The Nucleotide Sequence of the *LPDI* Gene Encoding Lipoamide Dehydrogenase in *Saccharomyces cerevisiae*: Comparison between Eukaryotic and Prokaryotic Sequences for Related Enzymes and Identification of Potential Upstream Control Sites

By JOE ROSS,* GRAEME A. REID AND IAN W. DAWES

Department of Microbiology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JG, UK

(Received 30 November 1987; revised 29 January 1988)

The complete nucleotide sequence of the *LPDI* gene, which encodes the lipoamide dehydrogenase component (E3) of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase multienzyme complexes of *Saccharomyces cerevisiae*, has been established. The flanking region 5' to the *LPDI* gene contains DNA sequences which show homology to known control sites found upstream of other yeast genes. The primary structure of the protein, determined from the DNA sequence, shows strong homology to a group of flavoproteins including *Escherichia coli* lipoamide dehydrogenase and pig heart lipoamide dehydrogenase. The amino acid sequence also reveals the presence of a potential targeting sequence at its N-terminus which may facilitate transport to and entry into mitochondria.

**INTRODUCTION**

Lipoamide dehydrogenase (EC 1.8.1.4) is a component of the multienzyme complexes pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase, which catalyse the oxidative decarboxylation of pyruvate and 2-oxoglutarate to acetyl-CoA and succinyl-CoA respectively (Reed, 1974). Both complexes contain multiple copies of three component enzymes: pyruvate dehydrogenase (EC 1.2.4.1) or 2-oxoglutarate dehydrogenase (EC 1.2.4.2) (E1); dihydrolipoamide acetyltransferase (EC 2.3.1.12) or dihydrolipoamide succinyltransferase (EC 2.3.1.61) (E2) and lipoamide dehydrogenase (E3). The E1 and E2 components are specific to their respective complexes whereas lipoamide dehydrogenase has been shown to be functionally interchangeable between both (Mukherjee *et al.*, 1965), and in *Escherichia coli* and *Saccharomyces cerevisiae* is encoded by a single gene (Guest & Creaghan, 1973; Dickinson *et al.*, 1986).

In higher eukaryotes lipoamide dehydrogenase has an additional role in the multienzyme complexes which specifically catalyse the oxidative decarboxylation of branched-chain 2-oxoacids derived from leucine, valine and isoleucine by transamination (Lawson *et al.*, 1983). The reversible oxidative decarboxylation of glycine has also been shown to involve lipoamide dehydrogenase in the aerobic bacterium *Arthrobacter globiformis* (Kochi & Kikuchi, 1976), in the anaerobe *Peptococcus glycinophilus* (Robinson *et al.*, 1973) and in rat liver mitochondria (Kochi & Kikuchi, 1976).

Regulation of the synthesis of the E1, E2 and E3 components of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase in *E. coli* has been studied extensively (Guest & Rice, 1984; Spencer & Guest, 1985), and it has been shown that the *E* and *F* genes of the *ace* operon and the *A* and *B* genes of the *suc* operon encode the E1 and E2 subunits of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase complexes respectively. Lipoamide dehydrogenase is encoded by a single gene, *lpd*, which is linked to the *ace* operon but can operate under the control of its own promoter. Transcription from *sucAB*, *aceEF* and *lpd* genes shows differential
regulation in response to the nature of the carbon source, thereby ensuring the necessary ratio of subunit components when pyruvate and/or 2-oxoglutarate dehydrogenase enzyme complexes are required. The absence of operons in eukaryotic organisms dictates that the regulation of synthesis of the E1, E2 and E3 components in yeast must differ from the system found in *E. coli*.

The role of lipoamide dehydrogenase in at least two multienzyme complexes raises several interesting questions concerning the regulation of synthesis and control of distribution of an enzyme which functions within two different complexes. The enzyme is encoded in the nucleus but must be transported to its site of action in the mitochondrion. Biochemical studies of the initiation of sporulation in *S. cerevisiae* have shown that the regulation of 2-oxoglutarate dehydrogenase activity may be of crucial importance in the switch of cells from vegetative growth via mitosis to the developmental processes of meiosis and sporulation (Dickinson *et al*., 1986). In *S. cerevisiae* the enzyme has also been shown to be subject to catabolite repression (Roy & Dawes, 1987).

Determination of the DNA sequence for the genes which encode the E1, E2 and E3 components of both pyruvate and 2-oxoglutarate dehydrogenase is necessary in order to study how their syntheses are regulated at the molecular level.

