Nucleotide sequence of the promoter region of the citrate synthase gene (*gltA*) of *Escherichia coli*

Elizabeth P. Hull, Margaret E. Spencer, David Wood and John R. Guest*

Department of Microbiology, Sheffield University, Sheffield S10 2TN, England

Received 18 April 1983

The *gltA* gene, specifying the citrate synthase (EC 4.1.3.7) of *Escherichia coli*, has been isolated and the nucleotide sequence of a 752 basepair segment containing the *gltA* promoter and encoding 96 amino-terminal residues of the protein has been defined using the dideoxy/M13 method. The results confirm the location and transcriptional polarity of the *gltA* gene and indicate that the *gltA* transcript may contain a long leader sequence of 302–306 nucleotides upstream from the coding region.

Citrate synthase gene *Promoter region* *DNA sequence* Escherichia coli

1. INTRODUCTION

Citrate synthase (EC 4.1.3.7) catalyses the condensation reaction between acetyl-CoA and oxaloacetate and, as the first enzyme of the Krebs cycle, it performs essential functions both in energy generation and carbon assimilation. It has been the subject of detailed enzymological and physiological investigations [1,2] but relatively little is known about the regulation of citrate synthase gene expression. In *Escherichia coli* citrate synthase is encoded by the *gltA* gene, which is located at 16.2 min in the linkage map (fig.1), very close to 4 genes encoding other Krebs cycle enzymes: *sdhA,B*, succinate dehydrogenase (large and small subunits); *sucA*, 2-oxoglutarate dehydrogenase, and *sucB*, dihydrolipoamide succinyltransferase [3–6]. Hybrid plasmids containing the *gltA* gene have been selected by nutritional complementation of a *gltA* mutant and the *gltA* gene has been located in a 3.1 kilobasepair (kb) sub-fragment (*Hin*dIII–*Eco*RI) of a larger cloned region that extends over 31 kb of bacterial DNA [5,7]. Enzyme amplifications of up to 15-fold have been observed with plasmid-containing strains [7,8]. Post-infection labelling studies with λ*gltA* transducing phages have shown that the *gltA* gene is expressed with an anti-clockwise polarity [5]. This contrasts with the succinate dehydrogenase genes and the genes encoding the specific components of the 2-oxoglutarate complex, which are transcribed with clockwise polarities relative to the *E. coli* linkage map (fig.1). We are engaged in studies aimed at:

(i) Elucidating the molecular mechanisms which control the expression of Krebs cycle genes and lead to the elaboration of a coordinated metabolic cycle;

(ii) Comparing the structures and expression of the genes encoding several pairs of related enzymes; e.g., succinate dehydrogenase and fumarate reductase and the E1 and E2 components of the 2-oxoglutarate and pyruvate dehydrogenase complexes.

During the course of this work the nucleotide sequence of a 752 basepair segment of DNA containing the *gltA* promoter and encoding 96 amino-terminal residues of citrate synthase has been defined.

2. MATERIALS AND METHODS

2.1. *Bacterial strains, phages and plasmids*

The citrate synthase mutant W620 (*thi-1 pyrD36*

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V.
00145793/83/$3.00 © Federation of European Biochemical Societies

gltA6 galK30 rpsL129) and its *recA* derivative, JRG1499, were used in complementation tests. Strain C600 (*thr-1 leuB6 thi-1 supE44 tonA21 lacY1*) was used for assaying and propagating lambda phages [5] and JRG1370, a *recA* λ^R derivative of C600, was used routinely as the transformation host for plasmid isolation [9]. The characteristics of the λ*gltA* (λG112, λG113) and λ*gltAsdhABsucAB* (λG116–118) transducing phages have been described in [5]. Plasmids containing different segments of the *gltA–sucB* region, pGS64–66 and pGS91 (fig.1), were constructed by sub-cloning fragments from appropriate digests of *gltA–sucB* DNA into the *Bam*HI or the *Bam*HI plus *Sal*I sites of pBR322. The plasmid derivatives were characterized by detailed restriction analysis. The media and routine methods of genetic selection and gene manipulation have been described in [5,7,9].

