Mol Gen Genet (1991) 228:445–452
002689259100264B

MGG
© Springer-Verlag 1991

Molecular organisation of the malate synthase genes of *Aspergillus nidulans* and *Neurospora crassa*

Ruth A. Sandeman[1,*], **Michael J. Hynes**[1], **John R.S. Fincham**[2], **and Ian F. Connerton**[2,3]

[1] Department of Genetics, University of Melbourne, Parkville, Victoria, 3052, Australia
[2] Department of Genetics, Downing Street, University of Cambridge, Cambridge, CB2 3EH, UK
[3] Department of Microbiology, University of Reading, Reading, RG1 5AQ, UK

Received March 1, 1991

Summary. The sequencing and comparison of the genes encoding the glyoxylate bypass enzyme malate synthase of *Aspergillus nidulans* (*acuE*) and *Neurospora crassa* (*acu-9*) are presented. The predicted amino acid sequences of the *A. nidulans* and *N. crassa* enzymes are 538 and 542 residues respectively and the proteins are 87% homologous. In fungi, the malate synthase proteins are located in glyoxysomes and the deduced *acuE* and *acu-9* proteins both contain a C-terminal S-K-L sequence, which has been implicated in transport into peroxisomes. The *acuE* coding region is interrupted by four introns and the *acu-9* coding region is interrupted by one intron which occurs at the same position as the C-terminal *acuE* intron. The 5′ non-coding regions of the two genes were examined for short homologous sequences that may represent the binding sites for regulatory proteins. Pyrimidine-rich sequences with weak homology to the *amdI9* sequence, which has been implicated in *facB*-mediated acetate regulation of the *amdS* gene, were found but their functional significance remains to be determined.

Key words: *Aspergillus nidulans* – *Neurospora crassa* – Acetate – Malate synthase – Glyoxysome transport – Gene regulation – Introns

Introduction

In microorganisms the presence of the glyoxylate bypass of the tricarboxylic acid (TCA) cycle allows the utilization of acetate as a sole carbon source without depleting the cell's reserves of C4 compounds. The key enzymes of this bypass are isocitrate lyase (EC 4.1.3.1) and malate synthase (EC 4.1.3.2) (Kornberg 1966). In the ascomycete fungi *Aspergillus nidulans* and *Neurospora crassa* mutations have defined loci required for growth on acetate (Apirion 1965; Armitt et al. 1976; Flavell and Fincham 1968). A number of genes encoding enzymes involved in acetate utilisation have recently been cloned. The *acuD* gene of *A. nidulans,* encoding isocitrate lyase, was cloned by complementation (Ballance and Turner 1986) and the *facA*/*acuA* (acetyl CoA synthetase) and *acuE* (malate synthase) genes were cloned by differential hybridization screening of cDNA (Sandeman and Hynes 1989). The identity of these genes was confirmed by complementation of existing mutations. In *N. crassa* the acetyl CoA synthetase (*acu-5*) and malate synthase structural genes were cloned by differential hybridization to acetate-induced mRNA (Thomas et al. 1988). The identity of the *acu-5* gene was established by complementation of an *acu-5* mutant and the identity of the malate synthase gene was established by complementation of an *acuE* mutation in *A. nidulans,* as no malate synthase-deficient mutants had been isolated in *N. crassa* (Thomas et al. 1988). Recently, the RIP (repeat-induced point mutation) phenomenon in *Neurospora* (Selker et al. 1987; Selker and Garrett 1988) has been exploited to generate mutations at the *N. crassa* malate synthase locus using the cloned malate synthase genes of both *A. nidulans* and *N. crassa.* The new locus has been designated *acu-9* (Connerton 1990).

In *A. nidulans* mutations in the *facB* gene affect the acetate inducibility of isocitrate lyase and malate synthase, as well as acetyl CoA synthetase and acetamidase, *amdS* (Hynes 1977). It is likely that the *facB* gene encodes a positive regulatory protein that coordinately controls the *facA, acuD, acuE* and *amdS* genes by binding to the 5′ non-coding regions of these genes (Hynes and Davis 1986; Hynes et al. 1988). As yet, no mutations have been isolated that identity a *facB*-like gene in *N. crassa.*

The transformation of the *acu-9* clone into *A. nidulans* not only established the identity of this gene but also demonstrated *trans*-specific gene expression, which in some transformants was regulated by the presence of acetate (Thomas et al. 1988). In similar experiments, using an *acu-5* clone to transform an *A. nidulans*

* *Present address:* School of Agriculture, La Trobe University, Bundoora, Victoria, 3083, Australia

Offprint requests to: M. Hynes

facA⁻strain and *facA* DNA to transform an *N. crassa acu-5*⁻ strain, transformants were obtained that displayed acetate-induced expression of the acetyl CoA synthetase genes, suggesting that the cloned genes were responding to regulatory signals in the alternative host (Connerton et al. 1990). This is despite the fact that comparison of the sequences has that shown a number of differences exist between these two genes, particularly at the 5′ ends (Connerton et al. 1990). The *A. nidulans acuE* gene also remains regulated by acetate when used to complement an *N. crassa acu-9* mutant (I.F. Connerton, unpublished).

This paper describes the sequencing and structure of the malate synthase genes from *A. nidulans* and *N. crassa*. The 5′ non-coding regions of these genes are compared and examined for possible protein-binding sites. The *acuE* and *acu-9* amino acid sequences are also compared, together with other malate synthase gene sequences, and are examined for glyoxysomal targeting signals.

Materials and methods

Cloning and sequencing. The cloning and initial characterization of the *acuE* gene of *A. nidulans* have been described (Sandeman and Hynes 1989). Sequencing of the *acuE* gene was accomplished by the dideoxy chain-termination method (Sanger et al. 1977) using the single-stranded bacteriophage vector M13 and a universal primer (mp18 and mp19, Norrander et al. 1983). Subclones were either sequenced directly or subjected to progressive deletion using T4 DNA polymerase (Dale et al. 1985). The sequencing strategy is shown in Fig. 1 A and the entire gene was sequenced in both strands. These sequence data will appear in the EMBL/Genbank/DDBJ Nucleotide Sequence Data Libraries under the accession number X56671 (*acuE, Aspergillus nidulans*).

The cloning and initial characterization of the malate synthase gene of *N. crassa* have been described (Thomas et al. 1988). This gene was sequenced using the dideoxy chain-termination method (Sanger et al. 1977) with fragments cloned in pEMBL 18 and 19. Small *Kpn*I fragments from the centre of the gene were sequenced in both strands using chemical cleavage (Maxam and Gilbert 1977) because of secondary structure problems. The sequencing strategy is shown in Fig. 1 B and the entire gene was sequenced in both strands. These sequence data will appear in the EMBL/Genbank/DDBJ Nucleotide Sequence Data Libraries under the accession number X56672 (*acu-9, Neurospora crassa*).

Transcript mapping. S1 nuclease mapping of the 5′ end of the *acuE* gene of *A. nidulans* was performed according to Burke (1984) in order to determine the start point(s) of transcription. An M13 subclone containing a 414 bp *Pst*I−*Sal*I M13 subclone (coordinates −404 to +10, see Fig. 2) was used as a source of the ^{32}P-labelled single-stranded probe and hybridized to total RNA from either non-induced or acetate-induced wild-type strains. Two fragments of 90 and 72 bp were protected from S1 nuclease digestion by induced RNA but not by non-induced RNA.

The transcription startpoint(s) of the *acu-9* gene of *N. crassa* were determined by primer extension using the oligonucleotide AGGATCTTCCTCTGGTG which primes at coordinate +58, by the method described previously (Connerton et al. 1990). Extension products of 84 and 119 bp were observed with poly(A)$^+$ RNA from an acetate-induced culture.

