Molecular Cloning and Nucleotide Sequence of the 3-Isopropylmalate Dehydrogenase Gene of Candida utilis

By KAZUHIRO HAMASAWA, YUKUHARU KOBAYASHI, SHIGEYOSHI HARADA, KOJI YODA, MAKARI YAMASAKI* AND GAKUZO TAMURA

Department of Agricultural Chemistry, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

(Received 2 July 1986; revised 27 October 1986)

A 3-isopropylmalate dehydrogenase (3-IMDH, EC 1.1.1.85) gene was cloned from a gene library of Candida utilis. One of the plasmids, pYKL30, could complement Escherichia coli leuB and Saccharomyces cerevisiae leu2 auxotrophs; a 2.2 kb HindIII fragment subcloned in pBR322 could still complement the leuB mutation. Southern hybridization confirmed that this fragment was derived from C. utilis. An open reading frame of 1089 bp that corresponded to a polypeptide of 363 amino acids, one residue shorter than the 3-IMDH of S. cerevisiae, was found in the cloned fragment. The homology between the 3-IMDHs of C. utilis and S. cerevisiae was 76.2% in nucleotides and 85.4% in amino acids. In contrast, the homology between the 3-IMDHs of C. utilis and Thermus thermophilus was much smaller and was restricted to some regions of the gene.

INTRODUCTION

Cell mass of strains of Candida utilis has been recognized to be useful as a food resource since the early studies of Fink et al. (1936) on the fermentative nature of the strains. As C. utilis generally requires no growth factors such as vitamins and has a high growth rate, it has been cultured on a large scale with sulphite waste liquor containing hemicellulose as the major carbon source. Culture broth of some strains of C. utilis contains many medically useful materials such as RNA, glutathione, NAD and coenzyme A. Although C. utilis is highly useful in the fermentation industry, there is little genetic information on this organism mainly because of the difficulty in obtaining auxotrophic mutants. In this report, a gene coding for 3-isopropylmalate dehydrogenase (3-IMDH, EC 1.1.1.85) was cloned from a strain of C. utilis and examined in detail. This gene is homologous to LEU2 of Saccharomyces cerevisiae, which is frequently used in genetic studies. It should be of help to construct Leu+ mutants, by gene disruption, to be used as recipient strains with a selectable marker in cloning based on C. utilis.

METHODS

Abbreviations: 3-IMDH, 3-isopropylmalate dehydrogenase; ORF, open reading frame.
and the part of the gel containing fragments larger than 1 kb was cut out. The DNA was extracted by freezing and thawing. The cloning vector YRp7 (4 µg) was digested with BamHI. Then the two types of fragment were ligated with T4 DNA ligase and the ligation products were used to transform E. coli RR1. Ampicillin-resistant and tetracycline-sensitive transformants were selected, and tested for complementation of leuB auxotrophy on minimal medium.

Southern hybridization. Chromosomal DNAs of E. coli, S. cerevisiae and C. utilis were prepared as described previously (Struhl et al., 1979) and digested with the restriction enzymes EcoRI and BamHI. The fragments were transferred from agarose gel to a nitrocellulose filter (Southern, 1975). The hybridization probe DNA was labelled by using a nick-translation kit (Amersham).

Sequencing. DNA was sequenced by the dideoxy chain termination method of Sanger et al. (1977). The sequence kit of Takara Brewery Co. was used. All of the sequences were determined at least twice on the same strand. The sequences were analysed with a PC9801 F2 personal computer (Nippon Electronic Co., Japan) and the software Genetyx-III (Software Development Co., Japan).

RESULTS

Cloning of the 3-IMDH gene of C. utilis

The E. coli RR1 transformants containing C. utilis chromosomal DNA were screened for complementation of the leuB6 auxotrophy. Fortunately, among the 492 transformants first obtained, we found two that grew well in minimal medium lacking leucine, one of which had a 10.7 kbp plasmid. This plasmid, pYKL30, also complemented the leucine deficiency of the S. cerevisiae leu2-3 leu2-112 auxotroph. The restriction map of plasmid pYKL30 is shown in Fig. 1. Localization of the 3-IMDH coding region in pYKL30 was examined as follows. The fragment containing TRP1 and ARS1 was removed by HindIII digestion followed by recircularization to
Fig. 2. Determination of the coding region of the *C. utilis* 3-IMDH gene. Bold lines represent the nucleotide sequence derived from *C. utilis*. Open lines represent the DNA of pBR322. Light lines show the deletion region in each derivative of pYKL40. The activity of complementation of the *E. coli* *leuB*6 auxotroph is indicated by Leu+ or Leu-. The line with arrowheads shows the coding region of the 3-IMDH.