2.2. *DNA sequencing*

DNA sequence analysis was by the dideoxy chain-termination method using M13 templates prepared on *E. coli* JM101 and synthetic 17-mer primer [10,11]. The sequence of the *gltA* promoter region was obtained by applying the strategy adopted for sequencing the 13000 basepair segment of DNA extending from the *Nru*I site in *gltA* to the *Sal*I site beyond *sucB* (fig.1). This has involved a combination of (a) directed cloning for specific fragments that are readily cloned into the versatile M13 vectors, M13mp8–9 [12] and (b) shot-gun cloning for the products of ultrasonic fragmentation, cloned into *Sma*I digested and phosphatased M13mp8 DNA [13], so that fully-overlapping sequence derived from both strands is obtained. The sources of DNA containing the *gltA* promoter region were random fragments from the *Nru*I–*Xho*I segment of pGS91 and specific fragments from the bacterial insert of λG117.

3. RESULTS AND DISCUSSION

3.1. *Location of the citrate synthase gene*

A genetic map showing the organisation and transcriptional polarities of 5 genes encoding Krebs cycle enzymes is presented in fig.1 together with a simplified restriction map of the corresponding segment of DNA. The citrate synthase of *E. coli* has a subunit M_r 46000–47000 [5,8], which

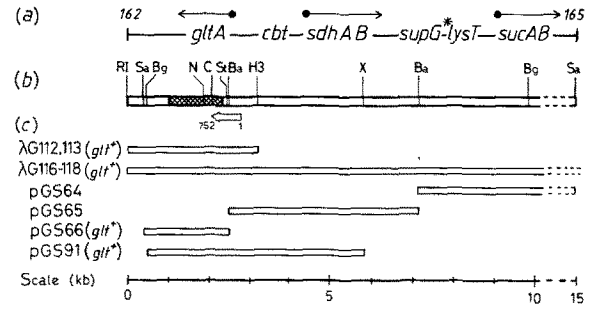

Fig.1. (a) Genetic map showing the relative positions and polarities of expression of the citrate synthase gene (*gltA*) and four other Krebs cycle genes (*sdhA,B, sucA,B*) in the 16 min region *E. coli* linkage map [5,6]. (b) Restriction map of the corresponding segment of DNA indicating the cleavage sites of the enzymes used for fragment isolation and for sub-cloning in pBR322 and M13 vectors. (c) Segments of DNA cloned in λ vectors [5,7] and into the *Bam*HI and *Sal*I sites of the *tet* gene of pBR322. The position of the *gltA* coding region (about 1.3 kb) is indicated by the shaded portion of the restriction map and the arrow defines the limits of the nucleotide sequence presented in fig.2. *Abbreviations:* Ba, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; H3, *Hin*dIII; N, *Nru*I; RI, *Eco*RI; Sa, *Sal*I; St, *Sst*I; X, *Xho*I.

corresponds to a coding region of approximately 1.3 kb. The location of the *gltA* gene has been identified by studies with λ*gltA* transducing phages containing the 3.1-kb *Eco*RI–*Hin*dIII fragment of bacterial DNA [5]. Lysogenic and dilysogenic derivatives of *E. coli* W620 (*gltA*) having λ*gltA* prophages at the bacterial *att*λ site exhibit a Glt^+ phenotype. Under these conditions, the phage-cloned gene is flanked by silent phage promoters, and this indicates that the 3.1-kb fragment contains both the *gltA* structural gene and the *gltA* promoter. Further studies have shown that the citrate synthase lesion is complemented when *gltA* mutants are transformed to Amp^R by plasmids pGS66 and pGS91 (fig.1). These plasmids contain even less of the critical region and indicate that the *gltA* structural gene is located in the 2.0 kb *Bgl*II–*Bam*HI segment of the restriction map.

3.2. *The nucleotide sequence*

The overall strategy for defining the complete nucleotide sequence of the citrate synthase promoter region and the succinate dehydrogenase and 2-oxoglutarate dehydrogenase complex genes has been outlined in section 2.2. The sequence of 752

basepairs containing the putative *gltA* promoter and the first 96 codons of the structural gene is presented in fig.2. It is completely consistent with the physical map obtained from restriction analyses with phages and plasmids containing the corresponding segment of DNA (fig.1,2). The coding region was identified using the computer program FRAMESCAN [14], which detected only one open reading frame. This starts at position 459 and exhibits a consistently high score with respect to the codon preferences of strongly-expressed *E. coli* genes. The coding region is also expressed with the polarity that was predicted for the *gltA* gene from post-infection labelling studies with λ*gltA* transducing phages [5]. The sequence analysis was not extended further into the *gltA* structural gene because this is being investigated by H.W. Duckworth (personal communication): the sequences of other segments of the *gltA–sucB* region will be reported elsewhere.