Computer analysis. Homology matrix analysis, comparisons of the 5′ non-coding regions and subsequence analysis of the *acuE* and *acu-9* genes were performed using the DNA Inspector II$^+$ and MacVector programs on an Apple Macintosh computer.

Results

Structure of the malate synthase genes

Restriction maps of the malate synthase genes of *A. nidulans* (*acuE*) and *N. crassa* (*acu-9*) are presented in Fig. 1

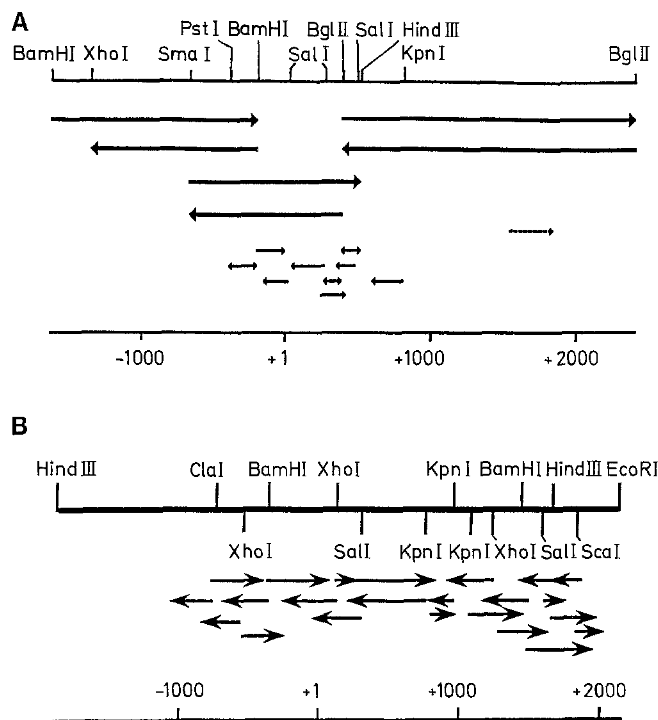

Fig. 1 A and B. Restriction maps and sequencing strategies of the *acuE* and *acu-9* genes. **A** A restriction map of the *acuE* gene. M13 subclones that were deletion cloned and sequenced are shown immediately below the restriction map. Smaller M13 subclones that were sequenced are shown below the deletion subclones. The extent of the sequenced region is shown by the *line* and the direction of sequencing by the *arrow*. The *dashed line* and *arrow* show the position and direction of a cDNA clone that was sequenced. +1 marks the start of translation of the *acuE* gene. **B** A restriction map of the *acu-9* gene. pEMBL18 and 19 subclones that were sequenced are shown below the restriction map. The extent of the sequenced region is shown by the *line* and the direction of sequencing by the *arrow*. +1 marks the start of translation of the *acu-9* gene

CTCGAGATTCCTGTCCCGCCGCTATCCAAGGGTGAGGATATGGGGGGCGAGACGATAAGGGAGTACGGCTTGCTAAGAATTTGGGTCGGCGATCGCCACGTTT −1253
ACGCGAAAAGACAGTGTCGCACGGGAAGTCTTTGGCGATGCTTTCGTTGATCACTTTGGTGGGACTCGGGAACATGAACTGCGCTTGTGGGAAGAGGCCGTAA −1150
CGGACTGGTATGCTCTACCTAATGGGATGTTCAATGGCATATGCTGACTTGATATAGGGAGGTGCGACGGTATATTGAGACAGTCTAGTGACCGGGCCTCACC −1047
GATCAAAGTTTAGGAGCTTGGAAGACAAGGCTGAAGGTTGACAGATGACCCGTTCGATCCCGAACTCGGTCAGCGCGTAGATTCCGGGGTCGATCCGGAGCTG −944
GGATGTATCCATATTATGATATCACATGTTAGGATAGCTCATGTTGGTCTGTTCATGGGTTTTCACGTGCCAGCGGTGCTGTTTTTCAGCGATTGAGGTATAG −841
ATACACAATTGTCGATGTTGGTCAAGTTTGGTTCCTTTCCTGCTAGCAGTGACTCACCGCGATATCGAGTCAATGTAATAATTATAGTAGCAGCTGGACAAAC −738
CTCGGAAGATGTCGGTGACAGAGCTTCAGTATGACGGACGAAAGGTGAAACTGCCCGGGCGATCAACCGCGGCACTGGGAGCTCATAGGAGCTGCTCGCTCGG −635
GCGTGCGGGGAGTTTTCCACCTTGTTTGCCTCCTGGCTGGGCAGCGTCACGGAGAAACATATGTAGCCACTCTGCATCTGCAATTGTAGACCAACATGTTGTA −532
GATTCGTTCTATTAAAACTCGTAATTGATAAACTGTAGCCTTCAGTCGGTCGAGAGCTCGGTCGTCTCGGTCGCTCGGTCGACGCCGTTCGCCGTGTAGAATG −429
CTGGGACTCGCAGCATTAAGAGCTCTGCAGTCGGCCGAGTAGCTTCGTATCTCGTGACTTTCCCGCGGGAAAATCACATGACTAGATACCGGATTCGGCCAGA −326
TCCGTTCATCCTGACGCTTTCAACTTCACGATCAACCGTCTGAATGGCTCAATTCCTATCGACCCTGCTCCATCCTTCCTGATTTGCCCGCGCTCATGGATCC −223
ATTCCAAAGTCTCCCAATCACCGGCGTCGATTGGACCGCGGTTGTTCATCGCTCACTGACGTAGTCACAATGTGCTCCTGCCGAACTGACTGTTATCGGCATG −120
CTTCATATATATATAAGTCTGGTCTAGGCTCGCCCCCGGATTCTTGATTGACTTTTCTTCTTTCCAAACTTCGCTACACTTCTCCATCTCTTCTATAAATGGAT −18

met ser gln val asp ala gln leu lys asp val ala ile leu gly ser val ser asn glu ala
AAGTGAACGCTAGCATC ATG TCT CAG GTC GAC GCC CAG CTT AAG GAT GTG GCC ATC CTC GGC TCC GTG AGC AAC GAA GCC +63

arg lys ile leu thr lys glu ala cys ala phe leu ala ile leu his arg thr phe asn pro thr arg lys ala leu
CGC AAG ATC CTC ACA AAG GAA GCC TGT GCT TTC CTC GCC ATC TTA CAC CGT ACC TTC AAC CCT ACT CGC AAG GCT CTC +141

leu gln arg arg val asp arg gln ala glu ile asp lys gly his leu pro asp phe leu pro glu thr lys his ile
CTG CAG CGC CGC GTT GAT CGT CAG GCC GAG ATC GAC AAG GGC CAC CTC CCC GAC TTC CTC CCA GAG ACC AAG CAC ATC +219

arg asp asp pro ser trp lys gly ala pro pro ala pro gly leu val asp arg arg val glu ile thr gly pro thr
CGC GAT GAT CCC AGC TGG AAG GGA GCT CCC CCA GCG CCT GGT CTC GTC GAC CGT CGC GTT GAG ATC ACT GGT CCT ACC +297

asp arg lys met val val asn ala leu asn ser asp va
GAC CGG AAG ATG GTC GTC AAC GCA CTG AAC TCG GAT GT **gtggac**ttacatggctgatttcgagggtatgcatcacagatctgcaaaca +385

l ala pro thr trp asp asn met ile asn gly gln ile asn leu tyr
tttcgatgagtaac**agctaac**ggcgtgaagactc**cag** C GCC CCT ACT TGG GAT AAC ATG ATC AAC GGC CAG ATC AAC CTT TAC +468