Fig. 3. Strategy for sequencing the 3-IMDH coding and flanking regions. Fragments from pYKL45 were inserted in the multiple cloning site of pUC18 vector. Arrows beneath the bold lines represent the direction and extent of sequences determined. The arrow at the bottom shows the 3-IMDH coding region. Vertical lines represent the positions of restriction sites.

Make it easy to examine the region. This plasmid was named pYKL40. It should have no autonomously replicating sequence (ARS) that works in *S. cerevisiae*, because when *S. cerevisiae* AH22 was transformed with 1 μg pYKL40, less than Leu+ prototrophs appeared. Plasmid pYKL40 was further digested with various restriction enzymes and recircularized to delete non-essential parts. The 3-IMDH activity of the deletion plasmids was examined by complementa-
K. Hamasawa and Others

Sau3A///BamHI

```
GATCAGGGAGCAGACTGATAGAGAACAGAATACGCGCTCAGAAGCCGACATTCTAGCTTATGCTGACATCT
AGCTTCTGAGTTCTGAAGACTGCTGGTCTGGAAGTTGTTGTGAGTTCTCAGGACATCTCAGAGTCTGAG
CCTGAGGCCCCAGGTGGTGAGTGCCCCGAGCAGGAGGACCCGGCTGCCGAGAGAGCTCCTTCTGCT
TCTACTCCGCCTCTCCTGCTTTCCCTGCTCTCTAGCTTCGCTCCTTCCTTCTCTCTCTCTCTCTCTC
```

-300

```
-200

```

-100

```
-10

```

-1

```

YetProGluLYsThrlleValValLeuProGlv

```

I00

```

AspHlsVaIGIyThrGIuIleThrAlaGIuAlalleLy~ValLeuLysAlalleGIuGIuVaILYsProGIuIleGIuPGlu

