunitz Trypsin Inhibitor Is Encoded by a Multigene Family

We described previously the isolation of a Kunitz trypsin inhibitor cDNA plasmid, designated as A-37, from a library of mid-maturation stage embryo mRNA (Goldberg et al., 1981). DNA sequence analysis revealed that the mRNA represented by A-37 encoded a protein with only 70% homology to the major Kunitz trypsin inhibitor present in soybean seeds (data not shown). To isolate a cDNA clone representing the major Kunitz trypsin inhibitor mRNA, we constructed a second library using mRNA isolated from immunoselected Kunitz trypsin inhibitor polysomes (Vodkin, 1981; J.J. Harada and R.B. Goldberg, unpublished results). We screened the KTi-enriched embryo mRNA library at a reduced criterion (32°C, $\Delta T_m = -28°$) with the A-37 cDNA plasmid and identified a second Kunitz trypsin inhibitor plasmid, designated as pKT3. The pKT3 DNA sequence indicated that it represented the major Kunitz trypsin inhibitor mRNA (Jofuku et al., 1989).

We hybridized the A-37 and pKT3 cDNA plasmids with DNA gel blots to estimate the Kunitz trypsin inhibitor gene copy number. Figure 1A shows that, under moderately stringent hybridization conditions, A-37 hybridized with a single-copy 4.0-kb EcoRI DNA fragment (42° lane). At the same criterion, pKT3 hybridized with two single-copy EcoRI DNA fragments, 5.5 kb and 0.6 kb in length (Jofuku et al., 1989). By contrast, a mixed A-37 and pKT3 probe hybridized with greater than 10 EcoRI fragments under less stringent hybridization conditions (Figure 1A, 22° lane). Together, these results indicate the presence of a large number of Kunitz trypsin inhibitor DNA sequences in the soybean genome, and that genes represented by the A-37 and pKT3 cDNA plasmids are present on different EcoRI fragments.

Kunitz Trypsin Inhibitor Genes Are Organized as Tandem Pairs

We used the A-37 and pKT3 cDNA plasmids to screen two λ Charon 4 soybean genome libraries for Kunitz trypsin inhibitor genes (see Methods). Figure 1B shows that we

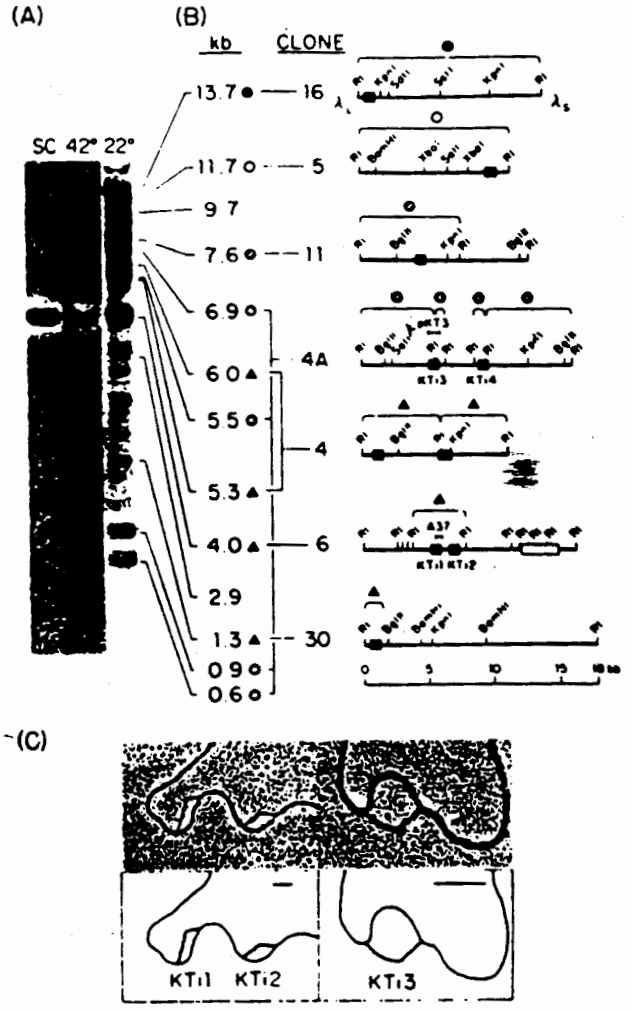


Figure 1. Soybean Kunitz Trypsin Inhibitor Gene Structure and Organization.

(A) Kunitz trypsin inhibitor gene representation in the soybean genome. Dare variety leaf DNA was digested with EcoRI, fractionated by electrophoresis, transferred to nitrocellulose, and hybridized with the Kunitz trypsin inhibitor cDNA plasmid A-37 (Goldberg et al., 1981; Jofuku et al., 1989) at either 42°C ($\Delta T_m = -18°$) or 22°C ($\Delta T_m = -38°$). SC lane contained a single-copy equivalent of EcoRI-digested λClone 6 DNA.

(B) Kunitz trypsin inhibitor gene sequence organization. Boxes and arrows represent gene locations and transcriptional orientations, respectively. The open box represents a non-seed protein gene described previously by Walling et al. (1986). Lines above the λClone 6 and λClone 4A phages indicate gene sequences represented by the cDNA plasmids A-37 and pKT3, respectively. Brackets indicate the EcoRI DNA fragments that hybridized with labeled A-37 and pKT3 plasmid DNAs at 22°C ($\Delta T_m = -38°$). The circles and triangles correlate phage DNA fragments with soybean DNA segments detected by the DNA gel blot shown in **(A)**.

(C) Kunitz trypsin inhibitor gene structure. λClone 6 and λClone 4A DNAs were hybridized with embryo mRNA under conditions that form R-loops (Fischer and Goldberg, 1982). Bars represent 1 kb.

identified seven recombinant phages that contained 10 distinct Kunitz trypsin inhibitor genes. Collectively, the DNA inserts present in these phages represented more than 145 kb of the soybean genome and contained greater than 90% of the EcoRI DNA fragments detected by A-37 and pKT3 in the DNA gel blot shown in Figure 1A.

Restriction endonuclease mapping and localization of Kunitz trypsin inhibitor DNA sequences within the recombinant phages demonstrated that several genes were clustered in tandem pairs. For example, Figure 1B shows that λClone 6 contains two Kunitz trypsin inhibitor genes, designated as KTi1 and KTi2, that are separated by less than 0.5 kb and are present on the 4.0-kb EcoRI DNA fragment that hybridized strongly with the A-37 plasmid (Figure 1A, SC and 42° lanes). Similarly, Figure 1B shows that λClone 4A possesses two Kunitz trypsin inhibitor genes, designated as KTi3 and KTi4. The KTi3 gene is present on the 5.5-kb and 0.6-kb EcoRI DNA fragments that hybridized strongly with the pKT3 plasmid, suggesting that KTi3 is the major Kunitz trypsin inhibitor gene (Jofuku et al., 1989). λClone 4 also contains two Kunitz trypsin inhibitor gene sequences (Figure 1B). Together, these results indicate that at least six of the 10 cloned Kunitz trypsin inhibitor genes are closely linked as tandem pairs in the soybean genome.

The KTi1, KTi2, and KTi3 Kunitz Trypsin Inhibitor Genes Do not Contain Introns

We characterized the KTi1, KTi2, and KTi3 Kunitz trypsin inhibitor genes by R-loop analysis (Fischer and Goldberg, 1982) to visualize their structures in the electron microscope. Figure 1C shows that these genes do not contain introns detectable by electron microscopy. We sequenced the KTi1, KTi2, and KTi3 genes and their contiguous regions to verify these structures directly. The KTi1 and KTi2 DNA sequences are presented in Figure 2, while the KTi3 gene sequence is presented in Jofuku et al. (1989). A schematic representation of all three genes is shown in Figure 3, along with relevant consensus sequences and a comparison of homologous gene regions. A comparison of the translated protein sequences with the sequence of the major Kunitz trypsin inhibitor protein (Kim et al., 1985) is shown in Figure 4.

Figures 2 and 3 show directly that introns are absent from all three Kunitz trypsin inhibitor genes. Analysis of the protein sequences encoded by the Kunitz trypsin inhibitor genes indicates that the KTi3 gene encodes a protein identical to the major Kunitz trypsin inhibitor found in soybean seeds (Figure 4; Jofuku et al., 1989). By contrast, the KTi1 and KTi2 genes encode proteins with only 70% homology with the major Kunitz trypsin inhibitor protein (Figure 4). Together, these results indicate that the KTi1, KTi2, and KTi3 Kunitz trypsin inhibitor genes have simple structures, and that they encode distinct proteins.

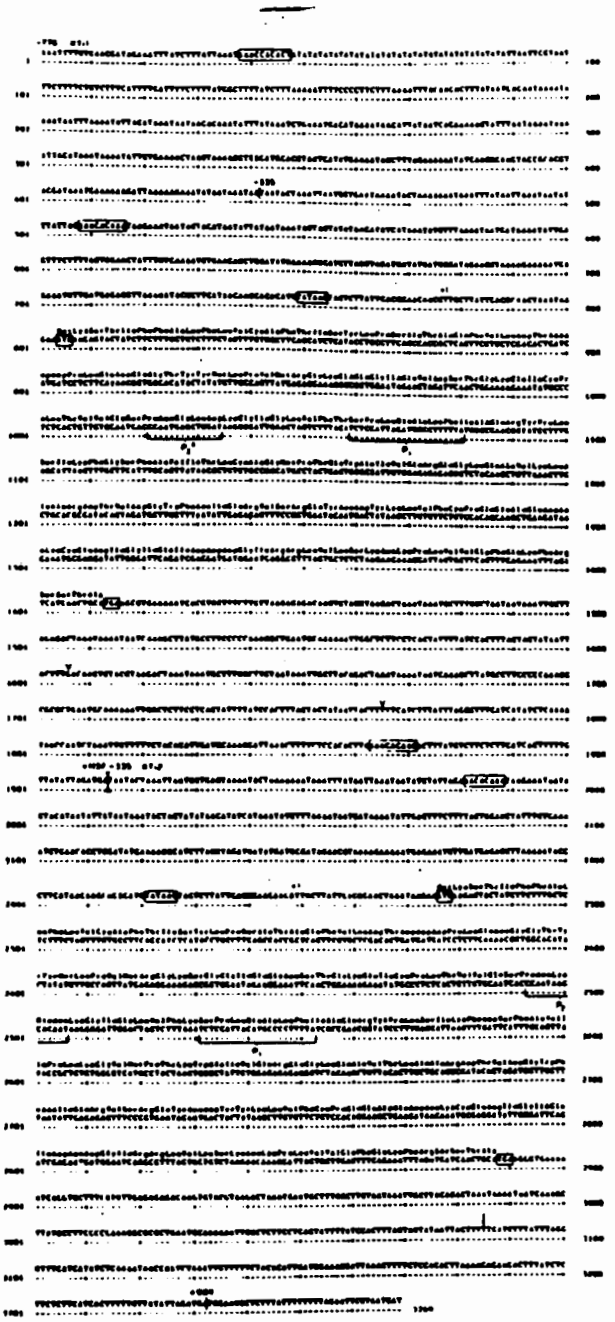


Figure 2. KTi1 and KTi2 Kunitz Trypsin Inhibitor Gene Sequences.

Boxes enclose TATA-like sequences, translation start and stop codons, and a CACA sequence motif (Goldberg et al., 1989). Brackets at −335, +1137/−335, and +984 enclose sequences with >90% similarity between the two gene regions. 5′ and 3′ ends of an insertion that arose from a duplication of 3′ KTi1 gene sequences (nucleotides +674 to +829) are designated by the V. An arrow shows the position at which this insertion sequence would occur within the homologous KTi2 gene region (nucleotide +836).

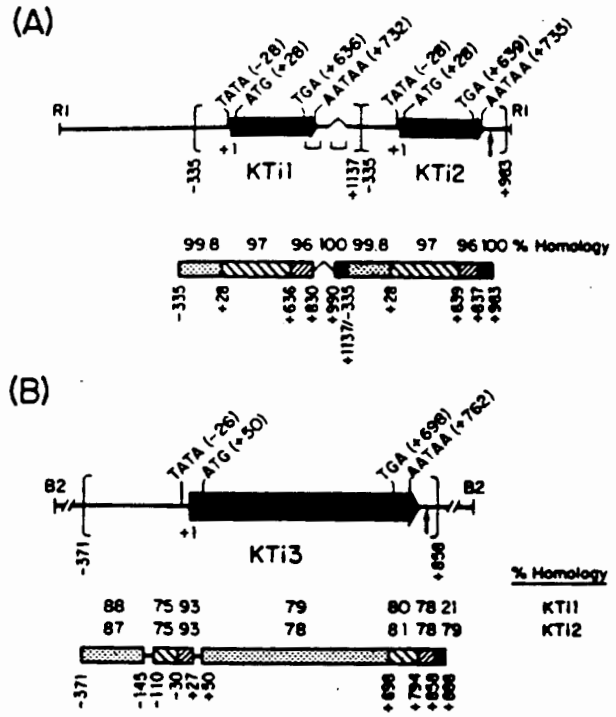


Figure 3. Kunitz Trypsin Inhibitor Gene Similarities.

Relevant consensus sequences and their locations are shown above the genes. Brackets below the gene regions indicate KTi1 nucleotides +674 to +829 that are duplicated to form a portion of the 5′-flanking region of KTi2. Arrows indicate the KTi2 and KTi3 3′ regions that do not contain the duplicated KTi1 DNA segment. Brackets enclosing gene regions indicate sequences that are highly homologous between the KTi1, KTi2, and KTi3 genes.

(A) KTi1 and KTi2 gene sequence comparison. Boxes filled with similar patterns indicate compared gene regions. Percent homology and nucleotide reference points are shown above and below the designated gene region, respectively.

(B) KTi1, KTi2, and KTi3 gene sequence comparisons. Percent homology of KTi3 gene regions with corresponding KTi1 or KTi2 sequences is shown above the designated region. The 5′ and 3′ ends of compared gene regions are shown below the designated gene region. Thin lines between regions indicate gaps introduced into the KTi3 gene sequence to allow for maximum alignment of all three gene sequences.

The KTi1 and KTi2 Kunitz Trypsin Inhibitor Genes Are Nearly Identical to Each Other

Figure 3A shows that the KTi1 and KTi2 genes are closely related to each other. For example, the coding regions of these genes differ by only 18 out of 608 nucleotides, indicating approximately 97% sequence similarity (Figure 3A). In addition, the KTi1 and KTi2 5′-flanking regions differ by only 1 nucleotide up to position −335. Upstream

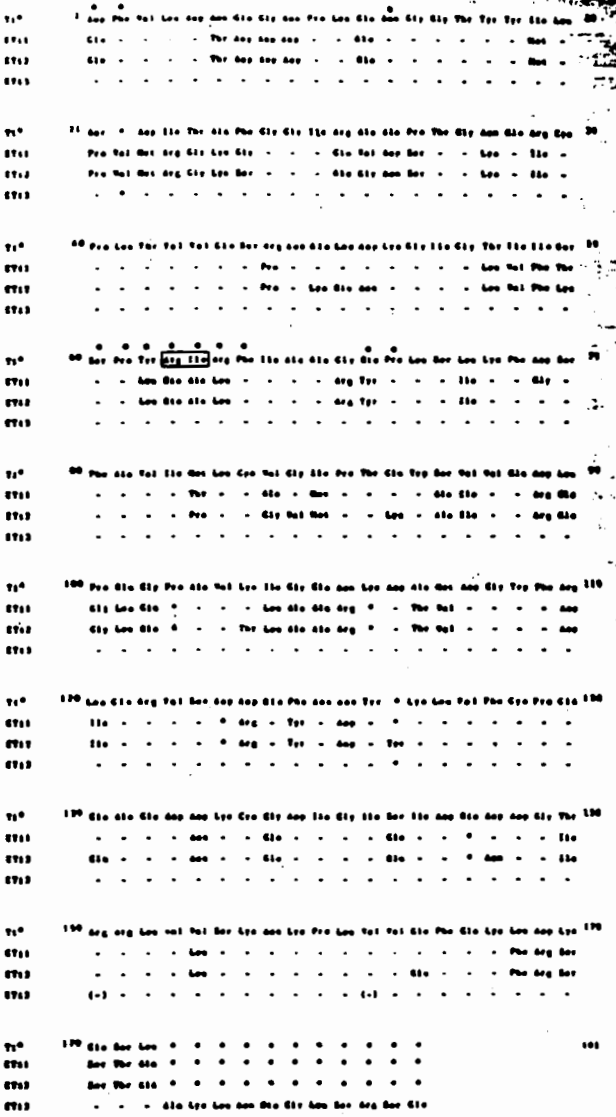


Figure 4. Kunitz Trypsin Inhibitor Amino Acid Comparisons.

The amino acid sequence of the Kunitz trypsin inhibitor *Tf*[a] allelic form (Kim et al., 1985) is shown for reference. The translated protein sequences of the KTi1, KTi2, and KTi3 genes are shown below the *Tf*[a] sequence. Dashes indicate amino acid sequence identities with the *Tf*[a] allelic form. Asterisks designate gaps introduced into a given amino acid sequence to obtain maximum alignment of all four proteins. Closed circles above the sequence designate Kunitz trypsin inhibitor amino acids that form possible contacts with specific amino acids of porcine trypsin (Sweet et al., 1974). Box encloses amino acids Arg-63 and Ile-64. The peptide bond between these two amino acids is cleaved by trypsin (Ozawa and Laskowski, 1966). Substitution and/or deletion of either amino acid inactivates Kunitz trypsin inhibitor activity (Ryan, 1981).

of this position, the KTi1 5′ region continues for at least 1 kb, while the KTi2 5′ region merges with the KTi1 3′ region (Figure 3A). By contrast, Figure 3B shows that the KTi3 coding region is only 80% similar to the corresponding KTi1 and KTi2 gene regions. A slightly higher degree of similarity is found between the KTi3 and KTi1/2 5′-flanking regions. Together, these findings indicate that the KTi1 and KTi2 genes are almost exact copies of each other, and that the KTi3 gene represents a divergent relative.

Kunitz Trypsin Inhibitor Gene Expression Is Regulated during Embryogenesis

We hybridized A-37 plasmid DNA with dot blots containing mRNAs from 14 different embryonic stages to investigate the combined KTi1 and KTi2 Kunitz trypsin inhibitor gene expression patterns during embryogenesis. We showed that, under our DNA/RNA hybridization conditions (42°C, $\Delta T_m = -18°$), there was less than 10% cross-reaction between A-37 and KTi3 mRNA and between pKT3 and KTi1 and KTi2 mRNAs (G. de Paiva and R.B. Goldberg, unpublished results). Figure 5A shows that the KTi1/2 mRNAs accumulated early in embryogenesis, reached a maximum level at mid-maturation [77 days after flowering (DAF)], and then decayed prior to dormancy.

To determine the prevalence of KTi3 mRNA relative to the KTi1 and KTi2 mRNAs, we hybridized A-37 and pKT3 plasmid DNAs with mid-maturation stage embryo mRNA gel blots under the same hybridization conditions. Figure 5B shows that both probes hybridized with a prevalent 0.9-kb embryo mRNA. However, the hybridization signal obtained with the pKT3 probe was approximately fivefold more intense than that obtained with A-37. This result indicates that KTi3 mRNA is more prevalent than both the KTi1 and KTi2 mRNAs at mid-maturation.

To determine whether the difference in KTi1/2 and KTi3 mRNA levels reflected the transcriptional activities of the KTi1/2 and KTi3 genes, we hybridized embryo ^{32}P-nuclear RNA synthesized in vitro (Walling et al., 1986) with DNA gel blots containing KTi1/2 and KTi3 gene fragments. Figure 6 shows that ^{32}P-nuclear RNA hybridized approximately fivefold more intensely with the KTi3 DNA fragment than with the KTi1/2 DNA fragment. Together, these data indicate that Kunitz trypsin inhibitor genes are regulated temporally during embryogenesis, that the KTi3 gene is expressed at a higher level than both the KTi1 and KTi2 genes, and that the difference in KTi1/2 and KTi3 gene expression levels is due primarily to a higher KTi3 transcription rate.

The KTi1 and KTi2 Kunitz Trypsin Inhibitor Genes Are Expressed at Different Levels during Embryogenesis

Figures 2 and 3 show that the KTi1 and KTi2 Kunitz trypsin inhibitor genes and their flanking regions are nearly identical to each other, suggesting that they may have similar expression programs. We used both primer extension and mRNA gel blot studies with two gene-specific oligonucleotide probes, designated as P_1 and P_2, to determine the individual KTi1 and KTi2 gene expression levels in mid-maturation stage embryos. The nucleotide sequences of

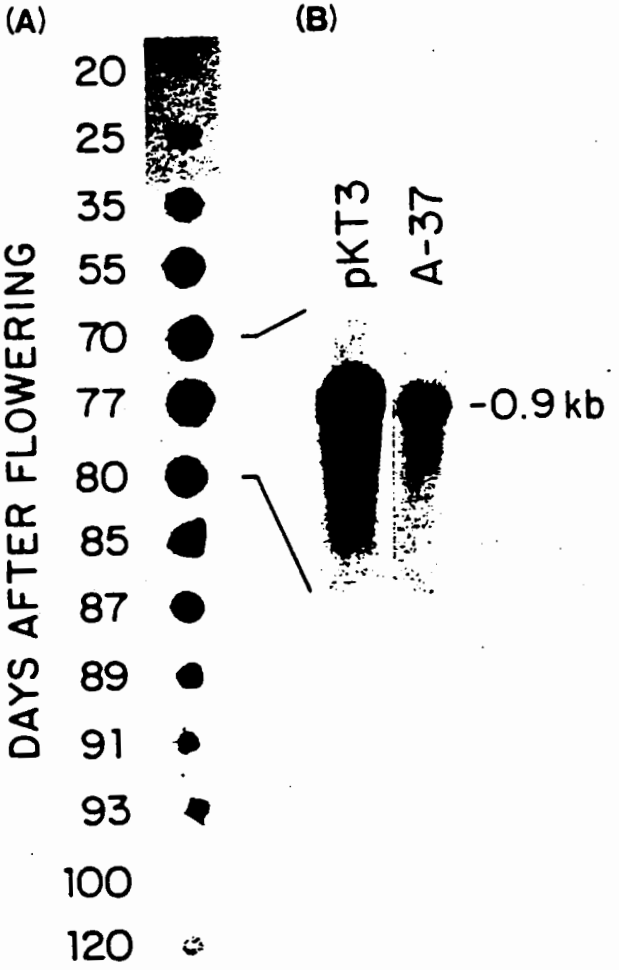


Figure 5. Kunitz Trypsin Inhibitor Gene Expression during Soybean Embryogenesis.

(A) Accumulation of KTi1/2 mRNAs during soybean embryogenesis. 0.2 μg of embryo mRNA from different developmental stages was spotted onto nitrocellulose and hybridized with labeled A-37 plasmid DNA (Figure 1). Soybean embryo developmental stages relative to days after flowering were described by Goldberg et al. (1989).

(B) Representation of KTi1/2 and KTi3 mRNAs in soybean embryos. Two micrograms of soybean mid-maturation stage embryo mRNA was fractionated on methylmercury hydroxide agarose gels, transferred to nitrocellulose, and hybridized with labeled A-37 and pKT3 plasmid DNAs (Figure 1).

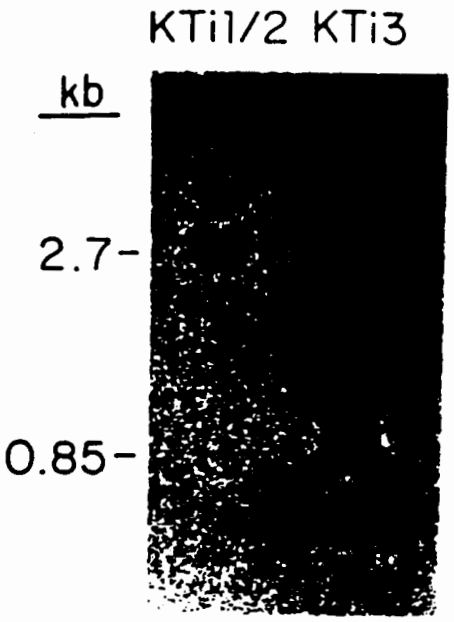


Figure 6. Kunitz Trypsin Inhibitor Gene Transcription in Soybean Embryo Nuclei.

Plasmid DNAs were digested with relevant restriction endonucleases, fractionated by electrophoresis on agarose gels, and transferred to nitrocellulose. The DNA gel blots were then hybridized with ^{32}P-nuclear RNAs synthesized from 60 DAF soybean embryo nuclei. We showed previously that at this developmental stage soybean seed protein gene transcriptional activities reach maximum levels (Walling et al., 1986). KTi1/2 and KTi3 refer to lanes containing 0.6 μg of KTi1 and KTi3 plasmid DNA inserts that are 2.7 kb and 0.85 kb in length, respectively.

the P_1 and P_2 oligonucleotides and their locations within the KTi1 and KTi2 genes are shown in Figure 2 and are schematically represented in Figure 7A.

We hybridized the 22-nucleotide P_1 oligonucleotide with mid-maturation stage mRNA, and then extended the primer with reverse transcriptase in the presence of ^{32}P-TTP. The P_1 oligonucleotide sequence is present in both the KTi1 and KTi2 genes (Figure 2). However, as seen in Figures 2 and 7B, the KTi1 and KTi2 Kunitz trypsin inhibitor genes differ by 1 nucleotide in the region 5′ to the shared P_1 oligonucleotide sequence (5′-ACA-3′ versus 5′-AAA-3′). Because of the cytosine residue in the KTi1 mRNA sequence, extension of a P_1 oligonucleotide/KTi1 mRNA hybrid will terminate 2 nucleotides before that of the analogous P_1/KTi2 hybrid in the presence of only ^{32}P-TTP (Figure 7B). Comparison of the P_1/embryo mRNA primer extension products enables the KTi1 and KTi2 mRNA levels to be estimated individually, even though the two mRNAs differ by only a few nucleotides (Figures 2 and 3).

Figure 7C shows that the expected 23-nucleotide and 25-nucleotide primer extension products were synthesized

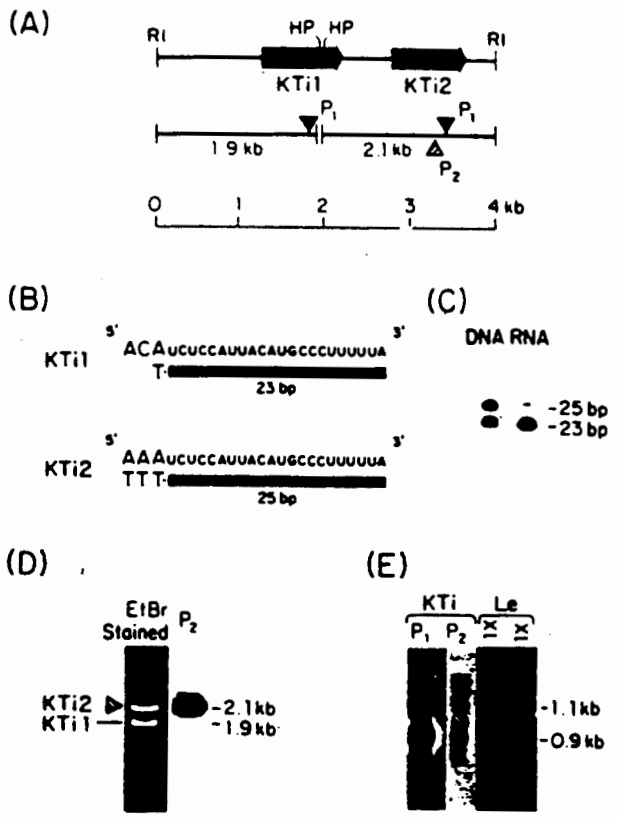


Figure 7. KTi1 and KTi2 Gene Expression in Soybean Embryos.

(A) Organization of the KTi1 and KTi2 Kunitz trypsin inhibitor genes. A schematic representation of the pBR325 recombinant plasmid pE4 is shown for reference. Arrows indicate the sites within the DNA fragments that are recognized by primers P_1 and P_2 (Figure 2).

(B) Experimental strategy for primer extension analysis using P_1. Dark boxes represent the 22-nucleotide P_1 primer. The 23-nucleotide and 25-nucleotide products expected from extending the P_1:mRNA hybrids with reverse transcriptase are shown below the corresponding mRNA sequence.

(C) Primer extension analysis. Five micrograms of embryo mRNA were hybridized with primer P_1, and DNA was synthesized in the presence of only ^{32}P-TTP (lane RNA). A control reaction was carried out using 1 μg of denatured pE4 DNA template (lane DNA).

(D) Hybridization of primer P_2 to DNA gel blots. pE4 DNA was digested with EcoRI and HpaI, fractionated by electrophoresis, and transferred to nitrocellulose. The DNA gel blot was then hybridized with labeled P_2 (lane P_2). KTi1 and KTi2 refer to the 1.9-kb and 2.1-kb EcoRI/HpaI DNA fragments (A).

(E) Hybridization of P_1 and P_2 primers with mRNA gel blots. Five micrograms of embryo mRNA were fractionated on agarose gels, transferred to nitrocellulose, and hybridized with labeled P_1 and P_2. KTi lanes refer to reactions with P_1 and P_2. Le lane refers to reactions with a 25-nucleotide primer that represents nucleotides +102 to +127 of the *Le1* lectin gene (Goldberg, Hoschek, and Vodkin, 1983). The 0.1X and 1X lanes contained 0.5 ng and 5 ng of mid-maturation stage (77 DAF) embryo mRNA. 1.1 kb and 0.9 kb refer to lectin and Kunitz trypsin inhibitor mRNA sizes, respectively.

in equimolar amounts using KTi1 and KTi2 DNA as templates, respectively (DNA lane). By contrast, the 23-nucleotide KTi1 mRNA extension product was synthesized at a 20-fold higher level than the 25-nucleotide KTi2 mRNA product in the presence of mid-maturation stage embryo mRNA (RNA lane). Experiments with mRNAs from different embryonic stages indicated that the difference in KTi1 and KTi2 mRNA levels occurred throughout embryogenesis (G. de Paiva and R.B. Goldberg, unpublished results).

To verify this result with a different procedure, we hybridized labeled P_1 and P_2 oligonucleotide probes with embryo mRNA gel blots. Under the hybridization conditions used (see Methods), the 14-nucleotide P_2 oligonucleotide probe should hybridize specifically with KTi2 mRNA because of a 3-nucleotide difference in the corresponding KTi1 mRNA region (see Figure 2). Figure 7D shows that, as expected, P_2 hybridized with KTi2 DNA and did not hybridize detectably with KTi1 DNA (P_2 lane). By contrast, the shared P_1 probe hybridized with KTi1 and KTi2 DNAs with equal intensities (data not shown).

Figure 7E shows that both oligonucleotide probes hybridized with a 0.9-kb embryo mRNA. However, the P_2 probe produced a 10-fold lower hybridization signal with mid-maturation stage embryo mRNA than did the P_1 probe (compare lanes P_1 and P_2), confirming that the KTi2 mRNA is less prevalent than the KTi1 mRNA. By reference to seed lectin mRNA standards (Figure 7E, Le lanes), we estimated that the KTi1 and KTi2 mRNAs represented 0.75% and 0.075% of the mid-maturation stage embryo mRNA mass, respectively. Because the KTi3 mRNA is at least 5 times more prevalent than the KTi1 and KTi2 mRNAs (Figure 5), it represents approximately 4% of the mRNA mass at this developmental stage. Together, these data show that the KTi1 and KTi2 Kunitz trypsin inhibitor genes are expressed at different quantitative levels during embryogenesis even though these genes have nearly identical DNA sequences and flanking regions.

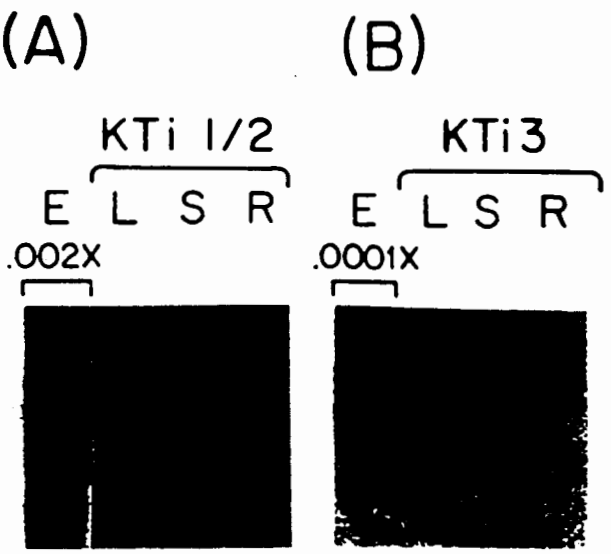


Figure 8. Kunitz Trypsin Inhibitor Gene Expression in Mature Plant Organ Systems.

Five micrograms of soybean leaf (L), stem (S), and root (R) mRNAs was fractionated on methylmercury hydroxide agarose gels and transferred to nitrocellulose paper. The mRNA gel blots were hybridized with Kunitz trypsin inhibitor gene probes and were then subjected to a 42°C washing criterion.

(A) KTi1/2 gene expression in mature plant organ systems. KTi1/2 indicates hybridization with labeled pE4 plasmid DNA (Figure 7). Lane E contained 10 ng of mid-maturation stage embryo mRNA, or the equivalent of 30 molecules of Kunitz trypsin inhibitor mRNA per cell (Figure 7; Goldberg et al., 1981).

(B) KTi3 gene expression in mature plant organ systems. KTi3 represents reaction with labeled pKT3 plasmid DNA (Figure 1). Lane E contained 0.5 ng of mid-maturation stage mRNA or approximately two molecules of KTi3 mRNA per cell (Figures 5 and 7).

Kunitz Trypsin Inhibitor Genes Are Expressed in Mature Plant Organ Systems

We hybridized Kunitz trypsin inhibitor gene probes with leaf, stem, and root mRNA gel blots to determine whether the Kunitz trypsin inhibitor genes were expressed in organ systems of the mature plant. As seen in Figure 8A, a 4.0-kb KTi1/2 gene probe (Figure 7A) hybridized with a low-prevalence 0.9-kb RNA present in leaf, stem, and root mRNAs. By contrast, Figure 8B shows that the KTi3 gene probe produced a 0.9-kb signal with leaf mRNA but not detectably with stem or root mRNAs. Relative to embryo mRNA standards (Figure 8, E lanes), we estimated that the KTi1, KTi2, and KTi3 mRNAs were present in mature plant organ systems at concentrations at least 1000-fold lower than those observed in mid-maturation stage embryos (compare L, S, R, and E lanes). Together, these results indicate that the KTi1/2 and KTi3 Kunitz trypsin inhibitor genes are expressed in specific organ systems of the mature plant in addition to being expressed during embryogenesis.

Kunitz Trypsin Inhibitor Genes Are Expressed Correctly during Tobacco Seed Development

Tobacco Transformants Contain One Kunitz Trypsin Inhibitor Locus

We transformed tobacco plants with the KTi1, KTi2, and KTi3 Kunitz trypsin inhibitor genes to begin to localize sequences responsible for their developmental-specific expression patterns. Figure 9 schematically shows the Kunitz trypsin inhibitor gene regions transferred to tobacco plants. We utilized three DNA fragments: (1) the 4.0-kb EcoRI DNA fragment containing both the KTi1 and KTi2

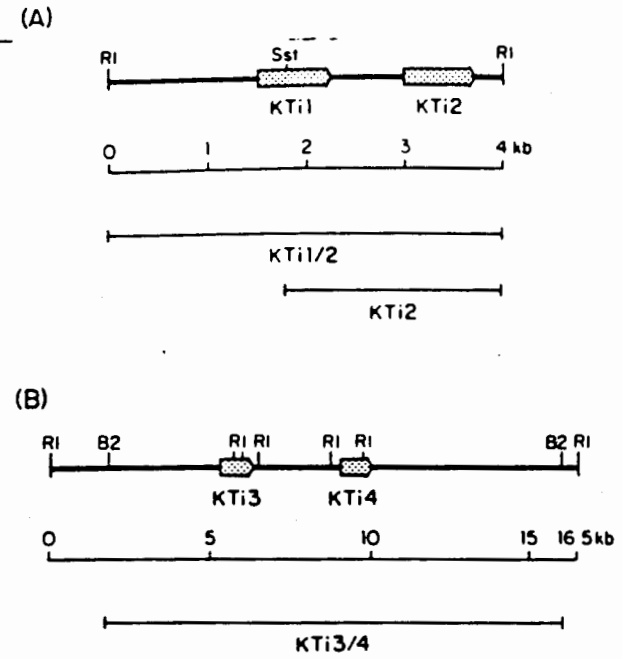


Figure 9. The KTi1/2 and KTi3/4 Kunitz Trypsin Inhibitor Gene Regions.

(A) Schematic representation of the KTi1 and KTi2 gene region. RI and Sst refer to EcoRI and SstI restriction endonucleases, respectively. Boxes represent gene locations, and arrows show transcriptional orientations (Figure 1). Lines below the gene region represent the 4.0-kb KTi1/2 and 2.2-kb KTi2 DNA fragments utilized in the gene transfer studies.

(B) Schematic representation of the KTi3 and KTi4 gene region. RI and B2 refer to EcoRI and BglII restriction endonucleases, respectively. Boxes represent gene locations and transcriptional orientations, respectively (Figure 1). Line below the gene region represents the 12.5-kb BglII DNA fragment used in the gene transfer studies.

genes (Figures 1, 3A, 7A, and 9A), (2) a 2.2-kb SstI fragment containing only the KTi2 gene (Figure 9A), and (3) a 12.5-kb BglII fragment containing the KTi3 gene (Figures 1, 3B, and 9B). In addition to the KTi3 Kunitz trypsin inhibitor gene, the 12.5-kb BglII DNA fragment contains another Kunitz trypsin inhibitor gene, designated as KTi4 (Figures 1 and 9B). The DNA sequence of the KTi4 Kunitz trypsin inhibitor gene has not yet been determined.

We selected several independent transformants for each Kunitz trypsin inhibitor gene region, and then determined genetically the number of Kunitz trypsin inhibitor loci present in each transformant by scoring the kanamycin-resistance (Kn^R) phenotype conferred by linked vector gene sequences (Hain et al., 1985). Seeds obtained from self-fertilized transformed plants segregated 3:1 relative to the Kn^R marker, indicating the presence of a single Kunitz trypsin inhibitor locus in each transformed tobacco genome (data not shown).

To verify this result directly, we hybridized gel blots containing DNA from transformants containing the KTi1 and KTi2 Kunitz trypsin inhibitor genes (Figure 9A) with the 4.0-kb KTi1/2 gene probe shown in Figures 7A, 9A, and 10A. Figure 10B shows that no detectable signal was obtained with untransformed tobacco DNA (lane C). By contrast, low-copy soybean DNA fragments of the expected size (4.0 kb) were present in transformants containing the KTi1/2 gene region (Figure 10B, RI lanes). In addition, only one DNA fragment was obtained with a restriction enzyme that does not digest either the vector or the KTi1/2 DNA (Figure 10B, Hp lanes). Together, these findings suggest that each tobacco transformant contained a single unrearranged Kunitz trypsin inhibitor gene region.

The KTi1 and KTi2 Kunitz Trypsin Inhibitor Genes Are Expressed Differentially in Tobacco Seeds

We hybridized tobacco seed mRNA gel blots with the 4.0 kb KTi1/2 gene probe shown in Figure 11A to determine whether the KTi1 and KTi2 Kunitz trypsin inhibitor genes were correctly expressed during tobacco seed development. Figure 11B shows that no detectable signal was observed with untransformed tobacco mRNA (C lane). By contrast, the 4.0-kb KTi1/2 gene probe produced a 0.9-kb signal with seed mRNAs from tobacco plants containing either the KTi1 and KTi2 gene (Figure 11B, KTi1/2 lane) or the KTi2 gene alone (Figure 11B, KTi2 lane). As predicted from the results obtained with soybean embryo mRNA (Figure 7), the KTi2 mRNA was present at a lower level than the KTi1 mRNA in the transformed tobacco seeds (compare Figure 11B, KTi1/2 and KTi2 lanes). The difference in KTi1/2 and KTi2 mRNA levels was not paralleled by a similar difference in vector Kn^R mRNA levels in seeds of the same transformants (data not shown). This suggests that the observed mRNA prevalence differences were due to actual differences in KTi1 and KTi2 Kunitz trypsin inhibitor gene expression levels and not due to position effects.

The KTi1 and KTi2 Kunitz Trypsin Inhibitor Genes Are Regulated Temporally during Tobacco Seed Development

We isolated transformed tobacco seed mRNAs at different times during seed development and hybridized each mRNA population with the 4.0-kb KTi1/2 gene probe to determine whether the KTi1 and KTi2 Kunitz trypsin inhibitor genes were temporally regulated during tobacco seed development. As shown in Figure 11C, the KTi1/2 mRNAs

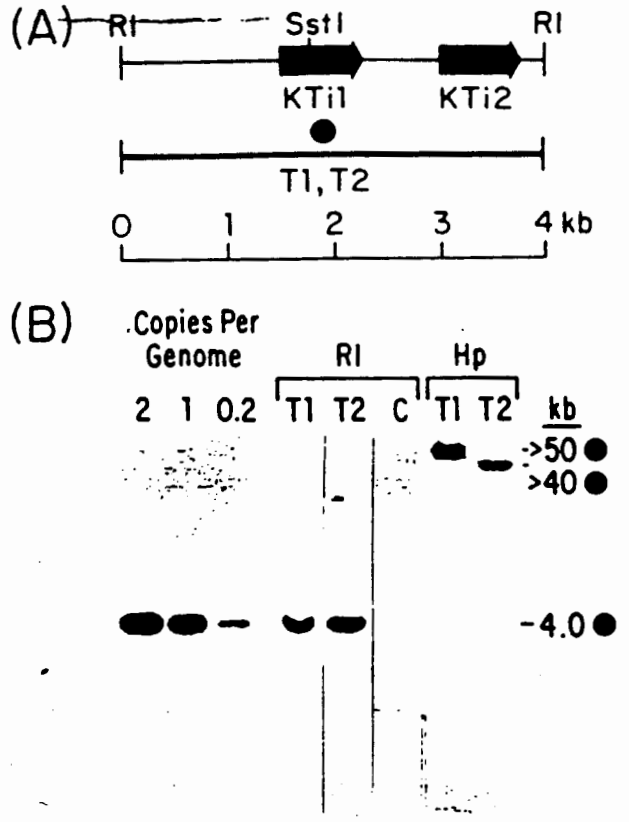


Figure 10. Representation of the KTi1/2 Kunitz Trypsin Inhibitor Gene Region in Transformed Tobacco Plants.

Five micrograms of leaf nuclear DNA from transformed tobacco plants containing the KTi1 and KTi2 trypsin inhibitor genes was digested with the indicated restriction endonucleases, fractionated by electrophoresis, transferred to nitrocellulose, and hybridized with the 4.0-kb EcoRI DNA fragment from plasmid pE4 (Figure 7).

(A) Map of relevant KTi1 and KTi2 gene restriction endonuclease sites and locations. RI and SstI refer to the EcoRI and SstI restriction endonucleases, respectively. Boxes and arrows represent gene locations and transcriptional orientations, respectively (Figure 1). The bracketed line below the restriction map indicates the 4-kb KTi1/2 DNA fragment that was transferred to tobacco and is present in KTi1/2 transformants designated T1 and T2 (Figure 9).

(B) Digestion of tobacco DNAs with EcoRI and HpaI. RI and Hp refer to tobacco DNA digested with EcoRI and HpaI restriction endonucleases, respectively. T1 and T2 refer to DNAs from transformed tobacco plants T1 and T2, respectively. C refers to untransformed tobacco DNA. Reconstruction lanes contained 0.2, 1, and 2 copy equivalents of EcoRI-digested λClone 6 phage DNA (Figure 1). DNA copy equivalents were calculated using a tobacco genome size of 2.4×10^6 kb (Okamuro and Goldberg, 1985). Sizes of restriction fragments in kilobases are shown to the right of the autoradiograms. Circles refer to tobacco DNA fragments that contain the KTi1/2 gene region shown in (A).

accumulated during the mid- to late-maturation periods of seed development (3 wk, 4 wk lanes). Similar accumulation patterns were obtained with seed mRNAs from several independent tobacco plants containing both the KTi1 and KTi2 genes, or the KTi2 gene alone (data not shown). At their peak prevalences (3 to 4 weeks), we estimated that the KTi1 and KTi2 mRNAs represented approximately 2×10^{-3} and 2×10^{-4} of the levels observed in mid-maturation stage soybean embryos, respectively (compare tobacco and soybean mRNA lanes, Figures 11B and 11C). Taken together, these data indicate that (1) the KTi1 and KTi2 genes are regulated temporally during tobacco seed development, (2) the KTi1 and KTi2 gene expression levels in tobacco seeds are lower than those observed in soybean embryos, and (3) that the KTi1 gene is expressed at a higher level than the KTi2 gene in tobacco seeds.

The KTi3 and KTi4 Kunitz Trypsin Inhibitor Genes Are Expressed during Tobacco Seed Development

We hybridized tobacco seed mRNA gel blots with pKT3 plasmid DNA and the 0.9-kb KTi4 gene probe shown in Figure 11D to determine whether the KTi3 and KTi4 Kunitz trypsin inhibitor genes were also expressed correctly during tobacco seed development. As seen in Figures 11E and 11F, both gene probes produced 0.9-kb signals with seed mRNAs from KTi3/4 transformed tobacco plants (KTi3/4 lanes). Both the KTi3 and KTi4 mRNAs accumulated during the mid- to late-maturation stages of tobacco seed development (Figures 11E and 11F, KTi3/4 3 wk, 4 wk lanes). Densitometric analysis indicated that KTi3 mRNA represented approximately 0.1% of the tobacco seed mRNA mass, and was higher in concentration than that estimated for the KTi1 and KTi2 mRNAs (Figures 11B and 11C). Together, these results demonstrate that each Kunitz trypsin inhibitor gene is regulated temporally during tobacco seed development, and that the quantitative differences in KTi1, KTi2, and KTi3 gene expression levels observed in soybean embryos (Figures 5 and 7) are maintained in developing tobacco seeds.

Kunitz Trypsin Inhibitor Protein Is Present in Transformed Tobacco Seeds

We isolated proteins from mature tobacco seeds and then reacted protein gel blots with antibodies raised against the major Kunitz trypsin inhibitor to determine whether the Kunitz trypsin inhibitor genes were expressed at the protein level. As shown in Figure 12 (T1, T2 lanes), the Kunitz trypsin inhibitor antibodies reacted with several proteins from seeds of transformants containing the KTi3 and KTi4 Kunitz trypsin inhibitor genes (Figure 11D). By contrast, the Kunitz trypsin inhibitor antibodies did not react detect-

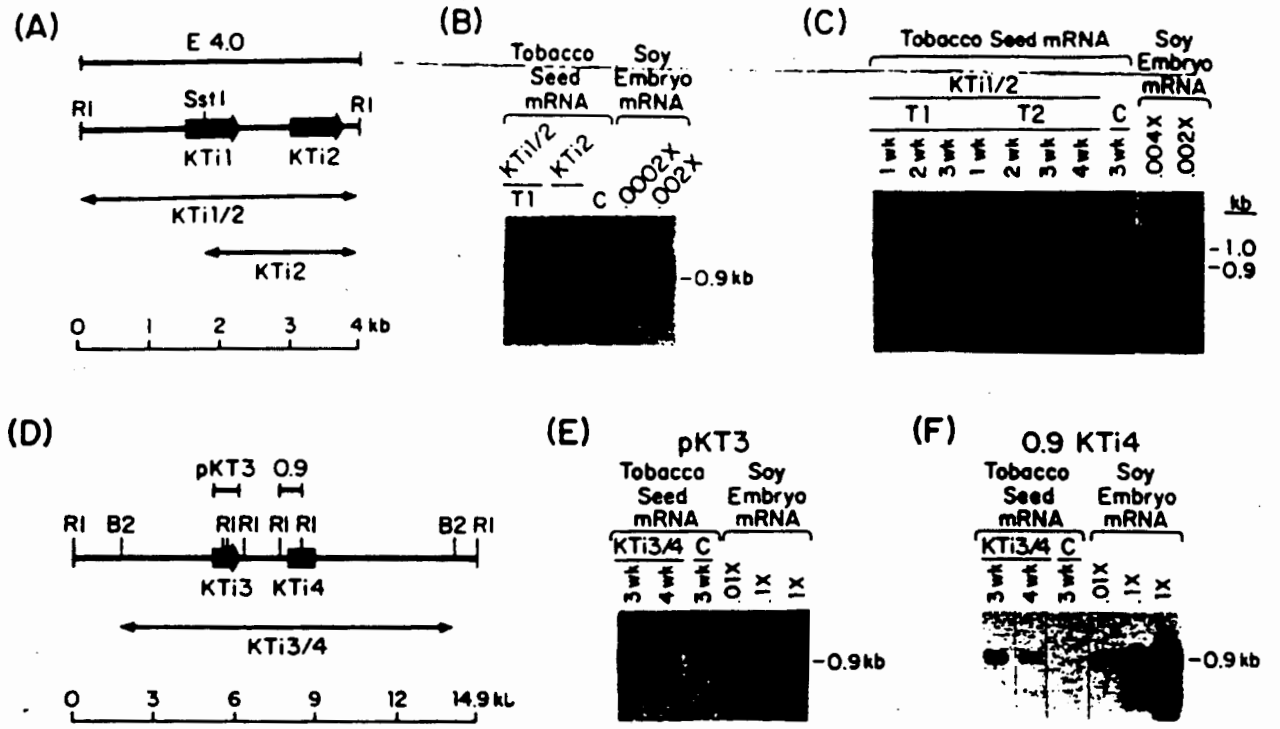


Figure 11. Kunitz Trypsin Inhibitor Gene Expression during Tobacco Seed Development.

(A) The KTi1 and KTi2 gene region. The 4.0-kb KTi1 and KTi2 EcoRI DNA fragment described in Figures 7 and 9 is shown schematically for reference. RI and SstI refer to EcoRI and SstI restriction endonucleases, respectively. Boxes and arrows represent gene locations and transcriptional polarities, respectively. The bracketed line E4.0 refers to the 4.0-kb EcoRI DNA fragment utilized as a hybridization probe in mRNA gel blot experiments shown in **(B)** and **(C)**. KTi1/2 and KTi2 indicate the Kunitz trypsin inhibitor DNA fragments transferred to tobacco plants (Figure 9).

(B) KTi1 and KTi2 gene expression in transformed tobacco seeds. Five micrograms of tobacco seed mRNA harvested 3 weeks after pollination was fractionated by electrophoresis, transferred to nitrocellulose, and hybridized with the E4.0 probe described in **(A)**. KTi1/2 refers to tobacco seed mRNA from transformed plant T1 that contains both KTi1 and KTi2 genes (Figure 9). KTi2 refers to seed mRNA from a pool of five independent tobacco transformants. Genetic analysis indicated that four of the five plants contained the KTi2 trypsin inhibitor gene region described in Figure 9 (data not shown). C refers to untransformed tobacco seed mRNA. The soy embryo .0002X and .002X lanes contained 1 ng and 10 ng of mid-maturation stage embryo mRNA or the equivalent of three and 30 molecules of KTi1/KTi2 Kunitz trypsin inhibitor mRNA, respectively (Figure 7).

(C) KTi1 and KTi2 gene expression in developing tobacco seeds. mRNA gel blots containing 5 μg of tobacco seed mRNA from different developmental stages were hybridized with the labeled E4.0 DNA fragment. T1 and T2 refer to seed mRNAs from KTi1/2 transformed tobacco plants T1 and T2, respectively (Figure 10). C refers to untransformed tobacco seed mRNA. Wk refers to week after pollination. The soy embryo lanes .004X and .002X contained 10 ng and 20 ng of mid-maturation stage embryo mRNA, and represented Kunitz trypsin inhibitor mRNA prevalences of 0.01% and 0.02%, respectively.

(D) The KTi3 and KTi4 Kunitz trypsin inhibitor gene region. A schematic representation of the KTi3/4 gene region described in Figure 9 is shown for reference. RI and B2 refer to the EcoRI and BglII restriction endonucleases, respectively. Boxes and arrows represent approximate gene locations and transcriptional polarities, respectively. The bracketed lines pKT3 and 0.9 refer to the KTi3 and KTi4 gene probes used in mRNA gel blot experiments described in **(E)** and **(F)**, respectively. KTi3/4 indicates the 12.5-kb DNA fragment used for the gene transfer studies described in Figure 9.

(E) KTi3 gene expression in transformed tobacco seeds. mRNA gel blots containing 0.5 μg of tobacco seed mRNA were hybridized with labeled pKT3 plasmid DNA. KTi3/4 and C refer to transformed and untransformed tobacco seed mRNAs, respectively. Wk refers to week after pollination. The soy embryo .01X, .1X, and 1X lanes contained 0.005 μg, 0.05 μg, and 0.5 μg of mid-maturation stage embryo mRNA and correspond to KTi3 mRNA prevalences of 0.045%, 0.45%, and 4.5%, respectively (Figures 5 and 7).

(F) KTi4 gene expression in tobacco seeds. mRNA gel blots containing 0.25 μg of tobacco seed mRNAs were reacted with the 0.9-kb KTi4 gene probe. KTi3/4 and C refer to transformed and untransformed tobacco seed mRNAs, respectively. The soy embryo .01X, .1X, and 1X lanes contained 0.0025 μg, 0.025 μg, and 0.25 μg of mid-maturation stage mRNA, respectively.

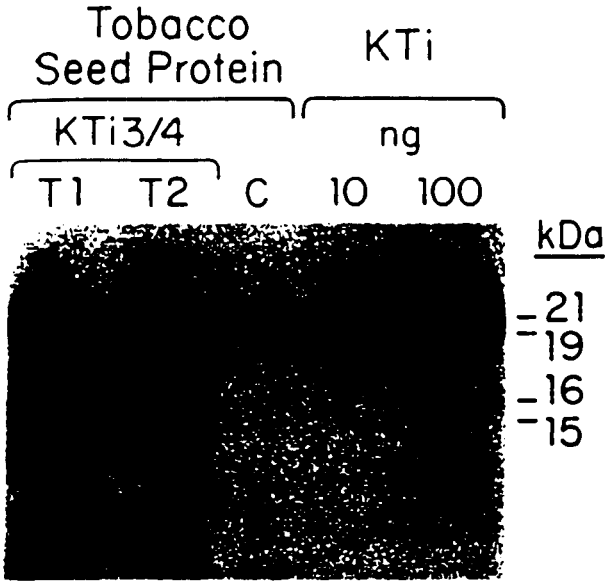


Figure 12. Presence of Soybean Kunitz Trypsin Inhibitor Protein in Transformed Tobacco Seeds.

