

Nucleotide sequence of ORF2: an open reading frame upstream of the tRNA ligase gene

George A.Komatsoulis, Shawn K.Westaway and John N.Abelson

Division of Biology 147-75, California Institute of Technology, Pasadena, CA 91125, USA
Submitted October 1, 1987

Accession no.Y00473

During the course of sequencing the tRNA ligase gene from the yeast Saccharomyces cerevisiae, an open reading frame, ORF2, was discovered in its vicinity (1). ORF2 was found to be the locus of a single copy, essential gene (S. Westaway and J. Abelson, unpublished observations) with a transcript starting only 125 nucleotides upstream of tRNA ligase (1). The proximity of ORF2 to tRNA ligase suggested that their products might be transcriptionally coregulated and that its product might participate in the processing of tRNA precursors. For this reason we completely sequenced both strands of the ORF2 DNA using the method of Sanger, et. al. (2). The sequence is shown below, with the corresponding amino acid sequence of the putative protein product of 623 amino acids (mw 71,300). The left arrows show the transcription starts of tRNA ligase, and the right arrows show the initiation sites of ORF2 transcription. The protein is not significantly homologous to any protein on the NBRF database, leaving its function in question.

1 TGC GTG AGC GAT CCTCTCTTCG TAGTAATAT TTATGATGA CTAAAGGCA ACTGTAGATT TCGAAAGGA GAAGACTAT GGAATCTTC 202
285 Ser Thr Met
193 GAGCAAGT CTGAGAGA TGGATGAT TAACTACT TGTCTATAT TCTCTGGGA TAAAGTATA CTGAATTT ATGATAGCA AATGCTGCG 202
202 GCGCGCTG GAGATAAA GATATATTC TGAAGACTG TCGTAAAGC TATTGAAA CGAAGTGA TGAACCTGA ANAGUAGAG ACTGATATG 202
385 AAAGCTGAT TAAAGTCT TGCTTTTATA ATAGAAAGG GATTATAGAA ATG GCG GGT GAA ATT GAG ATT GAA GCT GCA TTA 191
390 CTT TTA TCG CAG GAG TTA GAA ACA ACT AAA CTC ACT TTC GAA ATC AAT AAA TCT TTG AAA AAT GCA GCT ACA TCC 472
474 AAT GAA TCC ACT GAA CTC CTC ACT GCT ATT CTT GCT AGA AAT AAT GTT TTA ACC ACA TTA GAA ACA AAT ATT GAA ACT 553
553 AAT GAA TCC ACT GAA CTC CTC ACT GCT ATT CTT GCT AGA AAT AAT GTT TTA ACC ACA TTA GAA ACA AAT ATT GAA ACT 553
554 TTG AAT TGC CTT GGC TCA CTT AAG GAT CTA GCG ACA GAA GGT TCC AAG GAT CAG ATT GAG CAG ATT TTA GAA AAG GGT ATT GAA 534
534 AAT GAA TCC ACT GAA CTC CTC ACT GCT ATT CTT GCT AGA AAT AAT GTT TTA ACC ACA TTA GAA ACA AAT ATT GAA ACT 534
615 CTT GCT TTA AAG CAA TAC ACC CAA CTA CTA GAT AAG CTA GAT GAT ATG TTG GAA GAG ATT GAG TCC GAA GAA GCG AAT CCA 715
715 GAG 715
716 GAA AAC AAC TGA TGA TTC GAT GGC ATT TTA ACT CAG TCG GAA GAA TTG ATC AAA CCG ACC GCG GGT CAA CTA GAA CTA 706
706 GAG 706
707 TTT ATT TCA ATT TTC AAC AGT ATT AAA CCG TTT GAT CCA CAA CTT ATT TTC ACC AAA AAG TCG CTA TTT CCA TAC CAG 677
708 GAG CAG CAG TTA GCG GCT TTA TCG TCG ATT TTA GAT TAT TTT GAT GGA AAT TCA GAA GGT TTT ATT ATA CAG CAG ATA CTC 958
958 GAG 958
