The *ampS*–*nprE* (124°–127°) region of the *Bacillus subtilis* 168 chromosome: sequencing of a 27 kb segment and identification of several genes in the area

Peggy Winters,^{1,2} Robert Caldwell,¹ Laura Enfield¹ and Eugenio Ferrari¹

Author for correspondence: Robert Caldwell. Tel: +1 415 846 5830. Fax: +1 415 845 6509.
e-mail: rcaldwell@genencor.com

¹ Genencor International, Inc., Department of Molecular Biology, 925 Page Mill Road, Palo Alto, CA 94304, USA

² Department of Biology, San Francisco State University, 1600 Holloway, San Francisco, CA 94132, USA

A stretch of DNA approximately 27 kb in length, adjacent to the *nprE* gene of *Bacillus subtilis*, has been sequenced. The sequenced fragment carries a total of 23 ORFs. Of these, 15 could be assigned probable functions based on homologies to characterized genes either in *B. subtilis* or in other organisms. The sequencing of this region has also allowed us to assign to this area *adeC* and *strB*, previously located on the other side of *nprE*, between *nprE* and the *pyr* operon.

Keywords: *Bacillus subtilis*, genome sequencing, mapping, *nprE*, *strB*

INTRODUCTION

Sequence analysis of whole bacterial genomes is generating a wealth of information that will positively impact a number of scientific areas. Some of the advantages and potential impacts of systematic sequencing, such as a better understanding of genome organization and evolution, identification of genes silent under laboratory conditions and reverse genetics approaches, have already been discussed (Glaser *et al.*, 1993; Sorokin *et al.*, 1993). This is true not only for the sequencing of poorly characterized microbes but also for the sequencing of bacteria such as *Bacillus subtilis* and *Escherichia coli* which are the paradigms of microbial genetics. The positive effects of this effort, however, will be felt beyond the areas of bacterial genetics and microbiology in general.

Systematic sequencing has already led to the serendipitous finding of genes whose presence was unsuspected in *B. subtilis*, for example polyketide synthases (Scotti *et al.*, 1993). One added advantage in the case of *B. subtilis* is that members of the *Bacillus* family play an essential role in the expression and secretion of industrial enzymes. In fact more than 50 % of enzymes with industrial applications are produced by *Bacillus* spp. The availability of the blueprint of *B. subtilis* could allow further development of this bacterium as a general host for enzyme secretion. Furthermore, the identification of new genes and bio-

chemical pathways will very likely expand the utilization of bacteria for other industrial applications beside enzyme production. Knowledge of how bacterial regulatory networks function will allow a more rational approach to modifying their biochemical pathways.

METHODS

Bacterial strains and plasmids. *Bacillus subtilis* 168 (*trpC2*) was the source of all chromosomal DNA used in this study. *B. subtilis* SR 265 (*strB47*), JH 503 (*strB3 ura*; Staal & Hoch, 1972) and 1A237 (*ura3 fruB*; Bacillus Genetic Stock Center, Ohio, USA) were used for genetic mapping. BG2872 (*trpC2* Cm^R) a derivative of *B. subtilis* 168 carrying pJM103 integrated approximately 1 kb downstream of the *ampS* gene, was also used for mapping studies. Subclones for sequencing were in plasmids pGEM7 (for nested deletion set constructions; Promega, used according to supplier's protocols) and pUC18/19 or pJM103 (Perego, 1993). pJM103 chromosomal integrants were used for DNA walking and for generating transformation and/or transduction mapping data. Plasmid constructions were transformed into *E. coli* strain PY1182 (MM294 *pcnB180* *zad::Tn10*), made competent by the CaCl₂ method (Sambrook *et al.*, 1989) and selected on LB agar at 50 µg carbenicillin ml⁻¹.

Genetic manipulations. *Bacillus* competent cells were prepared by the method of Anagnostopoulos & Spizizen (1961). PBS1 transductions were performed according to Hoch *et al.* (1967). Selection of chloramphenicol-resistant transformants and transductants in *B. subtilis* 168 was carried out on LB agar supplemented with 5 µg chloramphenicol ml⁻¹. Streptomycin resistance was screened for as described by Staal & Hoch (1972).