Sixty micrograms of tobacco seed protein (12 seed equivalents) was fractionated by electrophoresis on polyacrylamide gels, transferred to nitrocellulose, and reacted with Kunitz trypsin inhibitor antibodies. Kunitz trypsin inhibitor protein was detected by using a horseradish peroxidase immunoassay. T1 and T2 refer to seed proteins from KTi3/4 transformed tobacco plants T1 and T2, respectively. C refers to untransformed tobacco seed protein. The KTi lanes contained 10 ng and 100 ng of purified soybean Kunitz trypsin inhibitor protein.

ably with seed proteins from tobacco plants transformed with the KTi1 and KTi2 genes (data not shown), or with proteins from untransformed tobacco seeds (Figure 12, C lane). The 19-kD and 21-kD seed protein bands from KTi3/4 transformants were comparable in size with those obtained with purified Kunitz trypsin inhibitor protein (KTi lanes) and with total soybean seed protein (data not shown). We presume that the 15-kD and 16-kD proteins detected with transformed seeds represent degradation products (Figure 12, T1, T2 lanes). Relative to known Kunitz trypsin inhibitor standards, we estimated that the Kunitz trypsin inhibitor represented 0.2% of the tobacco seed protein mass, or approximately 1/20th of the level found in soybean. Together, these data demonstrate that at least one Kunitz trypsin inhibitor gene (KTi3) is expressed at the protein level in transformed tobacco seeds.

Kunitz Trypsin Inhibitor Genes Are Expressed in Tobacco Organ Systems

We hybridized tobacco leaf, stem, and root mRNA gel blots with Kunitz trypsin inhibitor gene probes (Figures 11A and 11D) to establish whether the Kunitz trypsin inhibitor genes were expressed in organ systems of the mature tobacco plant. Figures 13A and 13B show that the 4.0-kb KTi1/2 gene probe hybridized with 0.9-kb leaf and stem mRNAs from tobacco plants containing both the KTi1 and KTi2 genes (TOB-T; L, S lanes, Figure 13A), or the KTi2 gene alone (TOB-T; L, S lanes, Figure 13B). By contrast, no detectable signals were obtained with either KTi1/2 or KTi2 transformed root mRNAs (TOB-T; R lanes, Figures 13A and 13B), or with untransformed leaf, stem, and root mRNAs (TOB-C lanes, Figure 13A). Figures 13C and 13D show that the KTi3 and KTi4 gene probes produced 0.9-kb signals with tobacco leaf mRNA from plants containing the KTi3 and KTi4 Kunitz trypsin inhibitor genes (TOB-T; L lanes), and did not react detectably with either stem or root mRNAs from the same plants (TOB-T; S and R lanes). Together, these results demonstrate that the Kunitz trypsin inhibitor genes are expressed in tobacco mature plant organ systems, and that the organ system expression programs for individual Kunitz trypsin inhibitor genes differ as in soybean (Figure 8).

DISCUSSION

Many Kunitz Trypsin Inhibitor Genes Are Present in the Soybean Genome

Kunitz trypsin inhibitor genes represent one of several seed protein gene families that are highly regulated during the plant life cycle (Goldberg et al., 1989; Shotwell and Larkins, 1989). Both DNA gel blot studies and characterization of genomic clones indicate that there are at least 10 Kunitz trypsin inhibitor genes in the soybean genome (Figure 1). At least four of these genes (KTi1, KTi2, KTi3, and KTi4) are expressed at the mRNA level in both soybean and transformed tobacco plants (Figures 5, 7, and 11). cDNA clones have been identified that are specific for two additional Kunitz trypsin inhibitor genes that are present on λClone 4 and λClone 30 phages (Figure 1B; G. de Paiva, K.D. Jofuku, and R.B. Goldberg, unpublished results). Thus, we have evidence that at least six of the 10 Kunitz trypsin inhibitor genes are active in soybean embryos.

Genetic studies indicated that a single gene is responsible for encoding a seed protein with Kunitz trypsin inhibitor *activity* (Orf and Hymowitz, 1979). Three allelic forms of this gene (Ti^a, Ti^b, Ti^c) encode proteins that differ by only a few amino acids (Orf and Hymowitz, 1979; Kim et al., 1985). Soybean lines have been identified that have reduced amounts of Kunitz trypsin inhibitor protein and lack detectable Kunitz trypsin inhibitor *activity* (Orf and Hymowitz, 1979). The null phenotype is due to a mutation in the Kunitz trypsin inhibitor structural gene (e.g., Ti^a), and

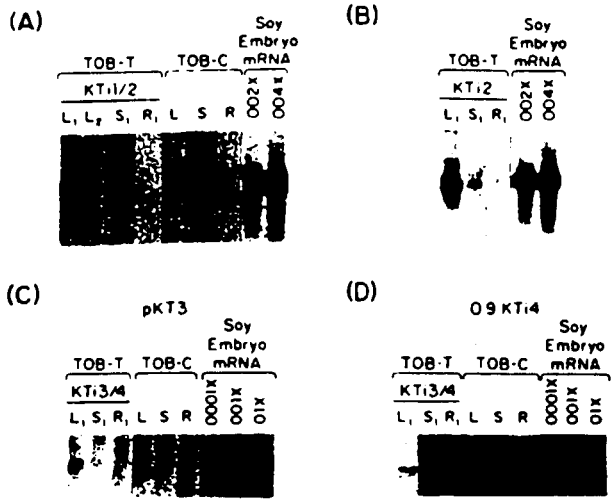


Figure 13. Kunitz Trypsin Inhibitor Gene Expression in Mature Tobacco Plant Organ Systems.

mRNA gel blots containing 5 μg of tobacco leaf (L), stem (S), and root (R) mRNAs were hybridized with the Kunitz trypsin inhibitor KTi1/2 E4.0, pKT3, and 0.9-kb KTi4 gene probes shown in Figure 11. TOB-T and TOB-C refer to transformed and untransformed tobacco mRNAs, respectively.

(A) KTi1/2 gene expression in transformed tobacco leaf, stem, and root. KTi1/2 L_1, S_1, and R_1 refer to mRNAs from KTi1/2 transformed tobacco plant T1, and L_2 refers to leaf mRNA from KTi1/2 transformant T2 (Figures 10 and 11). The .002X and .004X reconstruction lanes contained 10 ng and 20 ng of soybean mid-maturation stage embryo mRNA and correspond to Kunitz trypsin inhibitor mRNA prevalences of 0.006% and 0.01%, respectively (Figures 5 and 7).

(B) KTi2 gene expression in transformed tobacco leaf, stem, and root. KTi2 L_1, S_1, and R_1 refer to mRNAs from KTi2 tobacco transformant T1 (Figure 11). The .002X and .004X lanes contained soybean mid-maturation stage embryo mRNA and correspond to Kunitz trypsin inhibitor mRNA prevalences of 0.006% and 0.01%, respectively.

(C) KTi3 gene expression in mature tobacco plant organ systems. KTi3/4 L_1, S_1, and R_1 lanes contain leaf, stem, and root mRNAs from KTi3/4 transformed tobacco plant T1. The .0001X, .001X, and .01X lanes contained soybean mid-maturation stage embryo mRNA and correspond to approximately 2, 20, and 200 KTi3 mRNA molecules per cell (Figures 5 and 7).

(D) KTi4 gene expression in tobacco plant organ systems. KTi3/4 L_1, S_1, and R_1 refer to mRNAs from KTi3/4 transformed plant T1. The .0001X, .001X, and .01X reconstruction lanes contained 0.0005 μg, 0.005 μg, and 0.05 μg of soybean mid-maturation stage (77 DAF) embryo mRNA, respectively.

is inherited as a recessive allele (*ti*; Orf and Hymowitz, 1979). These results suggest that most of the Kunitz trypsin inhibitor genes identified in our experiments (Figure 1) do not encode proteins with trypsin inhibitor activity.

DNA sequence analysis of the KTi3 Kunitz trypsin inhibitor gene (Figure 4; Jofuku et al., 1989) indicates that it encodes a protein identical to that of the Kunitz trypsin inhibitor Ti^a allelic form (Kim et al., 1985). A null line that lacks Kunitz trypsin inhibitor *activity* has three mutations in the KTi3 gene (Jofuku et al., 1989). These mutations cause a frameshift that results in premature termination of KTi3 mRNA translation, and lead to a 100-fold reduction of KTi3 mRNA in soybean embryos (Jofuku et al., 1989). The KTi1 and KTi2 mRNAs are unaffected by the KTi3 gene mutation (Jofuku et al., 1989), and are probably responsible in part for the residual Kunitz trypsin inhibitor cross-reacting material found in mutant seeds (K.D. Jofuku and R.B. Goldberg, unpublished results).

The KTi1 and KTi2 gene sequences (Figure 2) indicate that they encode proteins with only 70% similarity to the Kunitz trypsin inhibitor Ti^a form (Figure 4). Inspection of the KTi1 and KTi2 protein amino acid sequences (Figure 4) shows that they lack the Arg-63 and Ile-64 amino acids responsible for trypsin inhibitor activity (Laskowski and Kato, 1980; Ryan, 1981). Nor do they contain other amino acids at these sites that would cause the KTi1 and KTi2 proteins to have chymotrypsin or elastase inhibitor activities (Ryan, 1981). These findings correlate well with the genetic analysis of Kunitz trypsin inhibitor null lines (Orf and Hymowitz, 1979), and strongly suggest that the KTi3 Kunitz trypsin inhibitor gene is responsible for encoding most of the Kunitz trypsin inhibitor activity *and* protein found in soybean seeds.

Several Kunitz Trypsin Inhibitor Genes Are Tandemly Linked

We do not know yet how all 10 Kunitz trypsin inhibitor genes are organized with respect to each other in soybean chromosomes because we have not attempted to link the genomic clones by either DNA walking studies (Harada, Barker, and Goldberg, 1989; Nielsen et al., 1989) or by genetic studies with restriction fragment length polymorphisms (Cho, Davies, and Nielsen, 1989). Analysis of individual Kunitz trypsin inhibitor genomic clones (Figure 1B), however, demonstrated that several Kunitz trypsin inhibitor genes are linked in tandem pairs. The most striking example of this form of organization is the KTi1 and KTi2 gene pair (Figures 1B and 1C). These genes are approximately 0.5 kb apart, are in the same transcriptional orientation, and have >95% nucleotide sequence similarity in their 5′, coding, and 3′ regions (Figures 2 and 3). The KTi1 and KTi2 Kunitz trypsin inhibitor genes also show extensive similarity to the KTi3 5′, coding, and 3′ regions (Figure 3). These results suggest that the KTi1, KTi2, and KTi3 genes probably originated from the duplication of an ancient relative, and that the KTi1 and KTi2 gene pair probably reflects a recent duplication and/or gene conversion event.

Kunitz Trypsin Inhibitor Genes Have Different Expression Programs

A significant aspect of our results is the observation that individual Kunitz trypsin inhibitor genes have different qualitative and/or quantitative expression programs. All Kunitz trypsin inhibitor genes investigated encode 0.9-kb mRNAs that are temporally regulated in soybean embryos and in seeds of transformed tobacco plants (Figures 5, 7, and 11). In both soybean and tobacco, however, individual Kunitz trypsin inhibitor mRNAs accumulate to different levels during seed development (Figures 5, 7, and 11). The KTi3 mRNA is the most prevalent (Figures 5 and 11), and is followed in prevalence by the KTi1 and KTi2 mRNAs (Figures 7 and 11). Transcription studies demonstrated that the difference in KTi3 and KTi1/2 mRNA levels is due primarily to a difference in the transcription rates of their corresponding genes (Figure 6).

Kunitz trypsin inhibitor genes are expressed at low levels in organ systems of soybean (Figure 8) and transformed tobacco plants (Figure 13). The reduced organ system expression levels are due in part to lower Kunitz trypsin inhibitor gene transcription rates in the leaf, root, and stem (Walling et al., 1986). Expression of the Kunitz trypsin inhibitor gene family in mature plant organ systems is similar to that observed for the lectin gene (Okamuro, Jofuku, and Goldberg, 1986; Goldberg et al., 1989), but differs from the glycinin and β-conglycinin storage protein genes that are expressed only during embryogenesis (Harada et al., 1989; Nielsen et al., 1989).

KTi1 and KTi2 mRNAs are present in the soybean leaf, stem, and root (Figure 8). In transformed tobacco plants these mRNAs are detected in both the leaf and stem but not the root (Figure 13). The absence of KTi1 and KTi2 mRNAs in transformed tobacco roots may indicate that the KTi1 and KTi2 genes are expressed at levels undetectable by our mRNA gel blot procedure, that position effects altered the KTi1 and KTi2 expression programs, or that the hybridization signal obtained with soybean root mRNA was due to expression of a related Kunitz trypsin inhibitor family member. By contrast with these findings, KTi3 and KTi4 mRNAs are detectable only in the leaf of both soybean (Figure 8) and transformed tobacco plants (Figure 13). We conclude from these observations that individual Kunitz trypsin inhibitor genes can have different qualitative expression programs in mature plant organ systems.

In situ localization studies shown in the accompanying paper (Perez-Grau and Goldberg, 1989) indicate that other differences occur in Kunitz trypsin inhibitor gene expression programs. These studies showed that KTi3 mRNA is more prevalent than the KTi1/2 mRNAs in the soybean embryo axis. In addition, KTi3 mRNA accumulates in specific globular stage embryo cells long before the KTi1 and KTi2 mRNAs are visualized. Finally, KTi3 and KTi1/2 mRNAs are distributed differently within embryo cotyledon cells, and have distinct spatial accumulation programs.

Control of Kunitz Trypsin Inhibitor Gene Expression—A Simple Hypothesis

What are the molecular events responsible for specifying the individual Kunitz trypsin inhibitor gene expression programs? The simplest hypothesis is that all Kunitz trypsin inhibitor genes share elements that program their expression during embryogenesis (Goldberg, 1986). Quantitative differences in expression levels (e.g., KTi3 versus KTi1 and KTi2 genes) could be due to the number of embryo *cis*-control elements and/or the presence of additional elements that enhance gene transcription rates in embryo cells. By contrast, expression in mature plant organ systems may be due to the presence of other control elements that are specific for cells within the leaf, root, or stem (Goldberg, 1986). Different combinations of these elements would cause Kunitz trypsin inhibitor genes to have different expression programs in mature plant organ systems.

At the present time we have no evidence for or against this hypothesis. Recent experiments with the cauliflower mosaic virus 35S gene indicate, however, that a hierarchy of discrete control elements is responsible for programming 35S gene expression to specific tobacco cell and tissue types (Benfey, Ren, and Chua, 1989). The Kunitz trypsin inhibitor gene family provides a unique opportunity to dissect control elements responsible for programming gene expression at all stages of the plant life cycle. What these elements are, and how they interact with regulatory gene products, remain to be determined.

METHODS

Isolation of Kunitz Trypsin Inhibitor Phages and Plasmids

The isolation and characteristics of the A-37 Kunitz trypsin inhibitor cDNA plasmid were described previously (Goldberg et al., 1981; Jofuku et al., 1989). The pKT3 cDNA plasmid representing the major Kunitz trypsin inhibitor gene KTi3 was isolated from a soybean mid-maturation stage embryo cDNA library (J.J. Harada and R.B. Goldberg, unpublished results) as described previously (Jofuku et al., 1989). Recombinant λ Charon 4 phages containing Kunitz trypsin inhibitor genes were isolated from a soybean Forrest variety EcoRI genome library (Fischer and Goldberg, 1982), and an AluI/HaeIII linker library of soybean Dare variety DNA. The linker library was constructed according to the procedure of Maniatis et al. (1978).

DNA Isolation and Labeling

Plant, phage, and plasmid DNAs were isolated as described previously (Jofuku and Goldberg, 1988). Synthetic oligonucleo-

tides were synthesized and purified according to established procedures (Matteucci and Caruthers, 1981; Vorndam and Kerschner, 1986). DNAs were labeled by nick translation under conditions specified by Bethesda Research Laboratories. Oligonucleotides were 5′ end-labeled as described by Maniatis et al. (1978).

Polysomal mRNA Isolation

Soybean and tobacco polysomal poly(A) mRNAs were isolated as described (Cox and Goldberg, 1988).

Seed Protein Isolation

Tobacco seed proteins were extracted according to Sano and Kawashima (1983).

Filter Hybridization Studies

DNA and mRNA gel blot experiments were carried out according to the procedures of Wahl, Stern, and Stark (1979) and Thomas (1983), respectively. mRNA gel blot studies with oligonucleotide probes were performed as described by Woods et al. (1982) and by Whitehead et al. (1983). Dot blot experiments were carried out as described by Thomas (1983). Protein gel blot studies were performed as described by Johnson et al. (1984).

R-Loop Analysis

R-loops were formed between Kunitz trypsin inhibitor phage DNAs and soybean mid-maturation stage embryo mRNAs as described previously (Fischer and Goldberg, 1982).

Primer Extension Analysis

A greater than 50-fold mass excess of a 22-nucleotide synthetic oligonucleotide primer (P_1; Figure 2) was hybridized with mid-maturation stage embryo mRNA or with denatured pE4 plasmid DNA (Figure 7) at 42°C in 10 mM Tris (pH 7.6), 300 mM NaCl, and 0.5 mM EDTA. The primer was then extended by reverse transcriptase using only ^{32}P-TTP (>600 Ci/mmol) as described by Murray, Hoffman, and Jarvis (1983). Labeled primer extension products were subjected to electrophoresis on a 20% polyacrylamide gel under denaturing conditions, and then visualized by autoradiography.

^{32}P-Nuclear RNA Synthesis

Synthesis of labeled RNA from isolated soybean embryo nuclei was carried out as described by Walling et al. (1986) and Cox and Goldberg (1988).

Tobacco Transformation and Plant Regeneration

Tobacco leaf discs and protoplasts were transformed with the KTi1/2, KTi2, and KTi3/4 gene regions (Figure 9) according to the procedures of Horsch et al. (1985), Okamuro et al. (1986), and Barker, Harada, and Goldberg (1988).

ACKNOWLEDGMENTS

We thank our former colleagues Drs. Bob Fischer, John Harada, Tom Sims, and Linda Walling, as well as Jessie Truettner and Roberta Schipper, for their invaluable advice and assistance during the course of this study. We also thank Dr. Bill Timberlake for many thoughtful suggestions, and Dr. Lila Vodkin for Kunitz trypsin inhibitor antibodies, immunoselected polysomes, and valuable comments. This research was supported by a United States Department of Agriculture grant (to R.B.G.). K.D.J. was supported by an ARCO Plant Cell Research Institute Predoctoral Fellowship.

Received September 11, 1989.

REFERENCES

Barker, S.J., Harada, J.J., and Goldberg, R.B. (1988). Cellular localization of soybean storage protein mRNA in transformed tobacco plants. Proc. Natl. Acad. Sci. USA **85**, 458–462.

Benfey, P., Ren, L., and Chua, N.-H. (1989). The CaMV enhancer contains at least two domains which can confer different developmental and tissue-specific expression programs. EMBO J. **8**, 2195–2202.

Cho, T.-J., Davies, C.S., and Nielsen, N.C. (1989). Inheritance and organization of glycinin genes in soybean. Plant Cell **1**, 329–337.

Cox, K.H., and Goldberg, R.B. (1988). Analysis of gene expression. In Plant Molecular Biology: A Practical Approach, C.H. Shaw, ed (Oxford: IRL Press), pp. 1–35.

Fischer, R.L., and Goldberg, R.B. (1982). Structure and flanking regions of soybean seed protein genes. Cell **29**, 651–660.

Goldberg, R.B. (1986). Regulation of plant gene expression. Philos. Trans. R. Soc. Lond. **B314**, 343–353.

Goldberg, R.B., Barker, S.J., and Perez-Grau, L. (1989). Regulation of gene expression during plant embryogenesis. Cell **56**, 149–160.

Goldberg, R.B., Hoschek, G., and Vodkin, L.O. (1983). An insertion sequence blocks the expression of a soybean lectin gene. Cell **33**, 465–475.

Goldberg, R.B., Hoschek, G., Ditta, G.S., and Breidenbach, R.W. (1981). Developmental regulation of cloned superabundant embryo mRNAs in soybean. Dev. Biol. **83**, 218–231.

Hain, R., Stabel, P., Czernilofsky, A.P., Steinbiss, H.H., Herrera-Estrella, L., and Schell, J. (1985). Uptake, integration, expression, and genetic transmission of a selectable chimeric gene by plant protoplasts. Mol. Gen. Genet. **199**, 161–168.

Harada, J.J., Barker, S.J., and Goldberg, R.B. (1989). Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. Plant Cell **1**, 415–425.

[illegible], B.A., Gatehouse, A.M.R., Sheerman, S.E., Barker, R.F., [illegible], Boulter, D. (1986). A novel mechanism of insect resistance engineered into tobacco. Nature 330, 160–163.

Horisberger, M., and Tacchini-Vonlanthen, M. (1983). Ultrastructural localization of Kunitz trypsin inhibitor on thin sections of soybean by the gold method. Histochemistry 77, 37–50.

Horsch, R.B., Fry, J.E., Hoffman, N.L., Eichholtz, D., Rogers, S.G., and Fraley, R.T. (1985). A simple and general method for transferring genes into plants. Science 227, 1229–1231.

Jofuku, K.D., and Goldberg, R.B. (1988). Analysis of plant gene structure. In Plant Molecular Biology: A Practical Approach, C.H. Shaw, ed (Oxford: IRL Press), pp. 37–66.

Jofuku, K.D., Schipper, R.D., and Goldberg, R.B. (1989). A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. Plant Cell 1, 427–435.

Johnson, D.A., Gautsch, J.W., Sportsman, J.R., and Elder, J.H. (1984). Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. Gene Anal. Technol. 1, 3–8.

Kim, S.-H., Hara, S. Hase, S., Ikenaka, T., Toda, H., Kitamura, K. and Kaizuma, N. (1985). Comparative study on amino acid sequences of Kunitz-type soybean trypsin inhibitors, *Ti* ᵃ, *Ti*ᵇ, and *Ti*ᶜ. J. Biochem. 98, 435–448.

Laskowski, M., Jr., and Kato, I. (1980). Protein inhibitors of proteinases. Annu. Rev. Biochem. 49, 593–629.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K., and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eukaryotic DNA. Cell 15, 687–701.

Ma[illegible]ci, M.D., and Caruthers, M.H. (1981). Synthesis of deoxy[illegible]eotides on a polymer support. J. Am. Chem. Soc. 103, 3185–3191.

Murray, M.G., Hoffman, L.M., and Jarvis, N.P. (1983). Improved yield of full-length phaseolin cDNA clones by controlling premature anticomplementary DNA synthesis. Plant Mol. Biol. 2, 75–84.

Nielsen, N.C., Dickinson, C.D., Cho, T.-J., Thanh, V.H., Scallon, B.J., Fischer, R.L., Sims, T.L., Drews, G.N., and Goldberg, R.B. (1989). Characterization of the glycinin gene family in soybean. Plant Cell 1, 313–328.

Norioka, N., Hara, S., Ikenaka, T., and Abe, J. (1988). Distribution of the Kunitz and the Bowman-Birk family proteinase inhibitors in leguminous seeds. Agric. Biol. Chem. 52, 1245–1252.

Okamuro, J.K., and Goldberg, R.B. (1985). Tobacco single-copy DNA is highly homologous to sequences present in the genomes of its diploid progenitors. Mol. Gen. Genet. 198, 290–298.

Okamuro, J.K., Jofuku, K.D., and Goldberg, R.B. (1986). Soybean seed lectin gene and flanking nonseed protein genes are developmentally regulated in transformed tobacco plants. Proc. Natl. Acad. Sci. USA 83, 8240–8244.

Orf, J.H., and Hymowitz, T. (1979). Inheritance of the absence of Kunitz trypsin inhibitor in seed protein of soybeans. Crop Sci. 19, 107–109.

Ozawa, K., and Laskowski, M., Jr. (1966). Reactive sites of trypsin inhibitors. J. Biol. Chem. 241, 3955–3961.

Perez-Grau, L., and Goldberg, R.B. (1989). Soybean seed protein genes are regulated spatially during embryogenesis. Plant Cell 1, 1095–1109.

Ryan, C.A. (1981). Proteinase inhibitors. In Biochemistry of Plants, A. Marcus, ed (New York: Academic Press), Vol. 6, pp. 351–370.

Ryan, C.A. (1988). Proteinase inhibitor gene families: Tissue specificity and regulation. In Temporal and Spatial Control of Plant Genes. D.P.S. Verma and R.B. Goldberg, eds (Vienna: Springer-Verlag), pp. 223–233.

Sano, M., and Kawashima, N. (1983). Isolation and partial characterization of the major seed protein from *Nicotiana tabacum*, and their accumulation during development. Agric. Biol. Chem. 47, 1305–1310.

Shotwell, M.A., and Larkins, B.A. (1989). The biology of seed storage proteins. In The Biochemistry of Plants, A. Marcus, ed (New York: Academic Press), Vol. 15, pp. 297–345.

Stahlhut, R.W., and Hymowitz, T. (1983). Variation in the low molecular weight proteinase inhibitors of soybean. Crop Sci.23, 766–769.

Sweet, R.M., Wright, H.T., Janin, J., Chothia, C.H., and Blow, D.M. (1974). Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6-Å resolution. Biochemistry 13, 4212–4228.

Tan-Wilson, A.L. (1988). Subclassification of soybean Bowman-Birk isoinhibitors. J. Am. Oil Chem. Soc. 65, 1475–1478.

Thomas, P.S. (1983). Hybridization of denatured RNA transferred or dotted to nitrocellulose. Methods Enzymol. 100, 255–265.

Vodkin, L.O. (1981). Isolation and characterization of messenger RNAs for seed lectin and Kunitz trypsin inhibitor in soybeans. Plant Physiol. 68, 766–771.

Vodkin, L.O., and Raikhel, N.V. (1986). Soybean lectin and related proteins in seeds and roots of Le^+ and Le^- soybean varieties. Plant Physiol. 81, 558–565.

Vorndam, A.V., and Kerschner, J. (1986). Purification of small oligonucleotides by polyacrylamide gel electrophoresis and transfer to diethylaminoethyl paper. Anal. Biochem. 152, 221–225.

Wahl, G.M., Stern, M., and Stark, G.R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA 76, 3693–3697.

Walling, L., Drews, G.N., and Goldberg, R.B. (1986). Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. Proc. Natl. Acad. Sci. USA 83, 2123–2127.

Whitehead, A.S., Goldberger, G., Woods, D.E., Markham, A.F., and Colten, H.R. (1983). Use of a cDNA clone for the fourth component of human complement (C4) for analysis of a genetic deficiency of C4 in guinea pig. Proc. Natl. Acad. Sci. USA 80, 5387–5391.

Woods, D.E., Markham, A.F., Ricker, A.T., Goldberger, G., and Colten, H.R. (1982). Isolation of cDNA clones for the human complement protein factor B, a class III major histocompatibility complex gene product. Proc. Natl. Acad. Sci. USA 79, 5661–5665.

Genetic modification of the storage lipids of plants

Anthony J Kinney

Du Pont Experimental Station, Wilmington, USA

Edible vegetable oils often require extensive chemical modification to increase their utility. A number of genetically altered plants that produce ready-modified oils have recently been reported. Genetic modification of oil-producing plants has been achieved, both by chemically induced or radiation-induced mutagenesis, and by genetic engineering. Approaches that combine mutants and transgenics have great potential for the production of commercially valuable phenotypes.

Current Opinion in Biotechnology 1994, 5:144–151

Introduction

Many plants use oil as an energy-rich storage compound, and a number of these plants, such as soybean, canola, olive, sunflower and safflower, are utilized as a source of oils for human consumption. The major lipid component of vegetable oils is triacylglycerol, which consists of three fatty acids esterified to the hydroxyl groups of the three carbons of a glycerol molecule. In edible oils, the predominant triacylglycerols have fatty acids that are 16 and 18 carbons in length. The 16-carbon fatty acid is normally unsaturated (i.e. palmitic acid, 16:0) and, although the unsaturated 18-carbon equivalent (i.e. stearic acid, 18:0) is also found, the majority of 18-carbon fatty acids in edible oils normally contain one, two or three double bonds (i.e. oleic acid, 18:1; linoleic acid, 18:2; and linolenic acid, 18:3).

The usefulness and quality of the oil depends upon the relative amounts of the above five fatty acid species. Oils rich in oleic acid are oxidatively stable, tend to have good flavor characteristics and may also have positive health benefits, whereas oils rich in polyunsaturated fatty acids (i.e. linoleic and linolenic acids) have poor flavor and low stability [1••]. In soybean oils, for example, a strong negative correlation ($r=-0.89$) exists between linoleic acid content and flavor quality and intensity [2••]. Most unmodified vegetable oils contain greater than 50% polyunsaturated fatty acids [3], and these oils require partial chemical hydrogenation if they are to be heated during use. Extensive hydrogenation of vegetable oils is required when the oil is used as a replacement for some animal fats, such as in the production of margarines [P1].

Within the past ten years or so, the primary focus of genetic improvement of oilseed crops has shifted from increased production and disease resistance to the improvement of oil quality by genetic, rather than chemical, means. Initially, the genetic modification of oilseeds was achieved by germplasm selection and chemical mutagenesis. More recently, the cloning of a number of important plant genes encoding lipid biosynthetic enzymes has led to the first transgenic oilseed plants with a modified seed oil composition (see Table 1).

In this review, I briefly overview recent progress in mutagenesis and in the introduction of cloned genes into oilseed plants, concentrating mainly on the production of edible oils. The modification, by genetic engineering, of domesticated oilseed plants to produce specialized industrial oils (e.g. ricinoleic acid) has also recently received considerable attention, and is reviewed elsewhere [4•,5••].

Decreased linolenic acid content

The synthesis of linolenic acid from linoleic acid is the result of a desaturation reaction catalyzed by a membrane-associated omega-3 desaturase. Both microsomal and plastid isoforms of omega-3 desaturases are found, and, in oilseeds, both types probably contribute to the final triacylglycerol linolenic acid content [6••]. The main substrate for microsomal omega-3 desaturation is linoleoyl phosphatidylcholine, whereas that for plastid omega-3 desaturation is linoleic acid esterified mainly to galactolipids and sulfolipids (see Fig. 1).

Genetic mutations that lead to a reduced seed triacylglycerol linolenic acid content have been identified in flax [7•], canola [P2•] and soybean [8,9]. Both maize and sunflower appear to be naturally deficient in seed omega-3 desaturases because their seed triacylglycerol linolenic acid content is negligible [1% or less of the total triacylglycerol fatty acid content (FAC)]. The canola mutant *IMC01* has a seed triacylglycerol linolenic acid

Abbreviations

ACP—acyl carrier protein; CoA—coenzyme A; FAC—total triacylglycerol fatty acid content; KAS I—β-ketoacylsynthetase I; KAS II—β-ketoacylsynthetase II.

Table 1. Some of the soybean, canola and *Arabidopsis* mutants and transgenics with altered fatty acid content.

Genetic modification	Species	Phenotype	Reference
Mutation			
IMC01	Canola	Reduced linolenic acid (18:3) in seed	[P2•]
C1640	Soybean	Reduced linolenic acid (18:3) in seed (*fan 1*), which varies with temperature	[9,10]
FAD 3	*Arabidopsis*	Reduced linolenic acid (18:3) in all tissues	[6••,13]
A23	Soybean	Reduced linolenic acid (18:3) in seed (*fan 2*), which varies with temperature, but averages ~2% less than the wild type at similar temperature	[15]
*A16/A17**	Soybean	Reduced linolenic acid (18:3) (*fan 1* + *fan 2*), which constitutes ~2% of seed triacylglycerol	[15]
FAD 2	*Arabidopsis*	Reduced linoleic acid (18:2) and increased oleic acid (18:1)	[6••,13]
A5••	Soybean	Reduced linoleic acid (18:2) and increased oleic acid (18:1), the latter of which varies with temperature	[8]
IMC129†	Canola	Reduced linoleic acid (18:2) and raised linolenic acid (18:3). Leaves and roots also have increased oleic acid (18:1)	[22•,P5,P6•]
C1727	Soybean	Increased palmitic acid (16:0) to 17% (FAC)	[26]
A6	Soybean	Increased stearic acid (18:0) to 30% (FAC)	[25,27•]
C1726	Soybean	Reduced palmitic acid (16:0) to 8.5% (FAC)	[26]
Transgene			
Omega-3 desaturase antisense	Soybean	Reduced linolenic acid (18:3) in seed to <1.5% (FAC)	‡
Omega-6 desaturase antisense	Soybean and canola	Reduced linoleic acid (18:2) and increased oleic acid (18:1) [80% (FAC) or greater] in seed	§
Acyl-ACP thioesterase sense	Canola	Increased palmitic acid (16:0) and stearic acid (18:0) [>20% (FAC)] in seed	[P10•]
Delta-9 desaturase antisense	Canola	Increased stearic acid (18:0) [>30% (FAC)] in seed	[29••]

•*A16/A17* are segregants from a cross between *A5* and *A23*. ••Also contains the *FAN 1* mutation allelic to the *C1640* mutation (i.e. contains reduced linolenic acid). †A number of mutants with a similar phenotype have been described. ‡AJ Kinney, unpublished data. §WD Hitz, AJ Kinney, unpublished data. (See text for a more detailed explanation of these phenotypes.)

content of 2–3% (FAC), compared with a content of about 6–8% (FAC) in varieties such as Westar. The oil produced from these mutant seeds has improved organoleptic characteristics and an increased oxida-

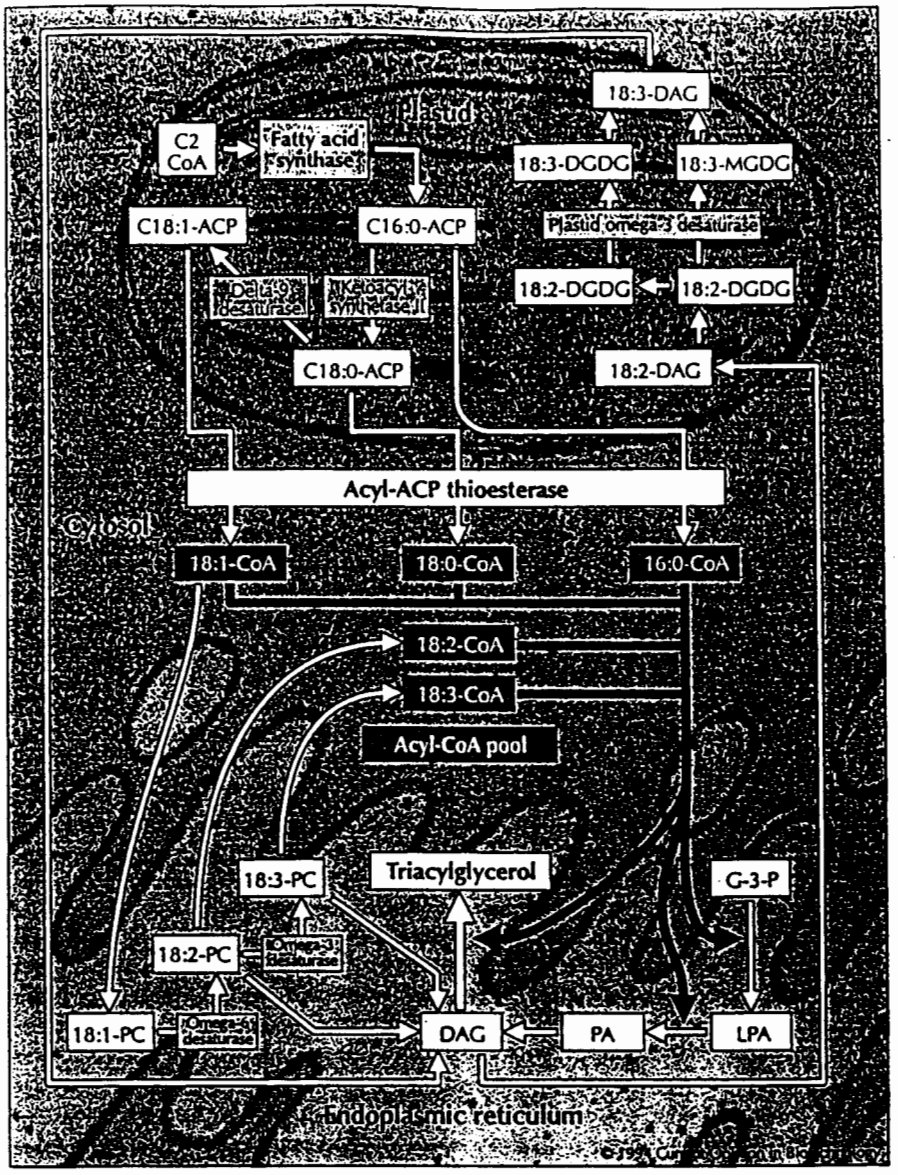


Fig. 1. Scheme illustrating some important reactions of fatty acid biosynthesis and desaturation in the developing oil seed. The level of triacylglycerol, the main storage lipid, is dependent on the balance of activities of a number of enzymes. Fatty acid biosynthesis and elongation involves important plastid enzymes, such as fatty acid synthase (in bold type, as it represents many enzymes), β-ketoacylsynthetase II and delta-9 desaturase. Plastid omega-3 desaturase catalyzes the desaturation of linoleic acid (18:2) to linolenic acid (18:3). (See text and [6••,13] for further details of these enzymes.) Fatty acids are liberated from ACP by acyl-ACP thioesterase and exported to the cytosolic pool (dark shading). This pool is also supplemented by the activities of omega-3 desaturase and omega-6 desaturase. Acylation and hydrolysis of these precursors in the endoplasmic reticulum result in the formation of triacylglycerol. Abbreviations include the following: CoA, coenzyme A; DAG, diacylglycerol; DGDG, digalactosyldiacylglycerol; G-3-P, glycerol-3-phosphate; LPA, lysophosphatidic acid; PA, phosphatidic acid; PC, phosphatidylcholine; and MGDG, monogalactosyldiacylglycerol.

tive stability [P2•]. A reduction in seed triacylglycerol linolenic acid content from about 7–8% (FAC) in the parent line (Century) to about 3–4% (FAC) was observed in the soybean mutant *C1640* when the plants were grown at 28°C during the day and 22°C during the night [9,10].

An allelic mutation has previously been described in the low linolenic/high oleic soybean line *A5* [8]. The locus at which this mutation occurred has been designated the *FAN* locus [9,11,12] and may be equivalent to the *FAD 3* locus of *Arabidopsis* [13]. It should be observed, however, that the soybean *fan* mutation is expressed only in roots and seeds [14], whereas the *Arabidopsis fad 3* mutation is constitutive. A second major gene controlling linolenic acid content of soybean seeds has recently been identified by a mutation at the *FAN 2* locus, in line *A23* [15•]. Field-grown *A23* plants have a linolenic acid content of about 5.5% (FAC), compared with 7.5% (FAC) in Century and about 4% (FAC) in *C1640* plants grown in similar plots. A cross between *C1640* and *A23*, however, yielded segregants with a seed linolenic acid content of less than 2.5% (FAC). The oil produced from two lines from this cross (i.e. *A16* and *A17*) was shown to be more stable to oxidation than oil from commmercial soybean lines, and it also had improved organoleptic properties [1••]. The soybean *fan* mutant phenotype in *C1640* is temperature sensitive [10,16•]. When these plants were grown either at 15°C during the day and 12°C during the night, or at 40°C during the day and 30°C during the night, they had a seed triacylglycerol linolenic acid content of 8% (FAC) or ~2% (FAC), respectively. It seems likely that a second desaturase gene locus is present in soybean, expression of which is induced by low temperatures. The effect of temperature on linolenic content is independent of the *fan* mutation [16•]. It is, thus, either the result of

a temperature-regulated plastid omega-3 desaturase or the result of a second temperature-regulated locus of a microsomal omega-3 desaturase, possibly corresponding to *FAN 2*.

The cDNA encoding the *Arabidopsis* microsomal omega-3 desaturase (*FAD 3*) has been cloned by chromosome walking [17••] and also by T-DNA tagging [18••]. The corresponding cDNAs from canola and soybean have been cloned using the *Arabidopsis* gene [18••]. Seed-specific antisense expression of the soybean cDNA in an Elite line (*A2872*) results in a reduction of seed triacylglycerol linolenic acid content to about 1.5% (FAC) in transgenic plants, compared with 8–10% (FAC) in the control *A2872* plants (AJ Kinney, unpublished data). The reduction is seen only in seed fatty acids and is independent of the plant growth temperature. The remaining 1.5% (FAC) linolenic acid is presumably the result of the plastid omega-3 desaturase, expression of which is not affected by antisense cDNA. It has become apparent from these transgenic studies, therefore, that the temperature-induced variation in linolenic acid content observed in *C1640* is probably the result of the activity of a second locus of a microsomal omega-3 desaturase. This second locus is effectively inhibited by the antisense cDNA construct in the seeds of transgenic soybeans. The cDNAs encoding plastid omega-3 desaturases have also been cloned from both canola and soybean [18••]. Thus, the possibility now exists of reducing to zero the seed triacylglycerol linolenic acid content of these oilseed crops.

Increased oleic acid content

It is likely that most of the polyunsaturated fatty acids in the seed triacylglycerols of oilseeds are the result of the initial desaturation of oleoyl-phosphatidylcholine by a microsomal omega-6 desaturase [19••]. In *Arabidopsis*, a mutation at the *FAD 2* locus results in a reduction in both linoleic and linolenic acids, as well as an increase in oleic acid. The *FAD 2* locus has recently been shown to be a structural gene for a microsomal omega-6 desaturase [19••,20••]. Mutants in 18:1 desaturation in the seed, and thus, with an increased seed oleic acid content, have been found in sunflower [21], safflower [P3], maize [P4•], soybean [8] and canola [22•,P5,P6•]. The seed triacylglycerol oleic acid content of these crops ranges, in commercial varieties, from around 18–22% (FAC) in sunflower and soybean to around 75% (FAC) in canola lines, such as Westar.

High 18:1 mutants of sunflower [21], in which oleic acid constitutes 80–90% (FAC), are now well established, and the high oleic soybean mutant *A5*, in which oleic acid constitutes ~40% (FAC), was discovered a number of years ago [8]. More recently, high oleic mutants of maize (>65% (FAC) oleic acid), of safflower (>80% (FAC) oleic acid), and of canola (>80% (FAC) oleic acid) have been reported [22•,P3,P4•,P5,P6•]. The oleic acid content of the high oleic acid maize and sunflower mutants appears to be temperature stable; however, the seed triacylglycerol oleic acid content of the high oleic soybean mutant *A5* does vary with temperature, and may be as low as 13% (FAC) in plants grown at 15°C during the day and 12°C during the night [10].

Canola mutants with an oleic acid content as high as 87% (FAC) have been described [22•] and, although this may be an environment-insensitive phenotype, analysis of a similar high oleic acid mutant of canola (WD Hitz, unpublished data) has revealed that both leaf and root oleic acid contents are also elevated, resulting in severe growth problems when the plants are grown at lower temperatures (20°C during the day and 10°C during the night). The abnormal growth resulting from the combined effect of low polyunsaturation of root and leaf lipids with low growth temperature has also been observed when *Arabidopsis fad 2* mutants were transferred from 20°C to 6°C [23••]. The gradual development of symptoms at the lower temperature has led to the suggestion that a number of diverse, subtle and largely unknown factors may be the cause of retarded growth, rather than a general deterioration in membrane integrity resulting from fluidity changes. Nevertheless, the need to restrict large changes in fatty acid unsaturation to the seeds of oilseed plants is clearly illustrated by these observations.

An *Arabidopsis* cDNA corresponding to the gene at the *FAD 2* locus has now been cloned from a T-DNA tagged mutant allelic to the chemically mutated *fad 2* [20••]. Equivalent cDNAs have also been cloned from canola and soybean [24••]. Transgenic canola and soybean containing antisense omega-6 desaturase cDNAs expressed only in the seed have now been made (WD Hitz, AJ Kinney, unpublished data) and will be a promising source of temperature insensitive high oleic canola and soybean seeds. In canola, for example, antisense desaturase plants produce seeds with an oleic acid content of >85% (FAC) (WD Hitz, unpublished data).

Increased and decreased saturated fatty acid content

Edible oils with increased amounts of palmitic acid or stearic acid would reduce the need for chemical hydrogenation in the production of vegetable oils products, such as margarines and confectionary fats [P7]. A number of mutants in soybean [25,26] and canola [P6•] with altered amounts of saturated fats are available. In soybean, for example, the mutant line *C1727* contains 17% (FAC) palmitic acid [26], and the the *A6* line contains over 30% (FAC) stearic acid [25], although the phenotype of the latter is also environmentally sensitive [27]. Comparable commercial varieties have about 12% (FAC) palmitic acid and 3% (FAC) stearic acid.

The quest for healthier food has also led to a search for vegetable oils with decreased saturate content and, accordingly, reduced saturate mutants have been reported in soybean [26], sunflower [P8•] and canola

[P6•]. The low palmitate mutant of soybean (*C1726*) has about 8.5% (FAC) palmitic acid [26], resulting in a little less than 12% of the total saturates. Generic canola, on the other hand, has only around 6% (FAC) saturates, and some recently described mutant lines have as little as 4% (FAC) total saturates [P6•]. Thus, as a healthy oil, canola has something of a head start over most other vegetable oils. It is unclear, in most of these altered saturate mutants, whether the mutations are in structural genes, and if they are, which structural genes are involved. Part of the problem is that the biochemical pathway leading to the synthesis of palmitic acid and stearic acid, and their subsequentc release into the cytoplasm, involves many enzymes and is not linear [28].

Palmitic, stearic and oleic acids are synthesized in the plastids of oilseeds while esterified to acyl carrier protein (ACP) (see Fig. 1). Palmitoyl-ACP (16:O) is made from the condensation of myristoyl-ACP (14:0) and carbon dioxide by four consecutive reactions. The specificity of this conversion is determined by the enzyme catalyzing the first rate-limiting reaction, the conversion of myristoyl-ACP to β-ketopalmitoyl-ACP, a reaction catalyzed by β-ketoacylsynthetase I (KAS I). KAS I also catalyzes the three condensation reactions involved in the conversion of octanoyl-ACP (8:0) to myristoyl-ACP. The conversion of palmitoyl-ACP to stearoyl-ACP is catalyzed by a simliar set of reactions with the condensation step catalyzed by a unique condensing enzyme, β-ketoacylsynthetase II (KAS II). Stearoyl-ACP (18:0) is desaturated to oleoyl-ACP (18:1) by a soluble, delta-9 desaturase. Acyl-ACP thioesterase releases fatty acids from ACP so they may be exported from the plastid to the cytoplasm, where, as acyl-CoA, they are further desaturated or assembled into triacylglycerol. Thus, the thioesterase competes with both KAS II and delta-9 desaturase for substrate. The ratio of fatty acids in the cytoplasm, and hence in triacylglycerol, therefore, is the result of a balance of the activities of a number of plastid enzymes.

The cloning of cDNAs for all of the above enzymes has been reported, and their relative abundance manipulated in transgenic oilseed plants [29••,P9,P10•,P11]. Antisense RNA suppression of the delta-9 desaturase in canola [29••] resulted in oil with up to 40% (FAC) stearic acid. An increase in total saturates to over 20% (FAC) has also been observed in canola oils by seed-specific overexpression of the acyl-ACP thioesterase cDNA ([P10•]; WD Hitz, unpublished data). Since the thioesterase is competing for substrate with KAS II and delta-9 desaturase, more thioesterase protein results in better competition and, thus, greater release of palmitate and stearate from ACP rather than enlongation or, in the case of stearate, desaturation. The increase in stearate was greater than that of palmitate such that the relative abundance of palmitate and stearate in the transgenic canola was about the same (around 10% (FAC) for each). This is probably because, in wild-type plants, thioesterase competes for substrate much more effectively with KAS II than with delta-9 desaturase and, thus, normally, less 18:0-ACP escapes from plastid. A greater effect on stearate (18:0) is seen, therefore, when this competition for both substrates is biased in favor of the thioesterase by increasing the relative activity of the thioesterase in the transgenic. It is, thus, theoretically possible to increase total saturates even more by combining the delta-9 desaturase suppression phenotype with the thioesterase overexpression phenotype, either by cotransformation or by crossing the two different transgenic lines.

Some biological limitations may, however, be placed on increasing the relative abundance of total saturates in plant oils. For example, the amount of saturated acyl-CoAs attached to triacylglycerols may be limited by the substrate preferences of the acyltransferases. Also, it is not known if solid fats can be catabolized as efficiently as liquid oils by germinating seeds, or if they can be catabolized at all. Since highly saturated oils are solid at the temperatures at which most plants are grown, a seed may find it impossible to germinate if it cannot mobilize its energy reserves. This may even be true of tropical high-saturate plant oils. Thus, although the total 16:0+18:0 saturate content of some palm oils can be greater than 50% (FAC), this is due entirely to the contribution of mesocarp oil to the total fatty acid content [3]. Most of the saturated fatty acids in the palm kernal are of shorter chain length (12:0 and 14:0). The longer chain saturates comprise, on average, around 12% (FAC) [and not more than 20% (FAC)] [3].

Transgenic lines with reduced saturates have been somewhat more elusive to obtain. A recent report has demonstrated a reduction [from 6% (FAC) to about 3.5% (FAC)] in the total saturate content of *Arabidopsis* seeds when KAS I and KAS II subunits, under the control of the napin promoter, are both overexpressed in the same plant (JL Bleibaum, A Genez, J Fayet-Faber, DW McCarter, GA Thompson, National Plant Lipid Cooperative, Plant Lipid Symposium, Minneapolis, Minnesota, July 29–31, 1993, Poster A18). A reduction in the total saturate content of total lipid has been observed when either the yeast [30••] or rat [31••] stearoyl-CoA desaturase was expressed, under the control of the cauliflower mosaic virus 35S promoter, in the cytoplasm of tobacco cells. In yeast and mammals, the stearoyl-CoA desaturase uses both palmitoyl-CoA and stearoyl-CoA as a substrate. Thus, in the transgenic tobacco, these plant cytoplasmic CoAs were desaturated to palmitoleoyl-CoA and oleoyl-CoA before being incorporated into complex lipids. In both cases, however, very little change in seed triacylglycerol fatty acid composition was seen. Probably as a consequence of using the 35S promoter, the greatest increases in 16:1 were observed in stems and roots, and to a lesser extent in the leaves. The decrease in total saturates in the seeds amounted to less than 2% (FAC) [i.e. from 11.4% (FAC) to 10.1% (FAC)] [30••]. These results do demonstrate, however, the potential for lowering triacylglycerol saturates in oilseed crops by expressing the yeast or rat desaturase gene under the control of seed-specific promoters, although it is unclear how the public will react to fields of transgenic plants expressing rat genes.

Conclusions

Although many of the modified-oil phenotypes produced through mutagenesis are similar to those produced by genetic engineering, most mutant phenotypes are either expressed only within a particular range of plant growth temperatures or the mutant plants themselves are environmentally sensitive. This is because plants have evolved to respond to adverse growth temperatures by changing their membrane fatty acid composition and, as a consequence of the altered acyl-CoA pool, these changes are also reflected in the composition of storage lipids.

In the case of polyunsaturated fatty acid biosynthesis, crops such as soybean and canola most likely have at least two genes for each of the major membrane-associated desaturases. One of these genes is probably expressed predominantly in the seed, and mutation at this locus (e.g. *fan* in soybean) would, thus, affect only seed lipids. The second locus is expressed constitutively and, in some cases, is temperature sensitive. Consequently, at low temperature, expression of this second locus is able to mask the effect of the seed-specific mutation. A combination, however, of a mutation at the second constitutuve locus with a mutation at the seed-predominant locus (e.g. the very high oleic canola) would effectively block polyunsaturate synthesis in both root and leaf. This would explain the poor growth of the double-mutant plant at low temperatures [23••]. One way of avoiding this would be to discover a mutation in a gene that regulates desaturase expression specifically in the seed, if such a regulator exists. Perhaps, however, a transgenic approach, either using seed-specific promoters to antisense desaturase genes or to *trans*-inactivate desaturase genes [32••], may be the best way to overcome these difficulties. In addition, transformation with a cloned cDNA or gene is probably the only efficient way of causing an increase in the expression of a specific gene.

Finally, although mutated plants and transgenic plants are often regarded as conceptually different entities, a transgenic phenotype is really nothing more than a highly specific 'mutation'. It is likely that the combination of these two approaches will enable further germplasm variation. Many of the mutants and the transgenics described above would be expected to complement each other. In the future, making crosses with mutants and transgenics may be the most efficient way of combining or enhancing oil traits.

Acknowledgement

The author thanks his colleague Bill Hitz for helpful comments on the manuscript and for providing unpublished data.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• NEFF WE, SELKE E, MOUNTS TL, RINSCH W, FRANKEL EN, ZIETOUN MAM: Effect of Triacylglycerol Composition and Structures on Oxidative Stability of Oils from Selected Soybean Germplasm. *JAOCS* 1992, 69:111–118.
This study considers how the position of fatty acids on the triacylglycerol molecule affects the stability of soybean oil.

2. •• LIU HR, WHITE PJ: Oxidative Stability of Soybean Oils with Altered Fatty Acid Compositions. *JAOCS* 1992, 69:528–532.
A study of the stability of oils with different fatty acid compositions obtained from the germplasm of a variety of soybeans. In the past, most studies were carried out using oils reconstituted from fractionated triacylglycerol species. This paper is also a good source of references to earlier work in the field.

3. GUNSTONE FD, HARWOOD JL, PADLEY FB: *The Lipid Handbook*. London: Chapman and Hall; 1986.

4. • KISHORE GM, SOMERVILLE CR: Genetic Engineering of Commercially Useful Pathways in Transgenic Plants. *Curr Opin Biotechnol* 1993, 3:176-180.
A short, but comprehensive and very readable review that covers many aspects of the genetic engineering of commercially important products, including industrial oils, from plants.

5. •• KNAUF VC: Progress in the Cloning of Genes for Plant Storage Lipid Biosynthesis. In *Genetic Engineering*, vol 15. Edited by Setlow JK. New York: Plenum Press; 1993:149–164.
A good review of recent progress in the genetic engineering of oilseeds. The manipulation of traits with relevance to industrial applicatons are particularly emphasized.