asp ala ile arg arg gln val asp phe lys gln gly gln lys glu tyr lys leu arg thr asp arg thr leu pro thr
GAT GCC ATC CGC CGC CAG GTC GAC TTC AAG CAG GGT CAG AAG GAG TAC AAG CTT CGC ACA GAC CGA ACC CTG CCC ACC +546

leu ile ala arg ala arg gly trp his leu asp glu lys his phe thr val asp gly glu pro ile ser gly ser leu
CTG ATT GCT CGT GCC CGT GGC TGG CAC CTC GAC GAG AAG CAC TTC ACT GTC GAT GGC GAG CCC ATC TCC GGC AGT CTG +624

phe asp phe gly leu tyr phe phe his asn ala lys glu leu val ala arg gly phe gly pro tyr phe tyr leu pro
TTC GAC TTT GGT CTG TAC TTC TTC CAC AAC GCC AAG GAA TTG GTG GCT CGC GGG TTC GGT CCT TAC TTC TAC CTT CCC +702

lys met glu ser his leu glu ala arg leu trp asn asp val phe asn leu ala gln asp tyr ile gly met pro arg
AAG ATG GAG TCT CAC CTC GAA GCT CGT CTG TGG AAC GAT GTC TTC AAC CTG GCT CAG GAC TAC ATT GGC ATG CCC CGC +780

gly thr ile arg gly thr val leu ile glu thr ile thr ala ala phe glu met glu glu
GGC ACC ATC CGT GGT ACC GTT CTG ATT GAA ACC ATC ACT GCT GCG TTT GAG ATG GAA GAG **gtgtgt**ttctttgcttgctc +861

ile ile tyr glu leu arg asp his ser ser gly leu asn cys gly arg trp asp
gtttagggcta**cgctaac**agctt**tag** ATC ATC TAC GAA CTC CGT GAC CAC AGC TCC GGT CTC AAC TGC GGC CGC TGG GAC +940

tyr ile phe ser phe ile lys lys phe arg gln his pro asn phe val leu pro asp arg ser asp val thr met thr
TAC ATC TTC TCC TTC ATC AAG AAA TTC CGC CAA CAC CCC AAC TTT GTC CTT CCT GAC CGC TCT GAT GTC ACC ATG ACC +1018

val pro phe met asp ala tyr val lys leu leu ile lys thr cys his lys arg gly val his ala met
GTA CCT TTC ATG GAT GCC TAC GTG AAG CTC CTC ATC AAG ACC TGT CAC AAG CGA GGA GTC CAC GCT ATG **gtatgc**cc +1095

gly gly met ala ala gln ile pro ile lys asp asn ala glu ala
tttcttttgcaaagttcggtatccg**tactaac**gtatg**cag** GGT GGA ATG GCC GCT CAA ATC CCC ATT AAA GAC AAC GCC GAG GCC +1180

asn asp lys ala met glu gly val arg ala asp lys leu arg glu val arg ala gly his asp gly thr trp val ala
AAC GAC AAG GCC ATG GAA GGC GTG CGC GCC GAT AAG CTC CGT GAA GTT CGT GCA GGC CAC GAC GGC ACA TGG GTT GCG +1258

his pro ala leu ala ser ile ala ser glu val phe asn lys tyr met pro thr pro asn gln met his val arg arg
CAC CCG GCT CTC GCT TCG ATT GCC AGT GAA GTT TTC AAC AAG TAC ATG CCC ACC CCC AAT CAG ATG CAC GTC CGC CGC +1336

glu asp val asn ile thr ala asn asp leu leu asn thr asn val pro gly lys ile thr glu asp gly ile arg lys
GAG GAC GTC AAC ATC ACC GCC AAC GAC CTC CTC AAC ACC AAC GTT CCC GGA AAG ATC ACC GAG GAC GGT ATC CGC AAG +1414

asn leu asn ile gly leu ger gyr get glu gly trp leu arg gly val gly cys ile pro ile asn tyr leu met
AAC CTG AAC ATC GGT CTC TCC TAC ATG GAG GGT TGG CTT CGT GGT GTC GGA TGT ATC CCT ATT AAC TAC CTG ATG +1489

glu asp ala ala thr ala glu val ser arg ser
gtaaggacccctgcactttcttagcccaagaaatga**tgttaat**gatgaaa**tag** GAG GAC GCC GCT ACC GCC GAA GTC TCC CGC AGC +1575

gln leu trp gln trp ala arg his gly val thr thr ser glu gly lys lys val asp lys ala tyr ala leu arg leu
CAG CTT TGG CAA TGG GCC CGC CAC GGT GTT ACC ACC TCT GAG GGC AAG AAG GTT GAC AAG GCT TAT GCC TTA CGT CTG +1653

leu lys glu gln ala asp ala leu ala ala lys gly pro lys gly asn lys phe gln leu ala gly arg tyr phe ser
CTG AAG GAA CAG GCC GAT GCC CTT GCA GCC AAG GGT CCT AAG GGC AAC AAG TTC CAG CTT GCT GGT CGC TAC TTT TCC +1731

glu aln val thr gly glu asp tyr ala asp phe leu thr ser leu leu tyr asn glu ile ser ser pro gly thr ala
GGC CAG GTT ACC GGT GAA GAT TAC GCC GAC TTC TTG ACC AGC CTG CTG TAC AAC GAG ATT TCG TCT CCG GGT ACT GCT +1809

ser lys leu ***
TCA AAG CTC TAA TTTCATCCTTTTTTGTTACCGGTTTCTTTATGGTCTTCTAATCTTGGGTGTTATGTTCGTTATGCTAGATATGTATGATTATGATT +1907
CCAAACTGATGGTTATTCTTTGATTCATCTTCAGGGTGTACGAGGCAAAGCATATCTATGCTTGATCTAGTATGCTATGCTGTGCTAAATTATCTATGTATAT +2010
CTAAAATGCAAACCAGATCCACCTTCGTATTTCATGCTTTATTTCATGTAGTAACCACAGGTTACTAACCCAAATATATACTTTTAAACAAACCAATATACTT +2113
GAAAAACCTGATCCCTCCAGAAATTTCATGGATCAAAAGGTTTACCCAATGGCAGGTTTCTTGAACTGTAGCCAACCCAAAAATAGAACTCATCCTTCGGCAG +2216
CACCCAATTCTGCTCACGCACACTCCAGTAGCTCAGATCT +2256