**METHODS**

*Sources of DNA.* The 3-7 kb *XhoI* fragment which contains the *LPD1* gene was obtained from pGP-R1 (Roy & Dawes, 1987) for 'shotgun' cloning into M13 vectors. Large-scale plasmid and phage RF2 DNA isolation was performed by alkaline-SDS lysis; the DNA was purified by centrifugation in CsCl/ethidium bromide gradients (Maniatis *et al*., 1982). Restriction fragments were separated by electrophoresis in agarose gels and extracted by electro-elution onto dialysis membrane (Smith, 1980).

*Cloning in the M13-based vector.* The 3-7 kb *XhoI* fragment was digested with three restriction enzymes for shotgun cloning into M13mp18 (Messing, 1983). *TaqI* and *MspI* fragments were cloned into the *AccI* site of M13mp18, and *Sau3AI* fragments were cloned into the *BamHI* site of the same vector. In addition, the 1-45 kb *XbaI* fragment was cloned into the *XbaI* site of M13mp18. Transfection of *E. coli* JM101 [Ap lac-pro supE thi/F' traD36 proAB lacP ZAM51] was performed according to published procedures (Winter & Fields, 1980).

*Nucleotide sequence analysis.* Single-stranded M13 DNA templates were prepared and sequenced by the dideoxy chain-termination method using a 17-nucleotide synthetic primer (Sanger *et al*., 1977). Two 15-nucleotide synthetic primers, based on information obtained from the *LPD1* sequence, were also used. The nucleotide sequence was compiled and analysed using the University of Wisconsin Genetics Computer Group programs (Devereux *et al*., 1984).

*Materials.* Restriction enzymes were purchased from Bethesda Research Laboratories, Boehringer, and Northumbria Biologicals. Phage T4 DNA ligase was from either Boehringer or Northumbria Biologicals. The 17-nucleotide synthetic primer was obtained from Pharmacia; the 15-nucleotide synthetic primers were purchased from OSWEL DNA Synthesis Service, Chemistry Department, Edinburgh University. DNA polymerase (Klenow fragment), dideoxynucleotides and deoxynucleotides were obtained from Boehringer, [α-32P]dATP from Amersham, and reverse transcriptase from Northumbria Biologicals.

**RESULTS AND DISCUSSION**

*The nucleotide sequence*

Based on deletion analysis of the plasmid pGP1 (Roy & Dawes, 1987), a YEpl3-based vector containing a 5-5 kb yeast DNA insert carrying the *LPD1* gene, a 3-7 kb *XhoI* fragment was chosen for 'shotgun' cloning analysis. A region of 2701 bp has been sequenced. All of the sequence was obtained from at least two independent clones, it was fully overlapping and was, with the exception of a 0-1 kb region 0-8 kb upstream of the coding region, derived from both DNA strands. The program MAP was used to identify the coding region representing the *LPD1* gene. One large open reading frame of 1-5 kb was found, which is consistent with the known size of the yeast lipoamide dehydrogenase subunit polypeptide (Wieland, 1983) and which exhibits a consistently high score with respect to preferred codon usage in *S. cerevisiae* determined using the programs CODONFREQUENCY and CODONPREFERENCE (data not shown). The complete and unambiguous sequence of the *LPD1* gene and the primary structure of lipoamide dehydrogenase, translated from the DNA sequence, are presented in Fig. 1.
The yeast lipoamide dehydrogenase gene

In the region 5'-distal to the LPD1 sequence are two potential open reading frames: one (at \(-571\)) which would be read in the opposite sense, is capable of encoding a polypeptide of 77 residues; the other (sequence not complete, terminating at \(-858\)) could encode a polypeptide of at least 76 amino acids. A search of the NBRF protein data base showed no strong homology between the two predicted amino acid sequences and any known protein.

Primary structure of yeast lipoamide dehydrogenase

The primary structure, translated from the nucleotide sequence of the LPD1 gene, contains 499 amino acid residues that correspond to a protein of \(M\), 54010 (54730 including the FAD cofactor). These \(M\), values are in good agreement with previous estimates (Wieland, 1983).

