A 12 kb nucleotide sequence containing the alanine dehydrogenase gene at 279° on the *Bacillus subtilis* chromosome

Bauke Oudega,¹ Micheline Vandenbol² and Gregory Koningstein¹

Author for correspondence: Bauke Oudega. Tel: +31 20 4447177. Fax: +31 20 4447229.
e-mail: oudega@bio.vu.nl

¹ Department of Molecular Microbiology, Institute of Molecular Biological Sciences, BioCentrum Amsterdam, Faculty of Biology, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

² Department of Microbiology, Faculty of Agronomy, 6 Avenue du Maréchal Juin, B-5030 Gembloux, Belgium

In the framework of the European project aimed at the sequencing of the *Bacillus subtilis* genome, a DNA fragment of 12315 bp was cloned and sequenced. The DNA fragment is located between *rrnB* (275°) and *pai* (284°). Twelve ORFs were predicted to encode putative proteins. Two of these (*ald* and *yukI*) coincided with known *B. subtilis* genes. The products of two other genes (*yukK* and *yukL*) showed significant similarity to known proteins present in databases, e.g. pyoverdine synthase of *Pseudomonas aeruginosa* and pristinamycine synthase D of *Streptomyces pristinaespiralis*.

Keywords: *Bacillus subtilis*, genome sequencing, 279° region, *ald*

Within the framework of the international project for the sequencing of the *Bacillus subtilis* genome (Kunst *et al.*, 1995), we are responsible for the sequencing of an approximately 100 kb chromosomal region between *rrnB* (275°) and *pai* (284°). The *ald* gene of *B. subtilis* has been mapped in this region at about 279° (Anagnostopoulos *et al.*, 1993). The sequence of the *ald* gene and flanking regions (1709 bp) was retrieved from the *SubtiList* database (Moszer *et al.*, 1995). A large chromosomal DNA fragment surrounding the *ald* gene was obtained using inverse Long Range (LR) PCR. Chromosomal DNA of *B. subtilis* was digested with *NarI*. The *ald* sequence of 1709 bp does not contain a *NarI* site (Siranosian *et al.*, 1993). After dilution of the digestion mixture, DNA fragments were circularized by ligation. For LR PCR experiments the primers 5'-TCACTGACCGTTTCATATGCGAT 3' and 5'-TTA-CAGGATGAATCATCTGTGGC 3' and the Expand Long Template PCR System (Boehringer Mannheim) were used. Ten independent PCR runs were carried out, which yielded a 12 kb DNA fragment.

The 12 kb DNA fragment was sequenced using a random sequencing strategy. The DNA fragment was sonicated and fragments between 500 and 1000 bp in length were purified and treated with Klenow enzyme and T4 DNA polymerase (Sambrook *et al.*, 1989). Blunt-end fragments were then inserted into the *SmaI* site of pUC18 and amplified in *Escherichia coli* XL1-blue cells.

A total of 120 clones were selected and sequenced on dsDNA using [³⁵S]dATP and Sequenase. Synthesized oligonucleotide primers and a cycle sequencing protocol were used on the original LR PCR fragment to complete the sequence. Each strand was sequenced at least twice and regions of uncertainty were re-sequenced in both directions using synthesized primers.

The sequence was analysed for the location of ORFs. The following computer programs and packages were used: GeneWorks 2.3, DNA Strider 1.0, Sequencer 2.1 and the GCG package. Twelve putative ORFs were detected, based on the length of an ORF starting from ATG, GTG or TTG and preceded by a ribosome binding site (Fig. 1 and Table 1). The putative genes were named *yukA*–*F*, *ald* and *yukI*–*M*, according to the

Fig. 1. Organization of ORFs of the *B. subtilis* 279° region. The scale (in kb) starts at the *NarI* site upstream of *ald* and ends with a *NarI* site downstream of *ald*. Besides the *NarI* (N) sites, a unique *SacI* (S) site and a unique *XmaI/SmaI* (X) site are indicated above the scale. ORFs are indicated by thick black arrows. The arrowheads show the direction of transcription/translation. 'T' indicates terminator structures identified previously; 'P' and the mirror image indicate putative promoter regions. A stem-loop structure between *yukI* and *yukK* is shown below the arrows. The dashed arrows indicate putative transcriptional units.

Abbreviation: LR PCR, Long Range PCR.

The EMBL accession number for the nucleotide sequence reported in this paper is Z82015.

Table 1. Putative ORFs in the 12 kb sequence containing the *ald* gene (279°)

ORF	Endpoints (nt)	Size of product (aa)	SD consensus sequence (upper case) and initiation codon (bold)*
<i>yukA</i>	< 1844	614	tttctttGGcGaTGttatgatg
<i>yukB</i>	1844 < 2719	291	atgaAGcGcAGGtcacgcatg
<i>yukC</i>	2747 < 4102	451	ttatgaAAGGAtcTGgtgaaatg
<i>yukD</i>	4117 < 4356	79	ctatgtAAGGgrGaTcctaaatg
<i>yukE</i>	4500 < 4718	72	acAGGAGGtaagaggaatg
<i>yukF</i>	5239 > 6543	434	aataaaAGGAaGGaTGagatg
<i>ald</i>	6611 > 7747	378	tatacAGGAGGagacagatatg
<i>yukI</i>	7860 > 8183	107	atcataaAGGAGGaTtcataatg
<i>yukJ</i>	8101 > 8538	145	tAtGGAtaTacaagatcaatg
<i>yukK</i>	8807 < 10225	472	cggctctccAGGCAcccgatg
<i>yukL</i>	10303 < 11445	380	gccAGAGGaacacagcagcatg
<i>yukM</i>	11533 < 11748	71	AcGcAtccTttattcaaatcatg

* SD, Shine–Dalgarno.

