http://www.pjbs.org

PJBS

ISSN 1028-8880

Pakistan Journal of Biological Sciences

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Pakistan Journal of Biological Sciences 5(8): 871-877, 2002

Molecular Cloning and Sequencing of D-mandelate Dehydrogenase Gene from *Rhodotorula graminis*

[1]Rosli Md.Illias, [2]Graem A. Reid, [3]Stephen K. Chapman, [4]Charles A. Fewson and [4]John S. Miles
[1]Department of Bioprocess Engineering, Faculty of Chemical Engineering and Natural Resources Engineering, Universiti Teknologi Malaysia, 81300, Skudai, Johor Malaysia
[2]Institute of Cell and Molecular Biology, University of Edinburgh, Maryland Road, Edinburgh EH9 3JR, Scotland, UK
[3]Department of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, Scotland, UK
[4]Institute of Medical and Life Sciences, University of Glasgow, Glasgow G12 8QQ

Abstract: The yeast *Rhodotorula graminis* can use D, L-mandelate as a source of carbon and energy. We have isolated the gene encoding D-mandelate dehydrogenase, one of the two enzymes that stereospecifically catalyze the first step in mandelate degradation. The sequences of the genomic DNA and a cDNA prepared by RT-PCR revealed the presence of three short introns within the coding region. The predicted amino acid sequence of D-mandelate dehydrogenase is 27-33% identical to other members of a large family of NAD^+-dependent 2-hydroacid dehydrogenase from a broad spectrum of bacteria and eukaryotes and it has a wide range of substrate specificities.

Key words: Mandelate dehydrogenase, *Rhodotorula*, gene, cloning, sequencing, yeast, bacteria

Introduction

The ability to utilize mandelate as a source of carbon and energy for growth has been found in a small but diverse range of bacteria and fungi (Fewson, 1988). The pathways for mandelate catabolism vary between different groups of organisms but the first step is generally an oxidation of mandelate (2-hydroxy-2-phenylacetate) to phenylglyoxylate. Several different types of enzymes have evolved to catalyze this reaction (Fewson, 1988). The yeast, *Rhodotorula graminis*, oxidizes both enantiomers of mandelate through the action of stereospecific mandelate dehydrogenase (Durham, 1984). L-mandelate dehydrogenase is a mitochondrial flavocytochrome b_2 that transfers electrons to cytochrome *c* (Yasin and Fewson, 1993, Smekal *et al.*, 1993). In contrast, D-mandelate dehydrogenase is a cytoplasmic NAD^+-dependent enzyme (Baker and Fewson, 1989). The purified protein is a homodimer of 38 KDa subunits and crystals diffracting to 0.25 nm have previously been obtained (Basak *et al.*, 1993). Amino acid sequence determination of about 30% of the polypeptide chain failed to reveal relationships with other known proteins. We have determined the complete sequence from the cloned cDNA to allow a better understanding of the evolution of D-mandelate dehydrogenase from an ancestral 2-hydroxy acid dehydrogenase. D-mandelate dehydrogenase has previously been used for the production of D-mandelate, which is a useful chiral synthon in the production of a range of pharmaceuticals, including semisynthetic β-lactam antibiotics (Yamazaki and Maeda, 1986; Vasik-Racki *et al.*, 1989, Hosono *et al.*, 1990). The yield of enzyme from *R. graminis* is limited but enzyme production could be greatly enhanced using a heterologous expression system. Production of recombinant enzyme would also allow the preparation of altered forms of D-mandelate dehydrogenase, for example with altered substrate specificity. The objective of this research is to clone, sequence and express D-mandelate dehydrogenase gene from *Rhodotorula graminis*.

Materials and Methods

Strains and plasmids: *Rhodotorula graminis* GX6000 (ATCC20804) was used as a source of RNA and DNA and was maintained, grown, harvested and stored as described previously (Baker and Fewson, 1989). *E. coli* strains TG1 and NF1 were used as a host for recombinant plasmids. The plasmids pTZ19r and pTZ18r were used for cloning (Rokeach *et al.*, 1988). This project was carried out at University of Edinburgh, Scotland from 1995-1997.

DNA isolation: Chromosomal DNA was isolated from a 10 ml culture of stationary phase culture grown in YPD medium [1% (w/w) yeast extract, 2% (w/v) peptone and 2% (w/v) glucose]. The cells were harvested by centrifugation and re-suspended in 1 ml of breakage buffer (0.9 M sorbitol/14 mM 2-mercaptoethanol /50 mM sodium phosphate buffer, pH 7.5). The cells were then disrupted by vortexing with acid-washed glass beads. To this suspension 50 μl of 0.5 M EDTA, pH 8.9 was added, vortexed briefly and then 50 μl of 10% SDS and 100 μl of proteinase K solution (5 mg/ml) were added to help lysis. The mixture was mixed well and incubated at 65°C for 30 min then extracted with 1:1 phenol: chloroform and the DNA was precipitated by the addition of 0.5 ml of absolute ethanol. Plasmid DNA and single-stranded plasmid DNA were isolated from *E. coli* transformants as described previously (Sambrook *et al.*, 1989; Viera and Messing, 1987). M13KO7 was used as helper phage for single stranded DNA production.

