RESEARCH COMMUNICATION

Cloning of cDNA encoding the bifunctional dehydroquinase·shikimate dehydrogenase of aromatic-amino-acid biosynthesis in *Nicotiana tabacum*

Carol A. BONNER* and Roy A. JENSEN

Department of Microbiology and Cell Science, 3103 McCarty Hall, University of Florida, Gainsville, FL 32611-0100, U.S.A.

Nicotiana tabacum cDNA encoding a bifunctional protein having catalytic domains for dehydroquinase and shikimate dehydrogenase was cloned and sequenced. Complementation of *Escherichia coli aroD* and *aroE* auxotrophs was successful. Amino acid sequencing located the N-terminus of the mature protein. The two catalytic domains exhibited greater amino acid identity with prokaryote homologues than with yeast and fungal homologues.

INTRODUCTION

In higher plants, dehydroquinase (DHQase) (EC 4.2.1.10) and shikimate dehydrogenase (SDH) (EC 1.1.1.25), are functional domains shared by a single protein (Boudet and Lecussan, 1974; Mousdale et al., 1987) of the common pathway of aromatic-amino-acid biosynthesis (steps 3 and 4). In microbial eukaryotes these domains comprise part of the larger pentafunctional AROM protein which catalyses steps 2–6 of the common pathway (Smith and Coggins, 1983). Genes encoding the latter proteins from *Saccharomyces cerevisiae* (*arom*1) and *Emericella* (*Aspergillus*) *nidulans* (*aromA*) have been sequenced (Charles et al., 1986; Duncan et al., 1987). Bacteria thus far studied possess monofunctional DHQase and SDH proteins. The corresponding monocistronic genes, *aroD* and *aroE*, have been cloned and sequenced in *Escherichia coli* (Duncan et al., 1986; Anton and Coggins, 1988). Gene sequences for *aroD* are also available from *Salmonella typhimurium* (Servos et al., 1991), *Bacillus subtilis* (Sorokin et al., 1993) and *Enterococcus faecalis* (Bensing and Dunny, 1993).

The present paper provides an initial report of the molecular cloning of cDNA encoding the bifunctional DHQase·SDH (AroD·E) from a higher plant. An analysis of the relationship with homologue sequences in both prokaryotes and lower eukaryotes is also given.

MATERIALS AND METHODS

Materials

A γZAP II *Nicotiana tabacum* cDNA library kit was purchased from Stratagene (La Jolla, CA, U.S.A.). All molecular-biological supplies, including the Magic Miniprep DNA purification kit, were purchased from Promega (Madison, WI, U.S.A.). AroD·E antibody was prepared by Kel-Farms (Alachua, FL, U.S.A.). Antibody-purification kit and Bradford reagent were purchased from Bio-Rad (Rockville Center, NY, U.S.A.). Dithiothreitol (DTT) was purchased from Research Organics (Cleveland, OH, U.S.A.). Dehydroquinate (90% pure) was chemically prepared (Haslam et al., 1963). Shikimate and other biochemicals were purchased from Sigma Chemical Co. (St. Louis, MO, U.S.A.). Other supplies were purchased from Fisher Scientific Co. (Orlando, FL, U.S.A.). *E. coli* strains AB1360 (*aroD*) and AB2834 (*aroE*) (Pittard and Wallace, 1966) were used.

