The Nucleotide Sequence of the *LPD1* 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 *LPD1* 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 *LPD1* 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). TagI 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 XhoI fragment was cloned into the XbaI site of M13mp18. Transfection of E. coli JM101 [λlac-pro E supE thi/F' traD36 proAB lacP ZAMZ51] 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 YEp13-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.
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.
(continued on facing page)
The yeast lipoarnide dehydrogenase gene

The yeast lipoarnide dehydrogenase gene

Fig. 1. Nucleotide sequence of the LPDI gene and deduced primary structure of lipoarnide dehydrogenase. The nucleotide sequence of a 2.7 kb region containing the sense strand of the LPDI 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..TATGT..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 LPDI 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 postulated to specify a site involved in polyadenylation in higher eukaryotes (Proudfoot & Brownlee, 1976) and seen in many yeast genes is found at position +12. The sequence TAG..TA(T)GT..TTT 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: CAGTATAGTATATATATTT.

Regulation of the LPDI 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 LPDI 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 lipoarnide dehydrogenase by general amino acid control may be less...
The yeast lipoamide dehydrogenase gene

(a) GCN4 binding site consensus sequence  \(rtTGACTCattt-t\)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPD1</td>
<td>CGTGAATCGTTTTTAG</td>
<td>-253</td>
</tr>
<tr>
<td>LPD1</td>
<td>GATGACTCGTTTTTAAA</td>
<td>-236</td>
</tr>
<tr>
<td>LPD1</td>
<td>TTTGACTCACCTCGAA</td>
<td>-108</td>
</tr>
</tbody>
</table>

(b) CYC1 UAS1  
CYC1 UAS2  
GAL1-10 control region sequence

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYC1</td>
<td>CTCTTTGCGCGGGGTMTT</td>
<td></td>
</tr>
<tr>
<td>CYC1</td>
<td>CTCTTTGCGAGCAGTTT</td>
<td></td>
</tr>
<tr>
<td>GAL1-10</td>
<td>CTCTTTGGAACTTTCAG</td>
<td></td>
</tr>
<tr>
<td>LPD1</td>
<td>CTCATTGGCGAGAATC</td>
<td>-189</td>
</tr>
</tbody>
</table>

(c) Heat-shock consensus sequence  \(C--GAA--TTC--G\)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPD1</td>
<td>CACGAATAGTCATG</td>
<td>-335</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 GAL1-10 control region. (c) Heat-shock promoter consensus sequence.

Fig. 2. Protein sequence comparison. The primary structures of S. cerevisiae lipoamide dehydrogenase (YLPDH), pig heart lipoamide dehydrogenase (PLPDH), E. coli lipoamide dehydrogenase (ECLPDH), human red blood cell glutathione reductase (HGR) and Pseudomonas aeruginosa mercuric reductase (PAMR) are aligned for maximum homology. The residue numbering is based on S. cerevisiae lipoamide dehydrogenase. A consensus line below the five sequences shows residues common to three or more of the amino acid sequences. The borders of the domains representing the four regions of strong homology are based on glutathione reductase and are marked by arrowheads below the consensus sequence. The 80 residues at the N-terminus of mercuric reductase have been omitted. The individual sequences show residues in agreement with the consensus sequence in upper case and all others in lower case. The alignment was performed using the program PRETTY with the default protein comparison file (Gribskov & Burgess, 1986), which regards certain amino acids as similar for the purposes of the consensus line. The asterisks in the pig heart lipoamide dehydrogenase sequence indicate regions of unknown sequence. Residues common to pig heart and yeast lipoamide dehydrogenase but not to the E. coli lipoamide dehydrogenase are shown by circles above the yeast 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 GAL1-10 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 CTCCGCGGAG 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 CAR1 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