Isolation of RNA: RNA was isolated from a 100 ml culture of *R. graminis* grown until mid-exponential phase in medium containing D, L-mandelate. The cells were harvested and re-suspended in 1 ml of TNE (50 mM Tris.HCl pH 7.5, 100 mM NaCl, 5 mM EDTA). Acid-washed glass beads were used to disrupt the cells with vigorous vortexing for 2 minutes. Then 4 ml of TNE, 0.2 ml 20% SDS and 4 ml phenol were rapidly added and the suspension was vortexed for an other 2 min. The mixture was spun for 15 min to separate the phases. The aqueous phase containing RNA was removed and extracted with 1:1, phenol:chloroform until a clear interface was achieved. The upper phase containing the RNA was removed and to this 0.1 volume of 3 M sodium acetate, pH 5.5 and 2 volumes of 100% ethanol were added to precipitate RNA. Before use in RT-PCR, contaminating DNA was removed from RNA by treatment with DNaseI. A 100 μl mixture containing 100 μg of total RNA and 10 units DNaseI in 20 mM Tris.HCl, pH 8.4, 12 mM $MgCl_2$ and 50 mM KCl was incubated at 37°C for 1 h. The reaction was stopped by heating to 65°C for 10 min and extracted with phenol-chloroform. The RNA was precipitated with ethanol, pellet dried and dissolved in 50 μl H_2O.

Polymerase chain reaction: A polymerase chain reaction was performed to synthesize a probe for D-mandelate dehydrogenase gene, containing: 1.7 μg *R. graminis* DNA; 1 μg each of the forward primer P33 (GGAATTCGAYTTYCARCARAARTTYGA; corresponding to the amino acid sequence AFQQKFE which is found at positions 21-27 from the N-terminus) and reverse P34 (CCGGATCCGRCCARTCRAANCCNGC; corresponding to AGFDWL which is found at position 7-12 of a tryptic peptide); 2.5 units *Taq* polymerase (Promega) in 1 x reaction buffer (Promega) in a final volume of 100 μg. The DNA was melted at 94°C for 2 min and extension at 72°C for 3 min was performed. A final extension at 72°C for 7 min was carried-out to complete the reaction.

Construction of *R. graminis* DNA libraries: Approximately 10 μg aliquots of *R. graminis* genomic DNA were digested with *Pst*I, *Bam*HI, *Eco*RI, *Sma*I, *Xba*I, *Sph*I and *Sac*I, separated on a 0.8% agarose gel and transferred to a nylon membrane (Hybond-N, Amersham, UK). The membrane was pre-hybridized for 1 h at 65°C, then hybridization was carried-out overnight at 65°C. The probe was prepared by isolating the 320 bp PCR product and labeling by random priming (Feinberg and Vogelstein, 1983). After washing with increasing stringency the membrane was allowed to dry and autoradiographed at −70°C. The genomic DNA library was constructed by digestion of genomic DNA with *Sac*I. The digested genomic DNA was ligated to plasmid pTZ19r cut with *Sac*I. About 10000 recombinants were screened by hybridization under the same conditions as for the Southern blotting.

First strand cDNA synthesis: A mixture of 1 μl of oligo (dT) $_{12-18}$ (500 μg/ml) and 15-20 μg of total *R. graminis* RNA (treated with DNaseI) in 10 μl of sterile distilled water was heated to 70°C for 10 min and then quickly chilled on ice. The contents of the tube were collected by brief centrifugation, mixed with 4 μl 5x first strand buffer (Gibco BRL, USA), 2 μl of 0.1 M DTT, 2 μl of 5 mM dNTPs and 1μl (200 units) Supercript™ II RNase H⁻ reverse transcriptase (Gibco BRL, USA) and incubated at 37°C for 1 h. The products were used immediately for PCR or stored at 20°C.

RT-PCR: The forward primer N1179 (CAAGGAATTCATGCCTCGCCCTCGCGT) and reverse primer R13 (CCACTGCAGTCAGTAGGCGCGAAAAGC) were designed for amplification of the complete D-mandelate dehydrogenase coding sequence. These incorporated cleavage sites for *EcoRI* and *Pst*I, respectively, to facilitate cloning of the PCR product. PCR was performed in a 50 μl reaction with 15 mM $MgCl_2$, 5 p mol of each primer, 200 μM dNTPs and 1 μl of the reverse transcriptase reaction products. After denaturing at 95°C for 5 min, 3 cycles were carried-out with: 95°C for 40 sec, 50°C for 30 sec and 72°C for 90 sec. A further 40 cycles were completed under the same conditions except that the annealing temperature was raised to 64°C. The reaction was completed by further 7 min incubation at 72°C.

DNA sequence determination and analysis: The DNA sequence was determined on both strands using dideoxy chain termination methods (Sanger *et al.*, 1977) with the Sequenase (US Biochemical Corp.) T7 polymerase. DNA sequence information was analyzed using the Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison and Wisc.

Western blotting: Proteins were separated by SDS-polyacrylamide gel electrophoresis and electrophoretically transferred to a nylon membrane (Hybond-N) as described previously (Haid and Suissa, 1984). D-mandelate dehydrogenase was detected using antiserum raised in rabbit followed by horseradish peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad) as secondary antibody. After washing, enzyme activity was visualized using *o*-dianisidine as a substrate.

Enzyme assay: D-mandelate dehydrogenase activity was determined following the reverse reaction in 1.4 ml assay mixtures containing 200 mM phosphate buffer (pH 5.85), 200 μM-NADH, 1 mM phenylglyoxylate and enzyme (crude cell extract). The oxidation of NADH was monitored at 340 nm.

Results and Discussion

PCR amplification of a gene fragment: N-terminal amino acid sequence was obtained from purified D-mandelate dehydrogenase and from three tryptic peptides, enabling us to design fully degenerate oligonucleotide primers for the polymerase chain reaction. A PCR with genomic DNA from *R. graminis* amplified a DNA fragment of 320 bp that was then cloned into M13mp19. The sequence of this fragment was shown to encode the N-terminal region of D-mandelate dehydrogenase. A small intron was also detected within this region by comparing the DNA sequence with amino acid sequence (Fig. 1).