Fig. 2. DNA and amino acid sequence of the *acuE* gene. The DNA sequence is numbered in relation to the translation startsite. The CAAT and TATAA boxes are *doubly underlined*, the transcription starts are *arrowed* and intron 5′ and 3′ splice sites and signals are in *bold face*

```
CTTTGTAAGTTTGGGTGACTGGCCGTTTCTTGACGCTTGAGCCTCAAGACGTTCAAGTTAGGTAGAGGG -905
ATGAAGAAAGACTAGAGAGGACGGTTTTAGAAAGAGCCTTTGATAGTTTACAGTGAATGCCGCCTTTTG -836
GTGAGGAGTGGATCAGGACAACAACAAACCACTTCTGTTGGAGACATAACAAAGTACAATGGAGACTTG -767
ACTTACACCGGTATGTTACAGCAGTTGACGATTCGTGAAAGGATGACCATCGATTTGCCTTACCGGTAT -698
CTCTACCTGGACCGGGACAGTCCTCCATAGCCACACGGAGACTGCCCCTAGCCCGAGAGCAGGAACCTC -629
GAGATGCGAGCTGATAAGGCCAACAGCAAAAGAGGCAAGATATGATTAACAGTCGTAACACAACCCCGC -560
CAAGGGGGCTCCATCAAGCAGACCAAAGACCACGCGCTGGTGCTATTGCCGAGCCATCGACTAAGATGC -491
ATGGCCCCGGTTCTTTTCGGTCATGCCTCGTTAGGACGGCGTTAGTGTAAAGTGCTCGATAGACGGACA -422
TCAATTGGATCCAGAAACCGAAGTTGCCTAGGTGAGTGACTCAAGCCCAGATGGGAAGTGGGTTGAAGG -363
CATGGGATCTTCACGAGACCTCTACAAGCTATCAATTCCTCCAGACAGATAAACAGACTCCTCTCCCCT -284
TCCTCCTTCTAGGTCCCTCTTCTCTCTTCACTGTCCCAAGTTATCCTATCTATACCCCATTCATCCAAT -215
CTACTTGCCCCTCCACCACCTCCTCCGGCAAACATCATCCTCTCTTCACATCCTCGGCACTTTTCGGTA -146
GCACGGGTTCTTCCTAGTTGTATGTTGCTGCAGACATTACTTAATAAGCTGTGAACTTGATTGCTGACT  -77
TTCTGACCTTTTGCTCAAGAACGATTTCATACCATTAGCATTTTCTTTTACAAAGATACCCACCACATA   -8
TTTCACA met ala ser val glu thr leu leu gln gly val thr ile ser gly
        ATG GCT TCC GTC GAG ACT CTT CTC CAG GGC GTC ACC ATC AGC GGC    +45
pro ile glu glu his gln arg lys ile leu thr pro gln ala leu ser phe
CCC ATT GAG GAG CAC CAG AGG AAG ATC CTG ACT CCC CAG GCC CTC TCC TTC    +96
val ala leu leu his arg ser phe asn gln thr arg lys asn leu leu glu
GTC GCC CTC CTC CAC CGT TCC TTC AAC CAG ACC CGC AAG AAC CTC CTC GAG   +147
arg arg his val arg gln ala glu ile asp arg gly val leu pro asp phe
CGC CGT CAC GTC CGC CAG GCC GAG ATC GAC CGC GGT GTC CTC CCC GAC TTC   +198
leu pro glu thr lys his ile arg glu asn pro thr trp lys gly ala ala
CTC CCC GAG ACC AAG CAC ATC CGT GAG AAC CCC ACA TGG AAG GGT GCC GCT   +249
pro ala ala pro pro leu val asp arg arg val glu met thr gly pro thr
CCC GCT GCG CCT CCG CTC GTC GAC CGT CGC GTT GAG ATG ACC GGT CCT ACC   +300
asp arg lys met val val asn ala leu asn ser asp val tyr thr tyr met
GAC CGG AAG ATG GTC GTC AAC GCA CTG AAC TCC GAT GTC TAC ACC TAC ATG   +351
ala asp phe glu asp ser ser ala pro thr trp ala asn met val asn gly
GCC GAC TTC GAG GAC TCC AGC GCC CCT ACC TGG GCC AAC ATG GTC AAC GGC   +402
gln val asn leu tyr asp ala ile arg arg gln ile asp phe lys gln gly
CAG GTG AAC CTC TAC GAT GCT ATC CGT CGC CAG ATC GAC TTC AAG CAA GGC   +453
pro lys glu tyr lys leu arg thr asp arg thr leu pro thr leu ile val
CCG AAG GAA TAC AAG CTC CGC ACC GAC AGG ACT CTT CCC ACC CTC ATC GTC   +504
arg pro arg gly trp his leu glu glu lys his val thr ile asp gly glu
CGC CCT CGT GGC TGG CAC CTT GAG GAG AAG CAC GTC ACC ATT GAC GGC GAG   +555
pro val ser gly ser leu phe asp phe gly leu tyr phe phe his asn ala
CCC GTC AGC GGT TCG CTT TTC GAC TTT GGT CTT TAC TTC TTC CAC AAC GCC   +606
lys glu leu val gln arg gly phe gly pro tyr phe tyr pro pro lys met
AAG GAG CTC GTC CAG CGC GGC TTC GGC CCC TAC TTC TAC CCT CCC AAG ATG   +657

glu ser his leu glu ala arg leu trp asn asp ala phe asn leu ala gln
GAG TCT CAC CTT GAG GCT CGT CTG TGG AAC GAT GCC TTC AAC CTG GCT CAG   +708
asp tyr val gly ile pro leu ser thr ile arg gly thr val leu ile glu
GAC TAC GTT GGC ATT CCC CTG AGC ACC ATC CGT GGT ACC GTC TTG ATC GAG   +759
thr ile thr ala ala phe glu met asp glu ile ile phe glu leu arg asn
ACA ATC ACT GCT GCG TTT GAG ATG GAC GAG ATC ATC TTC GAG CTC CGG AAC   +810
his thr ser gly leu asn arg gly gly trp asp tyr ile phe pro phe ile
CAC ACC TCC GGT CTC AAC CGC GGC GGT TGG GAT TAC ATC TTC CCC TTC ATC   +861
lys glu val arg arg phe pro asn phe val leu pro asp arg ser asp val
AAG GAA GTC CGC CGG TTC CCG AAC TTT GTC CTT CCT GAC CGC TCT GAT GTC   +912
thr met thr val pro phe met glu ala tyr val lys leu leu ile lys thr
ACC ATG ACG GTA CCT TTC ATG GAA GCC TAC GTC AAG CTC CTC ATC AAG ACC   +963
leu his arg leu val val his ala met gly met ala ala gln ile pro ile
CTT CAC AGA CTG GTT GTC CAC GCC ATG GGC ATG GCC GCT CAG ATC CCC ATC  +1014
lys asp asp lys ala ala asn asp lys ala met glu gly val arg ala asp
AAG GAC GAC AAG GCC GCC AAC GAC AAG GCC ATG GAG GGC GTT CGC GCC GAC  +1065
lys leu arg glu ala arg ala gly his asp gly thr trp val ala his pro
AAG CTC CGT GAA GCC CGT GCC GGC CAC GAC GGT ACC TGG GTT GCT CAC CCC  +1116
ala leu ala ser ile ala leu glu val phe asn lys his met pro thr pro
GCC CTT GCC AGC ATT GCC CTT GAG GTC TTC AAC AAG CAC ATG CCT ACC CCC  +1167
asn gln leu phe asn arg arg glu asp val lys ile gly gln gln asp leu
AAC CAG CTC TTC AAC CGC CGT GAG GAC GTC AAG ATT GGC CAG CAG GAC CTG  +1218
leu asn met asn val pro gly ser ser thr glu asp gly ile arg lys asn
CTC AAC ATG AAC GTC CCC GGC TCG AGC ACC GAG GAC GGT ATC CGC AAG AAC  +1269
leu asn thr gly leu gly tyr thr glu pro trp ile arg gly val gly cys
CTC AAC ACT GGT CTC GGC TAC ACG GAG CCT TGG ATC CGT GGT GTC GGT TGC  +1320
val pro ile lys  his pro gln
GTC CCC ATC AAG CAT CCT CAG gtatgtattctcttttctttcccttccacacagcaccc  +1379
                                 glu asp  ala ala thr ala glu val
atcgtgatcatcatactaacaaccacaacag GAG GAT GCC GCC ACC GCC GAG GTT      +1434
ser arg ser gln leu trp gln trp val lys his arg val thr thr ala glu
TCC CGC TCC CAG CTC TGG CAG TGG GTC AAG CAT CGC GTC ACC ACC GCC GAG  +1485
gly lys his val asp lys arg tyr pro leu lys leu leu lys glu ala asp
GGC AAG CAC GTC GAC AAG CGT TAT CCC CTC AAG CTT CTC AAG GAA GCA GAC  +1536
arg gln arg leu ala lys ala pro gln gly asn lys phe asn leu ala ala
CGA CAG AGA CTC GCC AAG GCT CCC CAG GGT AAC AAG TTC AAC CTT GCC GCT  +1587
gln tyr phe ala ser gln val thr gly glu asp tyr ala asp phe leu thr
CAG TAC TTC GCC AGC CAG GTG ACT GGC GAG GAC TAT GCC GAC TTC TTG ACC  +1636
cys leu leu tyr asn glu ile thr ser ala gly asn ser leu pro ala ser
TGC CTC CTT TAC AAC GAG ATT ACC AGT GCT GGT AAC TCG TTG CCT GCT TCC  +1689
lys leu ***
AAG CTC TAA ATGCTGAGTTCGGGATAATAAGGTCGAAGGACGATGATAAGGGGATGGCTGGGAA  +1753
AAGGAAAACGAACGGCAATA                                                  +1773
```

