ELSEVIER

Gene 190 (1997) 245–249

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Identification of an acetolactate synthase small subunit gene in two eukaryotes

Ronald G. Duggleby *

Centre for Protein Structure, Function and Engineering, Department of Biochemistry, University of Queensland, Brisbane 4072, Australia

Received 30 September 1996; received in revised form 27 November 1996; accepted 28 November 1996

Abstract

Acetolactate synthase catalyses the first step in branched-chain amino acid biosynthesis. The bacterial enzyme contains two large and two small subunits but there is only limited and circumstantial evidence for a small subunit in the eukaryotic enzyme. Here this evidence is summarised and protein sequences of two putative eukaryotic small subunits, from a yeast and a red alga, are presented. © Elsevier Science B.V. All rights reserved.

Keywords: Branched-chain amino acid; Chloroplast genome; Herbicide; Yeast genome

1. Introduction

Acetolactate synthase (ALS) is an essential enzyme in plants and many microorganisms because it catalyses the first step in the biosynthesis of branched-chain amino acids. In some bacteria it also plays a catabolic role, supplying acetolactate for the butanediol fermentation.

There appear to be two distinct forms of the enzyme that correspond to these functional roles. The anabolic enzyme contains FAD (Schloss et al., 1985) and is inhibited by the branched-chain amino acids (Weinstock et al., 1992) while the catabolic enzyme, sometimes referred to as the 'pH 6 acetolactate-forming enzyme', displays neither of these properties (Störmer, 1968; Peng et al., 1992). A further property of the anabolic enzyme is that it is inhibited by a number of compounds that are used as herbicides (Schloss et al., 1988). The remainder of this article concerns the anabolic enzyme only.

Many of the bacterial ALSs have been shown to be heterotetramers composed of two types of subunit, large and small. The latter subunit was first identified (Squires et al., 1983) for *Escherichia coli* isoenzyme III (ALSIII); DNA sequencing revealed an open reading frame that appeared to have a homologue in the operon that contains the gene for *E. coli* ALSII (Lawther et al., 1981). The protein product of the small subunit gene was later identified for *E. coli* ALSI (Eoyang and Silverman, 1984) and *Salmonella typhimurium* ALSII (Schloss et al., 1985).

The role of the small subunit is not entirely clear and it may be that it is involved in more than one way. For the various *E. coli* isoforms it has been shown that this subunit affects sensitivity to branched-chain amino acids (Eoyang and Silverman, 1986; Sella et al., 1993), specific activity (Lu and Umbarger, 1987), stability (Sella et al., 1993) and the kinetic properties (Weinstock et al., 1992).

Putative small subunit genes have been identified for a number of other bacterial species. This identification has been based mainly, and in most cases solely, on the presence of an open reading frame 3′ to the large subunit gene. In contrast, the presence of a small ALS subunit has never been demonstrated unequivocally in eukaryotes. Certainly no open reading frame nearby the large subunit gene has been identified but this is not surprising since operons are not a feature of eukaryotic genomes. However, there is some evidence that a small subunit may exist.

First, purified wheat ALS contains a low molecular weight component (Southan and Copeland, 1996) that could be a small subunit; on the other hand, it could be simply an impurity. Purified barley ALS has been reported to contain no small subunit (Durner and Böger, 1988) on the basis of SDS-PAGE. However, it is conceiv-

* Corresponding author. Tel.: +61 7 33654615; Fax: +61 7 33654699; e-mail: duggleby@biosci.uq.edu.au

Abbreviations: ALS, acetolactate synthase.