Isolation and sequencing of the D-mandelate dehydrogenase gene: *Rhodotorula graminis* chromosomal DNA was digested with seven different restriction enzymes. None of the restriction enzymes used to digest the genomic DNA cut within 320 bp fragments. Southern blot hybridization was carried-out on the digested chromosomal DNA. The blot was probed with 320 bp fragment that had been ^{32}P-labeled by random priming. The audio radiograph of the Southern blot (Fig. 2) showed that at least part of the D-mandelate dehydrogenase gene was contained within 4.4 kb *Sac*I fragment (Fig. 2, lane 8).

A genomic library was constructed from *R. graminis* DNA digested to completion with *Sac*I. The digested DNA was ligated to pTZ19r that had also been cut with *Sac*I. Transformants containing plasmids with inserts were identified as white colonies on plates containing X-gal and IPTG. Colony blotting using the same probe as for Southern blot screened approximately 10000 recombinants. A single positive clone (pRI1) was identified. Plasmid from this positive clone was purified, cut with *Sac*I and shown to contain an insert of the expected size, 4.4 kb.

The entire D-mandelate dehydrogenase coding region was sequenced on both DNA strands. Subclones of the 4.4 kb *Sac*I fragment in pTZ18r and pTZ19r were sequenced with a universal primer. The remaining sequence was obtained using primers that were designed according to the experimentally determined sequence. The complete coding sequence is contained within the cloned *Sac*I fragment. The 1630 bp of assembled sequence includes the PCR fragment previously sequenced. Translation in all three reading frames identified peptide sequences corresponding to those that had been determined experimentally but in different reading frames, indicating the presence of further introns within the coding region, in addition to the one already identified in 320 bp PCR fragment. Isolation of cDNA was therefore necessary for the unambiguous determination of the protein coding sequence and for expressions of recombinant D-mandelate dehydrogenase.

Isolation and sequencing of cDNA: Two PCR primers were designed based on the known N-terminal sequence of the protein and the C-terminal sequence predicted from the genomic DNA. An *Eco*RI cleavage site was incorporated in the forward primer and a *Pst*I restriction site in the reverse primer to facilitate the cloning of product. Total RNA from *Rhodotorula graminis* was used as a template for synthesis of single-stranded cDNA by reverse transcription (see Materials and Methods) and this was then used in PCR. The polymerase chain reaction was carried-out at 95°C for 5 min initial denaturing then 35 cycles of: 95°C for 40 s denaturing, 64°C annealing for 30 s and 72°C extension for 1.5 min. Finally another 5 min extension at 72°C was carried-out to complete the reaction. The resulting fragment of approximately 1 kb was treated with Klenow fragment then cut with *Eco*RI and ligated with pTZ19r that had been cut with *Eco*RI and *Sma*I to generate the recombinant plasmid pR13. The cDNA in pR13 was recloned into pTZ18r to obtain the alternative orientation (pR14) for sequencing the second strand. The same primers that were used to sequence the genomic DNA encoding D-mandelate dehydrogenase were also used to sequence the cDNA with single-

```
P33
EcoRI
GAATTCGACTTTCAGCAGAAATTTGAAGTCATCCCTGCCAACCTGACCACGCACGACGGG
       AspPheGlnGlnLysPheGluValIleProAlaAsnLeuThrThrHisAspGly

TTTAAACAGGCCCTGCGCGAGAAGCGGTGCGTTCTGTTCCTGGCCCTCACGCGGTTCTTC
PheLysGlnAlaLeuArgGluLysAr

CGCTGACCGCATTCTCCGCCGCATCACTTGTGCTTCTCCCCGTCGTACGCAGCTATGGCG
                       INTRON                          gTyrGlyA

ACTTCGAAGCCATCATCAAGCTTGCCGTCGAGAACGGCACCGAGAGCTATCCCTGGAAGC
spPheGluAlaIleIleLysLeuAlaValGluAsnGlyThrGluSerTyrProTrpAsnA

CCGACCTCATCTCGCACCTCCCTTCGTCCCTCAAAGTCTTTGCCGCCGCCGGCGCAGGTT
laAspLeuIleSerHisLeuProSerSerLeuLysValPheAlaAlaAlaGlyAlaGlyP

TTGATTGGCCGGATCC
heAspTrp    BamHI
               P34
```

Fig.1: The sequence of a 320 bp D-mandelate dehydrogenase gene fragment amplified from *R. graminis* genomic DNA. The regions corresponding to the PCR primer (P33 and P34) sequences are underlined – in the case of P34 the sequence shown is the compliment of the primer sequence. The amino acid sequence predicted from the DNA sequence matches perfectly the sequenced determined directly from the protein assuming the presence of an intron as indicated in bold.