Screening of *N. tabacum* cDNA library and sequencing cDNA clones

Approx. 10^6 plaque-forming units were transferred to isopropyl thio-β-D-galactoside-treated nitrocellulose filters and screened with AroD·E-specific polyclonal antibody. Antibody was raised in a rabbit by injection of a 1000-fold pure AroD·E sample purified from *N. silvestris* suspension cells [growth conditions were as described by Bonner et al. (1988a)] by a series of chromatographic steps which included a NADP-specific (2′,5′-ADP–Sepharose-4B; Pharmacia, Piscataway, NJ, U.S.A.) affinity column (C. A. Bonner and R. A. Jensen, unpublished work). Antibody was purified on DEAE Affi-Gel Blue Econo-Pac 10DG columns, followed by incubation with an *E. coli* lysate and precipitation of contaminating non-specific serum proteins. Five putative cDNA clones were purified by subsequent secondary and tertiary screening. These were further prepared for DNA purification by a Magic Miniprep DNA kit. Sequencing of purified cDNA was performed on an automated Li-Cor (Lincoln, NE, U.S.A.) sequencer utilizing fluorescently labelled dideoxynucleotides. The cDNA clone labelled SP3 was subcloned to obtain fragments for complete sequencing of the 2 kb cDNA in both directions.

Expression of cDNA encoding the AroD·E bifunctional protein

The pBluescript plasmids carrying the cloned cDNA inserts were transformed into competent *E. coli aroD* and *aroE* auxotrophic cells. Cells were grown on minimal medium plates (plus 100 μg/ml ampicillin) without aromatic amino acids. Prototrophic transformants were inoculated into 1-litre flasks with 200 ml of Luria–Bertani/ampicillin medium, incubated for 8 h at 37 °C, washed and resuspended in potassium phosphate buffer, pH 7.6, containing 20% glycerol and 1.0 mM DTT. Cells were sonicated, centrifuged at 150000 *g* at 4 °C for 1 h, and desalted on PD-10 Sephadex columns (Pharmacia, Piscataway, NJ,

Abbreviations used: DHQase·SDH, dehydroquinase·shikimate dehydrogenase; DTT, dithiothreitol.

* To whom correspondence should be addressed.

The nucleotide sequence of *Nicotiana tabacum* cDNA encoding bifunctional DHQase·SDH will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number L32794.

U.S.A.). SDH (Bonner et al., 1988b) and DHQase (Haslam et al., 1963) were assayed as referenced.

N-terminal amino acid sequencing

N-terminal amino acid sequencing of purified AroD·E was determined with an Applied Biosystems 470A Protein Sequencer with an on-line 120A PTH analyser via the services of the Protein Chemistry Core Facility (PCCF) (University of Florida). A mini-blot protocol for transfer of protein from SDS/PAGE on to PVDF membrane for analysis was provided by PCCF.

RESULTS AND DISCUSSION

Cloning and sequencing cDNA encoding the bifunctional AroD·E

Three of five putative cDNA clones obtained by screening a *N. tabacum* cDNA library with specific AroD·E polyclonal antibody functionally complemented both *E. coli aroD* and *aroE* auxotrophs. Sequencing of the 5′ and 3′ ends revealed that the three cDNA clones were identical. Extracts prepared from *E. coli* transformants carrying SP3 cDNA inserts possessed high catalytic activities for DHQase and SDH (Table 1); 15-fold greater activities were observed for DHQase and SDH in extracts from transformants compared with activities from wild-type extract.

The nucleotide and deduced amino acid sequences of SP3-cDNA are shown in Figure 1. An untranslated 3′ region of 277 bases follows a TAA stop codon, and three to four probable polyadenylation sites are marked prior to the beginning of a poly(A) tail. The seven N-terminal sequenced amino acid residues match the deduced amino acid sequence starting at position number 24. Thus the first 23 residues are presumed to be the carboxy portion of a truncated transit peptide. G^{24}–P^{259} is defined as the AroD· domain and D^{260}–A^{564} is defined as the ·AroE domain. The molecular mass was calculated for the mature protein (60388 Da) and for each domain of the mature protein (AroD· at 25737 Da and ·AroE at 34651 Da) by GCG sequence analysis (Devereux et al., 1987).

Table 1 DHQase and SDH activities in transformed *aroD* and *aroE* mutant strains of *E. coli*.