6. •• HEINZ E: Biosynthesis of Polyunsaturated Fatty Acids. In *Lipid Metabolism in Plants*. Edited by Moore TS Jr. Boca Raton: CRC Press; 1993:36–90.
The most recent and comprehensive review of polyunsaturated fatty acid biosynthesis in plants. This volume also contains many other interesting reviews on plant lipid biochemistry, including the biosynthesis of unusual fatty acids.

7. • STYMNE S, TONNET ML, GREEN AG: Biosynthesis of Linolenate in Developing Embryos and Cell Free Preparations of High Linolenate Linseed and Low Linolenate Mutants. *Arch Biochem Biophy* 1992, 294:557–563.
An interesting biochemical study of a flax mutant with reduced linolenic acid content. The authors conclude that the phenotype is a result of reduced omega-3 desaturase enzyme activity.

8. GRAEF GL, FEHR WR, MILLER LA, HAMMOND EG, CIANZIO SR: Inheritance of Fatty Acid Composition in a Soybean Mutant with Low Linolenic Acid. *Crop Sci* 1988, 28:55–58.

9. WILCOX JR, CAVINS JF: Gene Symbol Assigned for Linolenic Acid Mutant in the Soybean *J Hered* 1987, 78:410.

10. RENNIE BD, TANNER JW: Fatty Acid Composition of Oil from Soybean Seeds Grown at Extreme Temperatures *JAOCS* 1989, 66:1622–1624.

11. NICKELL AD, WILCOX JR, CAVINS JF: Genetic Relationship between Loci Controlling Palmtic and Linolenic Acids in Soybean. *Crop Sci* 1991, 31:1169–1171.

12. RENNIE BD, TANNER JW: New Allele at the Fan Locus in the Soybean Line A5. *Crop Sci* 1991, 31:297–301.

13. OHLROGGE JB, BROWSE J, SOMERVILLE CR: The Genetics of Plant Lipids. *Biochim Biophys Acta* 1991, 1082:1–26.

14. WANG XM, NORMAN HA, ST JOHN JB, YIN T, HILDEBRAND DF: Comparison of Fatty Acid Composition in Tissues of Low Linolenate Soybean. *Phytochemistry* 1989, 28:411–414.

15. FEHR WR, WELKE GA, HAMMOND EG, DUVICK DN, CIANZIO SR: Inheritance of Reduced Linolenic Acid Content in Soybean Genotypes A16 and A17. *Crop Sci* 1992, 32:903–906.

16. • WILCOX JR, CAVINS JF: Normal and Low Linolenic Acid Soybean Strains: Response to Planting Data. *Crop Sci* 1992, 32:1248–1251.
A study showing that the phenotype of low linolenic acid soybean mutants is affected by the growth temperature of the plants.

17. •• ARONDEL V, LEMIEUX B, HWANG I, GIBSON S, GOODMAN H, SOMERVILLE CR: Map-Based Cloning of a Gene Controlling Omega-3 Fatty Acid Desaturation in Arabidopsis. *Science* 1992, 258:1353–1355.
The first example of the use map-based cloning in plants, and one of the first membrane-associated fatty acid desaturases to be cloned from higher plants.

18. •• YADAV N, WIERZBICKI A, AEGERTER, M, CASTER CS, PEREZ-GRAU L, KINNEY AJ, HITZ WD, BOOTH RB, SCHWEIGER B, STECCA KL, *ET AL.* : Cloning of Higher Plant Omega-3 Fatty Acid Desaturases. *Plant Physiol* 1993, 103:467–476.
A landmark report detailing the cloning of cDNAs for membrane-associated fatty acid desaturases from economically important oilseed crops. One of the first examples of structural gene cloning by T-DNA insertional mutagenesis.

19. •• MIQUEL M, BROWSE J: Arabidopsis Mutants Deficient in Polyunsaturated Fatty Acid Synthesis *J Biol Chem* 1992, 267:1502–1509.
Reports the biochemical characterization of a mutant deficient in a microsomal omega-6 desaturase. The deficiency is confirmed by direct enzyme assays of the omega-6 desaturase in wild-type and mutant *Arabidopsis*. The authors demonstrate that this desaturase is responsibe for most of the polyunsaturated fatty acid biosynthesis in higher plants.

20. •• LIGHTNER J, OKULEY J, FELDMAN K, YADAV N, BROWSE J: The Arabidopsis Fad 2 Gene Encodes the Enzyme that is Essential for Polyunsaturated Lipid Synthesis. *Plant Cell* 1994, in press.
Reports the cloning of the cDNA and gene corresponding to the *FAD 2* locus in *Arabidopsis*. The gene encodes a desaturase enzyme with a predicted amino acid sequence related to other plant desaturase sequences. Mapping of the gene to the *FAD 2* locus confirms that the *fad 2* mutation is in the structural omega-6 desaturase gene.

21. PURDY RH: High Oleic Sunflower: Physical and Chemical Characteristics. *JAOCS* 1986, 63:773–775.

22. • AULD DL, HEIKKINEN MK, ERICKSON DA, SERNYK JL, ROMERO JE: Rapeseed Mutants with Reduced Levels of Polyunsaturated Fatty Acids and Increased Levels of Oleic Acid *Crop Sci* 1992, 32:657–662.
Describes rapeseed mutants with a relative oleic acid content of greater than 80% (FAC).

23. •• MIQUEL M, JAMES D, DOONER H, BROWSE J: Arabidopsis Requires Polyunsaturated Lipids for Low-Temperature Survival. *Proc Natl Acad Sci USA* 1993, 90:6208–6212.
A fascinating paper clearly showing that, in higher plants, polyunsaturated fatty acids are required for growth at low temperatures. The report also questions the commonly postulated notion that the primary cause of low temperature damage is a result of the disruption of membrane integrity.

24. •• HITZ WD, CARLSON TJ, BOOTH R, KINNEY AJ, STECCA KL, YADAV NS: Cloning of a Higher Plant Plastid Omega-6 Desaturase cDNA and its Expression in a Cyanobacterium. *Plant Physiol* 1994, in press.
In addition to reporting the cloning of a plastid omega-6 desaturase cDNA, this report shows the evolutionary relationships among a number of omega-3 and omega-6 membrane-associated desaturase proteins from soybean and canola.

25. BUHECK DM, FEHR WR, HAMMOND EG: Inheritance of Palmitic and Stearic Mutants of Soybeans. *Crop Sci* 1989, 29:652–656.

26. ERICKSON EA, WILCOX JR, CAVINS JF: Inheritance of Altered Palmitic Acid Percentage in two Soybean Mutants *J Hered* 1988, 79:465–468.

27. • SCHNEBLY SR, FEHR WR: Effect of Years and Planting Dates on Fatty Acid Composition of Soybean Genotypes. *Crop Sci* 1993, 33:716–719.
A study showing that the phenotype of a number of soybean mutants, including high oleic mutants, is affected by the growth temperature of the plants.

28. BROWSE J, SOMERVILLE CJ: Glycerolipid Synthesis: Biochemistry and Regulation *Annu Rev Plant Physiol Mol Biol* 1991, 42:467–506.

29. •• KNUTZON DS, THOMPSON GA, RADKE SE, JOHNSON WB, KNAUF VC, KRIDL JC: Modification of Brassica Seed Oil by Antisense Expression of a Stearoyl-ACP Desaturase Gene. *Proc Natl Acad Sci USA* 1992, 89:2624–2628.
Transgenic canola plants are described that display a large increase in stearic acid in their seed oil. The modification is achieved by transforming canola plants with an antisense delta-9 desaturase construct.

30. •• POLZSHOCK JJ, CHIN C-K, MARTIN CE: Expression of the Yeast Delta-9 Fatty Acid Desaturase in Nicotiana Tabacum. *Plant Physiol* 1992, 100:894–901.
See [31••].

31. •• GRAYBURN WS, COLLINS GB, HILDEBRAND DF: Fatty Acid Alteration by a Delta-9 Desaturase in Transgenic Tobacco Tissue. *Biotechnology* 1992, 10:675–678.
This report and [30••] demonstrate the potential for expressing genes from non-plant species in higher plants in order to modify fatty acid content.

32. •• KOOTER JM, MOL JNM: Trans-Inactivation of Gene-Expression in Plants. *Curr Opin Biotechnol* 1993, 4:166–171.
A very good summary of our current knowledge concerning inactivation of plant gene expression via sense-suppression with homologous transgenes.

Patents

- • of special interest
- •• of outstanding interest

P1. UNILEVER NV: Edible Fats Containing Specified Triglycerides Useful for Preparing Margarines with Good Spreadability, Plasticity and Melting Behaviour. 13/6/84 EP129293.

P2. • DU PONT DE NEMOURS & CO: Brassica napus Plant Produces Seed Having Reduced Linolenic Acid and Glucosinolate and Yields Oil with Low Sulfur, Improved Sensory Characteristics and Increased Oxidative Stability 30/9/92 WO9306714.
Describes chemically induced mutations in canola plants that reduce the linolenic acid content of the seed oil. The resulting oil has improved flavor, cooking and storage properties.

P3. RESEARCH & DEVELOPMENT INSTITUTE INC: Safflower Products with High Oleic or Linoleic Content for Food Industry or for Use in Resin or Paint. 22/8/91 WO9111906.

P4. • LUBRIZOL CORP: Subgroup of Corn Varieties Produces Corn Seed Having Oleic acid Value of at Least 65% by Weight Total Fatty Acid of Seed. 6/2/92 WO9201367.
Describes corn plants produced by mutagenesis that have a seed oleic acid content of 65% (FAC) compared with about 30% (FAC) in generic corn.

P5. ALLELIX INC: Production of Improved Rapeseed Exhibiting an Enhanced Oleic Acid Content. 12/7/89 EP323753.

P6. • DU PONT DE NEMOURS & CO: New Brassica napus Seeds, Plants and Oils for Food Products Have High Oleic Acid, Low Linoleic Acid, High or Low Palmitic Acid, Low Stearic Acid and Low Linolenic Acid. 19/3/92 WO9203919.

Describes a number of canola mutants with altered oil fatty acid content, including *IMC129*, which has an oleic acid content of greater than 80% (FAC).

P7. UNILEVER NV: Edible Plasticised Emulsion Spreads Have Fat Phase Comprising Non-Chemically Modified Vegetable Fat. 23/5/90 EP369519.

P8. PIONEER HI-BRED INTERNATIONAL INC: Sunflower Oil with Low Levels of Saturated Fatty Acids Produced by Crushing Sunflower Having Low Levels of Saturated Fatty Acids. 29/7/92 EP496504.

A sunflower mutant with reduced palmitic and stearic acid content.

P9. DU PONT DE NEMOURS INC: DNA Encoding Soybean Stearoyl-ACP Desaturase Enzyme and Precursor and Chimeric Genes, for Plant Transformation and Control of Levels of Saturated and Unsaturated Fatty Acids in Edible Oils. 12/12/91 WO9118985.

P10. DU PONT DE NEMOURS INC: Modifying Plant Oil Composition by using Nucleic Acid Fragments Encoding Soybean Seed Acyl-ACP Thioesterase, Providing Health Benefits to Humans in Reducing Risk of Coronary Heart Disease. 9/7/92 WO9211373.

Describes the cloning of an oleoyl-ACP thioesterase from soybean and other plants. The thioesterase can be used to increase (via sense expression) or decrease (via antisense expression) the saturate content of plant oils.

P11. CALGENE INC: New Plant Beta-Ketoacyl Synthase Proteins Obtained from Ricinus communis Useful for Modifying Fatty Acid Composition. 5/3/92 WO9203564.

AJ Kinney, Du Pont Experimental Station, PO Box 80402, Wilmington, Delaware 19880-0402, USA.

The Plant Cell, Vol. 6, 147–158, January 1994

Arabidopsis *FAD2* Gene Encodes the Enzyme That Is Essential for Polyunsaturated Lipid Synthesis

John Okuley,[a] Jonathan Lightner,[a] Kenneth Feldmann,[b] Narendra Yadav,[b] Ellen Lark,[a] and John Browse[a,1]

[a] Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164-6340
[b] Agricultural Products and Central Research and Development, E.I. Du Pont de Nemours & Co., Experimental Station, P.O. Box 80402, Wilmington, Delaware 19880-0402

The polyunsaturated fatty acids linoleate and α-linolenate are important membrane components and are the essential fatty acids of human nutrition. The major enzyme responsible for the synthesis of these compounds is the plant oleate desaturase of the endoplasmic reticulum, and its activity is controlled in Arabidopsis by the *fatty acid desaturation 2* (*fad2*) locus. A *fad2* allele was identified in a population of Arabidopsis in which mutations had been created by T-DNA insertions. Genomic DNA flanking the T-DNA was cloned by plasmid rescue and used to isolate cDNA and genomic clones of *FAD2*. A cDNA containing the entire *FAD2* coding sequence was expressed in *fad2* mutant plants and shown to complement the mutant fatty acid phenotype. The deduced amino acid sequence from the cDNA showed homology to other plant desaturases, and this confirmed that *FAD2* is the structural gene for the desaturase. Gel blot analyses of *FAD2* mRNA levels showed that the gene is expressed throughout the plant and suggest that transcript levels are in excess of the amount needed to account for oleate desaturation. Sequence analysis identified histidine-rich motifs that could contribute to an iron binding site in the cytoplasmic domain of the protein. Such a position would facilitate interaction between the desaturase and cytochrome b_5, which is the direct source of electrons for the desaturation reaction, but would limit interaction of the active site with the fatty acyl substrate.

INTRODUCTION

The polyunsaturated fatty acids linoleate (Δ9, 12–18:2) and α-linolenate (Δ9, 12, 15–18:3) are synthesized by plants but not by most other higher eukaryotes. Both of these fatty acids are essential components of human nutrition, because in mammals they act as precursors not only of membrane lipids but also of families of signaling molecules including the prostaglandins, thromboxanes, and leukotrienes (Smith and Borgeat, 1985). In many higher plants, 18:2 and 18:3 account for more than 70% of the fatty acids in leaf cells and 55 to 70% of the fatty acids in nonphotosynthetic tissues such as roots (Harwood, 1980). In both plants and animals, polyunsaturated acyl structures are considered to be essential membrane components, in part because they are virtually ubiquitous in the membranes of higher eukaryotes. Experimental manipulations designed to reduce membrane polyunsaturation also point to a critical role for these structures in ensuring proper membrane function and organism viability (Hugly and Somerville, 1992; Miquel et al., 1993).

In angiosperms as a whole, the vast majority of polyunsaturated lipid synthesis passes through a single enzyme, the 18:1 desaturase of the endoplasmic reticulum. Although a chloroplast 18:1 desaturase also operates in leaf cells of at least some plants, it is likely that the endoplasmic reticulum enzyme is quantitatively more important even in these cells (Browse and Somerville, 1991; Miquel and Browse, 1992). Furthermore, it is responsible for more than 90% of the polyunsaturated fatty acid synthesis in nonphotosynthetic tissues, such as roots, and in the developing seeds of oil crops, including soybean, sunflower, and canola, in which fatty acids are stored as triacylglycerol oils. Thus, one important function of the endoplasmic reticulum 18:1 desaturase is to provide 18:2 and (following further desaturation) 18:3 required for the correct assembly of cellular membranes throughout the plant. Just as importantly, the enzyme provides the polyunsaturated fatty acids found in vegetable oils that in turn are the major source of essential fatty acids in most human diets.

Mutants of Arabidopsis at the *fatty acid desaturation 2* (*fad2*) locus are deficient in activity of the endoplasmic reticulum desaturase (Miquel and Browse, 1992). Biochemical and genetic studies of these mutants have been important to our understanding of this desaturation step because the enzyme, like most of the plant desaturases, is an integral membrane protein that has been difficult to solubilize and, therefore, to investigate by traditional enzymological methods. In the absence of a purified enzyme, genetic techniques can also

[1] To whom correspondence should be addressed.

provide an alternative means to clone the relevant genetic locus. For example, cDNAs corresponding to the *FAD3* gene have been obtained by both map-based cloning (Arondel et al., 1992) and by gene tagging (Yadav et al., 1993). The homology of these cDNAs to other desaturase genes indicates that they correspond to the structural gene encoding the endoplasmic reticulum 18:2 desaturase. The *FAD3* cDNAs have made possible the cloning of at least three distinct, additional desaturase genes. Each of these contained a 5′ peptide sequence with the characteristics of a chloroplast transit peptide and, on this basis, the genes are thought to encode three of the chloroplast desaturases (Iba et al., 1993; Yadav et al., 1993; W. Hitz, J. Okuley, N. Yadav, and J. Browse, unpublished data). Extensive efforts using the four available gene sequences have, to date, failed to isolate a gene encoding the endoplasmic reticulum 18:1 desaturase.

In this study, we describe the identification of a *fad2* allele in a population of Arabidopsis in which mutations have been generated by T-DNA insertion (Feldmann, 1991) and the subsequent cloning and characterization of the wild-type *FAD2* gene. The expansion of the family of genes encoding plant membrane desaturases now permits some discussion of a possible structure for the 18:1 desaturase.

RESULTS

Identification of a *fad2* Allele in a T-DNA Line

To initiate our screen, we grew the 1800 T-DNA insertional lines (T_3 generation) that were then available and directly measured the overall fatty acid composition of leaf tissues by gas chromatography (Browse et al., 1986b; Feldmann et al., 1989). Because the T-DNA–containing lines segregated for the insert (and thus for any resultant mutation), it was necessary for us to sample several individuals from each line. To simplify the procedure, we harvested single leaves from 10 plants of one line and pooled these for analysis. Based on reconstruction experiments using the chemically induced Arabidopsis lipid mutants, we expected to be able to successfully identify any one of seven mutants in a segregating population.

Among the pooled leaf samples analyzed in this way, one sample from line 658 was identified as having an increase in 18:1. The increase was considerably smaller than would be expected from the data on chemically induced *fad2* mutants (Miquel and Browse, 1992)—18:1 in the mutant pool was only 5.5% compared with 2.8% in neighboring, wild-type pools. In principle, this result might have arisen from underrepresentation of homozygous mutants within the small number of leaves that had been pooled for analysis. However, as shown in Table 1, analysis of tissue samples from individual plants of the 658 line demonstrated that whereas the line was segregating for 18:1 content, the most extreme phenotypes (homozygous for the lipid mutation) contained considerably lower proportions of 18:1 than previously described for *fad2* mutants (James and Dooner, 1990; Lemieux et al., 1990). Reciprocal crosses between homozygous individuals from line 658 and the *fad2-1* mutant (Lemieux et al., 1990) produced F_1 progeny with fatty acid compositions intermediate between those of the two parents. The lack of complementation in these crosses demonstrated that the lipid mutation in line 658 is an allele of *fad2*, which we designated *fad2-5*.

When 200 T_3 seeds of line 658 were germinated on agar containing 50 μg/mL kanamycin, only four kanamycin-sensitive individuals were identified. This proportion of kanamycin-sensitive seedlings is a good approximation to a 63:1 ratio and indicated to us that the line contained three unlinked T-DNA inserts. In this and two other experiments, a total of 56 kanamycin-sensitive seedlings were identified. Fifty-three of these were analyzed for fatty acid composition and at least seven displayed 18:1 levels that were higher than those observed in wild-type seedlings grown under these conditions.

To further test whether the *fad2-5* mutation might be the result of T-DNA insertion, we isolated a derivative line that segregated for a single locus for both kanamycin resistance

Table 1. Fatty Acid Compositions of Root, Leaf, and Seed Tissues of the Wild-Type, *fad2-5*, and *fad2-1* Mutant Arabidopsis Plants[a]

Fatty Acid	Root			Leaf			Seed		
	WT	*fad2-5*	*fad2-1*	WT	*fad2-5*	*fad2-1*	WT	*fad2-5*	*fad2-1*
16:0	24.7	12.7	14.0	13.7	10.0	13.9	10.2	8.5	8.6
16:1	1.2	1.3	2.3	2.4	4.1	2.2	–[b]	–	–
16:3	–	–	–	16.0	15.4	18.5	–	–	–
18:0	3.2	2.0	1.7	0.4	0.5	0.5	2.5	4.0	4.3
18:1	6.8	27.9	55.9	2.3	15.0	20.9	15.4	37.7	53.5
18:2	29.8	20.0	6.4	14.5	4.6	3.8	32.7	8.1	3.2
18:3	29.1	20.4	12.8	50.8	45.0	39.6	20.3	11.3	5.5
20:1	–	–	–	–	–	–	16.6	26.0	23.9

[a] Data for wild type (WT) and *fad2-1* are from Lemieux et al. (1990). Root tissue was harvested from plants grown in liquid medium (Miquel and Browse, 1992). Leaves from 17-day-old plants were analyzed. Seeds were sampled from mature siliques. Data are mol %.
[b] Dashes indicate that the fatty acid was not detected.

and the mutant fatty acid phenotype. Approximately 100 individual T_3 plants were grown to maturity, and seeds were collected. One sample of seed from each T_3 plant was tested for the ability to germinate and grow in the presence of kanamycin. In addition, the fatty acid compositions of 10 additional individual seeds from each line were determined. A T_3 plant, 658-75, was identified whose progeny seeds segregated 28 kanamycin sensitive to 60 kanamycin resistant and seven seeds with low or intermediate 18:1 (15 to 22%) to two with high 18:1 (35 to 38%). Approximately 400 T_4 progeny of the derivative line 658-75 were grown, and their leaf fatty acid compositions were determined. From these, 91 plants were identified as being homozygous for the *fad2-5* mutation. The remaining plants could not be definitively assigned to wild-type and heterozygous classes on the basis of leaf fatty acid composition and, thus, could not be used to test linkage between the T-DNA markers and the fatty acid phenotype. Eighty-three of the 91 homozygous plants were tested for the presence of nopaline, which is a second, easily scored marker of the T-DNA (Errampalli et al., 1991). All 83 plants were positive for the presence of nopaline. Thus, this experiment indicated cosegregation of the *fad2* locus with the T-DNA insert present in the 658-75. Therefore, we set out to isolate plant DNA flanking the site of T-DNA insertion.

Cloning a cDNA That Spans the Site of T-DNA Insertion

The modified T-DNA used to generate the mutant population contains the origin of replication and the ampicillin resistance gene of plasmid pBR322 (Feldmann and Marks, 1987). This feature permits the recovery of T-DNA–plant DNA junction fragments as plasmids in *Escherichia coli* by the method of plasmid rescue (Behringer and Medford, 1992). Genomic DNA from homozygous mutant segregants of the 658-75 line was digested to completion with either the BamHI or SalI restriction enzyme and allowed to religate at a dilute concentration to promote self-ligation. The ligation products were used to transform *E. coli* cells that were then subjected to ampicillin selection. No ampicillin-resistant colony was obtained from the experiment with SalI-digested DNA, but a single ampicillin-resistant colony was identified from the plasmid rescue of BamHI-digested DNA and designated pTF-658. Restriction analysis of pTF-658 with BamHI, SalI, and EcoRI restriction enzymes indicated that in addition to the expected 14.2-kb portion of the T-DNA, it contained a 1.6-kb EcoRI-BamHI fragment of putative plant DNA that would lie adjacent to the left T-DNA border (which contains an EcoRI site). This EcoRI-BamHI fragment was subcloned into a pBluescript SK– vector to yield plasmid pSI658.

When the 1.6-kb fragment was radiolabeled and used to probe gel blots of genomic DNA from wild-type Arabidopsis at high stringency, strongly hybridizing bands were observed, which confirmed that the flanking sequence was plant DNA. For DNA from plants of the segregating 658-75 line, the same bands were present, but, in addition, a set of distinct bands was visible (data not presented). This finding indicated that the 1.6-kb fragment is indeed part of the locus that is interrupted by the T-DNA. The 1.6-kb EcoRI-BamHI fragment was used as a radiolabeled hybridization probe to screen an Arabidopsis λYES cDNA library (Elledge et al., 1991). Four of the positively hybridizing plaques identified were subjected to plaque purification. Plasmids containing the cDNA sequences were then generated by utilizing the cre-lox recombination feature of the λYES library (Elledge et al., 1991). The four cDNAs ranged in length from 1 to 1.4 kb. Restriction enzyme mapping and partial sequence determination indicated that they all represent the same gene. Nucleotide sequencing of two of the cDNAs (pF2a and pF2b) produced the 1372-bp sequence shown in Figure 1A that contains an open reading frame that encodes a predicted protein of 383 amino acids.

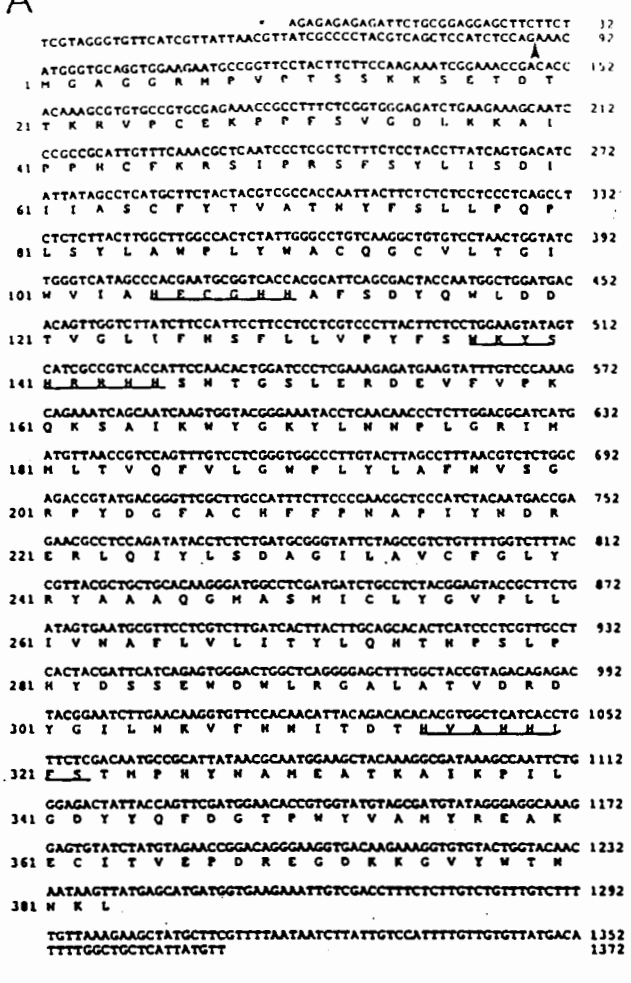
A genomic clone corresponding to the cDNA was isolated by screening an Arabidopsis library using the insert from pF2b as a radiolabeled probe. The 6-kb HindIII insert from a purified, positively hybridizing plaque was subcloned into a pBluescript SK– vector. Sequencing of 3 kb of the genomic clone spanning the sequence shown in Figure 1A revealed the presence of a single intron of 1134 bp at a position between nucleotides 88 and 89 of the cDNA. Partial sequencing of the 1.6-kb EcoRI-BamHI genomic border fragment insert in pSI658 from the EcoRI end showed that, for the first 61 nucleotides, it is colinear with the left T-DNA border (except for a deletion of nine contiguous nucleotides at position 42 in the border fragment) and colinear from nucleotides 57 to 104 with that of nucleotides 41 to 88 of the cDNA. This result suggests that the left end of the T-DNA disrupted the gene within a 4-bp (5′-TGTT-3′) region of homology between the T-DNA border and host DNA and shows that the T-DNA disrupted the gene in the transcribed region and 5′ to the intron in the untranslated sequence. The structure of the *fad2-5* locus together with the cloned flanking sequence and three of the cDNA inserts are shown diagrammatically in Figure 1B.

DNA from a genomic phage clone was also used to detect and map a restriction fragment length polymorphism in a gel blot containing HindIII-digested genomic DNA from inbred lines derived from a cross between *Arabidopsis thaliana* (ecotype Wassileskija) and *A. thaliana* marker line W100 (ecotype Landsberg *erecta* background) essentially as described previously by Reiter et al. (1992). A single genetic locus corresponding to this gene was positioned on the upper arm of chromosome 3 between the cosmid c3838 and λAT228 restriction fragment length polymorphism markers and just 4.5 centimorgans proximal to the *fad7* locus previously mapped in the same population (Yadav et al., 1993).

Complementation of the *fad2-1* Mutant

To establish definitively that the cloned gene represents the *fad2* locus, we transformed the cDNA into *fad2-1* mutant plants. The 1.4-kb EcoRI fragment containing the cDNA was isolated

A



B

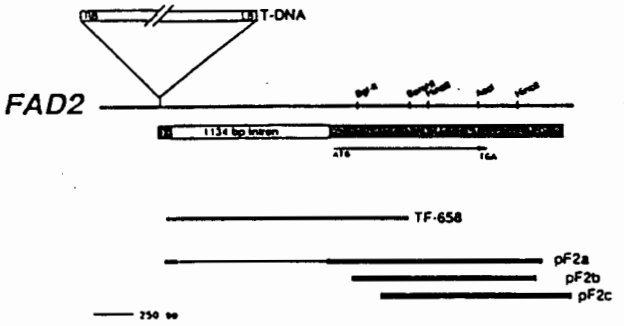


Figure 1. Structure and Organization of the Arabidopsis *FAD2* Gene.

(A) Nucleotide and derived amino acid sequences of a *FAD2* cDNA. The 4-bp sequence that is homologous to a sequence in the T-DNA border and into which T-DNA insertion appears to have occurred is shown in boldface letters. The arrowhead indicates the site of the intron found in the genomic sequence. The three histidine-rich sequences that show homology to other membrane-bound desaturases are underlined. The GenBank accession number of the sequence is L26296.

(B) The genome structure at the *fad2-5* locus showing the T-DNA insert, intron, open reading frame, and partial restriction map. The approximate lengths and locations of the 1.6-kb flanking sequence from the plasmid rescue (TF-658) and three cDNAs (pF2a, pF2b, and pF2c) are shown in the lower part of the figure. RB, right border; LB, left border.

from plasmid pF2a and subcloned into the EcoRI site of the binary vector pGA748. One of the resulting plasmids, designated pGA-Fad2, contained the cDNA in the sense orientation behind the cauliflower mosaic virus 35S promoter. The plasmid pGA-Fad2 was transformed into Agrobacterium strain R1000 carrying the Ri plasmid to generate transformants R1000/pGA-Fad2. These transformants and R1000 cells transformed with pGA748 alone were used to transform *fad2-1* mutant Arabidopsis using a stem-transformation protocol (Arondel et al., 1992). Wild-type Arabidopsis was also transformed with the empty vector control.

Hairy roots emerging from explants were maintained for 4 days on medium containing cefotaxime and carbenicillin for counterselection of the Agrobacterium and then transferred to the same medium containing kanamycin to select roots that had been cotransformed with the kanamycin resistance gene from the pGA748 vector. Six days after this transfer, the 12 kanamycin-resistant root cultures from the experiment employing the pGA-Fad2 construct were sampled together with three cultures derived from *fad2-1* and one culture derived from wild-type Arabidopsis control transformations. The fatty acid composition of each sample was determined. Figure 2 shows the desaturase products, 18:2 and 18:3, plotted as a percentage of the total fatty acids in each sample. It is clear that the fatty acid phenotype in roots from *fad2-1* (11 to 21% 18:2 + 18:3) was very substantially complemented in all the kanamycin-resistant transformants with the exception of one (number 5), which showed a proportion of 18:2 + 18:3 that was only a little higher than the mutant controls. In some transformants, the proportion of 18:2 + 18:3 was slightly higher than that in the wild-type control.

Limited Transcript Levels in *fad2-5* Allow Considerable Desaturation

The comparison of the fatty acid compositions of roots, shoots, and seeds of *fad2-5* mutant plants with similar tissues from the *fad2-1* mutant in Table 1 demonstrates the leaky nature of the mutation found in the T-DNA line. If the proportions 18:2 plus 18:3 (the two products of the 18:1 desaturase) are used as an approximate, relative measure of desaturase activity, then *fad2-5* is estimated to have 53 ([40.4−19.2]/[58.9−19.2]), 52, and 24% of wild-type activity in roots, leaves, and seeds, respectively. We examined the steady state levels of *FAD2* RNA in wild-type, *fad2-5*, and *fad2-1* tissues. As shown in Figure 3, the level of *FAD2* RNA in both shoots and roots of *fad2-1*

was indistinguishable from that detected in corresponding wild-type tissues, which suggests that the *fad2-1* allele produces a mutant protein without affecting the size or stability of the mRNA. In contrast, the *fad2-5* plants showed greatly reduced levels of an RNA of a similar length to the wild-type molecule. However, this transcript was barely detectable in shoot tissue and was only at a slightly higher level in roots. Densitometry of the bands shown in Figure 3 suggested that the transcript level in roots of *fad2-5* plants is no more than 12 to 15% of the wild type. This very low transcript level was nevertheless able to effect more than half of the 18:1 desaturation attributable to the *FAD2* gene product in wild-type Arabidopsis roots. Because the bands shown in Figure 3 were the only regions of the gel blot to hybridize to the radioactive cDNA probe, it is probable that they represent the total level of transcript available for synthesis of the desaturase.

FAD2 Transcript Level Is Not Increased in Response to Chilling

An active *FAD2* gene product is essential for the survival of Arabidopsis at low temperatures (Miquel et al., 1993). Furthermore, observations in plants and other organisms suggest that the level of membrane lipid unsaturation increases with decreasing temperature, possibly as a regulated response to chilling treatment (Somerville and Browse, 1991; Thompson, 1993). To determine whether transcriptional regulation of *FAD2* occurs in Arabidopsis in response to chilling treatment, we transferred wild-type Arabidopsis plants from 22 to 6°C. At the time of transfer and at different times after transfer, samples of leaf material were harvested, and the level of *FAD2* transcript within the total RNA was determined by gel blot analysis. The data in Figure 4 indicate that the steady state level of *FAD2* transcript, relative to both the estimated amount of total RNA and the level of 28S RNA transcript, remained approximately constant for at least 3 days after transfer to the cold. Thus, transcriptional regulation of the *FAD2* gene apparently plays no role in acclimation of Arabidopsis to low temperatures. However, our results leave open the possibility that 18:1 desaturase activity could be regulated at either the translational or enzyme levels.

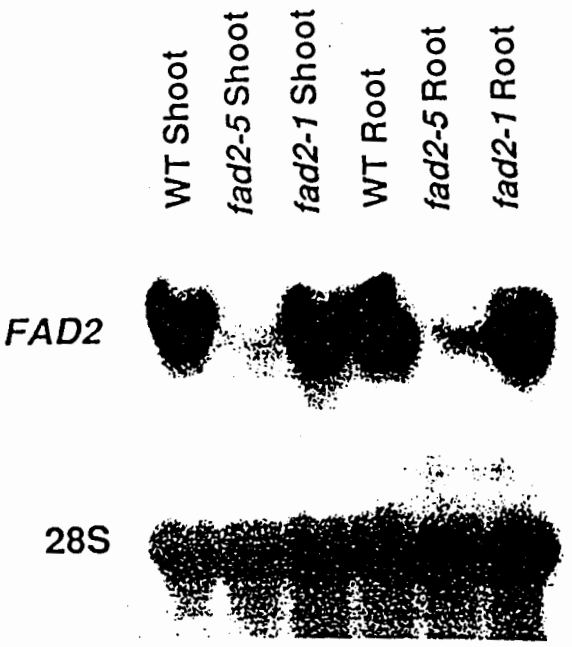


Figure 3. Transcript Levels in Wild-Type and *fad2* Mutant Arabidopsis.

Ten micrograms of total RNA per lane from shoots and roots of wild-type (WT), *fad2-5*, and *fad2-1* Arabidopsis plants was subjected to gel blot analysis using the pF2a insert as a probe (*FAD2*). Blots were stripped and reprobed with the Arabidopsis 28S rRNA gene (28S) as a loading control.

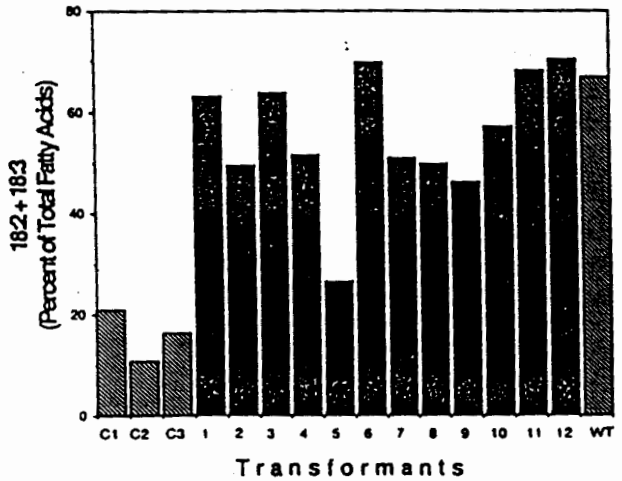


Figure 2. Polyunsaturated Fatty Acid Content of Transgenic *fad2-1* Plants.

The results shown are 18:2 + 18:3 (as percent of total fatty acids) for samples of hairy-root cultures from 12 independent transformants of the Arabidopsis *fad2-1* mutant with the pF2a cDNA insert. Control transformants (empty vector) of *fad2-1* (C1 to C3) and wild-type (WT) plants are included for comparison.

DISCUSSION

Cosegregation Analysis

The *fad2-5* allele that we isolated in the T-DNA insertion line 658 exhibited a leaky fatty acid phenotype (Table 1; Lemieux et al., 1990). More importantly, some of the seedlings that died on kanamycin-containing medium were homozygous or heterozygous for the *fad2-5* mutant phenotype. Both of these results argue against the *fad2-5* mutation being caused by

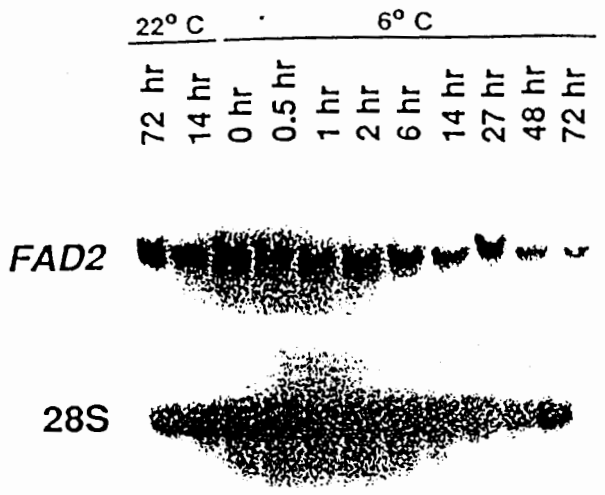


Figure 4. Transcript Levels in Wild-Type Plants after Transfer to 6°C.

Ten micrograms of total RNA per lane from plants maintained at 22°C or transferred to 6°C for up to 72 hr was subjected to gel blot analysis using the pF2a insert as a probe (*FAD2*). Blots were stripped and reprobed with the Arabidopsis 28S rRNA gene (28S) as a loading control.

T-DNA insertion. In particular, the identification of the mutant fatty acid phenotype in plants that died during kanamycin selection would normally be taken as strong evidence that the *fad2* gene was not tagged. However, our successful cloning of the *FAD2* locus demonstrated that these dead seedlings probably contained the T-DNA insert. It must be assumed that either the neomycin phosphotransferase II gene, which is responsible for kanamycin resistance in the T-DNA, was inactivated in these individuals (perhaps as a result of DNA methylation) or that the seedlings were inviable for unrelated reasons. It should also be noted that our second cosegregation analysis—the identification of nopaline in all 83 homozygous *fad2-5* individuals that were tested—provides only a very weak test of linkage. Even if the T-DNA and the *fad2* locus were as much as 5 centimorgans apart, there would still be a 40% probability of our obtaining the observed result. These results illustrate some of the difficulties involved in trying to decide whether to clone the T-DNA flanking regions in the search for a mutant locus.

Leaky *fad2-5* Allele Results from T-DNA Insertion in the 5′ Untranslated Region of the *FAD2* Gene

The comparison of the fatty acid compositions of *fad2-5* tissues with those of *fad2-1* and the wild type (Table 1; Lemieux et al., 1990) indicated that the homozygous T-DNA mutant contained considerable 18:1 desaturase activity. Despite the location of the T-DNA in the transcribed (but untranslated) region of the *FAD2* gene, a low level of *FAD2* transcript was detected in roots and shoots of *fad2-5* plants. Presumably, this low level of transcript is responsible for the increased 18:1 desaturation relative to other alleles of *fad2*.

The T-DNA from the 3850:1003 plasmid is large (>16 kb), and the presence of a 14.2-kb BamHI-EcoRI fragment in our rescued plasmid, although not conclusive, suggests that at least one copy of the T-DNA was incorporated into the genome at the site of insertion. For this reason, it is highly unlikely that the endogenous *FAD2* promoter is responsible for the synthesis of *FAD2* mRNA that we observed. Although we cannot rule out this possibility or the possibility that a promoter exists within the 1134-bp intron (Figure 1B), our findings strongly indicate that there is a weak but functional promoter within the left border region of the T-DNA. The promoter controls the synthesis of an RNA which, when processed, is of a similar length to the transcript produced in wild-type and *fad2-1* plants. Interestingly, a T-DNA insertion in the *fad3* gene of line 3707 of the T-DNA population also produced a leaky mutation (Yadav et al., 1993). In this case, the site of insertion is 612 bp 5′ to the initiation codon, but it is the right border of the T-DNA that is proximal to the *FAD3* open reading frame. From these results, we inferred that the residual activity of the endoplasmic reticulum 18:2 desaturase observed in homozygous mutants from line 3707 is mediated by a truncated form of the endogenous promoter confined to the 612-bp region 5′ to the initiation codon of the *FAD3* gene.

The ability of the insert from pF2a to complement the fatty acid phenotype of the *fad2-1* mutant confirmed the identity of the cloned gene. Given the low transcript level associated with the leaky *fad2-5* phenotype, it is not surprising that all but one of the transgenic hairy-root cultures produced more than 60% of the level of 18:2 plus 18:3 found in wild-type plants (Figure 2). At least two of the transgenic cultures exhibited slightly higher proportions of 18:2 plus 18:3 and lower 18:1 than the wild-type control, suggesting the possibility that *FAD2* overexpression can lead to more extensive desaturation of 18:1. However, the nearly complete (>90%) desaturation of 18:1 that occurred in wild-type Arabidopsis root cultures limits our ability to detect additional synthesis of polyunsaturated fatty acids as a result of *FAD2* overexpression. We would anticipate that the effects of overexpression might be more dramatic in plant species that contain higher levels of 18:1.

In both root and shoot tissues of wild-type Arabidopsis, the levels of *FAD2* transcript appeared to be high, considering that much lower transcript levels will support 50% of the desaturation attributable to the endoplasmic reticulum 18:1 desaturase in the wild type (Figure 3 and Table 1). This result suggests that in wild-type plants, the *FAD2* transcript may be present severalfold in excess of the amount needed to account for 18:1 desaturase activity and that there may be considerable translational and post-translational control of expression. Because the desaturation of 18:1 is the critical step in polyunsaturated lipid synthesis, it is possible that this excess is maintained to ensure that the enzyme activity is never limited by availability of transcript. Such a concept is consistent with the observation that the level of *FAD2* transcript does not increase following

transfer to low temperatures even though the membrane lipids of Arabidopsis, like those of many plants, became more unsaturated with decreasing temperature (Browse et al., 1986a).

FAD2 Sequence Shows Low Homology to Other Fatty Acid Desaturases

The homologies between the open reading frame of *FAD2* and previously described desaturases are low at both the nucleotide and protein levels. Comparisons of the predicted protein sequences of eight different fatty acid desaturases are included in Table 2. Only the comparisons with the Arabidopsis FAD3 and FAD7 (formerly FADD) sequences indicate moderate homologies (35 to 40% identity; 55 to 60% similarity), whereas the Brassica FAD6 homolog and cyanobacterial DesA sequences show weak homology (24% identity). As shown in Table 2, the FAD2 sequence does not show any more homology to the other membrane desaturase sequences than it does to the very divergent, soluble enzyme that catalyzes the desaturation of 18:0-ACP.

The overall low identity in these comparisons explains why *FAD3*, *FAD7*, and other cDNAs failed to identify the *FAD2* gene when they were used as heterologous probes. A second strategy that we and others employed in the search for *FAD2* was to design oligonucleotide probes to a nine–amino acid sequence (FVL/VGHDCGH; residues 97 to 105 in Arabidopsis FAD3) that is conserved between the higher plant 18:2 desaturases and the *desA* gene of *Synechocystis* strain PCC 6803 (Wada et al., 1990; Yadav et al., 1993). The corresponding sequence in FAD2 (WVIAHECGH; residues 101 to 109) is sufficiently divergent at the nucleotide level to preclude the possibility of successfully using such oligonucleotide probes to identify *FAD2* cDNA clones.

A considerable difference between the *FAD2* and *FAD7* genes is also seen in the structures of the genomic clones. The *FAD7* gene contains seven introns that vary in length from 79 to 301 bp and that are all located within the coding sequence (Iba et al., 1993). By contrast, the only intron in the *FAD2* sequence is large (1134 bp) and located 5′ to the open reading frame but still in the transcribed region of the gene. Unlike the Arabidopsis *FAD3* gene product but like the soybean and mung bean FAD3 homologs (Yadav et al., 1993), the predicted sequence of the FAD2 protein lacks either of the lysine-rich, carboxy-terminal motifs that have been suggested to represent the retention signal for integral membrane proteins in the endoplasmic reticulum (Jackson et al., 1990). The predicted FAD2 protein also lacks a recognizable signal sequence for targeting to the endoplasmic reticulum. This indicates that the protein may insert into the membrane post-translationally as does the rat 18:9 desaturase (Thiede et al., 1985).

Table 2. Homology between the Deduced Amino Acid Sequences of Fatty Acid Desaturases

Sequence	Identity[a]	Similarity[a]
	%	%
Arabidopsis FAD3	37.2	58.2
Arabidopsis FAD7	34.4	58.5
Brassica FAD6 homolog	24.3	48.6
Synechocystis Des A	23.8	50.5
Synechocystis Δ6	18.0	48.8
Brassica 18:0-ACP	19.1	41.7
Yeast-18:0-CoA	19.1	43.3
Rat-18:0-CoA	17.0	44.8

[a] Comparisons are with the FAD2 sequence employing the Gap program of the GCG package (Devereux et al., 1984) using a gap weight of 3.0 and length weight of 0.1.

Identification of Putative Iron Binding Motifs

The fatty acid desaturases fall into a general class of enzymes that contain iron that is not incorporated within a heme prosthetic group (Heinz, 1993). Recently, spectroscopic analysis has been used to demonstrate that the soluble 18:0-acyl-carrier-protein (18:0-ACP) desaturase is a member of a class of diiron-oxo proteins in which the two atoms of iron, bridged by an oxygen, are coordinated into the protein by histidine and carboxylic acid (aspartate and glutamate) residues (Fox et al., 1993). Because of the very close similarity between the reactions catalyzed by the membrane-bound desaturases and the desaturation of 18:0-ACP, it is reasonable to consider the possibility that the integral membrane proteins also contain a diiron-oxo reaction center, even though some data suggest that the mammalian membrane-bound 18:0–Coenzyme A desaturase may contain only one molecule of iron per polypeptide (Strittmatter et al., 1974).

Two quite different coordination structures are known from higher resolution x-ray crystallography of soluble diiron-oxo proteins. In the free radical protein of ribonucleotide reductase, the diiron cluster is ligated to two histidines and four acidic amino acids (Nordlund et al., 1990). The primary sequence of this protein contains two EXXH motifs (the critical residues are on the same face of an α-helix) that provide four of these six ligands. The same replicate EXXH motifs are present in unrelated enzymes (including the 18:0-ACP desaturase) that are recognized from their spectroscopic characteristics as diiron-oxo proteins (Fox et al., 1993). The structure of hemerythrin (Stenkamp et al., 1984) reveals a diiron cluster that is ligated through five histidines and two acidic residues. Three pairs of residues in the primary coordination sphere (H54 E58, H73 H77, and H106 D111) are three or four residues apart in the primary sequence, but the detailed structure of this active site is quite distinct from that found in ribonucleotide reductase (Nordlund et al., 1990). It should be noted that the EXXH motifs are found in enzymes that catalyze reactions for which high-valent iron-oxo structures have been proposed as catalytic intermediates, whereas hemerythrin is an O_2 binding protein with no catalytic capacity.

With these considerations in mind, we started our analysis of the putative sequences of FAD2 and the other desaturases

by attempting to identify possible iron binding motifs. Comparisons of the higher plant, membrane-bound desaturases FAD2, FAD3, FAD7 (Figure 5), and FAD6 (W. Hitz, personal communication) revealed sizable blocks of homology and similarity but only relatively short stretches where there was perfect agreement across all four sequences. The FAD2 sequence contains only one EXXH motif (residues 106 to 109; in the other plant desaturases this is present as DXXH) so that the protein does not meet the criteria of a putative diiron-oxo protein of the ribonucleotide reductase type. The other three plant sequences each contain a second D/EXXH motif, but these second copies are at very different positions in the various proteins. This suggests that they are not likely to be part of a conserved active site.

As a second stage of comparison, we assembled a total of eight distinctly different sequences including the four higher plant sequences together with the yeast OLE1 (Stukey et al., 1990), *Synechocystis* Δ6 and Δ12 (Wada et al., 1990; Reddy et al., 1993), and the rat 18:0-CoA (Thiede et al., 1986) desaturases. The wide evolutionary and biochemical divergence represented by these eight sequences allowed the overall homology to degenerate to a point at which it was possible to identify just three regions of strong conservation among

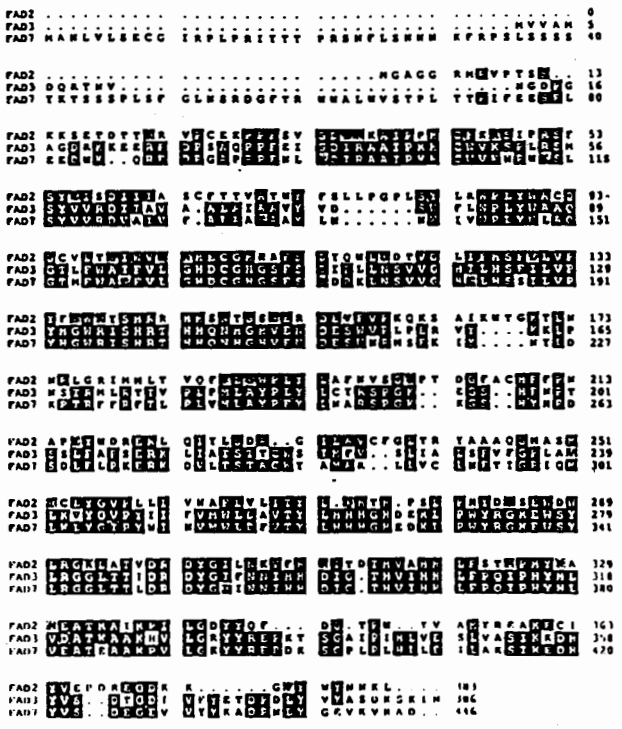


Figure 5. Comparison of the Deduced Amino Acid Sequences of Three Arabidopsis Fatty Acid Desaturases.

Identical and similar residues are shown on backgrounds of black and gray, respectively. The *FAD3* gene encodes the endoplasmic reticulum 18:2 desaturase; the *FAD7* gene (previously *FADD*) encodes a chloroplast 18:2 desaturase.

these integral membrane desaturases. The most strongly conserved motif is represented in FAD2 by WKYSHRRHH (residues 137 to 145). The tryptophan and all three histidine residues are present with the same spacing in all eight sequences, except that in the *Synechocystis* Δ12 desaturase, the spacing in the HXXHH box is increased by one residue (WRYRHNYLHH; residues 119 to 128) (Reddy et al., 1993). Interestingly, the tryptophan and the second histidine (with the same spacing as FAD2) are present in the two soluble plant acyl-ACP desaturases (Shanklin and Somerville, 1991; Cahoon et al., 1992) where the histidine is a critical residue in one-half of the proposed iron binding site (Fox et al., 1993).

The HXXHH motif is repeated toward the carboxy terminus of each sequence (in FAD2 as HVAHHLFS, residues 315 to 322). Again, all three histidine residues and the spacing are conserved in all eight membrane desaturases, except that the *Synechocystis* Δ12 desaturase once again has one extra residue in the spacing (HQVTHHLFP, residues 302 to 310). Finally, the FAD2 sequence HECGHHSF (residues 105 to 112) showed conservation of the first two histidines in all the higher plant and cyanobacterial enzymes as well as aspartate (glutamate in FAD2) as the second residue. In both the yeast and rat sequences, the spacing between the two histidines in this region is increased (HRLWSH in both), and there is no D/EXXH sequence present.

As discussed above, we failed to identify conserved, paired D/EXXH motifs within the eight sequences that were compared. It is possible that the positioning of the motifs varies between the different enzymes for functional reasons and/or that the ligands required to coordinate a diiron cluster are recruited from distinct individual sites throughout each protein sequence. However, the large number of highly conserved histidines among the blocks of homology that were identified suggest that if the membrane desaturases are diiron-oxo proteins, then a histidine-rich primary coordination shell for the iron cluster should be considered as a possibility, even though such a structure has, to date, only been identified in a protein that does not have a catalytic function. Alternatively, the membrane-bound desaturases may not be diiron-oxo proteins and the conserved histidine motifs may represent a different type of iron binding site. Of the three histidine-rich motifs that we have identified, the second and third in the protein sequence show the greatest degree of conservation. The spacing would place all three histidines on the same face of an α-helix (except for the *Synechocystis* Δ12 desaturase), although there is no evidence that the proteins are helical in these regions.

The Problem of Getting the Substrate Acyl Chain to the Active Site

The hydropathy profile calculated for the predicted FAD2 protein and displayed in Figure 6 is similar in several respects to those of other membrane-bound desaturases (Stukey et al., 1990; Wada et al., 1990; Reddy et al., 1993; E. Lark and J. Browse, unpublished data). Two long stretches (>45 residues

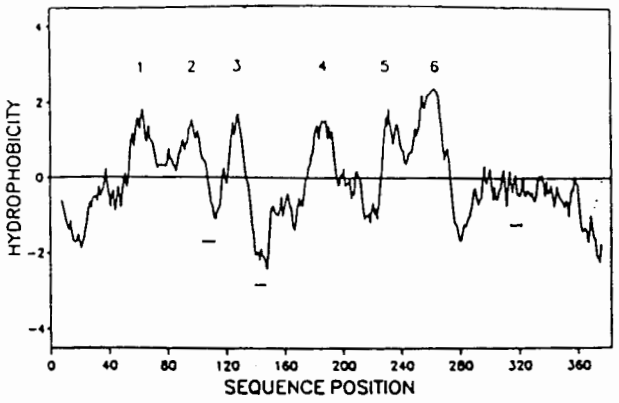


Figure 6. Hydropathy Plot of the Arabidopsis FAD2 Sequence.