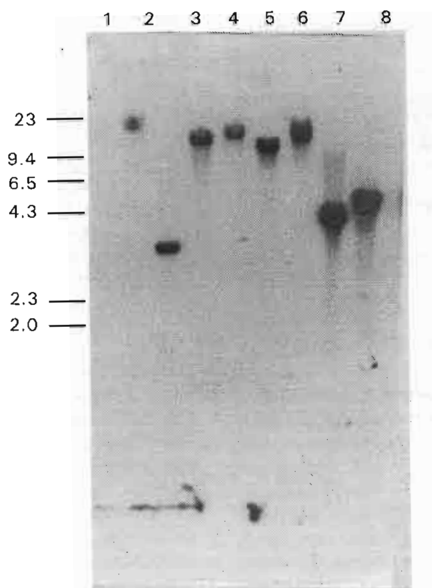

Fig. 2: Southern blot hybridization of *R. graminis* genomic DNA. Aliquots of *R. graminis* DNA were subjected to 0.8% agarose gel electrophoresis without treatment (lane 1) or after digestion with *Pst*I (lane 2), *Bam*HI (lane 3), *Eco*RI (lane 4), *Sma*I (lane 5), *Xba*I (lane 6), *Sph*I (lane 7) or *Sac*I (lane 8). The gel was blotted onto Hybond-N nylon membrane and hybridized with the cloned 320 bp PCR fragment that had been labeled with [^{32}P]-dCTP by random priming.

```
   1 tctcgaccagttcccgcacaaagtgggggcgatttgtgctggtggggcgcccggagctggcccctctgcttcttacagtccacttgcgct
     ..........................................................................................

  91 acgctgcttcttgtcgtgcccctttcttgctgctgtccgacttccgccatgcctcgccctcgcgtccttctcctcggcgaccccgctcgg
     ..............................................atgcctcgccctcgcgtccttctcctcggcgaccccgctcgg
                                                    M  P  R  P  R  V  L  L  L  G  D  P  A  R

 181 cacctcgacgacctctggagcgacttccagcaaaagttcgaagtcatccctgccaacctgaccacgcacgacgggtttaaacaggccctg
     cacctcgacgacctctggagcgacttccagcaaaagttcgaagtcatccctgccaacctgaccacgcacgacgggtttaaacaggccctg
     H  L  D  D  L  W  S  D  F  Q  Q  K  F  E  V  I  P  A  N  L  T  T  H  D  G  F  K  Q  A  L

 271 cgcgagaagcggtgcgttctgttcctggccctcacgcggttcttccgctgaccgcattctccgccgcatcacttgtgcttctccccgtcg
     cgcgagaagcg.....................................INTRON 1..................................
     R  E  K  R

 361 tacgcagctatggcgacttcgaagccatcatcaagcttgccgtcgagaacggcaccgagagctatccctggaacgccgacctcatctcgc
     .......ctatggcgacttcgaagccatcatcaagcttgccgtcgagaacggcaccgagagctatccctggaacgccgacctcatctcgc
             Y  G  D  F  E  A  I  I  K  L  A  V  E  N  G  T  E  S  Y  P  W  N  A  D  L  I  S  H

 451 acctcccttcgtccctcaaagtctttgccgccgccggcgcaggcttcgactggctcgacctcgacgcactcaacgagcgcggagtcgcgt
     acctcccttcgtccctcaaagtctttgccgccgccggcgcaggcttcgactggctcgacctcgacgcactcaacgagcgcggagtcgcgt
       L  P  S  S  L  K  V  F  A  A  A  G  A  G  F  D  W  L  D  L  D  A  L  N  E  R  G  V  A  F

 541 ttgccaactcgcgcggcgcaggcgacacggcgacatccgatctcgcgctgtacctcatcctgtccgtcttccgcctcgcgagctactcgg
     ttgccaactcgcgcggcgcaggcgacacggcgacatccgatctcgcgctgtacctcatcctgtccgtcttccgcctcgcgagctactcgg
       A  N  S  R  G  A  G  D  T  A  T  S  D  L  A  L  Y  L  I  L  S  V  F  R  L  A  S  Y  S  E

 631 agcgcgccgcgcgcacgggcgaccccgagacgttcaaccgcgtgcacctcgagattggcaagtcggcgcacaacccgcgcgggcacgtcc
     agcgcgccgcgcgcacgggcgaccccgagacgttcaaccgcgtgcacctcgagattggcaagtcggcgcacaacccgcgcgggcacgtcc
       R  A  A  R  T  G  D  P  E  T  F  N  R  V  H  L  E  I  G  K  S  A  H  N  P  R  G  H  V  L

 721 tcggcgcggtcgggctcggcgcgatccagaaggagatcgcgaggaaggcggtgcatggccttgggatgaagctcgtctactacgacgtcg
     tcggcgcggtcgggctcggcgcgatccagaaggagatcgcgaggaaggcggtgcatggccttgggatgaagctcgtctactacgacgtcg
       G  A  V  G  L  G  A  I  Q  K  E  I  A  R  K  A  V  H  G  L  G  M  K  L  V  Y  Y  D  V  A

 811 cgcctgccgacgcggagacggagaaggcgctcggtgctgagcgcgtcgactcgctcgaagagctggcgaggaggagcgactgtgtcagcg
     cgcctgccgacgcggagacggagaaggcgctcggtgctgagcgcgtcgactcgctcgaagagctggcgaggaggagcgactgtgtcagcg
       P  A  D  A  E  T  E  K  A  L  G  A  E  R  V  D  S  L  E  E  L  A  R  R  S  D  C  V  S  V

 901 tgtcgggtgagtctttccacttcgtcatcggggaacccgtcctcttggcactcaccagtatgcaccacacagtgccgtatatgaagttga
     tgtcgg...............................INTRON 2...........................tgccgtatatgaagttga
       S  V                                                                   P  Y  M  K  L  T

 991 cgcaccatctcattgacgaagccttcttcgccgcgatgaagcccggctcgcgcattgtcaatactgcgcgtggccccgtcatctcgcagg
     cgcaccatctcattgacgaagccttcttcgccgcgatgaagcccggctcgcgcattgtcaatactgcgcgtggccccgtcatctcgcagg
       H  H  L  I  D  E  A  F  F  A  A  M  K  P  G  S  R  I  V  N  T  A  R  G  P  V  I  S  Q  D

1081 acgcacttatcgccgcgctcaagtcgggcaagctgctcagtgcaggcctcgacgtgcacgagttcgagccacaggtgtccaaggaactca
     acgcacttatcgccgcgctcaagtcgggcaagctgctcagtgcaggcctcgacgtgcacgagttcgagccacaggtgtccaaggaactca
       A  L  I  A  A  L  K  S  G  K  L  L  S  A  G  L  D  V  H  E  F  E  P  Q  V  S  K  E  L  I

1171 tcgagatggtgcgttcgagccctgccactgctctgcccgccttccatactttcagatcgtgctgacctatccatgtgtctgcagaagcac
     tcgagatg.................................INTRON 3...................................aagcac
       E  M                                                                                K  H

1261 gtcacgctcacgacgcacatcggcggcgtggcgatcgagaccttccacgagttcgagcggctcaccatgaccaacatcgaccgcttcctc
     gtcacgctcacgacgcacatcggcggcgtggcgatcgagaccttccacgagttcgagcggctcaccatgaccaacatcgaccgcttcctc
     V  T  L  T  T  H  I  G  G  V  A  I  E  T  F  H  E  F  E  R  L  T  M  T  N  I  D  R  F  L

1351 ctgcaaggcaagcccttgctgaccgtgcgtctttgcccgaactctctgcgccccgcttttcgcgcctactgacgctctggcgctgctgca
     ctgcaaggcaagcccttgctgaccgtgcgtctttgcccgaactctctgcgccccgcttttcgcgcctactga
     L  Q  G  K  P  L  L  T  V  R  L  C  P  N  S  L  R  P  A  F  R  A  Y  *

1441 gcctgcgggcaaggtgtttgcgccgtccagcgctgaatgacctgctcggacagattcagatcgatactgcaggaaggagtgcacgaggag

1531 ggcagcaccctgacaagagcgctccgtctgcaatgaaccttacaggatgtcgcaatgaaactcagggagacagcggaacagcggaagttg

1621 ctgctgaaat
```