Enzyme source	Transformed pBluescript	Specific activities* DHQase	SDH
AB1360 *aroD*	−	0	1.3
AB1360 *aroD*	+	15.0	18.7
AB2834 *aroE*	−	0.7	0
AB2834 *aroE*	+	11.4	17.9

* Specific activity is defined as nmol·min^{-1}·mg of protein^{-1}.

```
    CCATTTTTGTGCTCTACAAGTTGGTTGCTAATGGAGTTGGTAGTGGATTCAGGGGTGAGGAAGATGGAGGGGGAGGCAATGACGAGGAACGAAACACTAATTTGTGCACCAATCATGGCA  120
  1 P  F  L  C  S  T  S  W  L  L  M  E  L  V  V  D  S  G  V  R  K  M  E  G  E  A  M  T  R  N  E  T  L  I  C  A  P  I  M  A

    GACACAGTGGATCAAATGTTGAATCTAATGCAAAAGGCTAAAATTAGTGGTGCTGATCTTGTGGAAGTTCGATTGGATAGCTTGAAAAGCTTTAATCCTCAATCAGATATCGATACTATT  240
 41 D  T  V  D  Q  M  L  N  L  M  Q  K  A  K  I  S  G  A  D  L  V  E  V  R  L  D  S  L  K  S  F  N  P  Q  S  D  I  D  T  I

    ATCAAACAGTCCCCTTTGCCTACCCTTTTCACTTACAGGCCCACTTGGGAAGGGGGTCAGTATGCTGGTGATGAAGTGAGTCGACTGGATGCACTTCGAGTAGCAATGGAGTTGGGAGCT  360
 81 I  K  Q  S  P  L  P  T  L  F  T  Y  R  P  T  W  E  G  G  Q  Y  A  G  D  E  V  S  R  L  D  A  L  R  V  A  M  E  L  G  A

    GATTACATTGATGTTGAGCTAAAGGCTATTGACGAGTTCAATACTGCTCTACATGGAAATAAATCAGCAAAATGCAAAGTTATTGTTTCTTCTCACAACTATGATAATACACCATCATCT  480
121 D  Y  I  D  V  E  L  K  A  I  D  E  F  N  T  A  L  H  G  N  K  S  A  K  C  K  V  I  V  S  S  H  N  Y  D  N  T  P  S  S

    GAGGAGCTCGGCAATCTAGTAGCAAGAATACAGGCATCTGGAGCTGACATTGTGAAGTTTGCAACAACTGCACTGGATATCATGGATGTTGCACGTGTATTCCAAATTACTGTACATTCT  600
161 E  E  L  G  N  L  V  A  R  I  Q  A  S  G  A  D  I  V  K  F  A  T  T  A  L  D  I  M  D  V  A  R  V  F  Q  I  T  V  H  S

    CAAGTACCAATAATAGCCATGGTCATGGGAGAGAAGGGTTTGATGTCTCGAATACTTTGTCCAAAATTTGGTGGATACCTCACATTTGGTACTCTTGAAGTGGGAAAGGTTTCGGCTCCT  720
201 Q  V  P  I  I  A  M  V  M  G  E  K  G  L  M  S  R  I  L  C  P  K  F  G  G  Y  L  T  F  G  T  L  E  V  G  K  V  S  A  P

    GGGCAACCAACAATTAAAGATCTTTTGAATATATACAATTTCAGACAGTTGGGACCAGATACCAGAATATTTGGCATTATCGGGAAGCCTGTTAGCCATAGCAAATCACCTTTATTGTAT  840
241 G  Q  P  T  I  K  D  L  L  N  I  Y  N  F  R  Q  L  G  P  D  T  R  I  F  G  I  I  G  K  P  V  S  H  S  K  S  P  L  L  Y
                                                                 └→

    AATGAAGCTTTCAGATCAGTTGGGTTTAATGGTGTTTATATGCCTTTGCTGGTTGATGATGTTGCAAATTTCTTTCGGACTTACTCATCTTTAGATTTTGCTGGCTCAGCTGTAACAATT  960
281 N  E  A  F  R  S  V  G  F  N  G  V  Y  M  P  L  L  V  D  D  V  A  N  F  F  R  T  Y  S  S  L  D  F  A  G  S  A  V  T  I

    CCTCACAAGGAAGCCATTGTTGACTGCTGTGATGAGTTGAATCCTACCGCTAAAGTAATAGGGGCTGTCAATTGTGTCGTAAGCCGACTCGATGGGAAGTTGTTTGGTTGCAATACAGAC 1080