Fig. 3. DNA and amino acid sequence of the *acu-9* gene. The DNA sequence is numbered in relation to the translation startsite. The TATAA box is *doubly underlined*, the transcription starts are *arrowed* and intron 5′ and 3′ splice sites and signals are in *bold face*

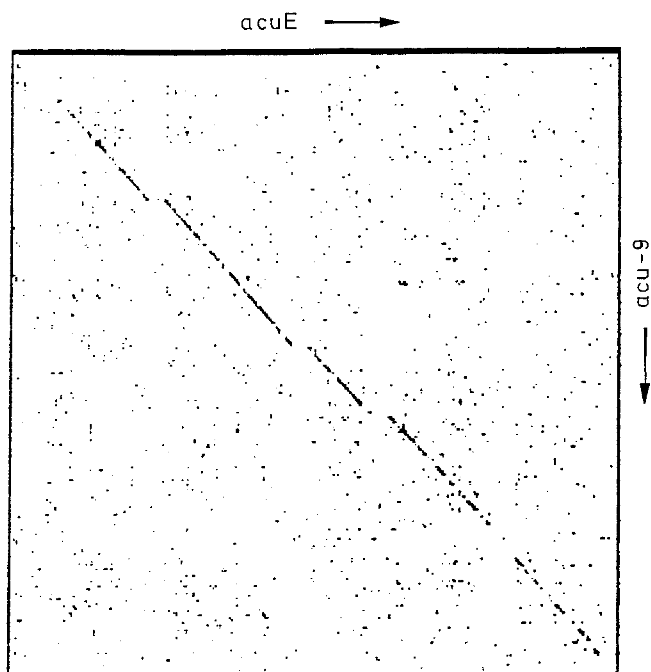

Fig. 4. DNA homology matrix analysis of the *acuE* and *acu-9* genes. The matrix was constructed by comparing the DNA sequences of the protein coding regions of both genes using a search element length of 8 bp with a maximum mismatch of 1 bp

and the complete nucleotide sequences are presented in Figs. 2 and 3 respectively. The *A. nidulans acuE* gene encodes a protein of 538 amino acids with a molecular weight of 60744 daltons. The *N. crassa acu-9* gene encodes a protein of 542 amino acids with a molecular weight of 61364 daltons. The two genes show a similar bias against adenosine in the third position of codons.

S1 nuclease mapping and primer extension of the 5′ ends of the malate synthase genes showed that there were two transcription startpoints for each gene (Figs. 2 and 3). The *acuE* gene has transcription startpoints 59 and 77 bp upstream of the translation initiation codon and the *acu-9* gene has transcription startpoints 23 and 58 bp upstream of the translation initiation codon. A CT-rich sequence is found between the transcription startpoints in both genes (Figs. 2 and 3). Both genes also contain TATAA-like sequences at the expected positions prior to the transcription startpoints and the *acuE* gene also contains a CCAAT-like sequence (Figs. 2 and 3).

The *acuE* coding region is interrupted by four introns ranging in size from 46 to 54 bp. The positions of these introns were determined by examining open reading frames and consensus 5′ and 3′ splice site and 3′ splice signal sequences. The positions were then confirmed by comparing the *acuE* sequence with the *Eschericia coli aceB* sequence (Byrne et al. 1988). The fourth intron in *acuE* was also localized by sequencing a cDNA clone from the 3′ end of the gene (see Fig. 1A). The *acu-9* coding region is interrupted by a single intron of 69 bp that corresponds in position with the fourth *acuE* intron. The position of this intron was determined by S1 nuclease mapping. The 5′ and 3′ splice site and 3′ splice signal sequences from the introns of both genes conform to consensus sequences determined from sequenced fungal genes (Gurr et al. 1988). DNA homology matrix analysis displays the high degree of similarity and the positions of introns of the malate synthase genes of *A. nidulans* and *N. crassa* at the DNA level (Fig. 4).

Comparison of the malate synthase amino acid sequences

A comparison of the predicted amino acid sequences of the *acuE, acu-9* and *aceB* genes and the *Brassica napus* (Comai et al. 1989) and *Cucumis sativus* (Smith and Leaver 1986) sequences are shown in Fig. 5. The *acuE* and *acu-9* coding regions are very similar, with 87% conserved residues, although there is a drop in conservation towards the N-terminal end of the protein (43% conserved residues in the first 21 amino acids). When the *aceB, B. napus* and *C. sativus* malate synthase sequences are compared with the fungal proteins the overall homology between the proteins is approximately 50% (Fig. 5). The sites of introns of the fungal genes lie within regions that are conserved between all five proteins.

Discussion

Comparisons of the DNA and amino acid sequences of the malate synthase genes of *A. nidulans* and *N. crassa* have shown that there is a high degree of conservation between these genes. However, there are a number of structural differences, both in the number of introns and in the 5′ non-coding regions of the genes. Despite these obvious diferences the *acuE* and *acu-9* genes are functionally interchangeable in both species as regards both protein product and, to a certain extent, regulation (Thomas et al. 1988). Thus, differences in the number of introns and in the 5′ untranscribed sequences have little effect on the expression of these genes in the alternative host.

Differences in structural organisation have been observed on comparison of a number of genes from *A. nidulans* and *N. crassa*. In the genes of the quinic acid utilization cluster there is a high degree of conservation between homologous genes. However, the *qutB* gene of *A. nidulans* is interrupted by two introns that are not observed in the homologous *qa-y* gene of *N. crassa*. Similarly, the *qutG* gene of *A. nidulans* has four introns while the homologous *qa-x* gene has two introns (Hawkins et al. 1988; Geever et al. 1989). In both cases these introns fall within regions of the genes that are otherwise highly conserved in both DNA and amino acid sequence (Hawkins et al. 1988; Geever et al. 1989).