DNA manipulations. *B. subtilis* 168 cells were grown to mid-exponential phase in LB medium, harvested, lysed with hen egg

The GenBank accession number for the nucleotide sequence reported in this paper is U51911.

Fig. 1. Overview of the area sequenced. The physical map, showing all the ORFs in the 26.96 kb contig, is expanded relative to the genetic map of the region. Unique *SalI* and *BamHI* restriction sites used in cloning steps are indicated.

white lysozyme and chromosomal DNA extracted as described by Marmur (1961). Plasmids were purified from PY1182 transformants using Qiagen spin columns. Ligations, restriction digestions and Southern blots were performed by standard procedures (Sambrook *et al.*, 1989). Restriction enzymes used in this work were purchased from Boehringer Mannheim and/or New England Biolabs.

DNA sequencing. All DNA sequences were determined on a Model 373A DNA Sequencer (Applied Biosystems) employing either standard *lac* and -48 reverse or custom synthesized oligonucleotides (Gibco BRL). Both the *Taq* Dye Terminator Cycle Sequencing Kit and the *Taq* Dye Primer Cycle Sequencing Kit with fluorescently labelled standard primers were used (Applied Biosystems). Sequences were determined on both strands, with a minimum redundancy of two on each strand, except for the previously reported *nprE* (Yang *et al.*, 1984) and *pdh-slp* regions (Hemilä *et al.*, 1990; Hemilä, 1991) where one strand only was determined to verify deposited sequences. In these regions, any conflicts found in one strand were then confirmed by second strand sequencing. Contiguity of separately rescued DNA sequences was confirmed either by Southern analysis or by sequencing across junctions in PCR products generated from genomic DNA using flanking oligonucleotides.

Data analysis. DNA sequences were compiled and assembled using the SEQMAN program from DNASTAR. Potential coding regions were determined using the ORF search feature of DNASTAR's EDITSEQ program on both strands with ATG, TTG, GTG and CTG selected as potential start codons and a minimal ORF size of 25 codons. Any putative ORFs identified

in this way were subsequently examined for the presence of potential Shine-Dalgarno (SD) sequences at appropriate distances upstream. Translations of ORFs identified as described above were then compared to known protein sequences in SwissProt using FASTA and BLAST programs, or to GenBank/EMBL DNA sequences using TFASTA in the Genetics Computer Group sequence analysis software programs.

RESULTS AND DISCUSSION

Cloning and sequencing

Clones covering the *ampS*–*nprE* region were isolated by rescue of contiguous sequences from *B. subtilis* 168 carrying pJM103 subclones integrated via homologous recombination. The first fragment was isolated by digesting with *Bam*HI chromosomal DNA extracted from a derivative of strain 168 carrying a plasmid integrated at the 3' flank of the *nprE* gene. The rescued fragment was approximately 10 kb long and reached the unique *Bam*HI site in the *pdhA* gene. The second walk, in which the chromosomal DNA of a 168 derivative carrying a plasmid integrated at the 3' *pdhA*/5' *pdhB* junction was digested with *Sal*I, allowed the rescue of an additional 14 kb, to the middle of the *ampS* gene. The 3' end of the *ampS* gene was isolated by an inverse PCR walk using universal plasmid-based sequence primers designed for Long Range PCR from a pJM103 subclone integrated at the 5' end of the *ampS* gene. An additional pJM103 subclone of the 3' end

Table 1. Features of putative ORFs and known genes in a 27 kb contiguous region between *ampS* (124°) and *nprE* (127°) on the *B. subtilis* chromosome