0378-1119/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved
PII S0378-1119(97)00002-4

```
Bfl     ----------------MANSDVTRHILSVLVQDVDGIISRVSGMFTRRAFNLVSLVSAKT   44
Ccr     MTANVQPAPASAYDLSPKDQAEQSATFALLVDNEPGVLHRVVGLFAARGYNIESLTVAET   60
Cgl     ----------------MANSDVTRHILSVLVQDVDGIISRVSGMFTRRAFNLVSLVSAKT   44
EcoH    ----------------------MRRILSVLLENESGALSRVIGLFSQRGYNIESLTVAPT   38
EcoM    ----------------------MQHQVNVSARFNPETLERVLRVVRHRGFHVCSMNMAAA   38
EcoN    ----------------MQNTTHDNVILELTVRNHPGVMTDVCGLFARRAFNVEGILCLPI   44
Mav     -------------------MSPQTHTLSVLVEAKPGVLARVAALFSRRGFNIESLAVGAT   41
Sav     ---------------------MSKHTLSVLVENKPGVLARITALFSRRGFNIDSLAVGVT   39
Sty     ----------------------MRRILSVLLENESGALSRVIGLFSQRGYNIESLTVAPT   38
                                            .        .  .  .   * ...  .

Bfl     E-THGINRITVVVD-ADELNIEQITKQLNKLIPVLKVVRLDEETT-IARAIMLVKVSADS  101
Ccr     DRKAHTSRITVVTR-GTRHVLDQIEAQLNKVVNVRRVHDVTRDPNGVERELALVKVRGSG  119
Cgl     E-THGINRITVVVD-ADELNIEQITKQLNKLIPVLKVVRLDEETT-IARAIMLVKVSADS  101
EcoH    D-DPTLSRMTIQTV-GDEKVLEQIEKQLHKLVDVLRVSELGQGAH-VEREIMLVKIQASG   95
EcoM    S-DAQNINIELTVA--SPRSVDLLFSQLNKLVDVAHVAICQSTTT--SQQIRA-------   86
EcoN    Q-DSDKSHIWLLVN--DDQRLEQMISQIDKLEDVVKVQRNQSDPTMFNKIAVFFQ-----   96
Mav     E-QKDMSRMTIVVS-AEETPLEQITKQLNKLINVIKIVELEDGNS-VSRELALIKVRADA   98
Sav     E-HPDISRITIVVNVIEALPLEQVTKQLNKLVNVLKIVELEPSAGRAGGELVLVKVRADN   98
Sty     D-DPTLSRMTIQTV-GDEKVLEQIEKQLHKLVDVLRVSELGQGAH-VEREIMLVKMEASG   95
                . .         .. .  *. *.  * ..

Bfl     TNRPQIVDAANIFRARVVDVAPDSVVIESTGTPGKLRALLDVMEPFG-IRELIQSGQIAL  160
Ccr     VDRLEALRIAEIFRAKPVDTTLESFVFEISGAPSKIDKFLDLMRPLG-LVELSRTGVLSI  178
Cgl     TNRPQIVDAANIFRARVVDVAPDSVVIESTGTPGKLRALLDVMEPFG-IRELIQSGQIAL  160
EcoH    YGRDEVKRNTEIFRGQIIDVTPSLYTVQLAGTSGKLSAFLASIRDVAKIVEVARSGVVGL  155
EcoM    ------------------------------------------------------------   86
EcoN    ------------------------------------------------------------   96
Mav     GTRSQVIEAVNLFRAKVIDVSPEALTIEATGDRGKIEALLRVLEPSV-SVRSS-NREWCR  156
Sav     ETRSQIVEIVQLFRAKTVDVSPEAVTIEATGGSDKLEAMLKMLEPFR-HQGARQSGTIAI  157
Sty     YGREEVKRNTEIFRGQIIDVTPTLYTVQLAGTSDKLDAFLASLRDVAKIVEVARSGVVGL  155

Bfl     NRGPKTMAPAKI-------                                           172
Ccr     ERGFEGM------------                                           185
Cgl     NRGPKTMAPAKI-------                                           172
EcoH    SRGDKIMR-----------                                           163
EcoM    -------------------                                            86
EcoN    -------------------                                            96
Mav     CPGPRGIGTAK--------                                           167
Sav     GRGARSITDRSLRPLDRSA                                           176
Sty     SRGDKIMR-----------                                           163
```

Fig. 1. Alignment of selected ALS small subunit protein sequences. Sequences were obtained from GenBank and aligned using the ClustalW (Thompson et al., 1994) program. An asterisk indicates a totally conserved residue, while a full stop denotes a position where there are conservative substitutions. Abbreviations used are: Bfl, *Brevibacterium flavum* MJ233; Ccr, *Caulobacter crescentus*; Cgl, *Corynebacterium glutamicum*; EcoH, *E. coli ilvH* (ALSIII); EcoM, *E. coli ilvM* (ALSII); EcoN, *E. coli ilvN* (ALSI); Mav, *Mycobacterium avium*; Sav, *Streptomyces avermitilis*; Sty, *S. typhimurium*.

able that it could be lost during multistep purification; in this context it is relevant that the various *E. coli* isoforms have differing affinities for their respective small subunits and that, for ALSIII, the small subunit is readily lost (Sella et al., 1993). In addition, even when a small subunit is present, it is not easily observed by SDS-PAGE (De Rossi et al., 1995) because it migrates as a rapidly moving, diffuse band that stains only weakly with Coomassie blue.