The plot was made according to the method of Kyte and Doolittle (Devereux et al., 1984). Six candidate membrane-spanning sequences are numbered. The short lines indicate regions corresponding to the three histidine-rich sequences that show homology to other membrane-bound desaturases.

each) of hydrophobic residues are present (55 to 107 and 228 to 273 in FAD2) together with shorter hydrophobic sequences near the center of the molecule.

From comparisons of the yeast and rat proteins, Stukey et al. (1990) proposed a model in which the long hydrophobic sequences form two membrane-traversing loops, each comprised of two membrane-spanning, α-helical segments. In this structure, the bulk of the protein (240 of 334 residues for the rat 18:0-CoA desaturase) is on the cytosolic side of the endoplasmic reticulum membrane. However, in the FAD2 hydropathy profile, there are two additional hydrophobic stretches that are very good candidates for single-pass membrane-spanning segments (regions 3 and 4 in Figure 6). The hydropathy plots of other plant desaturases also show one or more potential (single-pass) membrane-spanning sequences in addition to the proposed loops (data not presented). One of these (region 3 shown in Figure 6) corresponds to a comparable peak in the profiles of the predicted yeast and rat proteins (Stukey et al., 1990). However, acceptance of all six potential membrane-spanning regions in the FAD2 structure would make it impossible to orient all the conserved histidine-containing motifs on the cytosolic side of the membrane. The most highly conserved motif (residues 137 to 145) would always be on the side of the membrane away from the other two. One possible explanation that would retain the proposed rat/yeast structure for the plant and cyanobacterial enzymes is that the additional hydrophobic regions are involved in other aspects of protein structural determination or interaction with the substrate membrane lipids. Potentially, it will be possible to resolve these questions for one or more of the desaturase proteins using direct measurements of membrane topology (Jennings, 1989).

Regardless of the actual topology of the 18:1 desaturase, it is clear from Figure 6 that all the putative active site histidines are located in relatively hydrophilic regions of the protein. Presumably, such exposed locations would facilitate interactions of the active site with cytochrome b_5 (the immediate electron donor to the desaturase) on the cytoplasmic surface of the membrane (Heinz, 1993). However, the enzyme will use the two electrons from cytochrome b_5 to abstract two hydrogens of an 18:1 acyl chain that are on carbons 12 and 13 and, thus, normally 25 Å or so below the membrane surface. Moving the electrons through the protein toward the center of the bilayer and drawing a glycerolipid molecule out of the membrane to effect desaturation are both energetically challenging options, although a precedent for partial removal of a glycerolipid substrate from the bilayer is well established in the case of phospholipase A_2 enzymes (Scott et al., 1990). The most straightforward explanation for this apparent conflict may lie in the extremely dynamic nature of the membranes' fatty acid core (Venable et al., 1993). The flexibility of the acyl chains allows the C12–C13 region of the desaturase substrate (where the double bond is to be placed) a small but significant probability of being located close to the membrane surface at any given time. The rate of oleate turnover by desaturation in Arabidopsis and other higher plants is slow, with a half-time, calculated from in vivo labeling experiments (Miquel and Browse, 1992), of 2 to 4 hr. Consequently, the frequency with which the C12–C13 region of any particular 18:1 acyl chain would be located close to the membrane surface might not limit the rate of desaturation (Heinz, 1993).

Conclusion

The cloning of the Arabidopsis *FAD2* gene and the unusual structure and expression pattern of the *fad2-5* insertional allele have provided information on the molecular-genetic control of the 18:1 desaturation step carried out by the product of the *FAD2* gene. Sequence analysis of the predicted FAD2 protein and other membrane-bound fatty acid desaturases has permitted useful speculation about the structure and location of a putative active site. In addition, the availability of genes encoding the endoplasmic reticulum 18:1 and 18:2 desaturases (Arondel et al., 1992; Yadav et al., 1993) should now permit the manipulation of tissue fatty acid compositions through overexpression and the use of antisense techniques. Such approaches will contribute to our understanding of how membrane lipid composition affects plant function and may lead to the useful manipulation of seed storage lipids to produce modified vegetable oils.

METHODS

Screening of the Arabidopsis T-DNA Population

Approximately 30 plants from each of 1800 lines in a population of *Arabidopsis thaliana* (ecotype Wassilewskija) that were transformed

with a modified T-DNA (Feldmann and Marks, 1987) were grown in 3-in-diameter pots in controlled environment chambers. Ten leaves (each from a different plant) were harvested from a pot and pooled for analysis (Browse et al., 1986b) to provide a single, average fatty acid composition for each segregating line. Once line 658 had been identified as segregating for increased 18:1, homozygous mutant segregants of this line were used in reciprocal crosses with plants containing the *fatty acid desaturation 2-1* (*fad2-1*) allele (Lemieux et al., 1990; Miquel and Browse, 1992) and the F_1 progeny analyzed for fatty acid composition. The *fad2-1* allele is derived from the Columbia ecotype. However, wild-type plants of Columbia and Wassilewskija exhibit very similar leaf fatty acid compositions so that comparisons of mutants in these two genetic backgrounds are entirely valid.

Two markers were used to follow segregation of the T-DNA inserts. Activity of the nopaline synthase gene was determined by assaying for the presence or absence of nopaline in extracts of single leaves using the method of Errampalli et al. (1991). Activity of the neomycin phosphotransferase II gene was determined by germinating seeds on agar medium containing 50 μg/mL kanamycin (Feldmann et al., 1989). The derivative line, 658-75, was identified as the T_4 progeny of a single T_3 plant that was segregating for the fatty acid phenotype and for a single kanamycin resistance locus.

Isolation of a T-DNA–Plant DNA Junction Fragment

Genomic DNA was prepared from leaves of homozygous mutant individuals of the 658-75 line by the method of Rogers and Bendich (1985). A sample (1.5 μg) of this DNA was digested in a 50-μL volume with 0 units of BamHI (Bethesda Research Laboratories) according to the manufacturer's instructions. Following purification by phenol extraction and ethanol precipitation, the digested DNA was resuspended at 2.5 μg/mL in a ligation buffer containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP and 4 units of T4 DNA ligase and incubated for 16 hr at 16°C. Competent cells (Bethesda Research Laboratories) were transfected with 10 ng of ligated DNA per 100 μL according to the manufacturer's specifications. Transformants were selected on Luria–Bertani plates (10 g Bacto Tryptone, 5 g Bacto yeast extract 5 g NaCl, 15 g agar per liter, pH 7.4) containing 100 μg/mL ampicillin. After overnight incubation at 37°C, the plates were scored for ampicillin-resistant colonies.

A single, ampicillin-resistant colony was used to start an overnight culture in Luria–Bertani medium containing 25 mg/L ampicillin. Plasmid DNA from pTF-658 was isolated from harvested cells by an alkaline lysis method (Ausubel et al., 1991) to provide DNA for restriction analysis and for the synthesis of labeled probes. The 1.6-kb EcoRI-BamHI fragment from the pTF658 plasmid was also subcloned into the pBluescript SK– vector (Stratagene) as pSI658.

Isolation of cDNA and Genomic Clones Corresponding to the *FAD2* Gene

The EcoRI-BamHI fragment from plasmid pTF-658 was purified and labeled with α-^{32}P-dCTP using a random priming kit (Bethesda Research Laboratories) according to the manufacturer's recommendations. Approximately 17,000 plaques of an Arabidopsis cDNA library in the λYES vector (Elledge et al., 1991) were screened essentially as described by Ausubel et al. (1991). Fifteen positively hybridizing plaques were identified, and five of these were subjected to plaque purification. The cDNA clones were converted to plasmids by incubation in *Escherichia coli* BNN-132 cells (Elledge et al., 1991), and ampicillin-resistant plasmid clones were grown and used as a source of plasmid DNA as described above. Four cDNA clones (pF2a, pF2b, pF2c, and pF2d) that were confirmed as hybridizing to the EcoRI-BamHI fragment from pTF-658 were subjected to restriction analysis and partial sequencing. Because the four cDNAs contained sequences in common, the longest cDNA clone pF2a (1.4 kb) was sequenced in both directions using a DNA polymerase (Sequenase T7; U.S. Biochemical Corp.) according to the manufacturer's instructions, and primers were designed from newly acquired sequences as the experiment progressed.

The pF2b cDNA insert was used as a radiolabeled probe to screen an Arabidopsis genomic DNA library (Yadav et al., 1993). DNA from several pure, strongly hybridizing phage were analyzed by gel blot hybridization to the cDNA probe to identify a 6-kb HindIII insert fragment that contained the entire coding region of the gene. This fragment was subcloned into the pBluescript SK– vector and partially sequenced as described above. Analysis of the predicted protein sequences of *FAD2* and other desaturase gene products was performed using the GCG sequence analysis software package (Devereux et al., 1984).

Plant Transformation

The 1.4-kb cDNA fragment of pF2a was subcloned into the EcoRI site of the binary vector pGA748 (kindly supplied by G. An, Washington State University, Pullman). This vector contains the cauliflower mosaic virus 35S promoter, the nopaline synthase 3′ terminator sequence, and the *NPTII* gene that confers kanamycin resistance (An et al., 1988). A vector construct (pGA-Fad2), cloned in *E. coli* and identified as being correctly inserted in the sense orientation by restriction enzyme analysis, was transformed into *Agrobacterium tumefaciens* R1000 (a C58 strain carrying the Ri plasmid) using the freeze–thaw method (Holsters et al., 1978) to produce strain R1000/pGA-Fad2. This strain and a corresponding control (no insert) were used to transform *fad2-1* and wild-type Arabidopsis using a stem-transformation protocol (Arondel et al., 1992).

Measuring Levels of *FAD2* Transcript

Shoot material of wild-type, *fad2-1*, and *fad2-5* Arabidopsis plants was harvested from plants as the first flowers were opening. Root material was harvested from 2-week-old plants grown in liquid culture (Miquel and Browse, 1992). For the chilling experiment, plants were grown at 22°C, 100 to 120 μmol quanta m^2 sec^{-1} constant light and 60 to 70% relative humidity for 17 days before transfer to 6°C with the same light and humidity. Samples of shoot material were taken at intervals after transfer and from control plants left at 22°C. Harvested plant tissue was rapidly frozen in liquid nitrogen and the RNA extracted from it by a phenol–SDS method (Ausubel et al., 1991). In the separations shown here, 10 μg of total RNA was fractionated by electrophoresis on 1.2% (w/v) agarose 2.2 M formaldehyde gels and blotted onto nitrocellulose membranes (Schleicher & Schuell). The blots were probed sequentially with ^{32}P-labeled probes from the cDNA insert of pF2a and from the Arabidopsis 28S rRNA gene as described by Ausubel et al. (1991).

ACKNOWLEDGMENTS

This work was supported by grants from the U.S. Department of Agriculture (No. 92-37301-7728) and the National Science Foundation (No. DCB-9105550) to J.B. We are grateful to those who helped to screen the T-DNA population, particularly Florence Garlick and Mary Aegerter. We thank Elmer Heppard, Russell Booth, and Pablo Scolnik for help with sequencing and mapping the clones. We would also like to thank Keith Davis (Ohio State University) for generous use of facilities. Sequence analysis was performed using the facilities of the VADMS Center at Washington State University. We are also grateful to Bill Hitz and John Shanklin for discussing ideas about possible protein structures. The λYES Arabidopsis cDNA library was a generous gift from the laboratory of Ron Davis.

Received October 26, 1993; accepted November 29, 1993.

REFERENCES

An, G., Ebert, P.R., and Mitra, A. (1988). Binary vectors. In Plant Molecular Biology Manual, S.B. Gelvin, and R.A. Schilperoort, eds (Dordrecht: Kluwer Academic), pp. 1–19.

Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M., and Somerville, C.R. (1992). Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. Science 258, 1353–1355.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds (1991). Current Protocols in Molecular Biology. (New York: John Wiley & Sons).

Behringer, F.J., and Medford, J.I. (1992). A plasmid rescue technique for the recovery of plant DNA disrupted by T-DNA insertion. Plant Mol. Biol. Rep. 10, 190–198.

Browse, J., and Somerville, C. (1991). Glycerolipid synthesis: Biochemistry and regulation. Annu. Rev. Plant Physiol. Plant Mol. Biol. 42, 467–506.

Browse, J., McCourt, P., and Somerville, C.R. (1986a). A mutant of Arabidopsis deficient in $C_{18:3}$ and $C_{16:3}$ leaf lipids. Plant Physiol. 81, 859–864.

Browse, J., McCourt, P.J., and Somerville, C.R. (1986b). Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. Anal. Biochem. 152, 141–145.

Cahoon, E.B., Shanklin, J., and Ohlrogge, J.B. (1992). Expression of a coriander desaturase results in petroselinic acid production in transgenic tobacco. Proc. Natl. Acad. Sci. USA 89, 11184–11188.

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

Elledge, S.J., Mulligan, J.T., Ramer, S.W. Spottswood, M., and Davis, R.W. (1991). λYES: A multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *Escherichia coli* mutations. Proc. Natl. Acad. Sci. USA 88, 1731–1735.

Errampalli, D., Patton, D., Castle, L., Mickelson, L., Hansen, K., Schnall, J., Feldmann, K., and Meinke, D. (1991). Embryonic lethals and T-DNA insertional mutagenesis in Arabidopsis. Plant Cell 3, 149–157.

Feldmann, K.A. (1991). T-DNA insertion mutagenesis in Arabidopsis: Mutational spectrum. Plant J. 1, 71–82.

Feldmann, K.A., and Marks, M.D. (1987). *Agrobacterium*-mediated transformation of germination seeds of *Arabidopsis thaliana*: A non-tissue culture approach. Mol. Gen. Genet. 208, 1–9.

Feldmann, K.A., Marks, M.D., Christianson, M.L., and Quatrano, R.S. (1989). A dwarf mutant of Arabidopsis generated by T-DNA insertion mutagenesis. Science 243, 1351–1354.

Fox, B.G., Shanklin, J., Somerville, C., and Münck, E. (1993). Stearoyl-acyl carrier protein Δ^9 desaturase from *Ricinus communis* is a diiron-oxo protein. Proc. Natl. Acad. Sci. USA 90, 2486–2490.

Harwood, J.L. (1980). Plant acyl lipids: Structure, distribution, and analysis. In the Biochemistry of Plants: A Comprehensive Treatise, Vol. 4, P.K. Stumpf, ed (New York: Academic Press), pp. 1–55.

Heinz, E. (1993). Biosynthesis of polyunsaturated fatty acids. In Lipid Metabolism in Plants, T.S. Moore, ed (Boca Raton, FL: CRC Press), pp. 33–89.

Holsters, M., de Waele, D., Depicker, A., Messens, E., Van Montagu, M., and Schell, J. (1978). Transfection and transformation of *A. tumefaciens*. Mol. Gen. Genet. 163, 181–187.

Hugly, S., and Somerville, C. (1992). A role for membrane lipid polyunsaturation in chloroplast biogenesis at low temperature. Plant Physiol. 99, 197–202.

Iba, K., Gibson, S., Nishiuchi, T., Fuse, T., Nishimura, M., Arondel, V., Hugly, S., and Somerville, S. (1993). A gene encoding a chloroplast ω-3 fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy number of the *fad7* mutant of *Arabidopsis thaliana*. J. Biol. Chem. 268, 24099–24105.

Jackson, M.R., Nilsson, T., and Peterson, P.A. (1990). Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. EMBO J. 9, 3153–3162.

James, D.W., Jr., and Dooner, H.K. (1990). Isolation of EMS-induced mutants in Arabidopsis altered in seed composition. Theor. Appl. Genet. 80, 241–245.

Jennings, M.L. (1989). Topography of membrane proteins. Annu. Rev. Biochem. 58, 999–1027.

Lemieux, B., Miquel, M., Somerville, C., and Browse, J. (1990). Mutants of Arabidopsis with alterations in seed lipid fatty acid composition. Theor. Appl. Genet. 80, 234–240.

Miquel, M., and Browse, J. (1992). Arabidopsis mutants deficient in polyunsaturated fatty acid synthesis. J. Biol. Chem. 267, 1502–1509.

Miquel, M., James, D.J., Jr., Dooner, H., and Browse, J. (1993). Arabidopsis requires polyunsaturated lipids for low-temperature survival. Proc. Natl. Acad. Sci. USA 90, 6208–6212.

Nordlund, P., Sjöberg, B.-M., and Eklund, H. (1990). Three-dimensional structure of the free radical protein of ribonucleotide reductase. Nature 345, 593–598.

Reddy, A.S., Nuccio, M.L., Gross, L.M., and Thomas, T.L. (1993). Isolation of a Δ^6-desaturase gene from the cyanobacterium *Synechocystis* sp. strain PCC 6803 by gain-of-function expression in *Anabaena* sp. strain PCC 7120. Plant Mol. Biol. 27, 293–300.

Reiter, S.R., Williams, J.G., Feldmann, K.A., Rafalski, J.A., Tingey, S.V., and Scolnick, P.A. (1992). Global and local genome mapping

in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA 89, 1477–1481.

Rogers, S.O., and Bendich, A.J. (1985). Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol. Biol. 5, 69–76.

Scott, D.L., White, S.P., Otwinowski, Z., Yuan, W., Gelb, M.H., and Sigler, P.B. (1990). Interfacial catalysis: The mechanism of phospholipase A_2. Science 250, 1541–1546.

Shanklin, J., and Somerville, C. (1991). Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proc. Natl. Acad. Sci. USA 88, 2510–2514.

Smith, W.L., and Borgeat, P. (1985). The eicosanoids: Prostaglandins, thromboxanes, leukotrienes and hydroxy-eicosaenoic acids. In Biochemistry of Lipids and Membranes, D.E. Vance and J.E. Vance, eds (Menlo Park, CA: Benjamin/Cummings), pp. 325–360.

Somerville, C.R., and Browse, J. (1991). Plant lipids: Metabolism, mutants, and membranes. Science 252, 80–87.

Stenkamp, R.E., Sieker, L.C., and Jensen, L.H. (1984). Binuclear iron complexes of methemerythrin and azidomethemerythrin at 2.0-Å resolution. J. Am. Chem. Soc. 106, 618–625.

Strittmatter, P., Spatz, L., Corcoran, D., Rogers, M.J., Setlow, B., and Bedline, R. (1974). Purification and properties of rat liver microsomal stearyl coenzyme A desaturase. Proc. Natl. Acad. Sci. USA 71, 4565–4569.

Stukey, J.E., McDonough, V.M., and Martin, C.E. (1990). The *OLE1* gene of *Saccharomyces cerevisiae* encodes the Δ-9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene. J. Biol. Chem. 265, 20144–20149.

Thiede, M.A., Ozols, J., and Strittmatter, P. (1985). The induction and characterization of rat liver stearyl-CoA desaturase mRNA. J. Biol. Chem. 260, 14459–14463.

Thiede, M.A., Ozols, J., and Strittmatter, P. (1986). Construction and sequence of cDNA for rat liver stearyl-CoA desaturase. J. Biol. Chem. 261, 13230–13235.

Thompson, G.A. (1993). Responses of lipid metabolism to developmental change and environmental perturbation. In Lipid Metabolism in Plants, T.S. Moore, ed (Boca Raton, FL: CRC Press), pp. 591–619.

Venable, R.M., Zhang, Y., Hardy, B.J., and Pastor, R.W. (1993). Molecular dynamics simulations of a lipid bilayer and of hexadecane: An investigation of membrane fluidity. Science 262, 223–226.

Wada, H., Gombos, Z., and Murata, N. (1990). Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation. Nature 347, 200–203.

Yadav, N.S., Wierzbicki, A., Aegerter, M., Caster, C.S., Pérez-Grau, L., Kinney, A.J., Hitz, W.D., Booth, J.R.B., Jr., Schweiger, B., Stecca, K.L., Allen, S.M., Blackwell, M., Reiter, R.S., Carlson, T.J., Russell, S.H., Feldmann, K.A., Pierce, J., and Browse, J. (1993). Cloning of higher plant ω-3 fatty acid desaturases. Plant Physiol. 103, 467–476.

Reprinted from Nature, Vol. 313, No. 6005, pp. 810-812, 28 February 1985

Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter

Joan T. Odell, Ferenc Nagy & Nam-Hai Chua

Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399, USA

Although promoter regions for many plant nuclear genes have been sequenced, identification of the active promoter sequence has been carried out only for the octopine synthase promoter[1]. That analysis was of callus tissue and made use of an enzyme assay. We have analysed the effects of 5′ deletions in a plant viral promoter in tobacco callus as well as in regenerated plants, including different plant tissues. We assayed the RNA transcription product which allows a more direct assessment of deletion effects. The cauliflower mosaic virus (CaMV) 35S promoter provides a model plant nuclear promoter system, as its double-strand DNA genome is transcribed by host nuclear RNA polymerase II from a CaMV minichromosome[2]. Sequences extending to −46 were sufficient for accurate transcription initiation whereas the region between −46 and −105 increased greatly the level of transcription. The 35S promoter showed no tissue-specificity of expression.

The 35S promoter region was isolated as a *Bgl*II fragment extending from −941 to +208 with respect to the transcription start site mapped for the 35S RNA found in CaMV-infected turnip leaves[3]. The polyadenylation site for the 19S and 35S CaMV transcripts located at +180 (ref. 3) was deleted, as described in Fig. 1 legend, to eliminate any possible processing signals in the promoter fragment. A 3′ deleted promoter fragment extending to +9 was deleted at its 5′ end (see Fig. 1) and fragments extending to −343, −168, −105 and −46 were chosen for analysis.

An abbreviated human growth hormone gene (*hgh*)[4] was added as a test gene downstream to the 35S promoter deletion fragments. Information on plant cell recognition of animal gene splice and 3′ polyadenylation signals obtained from analysis of *hgh* RNA transcribed in transformed plant cells will be presented elsewhere (A. Hunt, N. Chu, J.T.O., F.N. and N.-H.C., in preparation). The 35S promoter-*hgh* chimaeric gene was inserted in the pMON178 tumour-inducing (Ti)-plasmid vector, a derivative of pMON120 (ref. 5). Included in this vector is the nopaline synthase (NOS) promoter placed 5′ to the neomycin phosphotransferase-II (*npt*-II) coding region (NOS promoter-*npt*-II gene), which is co-transferred with the 35S promoter-*hgh* gene into the tobacco genome and provides an internal standard for comparison of the activities from different 35S promoter deletion fragments.

Following tri-parental matings[5,6], *Agrobacterium tumefaciens* containing both chimaeric genes was used to infect SR1 *Nicotiana tabacum* cells by wounding[5] and co-cultivation[5,7]. Calli obtained by wounding were selected for kanamycin resistance (50 μg ml^{-1}) and hormone independence. Plants were regenerated from co-culture with selection for kanamycin resistance, making use of the short transfer property of the split-end vector system[8] which eliminates the T-DNA tumour genes.

Southern blots of DNA extracted from transformed calli (Fig. 2) showed the expected 1.58-kilobase (kb) *Eco*RI fragment containing the NOS promoter-*npt*-II gene and a 1.7–2.0-kb *Eco*RI fragment containing the 35S promoter-*hgh* gene, the size depending on the length of the specific promoter deletion fragment involved, indicating that no rearrangements had occurred in these two genes. Comparison with reconstructions of gene copy numbers in lanes 5–7 indicate that the copy number of these two genes varies between different transformed calli, but

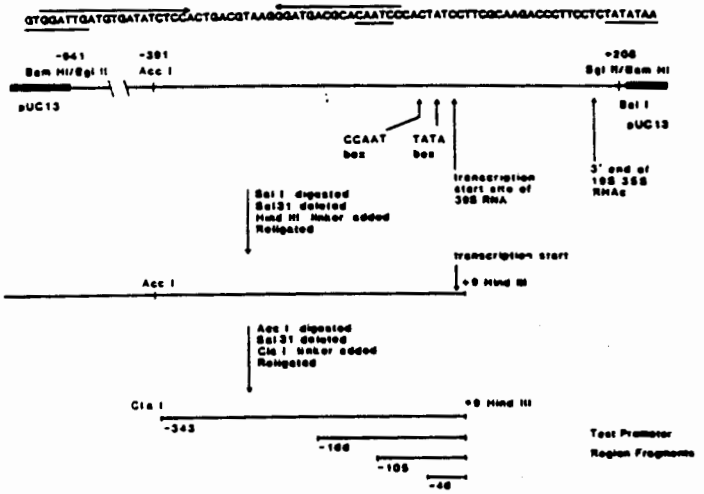


Fig. 1 Construction of 35S promoter region fragments. A 1.15-kb *Bgl*II fragment was subcloned from pCS101, a clone containing the entire Cabb-S CaMV genome[3], into the *Bam*HI site of pUC13. The resulting plasmid was linearized at the *Sal*I site in the pUC13 polylinker next to the 3′ end of the promoter fragment, digested with *Bal*31 exonuclease[11], ligated to *Hin*dIII linkers and recircularized. Clones were analysed for the extent of 3′ deletion by polyacrylamide gel sizing of the *Acc*I/*Hin*dIII fragments and finally by dideoxy sequencing[12] of subclones in pUC using the universal primer. The plasmid containing a 3′ deletion fragment with the *Hin*dIII linker at +9 was linearized with *Acc*I (site at −391), digested with *Bal*31 exonuclease, ligated to *Cla*I linkers and recircularized. Clones were analysed for the extent of 5′ deletion by polyacrylamide gel sizing of the *Cla*I/*Hin*dIII fragment, followed by dideoxy sequencing of subclones in pUC using either the universal primer or primer generation by exonuclease III digestion[13]. Above is the sequence of the −105 to −25 region of the 35S

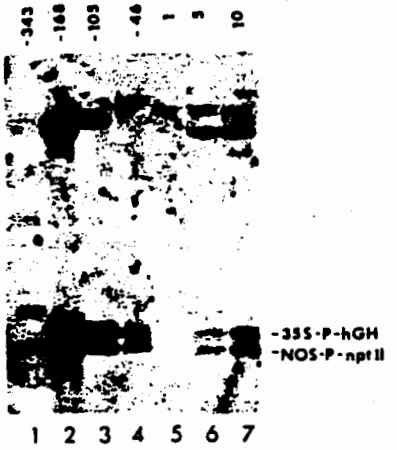


Fig. 2 Southern blot analysis of DNA from transformed tobacco calli. DNA was prepared, digested with *Eco*RI, electrophoresed on a 0.7% agarose gel and blotted onto a nitrocellulose filter[15]. A plasmid constructed to serve as the hybridization probe contains a *Bam*HI/*Sma*I *hgh* gene fragment and a *Bam*HI/*Bgl*II *npt*-II gene fragment cloned into pUC12 (GH-Neo24). The plasmid was nick translated [16] and hybridized to the Southern blot by the method of Thomashow *et al.*[17]. The following samples contain 15 μg of calli DNA transformed with: lane 1, −343 35S promoter-*hgh*; lane 2, −168 35S promoter-*hgh*; lane 3, −105, 35S promoter-*hgh*; lane 4, −46 35S promoter-*hgh*. Reconstructions of the NOS promoter-*npt*-II gene and 35S promoter-*hgh* gene copy numbers contain 15 μg of control untransformed plant DNA mixed with different amounts of the pMON178 plasmid containing the −105 35S promoter-*hgh* gene: lane 5, 17 pg = 1 copy; lane 6, 85 pg = 5 copies; lane 7, 170 pg = 10 copies. The bands near the top of the filter in lanes 1–4 result from hybridization of the pBR322 sequences in the GH-Neo24 probe plasmid to pBR322 sequences in the

that there is an equal number of the two genes in each callus, supporting the assumption that the two genes are co-transferred into the tobacco genome. Thus, the NOS promoter-*npt*-II gene provides a control for variables inherent in the transformation system including copy number and integration site.

Northern blots of poly(A)$^+$ RNA from transformed calli carrying each of the 35S promoter deletions showed the expected *npt*-II transcript and also a 2.3-kb transcript that contains *hgh* sequences (Fig. 3). The 35S promoter is responsible for transcription of the *hgh* RNA, as S_1 protection experiments indicate that the 5′ end of this RNA maps to the normal 35S transcription start site (Fig. 4). All the deleted promoters initiate transcription at the same site (data not shown). To compare the transcriptional activities of the different deletions, the ratio of *hgh* to *npt*-II transcripts was determined for each RNA sample. By quantitating silver grains eluted from bands on X-ray films[9], the *hgh*/*npt*-II ratios were: −46, 0.16/1; −105, 0.9/1; −168, 2.6/1; −343, 3.1/1. The *hgh*/*npt*-II ratio is the same for different RNA preparations from calli transformed with the same deletion fragment (data not shown). The −46 promoter fragment has an approximately 20-fold lower transcription level than the −343 fragment, whereas the −105 fragment has a 3-fold decrease and the −168 fragment has no significant decrease in activity. We also found that the −343 fragment and a −941 fragment have no significant difference in promoter activity (data not shown).

Regenerated transformed tobacco plants contain one to two copies per genome of the NOS promoter-*npt*-II and 35S promoter-*hgh* genes, determined by Southern blots (data not shown). This low copy number in plants, compared with the high copy numbers in calli (Fig. 2), could be explained by the transformation procedure (co-cultivation versus wounding) or by the type of T-DNA transfer event involved (short verus long). Northern blots show that leaves of plants transformed with each of the four 35S promoter deletions contain the same 2.3-kb *hgh* RNA found in transformed calli (Fig. 3*c*). By S_1 analysis the 5′

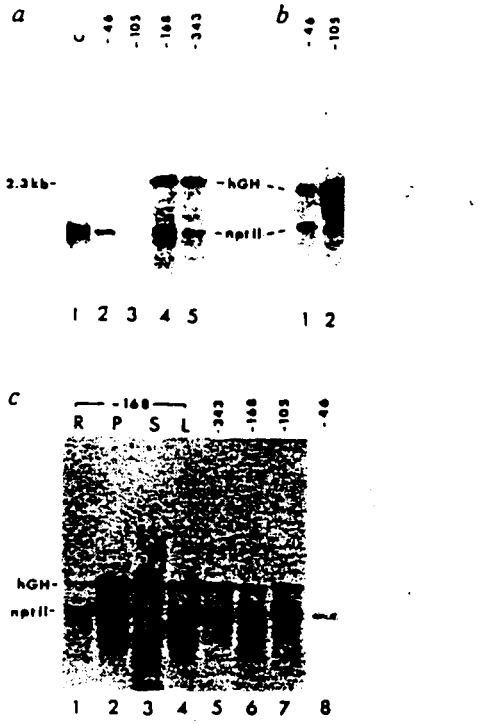


Fig. 3 Northern blot analysis of transformed tobacco calli transcripts. RNA was prepared from calli using guanidinium thiocyanate as a protein denaturant[18] and polyadenylated RNA was isolated by chromatography on poly(U)-Sepharose[19]. RNA samples were denatured with glyoxal, electrophoresed on a 1% agarose gel and transferred to nitrocellulose filters[20]. The filters were hybridized with nick-translated probes[16] for 40 h in 50% formamide, 5× SSC, 2× Denhardt's, 20 mM sodium phosphate, 0.1% SDS and 100 μg ml^{-1} denatured DNA. Filters were then washed in 2×SSC, 0.1% SDS at 23 °C followed by 0.1× SSC, 0.1% SDS at 55 °C for 1 h. *a*, Purified DNA fragments from the *npt*-II (*Bam*HI/*Hin*dIII) and *hgh* (*Bam*HI/*Sma*I) genes labelled with approximately equal specific activities (10^8 c.p.m. μg^{-1}) were hybridized to a filter containing 0.5–2 μg poly(A)$^+$ RNA from calli transformed with: lane 1, control pMON vector; lane 2, −46 35S promoter-*hgh*; lane 3, −105 35S promoter-*hgh*; lane 4, −168 35S promoter-*hgh*; lane 5, −343 35S promoter-*hgh*. *b*, Purified *npt*-II and *hgh* fragments were labelled so that the specific activity of the *npt*-II fragment was approximately 10-fold lower than that of the *hgh* fragment. Hybridization was to 2 μg poly(A)$^+$ RNA from calli transformed with: lane 1, −46 35S promoter-*hgh*; lane 2, −105 35S promoter-*hgh*. *c*, The GH-Neo24 plasmid described in Fig. 2 legend was used as the probe. Each RNA preparation was made from a pool of several plants derived from different transformation events, but carrying the same 35S promoter deletion fragment. Gel samples contain ~2 μg poly(A)$^+$RNA from: lanes 1–4, RNA from different parts of plants transformed with −168 35S promoter-*hgh*; lane 1, roots; lane 2, petals; lane 3, stems; lane 4, leaves; leaf RNA from plants transformed with: lane 5, −343 35S promoter-*hgh*; lane 6, −168 35S promoter-*hgh*; lane 7, −105 35S promoter-*hgh*; lane 8, −46 35S promoter-*hgh*

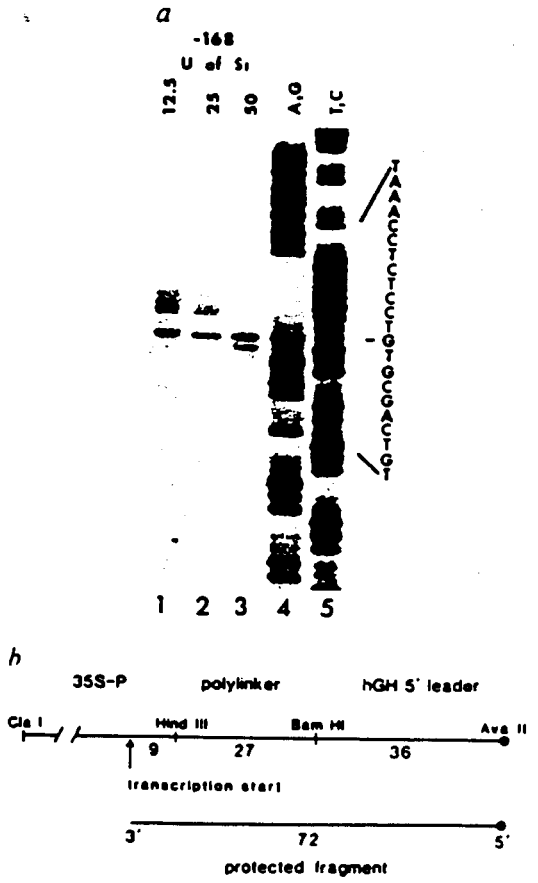


Fig. 4 *a*, S_1 nuclease mapping[21] of the 5′ end of the *hgh* transcript to the CaMV sequence. Total RNA (10 μg) from callus carrying the −168 35S promoter-*hgh* gene construction was hybridized to the probe diagrammed in *b* in 80% formamide at 37 °C. Samples were digested in 100 μl with the following amounts of S_1 nuclease: lane 1, 12.5 U; lane 2, 25 U; lane 3, 50 U. The same probe was sequenced using the Maxam and Gilbert technique[22]: lane 4, A+G reaction; lane 5, T+C reaction. All samples were electrophoresed on an 8% polyacrylamide/7 M urea gel. The DNA sequence[13] surrounding the *hgh* RNA 5′ end is shown to the right. *b*, Diagram of S_1 probe and region protected by *hgh* RNA. A *Cla*I/*Ava*II DNA fragment extending from the 5′ border of the 35S promoter fragment and into the 5′ leader region of the *hgh* transcript was 5′ end-labelled and strand separated on a polyacrylamide gel. Between the probe's labelled *Ava*II end and the 5′ end of the *hgh*

end of this leaf *hgh* RNA was found to be identical to that of the callus *hgh* RNA (data not shown). The effects of the deletions on promoter activity in transformed plant leaves (Fig. 3*c*) closely resemble results described for the transformed calli. The 35S promoter was also active in roots, petals and stems of transformed plants (Fig. 3*c*), with deletions having no specific effects on tissue expression (data not shown). The ratio of *hgh*/*npt*-II transcripts is constant in the different tissues. Both transcripts appear reduced in the root RNA preparation, but this could be due to varying amounts of ribosomal RNA contamination in the polyadenylated RNA preparations.

Here, we have shown that although the normal host range of CaMV is limited to members of the Cruciferae, the 35S promoter is active when isolated as a fragment from the viral genome and integrated into the tobacco genome. Thus, accurate transcription from the 35S promoter does not depend on any *trans*-acting viral gene products. The ability to regenerate tobacco plants from transformed protoplasts has allowed us to demonstrate that the 35S promoter is expressed in leaves, stems, roots and petals.

Promoter deletion analysis in transformed calli and plants showed that a −46 fragment, which does contain a TATA-box sequence (see Fig. 1), produces a low level of correctly initiated transcripts. The region between −46 and −105 which greatly increases the level of transcription contains a CAAT-box sequence, an inverted repeat region and a sequence resembling the consensus core for enhancers in animal systems ($GTGG^{AAA}_{TTT}G$) (ref. 10; see Fig. 1). We are investigating whether one or more of these features plays a substantial role in increasing the level of 35S promoter activity or could act to increase transcription from a foreign promoter.

We thank Dr Ken Richards for the CaMV clone, Dr Steve Rogers for pMON178 and Dr George Pavlakis for the abbreviated *hgh* gene. We also thank I. Roberson and K. Thurman for technical assistance and Dr M. Boutry and Dr G. Morelli for helpful discussions. This work was supported by a grant from the Monsanto Company. J.T.O. holds an NIH postdoctoral fellowship (5F32AI06342).

Received 3 September; accepted 4 December 1984.

1. Koncz, C. *et al. EMBO J.* **2**, 1597–1603 (1983).
2. Oliszewski, G., Hagen, G. & Guilfoyle, T. J. *Cell* **29**, 395–402 (1982).
3. Guilley, H., Dudley, R. K., Jonard, G., Balazs, E. & Richards, K. E. *Cell* **30**, 763–773 (1982).
4. Pavlakis, G. N. & Hamer, D. *Rec. Prog. Horm. Res.* **39**, 353–385 (1983).
5. Fraley, R. T. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4803–4807 (1983).
6. Ditta, G., Stanfield, S., Corbin, D. & Helinski, D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7347–7351 (1980).
7. Marton, J., Wullems, G. S., Molendijk, L. & Schilperoort, R. A. *Nature* **277**, 129–131 (1979).
8. Horsch, R. B. *et al. Science* **223**, 496–498 (1984).
9. Suissa, M. *Analyt. Biochem.* **133**, 511–514 (1983).
10. Gluzman, Y. & Shenk, T. (eds) *Current Communications in Molecular Biology* (Cold Spring Harbor, New York, 1983).
11. Sakonju, S., Bogenhagen, D. F. & Brown, D. D. *Cell* **19**, 13–26 (1980).
12. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
13. Guo, L.-H. & Wu, R. *Meth. Enzym.* **100**, 60–96 (1983).
14. Franck, A., Guilley, H., Jonard, G., Richards, K. & Hirth, L. *Cell* **21**, 285–294 (1980).
15. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
16. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **133**, 237–251 (1977).
17. Thomashow, M. F., Nutter, R. C., Montoya, A. L., Gordon, M. P. & Nester, E. W. *Cell* **19**, 1729–1739 (1980).
18. Glisen, V., Crkvenjakov, R. & Byus, C. *Biochemistry* **13**, 2633–2637 (1974).
19. Broglie, R., Bellemare, G., Bartlett, S. G., Chua, N.-H. & Cashmore, A. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7304–7308 (1981).
20. Thomas, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
21. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
22. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).

The Plant Cell, Vol. 1, 1095–1109, November 1989

Soybean Seed Protein Genes Are Regulated Spatially during Embryogenesis

Luis Perez-Grau and Robert B. Goldberg[1]

Department of Biology, University of California, Los Angeles, California 90024-1606

We used in situ hybridization to investigate Kunitz trypsin inhibitor gene expression programs at the cell level in soybean embryos and in transformed tobacco seeds. The major Kunitz trypsin inhibitor mRNA, designated as KTi3, is first detectable in a specific globular stage embryo region, and then becomes localized within the axis of heart, cotyledon, and maturation stage embryos. By contrast, a related Kunitz trypsin inhibitor mRNA class, designated as KTi1/2, is not detectable during early embryogenesis. Nor is the KTi1/2 mRNA detectable in the axis at later developmental stages. Outer perimeter cells of each cotyledon accumulate both KTi1/2 and KTi3 mRNAs early in maturation. These mRNAs accumulate progressively from the outside to inside of each cotyledon in a "wave-like" pattern as embryogenesis proceeds. A similar KTi3 mRNA localization pattern is observed in soybean somatic embryos and in transformed tobacco seeds. An unrelated mRNA, encoding β-conglycinin storage protein, also accumulates in a wave-like pattern during soybean embryogenesis. Our results indicate that cell-specific differences in seed protein gene expression programs are established early in development, and that seed protein mRNAs accumulate in a precise cellular pattern during seed maturation. We also show that seed protein gene expression patterns are conserved at the cell level in embryos of distantly related plants, and that these patterns are established in the absence of non-embryonic tissues.

INTRODUCTION

Seed proteins are encoded by a diverse gene set that is highly regulated during the plant life cycle (Goldberg, Barker, and Perez-Grau, 1989). Seed protein genes are regulated temporally during embryogenesis, are either repressed or active at low levels in the mature plant, and are often expressed at different levels in the cotyledon and axis (Goldberg et al., 1981a; Meinke, Chen, and Beachy, 1981; Ladin et al., 1987; Goldberg et al., 1989). Both transcriptional and post-transcriptional processes have been shown to regulate seed protein gene expression (Walling, Drews, and Goldberg, 1986; Goldberg et al., 1989). Gene transfer experiments have identified regions 5′ to several seed protein genes that are responsible, in part, for their developmental-specific expression programs (Chen, Schuler, and Beachy, 1986; Chen, Pan, and Beachy, 1988; Bustos et al., 1989; Chen et al., 1989; Goldberg et al., 1989; Jordano, Almoguera, and Thomas, 1989). In some cases, these regulatory regions have been shown to interact with DNA binding proteins that may play a role in seed protein gene transcription (Jofuku, Okamuro, and Goldberg, 1987; Allen et al., 1989; Bustos et al., 1989; Jordano et al., 1989; Riggs, Voelker, and Chrispeels, 1989).

In the accompanying paper (Jofuku and Goldberg, 1989), we showed that Kunitz trypsin inhibitor genes are expressed differentially during the soybean life cycle and in transformed tobacco plants. Individual members of the Kunitz trypsin inhibitor gene family have distinct qualitative and quantitative expression programs, suggesting that each gene has a unique *cis*-element combination that targets its activity to a specific differentiated state (Jofuku and Goldberg, 1989). A major unresolved question is how Kunitz trypsin inhibitor genes, and other seed protein genes, are regulated with respect to cell type during embryogenesis. That is, how do differentiation events that result in specific embryo cell types correlate with processes that lead to the accumulation of individual seed protein mRNAs?

In this study, we investigated the cellular expression programs of two Kunitz trypsin inhibitor gene classes, designated as KTi1/2 and KTi3 (Jofuku and Goldberg, 1989; Jofuku, Schipper, and Goldberg, 1989), and compared these programs with those of the β-conglycinin storage protein gene family (Barker, Harada, and Goldberg, 1988; Chen et al., 1988; Chen et al., 1989; Harada, Barker, and Goldberg, 1989). Our results show that seed protein genes are highly regulated with respect to cell type in both zygotic and somatic embryos, that seed protein mRNAs accumulate in a "wave-like" pattern during embry-

[1] To whom correspondence should be addressed.

ogenesis, and that similar events occur during the development of transformed tobacco seeds. We conclude that embryo cells become committed to express specific seed protein genes early in embryogenesis, that the seed protein mRNA localization patterns are conserved between distantly related plant species, and that these patterns can occur in the absence of surrounding non-embryonic seed tissue.

RESULTS

Kunitz Trypsin Inhibitor mRNA Is Present Early in Embryogenesis

In the accompanying paper (Jofuku and Goldberg, 1989), we described the general expression programs for two Kunitz trypsin inhibitor gene classes, designated as KTi1/2 and KTi3. The KTi1 and KTi2 genes are almost identical to one another, direct the synthesis of mRNAs that accumulate to different levels during embryogenesis, and probably encode proteins lacking trypsin inhibitor activity (Jofuku and Goldberg, 1989). By contrast, the KTi3 gene is only 80% similar to the KTi1 and KTi2 genes at the nucleotide level, is expressed at a higher level during embryogenesis, and encodes the major Kunitz trypsin inhibitor protein present in soybean seeds (Jofuku and Goldberg, 1989; Jofuku et al., 1989). All three Kunitz trypsin inhibitor genes are expressed at low levels in mature plant organ systems; however, the KTi1/2 and KTi3 gene expression patterns differ (Jofuku and Goldberg, 1989).

We hybridized KTi1 and KTi3 anti-mRNA probes (see Methods) in situ with seed sections containing globular, heart, and cotyledon stage embryos. The KTi1 anti-mRNA probe will hybridize with both KTi1 and KTi2 mRNAs under the in situ RNA/RNA hybridization conditions used (Cox and Goldberg, 1988; G. de Paiva and R.B. Goldberg, unpublished results). By contrast, the KTi3 anti-mRNA probe will hybridize only with KTi3 mRNA (G. de Paiva and R.B. Goldberg, unpublished results). Together, these probes measure the distribution of KTi1/2 and KTi3 mRNA molecules in developing seeds.

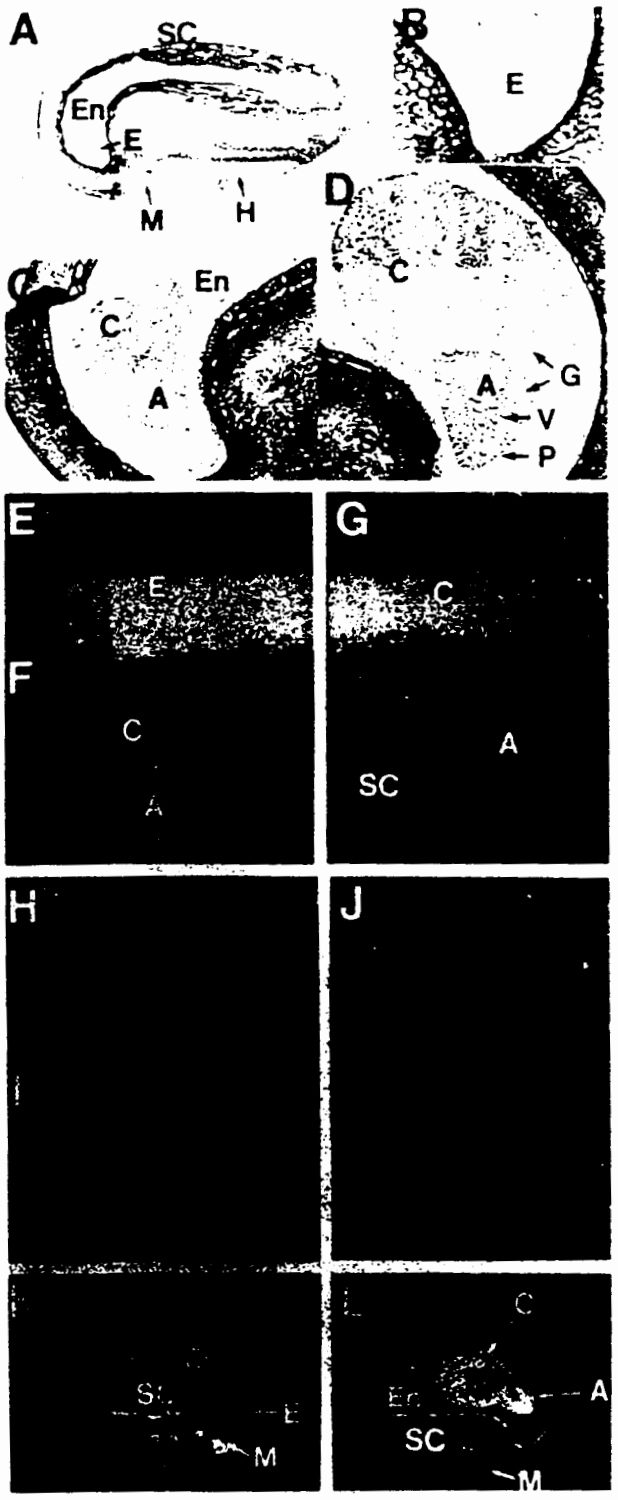
Figures 1A to 1D show bright-field photographs of longitudinal seed sections (see Methods) containing embryos in the initial stages of development. Figure 1A shows the structure of a soybean seed shortly after fertilization. A 500-cell globular stage embryo (E), shown at a higher magnification in Figure 1B, is surrounded by non-embryonic endosperm tissue (En) and seed coat tissue (SC). Both the micropyle (M) and hilum (H) seed coat regions can be distinguished at this developmental stage. Figure 1C shows that embryo cells are morphologically similar until the heart stage, when polarization into the cotyledons (C) and axis (A) occurs. As shown in Figure 1D, the cotyledons and axis become more prominent at the cotyledon stage, or period of rapid cell division (Goldberg et al., 1981b; Goldberg et al., 1989), and protoderm tissue (P), ground meristem tissue (G), and pre-vascular tissue (V) can be visualized in both embryonic organ systems.

Figures 1E to 1G show that the KTi1/2 anti-mRNA probe did not produce hybridization grains over background levels with RNA present in the globular stage embryo (Figure 1E), heart stage embryo (Figure 1F), or cotyledon stage embryo (Figure 1G). Nor were there grains above background within non-embryonic endosperm or seed coat tissue. By contrast, Figure 1H shows that the KTi3 anti-mRNA probe hybridized intensely with RNA in cells at the micropylar end of the globular stage embryo, producing a polarized pattern. As seen in Figures 1I and 1J, the KTi3 anti-mRNA probe also hybridized with RNA present within the axis of heart stage and cotyledon stage embryos, but did not hybridize detectably with RNA present within the cotyledon. Hybridization grains were visualized only over axis ground meristem regions, indicating that KTi3 mRNA was localized specifically within this tissue early in development.

We hybridized ^{3}H-poly(U) with the seed sections containing globular and cotyledon stage embryos to localize the distribution of total poly(A) RNA molecules during early development. As seen in Figures 1K and 1L, hybridization grains appeared over the embryo, endosperm, and seed coat regions. By contrast with the results obtained with the KTi3 anti-mRNA probe (Figures 1H and 1J), the distribution of grains was uniform throughout each embryo (Figures 1K and 1L). Together, our data show that the KTi1/2 and KTi3 mRNAs accumulate differentially during early seed development, that globular embryo cells destined to form part of the axis accumulate KTi3 mRNA, and that KTi3 mRNA is localized only within axis ground meristem cells early in embryogenesis.

Kunitz Trypsin Inhibitor mRNAs Do not Accumulate Simultaneously in All Embryo Cells during Maturation

We hybridized the KTi1/2 and KTi3 anti-mRNA probes in situ with seed sections containing maturation stage embryos to localize Kunitz trypsin inhibitor molecules at later developmental stages. Figures 2A to 2E show bright-field photographs of transverse seed sections (see Methods) containing maturation stage embryos. Two prominent cotyledons (C) can be visualized that are surrounded by endosperm tissue (En) and seed coat tissue (SC). Within each cotyledon, storage parenchyma cells can be distinguished from the more darkly stained vascular tissue (V). As maturation progresses, the cotyledons enlarge by cell expansion (Goldberg et al., 1981b; Goldberg et al., 1989) and fill the entire seed cavity (Figures 2A to 2D).



The KTi3 Kunitz Trypsin Inhibitor mRNA Accumulates in a Wave-Like Pattern

Figure 2F shows that the KTi3 anti-mRNA probe hybridized intensely with RNA present within the cotyledons of embryos at 25 days after flowering (DAF). As seen in Figure 2F, KTi3 mRNA was localized primarily within cells along the outer edge of each cotyledon, and the hybridization grains were uniformly distributed along the periphery. No hybridization grains above those produced with the KTi3 mRNA control probe (Figure 2O) were visualized within inner cotyledon cells or within non-embryonic seed coat and endosperm tissues. Figures 2G and 2H show that, as maturation continued, there was a progressive localization of KTi3 mRNA from the outer to inner margins of both cotyledons. By 70 DAF, KTi3 mRNA was distributed uniformly throughout each cotyledon. However, hybridization grains were not detectable within vascular tissue (Figures 2G and 2H), indicating that KTi3 mRNA was localized primarily within storage parenchyma cells. By contrast, Figure 2I shows that the KTi3 anti-mRNA probe did not produce a uniform hybridization pattern with embryos later in maturation (80 DAF). KTi3 mRNA was concentrated primarily along the perimeter of each cotyledon similar to

Figure 1. Kunitz Trypsin Inhibitor mRNA Localization during Early Soybean Embryogenesis.