```

Clal 200

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```

```
Candida utilis 3-IMDH gene

Candida utilis 3-IMDH gene

It was concluded that the 3-IMDH coding region was situated in a BamH1/Sau3A1–HindIII fragment of 2-2 kbp. Therefore, this fragment was recloned in pBR322 at the HindIII site to form pYKL45 (Fig. 1.). Southern hybridization experiments were done to test whether the sequence complementing the leuB6 mutation of E. coli corresponded to C. utilis DNA. HindIII-digested chromosomal DNAs from C. utilis clearly hybridized with pYKL40, while those from S. cerevisiae and E. coli did not (data not shown), so it was confirmed that the 2-2 kbp fragment originated from C. utilis.

DNA sequence of the 3-IMDH gene of C. utilis

To analyse the DNA sequence of the 3-IMDH gene of C. utilis, we sequenced the 2.2 kbp fragment of pYKL45 by the dideoxy chain termination method. Fig. 3 shows the sequence strategy. Since the fragment bears three unique cleavage sites of Clal, EcoRI and BglII, it was divided into four pieces by the enzymes. They were subcloned in the multiple cloning site of pUC18 vector (Yanisch-Perron et al., 1985). Sequencing of each fragment was repeated two or three times. The DNA sequence of the fragment (Fig. 4) consisted of 2209 bp and contained an open reading frame (ORF) of 1089 bp encoding 363 amino acids. A translatable amino acid sequence is shown under the nucleotide sequence in Fig. 4. The M, of the corresponding polypeptide was calculated to be 38700. A TATA-box-like sequence, the consensus sequence of the RNA polymerase II recognition site, could also be found at two places, namely at positions -258 and -30 (Sentenac & Hall, 1982).

DISCUSSION

The 3-IMDH gene of C. utilis was cloned and its nucleotide sequence was determined. This is the first report of the nucleotide sequence of a functional gene of C. utilis as far as we know. The sequence of 2209 bp was an alignment of four restriction fragments determined separately. The sequences of HindIII–Clal, Clal–EcoRI, EcoRI–BglII and BglII–HindIII fragments could be clearly read from one end to the other. However, if there were two proximal Clal sites or an EcoRI or BglII site which could not be detected by mapping with restriction endonucleases, they would be missed. To test this possibility we cleaved pYKL45 with one of ClaI, EcoRI and BglII completely, recircularized with ligase and used this DNA to transform E. coli RR1. As all the ampicillin-resistant transformants were also Leu+, there should be no missing fragments.

Nucleotide sequences of the 3-IMDH genes of S. cerevisiae (LEU2) and T. thermophilus (leuB) have already been determined. When the DNA sequence of the 3-IMDH gene of C. utilis was compared with that of S. cerevisiae, a considerable homology (76.2%) was found. Moreover, since only the third letter of the corresponding codon differs in many cases, the homology of amino acid sequences is as high as 85.4%. Thus the two sequences could be aligned in spite of the difference in species (Fig. 5). The 3-IMDH gene from Candida maltosa has been cloned (Kawamura et al., 1983), and its recently determined nucleotide sequence also has a high homology (76%) with the 3-IMDH gene of S. cerevisiae (M. Takagi, personal communication). The C. utilis 3-IMDH was one amino acid shorter than the S. cerevisiae 3-IMDH. In Fig. 5, the underlined sequence of 14 amino acids (288–301) in S. cerevisiae was considered by Andreadis et al. (1984) to be the substrate-binding site of the LEU1 and LEU2 gene products. The homology was 85.7% in this region.

The nucleotide sequence of the leuB gene of Thermus thermophilus (Kagawa et al., 1984) was compared with that of the 3-IMDH gene of C. utilis. There is only 12.9% homology in the ORFs, but in some parts more homologous stretches were found (Fig. 6). It is unknown what these

Fig. 4. Nucleotide and amino acid sequence of the 3-IMDH gene and its flanking regions. The amino acid sequence of the ORF was translated from the nucleotide sequence. Numbering of both nucleotides and amino acids start from the beginning of the coding sequence. TATA box-like sequences were boxed. Underlines in the 5' non-coding region indicate a G-C-rich palindromic sequence. Broken lines and boxes at the 3' non-coding region represents the presumptive consensus sequence of transcription termination.