3.3. *Translation initiation and the amino acid sequence*

The coding region starts with an AUG initiation codon (positions 460–462, fig.2). This is preceded by a well-placed ribosome-binding site (Shine-Dalgarno sequence), UAAGGAG (position 446–453), having 7 consecutive bases complementary to the 3′-end of 16 S ribosomal RNA (fig.2). The sequence around the initiation codon has most of the features that have been identified in typical translational initiation regions [15,16].

The amino-terminal sequence predicted from the nucleotide sequence agrees with the 20-residue sequence determined by automated Edman degradation of the purified enzyme [8], if it is assumed that the formylmethionine is removed post-translationally to expose alanine as the amino-terminal residue (fig.2). The only discrepancy concerns residue 10, which is encoded by an asparagine codon (AAC), rather than the aspartate that was detected protein-chemically. Peptides with sequences corresponding to the predicted sequence for residues 21–55 have also been characterized (H.W. Duckworth, personal communication).

The *gltA* coding region starts 158 basepairs from the *Bam*HI site and this is entirely consistent with earlier predictions that the citrate synthase structural gene is located in the 2.0-kb *Bgl*II–*Bam*HI

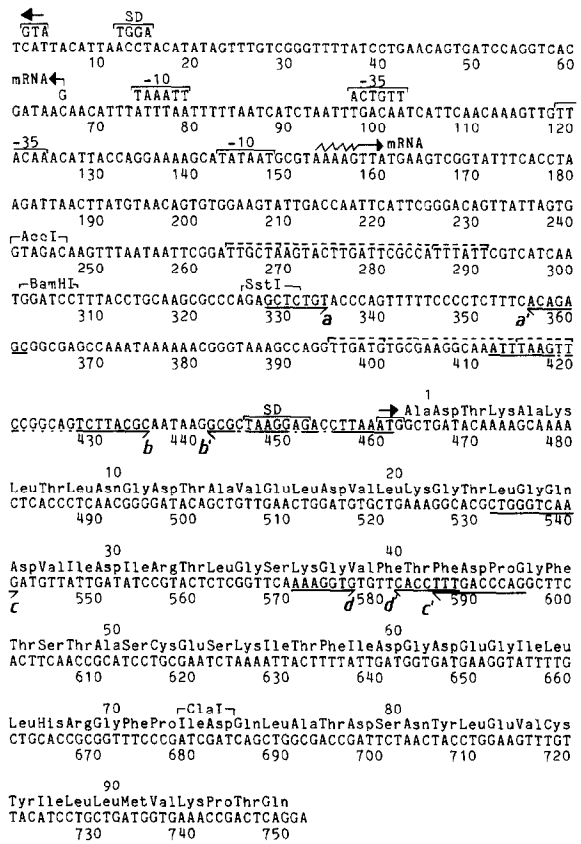

Fig.2. Nucleotide sequence of the promoter region of the *gltA* gene and the amino-terminal sequence of the citrate synthase of *E. coli*. The sequence shows 752 bases of the non-coding strand of DNA that contains the *gltA* promoter and 96 codons specifying the amino-terminal segment of citrate synthase. The nucleotide coordinates are numbered from an arbitrary site upstream of the structural gene. Relevant restriction sites are included and the proposed ribosome binding site (Shine-Dalgarno sequence, SD) and translation initiation site (⟶) are indicated. The −35 and −10 sites of the preferred *gltA* promoter sequence are also shown with potential sites for transcription initiation. The dashed lines denote sequences containing other potential promoter sequences (see text). Selected regions of dyad symmetry are underlined by converging arrows. The amino acid sequence is numbered from the amino-terminal alanine residue. Small segments of the complementary strand containing the promoter and translational initiation region of a gene that is expressed in the opposite direction to the *gltA* gene, are also included at the beginning of the sequence.

region, and that it is expressed anti-clockwise with respect to the linkage map (fig.1).