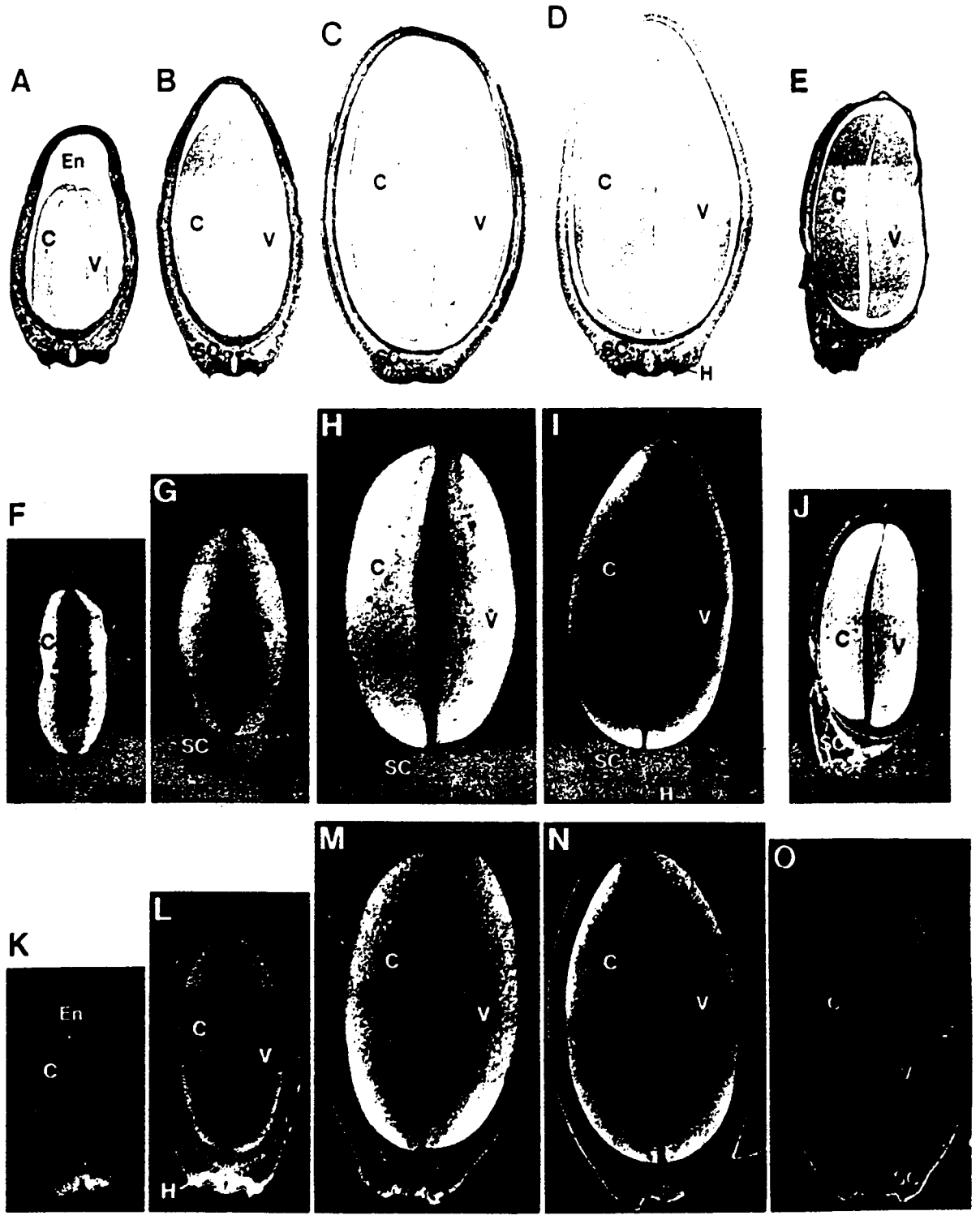
(A) Developing seed with a globular stage embryo. E, En, H, M, and SC refer to embryo, endosperm, hilum, micropyle, and seed coat, respectively. Actual seed length was 1.25 mm. Photograph was taken with bright-field microscopy.

(B) to **(D)** Bright-field photographs of a globular stage embryo **(B)**, a heart stage embryo **(C)**, and a cotyledon stage embryo **(D)**. Relative to the seed shown in **(A)**, the magnification factors for the globular, heart, and cotyledon stage embryos were ×4, ×2, and ×2, respectively. A, C, G, P, and V refer to axis, cotyledon, ground meristem, protoderm, and procambium or pre-vascular tissue, respectively.

(E) to **(G)** In situ hybridization of a KTi1/2 anti-mRNA probe with a globular stage embryo **(E)**, a heart stage embryo **(F)**, and a cotyledon stage embryo **(G)**. Magnification factors relative to the seed shown in **(A)** were ×4, ×2, and ×2 for the globular, heart, and cotyledon stage embryos, respectively. Photographs were taken by dark-field microscopy. White grains represent background hybridization levels, and were identical in density to those produced with the control KTi1/2 mRNA probe (data not shown).

(H) to **(J)** In situ hybridization of a KTi3 anti-mRNA probe with a globular stage embryo **(H)**, a heart stage embryo **(I)**, and a cotyledon stage embryo **(J)**. Magnification factors relative to the seed shown in **(A)** were ×4, ×2, and ×2 for the globular, heart, and cotyledon stage embryos, respectively. White areas represent regions containing RNA/RNA hybrids.

(K) and **(L)** In situ hybridization of ^{3}H-poly(U) with a seed containing a globular stage embryo **(K)** and a seed containing a cotyledon stage embryo **(L)**. Magnification factors for both photographs were ×0.8 relative to the seed shown in **(A)**.


A
En
C
V
B
C
V
SC
C
C
V
D
C
V
SC
H
E
C
V
H
F
C
G
SC
H
C
V
SC
I
C
V
SC
H
J
C
V
SC
K
En
C
L
C
V
H
M
C
V
N
C
V
O
C

the pattern observed with 25 DAF and 55 DAF embryos (Figures 2F and 2G).

Figure 2J shows that ^{3}H-poly(U) hybridized intensely with RNA present in embryonic and non-embryonic regions of a 55 DAF seed. Hybridization grains were uniform throughout each cotyledon and were present within both vascular and storage parenchyma tissues. This result contrasts the non-uniform KTi3 mRNA distribution observed within cotyledons of embryos at the same stage of development (Figure 2G). Hybridization of the KTi3 anti-mRNA probe with embryo longitudinal sections (data not shown) produced results identical to those obtained with the transverse sections (Figures 2F to 2I). These observations indicated that the KTi3 mRNA localization pattern was different from that of embryo total poly(A) RNA and was not confined to a single region of the developing embryo. Together, our results show that the KTi3 mRNA is localized specifically within embryo cotyledon storage parenchyma cells, that KTi3 mRNA accumulates progressively in a wave-like pattern from the outer to inner cotyledon margins, and that this pattern is reversed as KTi3 mRNA decays prior to seed dehydration.

KTi1/2 mRNAs Accumulate Later in Embryogenesis and Are Localized Primarily along the Border of Each Cotyledon

Figures 2K to 2N show the in situ hybridization results produced by the KTi1/2 anti-mRNA probe with adjacent sections of the same seeds used to localize KTi3 mRNA (Figures 2F to 2I). No hybridization grains above background were observed within 25 DAF seeds (Figure 2K), in contrast to the results obtained with the KTi3 anti-mRNA probe (Figure 2F). Figure 2L shows that, at 55 DAF, KTi1/2 mRNAs were localized within a thin band of outer cotyledon cells at the hilum end (H) of the seed. As shown in Figure 2M, KTi1/2 mRNAs spread to a wider band of cotyledon cells by 70 DAF, but remained concentrated primarily on the outer margins. Later in maturation, the KTi1/2 mRNAs retreated to a narrow zone of peripheral cells in a ring around each cotyledon (Figure 2N). Analogous to the results obtained with the KTi3 anti-mRNA probe (Figures 2G and 2H), hybridization grains were observed only within storage parenchyma tissue (Figures 2L to 2N).

Grain counts indicated that the hybridization signals obtained with the KTi1/2 anti-mRNA probe were at least fivefold to 10-fold lower than those obtained with the KTi3 anti-mRNA probe in the same outer cotyledon cells throughout maturation (data not shown). We infer from this result, and from those shown in Figure 2, that the KTi1/2 mRNAs have a lower prevalence than the KTi3 mRNA in mid-maturation stage embryo mRNA (Jofuku and Goldberg, 1989) because a smaller proportion of embryo cells contain the KTi1/2 mRNAs, and because there are fewer KTi1/2 mRNA molecules in cells containing both classes of Kunitz trypsin inhibitor mRNAs. Compared with the control hybridization obtained with a KTi1/2 mRNA probe (data not shown, see Figure 2O), no hybridization grains above background were observed at any stage of maturation within non-embryonic seed tissue (Figures 2K to 2N). Together, these results indicate that the KTi1/2 mRNAs appear later in development than the KTi3 mRNA, that the KTi1/2 mRNAs accumulate in an abbreviated wave-like pattern during maturation, and that KTi1/2 mRNAs remain concentrated within parenchyma cells along the outer perimeter of each cotyledon. Combined with the results obtained with younger embryos (Figure 1), these observations indicate that the KTi1/2 and KTi3 Kunitz trypsin inhibitor gene expression programs differ with respect to both timing and cell specificity during embryogenesis.

Kunitz Trypsin Inhibitor mRNAs Accumulate Differentially within the Embryonic Axis

We hybridized the KTi1/2 and KTi3 anti-mRNA probes in situ with mid-maturation stage axis sections to localize the

Figure 2. Kunitz Trypsin Inhibitor mRNA Localization in Maturation Stage Soybean Embryos.

(A) to **(E)** Bright-field photographs of seeds containing maturation stage embryos at 25 DAF **(A)**, 55 DAF **(B)**, 70 DAF **(C)**, 80 DAF **(D)**, and 55 DAF **(E)**. Actual length of the seed shown in **(A)** containing the 25 DAF embryo was 3.1 mm. Magnification factors for the seeds shown in **(B)**, **(C)**, and **(E)** were the same as that shown in **(A)**. The magnification factor for the seed shown in **(D)** was ×0.8 relative to that shown in **(A)**. C, En, H, SC, and V refer to cotyledon, endosperm, hilum, seed coat, and vascular tissue, respectively.

(F) to **(I)** In situ hybridization of a KTi3 anti-mRNA probe with maturation stage embryos at 25 DAF **(F)**, 55 DAF **(G)**, 70 DAF **(H)**, and 80 DAF **(I)**. Photographs were taken with dark-field microscopy. White grains represent regions containing RNA/RNA hybrids. Magnification factors correspond to those used in **(A)** to **(D)**.

(J) In situ hybridization of ^{3}H-poly(U) with a 55 DAF maturation stage embryo. Magnification factor corresponds to that used in **(E)**.

(K) to **(N)** In situ hybridization of a KTi1/2 anti-mRNA probe with maturation stage embryos at 25 DAF **(K)**, 55 DAF **(L)**, 70 DAF **(M)**, and 80 DAF **(N)**. Magnification factors correspond to those used in **(A)** to **(D)**. White regions within the 25 DAF seed **(K)** were identical to those produced with a KTi1/2 mRNA control probe (data not shown).

(O) In situ hybridization of a KTi3 mRNA probe with a maturation stage embryo (80 DAF). White regions represent background hybridization levels as well as dark-field light-scattering through the stained seed coat tissue. A KTi1/2 mRNA control probe produced the same results (data not shown). Magnification factor corresponds to that used in **(D)**.

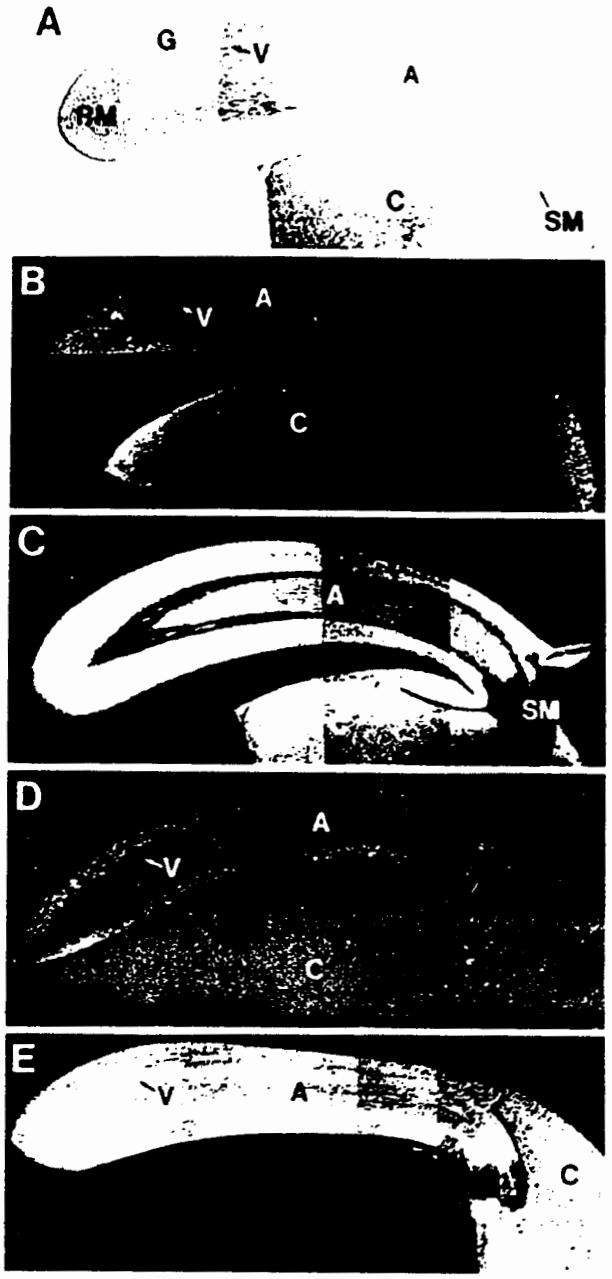


Figure 3. Kunitz Trypsin Inhibitor mRNA Localization in the Soybean Embryonic Axis.

(A) Bright-field photograph of a mid-maturation stage embryo axis (70 DAF). Actual axis length was 4.5 mm. Magnification factors of **(A)** to **(E)** were the same. A, C, G, RM, SM, and V refer to axis, cotyledon, ground meristem, root meristem, shoot meristem, and vascular tissue, respectively.

(B) to **(E)** In situ hybridization of a mid-maturation stage axis (70 DAF) with a KTi1/2 anti-mRNA probe (B), a KTi3 anti-mRNA probe **(C)**, a KTi3 mRNA probe (D), and ^{3}H-poly(U) (E). Photographs were taken by dark-field microscopy.

distribution of Kunitz trypsin inhibitor molecules within this embryonic organ system. Figure 3A shows a bright-field photograph of a longitudinal axis section (A) from a 70 DAF embryo and an adjacent cotyledon (C) region. Analogous to the organization of the cotyledon stage embryo axis (Figure 1D), the vascular tissue (V) is surrounded by ground meristem (G), and these tissues extend throughout the axis from the shoot meristem (SM) to the root meristem (RM).

Figure 3B shows that the KTi1/2 anti-mRNA probe did not produce any detectable hybridization grains above background over the axis, in contrast to the numerous grains observed on the cotyledon periphery of the same embryo (Figures 2M and 3B). RNA gel blot studies did detect the KTi1/2 mRNAs in the axis, but at a concentration 50-fold lower than that observed in cotyledon mRNA (G. de Paiva and R.B. Goldberg, unpublished results)—a level too low to be detected by our in situ hybridization procedure. By contrast, Figure 3C shows that the KTi3 anti-mRNA probe produced an intense hybridization signal throughout the length of the entire axis. This signal was equal in intensity to that produced within the cotyledons (Figures 2H and 3C), and was localized primarily over ground meristem tissue (G) and the root meristematic region (RM).

Figure 3D shows that in situ hybridization with a KTi3 mRNA control probe provided an estimate of background grains within the axis sections. The grain densities did not differ detectably from those obtained with the KTi1/2 anti-mRNA probe (Figure 3B), or from those obtained with the KTi3 anti-mRNA probe over the axis vascular and shoot meristem tissues (Figure 3C). As shown in Figure 3E, in situ hybridization with ^{3}H-poly(U) produced a uniform grain density over the entire axis, indicating that the poly(A) RNA concentrations were approximately the same in different axis regions and in vascular and ground meristem tissues. Together, these results indicate that KTi1/2 and KTi3 mRNAs have different prevalences in the axis, that the KTi3 axis and cotyledon mRNA prevalences are similar, and that KTi3 mRNA is localized primarily within axis ground meristem cells. Because the KTi3 anti-mRNA probe produced similar localization patterns with heart and cotyledon stage embryo axes (Figures 1I and 1J), our results also suggest that the cells containing KTi3 mRNA at the globular stage of development (Figure 1H) mark the lineage that gives rise to axis ground meristem at mid-maturation.

β-Conglycinin Storage Protein mRNA also Accumulates in a Wave-Like Pattern during Embryogenesis

We hybridized a CG-4 β-conglycinin anti-mRNA probe in situ with adjacent sections from the same seeds used for the Kunitz trypsin inhibitor localization studies (Figures 1

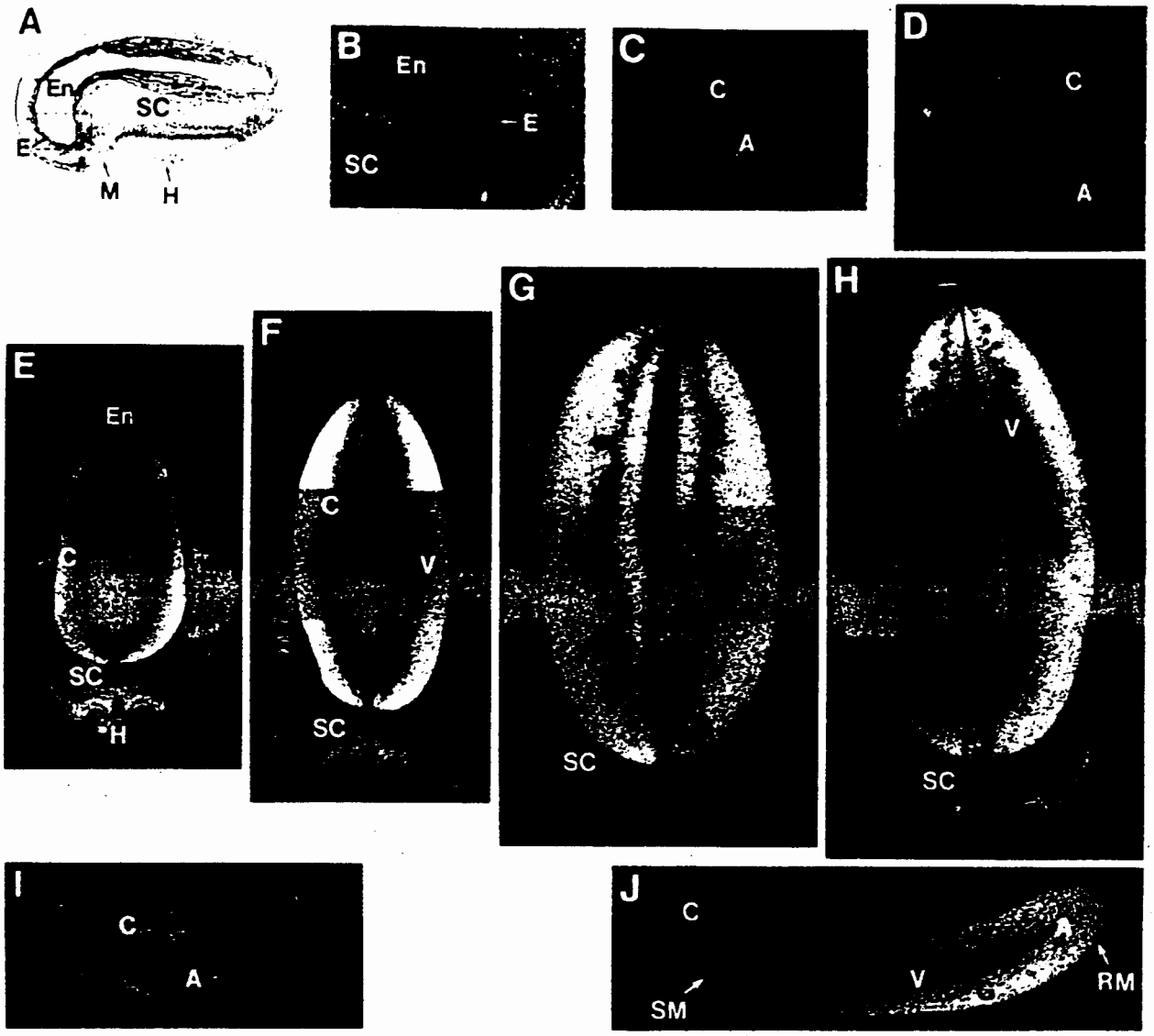


Figure 4. β-Conglycinin mRNA Localization during Soybean Embryogenesis.

(A) Bright-field photograph of a developing soybean seed with a globular stage embryo. E, En, H, M, and SC refer to embryo, endosperm, hilum, micropyle, and seed coat, respectively. Boxed area encloses the globular embryo. Actual seed length was 1.25 mm.

(B) to **(D)** In situ hybridization of a β-conglycinin anti-mRNA probe with a globular stage embryo **(B)**, a heart stage embryo **(C)**, and a cotyledon stage embryo **(D)**. A and C refer to axis and cotyledon, respectively. Photographs were taken by dark-field microscopy. Magnification factors relative to the seed shown in **(A)** were ×4, ×2, and ×2 for the globular, heart, and cotyledon stage embryos, respectively. White grains represent background hybridization levels and were identical in density to those produced with a CG-4 mRNA probe control (data not shown).

(E) to **(H)** In situ hybridization of a β-conglycinin anti-mRNA probe with maturation stage embryos at 25 DAF **(E)**, 55 DAF **(F)**, 70 DAF **(G)**, and 80 DAF **(H)**. V refers to vascular tissue. White areas within embryo represent regions containing RNA/RNA hybrids. White areas on seed coat are due to light-scattering effects of dark-field microscopy. Magnification factors relative to the seed shown in (A) were ×0.5, ×0.5, ×0.5, and ×0.4 for the 25 DAF embryo **(E)**, 55 DAF embryo **(F)**, 70 DAF embryo **(G)**, and 80 DAF embryo **(H)**, respectively.

(I) and **(J)** In situ hybridization of a β-conglycinin anti-mRNA probe with maturation stage embryonic axes at 55 DAF **(I)** and 70 DAF **(J)**. Actual lengths of the 55 DAF and 70 DAF axes were 2.5 mm and 5 mm, respectively. Magnification factors relative to the seed shown in **(A)** were ×0.4 for both **(I)** and **(J)**. White areas show regions containing RNA/RNA hybrids.

to 3) to compare the cellular distribution of different seed protein mRNAs. The CG-4 β-conglycinin gene encodes a 1.7-kb embryo mRNA (Barker et al., 1988; Harada et al., 1989). Under our in situ hybridization conditions (Barker et al., 1988; Cox and Goldberg, 1988), the CG-4 anti-mRNA probe will cross-react with a related 2.5-kb β-conglycinin mRNA, designated as CG-1 (Harada et al., 1989; S.J. Barker and R.B. Goldberg, unpublished results). Thus, in situ hybridization grains will reflect the distribution of both CG-1 and CG-4 β-conglycinin mRNAs, as well as other related β-conglycinin mRNAs (Harada et al., 1989).

Figure 4A shows a bright-field photograph of a longitudinal seed section similar to that shown in Figure 1A. As seen in Figures 4B to 4D, no hybridization grains above background were observed with the CG-4 anti-mRNA probe with seeds containing a globular stage embryo (Figure 4B), a heart stage embryo (Figure 4C), or a cotyledon stage embryo (Figure 4D). These results were similar to those obtained with the KTi1/2 anti-mRNA probe (Figures 1E to 1G). By contrast, Figure 4E shows that the CG-4 anti-mRNA probe hybridized intensely with RNA within the outer edges of 25 DAF embryo cotyledons similar to that produced with the KTi3 anti-mRNA probe (Figure 2F). As seen in Figures 4F and 4G, β-conglycinin mRNA remained localized primarily along the outer margins of 55 DAF embryo cotyledons (Figure 4F), and was distributed over the entire cotyledon at mid-maturation (Figure 4G). These results were similar to the wave-like accumulation pattern observed with KTi3 mRNA (Figures 2F to 2H), but differed from the abbreviated pattern seen with KTi1/2 mRNAs (Figures 2K to 2M). Figures 4F to 4H show that, like the KTi1/2 and KTi3 Kunitz trypsin inhibitor mRNAs, β-conglycinin mRNA was localized primarily within cotyledon storage parenchyma cells and was not detectable within vascular tissue or non-embryonic seed tissue. By contrast with both the KTi1/2 and KTi3 Kunitz trypsin inhibitor mRNAs (Figures 2I and 2N), Figure 4H shows that β-conglycinin mRNA remained uniformly distributed over both cotyledons late in seed maturation.

Figure 4I shows that β-conglycinin mRNA was present in the axis of 55 DAF embryos and was localized specifically within ground meristem tissue, as observed for the KTi3 Kunitz trypsin inhibitor mRNA (Figure 3C). As seen in Figure 4J, a similar β-conglycinin mRNA localization pattern was observed in the axis of 70 DAF embryos. At both developmental stages, hybridization grain densities were similar within the cotyledons and axis (Figures 4I and 4J). Recently, we showed that only the 2.5-kb CG-1 β-conglycinin mRNA is detected in the axis under more stringent in situ hybridization conditions (S.J. Barker and R.B. Goldberg, unpublished results). Thus, the axis hybridization signals observed with the CG-4 anti-mRNA probe probably reflect CG-1 mRNA molecules. Taken together, our findings indicate that β-conglycinin mRNA is not detectable during early embryogenesis, that β-conglycinin mRNA accumulates in a wave-like pattern during maturation, and that β-conglycinin and Kunitz trypsin inhibitor genes are expressed within similar embryo cell types.

Kunitz Trypsin Inhibitor mRNA Localization Patterns Are the Same in Zygotic and Somatic Embryos

We hybridized the KTi3 anti-mRNA probe in situ with somatic embryos to determine whether fertilization events and/or non-embryonic seed tissues were required to induce the mRNA localization patterns observed during seed development (Figures 1 to 3). Figures 5A to 5D show longitudinal sections of somatic embryos that were derived from embryogenic callus at various developmental stages. During the earliest stages of somatic embryogenesis (Figure 5A), embryos (E) are observed "budding" from the callus (CA). Later in development (Figures 5B to 5D), the somatic embryos become polarized and the axis (A) and cotyledons (C) can be visualized. Vascular tissue (V), meristematic tissue (SM and RM), and ground meristem tissue (G) can be distinguished from each other in the more mature somatic embryos (Figures 5B to 5D), and the organization of these tissues is similar to that observed in zygotic embryos (Figures 1 to 3). Somatic embryos similar to the one shown in Figure 5C have the potential to "germinate" and form plantlets in culture (data not shown).

Figures 5E to 5G show that the KTi3 anti-mRNA probe hybridized intensely with somatic embryos at all stages of development. As shown in Figure 5E, hybridization grains were observed within embryogenic callus tissue and within emerging globular- and heart-like embryos. Figures 5F and 5G show that the KTi3 anti-mRNA probe hybridized with equal intensity to RNA present within the axis and cotyledons. Analogous to the results observed with mid-maturation stage zygotic embryos (Figures 2 and 3), KTi3 hybridization grains were only observed over parenchyma and meristematic tissues and were not detected within vascular tissue (Figures 5E to 5G). By contrast, Figures 5H and 5L show that ^{3}H-poly(U) hybridized uniformly with RNA in all axis and cotyledon tissues, and Figures 5I to 5K show that a KTi3 mRNA control probe produced no hybridization grains above background at any stage of somatic embryogenesis. Together, these results indicate that, with respect to timing, organ system, and cell types, the KTi3 mRNA localization pattern is similar in somatic and zygotic embryos, and that this pattern can be established independently of non-embryogenic tissues and fertilization events.

Kunitz Trypsin Inhibitor mRNA Accumulates in a Wave-Like Pattern in Transformed Tobacco Seeds

We showed in the accompanying paper (Jofuku and Goldberg, 1989) that the KTi3 Kunitz trypsin inhibitor gene is expressed in transformed tobacco plants at the correct

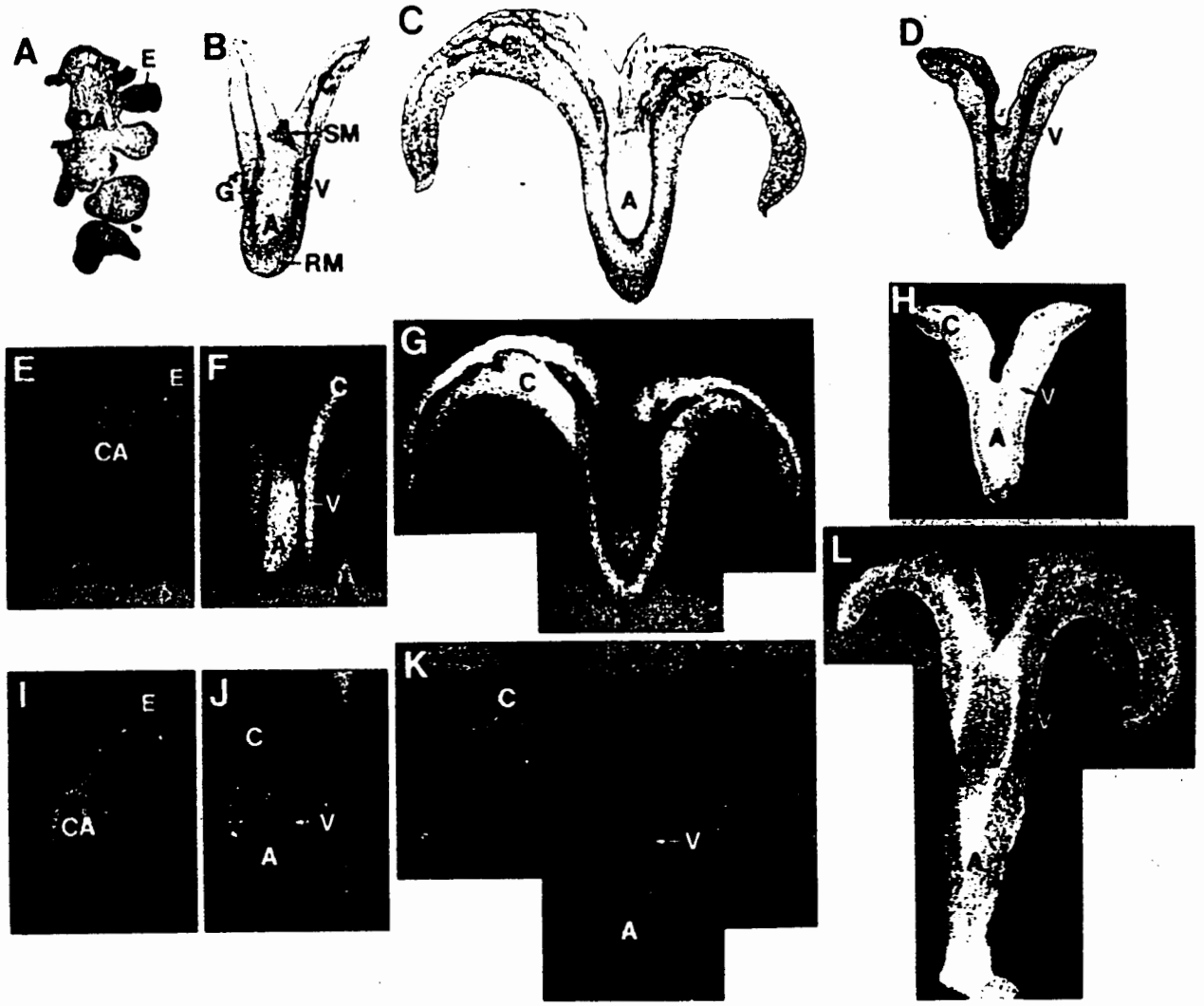


Figure 5. Localization of Kunitz Trypsin Inhibitor mRNA in Soybean Somatic Embryos.

(A) to **(D)** Bright-field photographs of embryogenic callus with a globular-like embryo **(A)**, young somatic embryo with emerging cotyledons **(B)**, mature somatic embryo with fully developed cotyledons and axis **(C)**, and young somatic embryo with emerging cotyledons **(D)**. A, C, CA, E, G, RM, SM, and V refer to axis, cotyledon, callus, embryo, ground meristem, root meristem, shoot meristem, and vascular tissue, respectively. Actual length of the embryo shown in **(D)** was 2.0 mm. Magnification factors relative to the embryo in **(D)** were ×2, ×1, and ×1 for the embryos in **(A)** to **(C)**, respectively.

(E) to **(G)** In situ hybridization of a KTi3 anti-mRNA probe with embryogenic callus containing globular-like embryos **(E)**, a young somatic embryo with emerging cotyledons **(F)**, and a mature somatic embryo with fully developed cotyledons and axis **(G)**. Magnification factors correspond with those shown in **(A)** to **(C)**. White areas represent regions containing RNA/RNA hybrids. Photographs were taken by dark-field microscopy.

(H) In situ hybridization of ^{3}H-poly(U) with a young somatic embryo with emerging cotyledons. Magnification factor corresponds with that used in **(D)**.

(I) to **(K)** In situ hybridization of a KTi3 mRNA probe with an embryogenic callus containing globular- and heart-like embryos **(I)**, a young somatic embryo with emerging cotyledons **(J)**, and a mature somatic embryo with fully developed cotyledons and axis **(K)**. Magnification factors correspond with those shown in **(A)** to **(C)**. White grains show background hybridization levels.

(L) In situ hybridization of ^{3}H-poly(U) with a mature somatic embryo containing fully developed cotyledons and axis. Magnification factor corresponds with that shown in **(D)**.

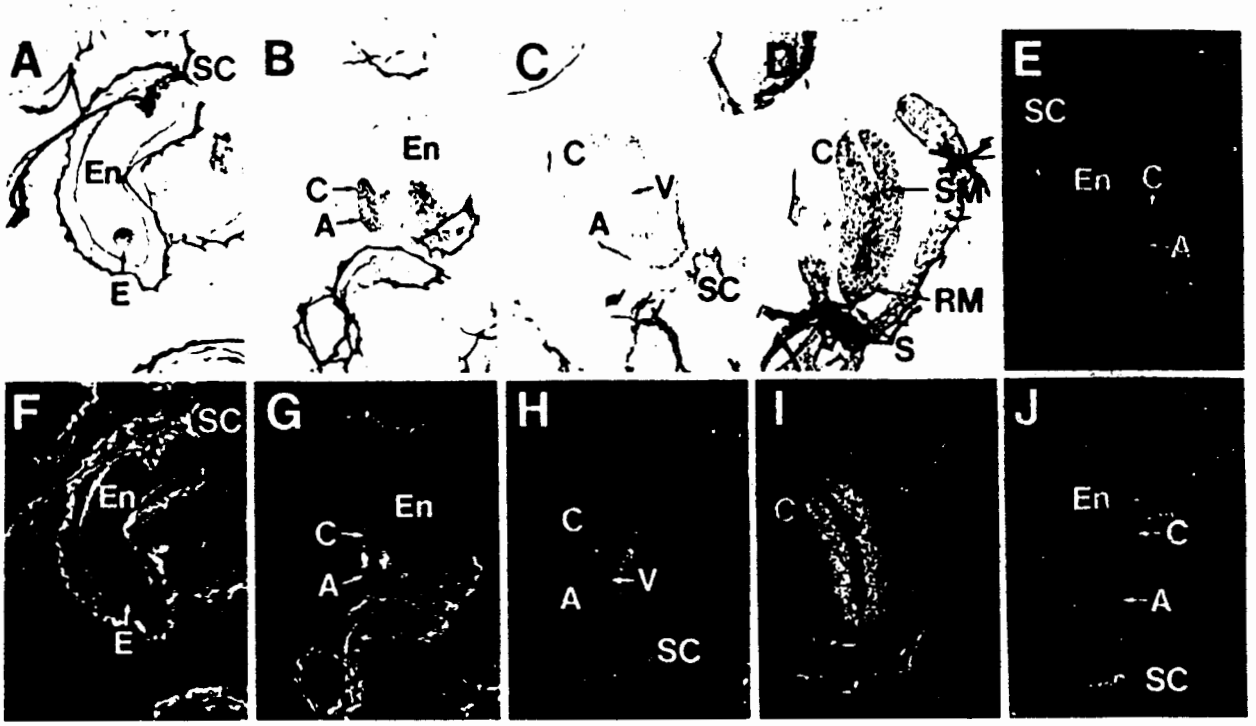


Figure 6. Localization of Kunitz Trypsin Inhibitor mRNA in Transformed Tobacco Seeds.

(A) to **(D)** Bright-field photographs of developing tobacco seeds containing a globular embryo **(A)**, a heart stage embryo **(B)**, and maturation or torpedo stage embryos [**(C)** and **(D)**]. Seeds were harvested from capsules at 9 DAP **(A)**, 16 DAP **(B)**, 18 DAP **(C)**, and 24 DAP **(D)**. Actual seed lengths were 0.75 mm. Magnification factors of **(A)** to **(D)** were the same. A, C, E, En, RM, S, SC, SM, and V refer to axis, cotyledon, embryo, endosperm, root menstem, suspensor, seed coat, shoot meristem, and vascular tissue, respectively. These plants were transformed with the 12.5-kb BglII fragment containing the KTi3 and KTi4 Kunitz trypsin inhibitor genes and were the same transformants analyzed previously (Jofuku and Goldberg, 1989).

(E) In situ hybridization of ^{3}H-poly(U) with a developing seed containing a late heart stage embryo (17 DAP). White areas represent regions containing RNA/RNA hybrids. Photograph was taken by dark-field microscopy. Magnification factor corresponds to that used in (A) tc (D).

(F) to **(I)** In situ hybridization of a KTi3 anti-mRNA probe with developing tobacco seeds containing a 9 DAP globular stage embryo (F), a 16 DAP heart stage embryo **(G)**, an 18 DAP maturation stage embryo **(H)**, and a 24 DAP maturation stage embryo **(I)**. Magnification factors correspond with those used in **(A)** to **(D)**. Hybridization grains shown in **(F)** did not differ in density with those produced with a KTi3 mRNA probe, and represent background hybridization levels.

(J) In situ hybridization of ^{3}H-poly(U) with a developing seed containing an 18 DAP maturation stage embryo. Magnification factor corresponds with that used in **(A)** to **(I)**.

stages of the life cycle. We hybridized the KTi3 anti-mRNA probe in situ with tobacco seed sections containing the KTi3 Kunitz trypsin inhibitor gene (Jofuku and Goldberg, 1989) to localize KTi3 mRNA molecules at different developmental stages. Figures 6A to 6D show longitudinal sections of tobacco seeds at different times after pollination (DAP). Figure 6A shows a globular embryo (E) that is embedded in non-embryonic endosperm tissue (En) and seed coat tissue (SC). As shown in Figure 6B, a young heart stage embryo has become polarized into axis (A) and cotyledon (C) regions. Figures 6C and 6D show that the axis and cotyledons become more prominent in maturation or torpedo stage embryos, the vascular tissue (V) and meristem tissue (SM and RM) can be visualized, and the embryo is still anchored to the seed coat by a small suspensor (S). In contrast with soybean seeds (Figures 1 and 2), endosperm persists throughout tobacco seed development and still surrounds the embryo late in maturation (Figures 6C and 6D).

Figure 6F shows that no detectable hybridization grains above background were observed over any region of a seed containing a globular stage embryo. By contrast, Figure 6G shows that the KTi3 anti-mRNA probe produced an intense hybridization signal on the outer margins of the heart stage embryo, mid-way between the axis and cotyledon termini. As shown in Figures 6H and 6I, hybridization

signals were evenly spread over both the axis and cotyledon regions, and progressed in a wave-like pattern from the outer to inner embryo margins during maturation. Figure 6I shows that by late maturation, KTi3 hybridization grains were present over most embryo cells, including those forming the entire axis root meristem region. Analogous to the results obtained with soybean zygotic embryos (Figures 1 to 3) and somatic embryos (Figure 5), no hybridization grains above background were observed within embryo vascular tissue (Figures 6H and 6I), or within non-embryonic tissues at any developmental stages (Figures 6F to 6I). This contrasts with the results shown in Figures 6E and 6J for ^{3}H-poly(U), in which hybridization grains were uniform over the entire embryo and were visualized within both endosperm and seed coat tissues. Together, these results indicate that the KTi3 gene expression program at the cell level is the same in soybean and transformed tobacco seeds, and that analogous cells along the margins of soybean and tobacco embryos become committed early in development to express seed protein genes.

DISCUSSION

Kunitz Trypsin Inhibitor mRNA Is Localized at One Pole of a Globular Stage Embryo

The molecular processes by which cells become differentiated from each other during plant development are not yet understood. All plant organ systems, including the embryo axis and cotyledons, are organized from three primary tissues—protoderm, ground meristem, and procambium (Steeves and Sussex, 1989). These tissues are formed early in embryogenesis; however, it is not known when their cell lineages are specified (Steeves and Sussex, 1989). Identification of a set of gene markers that are activated at the time of embryo cell specification should, in principle, open the door to unraveling this fundamental differentiation event at the molecular level (Davidson, 1989).

A significant aspect of our results is the observation that the KTi3 Kunitz trypsin inhibitor gene is expressed specifically within a small set of cells at the micropyle pole of the globular stage embryo (Figure 1H). These cells do not contain detectable levels of KTi1/2 Kunitz trypsin inhibitor mRNAs (Figure 1E), β-conglycinin mRNA (Figure 4B), or glycinin and lectin mRNAs (L. Perez-Grau and R.B. Goldberg, unpublished results). Previous studies showed that a globular stage embryo similar to the one shown in Figure 1B contains approximately 500 cells (Goldberg et al., 1981b). Globular embryos at earlier developmental stages also contain KTi3 mRNA molecules in the same micropyle region, but at a reduced level (L. Perez-Grau and R.B. Goldberg, unpublished results). Together, these observations indicate that a specific seed protein mRNA is distributed asymmetrically within one embryonic region only 8 to 9 cell divisions after fertilization has taken place.

At the developmental stage when KTi3 mRNA first appears, most embryo cells are similar to each other and cannot be distinguished using a morphological criterion (Figure 1B). Only protoderm cells along the periphery of the globular embryo can be observed, and the ground meristem and procambium tissues have not differentiated visibly (Figure 1B). It is not yet known how many genes are expressed at this stage of embryogenesis, or the extent to which differential gene activity occurs in various globular embryo cells and regions (Goldberg et al., 1989). However, our observation that KTi3 mRNA molecules accumulate at one pole of the globular embryo indicates that the embryo has already become polarized at the molecular level and that cells have become committed to express different gene sets.

The KTi3 mRNA Serves as a Marker for Axis Cell Specification

The presence of the KTi3 Kunitz trypsin inhibitor mRNA in cells at the micropyle pole of globular stage embryos provides a molecular marker for one of the most important events in plant development—the switch from a spherical globular embryo to a heart stage embryo with bilateral symmetry (Figures 1B and 1C). During this embryonic period, the cotyledons begin to differentiate, the root-shoot axis forms, the plant body visibly becomes polarized, and all three primary tissues become more prominent due to cell division events (Steeves and Sussex, 1989). Figure 1I shows that KTi3 mRNA molecules remain distributed asymmetrically within the embryo at the heart stage and are concentrated specifically within ground meristem tissue of the emerging axis. This localization pattern is maintained in the cotyledon stage embryo axis (Figure 1J) and within the fully differentiated axis of maturation stage embryos (Figure 3C). The persistence of KTi3 in the axis ground meristem through embryogenesis strongly suggests that this tissue is derived clonally from globular embryo cells containing KTi3 mRNA, and that KTi3 mRNA molecules serve as a marker for ground meristem cell specification and embryo polarization early in development. Uncovering the DNA control elements and protein factors responsible for KTi3 gene activation at the globular embryo micropyle pole should provide entry into the regulatory pathway responsible for the fate of one cell type during early embryogenesis.

Cotyledon Cells Are Specified before Seed Protein Genes Are Activated

In contrast with the mRNA localization pattern observed at the micropyle pole, KTi1/2 and KTi3 Kunitz trypsin inhibitor mRNAs (Figures 1E and 1H) and β-conglycinin mRNAs (Figure 4B) are not detected at the endosperm

pole of the globular embryo (Figures 1A and 4A). Nor are other seed protein mRNAs present, such as those encoding glycinin and lectin (L. Perez-Grau and R.B. Goldberg, unpublished results). Seed protein mRNAs are also not detectable in the cotyledons that emerge from the embryo endosperm pole at heart stage (Figures 1F and 1I; Figure 4C; L. Perez-Grau and R.B. Goldberg, unpublished results). Nor are they present within developing cotyledons of a cotyledon stage embryo (Figures 1G and 1J; Figure 4D; L. Perez-Grau and R.B. Goldberg, unpublished results).

Previously, we showed that seed protein gene families are either not detectably transcribed, or are transcribed at a very low level, at the late cotyledon stage of development when the embryo contains at least 100,000 cells (Goldberg et al., 1981b; Walling et al., 1986). At this developmental stage, both cotyledons are well formed and are larger than those present in the cotyledon stage embryo shown in Figure 1D. We conclude from these observations that events responsible for activating seed protein genes within the cotyledon are removed temporally from those leading to the cotyledon cell lineage specification prior to the heart stage of development, and that seed protein genes are transcriptionally activated after a large number of cotyledon cells have formed.

Seed Protein mRNAs Accumulate within the Cotyledons in a Wave-Like Pattern

Figures 2 and 4 show that Kunitz trypsin inhibitor and β-conglycinin mRNAs are first detected in cells along the outer margins of each cotyledon early in maturation. This result has also been observed for both lectin and glycinin mRNAs (L. Perez-Grau and R.B. Goldberg, unpublished results). As maturation stage embryos develop, each seed protein mRNA accumulates progressively in a wave-like pattern from the outer to inner edges of each cotyledon (Figures 2 and 4). Both the timing and extent to which the "wave" occurs is specific for each seed protein mRNA. For example, the KTi1/2 mRNAs accumulate later than the KTi3 and β-conglycinin mRNAs (Figures 2F and 2K; Figure 4E). In addition, the KTi1/2 mRNAs only display an abbreviated wave-like pattern because they remain concentrated along the cotyledon outer borders throughout maturation (Figures 2K to 2N). We conclude from these results that cells become committed to express seed protein genes in an ordered pattern during embryogenesis, that this pattern develops from the outer to inner edges of each cotyledon, and that the cell commitment pattern is slightly different for each seed protein gene.

Soybean seed protein mRNAs also accumulate in a wave-like pattern during the development of transformed tobacco embryos. Figure 6G shows that the KTi3 Kunitz trypsin inhibitor mRNA is localized along the outer edges of a heart stage tobacco embryo and then accumulates progressively to the inner edges as the embryo matures (Figures 6H and 6I). The same pattern is observed for lectin mRNA in independently transformed tobacco plants (L. Perez-Grau and R.B. Goldberg, unpublished results). We infer from these observations that the selective commitment of cells on the outer edge of maturation stage embryos to express seed protein genes is caused by molecular events that are conserved between distantly related plants and probably represent an intrinsic feature of cotyledon development.

Seed Protein mRNA Localization Patterns Are Established in the Absence of Non-Embryonic Tissues

Diffusible substances from seed coat and/or endosperm tissue, or those originally present within the embryo sac prior to fertilization, are probably not responsible for establishing the wave-like seed protein mRNA accumulation pattern because a similar pattern occurs during the development of somatic embryos. Figures 2, 3, and 5 show that the KTi3 Kunitz trypsin inhibitor mRNA is localized within the same cell types, tissues, and regions of soybean somatic and zygotic embryos. A similar somatic and zygotic embryo localization pattern is also observed for lectin mRNA (L. Perez-Grau and R.B. Goldberg, unpublished results). Both KTi3 and lectin mRNAs accumulate during somatic embryogenesis in a wave-like pattern from the outer to inner embryo margins (Figure 5; L. Perez-Grau and R.B. Goldberg, unpublished results). The simplest hypothesis to explain these results is that seed protein mRNAs accumulate progressively from the outer to inner cotyledon borders as a consequence of embryo-specific events that are set into motion at a precise time during embryogenesis.

Seed protein gene transcription rates increase dramatically during maturation and are responsible, to a large extent, for the accumulation of seed protein mRNAs within the embryo (Walling et al., 1986; Goldberg et al., 1989). Although we have no direct proof at the present time, this observation suggests that the wave-like accumulation pattern is due to the progressive transcriptional activation of seed protein genes in cotyledon cells during embryogenesis and reflects the presence of active seed protein gene transcription factors. Whether these putative factors are induced by the transmission of intercellular signals from one cotyledon cell to another, by a gradient of regulatory substances that radiate inward from the outer edge of each cotyledon, or by an intrinsic clock that indicates when cotyledon cells reach the same developmental age is not yet known.

Seed Protein Genes Have Cell-Specific Expression Patterns

Seed protein genes are differentially expressed within specific embryo cells and tissues, irrespective of the mecha-

nisms responsible for establishing seed protein mRNA accumulation patterns. Figures 2 and 4 show that both Kunitz trypsin inhibitor and β-conglycinin mRNAs are localized primarily within mid-maturation stage embryo cotyledon storage parenchyma cells and are not detectable within vascular tissue. Lectin and glycinin mRNAs (L. Perez-Grau and R.B. Goldberg, unpublished results), as well as *Arabidopsis* 12S storage protein mRNA (Pang, Pruitt, and Meyerowitz, 1988), are also present exclusively within cotyledon storage parenchyma cells. Within the axis, seed protein mRNAs are not detectable within vascular tissue and are preferentially concentrated within the ground meristem parenchyma cells (Figures 3C and 4J). By contrast, the relative concentration of total poly(A) RNA molecules appears to be the same within all cotyledon cells and regions (Figure 2J) and within all axis tissues (Figure 3E).

Soybean seed protein mRNAs are present within analogous cell types in transformed tobacco embryos. Figure 6I shows that the KTi3 mRNA is concentrated within parenchyma cells in the tobacco cotyledon and axis regions. β-Conglycinin, lectin, and glycinin mRNAs also have the same distribution patterns in transformed tobacco embryos (Barker et al., 1988; L. Perez-Grau and R.B. Goldberg, unpublished results). We conclude that the nonrandom distribution of seed protein mRNAs within different embryo cell types reflects the differential expression of seed protein genes at the cellular level, and that the mechanisms responsible for seed protein gene cell-specific expression patterns are conserved between distantly related plants.

How Are Seed Protein Gene Expression Patterns Established in Specific Embryo Cell Types?

The differential expression of seed protein genes at the cell level can be explained by a simple model that assumes that the embryo mRNA localization patterns reflect the distribution of factors capable of inducing seed protein gene transcription. These factors would then interact with seed protein gene control elements specific for individual embryo cell types. For example, absence of detectable seed protein mRNAs from embryo vascular tissue implies that factors capable of directing seed protein gene transcription are either inactive or absent from that embryo tissue type. A cascade leading to the segregation of active factors in different cell types might be triggered initially when the protoderm, procambium, and ground meristem cell lineages are specified during early embryogenesis.

A more difficult observation to account for is the expression of seed protein genes within the same cell type but at different times during development. For example, Figures 2 and 4 show that KTi1/2 Kunitz trypsin inhibitor mRNAs accumulate in cells on the outer border of each cotyledon after the KTi3 and β-conglycinin mRNAs are already present. Post-transcriptional processes could be responsible for these temporal differences, as well as other cell-specific seed protein gene expression patterns (Harada et al., 1989). Alternatively, factors capable of inducing the transcription of seed protein genes within specific cell types might be different from those that establish temporal transcription patterns (Davidson, 1989). Clearly, the precise mechanisms by which embryo cells are programmed to express specific seed protein genes remain to be determined.

METHODS

Plant Material

Soybean Dare variety plants were grown in the greenhouse, and seeds were harvested and staged as described previously (Goldberg et al. 1981a, 1981b; Goldberg et al., 1989). Dare somatic embryos (Christianson, Warnick, and Carlson, 1983; Lippmann and Lippmann, 1984) were obtained from Dr. T.S. Rangan (Phytogen Inc., Pasadena, CA). Transformed tobacco plants containing the KTi3 Kunitz trypsin inhibitor gene were described in the accompanying paper (Jofuku and Goldberg, 1989).

Seed and Embryo Preparation for in Situ Hybridization

Soybean seeds containing globular, heart, and cotyledon stage embryos were dissected from developing pods. Seed chalazal ends were sliced off to permit fixative penetration, and then the seeds were fixed with glutaraldehyde as described by Cox and Goldberg (1988). Seeds containing maturation stage embryos were removed from developing pods and staged. These seeds were then sliced horizontally through their centers into 2-mm to 4-mm pieces and fixed with glutaraldehyde. Somatic embryos were fixed in glutaraldehyde without slicing. Tobacco seeds were harvested from developing capsules and staged using several criteria including DAP, color, and dry weight (Barker et al., 1988). Tobacco seeds were fixed with formaldehyde without slicing as described previously (Barker et al., 1988). Fixed seeds and embryos were dehydrated, cleared, embedded in paraffin, and sliced into 10-μm sections as described by Cox and Goldberg (1988).

In Situ Hybridization Probes

Single-stranded ^{35}S-RNA probes were synthesized using the pGEM transcription system (Promega Biotec). The KTi1 probe represented nucleotides +9 to +504 of the KTi1 Kunitz trypsin inhibitor gene (Jofuku and Goldberg, 1989). In this region the KTi1 and KTi2 Kunitz trypsin inhibitor genes are greater than 97% similar (Jofuku and Goldberg, 1989). The KTi3 gene probe represented nucleotides −106 to +718 of the KTi3 Kunitz trypsin inhibitor gene region (Jofuku and Goldberg, 1989; Jofuku et al., 1989). The β-conglycinin probe represented nucleotides +960 to +2427 of the CG-4 β-conglycinin gene (Barker et al., 1988; Harada et al., 1989).

In Situ Hybridization Conditions

In situ hybridization experiments were carried out exactly as described by Cox and Goldberg (1988) and by Barker et al. (1988). In brief, ^{35}S-RNA probes were hydrolyzed to a modal size of approximately 0.2 kb and then hybridized with fixed seed or embryo sections for 14 hr at a 42°C, 0.3 M Na^+, 50% formamide hybridization criterion. Anti-mRNA probes were used to localize mRNA molecules, whereas mRNA probes were used to measure background hybridization. Following hybridization, the seed or embryo sections were incubated with RNase A, and then washed at a 57°C, 0.02 M Na^+ criterion (Barker et al., 1988; Cox and Goldberg, 1988). Slides containing hybridized tissue sections were coated with nuclear track emulsion (Kodak NTB2), exposed for 2 days to 4 days, developed, and then stained with 0.05% toluidine blue. Photographs were taken with Kodacolor VRG100 film using an Olympus AHBT microscope with either bright-field or dark-field illumination. Color prints were produced by Village Photo (Westwood, CA) using a standard automated developing and printing process. In most cases these prints were spliced together to reconstruct the seed and/or embryo visualized in the microscopic field.

ACKNOWLEDGMENTS

We thank Kathleen Cox and Gary Drews for teaching us the in situ hybridization procedure. We also thank Diane Jofuku for the transformed tobacco plants and Kunitz trypsin inhibitor genes, Susan Barker for the CG-4 β-conglycinin gene, and Dr. T.S. Rangan for the soybean somatic embryos. L.P.-G. was funded by a Fulbright Fellowship from the Spanish Ministry of Science and Education. This research was supported by a USDA grant to R.B.G.

Received October 11, 1989.

REFERENCES

Allen, R.D., Bernier, F., Lessard, P.A., and Beachy, R.N. (1989). Nuclear factors interact with a soybean β-conclycinin enhancer. Plant Cell **1**, 623–631.