Fig. 3: The complete sequence of the D-mandelate dehydrogenase-coding region. The sequence of the genomic DNA (top line) is aligned with cDNA sequence (second line), clearly showing the positions of the three introns. The predicted amino acid sequence is shown below the cDNA sequence. These sequences have been submitted to the EMBL database with the reference number of AJ001428.

Introns in DMDH gene from R.graminis

Introns	Position	5′	Branchpoint	3′
1	282- 367	GT.........	CTGAC	.........CAG
2	907- 972	GT.........	CTCAC	.........CAG
3	1179-1254	GT.........	CTCAC	.........CAG
Introns in S.Cerevisiae				
		GT.........	TACTAAC	.......CAG
Introns in Neurospore crassa				
		GTAAGT.........	ACTAACA	.......CAG
		GTACGT.........	GCTGACT	........CAG

Fig. 4: Comparison of *R. graminis* introns with other introns. The sequences of the 5′ and 3′ ends and the putative branch points of the introns in the D-MDH gene are indicated along with their position in the sequence numbering is as shown in Fig. 3. The corresponding sequences for other genes from *S. cerevisiae* (Ruby and Abelson, 1991) and *Neurosporra crassa* (Orbach *et al.*, 1986) are shown for comparison.

stranded pR13 and pR14 templates as appropriate. The complete sequences of the genomic DNA and cDNA are aligned in Fig. 3 along with the predicted amino acid sequence of D-mandelate dehydrogenase.

DNA sequence features: The G+C content of the cloned genomic DNA is high at 63%. Although this gave rise to some problems in reading the sequence with standard Sequenase reaction conditions, the use of dITP in place of dGTP in the sequencing reactions allowed the DNA sequence to be read unambiguously. Comparison of the genomic DNA and cDNA sequences identifies the positions of three introns within the D-mandelate dehydrogenase-coding region, varying in length from 66 to 86 nucleotides (Fig. 3). Each of these begins with the dinucleotide GT and ends with the trinucleotide CAG. These features are conserved in few other protein-coding genes that have been sequenced from *Rhodotorula* and *Rhodosporidium* species (Filpula *et al.*, 1988, Rasmussen and Orum, 1991). The introns within these genes also contain a conserved sequence, CTGAC that presumably defines the RNA splicing branch point and bears some similarity to the conserved TACTAAC branch point sequence found in *Saccharomyces* introns (Ruby and Abelson, 1991). The first and third introns in the D-mandelate dehydrogenase gene also contain this exact sequence, whereas no perfect match is found in the second introns (Fig. 4). However, in later case, the sequence CTCAC is found, which has a single mismatch compared with the consensus. We have identified similar divergence from the consensus sequence in several introns within the L-mandelate dehydrogenase coding sequence from *R. graminis* (Illias *et al.*, 1998).