959 CTT GGT GAA AGC ACT AAA TTA ATC CTC AAG TGC ATC GTC CTT GAA CTT TTT GCG AAA GAA ATC AGC ACT GAA AAA 1039
1039 GAG 1039
1040 GCG CCG TAT GCG AAG GCG ACT AAG GCG ATC AAC ACC TAC ACG GGC TTA TTA GCG TTC ATC GCG AAT AAA AAA CTA CTA 1120
1120 GAG 1120
1121 GTA GAG CAG CTC TAC TCT CAA TAT GAA GAA AAT AAA CCG TTT GAT CCA CAA CTT ATT TTC ACC AAA AAG TCG CTA TTT 1203
1203 GAG 1203
1204 CCG ACT GTC TTC GCT GCA AAT TTC AAA ATT CTA CCA ACG AAC CTC GAG AAC TTT GCA TTC TTT ATT TTT GAT CTA CTC 1282
1282 GAG 1282
1283 AGC ATA AAT GAT GAT CCA AAA TCT CTT CGA GCG AAG CAA CTA GAA AAC TAT AAT TTA TTG CAA GAT TCC ACG CAA GAA 1361
1361 GAG 1361
1362 CCG ACT GAA TCA ACA GAG TCG TTA TCG ACA GAT CTT GAT GCG ATT TTA AAG GCA AAT TTT ATT TCT ACT ATT CCG TCC 1443
1443 GAG 1443
1444 AAT AAT GGT CTT ACT GAA GCA ACT GTA GAT ACC ATC TCA ACA CTA ACA AAG TTC AAT GAG TAC AAA AAG GCA TGT TTA GCG 1525
1525 GAG 1525
1526 GCG ATC GAC AAT ATC ACA CCG GAA AAT TCG TTA GCA TCC AAT TAT AAG GAG AAA GAA TAC ACT TCA CAG AAC GAG CTT TTA 1606
1606 GAG 1606
1607 AAT TCG GAA GAT AAT GAT TTA CTA TCA TCA TTT TTT AAT ACT GAT TCC ATA GAC ACT TCG GCA GTT AAT CTT CAG ACA AAG 1687
1687 GAG 1687
1688 GCA CAA ATA GCA CTC ACT GCG AAC GAA CAG CAG GAT CCG GCT GAT GCG ATT TTA AAG GCA AAT TTT ATT TCT ACT ATT CCG 1748
1748 GAG 1748
1749 GGA TTC TTT ATT TTA ATC AAC CTC ACG CTT GTC GAG CAG ATC CTT GAA ANG TCA GAA TTA ACA TTT ATG TTA GCT GGA 1849
1849 GAG 1849
1850 GTC CAG TCC GAA GCG CTG AAG AAA GGT TAT AAT ACT TAT ATG ATC TCA GAT TCG ACA GAT TCC GCA ACT GCA AAT TTC 1930
1930 GAG 1930
1931 ATG TCT CTT CTT ATT GAT ACT ACG GCG AAC AAG TCA AAA CAG AAA CAA CAA ATA AAG CAG AAA TTT AAG AAT TCC AAT 2011
2011 GAG 2011
2012 GAA GGA TTC CAA GAT TTA CTA TCA AAA AAA CAG TAT AAG CTT TCA GAT GEA TCA TTT AAA CTA ACT TTC AAC TCA GAA 2092
2092 GAG 2092
2093 ATA ATA TCG TCG CTT GCG ATT GAT GAA ACA TCC TAC AAT ACA TAT AAA CAG TCT TCC AAG AAT GGT GAA AAG CAT ACT 2175
2175 GAG 2175
2176 AAT TAT ACC CTT GAG CAA CTA ACT ACT CTT AAT CAA TTA CTA CCA TAG ATTGGGAGCA TACTCTTTA AAGAGGAAA TTAAGAGAAA 2212
2212 GAG 2212
2265 GAGCTGAGC TAGAGATG TAGAGTATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA 2264
2265 GAGCTGAGC TAGAGATG TAGAGTATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA TAGATGATA 2264
2485 AAAGCTTGA AATAAATCTT TCCCTGGAT CAGATTC 2502

Acknowledgements: Computer resources were provided by BIONET™ National Computer Resource for Molecular Biology, whose funding is provided by the Biomedical Research Technology Program, Division of Research Resources, NIH (grant 1-U41RR-01685-05). This work was supported by a National Research Service Award (5 T32 GM07616-09) from the National Institute of General Medical Sciences to G.A.K. and S.K.W.

References:

- 1. Westaway, S., Phizicky, E., and Abelson, J. N. manuscript in preparation.
2. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) PNAS 74,5463-5467.