321 P  H  K  E  A  I  V  D  C  C  D  E  L  N  P  T  A  K  V  I  G  A  V  N  C  V  V  S  R  L  D  G  K  L  F  G  C  N  T  D

    TATGTGGGTGCAATCTCCGCCATTGAAGAAGCGTTGCAAGGCTCACAGCCTAGTATGTCTGGGTCTCCCTTAGCTGGTAAATTATTTGTGGTCATTGGTGCTGGTGGCGCTGGCAAGGCA 1200
361 Y  V  G  A  I  S  A  I  E  E  A  L  Q  G  S  Q  P  S  M  S  G  S  P  L  A  G  K  L  F  V  V  I  G  A  G  G  A  G  K  A

    CTTGCTTATGGTGCAAAGGAAAAGGGGGCTCGGGTGGTGATTGCTAACCGTACCTATGAACGAGCGAGAGAACTTGCTGATGTAGTTGGAGGTCAGGCTTTGTCTCTTGACGAGCTTAGC 1320
401 L  A  Y  G  A  K  E  K  G  A  R  V  V  I  A  N  R  T  Y  E  R  A  R  E  L  A  D  V  V  G  G  Q  A  L  S  L  D  E  L  S

    AATTTCCATCCAGAAAATGACATGATTCTTGCAAATACCACCTCCATTGGCATGCAACCAAAGGTTGATGATACACCAATCTTTAAGGAAGCTTTGAGGTACTACTCACTTGTATTTGAT 1440
441 N  F  H  P  E  N  D  M  I  L  A  N  T  T  S  I  G  M  Q  P  K  V  D  D  T  P  I  F  K  E  A  L  R  Y  Y  S  L  V  F  D

    GCTGTTTATACGCCCAAAATCACTAGACTCTTGCGGGAAGCTCACGAGAGTGGAGTAAAAATTGTAACAGGAGTTGAAATGTTTATCGGCCAGGCATATGAACAATATGAGAGATTTACA 1560
481 A  V  Y  T  P  K  I  T  R  L  L  R  E  A  H  E  S  G  V  K  I  V  T  G  V  E  M  F  I  G  Q  A  Y  E  Q  Y  E  R  F  T

    GGGCTTGCCAGCTCCAAAGGAACTTTTCAAGAAAATTATGGCTGGATATTGAGAGCAAGGTCTCTTTCCCTTTTCAATGCGGCCCTGCTAGTTACTTTTCCTCCTAAATCCCTACATAGT 1680
521 G  L  A  S  S  K  G  T  F  Q  E  N  Y  G  W  I  L  R  A  R  S  L  S  L  F  N  A  A  L  L  V  T  F  P  P  K  S  L  H  S

    TGTGTGATAGCAATGGTCTTAGATTCCTCTGCCCTACCATTTGTGCTTCGGAGGAATTAATTCGTTCCAGGTAAATCGTGATTTTCACCAAAACAAATTCTTGAGGATGTTCAGGAAGGC 1800
561 C  V  I  A  M  V  L  D  S  S  A  L  P  F  V  L  R  R  N  *

    AGTCAGACATACCAGTGGACAATCGCCGTCATTCTGGCTTATATTAGACTCTTGTAGCACTTCATTCTTTGACAACTATGGTATCTCTAATTGTGCTTTCATTAAACACAGATGTATCAG 1920

    TGTTTCTCATTGTGACCCCATACTTGGAATTCCTCTTGATTATCATTATTATTAAATCTTGTCACATTATTGTATGATTTGTATCAAAAAAAAA                          2014
                                  ----- -----                 -----
                                        -----
```

Figure 1 Nucleotide sequence (numbered on the right) of *N. tabacum aro*·E

The deduced amino acid sequence of AroD·E is numbered on the left. Residues identified at the N-terminus of the mature protein are underlined. The stop codon is designated by an asterisk, and four possible polyadenylation sites are underlined with a broken line. Residue number 260 (D) is the presumed start of the SDH domain, as denoted by the bent arrow.