Second, we have confirmed (Chang and Duggleby, unpublished) that expression of the *Arabidopsis thaliana* ALS-encoding gene in *E. coli* results in an enzyme that, unlike the enzyme from the plant itself, is insensitive to inhibition by branched-chain amino acids (Singh et al., 1992). The suggested explanation (Singh et al., 1992) is that the expressed enzyme lacks a small subunit, although no evidence was adduced to support this proposal. A number of other explanations of this observation are possible, such as different post-translational processing, including proteolysis, between prokaryotes and eukaryotes. The plant enzyme is located in the chloroplast and contains an amino-terminal sequence that is believed to be a chloroplast transit peptide (Mazur et al., 1987). Although the enzyme expressed in *E. coli* is processed to a similar size as the native enzyme (Singh et al., 1992), it is not known whether cleavage of the transit peptide is at the same site as in the plant. Expression of the yeast enzyme in *E. coli* also results in an enzyme that is kinetically distinguishable from the native enzyme (Poulsen and Stougaard, 1989); this difference has also been ascribed to the lack of the appropriate small subunit.

Third, over-expression of the *A. thaliana* ALS-encoding gene in tobacco (Odell et al., 1990) or oilseed rape (Ouellet et al., 1994) gives greatly elevated amounts of the corresponding mRNA, but much smaller increases in ALS activity. This lack of correlation could be interpreted to indicate that some other component, such as a small subunit, is limiting.

Although none of these lines of evidence for an ALS

```
Bfl   ------------------------------------------------------------    0
Ppu   ------------------------------------------------------------    0
Sce   MLRSLLQSGHRRVVASSCATMVRCSSSSTSALAYKQMHRHATRPPLPTLDTPSWNANSAV   60

Bfl   ---------MANSDVTRHILSVLVQDVDGIISRVSGMFTRRAFNLVSLVSAKTETHGINR   51
Ppu   ---------------MKHTLSVLVQDEAGVLSRISGLFARRGFNIASLAVGPAEQIGVS-   44
Sce   SSIIYETPAPSRQPRKQHVLNCLVQNEPGVLSRVSGTLAARGFNIDSLVVCNTEVKDLSR  120
                       * *  *** + *++** **  + *+**+ ** +   *    +

Bfl   ITVVVDADELNIEQITKQLNKLIPVLKVVRLDEETTIARAIMLVKVSAD-----------  100
Ppu   ITMVVQGDNRTIEQLTKQLYKLVNILNVQDVTNIPSVERELMLIKIQVN-----------   93
Sce   MTIVLQGQDGVVEQARRQIEDLVPVYAVLDYTNSEIIKRELVMARISLLGTEYFEDLLLH  180
       * * ++     **   *   *+    * + ++     *++  +

Bfl   -----------STNRPQIVD---------------------------AANIFRARVVDVA  122
Ppu   -----------SQNRIEALEVK----------------------------IFRANVVDIA  114
Sce   HHTSTNAGAADSQELVAEIREKQFHPANLPASEVLRLKHEHLNDITNLTNNFGGRVVDIS  240
                 *+        +                             *   ***+

Bfl   PDSVVIESTGTPGKLRALLDVMEPFGIRELIQSGQIALNRGPK--TMAPAKI--------  172
Ppu   EDLIIVEVTGDPGKIVAIEQLLTKFGIIEIARTGKISLVRTSKINTYLKDKVVAYNA---  171
Sce   ETSCIVELSAKPTRISAFLKLVEPFGVLECARSGMMALPRTPLKTSTEEAADEDEKISEM  300
      +   ++*    *  + *   +   **  * ++ *   * *+

Bfl   ---------                                                     172
Ppu   ---------                                                     171
Sce   VDISQLPPG                                                     309
```

Fig. 2. Alignment of a bacterial and two putative eukaryotic ALS small subunit protein sequences. The *B. flavum* MJ233 (Bfl), *P. purpurea* (Ppu) and *S. cerevisiae* (Sce) sequences were obtained from GenBank and aligned using the ClustalW (Thompson et al., 1994) program. An asterisk indicates a totally conserved residue, while a plus denotes a position where the two eukaryotic sequences are identical. The underlined segment represents the most highly conserved region.

small subunit in eukaryotes is alone convincing, taken together they suggest that further work is merited. Recent advances in genome sequencing have provided an opportunity to look for ALS small subunit genes. Here the presence of such genes in two eukaryotic species is reported.

