

Sequence and analysis of the citrulline biosynthetic operon *argC–F* from *Bacillus subtilis*

Mary O'Reilly and Kevin M. Devine

Author for correspondence: Kevin M. Devine. Tel: +353 1 7021872. Fax: +353 1 6798558.
e-mail: kdevine@vax1.tcd.ie

Department of Genetics,
Trinity College, Dublin 2,
Ireland

The citrulline biosynthetic operon *argC–F* located at 100° on the *Bacillus subtilis* chromosome contains seven open reading frames which encode all the enzymes required for the biosynthesis of citrulline. The operon is transcribed as a single transcription unit. The second cistron of the operon is homologous to ArgJ (ornithine acetyltransferase) from *Bacillus stearothermophilus* and *Neisseria gonorrhoeae*, suggesting that the acetylation of glutamate and the deacetylation of acetylornithine are carried out by a single enzyme in a cyclical pathway. The *argF* gene is an orthologue of *argF* from *Pseudomonas aeruginosa* and a paralogue of *arcB* from *P. aeruginosa* and *argFlargI* from *Escherichia coli*.

Keywords: *Bacillus subtilis*, citrulline biosynthetic operon, *argC–F*

The citrulline biosynthetic operon *argC–F* is located at approximately 100° on the *Bacillus subtilis* chromosome between the loci *metD* and *trpS*. A schematic diagram of the operon is presented in Fig. 1. The operon contains seven open reading frames (ORFs), *argC*, *argJ*, *argB*, *argD*, *carA*, *carB* and *argF*, encoding all the enzymes required for the biosynthesis of citrulline. These were identified, with the exception of *argC* and *argF*, by their homology with genes from *Escherichia coli*, *Neisseria gonorrhoeae* and *Bacillus stearothermophilus*. This operon is transcribed as a single transcription unit. There are three overlapping ORFs (*argB/argD*, *carA/carB* and *carB/argF*), suggesting that translation of these ORFs is coupled. The sequences of *argC* and *argF* were reported previously (Smith *et al.*, 1990; Mountain *et al.*, 1990), and our sequences concur with those reported.

It has been reported that DNA located between the *argC* and *argB* cistrons of this *B. subtilis* operon complements both *argA* and *argE* mutants of *E. coli* (Mountain *et al.*, 1984, 1986). There is, however, only one ORF between the *argC* and *argB* cistrons in our sequence. The putative product of this ORF is 64% identical at the amino acid level to ArgJ from *B. stearothermophilus* and 37% identical to the ArgJ protein of *Neisseria gonorrhoeae* (Fig. 2, Table 1; Martin & Mulks, 1992; Sakanyan *et al.*, 1993). The ArgJ protein is an ornithine acetyltransferase, an enzyme which uses acetylornithine (an intermediate in the citrul-

line biosynthetic pathway) as acetyl donor in the first reaction in the pathway, the acetylation of glutamate. Thus the acetylation of glutamate and the deacetylation of acetylornithine reactions in the biosynthesis of citrulline in *B. subtilis* appear to be carried out in a manner similar to that observed in *B. stearothermophilus* (Sakanyan *et al.*, 1992, 1993) and *N. gonorrhoeae* (Martin & Mulks, 1992), and differently from that found in *E. coli*, where two separate enzymes, ArgA and ArgE, carry out the two reactions (Cunin *et al.*, 1986).

The percentage amino acid identities between the ORFs from this operon and homologous proteins from *B. subtilis* and other organisms are shown in Table 1. The percentage amino acid identities with homologues from *B. stearothermophilus* range from 51% (ArgB) to 64% (ArgJ), and with homologues from *E. coli* from 33% (ArgB) to 42% (CarB). *B. subtilis* contains two carbamoyl phosphate synthetases, one involved in the biosynthesis of arginine (CarA, CarB, this work) and a second involved in pyrimidine biosynthesis (PyrAA, PyrAB, Quinn *et al.*, 1991). These homologues are 48% (CarA, PyrAA) and 53% (CarB, PyrAB) identical (Table 1). The subunits of the two *B. subtilis* carbamoyl phosphate synthetases are more similar to each other than either is to the *E. coli* enzyme, suggesting that the gene duplication event in *B. subtilis* occurred after the divergence of Gram-negative and Gram-positive bacteria.