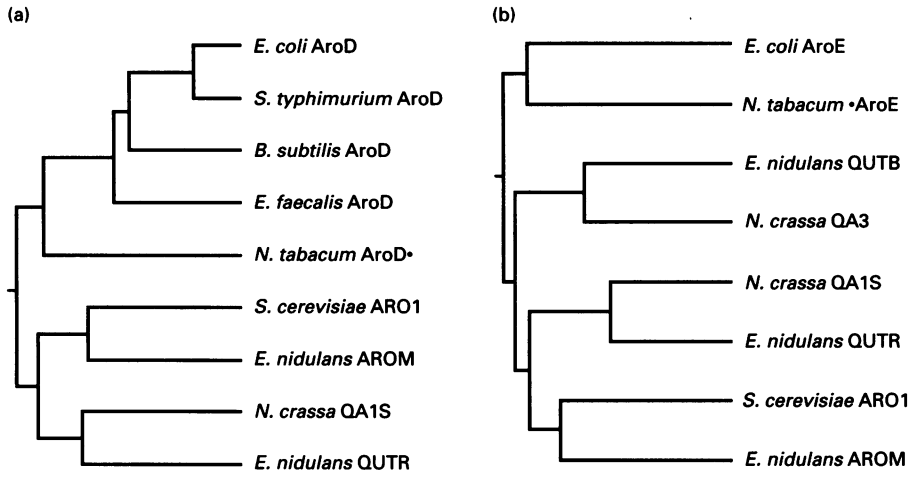

Figure 2 Dendrograms from the GCG PILEUP program showing the clustering arrangements of *N. tabacum* AroD· (a) and ·AroE (b) amino acid sequences with their homologues

PILEUP creates a multiple sequence alignment using a simplification of the progressive alignment method (Devereux et al., 1987).