Amino acid sequence: The cDNA encoding D-mandelate dehydrogenase predicts a protein of 351 amino acids with a calculated molecular weight of 38591 Daltons. This compares with the molecular weight of 38 KDa that was estimated by SDS-PAGE (Fewson and Baker, 1989). A computer search of protein sequence data banks with the program FASTA, using the D-mandelate dehydrogenase as the query sequence, indicated extensive amino acid sequence similarity with a range of D-2-hydroxy acid dehydrogenases. Alignment with other proteins in the Swissprot database demonstrated that *Rhodotorula graminis* D-mandelate dehydrogenase exhibits 27-33% identity to each of: the D-3-phosphoglycerate dehydrogenase from *Haemophilus influenzae* (SERA_HAEIN), D-glycerate dehydrogenase from *Hyphomicrobium methylovorum* (DHGY_HYPME), D-lactate dehydrogenase from *Lactobacillus delbrueckii* (LDHD_LACDE), formate dehydrogenase from *Hansenula polymorpha* (FDH_HANPO), D-3-phosphoglycerate dehydrogenase from *E. coli* (SERA_ECOLI), formate dehydrogenase from *Emericella nidulans* (FDH_NEUCR), formate dehydrogenase from *Neurospora crassa* (FDH_NEUCR), D-lactate dehydrogenase from *Lactobacillus casei* (DHD2_LACCA) and D-3-phosphoglycerate dehydrogenase from *S. cerevisiae* (SERX_YEAST). All of these enzymes utilize D-2-hydroxy acids as substrate with the exception of formate, which has a single carbon atom. D-mandelate dehydrogenase from *Rhodotorula graminis* thus clearly belongs to the D-isomer specific 2-hydroxy acid-dehydrogenase family. The two closest known relatives to D-mandelate dehydrogenase appear to be from other yeast but are sequences identified from genome sequencing projects. These are the putative products of a gene on the *Saccharomyces cerevisiae* chromosome XIV (EMBL accession number: Z71559; *Saccharomyces* Genome Database reference YNL274C) and a gene from *Schizosaccharomyces pombe* (accession number: D89185). The predicted amino acid sequences show 33% identity with D-mandelate dehydrogenase over 340 amino acids but are more closely related to each other with 46% identity. The substrates for the products of these two genes are currently unknown. Multiple alignments of the sequences of D-mandelate dehydrogenase with other members of the family clearly show their relations with several amino acid residues being strictly conserved in each protein (Fig. 5). The crystal structures of three members of D-2-hydroxy acid dehydrogenase family have been determined. D-lactate dehydrogenase from *Lactobacillus pentosus*, formate dehydrogenase from *Pseudomonas* sp. 101 (Lamzin *et al.*, 1992, Lamzin *et al.*, 1994) and D-glycerate dehydrogenase from *Hyphomicrobium methylovorum* (Golberg *et al.*, 1994) each have a two-domain structure that is typical of NAD^+-dependent enzymes. Sequence alignment indicates that the

```
dmdh ~~~MPRPRVL LLGDPARHLD DLWSDFQQKF EV..IPANLT THDGFKQALR   45
Scer MTLSGKPAAL LVGTLK.HAH KEWEALGKYA ELKTYSDG.T REDFLAKCK.
Spom ~~MSKKPIVL KLGKDA.FGD QAWGELEKIA DVITIPESTT REQFLREVKD
gldh ~~~~~~~~~~ ~~~~~~~~~S KKKILITWPL PEAAMARARE SYDVIAHGDD
fdh  DHYPGGQTLP TPKAIDFTPG QLLGSVSGEL GLRKYLESNG HTLVVTSDKD

                                                       *
dmdh EKRYGDFEAI IKLAVEN..G TESYPWNADL ISHLPSSLKV FAAAGAGFDW   93
Scer ..TEFQNVKA ICRTYNS..K FYMGIWDKEI IDNLPPSVKF ICHLGAGYET
Spom PQNKLSQVQV ITRTARS..V KNTGRFDEEL ALALPSSVVA VCHTGAGYDQ
gldh PKITIDEMIE TAKSVDALLI TLNEKCRKEV IDRIPENIKC ISTYSIGFDH
fdh  GPDSVFEREL VDADVVISQP FWPAYLTPER IAK.AKNLKL ALTAGIGSDH

      *      *                          *   *
dmdh LDLDALNERG VAFANSRGAG DTATSDLALY LILSVFRLAS YSERAARTGD  143
Scer VDVAACTARG IQVSHVPKAV DDATADVGIF LMLGALR..G FNQGIFELHK
Spom IDVEPFKKRH IQVANVPDLV SNATADTHVF LLLGALR..N FGIGNRRLIE
gldh IDLDACKARG IKVGNAPHGV TVATAEIAML LLLGSARRAG EGEKMIRTRS
fdh  VDLQSAIDRN VTVAEVTYCN SISVAEHVVM MILSLVRNYL PSHEWARKGG

                              *    *  *                  *
dmdh PETFNRVHLE IGKSAHNPRG HVLGAVGLGA IQKEIARKAV HGLGMKLVYY  193
Scer NNW.....NA NCKPSHDPEG KTLGILGLGG IGKTMAKRAR AFDMK.IVYH
Spom GNWPEAGPAC GSPFGYDPEG KTVGILGLGR IGRCILERLK PFGFENFIYH
gldh WPGWEPLELV GEKL....DN KTLGIYGFGS IGQALAKRAQ GFD.MDIDYF
fdh  WNIADCV... ..SHAYDLEA MHVGTVAAGR IGLAVLRRLA PFD.VHLHYT

                                           *     *
dmdh DVAPADAETE KALGAERVDS LEELARRSDC VSVSVPYMKL THHLIDEAFF  243
Scer NRTPLPEEEA E..GAEFV.S FDDLLAKSDV LSLNLPLNAH TRHIIGKPEF
Spom NRHQLPSEEE H..GCEYV.G FEEFLKRSDI VSVNVPLNHN THHLINAETI
gldh DTHRASSSDE ASYQATFHDS LDSLLSVSQF FSLNAPSTPE TRYFFNKATI
fdh  DRHRLPESVE KELNLTWHAT REDMYPVCDV VTLNCPLHPE TEHMINDETL

              * *****           **  *       ***     *
dmdh AAMKPGSRIV NTARGPVISQ DALIAALKSG KLLSAGLDVH EFEPQVSKE.  292
Scer QKMKRGIVIV NTARGAVMDE AALVEALDEG IVYSAGLDVF EEEPKIHPG.
Spom EKMKDGVVIV NTARGAVIDE QAMTDALRSG KIRSAGLDVF EYEPKISKE.
gldh KSLPQGAIVV NTARGDLVDN ELVVAALEAG RLAYAGFDVF AGEPNINEG.
fdh  KLFKRGAYIV NTARGKLCDR DAVARALESG RLAGYAGDVW FPQPAPKDHP

                 *
dmdh LIEMKHVTLT THIGGVAIET FHEFERLTMT NIDRFLLQGK PLLTVRLCPN  342
Scer LLENEKVILL PHLGTNSLET QYKMECAVLM NVKNGIVNDS LPNLVPEQRG
Spom LLSMSQVLGL PHMGTHSVET RKKMEELVVE NAKNVILTGK VLTIVPELQN
gldh YYDLPNTFLF PHIGSAATQA REDMAHQAND LIDALFGGAD MSYALA*...
fdh  WRTMPYNGMT PHISGTTLTA QARYAAGTRE ILECFFEGRP IRDEYLIVQG

dmdh SLRPAFRAY* ~~~~~~~
Scer DIE*~~~~~~ ~~~~~~~
Spom EDWPNESKPL V*~~~~~
gldh ~~~~~~~~~~ ~~~~~~~
fdh  GALAGTGAHS YSKGNAT
```