2. Results and discussion

There is limited similarity between the protein sequences of known ALS small subunits as illustrated in Fig. 1, which shows an alignment of several such sequences. In all, only four residues are totally conserved and a further 17 positions show conservative substitutions. A consensus sequence was derived from this alignment and compared to the individual sequences. The *B. flavum* and *C. glutamicum* sequences are most similar to the consensus and the former was used as a representative small subunit sequence. Various databases in GenBank were searched for conceptual translations into protein sequences that are similar to this *B. flavum* small subunit, using the BLAST program. In addition to known bacterial ALS small subunit genes, this search identified two well-matched eukaryotic genes: one (with a probability of arising by chance of 8.2×10^{-42}) from the chloroplast genome of the red alga *Porphyra purpurea* (Reith and Munholland, 1995) and the other ($P = 2.7 \times 10^{-25}$) from chromosome III of the yeast *Saccharomyces cerevisiae* (Oliver et al., 1992). The next best match ($P = 1.0$) corresponded to a fragment of a mouse lipoxygenase gene (Chen et al., 1994), mistranslated in a reverse reading frame. The *S. cerevisiae* and *P. purpurea* sequences are shown in Fig. 2, aligned with that of the ALS small subunit of *B. flavum*. These results clearly indicate that *S. cerevisiae* and *P. purpurea* contain a gene that could encode an ALS small subunit. A total of 38 residues are identical in all three sequences and, when only the two eukaryotes are compared, there are 66 identities. The most highly conserved region is a 26 residue sequence near the amino terminus, with the motif LVQXXXG$\phi\phi$SRϕSGXXXXRXFNϕXSL, where X is any amino acid while ϕ is one of V, L or I.

The *S. cerevisiae* sequence is substantially longer than the others; compared to that of *P. purpurea*, there are an additional 50 residues in the middle of the protein, a short extension at the carboxyl-terminus, and a 75 residue extension at the amino-terminus. It is suggested that the role of the latter is to act as a mitochondrial transit peptide, since it is known that *S. cerevisiae* ALS is a mitochondrial enzyme (Ryan and Kohlhaw, 1974). The transit peptides of other mitochondrial proteins frequently contain an arginine residue at position -2 relative to the cleavage site (von Heijne et al., 1989) and it is noted that this putative *S. cerevisiae* ALS small subunit contains an arginine at position 75, close to where homology with the *P. purpurea* sequence begins. Thus, it is proposed that the cleavage site is immediately after K76. It has also been noted (von Heijne et al., 1989) that mitochondrial transit peptides are enriched in A, L, R and S, but deficient in D and E. Similar characteristics are observed here, most notably for S

(which constitutes 18.4% of the first 76 residues but only 7.3% of the remaining 233 residues) and E (1.3% versus 9.4%).

Unlike ALS large subunits from plants (Mazur et al., 1987), the proposed *P. purpurea* ALS small subunit does not contain a chloroplast transit sequence. However, this is not necessary as the gene is located in the chloroplast genome. Thus it is suggested that in this plant, the large subunit is synthesised in the cytoplasm and transported to the chloroplast where it associates with the chloroplast-encoded small subunit. This arrangement is very similar to the situation often observed for ribulose 1,6-*bis*phosphate carboxylase, except that in that case it is the larger of the two subunits that is encoded by the chloroplast genome (Spreitzer, 1993).

Finding what appears to be an ALS small subunit gene in two eukaryotes as diverse as a yeast and a red alga suggests that small subunit genes will exist in other plants and fungi. However, the location of this gene, as well as that for the large subunit, may be variable. For example, it has been shown that in another red alga, *P. umbilicus*, an ALS large subunit is encoded by a chloroplast gene (Reith and Munholland, 1993). Further, the location of the *P. purpurea* ALS small subunit gene in the chloroplast may be unusual. We have searched for this gene in the complete chloroplast genomes of five other plants: *Nicotiana tabacum* (Shinozaki et al., 1986), *Oryza sativa* (Hiratsuka et al., 1989), *Pinus thunbergii* (Tsudzuki et al., 1992), *Marchantia polymorpha* (Ohyama et al., 1986) and *Odontella sinensis* (Kowallik et al., 1995). A total of 608 open reading frames were examined but the best match with the motif mentioned previously contained only 8 of the 17 conserved residues and bore no overall similarity to ALS small subunits; in contrast, the three sequences in Fig. 2 match in all 17 positions.

Because ALS is the target for several herbicides (Schloss et al., 1988), there has been considerable interest in transforming crop plants with herbicide-resistant forms of the enzyme (Odell et al., 1990; Ouellet et al., 1994). The success of this procedure is likely to be limited if a small subunit is an essential component of the plant enzyme. Thus, the work reported here may have significant practical implications. At present, there is no evidence that ALS small subunit genes exist in any eukaryotic species apart from *S. cerevisiae* and *P. purpurea*, or that even in these species the genes are actually expressed. Indeed, it is possible that these two genes serve an entirely different function that is unrelated to ALS activity. Ultimately the function of any DNA sequence, whose identity is based solely on homology, can only be proven by experiments designed to evaluate that function. In the case of these putative eukaryotic ALS small subunit genes, their function might be demonstrated by gene disruption or by co-expression with the large subunit genes. Current studies in this laboratory are examining these possibilities.