Lipoamide dehydrogenase, encoded in the nucleus and synthesized on cytoplasmic ribosomes, is a component of two multienzyme complexes which function in the mitochondrial matrix–inner membrane compartment. The presence of an N-terminal targeting sequence, which directs the protein to its correct subcellular location and facilitates its entry into the mitochondrion, has been shown in mammalian cells. Antibodies raised against the E1, E2 and E3 components of mitochondrial 2-oxoglutarate dehydrogenase from ox heart have been used to show the presence, in cultured pig kidney cells, of initial cytoplasmic translation products that are larger than the mature proteins (Hunter & Lindsay, 1986). The E3 component is synthesized as a polypeptide 10-20 amino acids larger than the mature protein. The primary structure, derived from the LPD1 nucleotide sequence of S. cerevisiae, supports the conclusion that, in this organism, the lipoamide dehydrogenase N-terminal region represents a mitochondrial targeting sequence. The first 20 amino acids are rich in seryl, threonyl and basic residues, and show a complete absence of acidic residues. All of these features are typical of mitochondrial targeting sequences in yeast (von Heijne, 1986). In E. coli such a targeting sequence would be unnecessary; and when the primary structures of E. coli and S. cerevisiae lipoamide dehydrogenase (Stephens et al., 1983) are aligned (see Fig. 2) the yeast sequence is seen to contain an additional 20 amino acids at its N-terminus which are absent from the inferred E. coli protein.

The primary structures of several flavoproteins have been determined either partially or completely. Pig heart lipoamide dehydrogenase, E. coli lipoamide dehydrogenase, human erythrocyte glutathione reductase and Pseudomonas aeruginosa transposon mercuric reductase amino acid sequences have previously been compared (Williams et al., 1984) and four regions were identified which show strong homology between the four proteins. Two of these (FAD-1 and FAD-2) are involved in the binding of FAD, and the others constitute a pyridine nucleotide binding domain and an 'interface' domain involved in the interaction between subunits of the enzyme. Fig. 2 shows the primary structure of these four proteins and of S. cerevisiae lipoamide dehydrogenase aligned for maximum homology using the computer programs GAP and PRETTY. The primary structure of S. cerevisiae lipoamide dehydrogenase is in full agreement with the established 'domain homology' found between the other proteins. Comparison of the three lipoamide dehydrogenase primary structures reveals that the strongest homology lies between those of pig heart and yeast. The higher level of homology is particularly noticeable in the dimer interface region concerned with the interaction of the two polypeptides in formation of the active dimeric form (see Fig. 2). This may indicate that there are differences between prokaryotes and eukaryotes in the recognition of the subunits during assembly, which is reflected in the different organization of the multienzyme complexes in terms of subunit composition, especially for pyruvate dehydrogenase (Reed, 1974). Despite these differences the overall degree of homology that exists between all three polypeptides indicates the heavy evolutionary constraints placed upon an enzyme which functions within two different multienzyme complexes. This is also reflected in the low immunogenicity of the yeast E3 subunit relative to the other enzymes of the pyruvate and 2-oxoglutarate dehydrogenase complexes (De Marcucci et al., 1985).

