Sequencing of a 9.9 kb Segment on the Right Arm of Yeast Chromosome VII Reveals Four Open Reading Frames, including \textit{PFK1}, the Gene Coding for Succinyl-CoA Synthetase (\(\beta\)-chain) and Two ORFs Sharing Homology with ORFs of the Yeast Chromosome VIII

PAULO GUERREIRO\textsuperscript{1}, DULCE AZEVEDO\textsuperscript{1}, TANIA BARREIROS\textsuperscript{1} AND CLAUDINA RODRIGUES-POUSADA\textsuperscript{1,*}

\textsuperscript{1}Laboratório de Genética Molecular, Instituto Gulbenkian de Ciência, Apartado 14, P-2781 Oeiras Codex, Portugal

Received 23 July 1996; accepted 26 August 1996

A 9.9 kb DNA fragment from the right arm of chromosome VII of \textit{Saccharomyces cerevisiae} has been sequenced and analysed. The sequence contains four open reading frames (ORFs) longer than 100 amino acids. One gene, \textit{PFK1}, has already been cloned and sequenced and the other one is the probable yeast gene coding for the \(\beta\)-subunit of the succinyl-CoA synthetase. The two remaining ORFs share homology with the deduced amino acid sequence (and their physical arrangement is similar to that) of the YHR161c and YHR162w ORFs from chromosome VIII. The sequence is in the EMBL data library under Accession Numbers Z73024, Z73025, Z73026, Z73028 and Z73029.

\textsuperscript{1}Correspondence to: Claudina Rodrigues-Pousada.

Contract grant sponsor: Commission of the European Community under the BIOTECH programme.

Contract grant sponsor: Calouste Gulbenkian Foundation.

INTRODUCTION

As a contribution to the European Union BIOTECH programme to sequence the \textit{Saccharomyces cerevisiae} genome, we have determined the sequence of a 9898 bp DNA fragment of the right arm of chromosome VII. The sequenced region is part of a 44 kb yeast segment inserted into the pWE15 cosmid that partially overlaps, at its left side, the insert of cosmid pEGH484 sequenced by H. Steensma’s group and, at its right, the insert of cosmid pUGH1273 sequenced by M. L. Carbone’s group, part of which was also sequenced by ourselves.

The fragment includes the previously sequenced \textit{PFK1} gene, the putative succinyl-CoA synthetase \(\beta\)-subunit encoding gene and two open reading frames (ORFs) sharing homology with other genes from yeast and rat. The latter presents a strong homology with ORFs from chromosome VIII and they both have the same physical arrangement as in this chromosome.
MATERIALS AND METHODS

Strains and plasmids

The pEGH301 cosmid was provided by H. Tettelin and A. Goffeau (DNA coordinators of chromosome VII) and consists of a 44 kb DNA fragment of the chromosome VII of *S. cerevisiae* strain S288C inserted in the pWE15 cosmid. Plasmid pUC19 was used for subcloning experiments. *Escherichia coli* strain XL1-Blue was the host used for transformation and amplification of plasmids. Standard techniques (Sambrook et al., 1989) were used for transformation of *E. coli*, restriction mapping, plasmid preparation, DNA ligation and other manipulations.

Sequencing strategy

Random libraries were generated from pEGH301 by digestion with the *Bg*II, *Cla*I, *Eco*RI, *Kpn*I and *Sfu*I restriction enzymes. Overlapping clones that cover the entire DNA fragment to be sequenced were chosen to generate subclones suitable to sequencing. The sequence of the 9898 bp of cosmid pEGH301 reported herein contains: the *Eco*RI-*Eco*RI, E18 (3·2 kb) subfragment, the *Kpn*I-*Kpn*I, K2 (2·0 kb) subfragment, the *Bg*II-*Bg*II, Bg7 (6·2 kb) subfragment, the *Sfu*I-*Sfu*I, Sf2 (1·7 kb), Sf12 (0·4 kb) and Sf11 (2·3 kb) subfragments, and the *Cla*I-*Cla*I, C2 (1·9 kb) subfragment. Clones suitable for sequencing were generated in two ways. Nested deletions were made using the Pharmacia Exonuclease III nested deletion kit in order to sequence one of the strands of the E18, Bg7, Sf11 and C2 subfragments. Direct deletions in all subfragments using suitable restriction enzymes were used for the remaining sequence. DNA sequencing was carried out by the dideoxy chain-termination method using T7 DNA polymerase (Pharmacia) and the ‘forward’ and ‘reverse’ primers. Synthetic oligonucleotides were also used to fill a few gaps. DNA strands were sequenced in both directions and gel compressions were solved by sequencing the fragment with Deaza G/A T7 sequencing mixtures (Pharmacia).

Sequence analysis software

Sequences from the individual reactions were assembled into longer contigs with the aid of DNASIS 5.0 analysis software (LKB-Hitachi). Comparison with the DNA sequences and deduced amino acid sequences to the GenBank and Swiss-Prot databases was done using MIPS facilities and the GCG package.

RESULTS AND DISCUSSION

The cosmid pEGH301 contains a fragment of about 44 kb located on the right arm of *S. cerevisiae* chromosome VII. Figure 1 presents the *Eco*RI restriction map of the fragment, the restriction map of the sequenced region and the location of the ORFs therein. The left extremity of the insert
overlaps H. Steensma’s cosmid pEGH484 and the right extremity overlaps M. L. Carbone’s cosmid pUGH1273.