The *Pseudomonas aeruginosa* genome contains two genes encoding ArgF-like proteins, *argF* (Itoh *et al.*, 1988) and *arcB* (Bauer *et al.*, 1987), which appear to reflect an ancient

The EMBL accession number for the nucleotide sequence reported in this paper is Z26919.

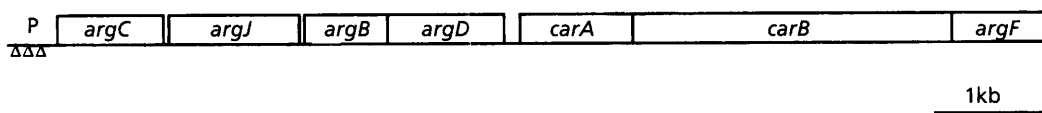


Fig. 1. Schematic representation of the citrulline biosynthetic operon from *B. subtilis*. Each ORF is represented by a box. A gap between boxes indicates that the ORFs do not overlap; a single vertical line indicates that these ORFs do overlap. The operon promoter is represented by P and the putative arginine boxes recognized by the arginine repressor are indicated by triangles (Δ).

```

B. SUBTIL M-IQLSEdqIVKVT-GDVSSPKGFQAKGVHCGLRYSKDLGVI ISETPAVSAAVYTQSHF
B. STEARO MTITKQYQGVAVADGTVVTPPEGFQAAGVNAAGLRYSKNDLGVILCDVPASAAAVYTQSHF
N. GONORR MAVNLTektEQLPDIDGIALYTAQAGVKKPG---HTDLTLIAVAAGSTVGAVFTTNRFF
*          *          *          *          *          *          *          *          *          *
B. SUBTIL QAAPIKVTQDSLKHGPTLKAVIVNSAIANACTGEQGLKDAYTMRESFASQLGIEPELVAV
B. STEARO QAAPLKVTQASLAVEQKQLQAVIVNRPCANACTGAQGLKDAYEMRELCAKQFGLALHHVAV
N. GONORR CAAPVHIKASHLDFDEdGVRALVINTGNANAGTGAQGRIDLAVCAAARQIGCKPNQVMP
***. . . . * . . . . . * * * * * * * * * * * * * * * * *
B. SUBTIL SSTGVIGEHLDMEKI HAGIELLKE--TPAGSGDFEEAILTDTVIKQTCYELAIGGK-TV
B. STEARO ASTGVIGEYLFMEKIRAGIKQLVPGVTMADAEAFQTALITDVTMKRACYQTTIDGK-TV
N. GONORR FSTGVILEPLPAKIIAALPKMQPAFWNEA---RAIMTDTVPKASREGKVGdQHTV
***** * * * * * * * * * * * * * * * * * * * * * *
B. SUBTIL TIGGARKGSGMIHPNMATMLGFVTTDAAIEEKALQKALRETTDVSFNQITVDGETSTNDM
B. STEARO TVGGAARKGSGMIHPNMATMLAFITTDANVSSPVLHAALRSITDVSFNQITVDGDTSTNDM
N. GONORR RATGIARKGSGMIHPNMATMLGF IATDAKVSQPVQLMTQEIADETFNITITVDGDTSTNDS
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *
B. SUBTIL VLVMANACAENECLTE-DHPDWPVFKALLLTCEDLAKIARdGEGATKLEAQVQGAKN
B. STEARO VVVMASGLAGNDELTP-DHPDWFENFYEARLKTCEdLAKQIAKdGEGATKLEIVRVGAKT
N. GONORR FVI IATGKNSQSEIDNIADPRYAQLKELLCSLALAQAIVRdGEGATKFI TVRVENAKT
. . . * . . . . * . . . . * * * * * * * * * * * * * * *
B. SUBTIL NLDANVIAKKIVGSNLVKTAVYGTdANWGRIIGAIGHSA-AQVTABEEVEVYLGQCLFK-
B. STEARO DEEAKKI AKQIVGSNLVKTAVYGDANWGRIIGAIGYSD-AEVNPDNDVAIGPMVMLK-
N. GONORR CDEARQAAAYAAARSPLVKTAF FASDFNLGKRLAAIGYADVADLDTDLVEMYLDLVAEH
* * * * * * * * * * * * * * * * * * * * * * * * * * *
B. SUBTIL NNEPQPFSEIAKEYLEGDEITIVIKMAEGDGNRAWGCdLTYDIKINASYRT
B. STEARO GSEPQPFSEEEAAAYLQqETVVI EVDLHIGdGVVAVWGCdLTYDYVKINASYRT
N. GONORR GGRAASYTEAQGQAVMSKDEITVRIKLHRGQAATVYTCdLSHGYSVINADYRS
. . . * . . . . * * * * * * * * * * * * * * * * *
    
```