3. Note added in proof

Recent examination of GenBank expressed sequence tags has identified three sequences (two from *A. thaliana* and one from rice) that may represent higher plant ALS small subunits. The last of these gives a very good match to the *P. purpurea* sequence; over residues 83–154 there are 46 identical, and 10 similar, amino acids. This EST is apparently encoded in the nuclues, as it is not present in the rice chloroplast genome.

References

Chen, X.S., Kurre, U., Jenkins, N.A., Copeland, N.G. and Funk, C.D. (1994) J. Biol. Chem. 269, 13979–13987.

De Rossi, E., Leva, R., Gusberti, L., Manachini, P.L. and Riccardi, G. (1995) Gene 166, 127–132.

Durner, J. and Böger, P. (1988) Z. Naturforsch. 43c, 850–856.

Eoyang, L. and Silverman, P.M. (1984) J. Bacteriol. 157, 184–189.

Eoyang, L. and Silverman, P.M. (1986) J. Bacteriol. 166, 901–904.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y. et al. (1989) Mol. Gen. Genet. 217, 185–194.

Kowallik, K.V., Stoebe, B., Schaffran, I., Kroth-Pancic, P. and Freier, U. (1995) Plant Mol. Biol. Rep. 13, 336–342.

Lawther, R.P., Calhoun, D.H., Adams, C.W., Hauser, C.A., Gray, J. and Hatfield, G.W. (1981) Proc. Natl. Acad. Sci. USA 78, 922–925.

Lu, M.F. and Umbarger, H.E. (1987) J. Bacteriol. 169, 600–604.

Mazur, B.J., Chui, C.F. and Smith, J.K. (1987) Plant Physiol. 85, 110–1117.

Odell, J.T., Caimi, P.G., Yadav, N.S. and Mauvais, C.J. (1990) Plant Physiol. 94, 1647–1654.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.

Oliver, S.G., van der Aart, Q.J.M., Agostoni-Carbone, M.L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J.P.G., Benit, P. et al. (1992) Nature 357, 38–46.

Ouellet, T., Mourad, G., Brown, D., King, J. and Miki, B. (1994) Plant Sci. 102, 91–94.

Peng, H.L., Wang, P.Y., Wu, C.M., Hwang, D.C. and Chang, H.Y. (1992) Gene 117, 125–130.

Poulsen, C. and Stougaard, P. (1989) Eur. J. Biochem. 185, 433–439.

Reith, M.E. and Munholland, (1993) J. Curr. Genet. 23, 59–65.

Reith, M.E. and Munholland, J. (1995) Plant Mol. Biol. Rep. 13, 333–335.

Ryan, E.D. and Kohlhaw, G.B. (1974) J. Bacteriol. 120, 631–637.

Schloss, J.V., Ciskanik, L.M. and Van Dyk, D.E. (1988) Nature 331, 360–362.

Schloss, J.V., Van Dyk, D.E., Vasta, J.F. and Kutny, R.M. (1985) Biochemistry 24, 4952–4959.

Sella, C., Weinstock, O., Barak, Z. and Chipman, D.M. (1993) J. Bacteriol. 175, 5339–5343.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K. et al. (1986) EMBO J. 5, 2043–2049.

Singh, B., Szamosi, I., Hand, J.M. and Misra, R. (1992) Plant Physiol. 99, 812–816.

Southan, M.D. and Copeland, L. (1996) Physiol. Plant. 96, 171–179.
Spreitzer, R.J. (1993) Annu. Rev. Plant. Physiol. Mol. Biol. 44, 411–434.
Squires, C.H., DeFelice, M., Deveraux, J. and Calvo, J.M. (1983) Nucleic Acids Res. 11, 5299–5313.
Störmer, F.C. (1968) J. Biol. Chem. 243, 3740–3741.
Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
Tsudzuki, J., Nakashima, K., Tsudzuki, T., Hiratsuka, J., Shibata, M., Wakasugi, T. and Sugiura, M. (1992) Mol. Gen. Genet. 232, 206–214.
von Heijne, G., Steppuhn, J. and Hermann, R.G. (1989) Eur. J. Biochem. 180, 535–545.
Weinstock, O., Sella, C., Chipman, D.M. and Barak, Z. (1992) J. Bacteriol. 174, 5560–5566.