Fig. 5: Multiple alignments of the sequenced D-mandelate dehydrogenase and other D-2-hydroxy acid dehydrogenases. The sequence of D-mandelate dehydrogenase from *R. graminis* (dmdh), D-glycerate dehydrogenase from *Hyphomicrobium methylovorum* (gldh; Swissprot: DHGY_HYPME), formate dehydrogenase from *Pseudomonas* sp.101 (fdh; Swissprot: FDH_PSESR) and the predicted products (Z71550 and D89185) of reading frames identified in the genomes of *Saccharomyces cerevisiae* (Scer.) and *Schizosaccharomyces pombe* (Spom.) were aligned using the PILEUP algoritm within the Wisconsin package. The asterisks above the aligned sequences denote residues that are identical in all five sequences. The numbering to the right indicates the dmdh residue number at the end of each line.

coenzyme-binding domain of D-mandelate dehydrogenase comprises residues 111-307. The 'catalytic' domain is formed by both N- and C-terminal portions of the polypeptide chain (residues 1-110 and 308-351). Within the family of D-2-hydroxy acid dehydrogenases, the NAD^+-binding domain is highly conserved than the catalytic domain. D-glycerate dehydrogenase was crystallized as the apo-protein whereas the structure of formate dehydrogenase was solved with both NAD^+ and azide present. Azide is an inhibitor that presumably binds in place of formate (Lamzin *et al.*, 1994). We can thus predict, by comparison with known structures and sequences that several residues are involved in NAD^+ binding (Gly170, Gly172, Gln175, Asp194, Pro229) which is achieved by a remarkably similar topological arrangements not only in D-2-hydroxy acid dehydrogenase but also in other NAD^+-dependent enzymes, despite a very low level of sequence similarity (Popov and Lamzin, 1994). In contrast, the catalytic subunits of the D-2-hydroxy acid dehydrogenase are quite different in both sequence and structure from other dehydrogenase. Several residues have been identified as important for catalysis in the D-2-hydroxy acid dehydrogenase by examination of crystal structures (Lamzin *et al.*, 1992, Goldberg *et al.*, 1994), chemical modification (Kochhar *et al.*, 1992b) and analysis of mutant enzymes (Kochhar *et al.*, 1992a). The catalytically important residues (His304, Glu286 and Arg257) are conserved in D-mandelate dehydrogenase.

Expression of D-mandelate dehydrogenase in *E. coli*: The D-mandelate dehydrogenase cDNA in pR13 was transferred into the expression vector pRC23 (Crowl *et al.*, 1985) as an *Eco*RI-*Bam*HI fragment to generate pR16. The plasmid directs expression under control of the P_L promoter from bacteriophage λ but expression is repressed at 30°C in *E. coli* NF1, which synthesizes the thermosensitive cI_{857} protein (Stanley and Luzio, 1984). After growth to mid-exponential phase ($O.D_{600}$ about 0.6), expression was induced by shifting to 42°C and continuing growth overnight. Expression of D-mandelate dehydrogenase was detected by Western blotting (data not shown). The protein was active with readily measurable activity for the reverse reaction; phenylglyoxylate-dependent oxidation of NADH. Specific activity of the recombinant D-mandelate dehydrogenase in crude extract is 0.0528 U/mg of total protein present. This activity was produced from a yield of 1370 mg total protein obtained from 17-gram wet cells. The development of this expression system opens the way to large-scale production of recombinant enzyme for more detailed biochemical and biophysical studies in addition to its potential use as a biocatalyst.

References

Baker, D.P. and C.A. Fewson, 1989. Purification and characterization of D (-)-mandelate dehydrogenase from *Rhodotorula graminis*. J. Gen. Microbiol., 135: 2305-2044.

Basak, A.K., D.P. Baker, C.A. Fewson and D.I. Stuart, 1993. Preliminary crystallographic study of D(-)-mandelate dehydrogenase from *Rhodotorula graminis*. J. Mol. Biol., 233: 781-783.

Crowl, R., C. Seamans, P. Lomedico and S. McAndrew, 1985. Versatile expression vectors for high-level synthesis of cloned gene products in *Escherichia coli*. Gene., 38: 31-38.

Durham, D.R., 1984. Initial reactions involved in the dissimilation of mandelate by *Rhodotorula graminis*. J. Bacteriol., 160: 778-780.

Feinberg, A.P. and B. Vogelstein, 1983. A technique for radio-labeling DNA restriction fragments to high specific activity. Ann. Biochem., 132: 6-13.

Fewson, C.A., 1988. Microbial metabolism of mandelate: a microcosm of diversity. FEMS Microbiol. Rev., 54: 85-110.