Barker, S.J., Harada, J.J., and Goldberg, R.B. (1988). Cellular localization of soybean storage protein mRNA in transformed tobacco seeds. Proc. Natl. Acad. Sci. USA **85**, 458–462.

Bustos, M.M., Guiltinan, M.J., Jordano, J., Begum, D., Kalkan, F.A., and Hall, T.C. (1989). Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T-rich, *cis*-acting sequence found upstream of a french bean β-phaseolin gene. Plant Cell **1**, 839–853.

Chen, Z.-L., Schuler, M.A., and Beachy, R.N. (1986). Functional analysis of regulatory elements in a plant embryo specific gene. Proc. Natl. Acad. Sci. USA **83**, 8560–8564.

Chen, Z.-L., Pan, N.-S., and Beachy, R.N. (1988). A DNA sequence element that confers seed specific enhancement to a constitutive promoter. EMBO J. **7**, 297–302.

Chen, Z.-L., Naito, S., Nakamura, I., and Beachy, R.N. (1989). Regulated expression of genes encoding soybean β-conglycinins in transgenic plants. Dev. Genet. **10**, 112–122.

Christianson, M.Z., Warnick, D.A., and Carlson, P.S. (1983). A morphologically competent soybean suspension culture. Science **222**, 632–634.

Cox, K.H., and Goldberg, R.B. (1988). Analysis of plant gene expression. In Plant Molecular Biology: A Practical Approach, C.H. Shaw, ed (Oxford: IRL Press), pp. 1–34.

Davidson, E.H. (1989). Lineage-specific gene expression and the regulative capacities of the sea urchin embryo: A proposed mechanism. Development **105**, 421–445.

Goldberg, R.B., Hoschek, G., Ditta, G.S., and Breidenbach, R.W. (1981a). Developmental regulation of cloned superabundant embryo mRNAs in soybean. Dev. Biol. **83**, 218–231.

Goldberg, R.B., Hoschek, G., Tam, S.H., Ditta, G.S., and Breidenbach, R.W. (1981b). Abundance, diversity, and regulation of mRNA sequence sets in soybean development. Dev. Biol. **83**, 201–217.

Goldberg, R.B., Barker, S.J., and Perez-Grau, L. (1989). Regulation of gene expression during plant embryogenesis. Cell **56**, 149–160.

Harada, J.J., Barker, S.J., and Goldberg, R.B. (1989). Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. Plant Cell **1**, 415–425.

Jofuku, K.D., Okamuro, J.K., and Goldberg, R.B. (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. Nature **328**, 734–737.

Jofuku, K.D., and Goldberg, R.B. (1989). Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. Plant Cell **1**, 1079–1093.

Jofuku, K.D., Schipper, R.D., and Goldberg, R.B. (1989). A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. Plant Cell **1**, 427–435.

Jordano, J., Almoguera, C., and Thomas, T. L. (1989). A sunflower helianthinin gene upstream sequence ensemble contains an enhancer and sites of nuclear protein interaction. Plant Cell **1**, 855–866.

Ladin, B.F., Tierney, M.L., Meinke, D.W., Hosangadi, P., Veith, M., and Beachy, R.N. (1987). Developmental regulation of β-conglycinin in soybean axes and cotyledons. Plant Physiol. **84**, 35–41.

Lippmann, B., and Lippmann, G. (1984). Induction of somatic embryos in cotyledonary tissue of soybean. Plant Cell Rep. **3**, 215–218.

Meinke, D.W., Chen, J., and Beachy, R.N. (1981). Expression of storage protein genes during soybean seed development. Planta **153**, 130–139.

ang, P.A., Pruitt, R.E., and Meyerowitz, E.M. (1988). Molecular cloning, genomic organization, expression, and evolution of 12S seed storage proteins of *Arabidopsis thaliana*. Plant Mol. Biol. 11, 805–820.

Riggs, C.D., Voelker, T.A., and Chrispeels, M.J. (1989). Cotyledon nuclear proteins bind to DNA fragments harboring regulatory elements of phytohemagglutinin genes. Plant Cell **1**, 609–621.

Steeves, T.A., and Sussex, I.M. (1989). Patterns of Plant Development. (Cambridge: Cambridge University Press).

Walling, L., Drews, G.N., and Goldberg, R.B. (1986). Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. Proc. Natl. Acad. Sci. USA. **83**, 2123–2127.

ISBN 0-87969-042-0

Cold Spring Harbor Symposia on Quantitative Biology
Volume XLIII (43) Part 1
DNA Replication and Recombination
1979

Complete Nucleotide Sequence of the *Escherichia coli* Plasmid pBR322

J. G. SUTCLIFFE*

The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

I have determined the 4362-nucleotide-pair sequence of the plasmid cloning vector pBR322 using the DNA-sequencing technique of Maxam and Gilbert (1977). The DNA structure has several interesting features that lead to testable predictions.

pBR322 contains DNA segments assembled in vitro from three naturally occurring plasmids (Bolivar et al. 1977c). The tetracycline resistance (Tc^r) phenotype derives from a fragment of the *Salmonella* plasmid pSC101 (Cohen and Chang 1977). The ampicillin resistance gene (*amp*^{r}) comes from the Tn*3* transposon originally carried by the *S. paratyphi* B plasmid R7268 (later known as R1*drd*19) (Datta and Kontomichalou 1965; Meynell and Datta 1967). The replicator region is derived from pMB1, an *Escherichia coli* plasmid bearing the genes for ColE1, colicin immunity, and the *Eco*RI restriction and modification enzymes (Betlach et al. 1976). Plasmid pMB1 appears to be related to the extensively studied plasmid ColE1; both plasmids replicate in the relaxed mode, require both RNA polymerase and DNA polymerase I for replication, are in the same incompatibility group, and share extensive DNA-sequence homology at the origin of replication. Figure 1 shows the detailed genealogy of pBR322.

Plasmid pBR322 is the most versatile cloning vector developed thus far. It has been approved as an EK-2 vector for *E. coli* cloning projects. There are seven known enzymes which cleave the pBR322 circle, each at a single site (Bolivar et al. 1977c, and this paper). Although all of these sites can be used for cloning foreign DNA fragments (inserts), three sites are especially valuable because their use inactivates genes. Inserts in the *Bam*HI and *Sal*I sites destroy the Tc^r phenotype normally conferred by the plasmid, but not ampicillin resistance. Such *amp*^{r}*tet*^{s} bacteria can be selected for by cycloserine enrichment (Bolivar et al. 1977a; Hamer and Thomas 1976). Similarly, inserts at the *Pst*I site often destroy ampicillin resistance, leaving the bacterium *amp*^{s}*tet*^{r}. These can be detected by replica plating. The *Pst*I site is particularly useful because its 3′ tetranucleotide extensions are ideal substrates for terminal transferase; hence, this site is excellent for cloning by the homopolymer tailing method (Lobban and Kaiser 1973). When G–C tailing is used, the *Pst*I restriction site is regenerated and the insert may be cut out with that enzyme.

Knowledge of the entire nucleotide sequence greatly enhances the value of pBR322 as a cloning vector, since, with the sequence in hand, it is relatively simple to map any new DNA fragments created by cloning. In addition, the plasmid contains three regions of interesting biological function (the Ap^r and Tc^r determinants and an origin of replication) which the DNA sequence will help to explain. The *amp*^{r} gene codes for a β-lactamase whose kinetic properties (Hall and Knowles 1976) and crystal structure (Knox et al. 1976) had been studied, but whose primary amino acid sequence had not been solved. That sequence is medically important for the design of better penicillins and can be deduced from the DNA sequence. The plasmid has a replicator region similar to ColE1. A comparison of these two origins should help to sort out which parts of the DNA are most necessary for plasmid replication. Very little is understood about the molecular features of tetracycline resistance and the primary sequence should be particularly helpful in this regard. Finally, the plasmid can serve as a rich source of DNA fragments of a variety of known sequences and lengths, and these should be useful in many sorts of biochemical investigations.

EXPERIMENTAL PROCEDURES

pBR322. Plasmid pBR322 was obtained from H. Boyer in strain RR1. It has been renamed pBR322GS to indicate that it is directly related to the sequenced DNA.

DNA sequencing. DNA-sequencing reactions were carried out by the procedure of Maxam and Gilbert with one exception: magnesium acetate was omitted from the hydrazine reaction stop solutions. This omission (A. Maxam, pers. comm.) eliminated the formation of a hydrazine-magnesium pellet, which was frequently noticed in the pyrimidine-specific reactions. The pellet often caused smeary electrophoresis lanes. In addition, partial solubilization of the hydrazine during the strand-scission step was consistently associated with incomplete cleavage. Incorporating this alteration into the published protocol made it possible

*Present address: Department of Cellular and Developmental Immunology, Scripps Clinic and Research Foundation, 10666 North Torrey Pines Road, La Jolla, California 92037.

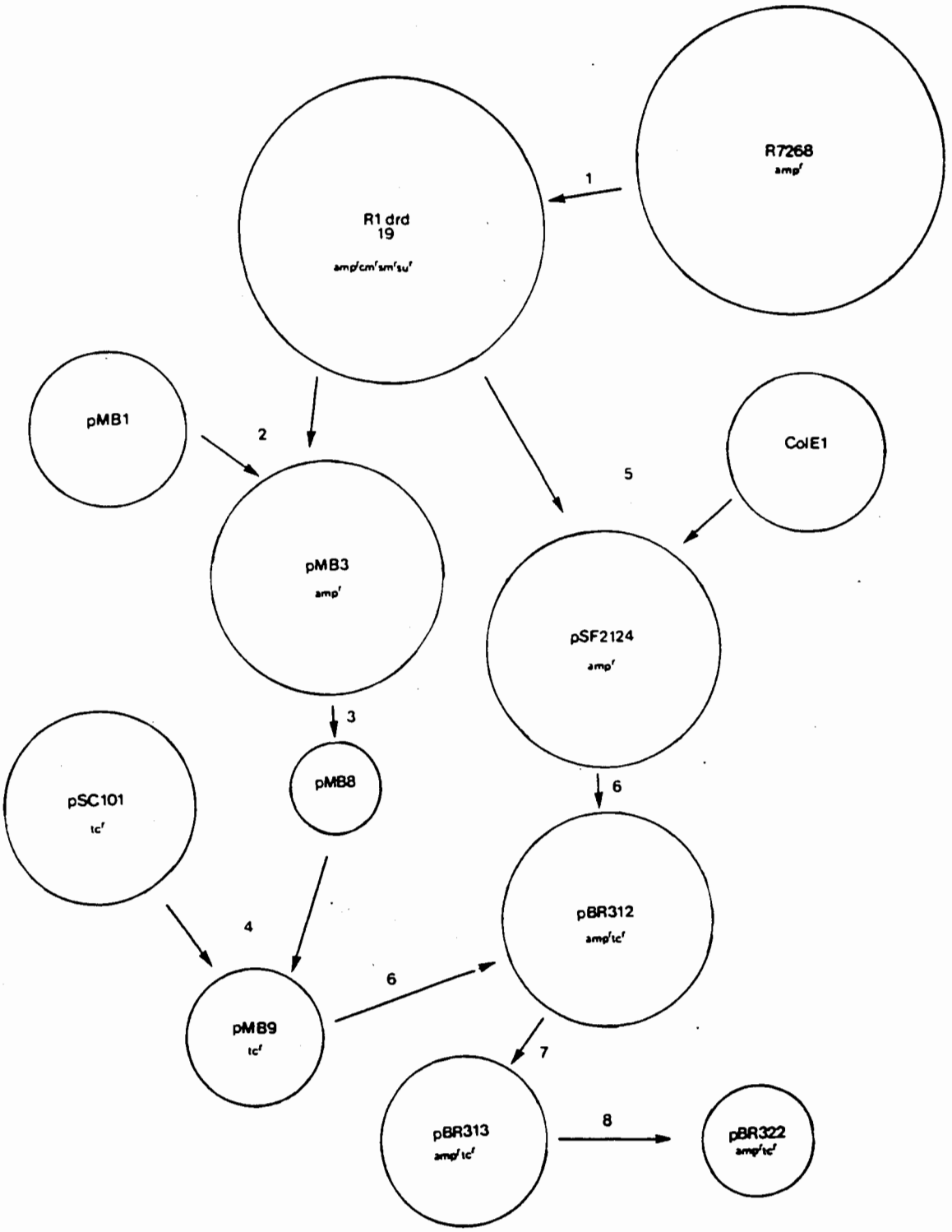


Figure 1. The ascent of pBR322. R7268 was isolated in London in 1963 (Datta and Kontomichalou 1965) and later renamed R1. (*1*) A variant, R1*drd*19, which was derepressed for mating transfer, was isolated (Meynell and Datta 1967). (*2*) The Ap transposon, Tn*3*, from this plasmid was transposed onto pMB1 to form pMB3 (Betlach et al. 1976). (*3*) This plasmid was reduced in size by *Eco*RI* rearrangement to form a tiny plasmid pMB8 (Rodriguez et al. 1976) which carries only colicin immunity. (*4*) *Eco*RI* fragments from pSC101 (Cohen and Chang 1977) were combined with pMB8 opened at its unique *Eco*RI site and the resulting chimeric molecule rearranged by *Eco*RI* activity to generate pMB9, a vector that has seen much service in the clone wars (Boyer et al. 1977). (*5*) In a separate event, the Tn*3* of R1*drd*19 was hopped to ColE1 to form pSF2124 (So et al. 1976). (*6*) The Tn*3* element was then transposed to pMB9 to form pBR312. (*7*) EcoRI* rearrangement of pBR312 led to the formation of pBR313 (Bolivar et al. 1977a), from which (*8*) two separate fragments were isolated and ligated together to form pBR322 (Bolivar et al. 1977c). During this series of constructions, R1 and ColE1 served only as carriers for Tn*3*. Material originally in pMB1 and pSC101 ended up in pBR322.

to read consistently sequence ladders substantially farther from the 5′ end. Although this explanation has never been demonstrated explicitly, I have a large number of examples of sequences run with and without magnesium and totally subscribe to this alteration.

For most of the sequences that I determined, DNA fragments labeled on both 5′ ends were subcut with a second restriction enzyme and the reaction products separated on preparative gels. This was usually a successful approach. To choose an enzyme appropriate for a given labeled fragment, several aliquots of fragment (amounting to a small percentage of the total labeled DNA), each with a different candidate enzyme, were digested and that enzyme which gave the largest and most well-resolved subfragments was selected. Strand separation seldom worked, but subcutting was quite reliable. New strand-separation technology may make that method of DNA preparation equally reliable.

RESULTS AND DISCUSSION

Strategy

The strategy used to sequence the plasmid was rather straightforward. Initially, I intended to sequence only the *amp′* gene and focused on that region. When I realized it was feasible to do the entire plasmid, I adopted the tactic of sequencing every obtainable 5′ end produced by those restriction enzymes that were accessible and that gave a pattern of well-resolved restriction bands. I thereby sequenced every 5′ end produced by *Hin*fI and almost all ends produced by *Hae*III, *Alu*I, and *Hpa*II. Some longer fragments required the use of other enzymes to produce a 5′ end within a particular stretch of DNA. It was necessary to use *Taq*I and *Mbo*I and two other enzymes, *Tha*I and *Ava*II, whose cutting sites had not been previously determined to sequence the entire 4362 base pairs (bp). The fragments from which the sequence is read are shown in Figure 2. It was possible to sequence the entire 4362 bp by relying on only eight enzymes to produce 5′ ends. (*Eco*RI, *Bam*HI, *Hin*dII, and *Hin*dIII were also used, but they need not have been.)

The gel patterns from which the DNA sequences are read are generally rather clear. Chemical artifacts or contaminating sequences in the background occasionally obscure the sequence pattern on an autoradiograph. When such problems occur, the entire run is usually lost as far as obtaining reliable sequence data, but these problems do not occur frequently enough to pose a substantial obstacle.

An artifact that cannot be circumvented directly is caused by intrastrand pairing during electrophoresis. The result is a compression of the DNA banding pattern such that the bands often superimpose upon each other, rendering that particular region of the sequence film unreadable. A solution for this aberrancy is to sequence the other DNA strand. If the same structural aberration appears it will be located at a different region of the DNA sequence (where the complementary sequence appears).

Intrastrand base pairing can occur when a sequence and its complement appear on the same DNA strand. Practically speaking, these electrophoresis problems were observed only when a GC-rich stem of four or more pairs could form. I ran sequencing gels very warm (1300 V on a 40-cm gel). P. Farabaugh (pers. comm.) observed stems with more A-T pairs when running at lower temperature. Higher running temperature reduces the stability of the stem interaction such that only stronger stems are observed. I observed eight compressions during the sequencing of pBR322.

Since the usefulness of a sequence depends upon its degree of accuracy, I maximized accuracy by taking several precautions. First, as can be seen in Figure 2, all parts of pBR322 were sequenced at least twice and about 75% of it on both strands to avoid strand-specific artifacts. Second, I sequenced across all restriction junctions (Fig. 2). This eliminated the possibility of omitting a small DNA fragment whose omission would result in a deletion in the sequence. Had I relied on restriction mapping to order fragments and depended on such an order to avoid sequencing across all restriction cuts, I could have had as many as five gaps in the final sequence.

Third, all sequence data were read on three separate occasions to eliminate errors in transcription. The freshly read sequences were noted separately and those determinations were compared with the master sequence. Discrepancies between separate readings were resolved by going back to the films. Such discrepancies (which were not infrequent) were of two sorts, misreading a film or unfaithful copying of the result. Both sorts were of a clerical nature, since film-reading errors were usually corrected by a subsequent check of the film. Occasionally, a region of a sequence was difficult to read on one film, but a film covering the complementary region on the other strand always led to unambiguous resolution of the sequence. Fourth, the data were not fit to a preconceived idea of what the sequence should be. That is, the final sequence was obtained for each region before it was compared with preexisting data. Finally, the printed sequence was checked extensively against the handwritten master sequence with the help of many people. All of these precautions were necessary because of the sheer volume of the data that had to be extracted. Chemical problems were minor compared with the clerical problems of handling the data.

The entire pBR322 DNA sequence is presented in Figure 3. A complete set of restriction maps along with a tabulation of fragment sizes for use as DNA size markers and for characterizing pBR322 inserts have been published elsewhere (Sutcliffe 1978b).

During the course of solving the pBR322 sequence, gel-separation technology improved so that longer runs (occasionally exceeding 300 nucleotides) could be read from a single labeled end. Variations in gel

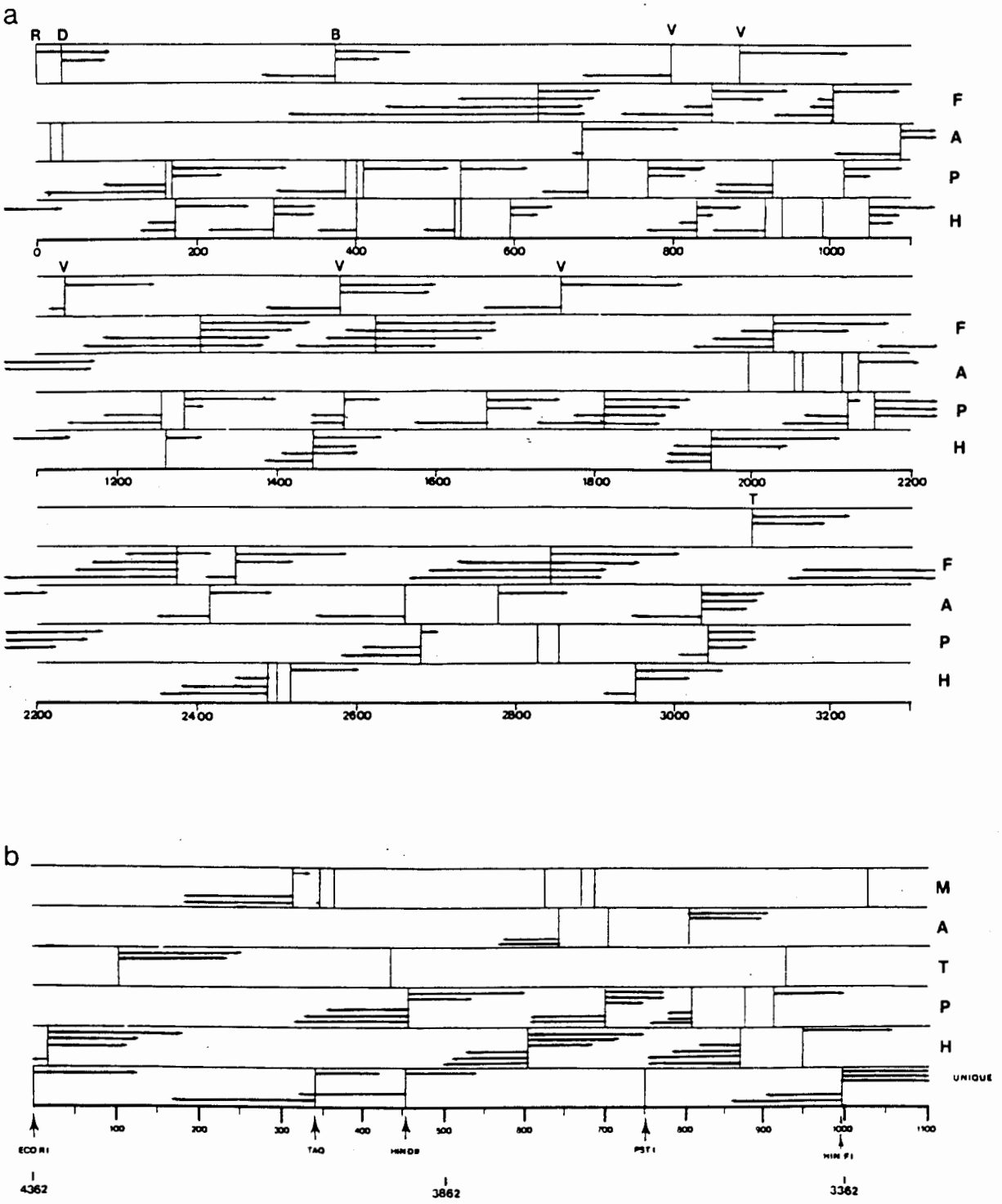


Figure 2. The overlapping sequences determined for pBR322. (*a*) Schematic restriction map of pBR322 for the enzymes *Hae*III(*H*), *Hpa*II(*P*), *Alu*I(*A*), and *Hin*fI(*F*). The sites in the top strip are the unique sites for *Eco*RI(*R*), *Bam*HI(*B*), and *Hin*dIII(*D*) and a few of the sites for *Ava*II(*V*) and *Tha*I(*T*). The map is numbered in base pairs. The arrows within the strip for each different enzyme show the extent of individual sequencing runs. The tail of the arrow is at the 5′ end of the sequenced fragment. It should be noted that all sequences were determined at least twice and all restriction sites were overlapped. (*b*) Similar data for the β-lactamase region (3259–4362). This figure is numbered in the counterclockwise, as well as the clockwise, direction (0–1100 is counterclockwise). *Mbo*I (*M*). (Reprinted, with permission, from Sutcliffe 1978a.)

AA)TTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAAT
AAGAGTACAAACTGTCGAATAGTAGCTATTCGAAATTACGCCATCAAATAGTGTCAATTTAACGATTGCGTCAGTCCGTGGCACATACTTTAGATTGTTA

GCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCG
CGCGAGTAGCAGTAGGAGCCGTGGCAGTGGGACCTACGACATCCGTATCCGAACCAATACGGCCATGACGGCCCGGAGAACGCCCTATAGCAGGTAAGGC

ACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCG
TGTCGTAGCGGTCAGTGATACCGCACGACGATCGCGATATACGCAACTACGTTAAAGATACGCGTGGGCAAGAGCCTCGTGACAGGCTGGCGAAACCGGC

CCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTG
GGCGGGTCAGGACGAGCGAAGCGATGAACCTCGGTGATAGCTGATGCGCTAGTACCGCTGGTGTGGGCAGGACACCTAGGAGATGCGGCCTGCGTAGCAC

GCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTT
CGGCCGTAGTGGCCGCGGTGTCCACGCCAACGACCGCGGATATAGCGGCTGTAGTGGCTACCCCTTCTAGCCCGAGCGGTGAAGCCCGAGTACTCGCGAA

GTTTCGGCGTGGGTATGGTGGCAGGCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTC
CAAAGCCGCACCCATACCACCGTCCGGGCACCGGCCCCCTGACAACCCGCGGTAGAGGAACGTACGTGGTAAGGAACGCCGCCGCCACGAGTTGCCGGAG

AACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGG
TTGGATGATGACCCGACGAAGGATTACGTCCTCAGCGTATTCCCTCTCGCAGCTGGCTACGGGAACTCTCGGAAGTTGGGTCAGTCGAGGAAGGCCACCC

CGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGA
GCGCCCCGTACTGATAGCAGCGGCGTGAATACTGACAGAAGAAATAGTACGTTGAGCATCCTGTCCACGGCCGTCGCGAGACCCAGTAAAAGCCGCTCCT

CCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAA
GGCGAAAGCGACCTCGCGCTGCTACTAGCCGGACAGCGAACGCCATAAGCCTTAGAACGTGCGGGAGCGAGTTCGGAAGCAGTGACCAGGGCGGTGGTTT

CGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCA
GCAAAGCCGCTCTTCGTCCGGTAATAGCGGCCGTACCGCCGGCTGCGCGACCCGATGCAGAACGACCGCAAGCGCTGCGCTCCGACCTACCGGAAGGGGT

TTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATC
AATACTAAGAAGAGCGAAGGCCGCCGTAGCCCTACGGGCGCAACGTCCGGTACGACAGGTCCGTCCATCTACTGCTGGTAGTCCCTGTCGAAGTTCCTAG

GCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATT
CGAGCGCCGAGAATGGTCGGATTGAAGCTAGTGACCTGGCGACTAGCAGTGCCGCTAAATACGGCGGAGCCGCTCGTGTACCTTGCCCAACCGTACCTAA

GTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAA
CATCCGCGGCGGGATATGGAACAGACGGAGGGGCGCAACGCAGCGCCACGTACCTCGGCCCGGTGGAGCTGGACTTACCTTCGGCCGCCGTGGAGCGATT

CGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATC
GCCTAAGTGGTGAGGTTCTTAACCTCGGTTAGTTAAGAACGCCTCTTGACACTTACGCGTTTGGTTGGGAACCGTCTTGTATAGGTAGCGCAGGCGGTAG

TCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGG
AGGTCGTCGGCGTGCGCCGCGTAGAGCCCGTCGCAACCCAGGACCGGTGCCCACGCGTACTAGCACGAGGACAGCAACTCCTGGGCCGATCCGACCGCCC

GTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGT
CAACGGAATGACCAATCGTCTTACTTAGTGGCTATGCGCTCGCTTGCACTTCGCTGACGACGACGTTTTGCAGACGCTGGACTCGTTGTTGTACTTACCA

CTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGT
GAAGCCAAAGGCACAAAGCATTTCAGACCTTTGCGCCTTCAGTCGCGGGACGTGGTAATACAAGGCCTAGACGTAGCGTCCTACGACGACCGATGGGACA

GGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAAC
CCTTGTGGATGTAGACATAATTGCTTCGCGACCGTAACTGGGACTCACTAAAAAGAGACCAGGGCGGCGTAGGTATGGCGGTCAACAAATGGGAGTGTTG

Figure 3. The pBR322 DNA sequence. The sequence is presented in double-strand form. The top strand is 5′→3′, the lower strand is complementary 3′→5′. The circular nature of the sequence is indicated by the nucleotides in parentheses at the ends. The sequence is numbered such that 0 is the middle of the unique *Eco*RI site and the count increases first through the *tet* genes, then pMB1 material, and finally through the Tn*3* region. This is clockwise numbering on the map in Fig. 4.

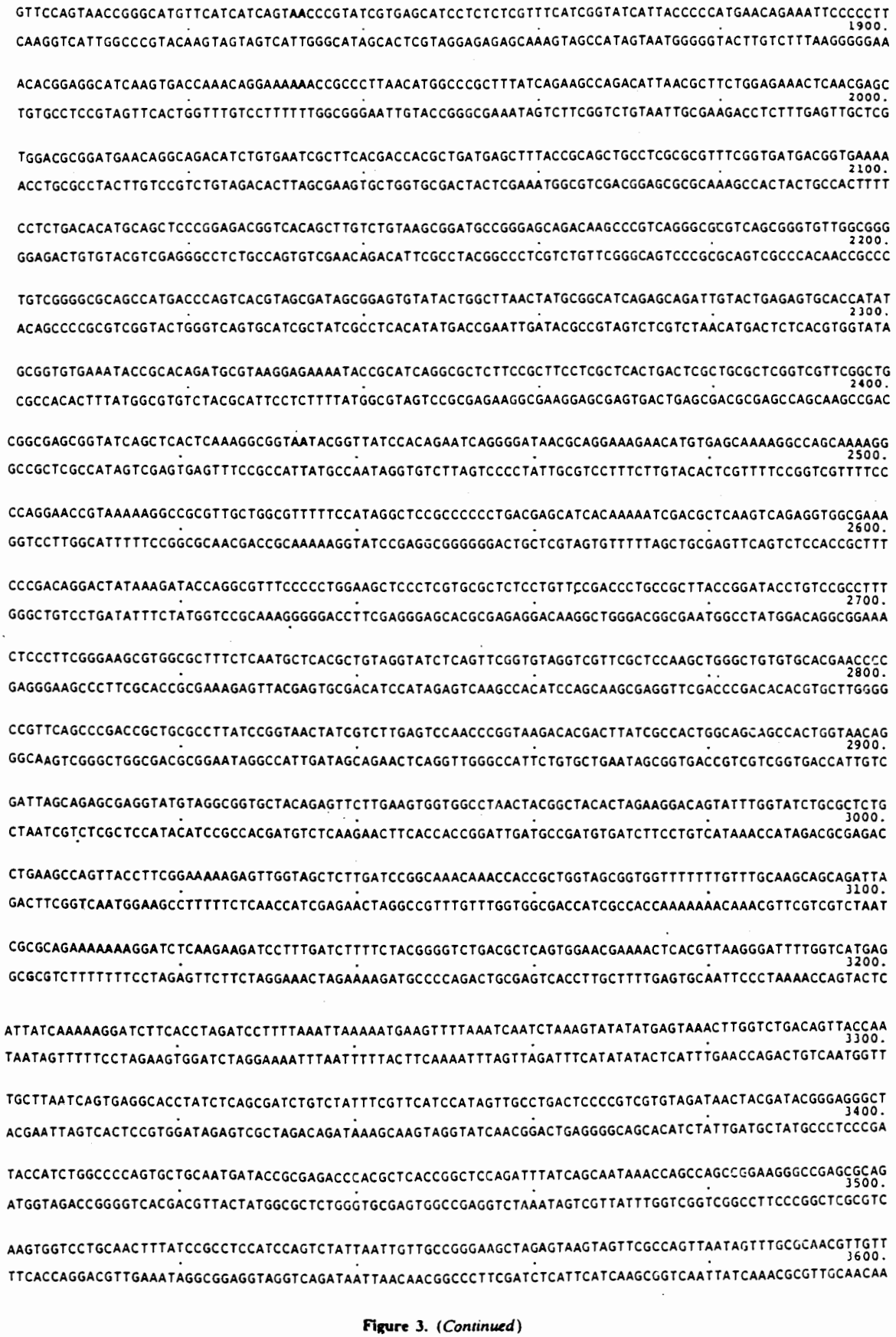


Figure 3. (*Continued*)

```
GCCATTGCTGCAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGT
                                                                                                  3700.
CGGTAACGACGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACA

TGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTC
                                                                                                  3800.
ACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAG

TCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC
                                                                                                  3900.
AGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACG

CCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
                                                                                                  4000.
GGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAATG

CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG
                                                                                                  4100.
GCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTC

GCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
                                                                                                  4200.
CGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACA

CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCA
                                                                                                  4300.
GAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGT

TTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAA(TTC)
AATAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGAAGTTCTT
```

Figure 3. (*Continued*)

thickness, length, running temperature, and acrylamide concentration all affect the amount of information that can be extracted from a single run (Sanger and Coulson 1978; W. Gilbert and R. Tizard, pers. comm.; A Maxam, in prep.; and this author). In particular, the best resolution is obtained on 1-m-long, 0.4 mm-thick gels which are run quite warm to the touch. Sequences away from the labeled end are better resolved on lower-percentage acrylamide gels. With multiple sample loadings, the 20% gels described in the original Maxam and Gilbert presentation of this technique are totally adequate for the first 100 bases from the labeled end and often further (on several occasions. I have read further than 170 bases from such gels). A more modern approach would utilize separate gels of different resolution capabilities, rather than multiple loadings on the same gel. Thus, at this time, repeating the project would require fewer sequencing runs simply because of improved gel technology.

It was undoubtedly easier and quicker to sequence many fragments almost indiscriminately than to spend time determining a restriction map which could eventually be read directly from the sequence. I developed this attitude while the project was in motion. Solving a sequence of similar size would now require less time than the 13 months spent on this project. Three of the arrows presented in Figure 2 came from one 1-m gel which was run by R. Tizard; the rest of the data were generated by myself. Analysis of the films and proofreading consumed a substantial portion of the time.

amp^r Gene

A detailed description of the *amp*^{r} gene was published earlier (Sutcliffe 1978a) but several points bear repeating in relation to the total nucleotide sequence.

1. The β-lactamase gene covers the unique *Pst*I site located 3611 bp clockwise from the unique *Eco*RI site (see Fig. 4).
2. The direction of translation of β-lactamase is counterclockwise and starts about 200 bp from the *Eco*RI site. The ribosome-binding site is shown in Figure 5.
3. The gene sequence indicates that β-lactamase is synthesized with a hydrophobic secretion signal of 23 amino acids at its amino-terminal end. The signal does not appear in the 263-amino-acid mature β-lactamase isolated from the cell periplasmic space (Ambler and Scott 1978; Sutcliffe 1978a). The protein sequence, as deduced from the DNA sequence, is shown in Figure 6.
4. The accuracy of the sequence across the gene is largely verified by partial protein sequencing (Ambler and Scott 1978).
5. A glutamine-to-lysine change near the amino-terminal end of the protein does not seem to alter the kinetic properties of β-lactamase.
6. When rat proinsulin mRNA reverse transcript was inserted at the *Pst*I site using homopolymer tailing and *E. coli* 1776 was transformed with this DNA, insulin antigen was found in the periplasmic space covalently fused to truncated β-lactamase antigen.

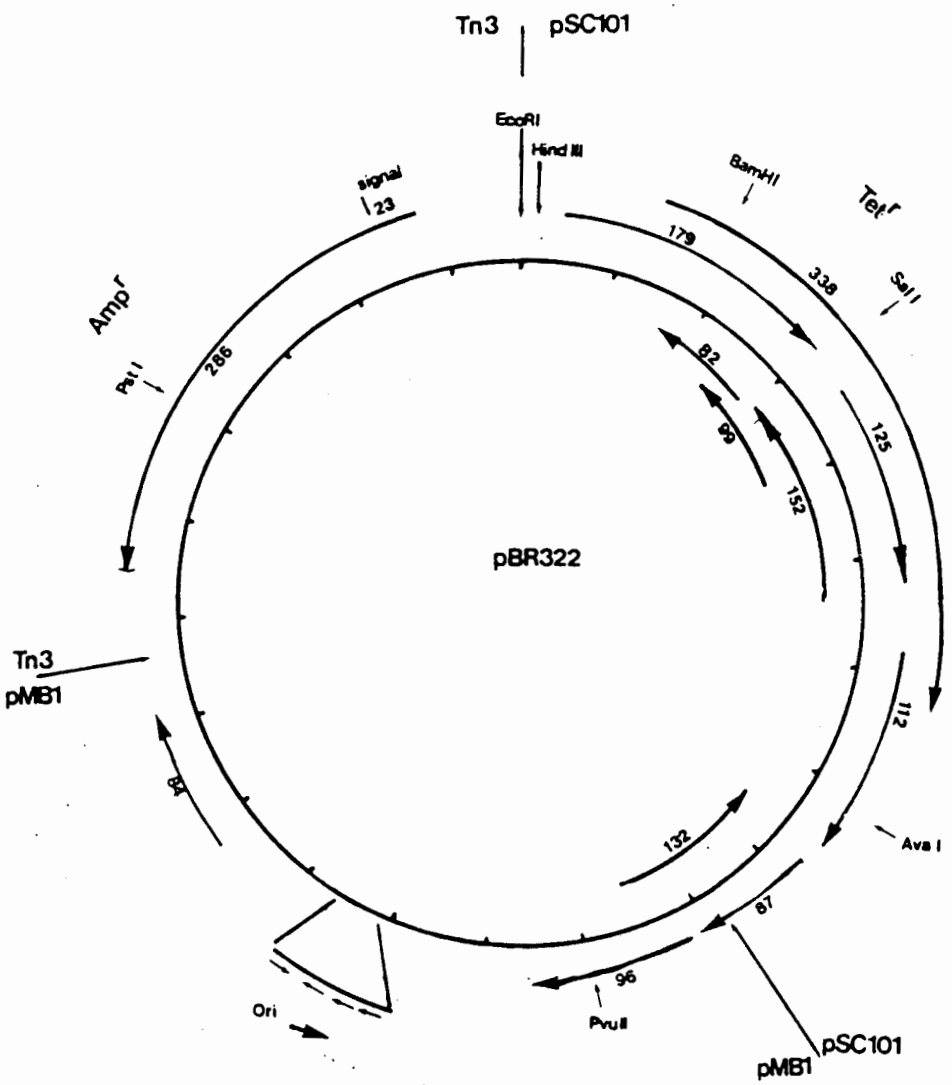


Figure 4. Biological map of pBR322. The boundaries of the DNA derived from pSC101, pMB1, and Tn*3* are indicated at the perimeter. Unique sites for the enzymes *Eco*RI, *Hin*dIII, *Bam*HI, *Sal*I, *Ava*I, *Pvu*II, and *Pst*I are indicated. Several proteins that could be encoded by the DNA sequence are indicated by the bold arrows inside and outside of the perimeter. The length in amino acids of each peptide is indicated on the arrow, whose tail is at the amino-terminal end. The 23-amino-acid-long signal for β-lactamase is indicated on the protein in the *amp*^{r} region. The origin of replication is indicated as a blowup and is presented in more detail in Fig. 7. The exact positions of the starts for the "genes" and the origin are found in the legends to Figs. 5 and 7. The pSC101/pMB1 junction is at 1762, and the pMB1/Tn*3* switch occurs at 3146.

Furthermore, many such fusion strains express low-level ampicillin resistance (Villa-Komaroff et al. 1978).

Tn*3* Inverted Repeat

The Ap^r transposon used in the construction of pBR322 originated from R1*drd*19 (So et al. 1976; Bolivar et al. 1977a). E. Ohtsubo et al. (pers. comm.) have completed the sequence of the inverted repeats of this transposon. A comparison between pBR322 and the Tn*3* repeat region shows total agreement between the R1*drd*19-derived Tn*3* "left-end" repeat and the pBR322 sequence from 3146 to 3448 (within the carboxyterminal end of the β-lactamase gene). This concurrence shows that this region has not been disrupted by the manipulations used to construct pBR322. Furthermore, the 5 bp immediately preceding the Tn*3* inverted repeat are not like any of those that have been found by Ohtsubo et al. (this volume) to be duplicated on the opposite ends of Tn*3*. This fact is consistent with their notion that, at least for Tn*3*, there is no site specificity for the integration of the transposon. This comparison also defines the extent of Tn*3* material on pBR322. The other end is slightly indefinite (since no sequence data from pSC101 are available) but it is somewhere close to the *Eco*RI site within the 200 bp between the *amp*^{r} and *tet*^{r} promoters.

```
                TAAGGAGGTGATC    SHINE-DALGARNO
A-179   AACGCAGTCAGGCACCGTGTATG  TAA (87)
A-87    TCACCGATACGCGAGCGAACGTG  TAA (1548)
A-96    CAACGTTCCAGTAACCCGGCATG  TGA (1818)
B-125   AACCCAGTCAGCTCCTTCCGGTG  TGA (697)
B-112   TATGCCGCCTCGGCGAGCACATG  TAG (1180)
C-338   TATGCGTTGATGCAATTTCTATG  TGA (260)
C-84    TTCGCTCCAAGCTGGGCTGTGTG  TGA (2789)
D-132   TTTTCCTGTTTGGTCACTTGATG  TAA (1912)
D-82    GAGCACCGCCGCCGCAAGGAATG  TAG (571)
E-286   AATATTGAAAAAGGAAGAGTATG  TAA (4153)
E-152   GATCCTTGAAGCTGTCCCTGATG  TAG (1080)
F-99    GAAAGCGGTCCTCGCCGAAAATG  TGA (788)
```

Figure 5. The hypothetical genes of pBR322. The 20 bases immediately preceding the initiator (ATG or GTG) of each of the "genes" discussed in the text are presented. The entry at the left (a letter "A" through "F" and a number) carries two pieces of information. The letter represents from which of the six possible reading frames (arbitrarily chosen A–C clockwise, D–F counterclockwise) the protein is translated. The number is the length in amino acids. The letters in the third column represent the nonsense triplet which terminates the protein. The number in parentheses corresponds to the position in the DNA sequence of the first base (A or G) in the initiator methionine codon. Above the sequence is the complement of the 16S RNA 3′ end (Shine and Dalgarno 1974; Steitz and Jakes 1975). Homologous runs of 3 or more bases in those which precede the initiator are underlined, se representing possible ribosome-binding sites. The β-lac- .nase (E-286) and D-82 have particularly good matches.

Origin of Replication

The ColE1 and pBR345 (also derived from pMB1) origin sequences have been worked out (Tomizawa et al. 1977; Bolivar et al. 1977b). Like pBR345, pBR322 is, except for 1 bp, homologous to ColE1 for 190 nucleotides upstream from the RNA/DNA junction observed to be the changeover from RNA primer to DNA chain when ColE1 is replicated in vitro (Tomizawa et al. 1977). Downstream from the replication start there is general, but not precise, homology with ColE1, and, except for an 11-bp-long *Hae*III fragment which was not reported for pBR345, the two pMB1-derived origins are the same. The sequence at the origin of replication is presented in Figure 7.

An interesting symmetry feature of the pBR322 origin sequence includes the 11-bp-long *Hae*III fragment which is part of a tandem repeat that is repeated again, largely conserved, once in the same orientation and once inverted (Figs. 4 and 7). The inverted sequence occurs exactly at the RNA/DNA junction. Figure 7 shows the striking degree of homology among these repeats. Although there is no evidence for the actual existence of secondary structure, stem-and-loop structures can be drawn. The repeats are not required for replication because Backman et al. (this volume) ve shown that plasmids can still replicate after this region has been deleted. Figure 8 shows a gel pattern which includes the 11-bp.

The extent of ColE1-like sequence in pBR322 was determined by comparison with unpublished data of H. Ohmori and J. Tomizawa (pers. comm.). From the

```
IN THE TET^R REGION

A-179:
MKSNNALIVILGIVILDAVGIGLVMPVLPGLLRDIVHSDSIASHYGVLLALYALMQFLCA
PVLGALSDRFGRRPVLLASLLGATIDYAIMATTPVLWILYAGRIVAGITGATGAVAGAYI
ADITDGEDRARHFGLMSACFGVGMVAGPWPGDCWAPSPCMHHSLRRRCSTASTYYWAAS

B-112:
MERVGMDCRRRPIPCLPPRVASRCMGPGHLDLNGSRRHLANGFTTPRIGANQFLRRTVNA
QTNPWQNISIASAISSSRTRRISGSVGSWPRVRMIVLLSLRTRLGWRGCLTG

B-125:
MGAGHDYRRRTYDCLLYHATRRTGAGSALGHFRRGPLSLERDDDRPVACGIRNLARPRSS
LRHWSRHQTFRREAGHYRRHGGRRAGLRLAGVRDARLDGLPHYDSSRFRRHRDARVAGHA
VQAGR

C-338:
MRTRSRSTVRPLWPPPSPARFATWSHYRLRDHGDHTRPVDPLRRTHRGRHHRRHRCGCWR
LYRRHHRWGRSGSPLRAHERLFRRGYGGRPVAGGLLGAISLHAPFLAAAVLNGLNLLLGC
FLMQESHKGERRPMPLRAFNPVSSFRWARGMTIVAALMTVFFIMQLVGQVPAALWVIFGE
DRFRWSATMIGLSLAVFGILHALAQAFVTGPATKRFGEKQAIIAGMAADALGYVLLAFAT
RGWMAFPIMILLASGGIGMPALQAMLSRQVDDDHQGQLQGSLAALTSLTSITGPLIVTAI
YAASASTWNGLAWIVGAALYLVCLPALRRGAWSRATST

D-82:
MVHARRWRPTVPRPRACHHTHAETSAHEPEVASPIFPIGDVGDIGASNRTCGAGDAGHDA
SGVEDPQDGCGRHDRVVDSGSK

E-152:
MVVIYLPGQHGLQRGHPDAAGSEKHHNGEGHPASRRERQQDVAQRVGRHAGDNGLLLAET
FGGGTSDEGLSEGVQDSEYRKRQADHRRAPAKAVLAENDPERCRHLSYELHDKEDSHKCG
DDSHAPRPPEGADWVEGSQGHRSTLSLMRLLH

F-99:
MTQSAAGTCPTSCMIKKTVISAATIVMPRAHRKELTGLKALKGIGRRSPLCDSCIRKQPS
SRLRPLSTAAARNGACKEMAPNSPPATGLPPYPRRNKRS

IN THE pMB1 REGION

A-96:
MFIISNPYREHPLSFHRYHYPHEQKFPLTRRHQVTKQEKTALNMARFIRSQTLTLLEKLN
ELDADEQADICESLHDHADELYRSCLARFGDDGENL

C-84:
MHEPPVQPDRCALSGNYRLESNPVRHDLSPLAAATGNRISRARYVGGATEFFKWWPNYGY
TRRTVFGICALLKPVTFGKRVGSS

β-LACTAMASE

E-286:
MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRP
EERFPMMSTFKVLLCGAVLSRVDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVREL
CSAAITMSDNTAANLLLTTIGGPLELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTM
PAAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRSALPAGWFIADKSGAGERGS
RGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW

pSC101-pMB1 JUNCTION

A-87:
MKRLLLQNVCDLSNNMNGLRFPCFVKSGNAEVSALHHYVPDLHRRMLLATLWNTYICINE
ALALTLSDFSLVPPHPYRQLFTLTTFQ

D-132:
MPPCKGEFLFMGVMIPMKREKMLTIRVTDDEHARLLERCEGKQLAVWMRRDQRKITQGGC
QRFVNTDVGVPQGSQQHPAMQIRNIMVQGADFRVSRLYETRKPKTIHVVAQVADVLQQQS
LHVRSRIGDSFC
```

Figure 6. The protein sequences of the pBR322 "genes." The nomenclature is explained in the legend to Fig. 5. The protein sequences are given by the one-letter amino acid code. A-179, B-112, B-125, C-338, D-82, E-152, and F-99 are all within the *tet* region. A-179 has a signal-like sequence (underlined) at its amino terminus. The Arg-His-rich regions are also underlined. A-96 and C-84 come from the pMB1 region. A-87 and D-132 are read across the pSC101/pMB1 junction. These two proteins may contain some residual information from a nonrequired protein encoded by one of the parental plasmids that was disrupted during the pBR322 construction. The β-lactamase sequence (E-286) has its secretion signal underlined.

end of the Tn*3* sequence (3146) to 1762, the sequences are similar, but they diverge thereafter. This defines the extent of pMB1-derived DNA and, therefore, pSC101 material must occupy the region between the *Eco*RI site and 1762. Within this pMB1 region two short peptides are hypothetically encoded by the DNA sequence, one, of 96 amino acids, downstream from the origin and the other, of 84 amino acids, upstream (see Fig. 4). Whether these proteins are actually translated remains unknown, but both hypothetical "genes"

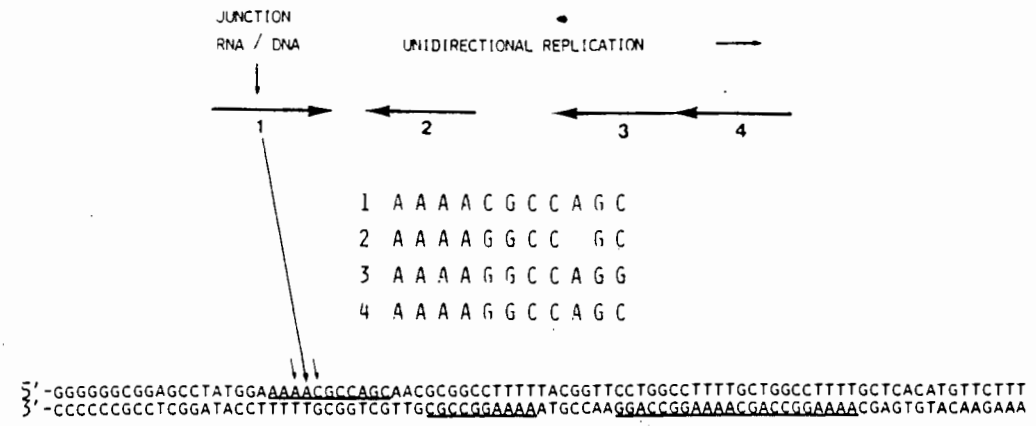


Figure 7. Origin of replication. The RNA-primer-to-DNA switch in vitro has been demonstrated to occur predominantly at the fourth A (and sometimes at the third A or the first C) in sequence 1 (Tomizawa et al. 1977). This is consistent with the assignment of the start of the nascent strand for trapped replication intermediates of pBR345 (Bolivar et al. 1977b). Replication proceeds unidirectionally in the direction indicated by the arrow. The three repeats are in the opposite orientation. There are 3 bp between 1 and 2, 7 bp between 2 and 3, and no base pairs between 3 and 4. The comparison of these four sequences shows that, of the 11 positions, 8 are occupied by the same base in all four of the repeats. The other 3 positions are the same in three of the four sequences. Sequence 2 had to be slipped in the ninth position in order to get a match at positions 10 and 11. The first A in sequence 1 corresponds to position 2536 of the plasmid sequence. Replication proceeds counterclockwise on the biological map (Fig. 4).

are preceded by sequences which have three base complementary matches with the Shine and Dalgarno (1974) sequence of the 3' end of the 16S ribosomal RNA. These could possibly function as ribosome-binding sites (see Fig. 5).

The "gene" for the peptide of 84 amino acids is located in a portion of the DNA sequence required for plasmid replication. Backman et al. (this volume) have recently shown that most of the 580 bp located upstream from the origin must be present for replication to occur. A replication protein necessary for plasmid maintenance encoded close to the origin could be the reason behind this observation. Only RNA polymerase, DNA polymerase I, and RNase H seem to be required in vitro for initiation of replication (Itoh and Tomizawa, this volume), so that cannot be the role of this protein.

No function is immediately suggested for the hypothetical peptide of 96 amino acids whose "gene" is downstream from the origin. In fact, it is unlikely that this is a required plasmid function because, as will be discussed, plasmid replication was not inactivated when DNA was inserted at the unique *Pvu*II site which lies within the "gene" for the peptide of 96 amino acids. Perhaps this protein is related to DNA transfer (H. Ohmori and J. Tomizawa, pers. comm.). The sequences of these two hypothetical proteins are presented in Figure 6.

tet^r Genes

Tetracycline resistance is lost when DNA fragments are inserted at either the *Bam*HI or *Sal*I site, and sometimes when inserts are cloned at the *Hin*dIII site (Bolivar et al. 1977c). The entire DNA sequence can code for only seven hypothetical polypeptides longer than 100 amino acids. Except for the β-lactamase gene, these cluster around the *Bam*HI and *Sal*I sites, overlap, and are read from all three triplet frames of both strands as indicated in Figures 4 and 5. These hypothetical genes were identified by locating all of the longest initiator (ATG or GTG) to nonsense triplet strings in all three reading frames of each strand. For the purpose of this discussion, read-through from weak TGA terminators will be ignored, although it could be relevant.

On the clockwise strand, four long hypothetical "genes" have been found which could code for peptides of 179, 338, 125, and 112 amino acids. The long protein of 338 amino acids spans both the *Bam*HI and *Sal*I sites; the protein of 179 amino acids spans the *Bam*HI site only. The protein of 338 amino acids also overlaps at least part of all of these other three "genes." On the other DNA strand, there are three possible "genes." Two (152 and 99 amino acids) span the *Sal*I site, while the third (82 amino acids) spans the *Bam*HI site. Therefore, inserts at either the *Bam*HI or *Sal*I site would interrupt three of these "genes." Two other hypothetical proteins (132 and 87 amino acids) can be read across the junction of the material derived from pMB1 and pSC101. It is unlikely that these two are biologically important, but they could represent residual parts of genes from one of the parent plasmids.

Which, if any, of these "genes" is active? When pMB9 or pSC101 is used to direct protein synthesis in minicells, tetracycline-related proteins are observed which have apparent molecular weights (depending on the citation) in SDS gels of about 34,000, 19,000, 17,000, 14,000, and 12,000 daltons (Meagher et al. 1977; Tait and Boyer 1978). The 34,000-dalton protein disappears when the plasmid has either *Bam* or *Sal* inserts and undoubtedly corresponds to the long pro-

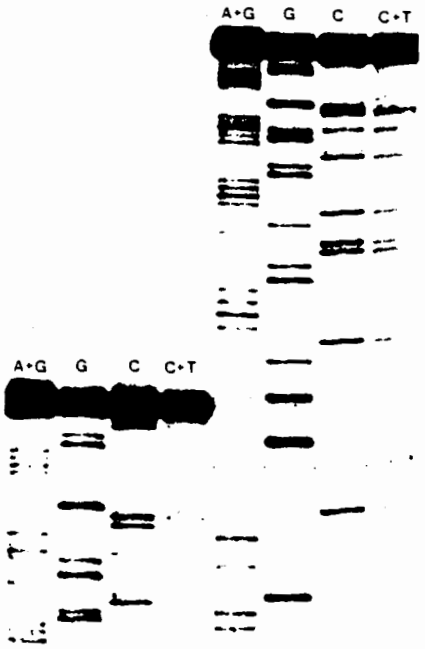


Figure 8. Sequence autoradiograph from the origin region. This gel profile shows the Maxam-Gilbert sequence ladder downstream from the origin of replication. The 5′ end of the sequenced fragment was the *Hin*fI site at position 2449. The longer region shown here starts at 2470. The sequence reads: AAGAACATGTGAGCAAAAGGCCAGCAAAA-GCgapAGGAACCGTAAAAAGGCC. The shorter region begins at position 2496 and confirms the other reading. There are three anomalies on this pattern. The first occurs between the T and G at position 2480. The light band between these darker bands is spillover from the material that was loaded earlier in the adjacent lanes and appears in the unreacted position in those adjacent four lanes. The gap (light C band) at position 2502 indicates a 5-methyl C (which reacts slowly with hydrazine) at that location (Ohmori et al. 1978). There is a light band that shows up in all four lanes between G at position 2505 and the following A. This was caused by a single-strand nick put into the DNA at this site by *Hin*fI. The nick appears at G↓AACC, which closely resembles the normal *Hin*fI cutting site G↓AXTC. Such nicks were often observed, particularly when *Hin*fI was used. These anomalies are interpretable for two reasons. First, it is obvious to a practiced eye that the banding pattern is irregular at these positions and, hence, one is alerted. Second, the region was sequenced more than once using different restriction fragments. The two false bands did not appear on other runs through this region.

tein of 338 amino acids decoded from the DNA sequence. The other proteins have not been well characterized in the literature, but could come from the "genes" which overlap the 338-amino acid protein (although the 14,000-dalton protein may come from outside the region of DNA on pBR322). Unfortunately, at least in my hands, the pBR322-directed minicell-protein profile is dominated by the β-lactamase protein such that the *tet*^{r} proteins are partially obscured. In my experiments, the β-lactamase appears as a singlet, but D. Vapnek (pers. comm.) has observed three bands in the range of 25,000–31,000 daltons whose mobilities shift in unison when the *Pst*I site carries an insert.

