

-1000 ggtctgatcg cttgtgtgcg agaggctgac catcatgccg agcctgttca gctccggtac aagctcgcgc ccgaaggcgg tcaggccgtt cccgctcgtg

-900 acaggctcga taggagagcc gtccccggcg gaagaggcga aggctgtgag aaaaaccctg caacagagcg gaagaaggcc gagaggctgc ctccgctcac

-800 aagtgtggca ggtgtgogtg agcgtaaggt accgcacgcc gagttgctga aagaggcgta ggatggccag cgagttcatc agtgatgctg agctgtagga

-700 agagcggggt agcgggagga tgcggtaga gactgttctg acaaaccctt ctagc**ccaat** caaggacgcg atcttccctt ccgcgaacgc actccgcacc

-600 tcgtccgccc ttcgagcgag cgccatctcc tcaggatagt gctccaccat ccgatgaata aggtcgacgc tctcgagtgc ccactcgacc gcgttctgtg

-500 gctcaaggaa gtctccccc gtcggctgcg catcacacgg agcatacgca acatggaaca agcctcctac gcgtccttga cgcaacttgc ggaggctcag

-400 atggcctgca agaccggtcg caagttcagg gaggacgtcg agcggctctc ggcagaggtg gcgagcgacg taggggaggt cgacgtgtcc gtctatgagt

-300 ggggtggcgtt tcaggatccc gcgggctca gcgagcaagg actcgttggg caaagtgaag gtgtactgat gagcggttgc **gaggccgcag agcagcgtg**

-200 **cgacgacggg** aagcttcggc acgagcatga ctgtgagtag tagtccaagg agaacagcgc agagtccgca ggagggcaca tggaggcaga gcgtggggcg

-100 gaggaggcag atggggagtc gcgctggggg acgagagggg gccgctcgac caactgctct **ctttegctct** **tgctgctgct** TGTA**CTGCTC** **GAACGACGCC**

1 **ATG**CATTAC AGAAGCGCGT CGTTGCTCTC GGATCAGGCG gtcgctcttt **ccctctctc** **cccacaccg** acagttctcg aggaggagta cagcagcgag

P1 M H S Q K R V V V L G S G

101 cgaggctgcc gagggggatc tgggttgacg **cagctcttga ttatacagTG ATTG**GCCTGA GCAGCGCCCT CATCTCGCT CGGAAGGGCT ACAGCGTGCA

P14 **V I G L S S A L I L A R K G Y S V H**

201 CATTGTGGCC CGGACTTGC CGGAGGACGT CTCGAGCCAG ACTTTCGCTT CGCGTGGGC TGTAGGTCGT CTCGTTCCGG CTCTCAGATG TCCCTGTGAG

P32 I V A R D L P E D V S S Q T F A S P W A

301 CTGACCATT CCATGTGTGA CGCAGGGCGC GAACTGGACG CCGTTCATGT CTCTCACGGA CGGTCTCTGA CAAGCAAAGT GGAAGAATT GACTTTgtgc

P52 G A N W T P F M S L T D G P R Q A K W E E L T L

401 gcttcccttc tgtgtctctt tttcactctg tgctaaccaa gctcgtgcta **cacagCAAGA** AGTGGGTGCA GCTTGTACCG ACCGGCCAGG TCATGTGGCT

P76 K K W V E L V P T G Q V M W L

501 CAAGGGGACG AGGCGGTTCC CGCAGAACGA GGATGGCTTG CTCGGGCATT GGTACAAGGA CATCACGCC AACgtggggt caccgctagt tctgctgctc

P91 K G T R R R F A Q N E D G L L G H W Y K D I T P N

601 ctcaacctca ccaggtaact gacgaccctc **cttcgctcgc acgcagTACC** GCCCCCTCCC TTCTCTCCGAA TGTCCCCCA ACTCGATCGG CGTAACTTAC