General features of the noncoding regions

The sequences upstream and downstream which flank the coding region of the LPD1 gene share several characteristics common to the 5'- and 3'-noncoding regions of many sequenced yeast genes.
Met Leu Arg Ile Arg Ser Leu Leu Arg Ala Phe Ser Ser Thr Val Arg Thr Leu Thr Ile Asn Lys 
ATG TTA AGA ATC AGA TCA GTC GTA AAT AAT AAG GCT GGC TTT TCG TCC ACA GTC AGC AAG ACC ATT ACC 
0 74
Ser His Asp Val Val Ile Gly Gly Gly Pro Ala Gly Tyr Val Ala Ile Lys Ala Gin Leu Gly Phe 
TCA CAT GAT GTC ATC GTC GGT GGC GCT GGT GTC GGT GTC GAT TCG GAA CAA TGG GGA
75 149
Asn Thr Ala Cys Val Gly Lys Gly Leu Gly Thr Cys Leu Leu Val Gly Cys Ile Pro Ser Lys Ala 
AAT GCA TTT TGA AAA AAG AGA GGC AAA TTG GCA GGC GTC ATT GTC AAT GGG TCG TCC AAA GCA 
150 224
Leu Leu Asn Ser His Leu Phe His Met His The Glu Ala Gin Arg Gly Ile Asp Val Alan Gly Asp 
CTG GTA AAT ATT TCT TAA TAT GAA ATG CAA GAC GAA AGG AGA AAT ATT CAC GGC GGT GAA
225 299
Ile Lys Ile Val Ala Asn Phe Gin Lys Ala Lys Asp Asp Ala Val Lys Gin Leu Thr Ile Gly Ile Gnu Leu 
ATC AAA ATT AAG GTA GCA AAG TCG GAT GGG GTC GGT GAA CTT CAT ATC TTA AGC TAT GGC GAT 
300 374
Leu Phe Lys Lys Asn Val Thr Thr Tyr Lys Gly Gly Ser Phe Gnu Asp Glu Thr Lys Arg Arg Val Thr 
CTG TTC AAA AAT AAT AGG GCC GTC ATG TAT AAA GGT ATT GAT TCA TGC GAA GAC GAG AAG CAA AGG ACT 
375 449
Pro Val Asp Gly Leu Gly Thr Val Lys Gly Asp His Ile Leu Asp Val Lys Asn Ile Val Ala Thr Gly 
CCT GGT GGA GAA AGA GGG ACT GTC AAG GAA GAC GAC GGA ATG ATA GAT GGA ATG AAG ACT ACG AGC GGC 
450 524
Ser Gly Thr Pro Phe Pro Gly Ile Gly Ile Gly Lys Val Val Ser Ser Gly Thr Ala Leu Ser Leu 
TCA GGG ATC GCC TTT GGG GCT GGT GAA AAG ATT GCT TCA ACA GAG GAG GAT GGC GGT GAA
525 599
Lys Gly Ile Pro Arg Leu Arg Gly Ile Gly Ile Gly Leu Met Gly Ser Val Ser Arg 
AAG GAA ATT CCC AAA AGA TTA ACC ATC GGT GGA GCA AGA ATT GCA GAT GGC GCA GGT GCC
600 674
Leu Gly Ser Lys Val Thr Val Val Gly Phe Gin Pro Gin Ile Gly Ala Ser Met Asp Gly Val Gly Val Ala 
AGG GCT TAC GTG GTA GAA TTT CAA CCT GAT ATT GCT GGT GGA GCT GAC GAG GCC GAC 
675 749
Thr Gin Lys Phe Lys Lys Gly Gly Gly Lys Asp Phe Lys Leu Ser Thr Lys Val Ile Ser Ala Lys Arg Asp 
ACC CAA AAG TIC TGG AAA AAG GAA GCT GAG TIC AAA TTA AGC AAA AAG TCT TCT GCA AGA AAG GAC 
750 824
Asp Lys Asp Val Val Glu Ile Val Glu Asp Thr Lys Thr Asn Lys Gin Asn Leu Glu Ala Glu Val Leu 
GAC AAG ACG GCT GTC ATT GTA GAA GAT AAC AAG AAA AAT AAG AAG CAA AAT GGA GGA GAA GGA
825 899
Leu Val Ala Val Gly Arg Pro Arg Pro His Ile Gly Leu Gly Lys Ala Gly Lys Val Asp Lys Arg 
GTC GGT GGT GGT GGA AAG ATT GCA GTA AAC TTA AAA GAA GAT GTA GCA GAA GAG AAA AGG 
900 974
Gly Arg Leu Val Ile Asp Asp Gin Phe Ser Lys Phe Pro His Lys Val Val Gly Asp Val Thr Phe Gly 
GGA GCC GTA GTC AGT GAT GAC AAA TTT AAT TGC AAG TGC TCC GCA CAC ATT AAA GAT GTA GGA GAT GCA
975 1049
(continued on facing page)
The yeast lipoamide dehydrogenase gene

Fig. 1. Nucleotide sequence of the LPD1 gene and deduced primary structure of lipoamide dehydrogenase. The nucleotide sequence of a 2.7 kb region containing the sense strand of the LPD1 gene as well as over 1 kb of flanking DNA is shown in the 5' to 3' direction. The specific features of the 5' flanking region are noted as follows: The 'TATA' consensus sequences are overlined; the sequences which show homology to known regulatory sequences are boxed. Specific features of the 3' flanking sequence are noted as follows: the proposed consensus polyadenylation signal, AATAAA, is denoted by overdots; the proposed polyadenylation/termination signal TAG...TATGG...TTT is denoted by underdots.

The noncoding regions of several yeast genes are AT rich and the regions 5' (−500 to −1) and 3' to the LPD1 coding region have overall A + T compositions of 63% and 74% respectively. The sequences TAATAA and TATAA found at positions −146 and −154 represent potential TATA boxes believed to be important in positioning transcription initiation by RNA polymerase II (Grosschedl & Birnstiel, 1980). An A is found at position −3 as has been reported for the majority of yeast genes sequenced (Kozak, 1981).

Downstream of the TGA translation termination codon an additional 114 bp has been sequenced. The motif AATAAA found at position +12: CAGTATAGTATATATGTT has been postulated to be a signal for transcription termination or polyadenylation (Zaret & Sherman, 1982), and a version of this is found at position +18: CAGTATAGTATATATTT.