A comparison of the acetyl CoA synthetase genes of *A. nidulans* and *N. crassa* also revealed similar differences in gene structure. The *facA* gene of *A. nidulans* contains six introns, two of which are contained within the 5′ leader mRNA and the N-terminal end of the protein where there is no conservation between the *A. nidulans* and *N. crassa* genes. Four more introns interrupt

```
N.c.  MASVETLLQG--VTISGP--IEEHQR------------KILTPQALSFVALLHRSFNQT-RKNLLERRHVRQAEIDRGVLPDFLPETKHIRENPTWKGAA
A.n.  MSQVDAQLKD--VAILGS--VSNEAR------------KILTKEACAFLAILHRTFNPT-RKALLQRRVDRQAEIDKGHLPDFLPETKHIRDDPSWKG-A
      *  *C  *    * * *   C C  *            **** C*  *C*C***C** * ** **C**  ******C* ************CC*C*** *
E.c.  MTEQATTT-D-ELAFTRP--YGEQEK------------QILTAEAVEFLTELVTHFTPQ-RNKLLAARIQQQQDIDNGTLPDFISETASIRDAD-WKIRG
B.n.  MEL-ETSVYRPNVAV-----YDSPDGVEVRGRYDQVFAKILTRDALGFVAELQREFRGHVRYAMECRREVKRR-YNSGAVPGFDPSTKFIRDGE-WVCAS
C.s.  MGSLGMYSESGLTKKGSSRGYDVPEGVDIRGRYDEEFAKILNKEALLFIADLQRTFRNHIKYSMECRREAKRR-YNEGGLPGFDPATKYIRDSE-WTCA-
      *                   C                   ** C*  *C   *   *    C  C   *       CC * C* *   *  **C   *

N.c.  PAAPPLV-DRRVEMTGPTDRKMVVNALNSDVYTYMADFEDSSAPTWANMVNGQVNLYDAIRRQIDFKQ--GPKEYKLRTDRTLPTLIVRPRGWHLEEKHV
A.n.  PPAPGLV-DRRVEITGPTDRKMVVNALNSDVWTYMADFEDSSAPTWDNMINGQINLYDAIRRQVDFKQ--GQKEYKLRTDRTLPTLIARARGWHLDEKHF
      * ** ** *****C***************** ************** **C***C*********C****  * *************** * *****C***C
E.c.  IPAD-LE-DRRVEITGPVERKMVINALNANVKVFMADFEDSLAPDWNKVIDGQINLRDAVNGTISYTNEAG-KIYQLKPNAV---LICRVRGLHLPEKHV
B.n.  VPPA--VADRRVEITGPVERKMIINALNSGAKVFMADFEDALSPSWENLMRGQVNLKDAVDGSITFNDKARNKVYKLN-DQVAK-LFVRPRGWHLPEAHI
C.s.  PVPPA-VADRRVEITGPVERKMIINALNSGAKVFMADFEDALSPNWENLMRGQINLKDAVDGTISFHDRVRNRVYKLN-DRTAK-LFVRPRGWHLPEAHI
              *****C*** C***CC****     C******    * * CC **C** **C   C C C    C * *  C     *C * ***** * *C

N.c.  TIDGEPVSGSLFDFGLYFFHNAKELVQR---GFGPYFYPPKMESHLEARLWNDAFNLAQDYVGIPLSTIRGTVLIETITAAFEMDEIIFELRNHTSGLNR
A.n.  TVDGEPISGSLFDFGLYFFHNAKELVAR---GFGPYFYLPKMESHLEARLWNDVFNLAQDYIGMPRGTIRGTVLIETITAAFEMEEIIYELRDHSSGLNC
      *C****C******************* *   ******* ************** *******C*C*  *****************C***C***C*C****
E.c.  TWRGEAIPGSLFDGALYFFHNYQALLAK---GSGPYFYLPKTQSWQGAAWWSGVFSYAEDRFNLPHGTIKATLLIETLPAVFQMDEILHALRDHIVGLNC
B.n.  LIDGEPATGCLVDFGLYFFHNYAKFRQTQGSGFGPGGYLPKMEHSREAKIWNSVFERAEKMAGIERGSIRATVLIETLPAVFQMNEILYELRDHSVGLNC
C.s.  FIDGEPATGCLVDFGLYFFHNHANFRRSQGQGYGPFFYLPKMEHSREAKIWNSVFERAEKMAGIERGSIRATVLIETLPAVFQMNEILYELRDHSVGLNC
         **   * *C*  ******          * **  * ** C    *  *   *  *C    C   C*C *C****C * *C*C**C  **C*  ***

N.c.  GGWDYIFPFIKEVRRFPNFVLPDRSDVTMTVPFMEAYVKLLIKTLHRLVVHAMG-MAAQIPIKDDKAANDKAMEGVRADKLREARAGHDGTWVAHPALAS
A.n.  GRWDYIFSFIKKFRQHPNFVLPDRSDVTMTVPFMDAYVKLLIKTCHKRGVHAMGGMAAQIPIKDNAEANDKAMEGVRADKLREVRAGHDGTWVAHPALAS
      * ***** *** C*  ******************C********* *C  ***** *********C  **************** ****************
E.c.  GRWDYIFSYIKTLKNYPDRVLPDRQAVTMDKPFLNAYSRLLIKTCHKRGAFAMGGMAAFIPSKDE-EHNNQVLNKVKADKSLEANNGHDGTWIAHPGLAD
B.n.  GRWDYIFSYVKTFQAHPDRLLPDRVLVGMGQHFMRSYSDLLIRTCHKRGVHAMGGMAAQIPIRDDPKANEMALDLVKKDKLREVRAGHDGTWAAHPGLIP
C.s.  GRWDYIFSYVKTFQAHPDRLLPDRVLVGMTQHFMRSYSDLLIRTCHRRGVHAMGGMAAQIPIRDDPKANEVALELVRKDKLREVKAGHDGTWAAHPGLIP
      * ***** CC* C   *C C****  * *   *C  *  ***C* *C    *** *** ** C*C   *C  CC *C **  *   ****** *** *

N.c.  IALEVFNKHMPTPNQLFN-RREDVK-IGQQDLLNMNVPGSSTEDGIRKNLNTGLGYTEPWIRGVGCVPIKHPQEDAATAEVSRSQLWQWVKHRVTTAEGK
A.n.  IASEVFNKYMPTPNQMHV-RREDVN-ITANDLLNTNVPGKITEDGIRKNLNIGLSYMEGWLRGVGCIPINYLMEDAATAEVSRSQLWQWARHGVTTSEGK
      ** ***** ******C   *****  *  C**** ****  ********** ** * * *C*****C**    **************** C* *** ***
E.c.  TAMAVFNDILGSRKNQLEVMREQDAPITADQLLAPCD-GERTEEGMRANIRVAVQYIEAWISGNGCVPIYGLMEDAATAEISRTSIWQWIHHQKTLSNGK
B.n.  ICMDAFSHMGNNPNQIKSMKRDDASAITEEDLLQIPR-GVRTLEGLRLNTRVGIQYLAAWLTGSGSVPLYNLMEDAATAEISRVQNWQWIRYGVEL-DGD
C.s.  ACMEVFTNMGNAPNQIRSMRRDDAANLTEEDLLQQPR-GVRTMEGLRLNTRVGIQYLAAWLTGAGSVPLYNLAEDAATAEISRVQNWAWLKYGVEL-DGD
           *        C     *CC      CC**     *  * C*C* *    C * *   *C * * C*C    *******C**   * * C      C*

N.c.  H----VDKRYPLKLLKEADRQRLAKAPQGNKFNLAAQY------FASQVTG-EDYADFLTCLLYNEITSAGNSLPASKL
A.n.  K----VDKAYALRLLKEQADALAAKGPKGNKFQLAGRY------FSEQVTG-EDYADFLTSLLYNEISSPG-T--ASKL
      C    *** * *C****      ** * ****C**  *      *  **** ******** ******C* * C  ****
E.c.  P----VTKALFRQMLGEEMKVIASELGE-ERFSQ-GR-FDDAARLMEQITTSDELIDFLTLPGYRLLA
B.n.  GLGVRVSKELFGRVVEEEMERIEKEVGK-DKFKR-GMYKEACKMFTKQCTAAE-LDDFLTLAVYDHIVAHYPIN-ASRL
C.s.  GLGVRVNKELFGRVVEEEMERIEREVGK--RFKK-GMYKEACKMFTRQCTAPN-LDDFLTLDAYNYIVIHHPRE-LSKL
           * * C   CC *             C*               * *  C C ****   *  C         *C*
```