K.HAMASAWA AND OTHERS

C. *utilis* - MPEKTIVVL PGDVGTEIT AEAIKVLKAI EEVKPEIKFN FQHILGGA A IDATGVPLPD

S. *cerevisiae* MSAPKIVVL PGDGQEt IT AEAIKVLKAI SDVRSNKFD FENHILGGA A IDATGVPLPD

DALEASKKAD AVLLGAVGP KWGTGVTPR QGLKIRKEL NLYANLPNP FASESLLDLS

EALEASKKVD AVLLGAVGP KWGTGSVPQA QGLKIRKEL QLYANLRLC FASDSSLDS

PIKAETVVKGT DVVVRELVG GIYFGRKED DSGVSADTTE TYSVPEVQRI TRMAAMFALQ

PIKPQFAKGT DVVVRELVG GIYFGMRKED DGQGVAWDES QTYVPEVQRI TRMAAFMALQ

HNPPPLTPSL DRAVLASSR LVRKVLVETI EKEFPQTVQ HQLIDSAAMI LIKYPTQMLN

HEPPLIPWISL DRAVLASSR LRKTVVEITI KNEFPQTVQ HQLIDSAAMI LVKNPQMLN

IVITSNMFGD IISDEASVIP GSLLLLPSAS LASLPDNSKA FGLYEPCHGS APDLPKNKVD

PIATILSAAM MLKLSDLIVE EGAVETAVK QVLDAKRTG DLKGTNTTE VGDAVEAVK

PIATILSAAM MLKLSDLNLPE EGKAIETAVK KVLDAKRTG DLGGNNTTE VGDAVEAVK

**KILA**

**KILA**

**Fig. 5.** Comparisons of the putative amino acid sequences of the ORFs between *C. utilis* and *S. cerevisiae*. The upper sequence represents *C. utilis* 3-IMDH and the lower one the *S. cerevisiae* 3-IMDH. Asterisks indicate homologous amino acids. The underlined region represents a possible substrate (3-isopropylmalate) binding site proposed by Andreadis et al. (1984).

partial homologies mean. Table 1 shows a comparison of codon usage among *C. utilis*, *S. cerevisiae* and *T. thermophilus* in the 3-IMDH coding region. *C. utilis* and *S. cerevisiae* are eukaryotes, and the codon usages are quite different from that of the prokaryote *T. thermophilus*, especially in the codons of leucine, glycine, proline and arginine. In the third letter of the codons, *C. utilis* preferred GC as compared to an AT choice in *S. cerevisiae*. As expected there is a high occurrence of GC residues in the third letter of the codons in the 3-IMDH of *T. thermophilus* (Table 1). It has been found that in the 5’ non-coding region of the 3-IMDH gene (*LEU2*) of *S. cerevisiae* a leader peptide rich in leucine precedes the translation initiation site, and a stem and loop structure can be formed at the end of the leader peptide (Yanofsky, 1983). This region resembles the attenuator structure which regulates some *E. coli* genes concerning amino acid synthesis. However, Erhart & Hollenberg (1983) later reported that this region is not relevant, because deletions in it did not affect the regulation of 3-IMDH gene expression. We searched the
Candida utilis 3-IMDH gene

(a) (46) -Gly-Gly-Ala-Ala-Ile-Asp-Ala- (52)
(b) (47) -Gly-Gly-Ala-Ala-Ile-Asp-Ala- (53)
(c) (41) -Gly-Gly-Ala-Ala-Ile-Asp-Ala- (47)

(a) (68) -Asp-Ala-Val-Leu-Leu-Gly-Ala-Val-Gly-Gly-Pro-Lys-Trp-Gly-
(b) (69) -Asp-Ala-Val-Leu-Leu-Gly-Ala-Val-Gly-Gly-Pro-Lys-Trp-Gly-
(c) (62) -Asp-Ala-Val-Leu-Leu-Gly-Ala-Val-Gly-Gly-Pro-Lys-Trp-Gly-

(a) (134) -Val-Arg-Glu-Leu-Val-Gly-Gly-Ile-Tyr-Phe-Gly-Glu-
(b) (135) -Val-Arg-Glu-Leu-Val-Gly-Gly-Ile-Tyr-Phe-Gly-
(c) (129) -Val-Arg-Glu-Leu-Val-Gly-Gly-Ile-Tyr-Phe-Gly-

(a) (259) -Glut-Pro-Ser-Leu-Gly-Leu-Pro-Ser-Ala-Ser-Leu-
(b) (260) -Pro-Ser-Leu-Gly-Leu-Pro-Ser-Ala-Ser-Leu-
(c) (249) -Pro-Ser-Leu-Gly-Leu-Pro-Ser-Ala-Ser-Leu-

(a) (284) -Glut-Pro-Cys-His-Gly-Ser-Ala-Pro-Asp-
(b) (265) -Pro-Cys-His-Gly-Ser-Ala-Pro-Asp-
(c) (246) -Glut-Pro-Cys-His-Gly-Ser-Ala-Pro-Asp-

(a) (300) -Pro-Ile-Ala-Thr-Ile-Leu-Ser-Ala-MET-MET-Leu-
(b) (301) -Pro-Ile-Ala-Thr-Ile-Leu-Ser-Ala-MET-MET-Leu-
(c) (284) -Pro-Ile-Ala-Thr-Ile-Leu-Ser-Ala-MET-MET-Leu-

Fig. 6. Homologous sequences in the ORFs of C. utilis (a), S. cerevisiae (b) and T. thermophilus (c). Identical amino acid sequences are boxed.

Table 1. Comparison of codon usage in the 3-IMDH gene of C. utilis, S. cerevisiae and T. thermophilus

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>Frequency C</th>
<th>Frequency S</th>
<th>Frequency T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>TTA  0 9 0</td>
<td>Gly GGT 22 28 1</td>
<td>Tyr TAT 0 0 1</td>
<td>Cys TGT 1 1 0</td>
</tr>
<tr>
<td></td>
<td>TTG 25 21 4</td>
<td>GGC 2 0 10</td>
<td>TAC 6 4 5</td>
<td>TGC 1 1 0</td>
</tr>
<tr>
<td></td>
<td>CTT 4 5 4</td>
<td>GGA 4 1 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTC 5 0 15</td>
<td>GGG 0 0 19</td>
<td>His CAT 1 3 0</td>
<td>Trp TGG 3 4 1</td>
</tr>
<tr>
<td></td>
<td>CTA 0 4 1</td>
<td>CAC 5 3 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGT 6 1 10</td>
<td>Pro CCT 0 7 4</td>
<td></td>
<td>Arg CGT 0 2 1</td>
</tr>
<tr>
<td></td>
<td>CCC 0 0 18</td>
<td>Gln CAA 7 9 1</td>