There is what appears (by analogy to published promoters) to be an excellent promoter sequence near the *Hin*dIII site (Boyer et al. 1977; Majors 1977). Transcripts from this site will include all of the clockwise-strand "genes." The proteins of 179, 338, and 112 amino acids have sequences that could be ribosome-binding sites (Shine and Dalgarno 1974), but I find no such sequence upstream from the initiator of the 125-amino acid protein (see Fig. 5). On the other strand, all three "genes" have a reasonable ribosome-binding-site-like sequence (Fig. 5). Protein chemistry will be required to prove a correlation between the "genes" and the proteins observed in minicells. The sequences of all of these hypothetical proteins are listed in Figure 6.

Two circumstantial observations tend to support the overlapping-genes model for tetracycline resistance. The phenotype of Tc^r is pleiotropic (Levy et al. 1978; Tait and Boyer 1978) and several different proteins functioning at once could account for this. Also, the protein sequences decoded from the DNA show that the hypothetical proteins are somewhat similar in amino acid composition as well as sharing stretches of Arg-His-rich runs. These runs could be of functional significance and implicate these "genes" as a family with functional relatedness. It is tempting to speculate that the Arg-His regions are important in tetracycline binding, since all of the proteins may interact with the antibiotic. The 338-amino acid protein resembles superficially a histone or protamine in that the first 89 amino acids are 28% arginines, and this suggests that it may be a nucleic-acid-binding protein. The amino-terminal end of the 179-amino acid protein looks like a hydrophobic secretion signal (Blobel and Dobberstein 1975), and therefore the protein could be membrane-associated. The features of this protein which resemble a secretion signal are the basic amino acid (lysine) at residue 2 followed by a stretch of 16 out of 19 aliphatic residues ending with two prolines separated by two amino acids (see Fig. 6). This structure is reminiscent of the β-lactamase signal (Sutcliffe 1978a), shown in Figure 6, which has an arginine at position 6 along with the rest of the structure, and also of a collection of other prokaryotic and eukaryotic signals, many of which share these properties (J. Knowles, pers. comm.).

Tetracycline inhibits protein synthesis. Although there is debate among investigators in the field over several points, tetracycline resistance has clearly been shown to involve several apparently separate molecular events. Tc^r cells accumulate less tetracycline than tetracycline-sensitive (Tc^s) cells at the same external tetracycline concentration (Tait and Boyer 1978; Levy et al. 1978). Tc^s cells exhibit a separate ATP-dependent, rapid tetracycline binding which is greatly reduced in Tc^r cells (Tait and Boyer 1978). Finally, Tc^r cells can synthesize protein in the presence of internal tetracycline concentrations that would totally inhibit Tc^s cells (Levy et al. 1978). There is some evidence that this last feature is due to a nonribosomal factor

since ribosomes from Tc^r cells show some ability to resist tetracycline in vitro but lose this ability after they have been washed (Levy et al. 1978). It seems likely, therefore, that more than one protein is involved in tetracycline resistance and that, for instance, the 179-amino acid protein (because of its postulated secretion signal) is involved in one of the tetracycline-exclusion events at the membrane and that the 338-amino acid protein (because of its postulated nucleic-acid-binding activity) is somehow involved in protein synthesis.

Cloning in pBR322

While studying the pBR322 DNA molecule, I found unique cutting sites for two enzymes that had not been reported previously (see Fig. 4). The *Pvu*II site at map position 2067 is suitable for flush-end cloning. A DNA fragment containing the *lac* operator was inserted into the unique *Pvu*II site by blunt-end ligation (Backman et al. 1976). Cells carrying the *lac* plasmid could be detected on indicator plates and these were Tc^s as well as Ap^r, so *Pvu*II provides a site for blunt-ending with no inactivation of either of the two drug resistances or replication. There is a unique *Ava*I site at position 1424. (See also Note Added in Proof.)

The β-lactamase secretion signal, when hooked in front of a new gene, can be used to secrete that protein product (Villa-Komaroff et al. 1978). The fact that the 179-amino-acid peptide in the *tet* region also appears to have a secretion signal may mean that genes cloned in the *Bam*HI site can also lead to secreted products.

Because the entire DNA sequence is known, it is now very simple to determine a detailed restriction map for an insert cloned into any of the pBR322 sites. When total chimeric plasmid DNA is cut by a restriction enzyme and the products are displayed on a gel, the fragments from the pBR322 portion of the molecule serve as internal size markers such that one can estimate the sizes of the insert bands accurately. Data which simplify this process have been published elsewhere (Sutcliffe 1978b). Plasmid pBR322 is quite a good source of DNA size markers. A word of caution: it has been noticed that the *Mbo*I-*Eco*RI fragment carrying the amino terminus of β-lactamase migrates on gels slightly anomalously compared with other *Mbo*I-generated markers; it migrates as though it were 345 bp, rather than 316 bp, long (K. Talmadge, pers. comm.). It is possible that other bands will occasionally exhibit anomalous mobility.

Accuracy of the Sequence

The confidence level of this sequence is of prime importance for two reasons. First, I have discussed several hypothetical gene products. Such discussions would be frivolous if there were not good reasons to expect the sequence to be accurate. Second, I am certain that pBR322 DNA will be used for many genetic and biochemical experiments because the DNA sequence is known. Other investigators must know how much to trust these data. The procedures described under Strategy make it likely that the sequence is entirely correct. Comparisons with other data raise the level of confidence even higher.

1. The sequence between the first and last amino acid of β-lactamase was shown to agree with the partial amino acid sequence for the protein (Sutcliffe 1978a; Ambler and Scott 1978). The one difference discovered as a result of comparison was shown to be a strain difference. Because the β-lactamase partial protein sequence was not consulted until the gene sequence was completed, I expect the rest of the pBR322 sequence to be comparably accurate.
2. The DNA sequence of the left-end inverted repeat determined by Ohtsubo et al. (this volume) for the Tn*3* of R1*drd*19 agrees completely with the pBR322 sequence.
3. The sequence from the end of Tn*3* to position 2360 corresponds to the origin of replication. The origin sequence for the related plasmid pBR345 has been solved (Bolivar et al. 1977b). These workers also have the sequence (unpubl.) for some of the rest of the pBR345 miniplasmid (H. Boyer, pers. comm.). A few differences appear, but in all cases the pBR322 data are very clear at these sites. In one instance the pBR345 sequence predicts an *Alu*I cut where there is none on the pBR322 sequence. *Alu*I cuts pBR345 but not pBR322 at that site, so these two sequences have diverged.
4. Small regions of the sequence have been verified by various other workers.

These direct sequence comparisons result in an independent check of about 2200 bp, or roughly 50% of the DNA. The fact that the restriction map matches the restriction digestion pattern (Sutcliffe 1978b) serves as a partial check on the rest of the plasmid. Therefore, the DNA sequence determined for pBR322 is probably totally correct. It is possible that there could be a few errors, but, because the same care was used for all parts of the sequencing, this is unlikely.

Evolution of pBR322

Unfortunately, the sequence of pBR322 is not static. Cases have been observed, in nonmutagenized DNA stocks, in which certain restriction patterns are altered (K. Backman, pers. comm.). There are two sorts of possible explanations for such findings. One is that the plasmid was only recently constructed in vitro and is still changing to reach its most stable form. The other, and perhaps more likely, explanation is that within a population of chloramphenicol-amplified DNA molecules, many molecules differ in one or a few positions from the norm. This was recently shown to be the case for the RNA phage Qβ (Domingo et al. 1978). When such plasmids are clonally amplified, a population that has temporarily fixed the particular difference

BLOBEL, G. and B. DOBBERSTEIN. 1975. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell. Biol.* **67**:835.

BOLIVAR, F., R. L. RODRIGUEZ, M. C. BETLACH and H. W. BOYER. 1977a. Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9. *Gene* **2**:75.

BOLIVAR, F., M. C. BETLACH, H. L. HEYNKER, J. SHINE, R. L. RODRIGUEZ, and H. W. BOYER. 1977b. Origin and replication of pBR345 plasmid DNA. *Proc. Natl. Acad. Sci.* **74**:5265.

BOLIVAR, F., R. L. RODRIGUEZ, P. J. GREEN, M. C. BETLACH, H. L. HEYNEKER, H. W. BOYER, J. H. CROSA, and S. FALKOW. 1977c. Construction and characterization of new cloning vehicles. II. A multi-purpose cloning system. *Gene* **2**:95.

BOYER, H. W., M. C. BETLACH, F. BOLIVAR, R. L. RODRIGUEZ, H. L. HEYNEKER, J. SHINE, and H. M. GOODMAN. 1977. The construction of molecular cloning vehicles. In *Recombinant molecules: Impact on science and society* (ed. R. F. Beer and E. G. Bassett), p. 9. Raven Press, New York.

COHEN, S. N. and A. C. Y. CHANG. 1977. Revised interpretation of the origin of the pSC101 plasmid. *J. Bacteriol.* **132**:734.

DATTA, N. and P. KONTOMICHALOU. 1965. Penicillinase synthesis controlled by infectious R factors in enterobacteriace. *Nature* **208**:239.

DOMINGO, E., D. SABO, T. TANIGUCHI, and C. WEISSMAN. 1978. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**:735.

HALL, A. and J. R. KNOWLES. 1976. Directed selective pressure on a β-lactamase to analyze changes involved in the development of enzyme function. *Nature* **264**:803.

HAMER, D. H. and C. A. THOMAS, JR. 1976. Molecular cloning of DNA fragments produced by restriction endonucleases Sal I and Bam I. *Proc. Natl. Acad. Sci.* **73**:1537.

KNOX, J. R., J. A. KELLY, P. C. MOEWS, and N. S. MURTHY. 1976. 5.5 Å Crystallographic structure of penicillin β-lactamase and radius of gyration in solution. *J. Mol. Biol.* **104**:865.

LEVY, S. B., L. MCMURRAY, P. ONIGMAN, and R. M. SAUNDERS. 1978. Plasmid-mediated tetracycline resistance in *E. coli*. In *Topics in infectious diseases* (ed. J. Drews and G. Hogenaur), vol. 2 p. 181. Springer-Verlag, New York.

LOBBAN, P. and A. D. KAISER 1973. Enzymatic end-to-end joining of DNA molecules. *J. Mol. Biol.* **78**:453.

MAJORS, J. 1977. Control of the *E. coli lac* operon at the molecular level. Ph.D. thesis, Harvard University, Cambridge, Massachusetts.

MAXAM, A. M. and W. GILBERT. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci.* **74**:560.

MEAGHER, R. B., R. C. TAIT, M. BETLACH, and H. W. BOYER. 1977. Protein expression in *Escherichia coli* mini cells by recombinant plasmids. *Cell* **10**:521.

MEYNELL, E. and N. DATTA. 1967. Mutant drug resistance factors of high transmissibility. *Nature* **214**:885.

OHMORI, H., J. I. TOMIZAWA, and A. MAXAM. 1978. Detection of 5-methylcytosine in DNA sequences. *Nucleic Acids Res.* **5**:1479.

RODRIGUEZ, R. L., F. BOLIVAR, H. M. GOODMAN, H. W. BOYER, and M. BETLACH. 1976. Construction and characterization of cloning vehicles. In *Molecular mechanisms in the control of gene expression* (ed. D. P. Nierlich et al.), p. 471. Academic Press, New York.

SANGER, F. and A. R. COULSON. 1978. The use of thin acrylamide gels for DNA sequencing. *FEBS Lett.* **87**:107.

SHINE, J. and L. DALGARNO. 1974. The 3'-terminal sequence of *E. coli* 16S RNA. *Proc. Natl. Acad. Sci.* **71**:1342.

SO, M., R. GILL, and S. FALKOW. 1976. The generation of ColE1-Ap^r cloning vehicle which allows detection of inserted DNA. *Mol. Gen. Genet.* **142**:239.

STEITZ, J. A. and K. JAKES. 1975. How ribosomes select initiator regions in mRNA. *Proc. Natl. Acad. Sci.* **72**:4734.

SUTCLIFFE, J. G. 1978a. Nucleotide sequence of the ampicillin resistance gene of *E. coli* plasmid pBR322. *Proc. Natl. Acad. Sci.* **75**:3737.

———. 1978b. pBR322 restriction map derived from the DNA sequence: Accurate DNA size markers up to 4361 nucleotides pairs long. *Nucleic Acids Res.* **5**:2721.

TAIT, R. C. and H. W. BOYER. 1978. On the nature of tetracycline resistance controlled by the plasmid pSC101. *Cell* **13**:73.

TOMIZAWA, J., H. OHMORI, and R. E. BIRD. 1977. Origin of replication of colicin E1 plasmid DNA. *Proc. Natl. Acad. Sci.* **74**:1865.

VILLA-KOMAROFF, L., A. EFSTRATIADIS, S. BROOME, P. LOMEDICO, R. TIZARD, S. NABER, W. CHICK, and W. GILBERT. 1978. A bacterial clone synthesizing proinsulin. *Proc. Natl. Acad. Sci.* **75**:3727.

Plant Physiol. (1993) 103: 467–476

Cloning of Higher Plant ω-3 Fatty Acid Desaturases[1]

Narendra S. Yadav*, Anna Wierzbicki, Mary Aegerter, Cheryl S. Caster, Luis Pérez-Grau, Anthony J. Kinney, William D. Hitz, J. Russell Booth Jr., Bruce Schweiger, Kevin L. Stecca, Stephen M. Allen, Marita Blackwell, Robert S. Reiter, Thomas J. Carlson, Sandra H. Russell, Kenneth A. Feldmann[2], John Pierce, and John Browse

Agricultural Products (N.S.Y., A.W., C.S.C., L.P.-G., A.J.K., J.R.B., B.S., K.L.S., M.B., J.P.) and Central Research and Development (W.D.H., S.M.A., R.S.R., T.J.C., S.H.R., K.A.F.), E. I. duPont de Nemours & Co., Experimental Station, P.O. Box 80402, Wilmington, Delaware 19880-0402; and Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164-6340 (M.A., J.B.)

Arabidopsis thaliana T-DNA transformants were screened for mutations affecting seed fatty acid composition. A mutant line was found with reduced levels of linolenic acid (18:3) due to a T-DNA insertion. Genomic DNA flanking the T-DNA insertion was used to obtain an *Arabidopsis* cDNA that encodes a polypeptide identified as a microsomal ω-3 fatty acid desaturase by its complementation of the mutation. Analysis of lipid content in transgenic tissues demonstrated that this enzyme is limiting for 18:3 production in *Arabidopsis* seeds and carrot hairy roots. This cDNA was used to isolate a related *Arabidopsis* cDNA, whose mRNA is accumulated to a much higher level in leaf tissue relative to root tissue. This related cDNA encodes a protein that is a homolog of the microsomal desaturase but has an N-terminal extension deduced to be a transit peptide, and its gene maps to a position consistent with that of the *Arabidopsis fad D* locus, which controls plastid ω-3 desaturation. These *Arabidopsis* cDNAs were used as hybridization probes to isolate cDNAs encoding homologous proteins from developing seeds of soybean and rapeseed. The high degree of sequence similarity between these sequences suggests that the ω-3 desaturases use a common enzyme mechanism.

The ω-6 and ω-3 fatty acid desaturases introduce the second and the third double bonds, respectively, in the biosynthesis of 18:2 and 18:3 fatty acids, which are important constituents of plant membranes. (The ω-3 and ω-6 designations refer to positions of the double bond from the methyl end of fatty acids.) They are also commercially important because the oxidative stability and nutritional value of seed oils is affected by the levels of these fatty acids. In leaf tissue, there are two distinct pathways for polyunsaturated fatty acid biosynthesis, one located in the microsomes and the other located in the plastid membranes. In nongreen tissues and developing seeds, the microsomal pathway predominates. In *Arabidopsis thaliana*, the microsomal ω-6 and ω-3 fatty acid desaturations are controlled by the *fad* 2 and *fad* 3 loci, respectively (Lemieux et al., 1990), and the plastid ω-6 and ω-3 fatty acid desaturations are controlled by the *fad* C and *fad* D loci, respectively (Browse and Somerville, 1991). It has been postulated that these loci correspond to structural genes for the desaturase enzymes, which have been recalcitrant to purification and study. Indeed, a cDNA encoding a *Brassica napus* microsomal ω-3 desaturase was recently cloned by homology to a fragment of *Arabidopsis* genomic DNA isolated by map-based cloning of the *fad* 3 locus (Arondel et al., 1992).

Genetic approaches for cloning plant genes encoding biochemically intractable products, such as membrane-associated desaturases, are becoming increasingly more refined and powerful, especially in studies that depend on the small, well-characterized genome of *A. thaliana*. In addition to map-based cloning, these methods include transposon tagging (Balcells et al., 1991) and T-DNA tagging (Feldmann, 1991; Walden et al., 1991). The T-DNA tagging method, in which insertion of the T-DNA of *Agrobacterium tumefaciens* into a gene of interest via transformation provides both the mutant phenotype and the means with which to clone the mutant allele, has been most successful to date.

We report here the isolation of the *Arabidopsis* microsomal ω-3 fatty acid desaturase gene by T-DNA tagging and the subsequent use of its cognate cDNA to manipulate the levels of polyunsaturated fatty acids in transgenic plant tissues as well as to isolate cDNAs from *Arabidopsis*, soybean, and rapeseed that encode homologs of the microsomal ω-3 desaturase, including putative plastid ω-3 desaturases.

MATERIALS AND METHODS

Screening of an *Arabidopsis* T-DNA Mutant Population

About 100 T_3 seeds (2 mg) of each of 6000 members of a population of T-DNA transformed lines of *Arabidopsis thaliana* (ecotype Wassilewskija) (Feldmann and Marks, 1987)

[1] This research was supported in part by National Science Foundation grant DCB-90084232 and U.S. Department of Agriculture grant 9237301-7728 to J.B.

[2] Present address: Department of Plant Sciences, University of Arizona, Tucson, AZ 87521.

* Corresponding author; fax 1-302-695-4296.

Abbreviations: cM, centimorgan; NOS, nopaline synthase gene; NPTII, neomycin phosphotransferase II; PCR, polymerase chain reaction; 16:0, palmitic acid; 16:1, palmitoleic acid; 18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid; 18:3, linolenic acid; 20:1, eicosenoic acid.

Northern Analyses

The following *Arabidopsis* tissues were harvested and frozen in liquid nitrogen: expanding leaves from the rosette stage of wild-type and line 3707 *Arabidopsis* plants grown side by side in soil in the greenhouse, 2-week-old whole wild-type *Arabidopsis* seedlings (with roots) grown in coarse sand, and in vitro cultured roots (Russell et al., 1992). RNA was isolated as described in Rerie et al. (1991). Poly(A)$^+$ mRNA was isolated using the PolyATtract mRNA isolation system (Promega Corp.). RNA was fractionated on 1.2% (w/v) agarose, 2.2 M formaldehyde gels, blotted onto Nytran membranes (Schleicher & Schuell), and cross-linked by UV irradiation (Ausubel et al., 1991). RNA blots were hybridized at 60°C with the ^{32}P-labeled CF3 and CFD cDNAs (see "Results") and an *Arabidopsis* actin probe in 6× SSC, 50 mM Tris-HCl, pH 8.0, 5× Denhardt's solution, 1.0% (w/v) SDS, 100 μg/mL of salmon sperm DNA, and were washed at 65°C to a stringency of 0.1× SSC. The actin probe, which served to normalize mRNA loadings, was obtained from RNA by reverse transcription and PCR amplification with primers corresponding to nucleotides 938 to 957 and nucleotides 1506 to 1524 of the *AAc1* gene (Nairn et al., 1988).

Plant Transformations

For constitutive expression in plants, the 1.4-kb CF3 cDNA (see "Results") was ligated in the sense orientation between the 35S promoter and the 3′ region of the NOS (Russell et al., 1992) (see Fig. 4). For seed-specific expression, the cDNA insert was cloned in sense orientation between the promoter for the gene for the α-subunit of β-conglycinin (Doyle et al., 1986) and the 3′ region of the phaseolin gene (Slightom et al., 1983). The chimeric genes were cloned adjacent to a plant-selectable marker between the T-DNA borders of a derivative of binary vector pZS94 (Russell et al., 1992) to provide plasmids pAW50 (β-conglycinin: CF3 cDNA) and pAW31 (35S: CF3 cDNA). The plant-selectable markers were sulfonylurea resistance (Russell et al., 1992) in plasmid pAW50 and kanamycin resistance (35S:NPT II:3′NOS, see Fig. 4) in plasmid pAW31. Plasmids pAW50 and pAW31 were transformed into *A. tumefaciens* strain LBA4404 (Hoekema et al., 1983) and *A. tumefaciens* strain R1000 (a C58 strain carrying an Ri-plasmid) (Moore et al., 1979), respectively, by the freeze/thaw method (Holsters et al., 1978).

Complementation of the mutation in line 3707 was carried out by transformation of root explants of line 3707 homozygous for the mutation by *Agrobacterium* strain LBA4404/pAW50 (Russell et al., 1992). Primary transformants (R_1) were selected on chlorsulfuron and transferred to individual containers as previously described (Russell et al., 1992). Individual R_2 seeds from two independent transgenic plants were analyzed for fatty acid composition.

Carrot (*Daucus carota* L.) cells were transformed by cocultivation of carrot root discs with *Agrobacterium rhizogenes* strains R1000 or R1000/pAW31 (Petit et al., 1986). Inoculated discs were incubated for 2 weeks at 25°C on an agar-solidified water medium and then transferred to a medium containing 500 mg/L of carbenicillin. Hairy roots that formed on the cut surfaces were excised and individually maintained on Murashige and Skoog minimal organics medium (Gibco) with 30 g/L of Suc and 500 mg/L of carbenicillin, with or without 50 μg/mL of kanamycin sulfate, and sampled for fatty acid composition.

RESULTS

An *Arabidopsis* Mutant Defective in ω-3 Desaturation Due to T-DNA Insertion

Since the T_3 seeds of T-DNA-transformed lines are segregating for the T-DNA insert, we combined approximately 100 T_3 seeds (2 mg) of each of 6000 members of the population of T-DNA-transformed lines of *A. thaliana* and determined the fatty acid content of each of the 6000 pooled samples (Browse et al., 1986). Based on our knowledge of chemically induced lipid mutants (James and Dooner, 1990; Lemieux et al., 1990), we expected to be able to identify mutants in a segregating line using this approach. Three lines with reduced 18:3 were identified, and each of these was shown to be segregating for the mutant phenotype. The first of these lines identified (line 3707) produced homozygous seeds that contained 3.1% 18:3 (Table I), suggesting that the mutation in this line is "leaky" compared with the previously described *fad* 3-2 mutant, which contains 1.9% 18:3 (James and Dooner, 1990). Individual plants of line 3707 were selfed, and 262 of the resultant T_4 progeny were grown and assayed for the presence of nopaline in leaf extracts (Errampalli et al., 1991). In addition, T_5 seeds from each of the T_4 plants were analyzed in bulk for fatty acid composition and for their ability to germinate in the presence of kanamycin (Feldmann et al., 1989). The analysis showed that the T_4 progeny of line 3707 fell into three classes (Table II), indicating that a single T-DNA insertion conditioned the low 18:3 phenotype in line 3707. The cosegregation of the mutant phenotype and T-DNA markers indicates with 95% certainty that the T-DNA and the mutation are no further than 1.2 cM apart. On the basis of this estimation of linkage, we proceeded to isolate plant DNA flanking the site of T-DNA insertion.

Arabidopsis ω-3 Desaturase Gene and cDNA

Since the modified T-DNA we used contains the origin of replication and the ampicillin resistance gene of plasmid pBR322 (Feldmann and Marks, 1987), the left T-DNA–plant DNA junction fragments from line 3707 were recovered as plasmids in *Escherichia coli* by the method of plasmid rescue

Table I. *Fatty acid composition of normal and mutant 3707 seeds*

The normal sample is an average of pooled T_3 seeds from T-DNA lines 3501 to 4000, the 3707 T_3 sample is pooled T_3 seeds from line 3707, and the 3707 T_4 sample is pooled T_4 seeds from a selfed progeny of line 3707 homozygous for the mutant fatty acid phenotype.

Sample	Percent Fatty Acid					
	16:0	18:0	18:1	18:2	18:3	20:1
Normal	7.4	3.0	17.0	29.3	16.1	20.2
3707 T_3	7.0	2.9	17.7	35.0	10.2	20.5
3707 T_4	6.4	3.0	15.9	42.4	3.1	23.6

Table II. *Co-segregation of mutant fatty acid phenotype and T-DNA markers in line 3707*

Leaves of 262 individual T_4 plants derived from line 3707 were tested for the presence of nopaline, and their selfed, progeny seeds (T_5) were tested for the mutant fatty acid phenotype and kanamycin resistance.

Number of T_4 Individuals	Phenotype
64	Nopaline absent in leaves; progeny (T_5) seeds show wild-type fatty acid composition and are all kanamycin sensitive
134	Nopaline present in leaves; progeny (T_5) seeds show heterozygous fatty acid composition similar to 3707 T_3 pool and segregate for kanamycin resistance
64	Nopaline present in leaves; progeny (T_5) seeds show the homozygous mutant fatty acid composition and are all kanamycin resistant

(Behringer and Medford, 1992). For this, line 3707 genomic DNA was isolated, digested with either *Sal*I or *Bam*HI restriction enzyme, self-ligated, and used to transform *E. coli* cells. This resulted in the isolation of plasmids pS1 and pB1 from the *Sal*I- and *Bam*HI-digested DNAs, respectively. Restriction analysis of these plasmids showed that, in addition to the expected fragments of the T-DNA, pS1 contained a 2.9-kb *Eco*RI-*Sal*I fragment and pB1 contained a 1.4-kb *Eco*RI-*Bam*HI fragment (in each case, the *Eco*RI site being in the left T-DNA border), and that the 1.4-kb *Eco*RI-*Bam*HI fragment in pB1 was contained within the 2.9-kb *Eco*RI-*Sal*I fragment in pS1. Southern analysis using a radiolabeled 1.4-kb *Eco*RI-*Bam*HI fragment as the hybridization probe showed that it hybridized to specific fragments of genomic DNA from both wild-type and line 3707 plants (data not shown). The nucleotide sequence of approximately 0.8 kb of the junction fragment starting from the *Eco*RI site in plasmid pS1 was determined; it was co-linear with the sequence of the T-DNA up to nucleotide position 65 in the left T-DNA border repeat (Yadav et al., 1982). The sequence beyond this point of divergence showed no significant identity to the T-DNA and revealed no extended open reading frame.

To isolate a cDNA corresponding to the site of T-DNA insertion, we isolated the corresponding genomic DNA from wild-type plants. For this we used the ^{32}P-labeled 1.4-kb *Eco*RI-*Bam*HI fragment as the hybridization probe to screen a phage library made to wild-type *Arabidopsis* genomic DNA. Southern analysis of DNA isolated from several positively hybridizing clones showed that a 5.2-kb *Hin*dIII fragment (Fig. 1) hybridized to the 1.4-kb *Eco*RI-*Bam*HI fragment. The 5.2-kb *Hin*dIII fragment was isolated and used, in turn, as a probe to screen an *Arabidopsis* cDNA library. Several positively hybridizing plaques were purified and their plasmids were excised. One plasmid contained a 1.4-kb cDNA, designated CF3, which was shown by Southern analysis to hybridize to a region of wild-type *Arabidopsis* genomic DNA present to the right of the site of T-DNA insertion as shown in Figure 1.

The nucleotide sequence of CF3 cDNA revealed a large open reading frame (nucleotides 46–1206) that encodes a 386-amino acid polypeptide, designated A3 (Fig. 2). Comparison of the deduced amino acid sequence A3 to that of the polypeptide encoded by the structural gene (*des A*) for a cyanobacterium fatty acid desaturase (Wada et al., 1990) revealed an overall identity of 26% and higher identity over shorter stretches of amino acids. This strongly suggested that CF3 cDNA encoded *Arabidopsis* microsomal ω-3 fatty acid desaturase. *Arabidopsis* polypeptide A3 showed 93 and 68% overall identity to the subsequently published polypeptide sequences of rapeseed ω-3 fatty acid desaturase (Arondel et al., 1992) and a mung bean cDNA, ARG1, made to IAA-induced mRNA (Yamamoto et al., 1992), respectively. However, it showed overall identities of only 21 and 17% to the microsomal stearoyl-CoA desaturases from rat (Thiede et al., 1986) and yeast (Stukey et al., 1990), respectively. No significant homology was observed with the soluble stearoyl-acyl

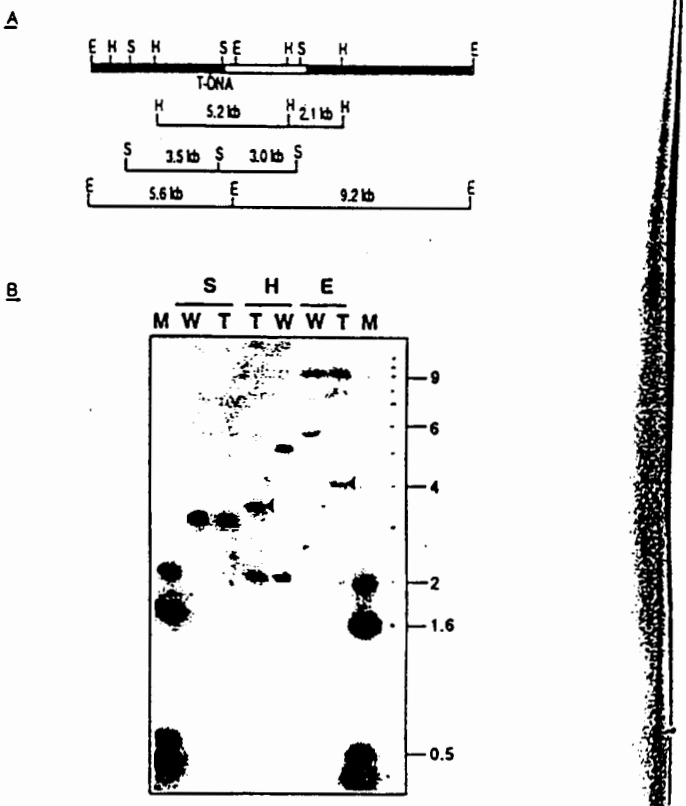


Figure 1. A, Restriction map of the region of wild-type *Arabidopsis* DNA containing the region hybridizing to the entire CF3 cDNA (open bar). T-DNA marks the site corresponding to position of the T-DNA insertion in line 3707. B, Southern analysis of genomic DNA from wild-type (W) and 3707 (T) *Arabidopsis* plants using ^{32}P-labeled CF3 cDNA as hybridization probe. The arrows show the novel, right junction fragments in line 3707 due to T-DNA insertion. Enzymes used are *Sal*I (S), *Hin*dIII (H), and *Eco*RI (E). Lane M contains 1-kb ladder DNA size markers.

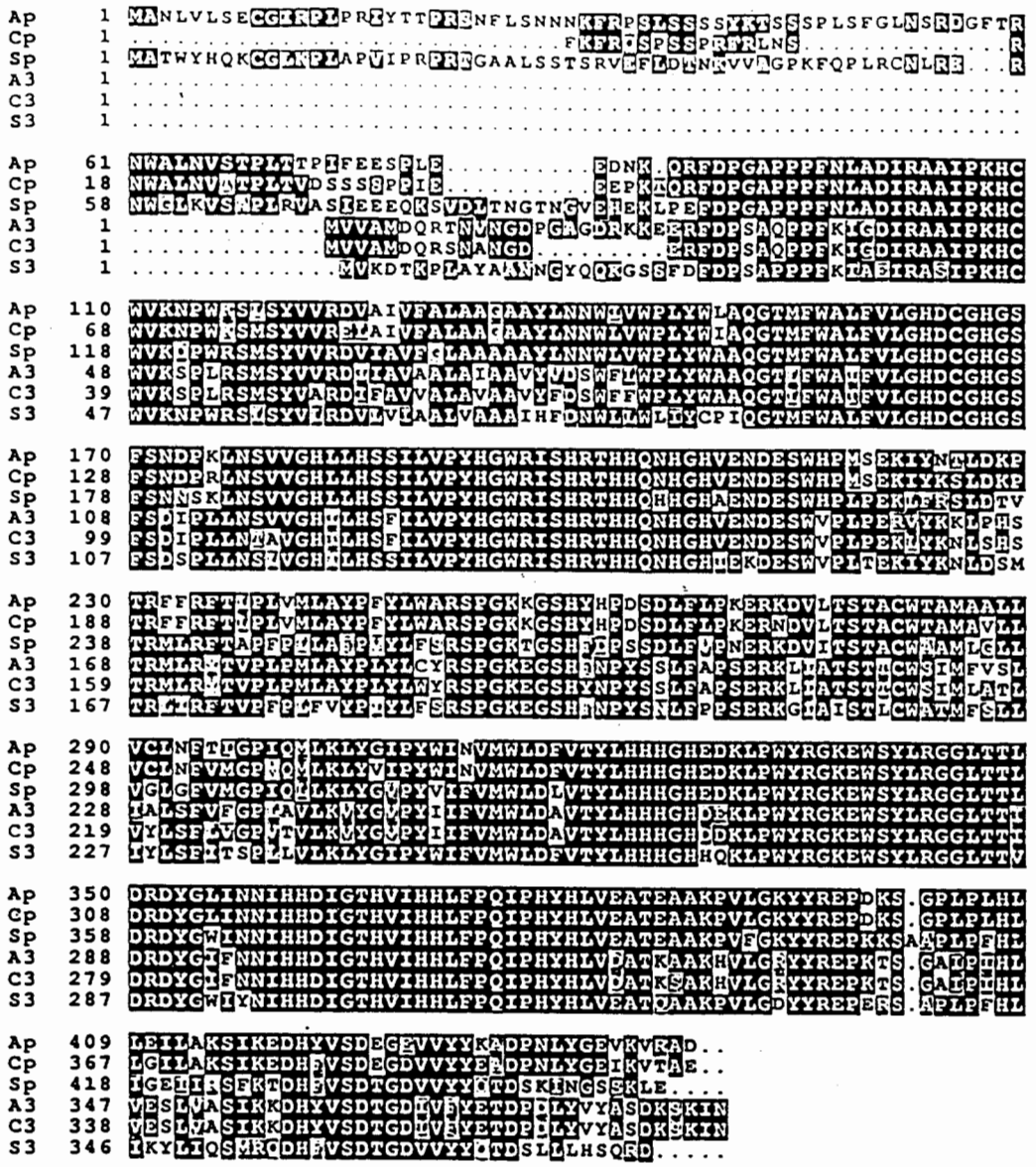


Figure 2. Comparison of the deduced amino acid sequences of higher plant ω-3 fatty acid desaturase homologs using standard one-letter amino acid codes. Identical and similar residues are shown on backgrounds of black and gray, respectively. Ap, Cp, Sp, A3, C3, and S3 refer to the deduced amino acid sequences encoded by *Arabidopsis* CFD cDNA, rapeseed BND cDNA, soybean GMD cDNA, *Arabidopsis* CF3 cDNA (*Fad 3*), rapeseed BN3 cDNA, and soybean GM3 cDNA, respectively.

carrier protein desaturases from higher plants (Shanklin and Somerville, 1991; Thompson et al., 1991).

Southern analysis of genomic DNA from wild-type and line 3707 plants using ^{32}P-labeled CF3 cDNA showed that the T-DNA integrated 5′ to the desaturase coding sequence (Fig. 1). To accurately determine the site of T-DNA integration, the right T-DNA–plant DNA junction fragment was isolated by screening a library made to genomic DNA from line 3707 with ^{32}P-labeled CF3 cDNA. Comparison of the nucleotide sequence of the wild-type ω-3 desaturase gene with those of the T-DNA-disrupted gene in the left and right junction fragments showed that the insertion in line 3707 resulted in a 56-bp deletion at the site of integration that occurred 612 bp 5′ to the initiation codon for the desaturase coding sequence.

Northern analysis of poly(A)$^+$ RNA isolated from leaf tissues of wild-type *Arabidopsis* and line 3707 homozygous for the low 18:3 phenotype showed that, relative to the wild-type tissue, the mutant tissue contained about one-fifth to one-tenth the amount of an apparently full-length ω-3 desaturase mRNA (Fig. 3). The above data, taken together with the leaky mutant phenotype in line 3707, suggest that the T-

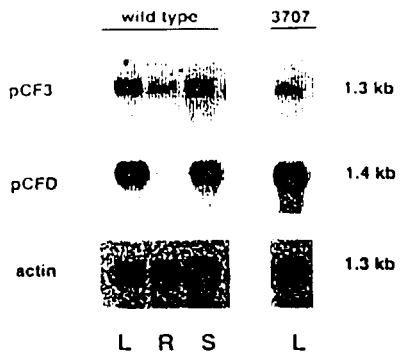


Figure 3. RNA gel-blot analysis of *Arabidopsis* using CF3 and CFD probes. Poly(A)$^+$ mRNAs (0.3 μg) isolated from expanding leaves (L) from wild-type and line 3707 rosette-stage plants, in vitro cultured wild-type roots (R), and 2-week-old wild-type seedlings (S) were separated on 1.2% agarose, 2.2 M formaldehyde gels, transferred to a nylon membrane, and hybridized to ^{32}P-labeled CF3 cDNA, CFD cDNA, or *AAct* actin probe. The sizes of the hybridizing mRNAs are indicated. The same blot was hybridized sequentially with CF3 and CFD probes to determine the relative prevalences of the corresponding mRNAs in these tissues. The actin probe was used as a loading control. Autoradiographs for CF3 and CFD probes were exposed for 12 h at −70°C, and that for the actin probe was exposed for 24 h at −70°C.

DNA insertion altered the quantitative expression of the microsomal ω-3 fatty acid desaturase without physically interrupting its mRNA.

Overexpression of *Arabidopsis* ω-3 Fatty Acid Desaturase in Transgenic Tissues

To confirm the identity of the gene product encoded by CF3 cDNA, the complete cDNA was introduced in the sense orientation behind a seed-specific promoter into line 3707. Five of six R_2 seeds from each of two independent transgenic plants tested showed more than a 10-fold increase in 18:3 level (Table III). The remaining seed from each transformant showed the mutant fatty acid phenotype.

CF3 cDNA was also introduced into carrot roots in the sense orientation behind the 35S promoter in binary vector pAW31 via the binary Ri plasmid transformation method. By this method, only a fraction (about half) of the hairy roots formed in the absence of the kanamycin selection will be transformed with both the Ri plasmid and the experimental plasmid, if present. The average 18:3 content in nine of the control hairy roots (transformed with R1000 strain without an experimental vector) was 9.2%, SE 0.3%. Of the 20 hairy roots transformed with R1000 strain containing pAW31 and grown in the absence of kanamycin, 6 showed a high content of 18:3 (average 62%, SE 0.4%), 2 showed an intermediate content (average 19%, SE 4%), and 12 showed the normal content (average 7.6%, SE 2.2%) compared with the control. Thus, overexpression of the cDNA in some carrot roots resulted in the conversion of up to 94% of the endogenous 18:2 into 18:3 (Fig. 4).

There was no significant change in fatty acids other than 18:2 and 18:3. Five roots of the high class (numbers 4, 19, 22, 23, and 25), one root of the intermediate class (number 36), one root with normal 18:3 (number 20), and one R1000 control root were tested for their ability to grow on kanamycin and for the presence of the chimeric gene in their genomic DNA. Roots 4, 19, 22, and 25 were kanamycin resistant, root 36 was weakly kanamycin resistant, and roots 20 and 23 and the R1000 control root were kanamycin sensitive. Southern analyses using ^{32}P-labeled 35S:CF3 cDNA:3′NOS chimeric gene showed that all roots of the high and intermediate classes contained the chimeric gene, whereas the root with normal 18:3 (number 20) and the R1000 control root did not (Fig. 4). It is unclear if the intermediate content of 18:3 in root 36 is related to the reduced intensity of hybridization to the 1.4-kb fragment in this root. Root 23 has an apparent deletion of approximately 0.5 kb in the 4.4-kb *Hin*dIII fragment, and its kanamycin sensitivity suggests that the deletion is in the 35S:NPTII:3′NOS chimeric gene.

An *Arabidopsis* cDNA Encoding a Homolog of the Microsomal ω-3 Desaturase

^{32}P-labeled CF3 cDNA was used as a hybridization probe at low stringency to screen the *Arabidopsis* etiolated hypocotyl cDNA library (Kieber et al., 1993). Several weakly hybridizing plaques were purified and their plasmids were excised and partially sequenced. The nucleotide sequence of 1550 bp of the cDNA insert, designated CFD, in one plasmid revealed an open reading frame encoding a 446-amino acid polypeptide, designated Ap, with an estimated molecular mass of 51 kD. Alignment of polypeptide Ap sequence with that of the *Arabidopsis* microsomal ω-3 desaturase showed an overall

Table III. *Fatty acid composition of seeds of line 3707 transformed with β-conglycinin promoter:CF3cDNA:3′NOS chimeric gene*

Wild-type and mutant 3707 compositions are each an average of three individual seeds. Samples A-1 to A-6 are individual R_2 seeds from one 3707 transformant, and samples B-1 to B-6 are individual R_2 seeds from an independent 3707 transformant. The level of each fatty acid is shown as a percentage of all five.

Sample	Percent Fatty Acid				
	16:0	18:0	18:1	18:2	18:3
Wild type	8.2	5.5	19.3	41.4	25.7
SE	0.20	0.32	0.46	0.30	0.08
Mutant 3707	8.7	5.1	19.8	61.8	4.5
SE	0.16	0.04	0.82	0.92	0.19
A-1	10.3	4.3	29.8	51.5	4.1
A-2	10.1	4.0	22.2	9.2	54.5
A-3	10.8	6.1	21.5	13.6	48.0
A-4	12.0	7.4	16.5	5.5	56.7
A-5	10.4	4.4	18.6	17.7	49.0
A-6	10.5	4.8	15.4	15.1	54.1
B-1	10.6	4.9	19.6	61.0	4.0
B-2	10.1	5.0	19.2	9.2	56.5
B-3	8.9	4.2	27.1	7.8	52.0
B-4	10.3	5.2	17.3	9.9	57.4
B-5	9.8	4.7	19.5	9.2	56.8
B-6	10.4	5.3	16.9	17.8	49.5

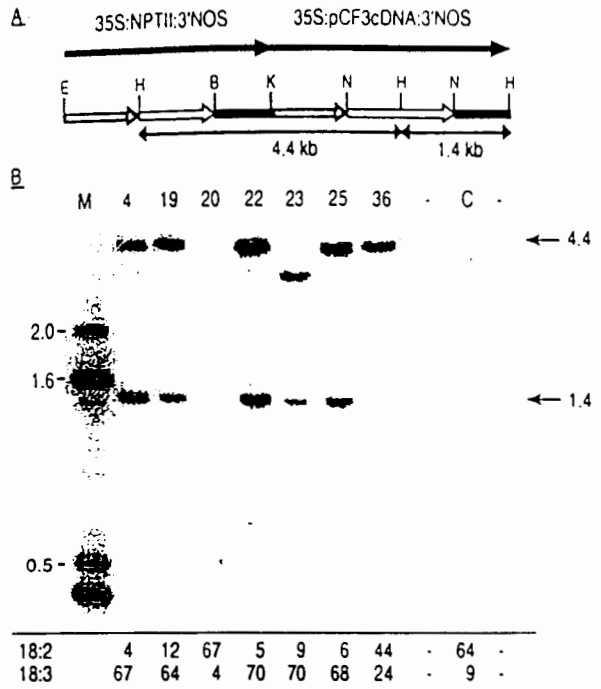


Figure 4. A, Physical map of a region of plasmid pAW31 showing the 35S:NPTII:3'NOS and 35S:CF3cDNA:3'NOS chimeric genes. The 35S promoter (35S) fragment is shown as gray boxes, the CF3 cDNA and the NPTII coding region are shown as open boxes, and the 3'NOS regions are shown as solid boxes. B, Genomic DNA gel-blot analysis of transgenic carrot hairy roots. *Hin*dIII-digested total DNA (100 ng) from individual root cultures (numbers are shown on top, lane C is R1000 control transformant) were electrophoresed on a 1% agarose gel, transferred to a nylon membrane, and hybridized with a ^{32}P-labeled CF3 cDNA. The sizes of standard mol wt markers (×1000) in lane M are shown on the left, and those of the observed *Hin*dIII fragments (×1000) are shown on the right. The levels of 18:2 and 18:3 as percent of total fatty acids in the different roots are shown below each lane.

similarity of 81% and identity of 66%. It also revealed an approximately 63-amino acid N-terminal extension (Fig. 2). The N-terminal Met is the only Met in the N-terminal extension and the extension has several characteristics of transit peptides of nuclear-encoded chloroplast proteins. These include a high content of hydroxylated residues (the sequence is 21% Ser), a low content of acidic residues (only one residue each of Glu and Asp compared with eight basic residues), and the highly conserved N-terminal dipeptide Met-Ala (de Boer and Weisbeek, 1991; von Heinji and Nishikawa, 1991).

Northern analyses of poly(A)$^+$ RNA from wild-type *Arabidopsis* leaf, seedling, and cultured root tissues showed that the mRNAs corresponding to CF3 and CFD cDNAs were differentially accumulated in leaf and root tissues. CF3 mRNA was about five times more abundant in leaf than in root tissues. CFD mRNA, in contrast, accumulated at a high level in leaf and seedling tissues and at only a trace level in root tissue. In addition, in leaf tissue CFD mRNA was about three times more abundant than CF3 mRNA, whereas in root tissue CF3 mRNA was much more abundant than CFD mRNA (Fig. 3).

CFD cDNA was hybridized to genomic DNA from *A. thaliana* (ecotype Wassileskija and marker line W100 ecotype Landesberg background) digested with *Eco*RI. A restriction fragment length polymorphism was identified and mapped as previously described (Reiter et al., 1992). A single genetic locus corresponding to this cDNA was positioned on the upper arm of chromosome 3 at a position 8 ± 2 cM proximal to cosmid c3838 restriction fragment length polymorphism marker, 9 ± 2 cM distal to the λ AT228 marker, and 39 ± 7 cM distal to the glabrous (*gl-1*) locus (Reiter et al., 1992).

Soybean and Rapeseed cDNAs Encoding Homologs of *Arabidopsis* CF3 and CFD cDNAs

cDNA libraries representing the mRNA population of soybean and rapeseed cotyledons actively engaged in oil biosynthesis were screened at a low stringency for cDNAs related to *Arabidopsis* CF3 and CFD cDNAs. The rapeseed library was screened with ^{32}P-labeled CF3 and CFD cDNAs. Several cross-hybridizing clones were purified in each case and subjected to nucleotide sequence analyses. Rapeseed cDNA, designated BN3, was 1336 bp and contained a large open reading frame that encodes a 377-amino acid polypeptide, designated C3. Rapeseed cDNA, designated BND, contained a 1416-bp sequence with an incomplete open reading frame that encodes a 404-amino acid polypeptide, designated Cp. Comparison of the amino acid sequences of rapeseed polypeptides C3 and Cp and *Arabidopsis* polypeptides A3 and Ap showed that C3 and Cp polypeptides are homologs of the *Arabidopsis* microsomal and putative plastid ω-3 desaturases, respectively. This conclusion is based on percent identity at the amino acid level (Table IV) and the presence or absence of an N-terminal amino acid sequence (Fig. 2), even though the deduced amino acid in BND cDNA contains only part of the putative transit peptide sequence.

The soybean cDNA library was screened with ^{32}P-labeled CF3 cDNA and one of the purified clones was shown to contain a cDNA insert, designated GM3, with a large open reading frame that encodes a 380-amino acid polypeptide, designated S3. A 1-kb *Hha*I fragment of GM3 cDNA was isolated and used to rescreen the soybean cDNA library at low stringency. This resulted in the isolation of a distinct cDNA, designated GMD, that contained an open reading frame that encodes a 454-amino acid polypeptide, designated Sp. Comparison of the soybean polypetides S3 and Sp with rapeseed polypeptides C3 and Cp and *Arabidopsis* polypep-

Table IV. *Percent identity at the amino acid level between different higher plant ω-3 fatty acid desaturase homologs*

The nomenclature of the homologs is the same as in Figure 2.

	Ap	C3	Cp	S3	Sp
A3	66	93	66	68	67
Ap		67	90	67	69
C3			68	68	68
Cp				68	74
S3					68

tides A3 and Ap showed that S3 and Sp polypeptides are homologs of the *Arabidopsis* microsomal and putative plastid ω-3 desaturases, respectively. This conclusion is based largely on the presence or absence of an N-terminal amino acid sequence (Fig. 2) but also partly on the percent identity at the amino acid level (Table IV). The N-terminal sequence in the deduced amino acid sequence of GMD cDNA is authentic, because the open reading frame in GMD cDNA has an in-frame termination codon 5′ to the initiation codon and its deduced amino acid sequence begins at the only Met residue in it. Since it also shares characteristics of transit peptides of nuclear-encoded chloroplast proteins, including the N-terminal dipeptide Met-Ala, we deduce that it is the transit peptide.

DISCUSSION

Although T-DNA tagging has been useful in cloning plant genes, the genes that have been isolated have been those that affect visible, easily scored multigenic traits (Feldmann, 1991). This report provides support for the utility of T-DNA tagging in cloning an arbitrary gene for which there is only a specific assay for gene function. Statistical analysis would suggest that, assuming random insertion of the T-DNA into the *Arabidopsis* genome (100 Mbp), screening the present population of approximately 12,000 *Arabidopsis* transformants containing approximately 16,000 T-DNA inserts (Feldmann, 1991) would provide a 30 to 40% chance of uncovering a mutant at an arbitrary locus of 2.5 to 3 kb. If, as some surmise (Walden et al., 1991), the T-DNA insertion is biased toward transcriptionally active regions of the genome, then the probability of uncovering active genes increases correspondingly.

The identification of the polypeptide encoded by CF3 cDNA as the microsomal ω-3 fatty acid desaturase is based on its complementation of the mutation in line 3707. Expression of the *Arabidopsis* enzyme in line 3707, under the control of a seed-specific promoter, resulted in a 12-fold increase in 18:3 content when compared with the untransformed mutant 3707, and a 2-fold increase in 18:3 content when compared with the wild-type control. Overexpression of the *Arabidopsis* enzyme in carrot hairy roots resulted in a more than 7-fold increase in the 18:3 content, and almost all endogenous 18:2 was converted to 18:3. Overexpression of the rapeseed enzyme in wild-type *Arabidopsis* roots was previously reported to result in a 1.6-fold increase in 18:3 content (Arondel et al., 1992). Thus, the reaction catalyzed by ω-3 desaturase appears to be a rate-limiting step in the biosynthesis of 18:3 in *Arabidopsis* seeds as well as in *Arabidopsis* roots and carrot hairy roots. This observation is supported by genetic studies with *fad* 3 mutants that indicate gene dosage-dependence of the *fad* 3 phenotype (Lemieux et al., 1990). If, as seems likely, the ω-3 desaturase enzyme is also rate limiting in agronomically important oilseeds such as rapeseed or soybean, then the alteration of the 18:3 content in the triacylglycerols of these plants by transgenic approaches should prove practicable.

Mutants of *Arabidopsis* with specific alterations in membrane lipid composition have provided considerable information about the effects of lipid structure on membrane function (Somerville and Browse, 1991). However, these mutants have invariably exhibited decreases in unsaturation relative to wild-type plants. The isolation of the ω-3 desaturase gene and the demonstration that overexpression of its coding sequence can result in very high 18:3 content in transgenic plants will now enable the study of the physiology and cell biology of plants in which the levels of membrane unsaturation are higher than normal.

CFD cDNA, which was isolated using CF3 cDNA as a hybridization probe at low stringency, encodes polypeptide Ap, which is a structural homolog of the microsomal fatty acid desaturase, but with an N-terminal extension (Fig. 2). This N-terminal sequence is deduced to be a transit peptide because it shares several characteristics of transit peptides of nuclear-encoded chloroplast proteins. These include a high content of hydroxylated residues, a low content of acidic residues, and the N-terminal dipeptide Met-Ala (de Boer and Weisbeek, 1991; von Heinji and Nishikawa, 1991). In addition, it is co-linear with, and shares limited homology to, the deduced transit peptide described for the soybean putative plastid ω-3 desaturase. mRNA corresponding to CFD cDNA accumulates at very high levels in leaf but not in root tissue (Fig. 3). Finally, CFD cDNA maps 39 ± 7 cM distal to the *gl-1* locus. Two plastid fatty acid desaturation mutations, *fad D* and *fad B*, were mapped 40 ± 6 and 28 ± 6 cM, respectively, distal to the *gl-1* locus (Hugly et al., 1991). Thus, the map position for the gene encoding CFD cDNA is consistent with that of the *Arabidopsis fad D* locus, which controls plastid ω-3 desaturation. Based on the above discussion, we postulate that the CFD cDNA is derived from the *fad D* locus and encodes the plastid ω-3 fatty acid desaturase. This conclusion will be confirmed by the biological expression of the CFD cDNA.