Fewson, C.A., 1992. Microbial metabolism and mandelate: occurrence, properties and evolution of mandelate dehydrogenase and other enzymes of the mandelate pathway. In The Evolution of Metabolic Function. R. P. Mortlock, (Ed.). pp: 115-141. Boca Raton: CRC Press.

Filpula, D., C.A. Vaslet, A. Levy, A. Sykes and R.L. Strausberg, 1988. Nucleotide sequence of gene for phenylalanine ammonia-lyase *from Rhodotorula rubra*. Nuc. Acid Res., 16: 11381.

Goldberg, J.D., T. Yoshida and P. Brick, 1994. Crystal structure of a NAD-dependent D-glycerate dehydrogenase at 2.4 Å resolutions. J. Mol. Biol., 236: 1123-1140.

Haid, A. and M. Suissa, 1984. Immunochemical detection of membrane proteins after sodium dodecyl sulfate polyacrylamide gel electrophoresis. Meth. Enzymol., 96: 192-205.

Hosono, K., S. Kajiwara, Y. Yamazaki and H. Maeda, 1990. Construction of a bioreactor for production of (R)-(-)-mandelate, a typical specialty chemical. J. Biotechnol., 14: 149-156.

Illias, R.M., R. Sinclair, D. Robertson, A. Neu, S.K. Chapman and G.A. Reid, 1998. L-mandelate dehydrogenase from *Rhodotorula graminis*: cloning, sequence and kinetic characterization of the recombinant enzyme and its independently expressed flavin domain. Biochem. J., 333: 107-115.

Kochhar, S., N. Chuard and H. Hottinger, 1992a. Glutamate 264 modulates the pH dependence of NAD^+-dependent D-lactate dehydrogenase. J. Biol. Chem., 267: 20298-20301.

Kochhar, S., P.E. Hunziker, P. Leong-Morgenthaler and H. Hottinger, 1992b. Primary structure, physicochemical properties and chemical modification of NAD^+-dependent D-lactate dehydrogenase. J. Biol. Chem., 267: 8499-8513.

Lamzin, V.S., A.E. Aleshin, B.V. Strokopytov, M.G. Yukhnevich, V.O. Popov, E.H. Harutyunyan and K.S. Wilson, 1992. Crystal structure of NAD-dependent formate dehydrogenase. Eur. J. Biochem., 206: 441-452.

Lamzin, V.S., Z. Dauter, V.O. Popov, E.H. Harutyunyan and K.S. Wilson, 1994. High resolution structures of holo- and apo-formate dehydrogenase. J. Mol. Biol., 236: 759-785.

Orbach, M.J., E.B. Porro and C. Yanofsky, 1986. Cloning and characterization of the gene for β-tubulin from a benomyl-resistant mutant of *Neurospora crassa* and its uses as dominant selectable marker. Mol. Cell. Biol., 6: 2452-2461.

Popov, V.O. and V.S. Lamzin, 1994. NAD^+-dependent formate dehydrogenase. Biochem. J., 301: 652-643.

Rasmussen, O.F. and H. Orum, 1991. Analysis of the gene for phenylalanine ammonia-lyase from *Rhodosporidium toruloides*. DNA Seq., 1: 207-211.

Rokeach, L.A., J.A. Haselby and S.O. Hoch, 1988. Molecular cloning of a cDNA encoding the human SM-D autoantigen. Proc. Nat. Acad. Sci. USA., 85: 4832-4836.

Ruby, S.W. and J. Abelson, 1991. Pre-mRNA splicing in yeast. Trend. Gen., 7: 79-85.

Sambrook, J., E.F. Fritsch and T. Maniatis, 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., S. Nicklen and A.R. Coulson, 1977. DNA sequencing with chain terminating inhibitors. Proc. Nat. Acad. Sci. USA., 74: 5463-5467.

Smekal, O., M. Yasin, C.A. Fewson, G.A Reid and S.K. Chapman, 1993. L-mandelate dehydrogenase from *Rhodotorula graminis*: comparison with L-lactate dehydrogenase (Flavocytochrome b_2) from *Saccharomyces cerevisiae*. Biochem. J., 290: 103-107.

Stanley, K.K. and J.P. Luzio, 1984. Construction of a new family of high efficiency bacterial vectors: identification of cDNA clones coding for human liver proteins. EMBO. J., 3: 1429-1434.

Vasik-Racki, D., M. Jonas, C. Wandrey, W. Hummel and M.R. Kula, 1989. Continuous (R)-mandelic acid production in an enzyme membrane reactor. Appl. Microbiol. Biotechnol., 31: 215-222.

Viera, J. and J. Messing, 1987. Production of single-stranded plasmid DNA. Meth. Enzymol., 153: 3-11.

Yamazaki, Y. and H. Maeda, 1986. Enzymatic synthesis of optically pure (R)-(-)-mandelic acid and other 2-hydroxy carboxylic acids- screening for the enzyme and its purification, characterization and use. Agric. Biol. Chem., 50: 2621-2631.

Yasin, M. and C.A. Fewson, 1993. L-(+)-mandelate dehydrogenase from *Rhodotorula graminis*: purification, partial characterization and identification as a flavocytochrome *b*. Biochem. J., 293: 455-460.

Yoshida, T., K. Yamaguchi, T. Hagishita, T. Mitsunaga, A. Miyata, T. Tanabe, H. Toh, T. Oshiro, M. Shimao and Y. Izumi, 1994. Cloning and expression of the gene for hydroxypyruvate reductase (D-glycerate dehydrogenase) from an obligate methylotroph *Hyphomicrobium methylovorum* GM2. Eur. J. Biochem., 223: 727-732.