Fig. 5. Comparison of the *acuE, acu-9, aceB, Brassica napus* and *Cucumis sativus* amino acid sequences. Identical amino acids are indicated by *, conservative substitutions by C. *Gaps,* indicated by –, have been introduced into the sequences to maximise the homology. Conservative substitutions were determined according to Dayhoff et al. (1972) and Kos et al. (1988). N.c. *Neurospora crassa*, A.n. *Aspergillus nidulans*, E.c. *Eschericha coli*, B.n. *B. napus*, C.s. *C. sativus*

the *facA* coding region, towards the C-terminal end of the protein, and occur in regions that are highly conserved between the *facA* and *acu-5* genes in both DNA and amino acid sequence. In contrast the *acu-5* gene of *N. crassa* contains only one intron which matches the position of the final *facA* intron (Connerton et al. 1990). Thus, the three comparisons *qut/qa*, *facA/acu-5* and *acuE/acu-9* all give similar results. In each case, the *A. nidulans* gene contains more introns in comparison with its *N. crassa* counterpart, with the intron(s) common to both confined to the downstream end of the coding sequence. The NADP-specific glutamate dehydrogenase genes *am* (*N. crassa*) and *gdhA* (*A. nidulans*) each have two introns, in precisely corresponding positions (Hawkins et al. 1989). Comparison of *A. nidulans areA* (Kudla et al. 1990) and *N. crassa nit-2* (Fu and Marzluf 1990a), coding for regulators of nitrogen catabolism, show a high degree of similarity in predicted amino acid sequences except that the *nit-2* product has an N-terminal sequence of 218 amino acid residues not present in the *areA* protein. Whereas *nit-2* contains two introns, one in the 218 codon extension, *areA* has none. In spite of the gross differences at the N-terminal ends of these proteins the *nit-2* gene has been successfully transformed into an *A. nidulans areA*$^{-}$ strain and complemented the lost *areA* protein activity (Davis and Hynes 1987).

It appears that in the evolution of the filamentous Ascomycetes introns have been acquired or lost rather freely. Whether gain or loss is the predominant event is unclear. One conclusion that does emerge is that the coding sequences flanking fungal introns are strongly conserved; there is no indication of imprecise removal or insertion, nor of any change in intron/exon boundaries such as might result from the adoption of alternative donor or receptor splice sites.

Despite the structural differences between homologous genes in *A. nidulans* and *N. crassa,* particularly in the case of the *facA/acu-5* and *areA/nit-2* genes where the N-terminal regions are very different (Connerton et al. 1990; Fu and Marzluf 1990a; Kudla et al. 1990), these genes perform similar functions in their respective organisms. The *facA/acu-5, qutE/qa-2* and *acuE/acu-9* genes have also been *trans*-genically expressed in the alternative hosts (Connerton et al. 1990; Hiatt and Case 1990). These results suggest that gross structural differences between homologous genes have little effect on their correct expression and function. Moreover, the fact that some induction of these genes can occur transgenically, e.g. *facA/acu-5* (Connerton et al. 1990) and *acuE/acu-9* (Thomas et al. 1988), suggests that similar regulatory proteins and binding sequences are shared by *A. nidulans* and *N. crassa.* The *nit-2* gene has been shown to function in vivo in *A. nidulans* (Davis and Hynes 1987), which is consistent with the recent demonstration of in vitro binding of *nit-2* protein to the *A. nidulans niiA/niaD* intergenic region (Fu and Marzluf 1990b).

The comparison of the 5′ non-coding regions of coordinately regulated genes has proved useful in identifying regulatory protein binding sites in a number of organisms. As the *acuE* and *acu-9* genes can be transgenically expressed and induced it might be expected that these two genes would share a common regulatory protein binding site. The *acuE* gene is known to be positively controlled by the presence of the *facB* protein and the inducer acetate at the level of transcription (Sandeman and Hynes 1989). The fact that induced expression of the *acu-9* gene in *A. nidulans* can be achieved would suggest that the *facB* protein can recognise a binding sequence in the 5′ region of the *acu-9* gene.

Analysis of the promoter region of the *amdS* gene of *A. nidulans,* which is also regulated by the *facB* gene, has revealed a sequence involved in *facB*-mediated acetate induction. This sequence was identified by a single base pair mutation, *amdI9*, that causes increased inducibility of the *amdS* gene by acetate that is *facB* dependent (Hynes 1975, 1977; Hynes et al. 1988). A 21 bp oligonucleotide including the *amdI9* sequence has been found to be sufficient to confer *facB*-dependent acetate inducibility on an *amdS-lacZ* fusion construct (T.G. Littlejohn and M.J. Hynes, unpublished data). Comparison of the 5′ regions of the *acuE* and *acu-9* genes with the *amdI9* region sequence indicates a number of related pyrimidine-rich sequences. Similar sequences are found in the 5′ regions of the *facA* and *acu-5* genes which are also regulated by acetate (Table 1; Connerton et al. 1990). These sequences are potential binding sites for the *facB,* or *N. crassa* equivalent, protein. However, the homologies are weak and detailed functional analysis is required to investigate their significance.

In organisms that contain a glyoxylate bypass the isocitrate lyase and malate synthase enzymes have been found to be located within a subcellular organelle, the glyoxysome, and this is also the case for *A. nidulans* and *N. crassa* (Dijkema and Visser 1987; Zimmerman and Neupert 1980; Thomas and Baxter 1987). Recent work on peroxisomally located proteins has identified a tripeptide at the C-terminus of these proteins that is both necessary and sufficient to direct protein transport into the peroxisome (Gould et al. 1987, 1988, 1989). The tripeptide from the luciferase gene has been extensively studied using site-directed mutagenesis to alter the sequence of the residues and immunofluorescence to examine the partitioning of the mutant proteins (Gould et al. 1989). It was found that the first two residues of the tripeptide could be either one of three possible amino acids but that the final leucine residue may be fixed. Thus, this peroxisomal targeting sequence is likely to conform to the sequence S/A/C−K/H/R−L (Gould et al. 1989). A survey of peroxisomal and glyoxysomal proteins showed that a number of the proteins contained this sequence at the C-terminus, including the cucumber and rapeseed malate synthase proteins (Comai et al. 1989; Smith and Leaver 1986). However, this sequence is not found at the C-terminus of all peroxisomally located proteins (Osumi and Fujika 1990). For example, acyl CoA oxidase of *Candida tropicalis* was shown to have two internal regions that directed transport into the peroxisome. Although this protein did contain an S−K−L sequence, the tripeptide was not associated with the targeting regions (Small et al. 1988). Thus, a number of different topogenic signals that direct protein transport into peroxisome-like organelles are likely to occur.

The malate synthases from both *A. nidulans* and *N. crassa,* which have been shown to be localised in glyoxysomes (Dijkema and Visser 1987; Zimmerman and Neupert 1980), do conform to the consensus defined by Gould et al. (1989). Both have the sequence serine-lysine-leucine at the C-terminus. This sequence is also shared by the *B. napus* (S−R−L) and *C. sativus* (S−K−L) proteins but is absent from the *aceB* protein (Fig. 5).

Table 1. *amdI9*-like sequence homologies in the 5′ regions of the *acuE, acu-9, facA* and *acu-5* genes

Gene	Position	Sequence	Strand
amdI9	− 219	CTGCAGCTTCCC	+
acuE	−1355	CTCGAGATTCCT	+
	−1174	GTTCATGTTCCC	−
	−1153	CGGCCTCTTCCC	−
	− 724	CGACATCTTCCG	−
	− 682	GGGCAGTTTCAC	−
	− 308	TTTCAACTTCAC	+
acu-9	− 626	CTCGAGGTTCCT	−
facA	−1349	CGCCTGCTTCCC	+
	− 97	CTGCATTTTCCC	+
acu-5	− 517	CTGCTTGTTCGC	−
	− 449	TTTCAGATTCCA	+

All sequences are positioned, from the start of the sequence, in comparison with the startpoint of translation. + represents the coding strand and − the non-coding strand

Acknowledgements. This work was supported in part by the Australian Research Grants Committee and a Commonwealth Postgraduate Research Award to R.A.S.