3.4. *Transcriptional initiation sites*

The sequence upstream from the coding region contains several potential RNA polymerase −35 sequences (recognition sites) and −10 sequences (Pribnow boxes), including examples of both consensus sequences: −35, TTGACA (96–101) and −10, TATAAT (143–148) [17]. The best combination involves a putative −35 sequence, TTACAA (position 119–124), located 18 basepairs upstream from the consensus −10 sequence (fig.2). If this is the functional promoter, the transcription initiation site is likely to be one of the purine nucleotides at positions 153–157 and the citrate synthase mRNA should contain a leader sequence of 302–306 untranslated nucleotides. Another possible promoter occurs at position 264–291 (fig.2). Further studies, including transcript mapping, will be needed to identify the functional promoter. It is interesting to note that pGS66, which expresses citrate synthase, contains the *gltA* coding region but not the preferred *gltA* promoter (fig.1). The sequence between the *Bam*HI site and the coding region contains one promoter-like sequence (395–420), but this is probably too short to be functional and it seems more likely that the *gltA* gene of pGS66 is transcribed by the plasmid-encoded *tet* promoter operating across the critical *Bam*HI site.

3.5. *Other features of the nucleotide sequence*

In the first 102 nucleotides there is another potential promoter region, a ribosome binding site and a translational initiation site for a gene that is expressed in the opposite direction to the *gltA* gene (fig.2). This segment of the sequence and the coding region associated with it will be described in a subsequent publication.

The sequence contains many regions of hyphenated dyad symmetry in both the coding and non-coding regions. Those capable of forming stable stem and loop structures in the mRNA transcript are identified by letters in fig.2, and their free energy values [18] are (kcal/mol): *aa'*, −8.0; *bb'*, −17.0; *cc'*, −13.3 and *dd'*, −10.0. The most stable structure (*bb'*) contains the ribosome binding site and it also possesses internal symmetries allowing the formation of alternative stem and loop structures by the pairing of positions 338–442 or 446–450 with 454–458. The significance of these structures is not clear but those in the mRNA leader sequence may be involved in *gltA* gene expression. The synthesis of citrate synthase, like other Krebs cycle enzymes, varies considerably with respect to growth substrate and is subject to catabolite and anaerobic repression [19,20]. Armed with the nucleotide sequence of the *gltA* promoter and with readily manipulated M13 clones containing this region, it should now be possible to investigate the factors controlling citrate synthase synthesis at the molecular level, and make comparisons with the other Krebs cycle genes that are under investigation in this laboratory.

ACKNOWLEDGEMENTS

This work was supported by project grant GR/B91396 from the Science and Engineering Research Council.

REFERENCES

[1] Srere, P.A. (1975) Adv. Enzymol. 47, 57–101.
[2] Weitzman, P.D.J. and Danson, M.J. (1976) Curr. Top. Cell. Regul. 10, 161–204.
[3] Ashworth, J.M., Kornberg, H.L. and Nothmann, D.L. (1965) J. Mol. Biol. 11, 654–657.
[4] Creaghan, I.T. and Guest, J.R. (1972) J. Gen. Microbiol. 71, 207–220.
[5] Spencer, M.E. and Guest, J.R. (1982) J. Bacteriol. 151, 542–552.
[6] Bachmann, B.J. and Low, K.B. (1980) Microbiol. Rev. 44, 1–56.
[7] Guest, J.R. (1981) J. Gen. Microbiol. 124, 17–23.
[8] Duckworth, H.W. and Bell, A.W. (1982) Can. J. Biochem. 60, 1143–1147.
[9] Guest, J.R., Roberts, R.E. and Stephens, P.E. (1983) J. Gen. Mic. 129, 671–680.
[10] Sanger, F., Coulson, A.C., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[11] Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmus, R.C. (1981) Nucleic Acids Res. 9, 1691–1706.
[12] Messing, J. and Vieira, J. (1982) Gene 19, 269–276.
[13] Deininger, P.L. (1983) Anal. Biochem. 129, 216–223.
[14] Staden, R. and McLachlan, A.D. (1982) Nucleic Acids Res. 10, 141–156.

[15] Atkins, J.F. (1979) Nucleic Acids Res. 7, 1035–1041.
[16] Stormo, D.G., Schneider, T.D. and Gold, L.M. (1982) Nucleic Acids Res. 10, 2971–2996.
[17] Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319–353.
[18] Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Grall, J. (1973) Nature New Biol. 246, 40–41.
[19] Gray, C.T., Wimpenny, J.W.T., Hughes, D.E. and Mossman, M.R. (1966) Biochim. Biophys. Acta 117, 22–32.
[20] Gray, C.T., Wimpenny, J.W.T. and Mossman, M.R. (1966) Biochim. Biophys. Acta 117, 33–41.