Rapeseed polypeptide C3 was identified as the microsomal ω-3 desaturase by its high (93%) identity at the amino acid level to *Arabidopsis* microsomal ω-3 desaturase. The rapeseed ω-3 desaturase reported in this study had a 96% amino acid sequence identity with the previously reported rapeseed ω-3 desaturase. It seems likely, therefore, that the two rapeseed polypeptides are isozymes. Soybean polypeptide S3 has 68% and 67% identity with *Arabidopsis* microsomal and putative plastid ω-3 desaturases, respectively. Since it lacks the N-terminal extension transit peptide, we postulate that it encodes the microsomal ω-3 desaturase.

Soybean polypeptide Sp contains an N-terminal extension deduced to be a transit peptide. The length of the deduced transit peptide in Sp is similar to that in Ap, the putative plastid ω-3 desaturase of *Arabidopsis*. Although there is little amino acid sequence identity with the *Arabidopsis* transit peptide, the extension has characteristics similar to those of transit peptides of nuclear-encoded chloroplast proteins. Thus, it is likely that soybean polypeptide Sp is a plastid ω-3 fatty acid desaturase. The rapeseed BND cDNA encodes a polypeptide, Cp, that was identified to be the plastid ω-3 desaturase based on a 90% identity to the *Arabidopsis* putative plastid desaturase, but the rapeseed cDNA is incomplete and encodes only a part of a putative transit peptide.

Our identification of the rapeseed and soybean polypeptides is supported by the phylogenetic analysis based on Hein's alignment algorithm (Hein, 1990). This algorithm

assumes that the sequences are related in some way and constructs a phylogeny based on evolutionary parsimony (Fig. 5). The analysis shows that the earliest divergence in ancestral relationships is between the group of sequences we have identified as microsomal ω-3 fatty acid desaturases and the group we have identified as putative plastid ω-3 fatty acid desaturases. Based on these homologies, it is also apparent that the previously unidentified mung bean cDNA (ARG1) (Yamamoto et al., 1992) encodes a mung bean microsomal ω-3 desaturase.

Microsomal ω-3 desaturases from *Arabidopsis* and rapeseed, both in this and the previously published report (Arondel et al., 1992), share the motif of two Lys residues positioned three and five residues from the C terminus that is believed to be sufficient for retention of transmembrane ER proteins (Jackson et al., 1990). This motif is absent from the putative plastid homologs from all three species. However, its significance is unclear because the soybean homolog S3 lacks it altogether, and the mung bean homolog (encoded by ARG1 cDNA) (Yamamoto et al., 1992) shows a Lys-Ser-Lys tripeptide at the C terminus. Additional soybean homologs of polypeptide S3 are being investigated.

Comparison of the deduced amino acid sequences of the different ω-3 desaturase homologs, of both the microsomal and the putative plastid types, shows that they have overall identities of 66% or greater at the amino acid levels (Table IV). It also shows that the percent identity between the microsomal and the putative plastid desaturases within each species is similar to that between the soybean and *Arabidopsis* microsomal (68%) or plastid (69%) homologs.

We were not successful in cloning the microsomal ω-6 fatty acid desaturase from the *Arabidopsis* cDNA library using ^{32}P-labeled CF3 and CFD cDNAs as probes under low stringency hybridization conditions in which CF3 and CFD cDNAs cross-hybridize. This suggests that the microsomal ω-3 desaturase is more closely related to the putative plastid ω-3 desaturase than it is to the microsomal ω-6 desaturase. There is evidence that microsomal desaturations use phosphatidylcholine as the lipid substrate and Cyt b_5 as the immediate electron donor (Smith et al., 1990), whereas plastid desaturations (Schmidt and Heinz, 1990) and cyanobacterial desaturations (Wada et al., 1993) use galactolipids as the lipid substrate and reduced Fd as the electron donor. Because microsomal and plastid desaturations use different lipid substrates and immediate electron donors, the high degree of similarity between the primary structures of the microsomal ω-3 desaturase and its putative plastid homolog suggests a conserved enzyme mechanism and common structural motifs that recognize the 18:2 fatty acyl moiety.

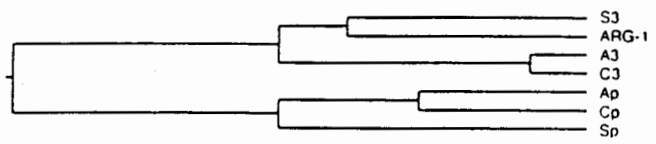


Figure 5. Phylogenetic tree of ω-3 desaturase homologs based on the gene conversion algorithm (Hein, 1990). The length of the branches is proportional to the evolutionary divergence. ARG1 is the auxin-induced mung bean cDNA (Yamamoto et al., 1992); other nomenclature is the same as in Figure 2.

ACKNOWLEDGMENTS

We thank J. Ecker for providing us with the *Arabidopsis* cDNA library, S. Coomber for the *Arabidopsis* wild-type genomic DNA library, and Kevin Ripp for the soybean cDNA library. We thank Florence Garlick, Jingrui Wu, and many others for help in screening the T-DNA population.

Received April 27, 1993; accepted June 21, 1993.
Copyright Clearance Center: 0032-0889/93/103/0467/10.
The GenBank accession numbers for the cDNA sequences reported in this article are L22961, L22931, L22963, L22962, L22965, and L22964 for CF3, CFD, BN3, BND, GM3, and GMD, respectively.

LITERATURE CITED

Arondel V, Lemiewx B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992) Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258: 1353–1355

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1991) Current Protocols in Molecular Biology. John Wiley & Sons, New York

Balcells L, Swineburne J, Coupland G (1991) Transposons as tools for the isolation of plant genes. Trends Biotechnol 9: 31–37

Behringer FJ, Medford JI (1992) A plasmid rescue technique for the recovery of plant DNA disrupted by T-DNA insertion. Plant Mol Biol Rep 10: 190–198

Browse J, McCourt PJ, Somerville CR (1986) Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. Anal Biochem 152: 141–145

Browse J, Somerville C (1991) Glycerolipid synthesis: biochemistry and regulation. Annu Rev Plant Physiol Plant Mol Biol 42: 467–506

de Boer AD, Weisbeek PJ (1991) Chloroplast protein topogenesis: import, sorting and assembly. Biochim Biophys Acta 1071: 221–253

Doyle J, Schuler M, Godette WD, Zenger V, Beachy RN, Slightom JL (1986) The glysolated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. J Biol Chem 261: 9228–9238

Errampalli D, Patton D, Castle L, Mickelson L, Hanson K, Schnall J, Feldmann K, Meinke D (1991) Embryonic lethals and T-DNA insertional mutagenesis in *Arabidopsis*. Plant Cell 3: 149–157

Feinberg AP, Volgestein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13

Feldmann KA (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. Plant J 1: 71–82

Feldmann KA, Marks MD (1987) *Agrobacterium*-mediated transformation of germination seeds of *Arabidopsis thaliana*: a non-tissue culture approach. Mol Gen Genet 208: 1–9

Feldmann KA, Marks MD, Christianson ML, Quatrano RS (1989) A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. Science 243: 1351–1354

Grimes HD, Overvoorde PJ, Ripp K, Franceschi VR, Hitz WD (1992) A 62-kD sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport. Plant Cell 4: 1561–1574

Hein J (1990) Reconstructing evolution of sequences subject to recombination using parsimony. Math Biosci 98: 185–200

Hoekema A, Hirsch PR, Hooykaas PJJ, Schilperoort RA (1983) A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature 303: 179–303

Holsters M, de Waele D, Depicker A, Messens E, Van Montagu M, Schell J (1978) Transfection and transformation of *A. tumefaciens*. Mol Gen Genet 163: 181–187

Hugly S, Kunst L, Somerville C (1991) Linkage relationships of mutations that affect fatty acid composition in *Arabidopsis*. J Hered 82: 4321

Jackson MR, Nilsson T, Peterson PA (1990) Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. EMBO J 9: 3153–3162

James DW Jr, Dooner HK (1990) Isolation of EMS-induced mutants in *Arabidopsis* altered in seed composition. Theor Appl Genet 80: 241–245

Kamalay JC, Goldberg RB (1980) Regulation of structural gene expression in tobacco. Cell 19: 935–946

Kieber JJ, Rothenberg M, Roman G, Feldmann KA, Ecker JR (1993) *CTR1*, a negative regulator of the ethylene response pathway in *Arabidopsis thaliana*, encodes a member of the Raf family of protein kinases. Cell 72: 427–441

Lemieux BMM, Somerville C, Browse J (1990) Mutants of *Arabidopsis* with alterations in seed lipid fatty acid composition. Theor Appl Genet 80: 234–240

Moore L, Warren G, Strobel G (1979) Involvement of a plasmid in the hairy root disease of plants caused by *Agrobacterium* rhizogenes. Plasmid 2: 617–626

Nairn CJ, Winesett L, Ferl RJ (1988) Nucleotide sequence of an actin gene from *Arabidopsis thaliana*. Gene 65: 247–257

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48: 443–453

Petit A, Berkaloff A, Tempe J (1986) Multiple transformations of plant cells by *Agrobacterium* may be responsible for the complex organization of T-DNA in crown gall and hairy roots. Mol Gen Genet 202: 388–393

Reiter SR, Williams JG, Feldmann KA, Rafalski JA, Tingey SV, Scolnick PA (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci USA 89: 1477–1481

Rerie WG, Whitecross M, Higgins TGV (1991) Developmental and environmental regulation of pea legumin genes in transgenic tobacco. Mol Gen Genet 225: 148–157

Russell SH, Hoopes JL, Odell JT (1992) Directed excision of a transgene from the plant genome. Mol Gen Genet 234: 49–59

Schmidt H, Heinz E (1990) Desaturation of oleoyl groups in envelope membranes from spinach chloroplasts. Proc Natl Acad Sci USA 87: 9477–9480

Shanklin J, Somerville C (1991) Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proc Natl Acad Sci USA 88: 2510–2514

Slightom JL, Sun SM, Hall TC (1983) Complete nucleotide sequence of a French bean storage protein gene: phaseolin. Proc Natl Acad Sci USA 80: 1897–1901

Smith MA, Cross AR, Jones OTG, Griffiths WT, Stymne S, Stobart K (1990) Electron-transport components of the 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine delta-12-desaturase in microsomal preparations from developing safflower (*Carthamus tinctorius*) cotyledons. Biochem J 272: 23–29

Somerville CR, Browse J (1991) Plant lipids: metabolism, mutants, and membranes. Science 252: 80–87

Stukey JE, McDonough VM, Martin CE (1990) The *OLE1* gene of *Saccharomyces cerevisiae* encodes the delta-9 fatty acid desaturase and can be functionally replaced by the rat stearoyl-CoA desaturase gene. J Biol Chem 265: 20144–20149

Thiede MA, Ozols J, Strittmatter P (1986) Construction and sequence of cDNA for rat liver stearyl-CoA desaturase. J Biol Chem 261: 13230–13235

Thompson GA, Scherer DE, Foxall-Van Aken S, Kenny JW, Young HL, Shintani DK, Kridl JC, Knauf VC (1991) Primary structures of the precursor and mature forms of stearoyl-acyl-carrier-protein desaturase from safflower embryos and requirement of ferredoxin for enzyme activity. Proc Natl Acad Sci USA 88: 2578–2582

von Heinji G, Nishikawa K (1991) Chloroplast transit peptides: the perfect random coil? FEBS Lett 278: 1–3

Wada H, Gombos Z, Murata N (1990) Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation. Nature 347: 200–203

Wada H, Schmidt H, Heinz E, Murata N (1993) In vitro ferredoxin-dependent desaturation of fatty acids in cyanobacterial thylakoid membranes. J Bacteriol 175: 544–547

Walden R, Hayashi H, Schell J (1991) T-DNA as a gene tag. Plant J 1: 281–288

Yadav N, Vanderleyden J, Bennett DR, Barnes WM, Chilton M-D (1982) Short direct repeats flank the T-DNA on a nopaline Ti-plasmid. Proc Natl Acad Sci USA 79: 6322–6326

Yamamoto KT, Mori H, Imaseki H (1992) Novel mRNA sequences induced by indole-3-acetic acid in sections of elongating hypocotyls of mung bean (*Vigna radiata*). Plant Cell Physiol 33: 13–20

Mol Gen Genet (1988) 212:105-111

LOCKE

MGG

General organization of the genes specifically involved in the diaminopimelate-lysine biosynthetic pathway of *Corynebacterium glutamicum*

Patrice Yeh[1,*], Armand Michel Sicard[2], and Anthony J. Sinskey[1]

[1] Department of Applied Biological Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Centre de Recherche de Biochimie et de Genetique Cellulaires du C.N.P.S., 118 route de Narbonne, F-31062 Toulouse Cedex, France

Summary. We utilized diaminopimelate-lysine mutants of *Escherichia coli* K12 to clone the genes specifically involved in the *Corynebacterium glutamicum* diaminopimelate-lysine anabolic pathway. From a cosmid genomic bank of *C. glutamicum* strain AS019, we isolated cosmids pSM71, pSM61 and pSM531, that are respectively able to complement *dapA*/*dapB*, *dapD*, and *lysA* mutants of *E. coli*. DNA hybridization analysis indicates that these complementing genes are located on the chromosome of *C. glutamicum* in at least three separate transcription units. Subcloning of parental cosmids in *dapA*, *dapD*, and *lysA* mutants of *E. coli* localized these genes, respectively, within 1.4, 3.4, and 1.8 kb fragments, cloned in an *E. coli*/*C. glutamicum* shuttle vector. Enzymatic analysis in *C. glutamicum* identified the *dapA*-complementing gene as L-2,3-dihydrodipicolinate synthetase (*dapA*), and the *lysA*-complementing gene as *meso*-diaminopimelate decarboxylase (*lysA*). In contrast, complementation of *E. coli dapD8*, presumably lacking L-Δ^1-tetrahydrodipicolinate synthetase (*dapD*), led us to clone a diaminopimelate-lysine anabolic gene of *C. glutamicum* which does not exist in *E. coli*: *meso*-diaminopimelate dehydrogenase. Although *meso*-diaminopimelate is crucial in lysine formation and in cell wall biosynthesis, expression of the genomic copies of the cloned genes, which encode activities involved at key branching points of the diaminopimelate-lysine pathway of *C. glutamicum*, appears constitutive with regard to the addition of diaminopimelate and/or lysine during cell growth.

Key words: *Corynebacterium glutamicum* – Diaminopimelate-lysine anabolic pathway – Heterologous complementation – Homologous expression

Introduction

Coryneform bacteria of the genera *Corynebacterium* and *Brevibacterium* comprise a large variety of gram-positive, non sporulating, rod- or club-shaped bacteria, widely dispersed in different ecosystems. They include both animal (*C. diptheriae*) and plant pathogens (*C. fascians*, *C. michiganese*), as well as non-pathogens of industrial interest used mainly for the production of L-amino acids including lysine and glutamate (*C. glutamicum*, *C. lilium*, *B. flavum*, *B. lactofermentum*).

Despite their economic importance, the in vitro DNA characterization of coryneform genes is still rather limited. Using a direct shotgun experiment, Follettie and Sinskey (1986) recently reported the cloning and nucleotide sequence of the *pheA* gene of *C. glutamicum*, and Ozaki et al. (1985) cloned the same gene from a different strain of this organism, providing valuable insights into our general knowledge of bacterial evolution (Ozaki et al. 1985; Follettie and Sinskey 1986). Although *meso*-diaminopimelate (*meso*-DAP) is a crucial component of the cell wall of *C. glutamicum* (Keddie and Cure 1978; Pitcher 1983), we were unable to isolate the corresponding auxotrophs of this organism and thus, cloning of the genes specifically involved in *meso*-DAP synthesis could not be achieved through a direct cloning procedure in *C. glutamicum*.

Complementation of *Escherichia coli* auxotrophs with coryneform genomic DNA has been achieved with *B. lactofermentum* DNA (Marquez et al. 1985; Del Real et al. 1985; Mateos et al. 1987), and in this laboratory with *C. glutamicum* DNA, as previously reviewed (Batt et al. 1985). It now appears that, due to the differences which might exist between the amino acid anabolic pathways of *E. coli* and *C. glutamicum* (see below), or *B. lactofermentum*, complementation of a well-characterized mutant of *E. coli* with coryneform genomic information does not necessarily result in the cloning of a coryneform activity equivalent to that lacking in *E. coli*. This observation points out the limits of such a heterologous approach. In addition, it is also necessary to overcome the restriction-modification system of coryneform bacteria for further studies of the genetic expression of the cloned genes back in their original environment.

Lysine biosynthesis proceeds in bacteria through the "DAP pathway", as opposed to the "dipicolinate pathway" present in some eucaryotic cells. As shown in Fig. 1, the presumptive DAP-lysine anabolic pathway of *C. glutamicum* shows major variations from that described and extensively studied in *E. coli* K12 (Stragier et al. 1983a; Patte 1983; Bouvier et al. 1984; Richaud C et al. 1984; Martin et al. 1986; Richaud F et al. 1986; Richaud C et al. 1987). First, the occurrence of the reversible *meso*-DAP dehydrogenase, reported in some gram-positive organisms (Misono et al. 1979), generates a partial bypass of the DAP-lysine anabolic pathway of *C. glutamicum* (Ishino et al. 1984). Second, the regulation of the entire pathway appears

* *Present address:* Genetica (Rhône-Poulenc S.A.), 160 quai de Polangis, F-94340 Joinville-le-pont, France

Offprint requests to: P. Yeh

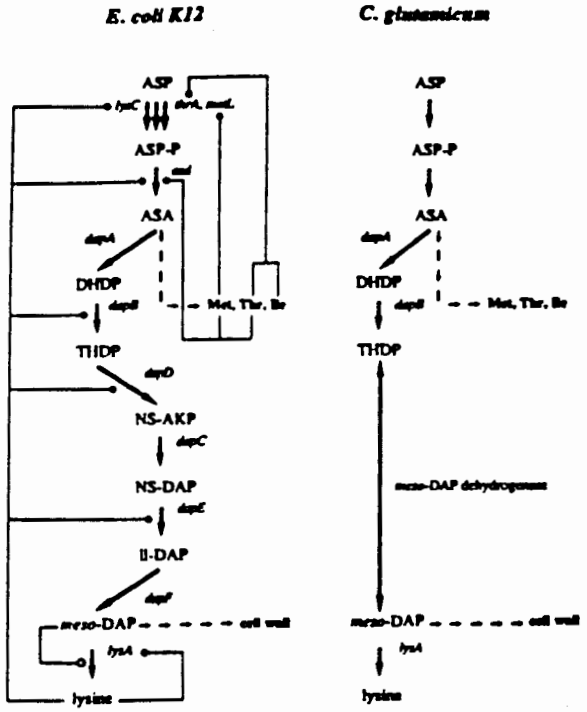


Fig. 1. Synthesis of *meso*-diaminopimelate (*meso*-DAP) and lysine in *Escherichia coli* K12 and *Corynebacterium glutamicum*. Abbreviations used: ASP, L-aspartate; ASP-P, L-aspartyl-phosphate; ASA, L-aspartate semialdehyde; DHDP, L-2,3-dihydrodipicolinate; THDP, L-Δ^1-tetrahydrodipicolinate; NS-AKP, L-N-succinyl-2-amino-6-ketopimelate; NS-DAP, LL-N-succinyl-2,6-diaminopimelate; *ll*-DAP, LL-diaminopimelate; *meso*-DAP, *meso*-diaminopimelate; Met, L-methionine; Thr, L-threonine; Ile, L-isoleucine. •, gene repression mediated by lysine; o, gene induction mediated by *meso*-DAP. For clarity, the allosteric inhibition of the enzymes has not been represented, and the pathways leading to Met, Thr, Ile, and cell wall biosynthesis have not been detailed (*consecutive arrows*)

to be different in *C. glutamicum* since a unique aspartokinase, inhibited through a concerted mechanism, has been characterized in *B. flavum* (Shiio and Miyajima 1969), which is quite different from the three independently regulated isozymic aspartokinases of *E. coli* K12 (*lysC*, *metL*, and *thrA*).

Since most previous work on the DAP-lysine anabolic pathway of coryneform bacteria has focused on the characterization of the enzymes involved in this pathway (Tosaka and Takinami 1978; Ishino et al. 1984), and on the study of feedback regulatory features of key enzymes (Shiio and Miyajima 1969), very little is known about the regulation of these genes at the transcriptional level. In fact, to our knowledge, none of the DAP-lysine anabolic genes of *C. glutamicum* has ever been reported to be inducible or repressible. In contrast, numerous lysine-mediated transcriptional regulatory mechanisms, probably involving DNA-protein interactions, have been described for most of the genes specifically involved in the DAP-lysine biosynthesis of *E. coli* K12 (Boy et al. 1979; Patte 1983; Bouvier et al. 1984; Richaud C et al. 1984). Interestingly, these genes are scattered along the chromosome in separate transcription units (Bukhari and Taylor 1971a; Bachmann 1983). In addition, a model has been proposed to explain the complex DAP-mediated induction of the expression of the *lysA* gene, via the product of the *lysR* regulatory gene located immediately upstream from the structural gene and transcribed divergently (Stragier et al. 1983a, b; Stragier and Patte 1983).

Genetic and enzymatic analyses are presented here which indicate that, for *C. glutamicum* as well, the genes specifically involved in DAP-lysine biosynthesis are scattered along the chromosome in at least three separate transcription units.

Materials and methods

Bacterial strains, plasmids, media and culture conditions. Bacterial strains and plasmids used in this study are listed in Table 1. *E. coli* K12 strains AT997 (*dapA*, see text), AT999 (*dapB17*), AT986 (*dapD8*) and AT2453 (*lysA22*) are from Dr. B. Bachmann, Dept. of Human Genetics, Yale University School of Medicine, Conn., USA. *E. coli* strain RLA81 (*lysR*) is from Dr. P. Stragier, Université Paris-Sud, Orsay, France.

C. glutamicum strain AS019 is a spontaneous rifampicin-resistant (Rif^r) mutant of a glutamate-producing strain (ATCC 13059), and thus, is very likely to harbor a wild-type DAP-lysine anabolic pathway. Strain AS376 is one of our ten lysine-requiring mutants of strain AS019. These lysine auxotrophs, which do not require *meso*-DAP for their growth in minimal medium (MMCG, see below), were obtained after 1-methyl-3-nitro-1-nitrosoguanidine mutagenesis, according to the procedure of Follettie and Sinskey (1986).

The cosmid genomic bank of *C. glutamicum* was constructed by M.T. Follettie; its utilization has been previously reviewed (Batt et al. 1985). Shuttle vector pWS124 was constructed by W.G. Shanabruch (unpublished results); its restriction map has been reported elsewhere (Follettie and Sinskey 1986).

LB medium (Maniatis et al. 1982), supplemented with 0.2% glucose, was used as rich medium for both *E. coli* and *C. glutamicum*. M9 medium (Maniatis et al. 1982) was used as minimal medium for *E. coli*, while MMCG was used as minimal medium for *C. glutamicum* (Follettie and Sinskey 1986). L-lysine (Sigma), DAP (Sigma, a mixture of *meso*, *ll* and *dd* isomers), and 14 g/l agar were added when appropriate. L-threonine and L-leucine were also from Sigma and used as indicated in the text. Ampicillin (Ap, 50 mg/l) and kanamycin (Km, 20 mg/l) were used when appropriate for *E.coli* and *C. glutamicum* cultures, respectively.

E. coli and *C. glutamicum* were grown under aerobic conditions at 37° and 30° C, respectively.

Isolation of DNA and transformation procedures. Plasmid and genomic DNA from *C. glutamicum* were isolated as previously described (Follettie and Sinskey 1986). *C. glutamicum* was transformed via polyethylene glycol-mediated uptake of DNA into lysozyme-treated cells, according to the procedure of Yoshihama et al. (1985).

Plasmid DNA from *E. coli* was extracted according to the procedure of Birnboim and Doly (1979). *E. coli* mutants were transformed by using the calcium chloride method

Table 1. Bacterial strains and plasmids used in this study

Strains	Genotype or relevant characteristics	Source or reference
Escherichia coli		
AT997	*dapA* mutant of *E. coli* K12 (see text) (previous genotype: *dapC15 thi-1 relA1 spoT1*)	This work Bukhari and Taylor (1971a)
AT999	*dapB17::Mu thi-1 relA1 spoT1*	Bukhari and Taylor (1971a)
AT986	*dapD8 thi-1 relA1 spoT1*	Bukhari and Taylor (1971a)
AT2453	*lysA22 thi-1 relA1 spoT1*	Bukhari and Taylor (1971a, b)
RLA81	*araD139 Δ(lacIPOZYA)U169 StrA' thiA1106 lysR::Mucts 1101*	Stragier et al. (1983a)
Corynebacterium glutamicum		
AS019	Spontaneous Rif' mutant of ATCC 13059	This laboratory
AS376	Lysine auxotroph isolated from AS019	This laboratory
Plasmids		
Bank	*Sau*3A large insert of AS019 genomic DNA cloned in the *Bam*HI site of cosmid pHC79	M.T. Follettie, this laboratory
pWS124	*C. glutamicum*/*E. coli* shuttle vector Km', Ap'	Follettie and Sinskey (1986)
pUC8	Ap', *lacP*(polylinker)*lacZ'* replicon ColE1	Vieira and Messing (1982)

of Cohen et al. (1973). Phenotypic complementation of these mutants resulted in their growth on M9 Ap plates after 48 h incubation.

DNA cloning experiments. All restriction enzymes, T4 DNA ligase and Klenow fragment were from New England Biolabs or International Biotechnologies Inc., and they were used as recommended by the suppliers. Calf intestine phosphatase (Boehringer, Mannheim) was used as described in Maniatis et al. (1982). λ DNA cut with *Eco*RI and/or *Hin*dIII (New England Biolabs) was used as a molecular weight standard, as well as a DNA concentration reference, when submitted to horizontal electrophoresis in agarose gels with TBE buffer (Maniatis et al. 1982). Fragment size purification was performed using an electroelution procedure.

DNA hybridization analysis. Cosmids pSM71, pSM61 and pSM531 were digested with restriction endonuclease *Sal*I and labeled with [α^{32}P]dATP (Amersham, 3000 Ci/mmol), by the nick translation procedure of Rigby et al. (1977). These radioactive probes were used in DNA hybridization analysis: each lane of a 0.9% horizontal agarose gel contained 2 μg of *Sal*I-digested genomic DNA from *C. glutamicum* strain AS019, or 10 ng of *Sal*I-restricted cosmid DNA. After electrophoresis, DNA was transferred onto nitrocellulose paper by the procedure of Southern (1975). Prehybridization, hybridization and stringent washing conditions were as described in Maniatis et al. (1982). λ DNA cut with *Hin*dIII was labeled by using a Klenow end-filling procedure as recommended by the supplier (New England Biolabs).

Preparation of crude extracts. *C. glutamicum* crude extracts were prepared under the following standard conditions: cells were grown in 100 ml MMCG, supplemented as indicated in the text, and collected at an OD_{600nm} corresponding to the late exponential growth phase (0.6–0.8). They were washed in the corresponding enzymatic measurement buffer, and sonicated on ice under standard conditions (see below). After centrifugation for 45 min at 12000 rpm in a JA-21 rotor, the supernatants were collected (crude extracts), and immediately assayed for protein concentration and enzymatic activity. Protein concentration was determined using a Bio-Rad assay with bovine serum albumin (Sigma) as standard. Under these conditions protein concentrations of the crude extracts were between 0.5 and 1 mg/ml.

Measurement of enzymatic activities. For a given crude extract, protein concentration and enzymatic activities were measured at least twice and the average value was calculated. At least two different crude extracts were assayed for a given experiment, and the reference crude extract (strain AS019 harboring plasmid pWS124) was always assayed in parallel for relative activity determination. Results are given for one typical experiment.

L-2,3-dihydrodipicolinate (DHDP) synthetase (*dapA*) was assayed by following the 270 nm procedure of Yugari and Gilvarg (1965). Its substrate (L-aspartic semialdehyde) was prepared by ozonolysis of L-allylglycine as previously described (Yugari and Gilvarg 1965). Reaction was initiated with the addition of L-aspartic semialdehyde (approximately 1 mM), in a final volume of 1 ml. Under initial conditions, one unit of activity is defined as the amount of enzyme which increases the OD_{270nm} by one unit per minute.

Meso-DAP dehydrogenase activity of the crude extracts was assayed spectrophotometrically at 25° C by following the reduction of $NADP^+$, essentially as described by Annie et al. (1985). Conditions were slightly altered: in a final volume of 1 ml, the reaction mixture contained the crude extract, 0.1 mM $NADP^+$, 200 mM sodium bicarbonate buffer, pH 10.5, and the reaction was initiated with the addition of DAP at a final concentration of 10 mM. One unit is defined as the amount of enzyme which catalyzes the formation of 1 nmol of NADPH per minute under initial conditions.

Meso-DAP decarboxylase (*lysA*) activity was determined by measuring $^{14}CO_2$ evolution from [1,7-^{14}C]DAP (Amersham, a mixture of *meso*, *ll*, and *dd* isomers, 117 mCi/mmol), essentially as described in Kelland et al. (1985). The assay mixture contained 50 mM potassium phosphate buffer, pH 8, 15 mM DAP (Sigma), 63 nCi of [1,7-^{14}C]DAP, 100 μM pyridoxal 5'-phosphate, 1 mM ethylenediaminetetraacetate, 1 μM β-mercaptoethanol, and the crude extract in a final volume of 1 ml. The reaction, initiated with the addition of enzyme (crude extract), was incu-

bated at 30° C with continual shaking, and the kinetics of [^{14}C]DAP decarboxylation were determined (three to five points). Under initial conditions, one unit is defined as the amount of enzyme that liberates 1 nmol of CO_2 per minute.

Results

Complementation of E. coli *K12 auxotrophs with* C. glutamicum *DNA*

From a cosmid genomic bank of *C. glutamicum* in cosmid pHC79, we isolated cosmids pSM71, pSM61, and pSM531, which are able to complement AT997 (*dapA*, see below)/AT999 (*dapB17*), AT986 (*dapD8*), and AT2453 (*lysA22*) mutants of *E. coli* K12, respectively (Bukhari and Taylor 1971a). Restriction analysis revealed that each of these cosmids harbored a 30 to 40 kb DNA insert (data not shown). The absence of cross-complementation between cosmids pSM71 (*dapA/dapB17*), pSM61 (*dapD8*), pSM531 (*lysA22*), and the three complementation classes they define in *E. coli* K12, suggests that the DAP-lysine anabolic genes of *C. glutamicum* encoded by these cosmids are dispersed along the chromosome in at least three separate transcription units. It is noteworthy that the *E. coli* DAP-lysine auxotrophs used in this experiment also map in different regions of the chromosome (Bukhari and Taylor 1971a; Bachmann 1983).

To our knowledge, the identity of the defective enzymes of these *E. coli* auxotrophs was unknown, with the exception of *E. coli* strain AT2453 (*lysA22*) lacking *meso*-DAP decarboxylase (Bukhari and Taylor 1971b). Surprisingly, *E. coli* AT997 (*dapC15* according to Bukhari and Taylor 1971a) is lacking DHDP synthetase (P. Stragier, personal communication). We confirmed the absence of DHDP synthetase, the *dapA* product, in this *E. coli* mutant (data not shown) and thus, the *dapC15* mutation of Bukhari and Taylor is actually located in the *dapA* locus.

From the genomic bank used in this study, we could not isolate a cosmid able to complement *E. coli* strain RLA81 (*lysR*), lacking the regulatory protein of the *lysA* gene (Stragier et al. 1983a).

Structural relationship between cosmids pSM71, pSM61 and pSM531

DNA hybridization analysis demonstrated the *C. glutamicum* genomic origin of the inserts carried by cosmids pSM531 (Fig. 2), pSM71 and pSM61 (data not shown).

Since the cosmid genomic bank of *C. glutamicum* was built in the *Bam*HI site of parental cosmid pHC79 (M.T. Follettie, personal communication), two unrelated recombinant cosmids should harbor two, but no more, related *Sal*I restriction fragments. Therefore, the autoradiogram given in Fig. 2, for which we used cosmid pSM531 (*lysA22*) as a probe, confirmed the absence of a physical relationship between the genomic DNA carried by cosmid pSM531, and the *Sau*3A genomic inserts carried by cosmids pSM71 (lane E) and pSM61 (lane D). The absence of a detectable relationship between these inserts was also confirmed by using cosmid pSM71 (*dapA/dapB17*), or pSM61 (*dapD8*), as a probe (data not shown).

From Fig. 2, it is also noteworthy that the immediate environment of the *lysA*-complementing gene from cosmid pSM531 (3.8 and 4.0 kb *Sal*I restriction fragments, Yeh et al. 1988) is found in the chromosome of *C. glutamicum* strain AS019 without detectable rearrangement (lanes A and B).

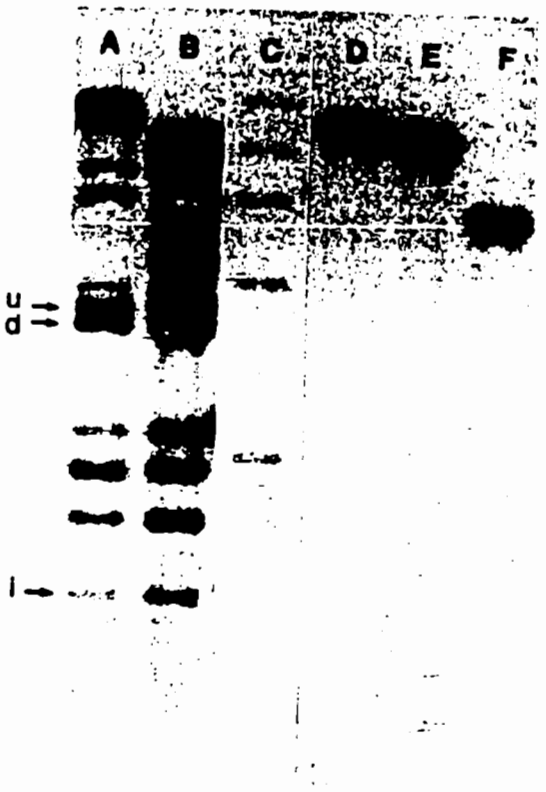


Fig. 2. DNA hybridization analysis of cosmid pSM531 (*lysA*). The probe (cosmid pSM531) was cut with *Sal*I, labeled as indicated in the text, and hybridized to *Corynebacterium glutamicum* genomic DNA cut with *Sal*I (lane A), itself (lane B), cosmid pSM61 (*dapD*) cut with *Sal*I (lane D), cosmid pSM71 (*dapA/dapB*) cut with *Sal*I (lane E), and parental cosmid pHC79 cut with *Sal*I (lane F). λ DNA cut with *Hin*dIII, labeled as indicated in Materials and methods, was used as molecular weight standard (lane C). The *Sal*I restriction fragments spanning the *lysA* gene are indicated as follows: u = upstream 4.0 kb *Sal*I fragment; d, downstream 3.8 kb *Sal*I fragment; i, internal 1.4 kb *Sal*I fragment (see text and Yeh et al. 1988)

Subcloning of the cosmids, and identification of the cloned genes

The complementing genes encoded by cosmids pSM71 (*dapA/dapB17*), pSM61 (*dapD8*) and pSM531 (*lysA22*) were subcloned in *E. coli* strains AT997 (*dapA*), AT986 (*dapD8*), and AT2453 (*lysA22*) respectively. This functional subcloning strategy localized these *dapA*-, *dapD8*- and *lysA22*-complementing genes within small DNA fragments inserted in pWS124, an *E. coli*/*C. glutamicum* shuttle vector stably maintained in both organisms (W.G. Shanabruch, unpublished results).

Table 2. Identification of the cloned genes and homologous expression level

	DHDP synthetase		DAP dehydrogenase		DAP decarboxylase	
	spec. act.	rel. act.	spec. act.	rel. act.	spec. act.	rel. act.
pWS124	0.21 ± 0.05	1.00	130 ± 15	1.00	28 ± 3	1.00
pRC3	1.5 ± 0.3	7 ± 3	NT		NT	
pDH24	NT		2500 ± 300	18 ± 3	NT	
pRS7	NT		NT		52 ± 6	1.9 ± 0.3

Corynebacterium glutamicum strain AS019 harboring plasmids pWS124, pRC3, pDH24, or pRS7 was grown in 100 ml MMCG, 20 mg/l kanamycin and specific activities (spec. act.) were determined and expressed in units per milligram total protein as defined in Materials and methods. rel. act., relative activities. NT, not tested. DHDP, L-2,3-dihydrodipicolinate; DAP, diaminopimelate

From bifunctional cosmid pSM71 (*dapA/dapB17*), we obtained plasmid pRC3 containing a 1.4 kb *Sau*3A insert (previously engineered, via a pUC8 vector, as a 1.4 kb *Bam*HI-*Sal*I fragment), cloned as a *Bam*HI-*Sal*I fragment in vector pWS124. Plasmid pRC3 (*dapA*) is not able to complement strain AT999 (*dapB17*); this is obviously due to the small size (1.4 kb) of the insert carried by this plasmid. From cosmid pSM61 (*dapD8*), we obtained plasmid pDH24 containing a 3.4 kb *Sau*3A fragment in the *Bam*HI site of vector pWS124, while from cosmid pSM531 (*lysA22*), we obtained plasmid pRS7 which contains a 1.8 kb *Sau*3A fragment cloned in the same restriction site of this shuttle vector.

Plasmids pRC3 (*dapA*), pDH24 (*dapD8*) and pRS7 (*lysA22*) were introduced back into *C. glutamicum* strain AS019 according to the procedure of Yoshihama et al. (1985), and their restriction patterns were identical in both organisms (data not shown). Enzymatic analysis of the cloned activities in *C. glutamicum* demonstrated their identity (Table 2): a significant increase in DHDP synthetase specific activity (four- to tenfold) was achieved with plasmid pRC3 (*dapA*-complementing plasmid), which led us to conclude that it encodes the corresponding enzyme (DHDP synthetase). Thus complementation of strain AT997 (*dapA*) was due to the ability of the *C. glutamicum dapA* gene to substitute its counterpart lacking in *E. coli*. A similar conclusion was reached with plasmid pRS7 (*lysA22*): its introduction in *C. glutamicum* strain AS019 led to a *significant* increase in *meso*-DAP decarboxylase specific activity (although only twofold overexpression) and thus, complementation of *E. coli* AT2453 (*lysA22*) was also due to sufficient expression of a *C. glutamicum* gene equivalent to the one lacking in this auxotroph (*lysA*). Furthermore, plasmid pRS7 is able to complement all of our ten lysine auxotrophs of *C. glutamicum* (e.g. strain AS376), providing genetic evidence for the identity of the cloned gene (refer to Fig. 1). On the basis of the *E. coli* system (Stragier et al. 1983a), the complementation ability to the *C. glutamicum lysA* gene does not provide genetic evidence for the phenotype of the *C. glutamicum* lysine auxotrophs used in this experiment.

Interestingly, the heterologous complementation of *E. coli* strain AT986 (*dapD8*) led us to clone a DAP-lysine anabolic gene of *C. glutamicum* which does not exist in *E. coli*: *meso*-DAP dehydrogenase (Misono et al. 1979). As shown in Table 2, 18-fold overexpression of this singular gram-positive activity was achieved after reintroduction of plasmid pDH24 (*dapD8*) into *C. glutamicum* strain AS019.

Table 3. Expression studies under defined growth conditions

Growth additions*	DHDP synthetase	DAP dehydrogenase	DAP decarboxylase
None	0.19 ± 0.05	135 ± 15	30 ± 4
DAP	0.15 ± 0.04	120 ± 14	22 ± 3
Lys	0.23 ± 0.06	130 ± 15	25 ± 3
DAP + Lys	NT	105 ± 12	NT
Thr	0.23 ± 0.06	NT	NT
Leu	0.22 ± 0.06	NT	NT

Results are given as the specific activities of the crude extract from *Corynebacterium glutamicum* strain AS019 harboring plasmid pWS124, expressed in units per milligram total protein (see Materials and methods)

* Cells were grown in 100 ml MMCG, 20 mg/l kanamycin, with the indicated amino acids at a final concentration of 5 mM. DHDP, L-2,3-dihydrodipicolinate; DAP, diaminopimelate; Lys, lysine; Thr, threonine. NT, not tested

Expression of the genes under defined growth conditions

Meso-DAP is involved in cell wall biosynthesis in *C. glutamicum* and in lysine formation (Fig. 1) and thus, regulation mechanisms might exist which could control the intracellular amounts of DAP and lysine in this bacterium. We assayed the expression of the chromosomal DHDP synthetase (*dapA*), *meso*-DAP dehydrogenase, and *meso*-DAP decarboxylase (*lysA*) genes of *C. glutamicum* strain AS019 under defined growth conditions. As shown in Table 3, we found that expression of these genes appears to be independent of the addition of DAP and/or lysine (5 mM) to the growth medium.

In addition, in contrast to what has been observed in a lysine-producing strain of *B. lactofermentum* (Tosaka et al. 1978), expression of the genomic DHDP synthetase (*dapA*) of *C. glutamicum* strain AS019 is not repressed by leucine (Table 3). Furthermore, although the *dapA* locus encodes an enzyme located immediately after the branch point of the pathway leading to DAP-lysine and threonine biosynthesis (Fig. 1), its expression also appears to be constitutive with regard to the addition of threonine (5 mM) during cell growth.

Discussion

By using a heterologous complementation strategy in various DAP or lysine auxotrophs of *E. coli* K12, we isolated

three cosmids, carrying non-overlapping genomic inserts of *C. glutamicum* strain AS019, and encoding DHDP synthetase (*dapA*), *meso*-DAP dehydrogenase (no corresponding locus in *E. coli*), and *meso*-DAP decarboxylase (*lysA*). These genes are distributed along the chromosome of this organism in at least three separate transcription units: *dapA* (and possibly *dapB*, see below), *lysA*, and the genetic locus coding for the singular lysine anabolic bypass reported in some gram-positive organisms, including *C. glutamicum* (Misono et al. 1979; Ishino et al. 1984). Since we have provided evidence that we have cloned the DAP-lysine biosynthetic bypass of *C. glutamicum* strain AS019, and since the emergence of the cryptic nature of this singular enzyme raises interesting questions regarding the evolution of the DAP-lysine anabolic pathway among bacteria, we propose to name this new locus *dapY* by analogy with the DAP biosynthetic genes of *E. coli*.

Furthermore, since DHDP reductase (*dapB*) occurs in *B. lactofermentum* (Tosaka and Takinami 1978), a closely related coryneform bacteria (Minnikin et al. 1978), the product of the *dapB* gene of *C. glutamicum*, present in bifunctional cosmid pSM71 (*dapA/dapB*) and absent in plasmid pRC3 (*dapA*), is probably encoded in the vicinity of the *dapA* locus. Therefore, it is very likely that we have cloned the four structural genes specifically involved in the presumptive DAP-lysine anabolic pathway of *C. glutamicum* AS019: *dapA*, *dapB*, *dapY*, and *lysA* (see Fig. 1).

We focused our study on the genes located at strategic branch points. As shown in Fig. 1, DHDP synthetase (*dapA*) is the first enzyme involved in the specific DAP biosynthetic pathway of *C. glutamicum*, while *meso*-DAP dehydrogenase (*dapY*) and *meso*-DAP decarboxylase (*lysA*) are directly involved, at least in vitro, in *meso*-DAP catabolism (Misono et al. 1979). Interestingly, expression of these genes is constitutive with respect to the addition of L-lysine (5 mM) to the growth medium (Table 3), while in *E. coli* most of the genes involved in DAP-lysine anabolism are repressed by this amino acid (see Fig. 1). In addition, it is noteworthy that expression of these key enzymes also appears to be constitutive with regard to DAP (Table 3).

A possible explanation of these results would be that, in contrast to what has been described in *E. coli* (see Fig. 1), no transcriptional regulation mechanism exists in *C. glutamicum* which can control metabolic flux toward cell wall biosynthesis and lysine formation. It could be possible that the *reversible* pH-dependent *meso*-DAP dehydrogenase of *C. glutamicum* (Ishino et al. 1984) is also involved in vivo in anabolism/catabolism of *meso*-DAP, leading to a modulation of the DAP-lysine pathway at the enzyme level. For example, intracellular pH might change during cell growth of *C. glutamicum*, and therefore could modulate the catalytic properties of this singular enzyme for its substrates: *meso*-DAP and L-Δ^1-tetrahydrodipicolinate (THDP, see Fig. 1).

Alternatively, although we have genetic evidence that lysine is able to enter *C. glutamicum* cells (non-conditional lysine auxotrophs have been isolated), the uptake of this amino acid by this gram-positive organism might not be efficient enough for detectable modulation of putative lysine-mediated regulation of target genes. In addition, since we do not have any genetic evidence that DAP is able to enter *C. glutamicum* cells efficiently (DAP auxotrophs of this organism have not been isolated), we cannot rule out the possibility that, as in *E. coli*, DAP is involved in the regulation of the DAP-lysine anabolic pathway of *C. glutamicum*. This is supported by the fact that DAP is not able to enter *Pseudomonas aeruginosa* efficiently (J.C. Patte, personal communication). This hypothesis is also supported by the fact that in *E. coli*, for which several different DAP auxotrophs have been reported (Bukhari and Taylor 1971a), expression of the cloned *meso*-DAP dehydrogenase of *C. glutamicum* is subject to a significant DAP-mediated repression effect, dependent on the presence of a 1.7 kb fragment adjacent to that encoding the structural activity (Yeh et al., manuscript in preparation), although expression of its genomic copy appears to be insensitive to the addition of this metabolite (Table 3). Additional studies on the uptake of amino acids by corynebacteria are therefore needed for correct interpretation of our regulation data.

The *dapA*, *dapY*, and *lysA* genes of *C. glutamicum* strain AS019 were cloned in plasmid pWS124 and reintroduced into their original genetic environment, leading to overexpression of the cloned activities: approximately 7-fold for DHDP synthetase (*dapA*, plasmid pRC3), 18-fold for *meso*-DAP dehydrogenase (*dapY*, plasmid pDH24, and 2-fold for *meso*-DAP decarboxylase (*lysA*, plasmid pRS7). Since plasmid pWS124 is thought to be present in *C. glutamicum* at approximately 15 copies per cell, the homologous expression levels of DHDP synthetase (approximately 7-fold) and *meso*-DAP dehydrogenase (18-fold) suggest that we have cloned these structural genes under the control of their original promoters. In fact, these values are in good agreement with the expression level of the constitutive *pheA* gene of *C. glutamicum* strain AS019 (6-fold), cloned in the same shuttle vector by Follettie and Sinskey (1986). In contrast, the expression level of *meso*-DAP decarboxylase encoded by plasmid pRS7 is quite low, although significant (2-fold overexpression). The possible reasons for such weak expression have been investigated elsewhere, leading to the full expression of the cloned *meso*-DAP carboxylase (15-fold overexpression, Yeh et al. 1988).

Acknowledgements. This work was supported by a grant from Rhône-Poulenc Sante, France. We thank Barbara Bachmann and Patrick Stragier for providing us with *E. coli* DAP or lysine auxotrophs, Max T. Follettie for kindly providing us with the cosmid genomic bank of *C. glutamicum* and for his constant interest throughout this work, as well as helpful scientific discussions. We also thank Monica M. Palcic for helpful advice in *meso*-DAP decarboxylase measurements, and Benoit Arcangioli, Julian Davies, and Rosemary Sousa for valuable criticisms of this manuscript and helpful suggestions. This work was presented in part at the Fifth International Symposium On The Genetics Of Industrial Microorganisms held in Split, Yugoslavia, 14–20 Sept. 1986.

References

Aanie T, Bartlett M, White PJ (1985) Species of bacillus that make a vegetative peptidoglycan containing lysine lack diaminopimelate epimerase but have diaminopimelate dehydrogenase. J Gen Microbiol 131:2145–2152

Bachmann B (1983) Linkage map of *Escherichia coli* K-12. Ed. 7. Microbiol Rev 47:180–230

Batt CA, Follettie MT, Shin HK, Yeh P, Sinskey AJ (1985) Genetic engineering of coryneform bacteria. Trends Biotechnol 3:305–310

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Bouvier J, Richaud C, Richaud F, Patte JC, Stragier P (1984)

Nucleotide sequence and expression of the *E. coli dapB* gene. J Biol Chem 259:14829–14834

Boy E, Richaud C, Patte JC (1979) Multiple regulation of DAP-decarboxylase synthesis in *E. coli* K-12. FEMS Microbiol Lett 5:287–290

Bukhari AI, Taylor AL (1971a) Genetic analyis diaminopimelic acid- and lysine-requiring mutants of *Escherichia coli*. J Bacteriol 105:844–854

Bukhari AI, Taylor AL (1971b) Mutants of *Escherichia coli* with a growth requirement for either lysine or pyridoxine. J Bacteriol 105:988–998

Cohen SN, Chang ACY, Hsu L (1973) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc Natl Acad Sci USA 69:2110–2114

Del Real G, Aguilar A, Martin JF (1985) Cloning and expression of tryptophan genes from *Brevibacterium lactofermentum* in *Escherichia coli*. Biochem Biophys Res Commun 133:1013–1019

Follettie MT, Sinskey AJ (1986) Molecular cloning and nucleotide sequence of the *C. glutamicum pheA* gene. J Bacteriol 167:695–702

Ishino S, Yamaguchi K, Shirahata K, Araki K (1984) Involvement of *meso*-α,ε-diaminopimelate D-dehydrogenase in lysine biosynthesis in *C. glutamicum*. Agric Biol Chem 48:2557–2560

Keddie RM, Cure GL (1978) Cell wall composition. In: Bousfield IJ, Callaly AG (eds) Coryneform bacteria. Academic Press Inc., London, pp 47–83

Kelland JG, Palcic MM, Pickard MA, Vederas JC (1985) Stereochemistry of lysine formation by *meso*-diaminopimelate decarboxylase from wheat germ: use of ^{1}H-^{13}C NMR shift correlation to detect stereospecific deuterium labelling. Biochemistry 24:3263–3267

Maniatis T, Fritsch FF, Sambrook J (1982) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Marquez G, Fernandez Sousa JM, Sanchez F (1985) Cloning and expression in *E. coli* of genes involved in the lysine pathway of *B. lactofermentum*. J Bacteriol 164:379–383

Martin C, Borne F, Cami B, Patte JC (1986) Autogenous regulation by lysine of the *lysA* gene of *Escherichia coli*. FEMS Microbiol Lett 36:105–108

Mateos LM, Del Real G, Aguilar A, Martin JF (1987) Cloning and expression in *Escherichia coli* of the homoserine kinase (*thrB*) gene from *Brevibacterium lactofermentum*. Mol Gen Genet 206:361–367

Minnikin DE, Goodfellow M, Collins MD (1978) Lipid composition in the classification and identification of coryneform and related taxa. In: Bousfield IJ, Callaly AG (eds) Coryneform bacteria. Academic Press, London, pp 85–160

Misono H, Togawa H, Yamamoto T, Soda K (1979) *Meso*-α,ε-diaminopimelate D-dehydrogenase: distribution and the reaction product. J Bacteriol 137:22–27

Ozaki A, Katsumata R, Oka T, Furuya A (1985) Cloning of the genes concerned in phenylalanine biosynthesis in *Corynebacterium glutamicum* and its application to breeding of a phenylalanine-producing strain. Agric Biol Chem 49:2925–2930

Patte JC (1983) Diaminopimelate and lysine. In: Hermann K, Somerville R (eds) Amino acid biosynthesis and genetic regulation. Addison-Wesley Publishing Co., Reading, MA

Pitcher DG (1983) Deoxyribonucleic acid base composition of *Corynebacterium diphtheriae* and corynebacteria with cell wall type IV. FEMS Microbiol Lett 16:291–295

Richaud C, Richaud F, Martin C, Haziza C, Patte JC (1984) Regulation of expression and nucleotide sequence of the *E. coli dapD* gene. J Biol Chem 259:14824–14828

Richaud C, Higgins W, Mengin-Lecreulx D, Stragier P (1987) Molecular cloning, characterization, and chromosomal localization of *dapF*, the *Escherichia coli* gene for diaminopimelate epimerase. J Bacteriol 169:1454–1459

Richaud F, Richaud C, Ratet P, Patte JC (1986) Chromosmal location and nucleotide sequence of the *E. coli dapA* gene. J Bacteriol 166:297–300

Rigby PW, Dieckmann M, Rhodes C, Berg P (1977) Labelling DNA to high specific activity *in vitro* by nick translation with DNA polymerase I. J Mol Biol 113:237–251

Shiio I, Miyajima R (1969) Concerted inhibition and its reversal by end products of aspartate kinase in *Brevibacterium flavum*. J Biochem 65:849–859

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Stragier P, Patte JC (1983) Regulation of diaminopimelate decarboxylase synthesis in *E. coli*: III. Nucleotide sequence and regulation of the *lysR* gene. J Mol Biol 168:333–350

Stragier P, Richaud F, Borne F, Patte JC (1983a) Regulation of diaminopimelate decarboxylase synthesis in *E. coli*: I. Identification of a *lysR* gene encoding an activator of the *lysA* gene. J Mol Biol 168:307–320

Stragier P, Danos O, Patte JC (1983b) Regulation of diaminopimelate decarboxylase synthesis in *E. coli*: II. Nucleotide sequence of the *lysA* gene and its regulatory region. J Mol Biol 168:321–331

Tosaka O, Takinami K (1978) Pathway and regulation of lysine biosynthesis in *Brevibacterium lactofermentum*. Agric Biol Chem 42:95–100

Tosaka O, Hirakawa H, Takinami K, Hirose Y (1978) Regulation of lysine biosynthesis by leucine in *Brevibacterium lactofermentum*. Agric Biol Chem 42:1501–1506

Vieira J, Messing J (1982) The pUC plasmids, an M13 mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268

Yeh P, Sicard AM, Sinskey AJ (1988) Nucleotide sequence of the *lysA* gene of *Corynebacterium glutamicum* and possible mechanisms for modulation of its expression. Mol Gen Genet 212:112–119

Yoshihama M, Higashino K, Eswara AR, Akedo M, Shanabrush WG, Follettie MT, Walker GC, Sinskey AJ (1985) Cloning vector system for *Corynebacterium glutamicum*. J Bacteriol 162:591–597

Yugari Y, Gilvarg C (1965) The condensation step in diaminopimelate synthesis. J Biol Chem 240:4710–4716

Communicated by J.W. Lengeler

Received July 20, 1987 / December 7, 1987