Regulation of the LPD1 gene

The upstream region from −358 to −102 contains a number of motifs which show homology to ones that have roles in the transcriptional regulation of other yeast genes. This region therefore probably contains elements which control the expression of the LPD1 gene. These motifs are illustrated in Fig. 3. At position −247 there is a TGACTC sequence with an adjacent run of T residues that conforms very closely to the consensus sequence obtained from the GCN4 protein binding sites studied to date (Hill et al., 1986). This protein mediates the general control of amino acid biosynthesis response that is involved in modulating during amino acid starvation the expression of some genes for amino acid synthesis (Struhl, 1982). It is interesting that a second, almost perfect direct repeat sequence (TGAATCGTTTTT), is also present 17 bp away at position −265. At −114 there is another TGACTC motif, although this conforms less well to the GCN4 consensus. All other yeast genes found so far to be subject to this GCN4-mediated control are concerned directly with synthesis of amino acids or charged tRNA species. Regulation of lipoamide dehydrogenase by general amino acid control may be less
The yeast lipoamide dehydrogenase gene

(a) GCN4 binding site consensus sequence

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPD1</td>
<td>-264</td>
<td>CGTGAATCGTTTTTAG</td>
</tr>
<tr>
<td>LPD1</td>
<td>-247</td>
<td>GATGACTCGTTTTTTAA</td>
</tr>
<tr>
<td>LPD1</td>
<td>-114</td>
<td>TTTGACTCACCTCAGGA</td>
</tr>
</tbody>
</table>

(b) CYC1 UAS1             CTCTTTGGCGGGGTTTTT
CYC1 UAS2             CTCTTTGGCGAGCGTTTT
GALI-10 control region sequence             CTCTTTGGAAACCTTTCAG

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPD1</td>
<td>-204</td>
<td>CTCATTGGCGAGAAGTC</td>
</tr>
</tbody>
</table>

(c) Heat-shock consensus sequence

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPD1</td>
<td>-348</td>
<td>CACGAATAGTCATG</td>
</tr>
</tbody>
</table>

Fig. 3. Alignment of potential regulatory sequences with known motifs. In each case the established regulatory sequence is displayed, with the upstream regions of the LPD1 gene which show homology shown directly below. (a) Consensus sequence for the general control regulatory site. Highly conserved nucleotides are shown as capitals, conserved residues are shown in lower case, and nonconserved residues are indicated by dashes. (b) Upstream activation sites (UAS1 and UAS2) of CYC1 and a sequence located in the GALI-10 control region. (c) Heat-shock promoter consensus sequence.

Surprising than it seems at first sight, since some of the tricarboxylic acid (TCA) cycle intermediates are substrates for the synthesis of such amino acids as glutamate, arginine, lysine and proline. There is one report that fumarase activity is derepressed in arginine bradytrophs; it has been suggested that this TCA cycle enzyme may also be under general amino acid control (Delforge et al., 1975).

A second putative control region, at position -204, together with its homology with the upstream activation sites of CYC1, is shown in Fig. 3. The function of this region is not known but both genes are involved in aerobic metabolism and are subject to catabolite repression (Guarente et al., 1984; Roy & Dawes, 1987); a related sequence with homology over 8 bases can be found in the GALI-10 control region (Johnston & Davis, 1984).

A third potential control site shown in Fig. 3 is the sequence at -352 which shows a 7 out of 8 match with the general heat-shock consensus sequence (Pelham, 1985).

Several further sequences are also of interest. One is an inverted repeat CTCCCGCGGAG at position -188 which is similar to a general repressor protein binding sequence proposed by...
T. Cooper (personal communication); a second is the sequence CACCTCGA at position -109 which is homologous to one found upstream of the ARG3 and CARI genes involved in arginine biosynthesis and degradation respectively (Crabeel et al., 1985). In addition to the above, the upstream region contains a number of direct repeats and inverted repeats which may also have a role in the regulation of LPD1 expression. One of these at -283 is very interesting since it is a 23 bp inverted repeat, containing direct repeats of 9 bp each within it. It is represented in a symmetrically truncated form at -361. We are currently investigating the functional significance of these upstream regions.

We are grateful to Lynn Kennedy for technical assistance, and to the SERC for a postgraduate studentship (J.R.) and research grant support.

REFERENCES


group in the electron transfer protein P² from Peptococcus glycophilus. *Journal of Biological Chemistry* 248, 5319–5323.