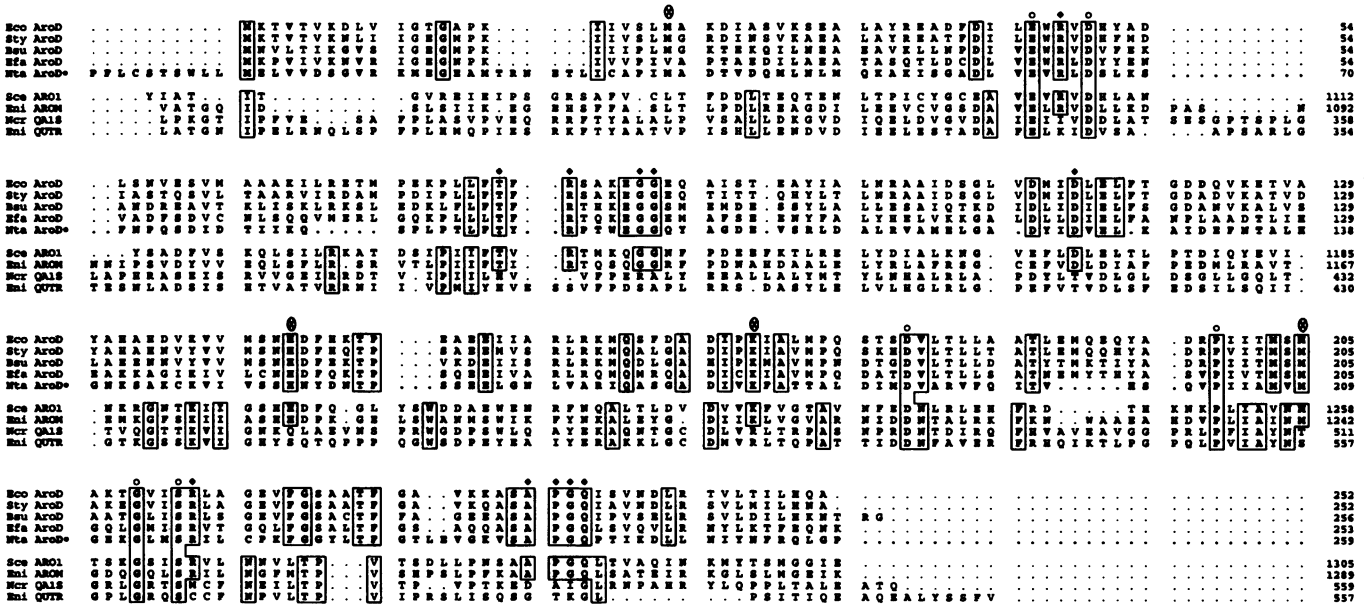

Figure 3 Multiple alignment of *N. tabacum* AroD· with known homologues

Invariant residues in each cluster (see Figure 2a) are boxed. The six residues that are conserved in all nine proteins are marked as ○. Active-site residues identified in *E. coli* are indicated by ⊗. Residues which are conserved in all seven catalytic proteins, but not in the two regulatory proteins, are indicated by ◆. Abbreviations: Eco, *E. coli*; Sty, *S. typhimurium*; Bsu, *B. subtilis*; Efa, *E. faecalis*; Nta, *N. tabacum*; Sce, *S. cerevisiae*; Eni, *E. nidulans*; Ncr, *N. crassa*.

Homology relationships of the AroD· and ·AroE domains

Dendrograms corresponding to multiple alignments of the functional domains of *Nta* AroD·E with their homologues were obtained from the GCG PILEUP program (Devereux et al., 1987), which uses the method of Feng and Doolittle (1987), and are shown in Figure 2. Two clusters are present in each dendrogram, with *Nta* AroD·E grouping with the monofunctional proteins. Thus each domain component of the bifunctional plant enzyme is evolutionarily closer to prokaryotic homologues than to homologues present in lower eukaryotes (yeast and fungi).

The DHQase domain of AroD·E is aligned with its homologues in Figure 3. Four amino acids were reported to be important catalytic residues for the *E. coli* AroD protein. His-143 (Deka et al., 1992), Lys-170 (Chaudhuri et al., 1991), Met-23 and Met-205 (Kleanthous and Coggins, 1990) are conserved in all homologues having AroD catalytic activity, except for Met-23 (which is conserved, however, in *N. tabacum*). It has been suggested that the repressor proteins which regulate quinate catabolism in *Neurospora crassa* (QA1S) (Geever et al., 1989) and in *E. nidulans* (QUTR) (Hawkins et al., 1992) evolved from the three domains of the pentafunctional protein corresponding to DHQase, SDH and shikimate kinase [reviewed by Hawkins et