P115 Y R P L P S S E C P P N S I G V T Y

701 GACACCCTCT CCGTCCACGC ACCAAAGTAC TGCCAGTACC TCGCGAGGGG ATTACAGAAG CTGCGCGGA CGTTTGAGAG ACGGACGGTC ACGTCGGTGT

P133 D T L S V H A P K Y C Q Y L A R G L Q K L G A T F E R R T V T S V

801 AGCAGGCGTT CGAGGGGGTG GATTGTGGTG TCAACGCTAC GGGTCTCGgt gctcgttacc ttgtctctca tgcgtcaagc agcttactaa cagtctcttc

P167 E Q A F E G V D L V V N A T G L

901 gcagGCGCA AGTCGATCGC TGGCATCGAC GACCAAGCCG CCGGCAAT CCGCGTCAA ACCGTCTCTG TCAAGTCCG CTGCAAGCGC TGCACGATGG

P182 G A K S I A G I D D Q A A E P I R G Q T V L V K S A C K R C T M

1001 ACTCGTCCGA CCCCTCTTCT CCCGCTTATA TCATCCCACG ACCAGTGGC GAAGTGTACT GCGGCGGAC GTACGGCGTG GCGGATGGG ACTTGTCCGT

P215 D S S D P S S P A **M** I I P R P G G E V I C G G T **M** G V G D W D L S

1101 CAACCTGAG ACGGCTCAGC GGATCTCAA GCACTGCCTG CGACTCGACC CGTCCATCTC GAGCGACGGC ACGATCGAGG GTATCGAAGT CCTCCGGCAC

P248 V N P E T V Q R I L K H C L R L D P S I S S D G T I E G I E V L R H

1201 AACGTTGGCT TGCACCCGC ACGTCGAGGC GGCCCCGCG TCGAGGCAGA ACGGCTCGCT CTCCCTCTCG ACCGGTCCAA ATCGCCCTC TCGCTTGCA

P281 N V G L **R** P A R R R G G P R V E A E R L V L P L D R S K S P L S L G

1301 AGGCACGAC AAGGGCGCA AAGGAGAAGG AGGTGACACT CGTGATCGC TATGGTCTT CGAGCGCCGG GTACCAGCAG AGTTGGGGCG CGGCGGAGGA

P315 K G T T R A A K E K E V T L V H A Y G F S S A G Y Q Q S W G A A E D

1401 TGTTGCGCTG CTTGTGAGG AGGCGTTCCA GCGGTACCAC GGCGCTGCAA GGGAGTCGAA GCTGTAGGGC GAGGCTTGAG **TGGTGTATT** GGGGCATTC

P348 V A L L V E E A F Q R Y H G A A R E **S K L** *

1501 ACCAGTTCG CCGCgtatcg agaagagagt gtggcgaccc ttgcaccgt ctgaccttgc caaccgctca actccttgc ctttggcctg cttgctctct

↑ PolyA site

1601 cggtctctcc tggatttctc aagctcctc ategaagctg teactcgca ggcacatcct teaagctgtc gtcgcccctt tctctctct cctctctctc

Addition file 1. Structure of *R. toruloides* DAO1 gene. Nucleotide positions are defined from the translation start codon “ATG” (in which the position of “A” is +1). Protein positions are labeled with the prefix “P”. The translational start codon within *NcoI* site (CCATGG) in the intron 1-containing promoters are and the ATG to ATC mutation are boxed. Exons are shown in capital letters and introns are in lowercase italics; the putative CAAT box is highlighted in grey; DRE1, *ct* box and IES1 are indicated with shaded text. The putative coenzyme binding consensus sequence (GXGXXG, where “X” is any amino acid) is underlined; the dimerization loop is double underlined; the amino acid residues H-bonded to the flavin O-2 atom (YQ) are in bold and italics; TG repeats upstream of the polyA site are highlighted in grey; the residues interacting with the

-carboxy and -amino groups of the substrate (Y₂₂₃, Y₂₃₈ and R₂₈₅) are boxed and highlighted in grey; the C-terminus peroxisomal targeting signal (**SKL**) is highlighted in grey. Abbreviations: *ct* box: pyrimidine-rich region; *tsp*: transcriptional start point; *PolyA site*: polyadenylation site.