ORF	SubtiList database name	Size of product (aa)	SD consensus sequence (upper case) and initiation codon (bold)
<i>ampS</i>	<i>ampS</i>	410	AGcttGGAGGaatgattc ATG
<i>ykpC</i>		44	gttAAGGAGGaaacag ATG
<i>mreBH</i>	<i>mreBH</i>	335	AGgAtaGAGaTtggGtgaaaac ATG
<i>abb</i>	<i>ylxT</i>	92	ttAAAGGAaGaaggGttttt ATG
<i>kinC</i>	<i>kinC</i>	428	AagcgGcAGGTGgTcat ATG
<i>ykqA</i>	<i>ylxU</i>	277	tGAgAGGAGaaaAataaa GTG
<i>ykqB</i>	<i>ylxV</i>	221	AtAAAGGAGcaccacg ATG
<i>adeC</i>	<i>adeC</i>	577	AaAAAGGAGGgGcctat TTG
<i>ykqC</i>		555	gtAttGGAGtTatgagcggct ATG
<i>ykrA</i>		257	AtAAAGGgGaaGtTGcg ATG
<i>ykrB</i>		184	ctttAGGAGtgaAgtttc TTG
<i>ykrC</i>	<i>ykyA</i>	205	AGAtcGGAGcTttTtgtgcc ATG
<i>pdbA</i>	<i>pdbA</i>	371	AGAAAGGAaGaGgTGacttagt ATG
<i>pdbB</i>	<i>pdbB</i>	325	cagAAGGAGtcGAaGtaagcc ATG
<i>pdbC</i>	<i>pdbC</i>	442	AactAGGAGGTcgaGaact GTG
<i>pdbD</i>	<i>pdbD</i>	470	ttAAtGGAGGcGtaatgtt ATG
<i>slp</i>	<i>slp</i>	124	AcAgAGGAGaaGATag ATG
<i>cad</i>	<i>cad</i>	490	AcAAAGGtGGgaAaaca TTG
<i>yktA</i>		88	tacAAGGAGtgaAgaaa ATG
<i>yktB</i>		212	AtgAAGGAGGaaAatc ATG
<i>yktC</i>		265	AGAAgaGAGGTacgaacga ATG
<i>yktD</i>		304	tttgtGGAGGatcatttc ATG
<i>nprE</i>	<i>nprE</i>	521	AaAAAGGgGGattTatt GTG
3' 16S rRNA (reverse complement)			AGAAAGGAGGTGATG

of the *nprE* gene was used to rescue the neutral protease gene from strain 168 to confirm previously published sequence data (Yang *et al.*, 1984). To confirm the direction of these walking steps, plasmids carrying sub-fragments from each of these walks were integrated and mapped with respect to *metC*, *fruB* and *ura*.

Gene sequence analysis

The sequences generated revealed the presence of 23 ORFs, 13 of which are transcribed in the direction of replication of the chromosome (Fig. 1). Out of these 23 putative genes, 17 utilize ATG, 3 use GTG, 3 TTG and none use CTG as start codons (Table 1). Twelve of the 23 could be assigned functions based on identity to previously reported sequences (Table 2). Comparison of the translations of the other 11 ORFs with both SwissProt and GenBank databases revealed significant homologies to proteins with known functions for only three of them: *ykpC*, *ykrB* and *yktC*. Of the remaining eight ORFs, three are identical to *B. subtilis* genes of unknown function, one is highly homologous to an unknown function ORF of *B. stearothermophilus* and four show no significant homologies to sequences in the databases.

Two relatively large contigs of this region have been sequenced previously. A few discrepancies were found in

the *pdb* operon between our sequence and the published sequence (Hemilä *et al.*, 1990), most likely due to differences in the strains used as the source of DNA, while no differences were found with the contig centred around the *kinC* gene as reported by Kobayashi *et al.* (1995).