References

Apirion D (1965) The two-way selection of mutants and revertants in respect of acetate utilization and resistance to fluoroacetate in *Aspergillus nidulans*. Genet Res 6:317–329

Armitt S, McCullough W, Roberts CF (1976) Analysis of acetate non-utilizing (*acu*) mutants in *Aspergillus nidulans*. J Gen Microbiol 92:263–282

Ballance DJ, Turner G (1986) Gene cloning in *Aspergillus nidulans*: isolation of the isocitrate lyase gene (*acuD*). Mol Gen Genet 202:271–275

Burke JF (1984) High-sensitivity S1 mapping with single-stranded (^{32}P) DNA probes synthesized from bacteriophage M13mp9 templates. Gene 30:63–68

Byrne C, Stokes HW, Ward KA (1988) Nucleotide sequence of the *aceB* gene encoding malate synthase A in *Escherichia coli*. Nucleic Acids Res 16:10924

Comai L, Baden CS, Harada JJ (1989) Deduced sequence of a malate synthase polypeptide encoded by a subclass of the gene family. J Biol Chem 264:2778–2782

Connerton IF (1990) Premeiotic disruption of the *Neurospora crassa* malate synthase gene by native and divergent DNAs. Mol Gen Genet 223:319–323

Connerton IF, Fincham JRS, Sandeman RA, Hynes MJ (1990) Comparison and cross-species expression of the acetyl-CoA synthetase genes of the ascomycete fungi, *Aspergillus nidulans* and *Neurospora crassa*. Mol Microbiol 4:451–460

Dale RMK, McClure BA, Houchins JP (1985) A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18s rDNA. Plasmid 13:31–40

Davis MA, Hynes MJ (1987) Complementation of *areA*$^-$ regulatory gene mutations of *Aspergillus nidulans* by the heterologous regulatory gene *nit-2* of *Neurospora crassa*. Proc Natl Acad Sci USA 84:3753–3757

Dayhoff MO, Eck RV, Park CM (1972) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5. National Biomedical Research Foundation, Washington DC. pp 89–99

Dijkema C, Visser J (1987) ^{13}N-NMR analysis of *Aspergillus* mutants disturbed in pyruvate metabolism. Biochim Biophys Acta 931:311–319

Flavell RB, Fincham JRS (1968) Acetate non-utilizing mutants of *Neurospora crassa*. I. Mutant isolation, complementation studies and linkage relationships. J Bacteriol 95:1056–1062

Fu Y-H, Marzluf GA (1990a) *nit-2*, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain. Mol Cell Biol 10:1056–1065

Fu Y-H, Marzluf GA (1990b) *nit-2*, the major positive-acting nitrogen regulatory gene of *Neurospora crassa*, encodes a sequence-specific DNA-binding protein. Proc Natl Acad Sci USA 87:5331–5335

Geever RF, Huiet L, Baum JA, Tyler BM, Patel VB, Rutledge BJ, Case ME, Giles NH (1989) DNA sequence, organization and regulation of the *qa* gene cluster of *Neurospora crassa*. J Mol Biol 207:15–34

Gould SJ, Keller G-A, Subramini S (1987) Identification of a peroxisomal targeting signal at the carboxy-terminus of firefly luciferase. J Cell Biol 105:2923–2931

Gould SJ, Keller G-A, Subramini S (1988) Identification of peroxisomal targeting signals located at the carboxy-terminus of four peroxisomal proteins. J Cell Biol 107:897–905

Gould SJ, Keller G-A, Hosken N, Wilkinson J, Subramini S (1989) A conserved tripeptide sorts proteins to peroxisomes. J Cell Biol 108:1657–1664

Gurr SJ, Unkles SE, Kinghorn JR (1988) The structure and organisation of nuclear genes of filamentous fungi. In: Kinghorn JR (ed) Gene structure in eukaryotic microbes, vol 23. SGM Special Publications, IRL Press, Oxford, pp 93–139

Hawkins AR, Lamb HK, Smith M, Keyte JW, Roberts CF (1988) Molecular organisation of the quinic acid utilization (*QUT*) gene cluster in *Aspergillus nidulans*. Mol Gen Genet 214:224–231

Hawkins AR, Gurr SJ, Montague P, Kinghorn JR (1989) Nucleotide sequence and regulation of expression of the *Aspergillus nidulans gdhA* gene encoding NADP dependent glutamate dehydrogenase. Mol Gen Genet 218:105–111

Hiatt KL, Case ME (1990) Induced expression of the *Aspergillus nidulans QUTE* gene introduced by transformation into *Neurospora crassa*. Mol Gen Genet 222:201–205

Hynes MJ (1975) A *cis*-dominant regulatory mutation affecting enzyme induction in the eukaryote *Aspergillus nidulans*. Nature 253:210–212

Hynes MJ (1977) Induction of the acetamidase of *Aspergillus nidulans* by acetate metabolism. J Bacteriol 131:770–775

Hynes MJ, Davis MA (1986) The *amdS* gene of *Aspergillus nidulans*: control by multiple regulation signals. BioEssays 5:123–128

Hynes MJ, Corrick CM, Kelly JM, Littlejohn TG (1988) Identification of the sites of action for regulatory genes controlling the *amdS* gene of *Aspergillus nidulans*. Mol Cell Biol 8:23589–2596

Kornberg HL (1966) The role and control of the glyoxylate cycle in *Escherichia coli*. Biochem J 99:1–11

Kos T, Kuijvenhoven A, Hessing HGM, Pouwels PH, van den Hondel CAMJJ (1988) Nucleotide sequence of the *Aspergillus niger trpC* gene: structural relationship with analogous genes of other organisms. Curr Genet 13:137–144

Kudla B, Caddick MX, Langdon T, Martinez-Rossi NM, Bennet CF, Sibley S, Davies RW, Arst HN Jr (1990) The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. EMBO J 9:1355–1364

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564

Norrander J, Kempe T, Messing J (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101–106

Osumi T, Fujiki Y (1990) Topogenesis of peroxisomal proteins. BioEssays 12:217–222

Sandeman RA, Hynes MJ (1989) Isolation of the *facA* (acetyl-CoenzymeA synthetase) and *acuE* (malate synthase) genes of *Aspergillus nidulans*. Mol Gen Genet 218:87–92

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Selker EU, Garrett PW (1988) DNA sequence duplications trigger gene inactivation in *Neurospora crassa*. Proc Natl Acad Sci USA 85:6870–6874

Selker EU, Cambareri EB, Jensen BC, Haack KR (1987) Rearrangement of duplicated DNA in specialized cells of *Neurospora*. Cell 51:741–752

Small GM, Szabo LJ, Lazarow PB (1988) Acyl-CoA oxidase contains two targeting sequences each of which can mediate protein import into peroxisomes. EMBO J 7:1167–1173

Smith SM, Leaver CJ (1986) Glyoxysomal malate synthase of cucumber: molecular cloning of a cDNA and regulation of enzyme synthesis during germination. Plant Physiol 81:762–767

Thomas GH, Baxter RL (1987) Analysis of mutational lesions of acetate metabolism in *Neurospora crassa* by ^{13}C nuclear magnetic resonance. J Bacteriol 169:359–366

Thomas GH, Connerton IF, Fincham JRS (1988) Molecular cloning, identification and transcriptional analysis of genes involved in acetate utilization in *Neurospora crassa*. Mol Microbiol 2:599–606

Zimmerman R, Neupert W (1980) Biogenesis of glyoxysomes: synthesis and intracellular transfer of isocitrate lyase. Eur J Biochem 112:225–233

Communicated by C.A. v. den Hondel