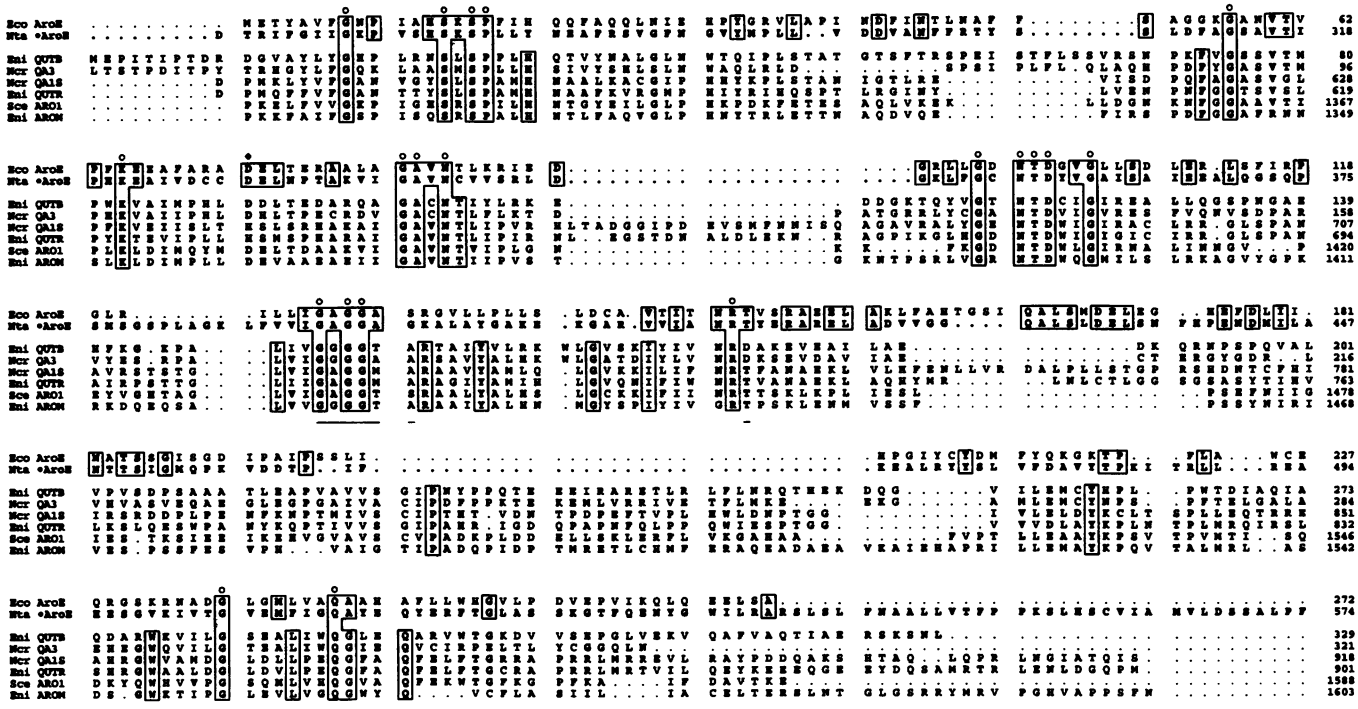

Figure 4 Multiple alignment of *N. tabacum* ·AroE with known homologues

Invariant residues in each cluster (see Figure 2b) are boxed. The 20 invariant residues are indicated by ○. The single residue conserved in the catalytic proteins, but not the regulatory proteins, is indicated by ◆. The glycine-rich region presumed to correspond to the Rossman fold for NAD(P)H binding is underlined, as is the terminal residue of the motif which is usually acidic in NAD^+-specific enzymes. See the legend to Figure 3 for abbreviations. The five C-terminal amino acids (VLRRN) are not shown.

al. (1993)]. These retained the ability to bind what previously were substrate molecules for function in a new role as regulatory agents. Thus, critical residues such as those marked ⊗ were altered to retain binding but not catalysis. Note that nine other residue positions are completely conserved in the catalytic proteins, but not in the two regulatory proteins. This implies that some or all of these residues may also be important for catalysis. Six residues are invariant for all nine proteins.

The SDH domain of AroD·E is aligned with its homologues in Figure 4. This domain is also homologous with the catabolic quinate dehydrogenase (QDH) of *E. nidulans* (QUTB) (Hawkins et al., 1988) and *N. crassa* (QA3) (Geever et al., 1989). Both of the latter are monofunctional and specific for NAD^+. Thus catabolic QDH exhibits homology with biosynthetic SDH, while catabolic dehydroquinases display no obvious homology with biosynthetic dehydroquinases [reviewed by Hawkins et al. (1993)]. This is striking in that the former differ in substrate specificity, whereas the latter do not.