Fig. 2. Alignment of the amino acid sequences of the ArgJ proteins from *B. subtilis*, (*B. subtil*) [this work] and *N. gonorrhoeae* (*N. gonorr*) [Martin & Mulks, 1992] and *B. stearothermophilus* (*B. stearo*) [Sakanyan et al., 1993] using the CLUSTAL V package (Higgins et al., 1992). Exact amino acid matches between all three proteins are indicated by an asterisk (*); conservative amino acid substitutions are indicated by a full stop (.)

Table 1. Percentage identity between the proteins encoded by the citrulline biosynthetic operon *argC-F* of *B. subtilis* and homologous proteins from other bacteria

<i>B. subtilis</i> <i>argC-F</i> protein	(F _{op})*	Percentage identity with proteins from:		
		Gram-negative bacteria†	<i>B. subtilis</i>	<i>B. stearo-</i> <i>thermophilus</i>
ArgC	(0.22)	ArgC ^{Ec} 37	-	ArgC‡ 59
ArgJ	(0.30)	ArgJ ^{Ng} 37	-	ArgJ 64
ArgB	(0.25)	ArgB ^{Ec} 33	-	ArgB 51
ArgD	(0.30)	ArgD ^{Ec} 38	-	ArgD‡ 62
CarA	(0.23)	CarA ^{Ec} 35	PyrAA 48	-
CarB	(0.25)	CarB ^{Ec} 42	PyrAB 53	-
ArgF	(0.26)	ArgF ^{Pa} 47	-	-
		ArcB ^{Pa} 42		
		ArgF ^{Ng} 42		
		ArgF ^{Ec} 40		
		ArgI ^{Ec} 40		

* Frequency of optimal codon usage.
 † Ec, Ng and Pa superscripts refer to proteins identified in *E. coli*, *N. gonorrhoeae* and *P. aeruginosa* respectively.
 ‡ Comparisons calculated using partial amino acid sequences.

duplication event (the encoded proteins exhibit only 40% identity). The *B. subtilis* argF product is more similar to the *P. aeruginosa* argF product (47% identity) than it is to the arcB product (42% identity). Thus *B. subtilis* argF and *P. aeruginosa* argF appear to be orthologous genes, while the *P. aeruginosa* arcB gene is paralogous. The *E. coli* argF and argI genes (whose products share 87% identity) represent a comparatively recent duplication in the *E. coli* lineage (Van Vliet *et al.*, 1984) and both appear to be orthologues of *P. aeruginosa* arcB (with which they share 58% identity) rather than to argF (40% identity). Thus it appears that the *B. subtilis* argF and *E. coli* argF/argI are paralogues.

The extent of codon usage bias in the *B. subtilis* genes was estimated by the frequency of optimal codons (F_{op} , Table 1) as defined by Sharp *et al.* (1990). The values obtained are typical of moderately/lowly expressed genes in *B. subtilis*; the *E. coli* homologues have similarly moderately biased codon usage (data not shown).

ACKNOWLEDGEMENTS

This work was carried out as part of the CEC programme to sequence the entire genome of *B. subtilis*. The work was supported by EOLAS Scientific Grant SC/90/126 and by SCIENCE Grant SCL*0211-C(EDB). The authors would like to thank Pascale Serron (INRA, Jouy-en-Josas, France), Paul Sharp and Andrew Lloyd (Irish National Centre for Bioinformatics) for valuable assistance.