<td>CGC 0 0 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCA 21 16 1</td>
<td>CAG 2 0 2</td>
<td>CGA 0 0 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCG 4 5 10</td>
<td>CCA 1 0 7</td>
<td></td>
<td>CGG 0 0 11</td>
</tr>
<tr>
<td>Phe</td>
<td>TTT 5 5 3</td>
<td>Asn AAT 1 5 0</td>
<td>AGA 10 9 0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TTC 4 5 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>ATT 4 8 0</td>
<td>Thr ACT 7 4 0</td>
<td>AAC 11 8 6</td>
<td>AGG 0 0 6</td>
</tr>
<tr>
<td></td>
<td>ATC 20 16 7</td>
<td>ACC 8 8 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATA 1 1 1</td>
<td>ACA 5 4 0</td>
<td>Lys AAA 3 10 0</td>
<td>Ser TCT 12 10 2</td>
</tr>
<tr>
<td></td>
<td>ACG 0 0 6</td>
<td>AAG 21 18 15</td>
<td>TCC 8 9 8</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>ATG 7 7 6</td>
<td>Asp GAT 9 17 1</td>
<td>Ala GCT 16 13 2</td>
<td>TCG 1 0 1</td>
</tr>
<tr>
<td></td>
<td>GAT 16 18 0</td>
<td>GAC 11 7 15</td>
<td>GCC 17 16 28</td>
<td>AGT 0 3 0</td>
</tr>
<tr>
<td></td>
<td>GTG 14 7 12</td>
<td>GCA 4 5 1</td>
<td>GCC 17 16 28</td>
<td>AGC 0 1 4</td>
</tr>
<tr>
<td></td>
<td>GTA 0 0 0</td>
<td>Glu GAA 5 20 3</td>
<td>GGC 2 2 11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GTG 2 4 20</td>
<td>GAG 21 3 28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C. C. utilis; S. S. cerevisiae; T. T. thermophilus.

5' non-coding region of the C. utilis 3-IMDH gene for similar structures, but could not find them. Martinez-Arias et al. (1983, 1984) suggested on the basis of experiments on deletion mutants, that a palindromic sequence upstream of the TATA box, 5'-TGAGAGGCCGAACCGGCTTTTCA-3', participates in the regulation of the expression of the S. cerevisiae LEU2 gene. In the 5' non-coding region of C. utilis similar sequences were located at -204 to -181, -164 to -151 and -143 to -129, but they were downstream of the TATA box at -256.
A tRNA<sup>Leu</sup> gene containing an intervening sequence has been reported in the 5' non-coding region of the <i>S. cerevisiae</i> LEU2 gene (Andreadis et al., 1982): we searched for a homologue of the <i>C. utilis</i> tRNA<sup>Leu</sup> sequence described by Murasugi & Takemura (1978) but failed to find a corresponding sequence. Zaret & Sherman (1982) suggested that the 3' non-coding regions of several genes contain transcription termination sequences. <i>S. cerevisiae</i> also has such a sequence in the 3' non-coding region of the LEU2 gene. In <i>C. utilis</i>, the consensus signal sequence for termination and poly(A) addition, TAA/TAG/TGA...1-140*(T-rich)*...TAGT/TATG...*(A-T-rich)*...TTT...poly(A) was also found in this region (Fig. 4). Many kinds of dehydrogenases have a structure which is thought to be part of the coenzyme-binding domains. Recently it was proposed that ADP, a component of NAD or flavin-adenine dinucleotide (FAD), would form a structure of β-strand-α-helix-β-strand which was termed an ADP-binding βββ-fold (Froman et al., 1984). The authors compared many ADP-binding proteins and found two common characteristics of all complexes: (i) the occurrence of glycine residues at the N-terminus of the helix (e.g. Gly*Gly***Gly), which results in a favorable interaction between the α-helix dipole and the negatively charged pyrophosphate moieties, and (ii) the occurrence of the eight amino acids at a specific position in a peptide fragment. The total length of this amino acid sequence varies between 29 and 31 residues. We searched the <i>C. utilis</i> 3-IMDH gene for a sequence with these characteristics, but failed to find it, so the amino acid sequence of the ADP-binding βββ-fold may be different from the others reported (Wierenga et al., 1985, 1986).

There are many homologous sequences among the 3-IMDHs of <i>C. utilis</i>, <i>S. cerevisiae</i> and <i>T. thermophilus</i> (Fig. 6). It is unknown what these homologies mean, but they may in some way concern essential sequences such as the catalytic domain or the nucleotide binding site.

As we now have the nucleotide sequence of the 3-IMDH gene of <i>C. utilis</i>, it may be possible to get Leu<sup>+</sup> auxotrophic mutants of <i>C. utilis</i> by the gene disruption technique (Rothstein, 1983). Then, the 3-IMDH gene will be useful as a selective marker in gene manipulation of <i>C. utilis</i>.

**REFERENCES**


Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel
Candida utilis 3-IMDH gene