Table 2. Similarity of predicted ORF products to known proteins

Product	Similar protein in database	Identity observed	BLAST score
Ald	L-Alanine dehydrogenase of <i>B. subtilis</i>	100 % over the entire length	–
YukI	YuxI, partial protein of <i>B. subtilis</i>	100 % in the N-terminal 70 residues	313
YukK	Pyoverdine synthetase D of <i>P. aeruginosa</i>	50 % in the C-terminal half	199
	Pristinamycin I synthase of <i>S. pristinaespiralis</i>	60 % in the C-terminal half	272
	Gramicidine S synthase 2 of <i>B. brevis</i>	50 % in the C-terminal half	222
YukL	Pristinamycine I synthase of <i>S. pristinaespiralis</i>	60 % over the entire length	444
	Pyoverdine synthase D of <i>P. aeruginosa</i>	60 % in the N-terminal half	625
	Siderophore biosynthase protein PbsC of <i>Pseudomonas</i> sp.	55 % over the entire length	419

rules agreed upon by the international consortium sequencing the *B. subtilis* genome. The initiation codon for all 12 ORFs was ATG. All putative genes had reasonable to good ribosome binding sites, except *yukM* which showed poor similarity to the consensus ribosome binding site. The direction of transcription and translation of three genes, *yukF* and *yukIJ*, was the same as that of *ald*, whereas that of eight other genes (*yukAB-CDE* and *yukKLM*) was in the opposite direction. The orientation of the fragment on the chromosome is not yet known since the 12 kb fragment has not been connected to other known markers.

The sequence of the 1709 bp *ald* fragment showed 44 errors as compared to the sequence of the 12315 bp fragment sequenced in this project. A total of 35 errors were located at the 5' end of the *ald* fragment, but did not affect the sequence of the *ald* gene itself. Four errors were located in the *ald*–*yukI* intergenic region and five were found in the 3' end of the 1709 bp *ald* fragment.

Comparison of the amino acid sequences of the putative gene products with protein databases [NCBI BLAST search (<http://www.ncbi.nlm.nih.gov>)] revealed that

only the sequences of the *ald* gene product and part of the sequence of YukI have been reported and characterized before (Siranosian *et al.*, 1993). The *ald* gene, also known as *spoVN*, encodes L-alanine dehydrogenase and is required for normal sporulation in *B. subtilis*. Its regulation is probably related to the metabolism of alanine (Siranosian *et al.*, 1993). The primary sequence of YukI, formerly known as YuxI, was only partly presented before. Only 65 aa of this gene product were reported; the C-terminal part was not found (Siranosian *et al.*, 1993). Our sequence data revealed that the total sequence of YukI contains 107 aa. The possible functions of two gene products, YukK and YukL, were predicted on the basis of a considerable degree of similarity to known sequences of other bacteria (Table 2). These results suggest that both YukK and YukL are involved in antibiotic biosynthesis. It should be noted that YukK and YukL are rather small as compared to the pyoverdine synthase and pristinamycine I synthase entries in the database and that the high degree of similarity is mainly restricted to a part of the proteins. YukK and YukL possibly form a part of a biosynthetic complex.

The other putative gene products showed no significant similarity to known products that could give information about their function in *B. subtilis*. The amino acid sequence of YukA is only partly known, the 3' end of the *yukA* gene is missing. The function of these products remains to be studied.

The genes *yukABCDE* might constitute one operon. No stem-loop structures, terminator structures or promoter regions were found between these genes. The start codon of *yukA* overlaps one base with the TGA stop codon of *yukB*. The TGA stop codon of *yukD* overlaps with the putative ribosome binding site of *yukC*. Upstream of *yukE* a structure resembling a promoter region [−10 (TATAAT) and −35 (TTGACA)] was found. This putative promoter region might be involved in transcription of the *yukABCDE* gene cluster. Another putative promoter region was found upstream of *yukF*. Downstream of this gene and in front of *ald* a possible terminator structure and a putative promoter region were predicted by Siranosian *et al.* (1993). These authors also predicted another terminator structure between *ald* and *yukI*. This structure might be involved in attenuation of the expression of *yukI* driven from the *ald* upstream promoter. A stem-loop structure was detected downstream of *yukJ*. *ald*, *yukI* and *yukJ* are transcribed in one direction and might constitute one operon. There is an 80 bp overlap between *yukI* and *yukJ*. Finally, *yukKLM* might also form one operon. Possible promoter regions were detected upstream of *yukM* and the stem-loop structure downstream of *yukK* might be a part of the terminator of this putative operon.

In conclusion the 12 kb DNA fragment sequenced might contain three transcriptional units, one of which might be involved in alanine metabolism and sporulation, and another one in antibiotic biosynthesis.

REFERENCES

- Anagnostopoulos, C., Piggot, P. J. & Hoch, J. A. (1993). The genetic map of *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 425–461. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Kunst, F., Vassarotti, A. & Danchin, A. (1995). Organization of the European *Bacillus subtilis* genome sequencing project. *Microbiology* **141**, 249–255.
- Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Moszer, I., Glaser, P. & Danchin, A. (1995). *SubtiList*: a relational database for the *Bacillus subtilis* genome. *Microbiology* **141**, 261–268.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Siranosian, K. J., Ireton, K. & Grossman, A. D. (1993). Alanine dehydrogenase (Ald) is required for normal sporulation in *Bacillus subtilis*. *J. Bacteriol* **175**, 6789–6796.

Received 7 November 1996; revised 27 December 1996; accepted 16 January 1997.