A conserved glycine-rich region of the proposed Rossman-fold motif for NAD(P)H binding (Wierenga et al., 1986) is underlined in Figure 4. The residue position known to be acidic for NAD^+-specific enzymes (as is Asp-179 of Eni QUTB), but not for $NADP^+$-specific enzymes, is underlined. Pairwise comparison of *N. tabacum* AroD·E with its catalytic homologues shows identities ranging from 24 to 32 % and from 27 to 36 % for DHQase and SDH functional domains, respectively (results not shown).

This work was supported in part by contract DE-FG05-86ER13581 from the United States Department of Energy. This is Florida Agricultural Experiment Station Journal Series No. R03795.

REFERENCES

Anton, I. A. and Coggins, J. R. (1988) Biochem. J. **249**, 319–326
Bensing, B. A. and Dunny, G. M. (1993) J. Bacteriol. **175**, 7421–7429
Bonner, C. A., Kenyon, C. and Jensen, R. A. (1988a) Physiol. Plant. **74**, 1–10
Bonner, C. A., Vrba, J. and Jensen, R. A. (1988b) Physiol. Plant. **73**, 451–456
Boudet, A. M., and Lecussan, R. (1974) Planta **119**, 71–79
Charles, I. G., Keyte, J. W., Brammar, W. J., Smith, M. and Hawkins, A. R. (1986) Nucleic Acids Res. **14**, 2201–2213
Chaudhuri, S., Duncan, K., Graham, L. D. and Coggins, J. R. (1991) Biochem. J. **275**, 1–6
Deka, R. K., Kleanthous, C. and Coggins, J. R. (1992) J. Biol. Chem. **267**, 22237–22242
Devereux, J., Haeberli, P. and Marquess, P. (1987) The Program Manual for the Sequence Analysis Software Package, Genetics Computer Group, Madison, WI
Duncan, K., Chaudhuri, S., Campbell, M. S. and Coggins, J. R. (1986) Biochem. J. **238**, 475–483
Duncan, K., Edwards, R. M., and Coggins J. R. (1987) Biochem. J. **246**, 375–386
Feng, D. F. and Doolittle, R. F. (1987) J. Mol. Evol. **35**, 351–360
Geever, R. F., Huiet, L., Baum, J. A., Tyler, B. M., Patel, V. B., Rutledge, B. J., Case, M. E. and Giles, N. H. (1989) J. Mol. Biol. **207**, 15–34
Haslam, E., Haworth, R. D. and Knowles, P. F. (1963) Methods Enzymol. **6**, 498–501
Hawkins, A. R., Lamb, H. K., Smith, M., Keyte, J. W. and Roberts, C. F. (1988) Mol. Gen. Genet. **214**, 224–231
Hawkins, A. R., Lamb, H. K. and Roberts,C. F. (1992) Gene **110**, 109–114
Hawkins, A. R., Lamb, H. K., Moore, J. D., Charles, I. G. and Roberts, C. F. (1993) J. Gen. Microbiol. **139**, 2891–2899
Kleanthous, C. and Coggins, J. R. (1990) J. Biol. Chem. **265**, 10935–10939
Mousdale, D. M., Campbell, M. S. and Coggins, J. R. (1987) Phytochemistry **26**, 2665–2670
Pittard, J. and Wallace, B. J. (1966) J. Bacteriol. **91**, 1494–1508
Servos, S., Chatfield, S., Hone, D., Levine, M., Dimitriadis, G., Pickard, D., Dougan, G., Fairweather, N. and Charles, I. (1991) J. Gen. Microbiol. **137**, 147–152
Smith, D. D. S. and Coggins, J. R. (1983) Biochem J. **213**, 405–415
Sorokin, A., Zumstein, E., Azevedo, V. Ehrlich, S. D. and Serror, P. (1993) Mol. Microbiol. **10**, 385–395
Wierenga, R. K., Terpstra, P. and Hol, W. G. J. (1986) J. Mol. Biol. **187**, 101–107

Received 17 May 1994/7 June 1994; accepted 8 June 1994