REFERENCES

- Baur, H., Stalon, V., Falmagne, P., Luethi, E. & Haas, D. (1987). Primary and quaternary structure of the catabolic ornithine carbamoyltransferase from *Pseudomonas aeruginosa*. *Eur J Biochem* **166**, 111–117.
- Cunin, R., Glansdorff, N., Pierard, A. & Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. *Microbiol Rev* **50**, 314–352.
- Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* **8**, 189–191.
- Itoh, Y., Soldati, L., Stalon, V., Falmagne, P., Terawaki, Y., Leisinger, T. & Haas, D. (1988). Anabolic ornithine carbamoyltransferase of *Pseudomonas aeruginosa*: nucleotide sequence and transcriptional control of the argF structural gene. *J Bacteriol* **170**, 2725–2734.
- Martin, P. R. & Mulks, M. H. (1992). Sequence analysis and complementation studies of the argJ gene encoding ornithine acetyltransferase from *Neisseria gonorrhoeae*. *J Bacteriol* **174**, 2694–2701.
- Mountain, A., Mann, N. H., Munton, R. N. & Baumberg, S. (1984). Cloning of a *Bacillus subtilis* restriction fragment complementing auxotrophic mutations in eight *Escherichia coli* genes of arginine biosynthesis. *Mol & Gen Genet* **197**, 82–89.
- Mountain, A., McChesney, J., Smith, M. C. M. & Baumberg, S. (1986). Gene sequence encoding early enzymes of arginine biosynthesis within a cluster in *Bacillus subtilis*, as revealed by cloning in *Escherichia coli*. *J Bacteriol* **165**, 1026–1028.
- Mountain, A., Smith, M. C. M. & Baumberg, S. (1990). Nucleotide sequence of the *Bacillus subtilis* argF gene encoding the ornithine carbamoyltransferase. *Nucleic Acids Res* **18**, 4594.
- Quinn, C. L., Stephenson, B. T. & Switzer, R. L. (1991). Functional organisation and nucleotide sequence of the *Bacillus subtilis* pyrimidine biosynthetic operon. *J Biol Chem* **266**, 9113–9127.
- Sakanyan, V., Kochikyan, A., Mett, I., Legrain, C., Charlier, D., Piérard, A. & Glansdorff, N. (1992). A re-examination of the pathway for ornithine biosynthesis in a thermophilic and two mesophilic *Bacillus* species. *J Gen Microbiol* **138**, 125–130.
- Sakanyan, V., Charlier, D., Legrain, C., Kochikyan, A., Mett, I., Piérard, A. & Glansdorff, N. (1993). Primary structure, partial purification and regulation of key enzymes of the acetyl cycle of arginine biosynthesis in *Bacillus stearothermophilus*: dual function of ornithine acetyltransferase. *J Gen Microbiol* **139**, 393–402.
- Sharp, P. M., Higgins, D. G., Shields, D. C., Devine, K. M. & Hoch, J. A. (1990). *Bacillus subtilis* gene sequences. In *Genetics and Biotechnology of Bacilli*, vol. III, pp. 89–98. Edited by M. M. Zukowski, A. T. Ganesan & J. A. Hoch. New York: Academic Press.
- Smith, M. C. M., Mountain, A. & Baumberg, S. (1990). Nucleotide sequence of the *Bacillus subtilis* argC gene encoding the N-acetylglutamate-gamma-semialdehyde dehydrogenase. *Nucleic Acids Res* **18**, 4595.
- Van Vliet, F., Cunin, R., Jacobs, A., Piette, J., Gigot, D., Lauwerreys, M., Piérard, A. & Glansdorff, N. (1984). Evolutionary divergence of genes for ornithine and aspartate carbamoyltransferases – complete sequence and mode of regulation of the *Escherichia coli* argF gene; comparison of argF with argI and pyrB. *Nucleic Acids Res* **12**, 6277–6289.

Received 1 November 1993; accepted 1 December 1993.