The DNA sequence of the 9898 bp fragment was determined and is deposited in the EMBL data library under accession numbers Z73024, Z73025, Z73026, Z73028 and Z73029. This region comprises a total of four complete ORFs longer than 100 amino acids, which are listed in Table 1. Other genetic elements, such as tRNAs, Ty elements or introns, were not found.

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
<th>Strand</th>
<th>No. of amino acids</th>
<th>CAI</th>
<th>FASTA opt. score</th>
<th>Best homology with</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>G8599</td>
<td>4049–1089</td>
<td>C</td>
<td>987</td>
<td>0·47</td>
<td>100% identical</td>
<td>PFK1, phosphofructokinase 1</td>
<td>—</td>
</tr>
<tr>
<td>G8610</td>
<td>6896–5193</td>
<td>C</td>
<td>568</td>
<td>0·16</td>
<td>1137</td>
<td>YHR161c</td>
<td>1 TM span</td>
</tr>
<tr>
<td>G8620</td>
<td>7651–8088</td>
<td>W</td>
<td>146</td>
<td>0·10</td>
<td>590</td>
<td>YHR162w</td>
<td>—</td>
</tr>
<tr>
<td>G8625</td>
<td>9634–8354</td>
<td>C</td>
<td>427</td>
<td>0·19</td>
<td>269</td>
<td>O44 protein (rat)</td>
<td>—</td>
</tr>
</tbody>
</table>

The G8599 ORF has already been sequenced and identified as the PFK1 gene, which encodes the α-subunit of the enzyme phosphofructokinase (Heinisch et al., 1989). Phosphofructokinase is an octameric enzyme composed of four α-subunits and four β-subunits that catalyses the conversion of fructose-6-phosphate and ATP to fructose-1,6-biphosphate and ADP in glycolysis. Seven differences have been found between the two sequences.
probably due to the use of different strains. One is a deletion located in the 3' non-coding region (one C is missing at 932/933 position) and the remainder are located near the C-terminal end of the coding region. These differences do not cause any amino acid substitution (G to A at positions 1620 and 1716, G to T at position 1692, G to C at position 1782, T to C at position 1794, and A to G at position 1875).

G8610

The G8610 ORF encodes a putative protein of 568 amino acids that shows 51.4% identity with YHR161c (Figure 2), a putative protein of 637 amino acids from chromosome VIII (Johnston et al., 1994). Indeed, the comparison between the sequence of chromosome VIII and the present DNA sequence reveals significant similarities. Interestingly enough, both the deduced amino acid sequence and the physical arrangement of G8610 and G8620 ORFs on chromosome VII are similar to those of YHR161c and YHR162w on chromosome VIII, indicating that duplications/transpositions had occurred during chromosome segregation. However, the low homology of less than 50% observed between the intergenic regions of G8610/G8620 (755 bp) and of YHR161c/YHR162w (981 bp), suggests that this may represent an old event. Studies of the expression of these ORFs have shown that their transcripts are not detected by Northern analysis in exponentially growing cells, even when amounts corresponding to 20 μg of total RNA were used.

G8620

The G8620 ORF encodes a protein that shows 73.6% identity with YHR162w (Johnston et al., 1994) and 45.9% identity with the rat O44 protein (Figure 3; Tsou et al., 1986), whose functions are unknown. However, the rat O44 gene was found to be transcribed into nearly 100 different mRNAs due to alternative transcription initiation sites and polyadenylation sites. A putative alternative splicing site was also detected that produces mRNAs coding for proteins with different C-termini. The O44 gene is expressed in all tissues tested but in brain it transcribes longer mRNAs than those transcribed in liver or kidney (Tsou et al., 1986).

G8625

The G8625 ORF encodes a protein of 427 amino acids that is strongly homologous to the β-subunit of the succinyl-CoA synthetase from eight organisms (Figure 4), showing a maximum of 55.7% identity with the rumen fungus Neocallimastix frontalis enzyme. This suggests that G7742 defines the yeast SCS-β gene. SCS is a mitochondrial heterodimeric enzyme of the citric acid cycle that catalyses the conversion of succinyl-CoA to succinate and CoA together with the formation of GTP (Bridger, 1974). Assuming that G8625 is the yeast succinyl-CoA counterpart, it will be the longest SCS so far detected, with an extended N-terminus and three unique residues at position 245–247. The first 30 residues of the protein share typical characteristics of presequences of imported proteins.
mitochondrial proteins, namely seven basic residues (M1, R4, K5, K12, R18, R28 and R29), five serines (S3, S6, S8, S11 and S17) and no acidic residues (Hurt and van Loon, 1986).

ACKNOWLEDGEMENTS
We thank H. Tettelin, coordinator of sequencing chromosome VII, for providing us with the recombinant cosmid. We wish to acknowledge K. Kleine at MIPS for his help in computer analysis. This work was supported, in part, by the Commission of the European Community under the BIOTECH programme for sequencing the yeast genome and by grants of the Calouste Gulbenkian Foundation given to C.R.-P.

REFERENCES