One intriguing finding is the presence of a short ORF, *ykpC*, encoding a 44 aa peptide which shows homology to and aligns exactly with the N terminus of *B. subtilis* OppB (Perego *et al.*, 1991). In spite of the length of the sequence, we believe that this is a true gene (or a pseudogene) based on the presence of a strong consensus SD sequence followed by a start codon, at the canonical distance, and the presence, at the 3' end, of a 35 bp long inverted repeat capable of forming a stem-loop structure with a ΔG of -14.0 kcal mol⁻¹. To eliminate the possibility of an artifact due to rearrangement during the cloning steps, we rescued the same area by direct PCR from strain 168 chromosomal DNA using primers flanking the *ykpC* gene and obtained a fragment of the expected size.

Genetic analysis and mapping

The presence of the *adeC* gene in the region being sequenced was surprising. This gene, in fact, had been mapped outside this area, between *nprE* and the *pyr* operon, based on its linkage to the *strB* gene (Nygaard *et*

Table 2. Similarity of predicted ORF products to known gene products

ORF	Similar protein in database	Database accession no.*	Match (%)†
<i>ampS</i>	Aminopeptidase AmpS, <i>B. subtilis</i> (partial)	GB: D37799	100 (291)
<i>ykpC</i>	Oligopeptide transport system Permease protein OppB, <i>B. subtilis</i>	SP: P24138	29 (41)
<i>mreBH</i>	Rod shape determining protein MreBH, <i>B. subtilis</i>	GB: D37799	100 (335)
<i>abb</i>	Product of <i>orf1</i> , <i>B. subtilis</i>	GB: D37799	100 (92)
	Transition state regulator AbrB, <i>B. subtilis</i>	GB: X12820	58 (92)
<i>kinC</i>	Histidine protein kinase KinC, <i>B. subtilis</i>	GB: D37799	100 (428)
<i>ykgA</i>	Product of <i>orf3</i> , <i>B. subtilis</i>	GB: D37799	100 (277)
<i>ykgB</i>	Product of <i>orf4</i> , <i>B. subtilis</i>	GB: D37799	100 (221)
<i>adeC</i>	Adenine deaminase, <i>B. subtilis</i>	GB: X83795	100 (577)
<i>ykgC</i>	None		
<i>ykrA</i>	Product of <i>orf1</i> , <i>B. stearothermophilus</i> , located in <i>pdbA</i> 5' region (partial, C terminus)	SP: P21878	66 (94)
<i>ykrB</i>	Polypeptide deformylase Def, <i>Mycoplasma genitalium</i>	SP: P47352	37 (173)
<i>ykrC</i>	Product of <i>ykyA</i> , <i>B. subtilis</i>	SP: P21884	99.5 (205)
<i>pdhA</i>	Pyruvate decarboxylase E-1 α subunit, <i>B. subtilis</i>	GB: M57435	99.7 (371)
<i>pdhB</i>	Pyruvate decarboxylase E-1 β subunit, <i>B. subtilis</i>	GB: M57435	100 (325)
<i>pdhC</i>	Dihydrolipoamide acetyltransferase E2, <i>B. subtilis</i>	GB: M57435	100 (442)
<i>pdhD</i>	Dihydrolipoamide dehydrogenase E3, <i>B. subtilis</i>	GB: M57435	100 (470)
<i>slp</i>	PAL-related lipoprotein, <i>B. subtilis</i>	GB: M57435	100 (124)
<i>cad</i>	Lysine decarboxylase, <i>B. subtilis</i>	GB: X58433	100 (490)
<i>yktA</i>	None		
<i>yktB</i>	None		
<i>yktC</i>	Extragenic suppressor protein SuhB, <i>E. coli</i>	SP: P22783	33 (253)
	<i>myo</i> -Inositol-1(or 4)-monophosphatase, <i>Homo sapiens</i>	SP: P29218	34 (253)
<i>yktD</i>	None		
<i>nprE</i>	Neutral protease NprE, <i>B. subtilis</i>	GB: K01985	100 (521)

* SP, SwissProt; GB, GenBank.

† The number of amino acids over which the percentage match was determined is shown in parentheses.

al., 1988). While we did not generate mapping data directly confirming the presence of *adeC* in this region of the chromosome, the transformation and transduction results, carried out throughout the project, give us confidence in this placement. Furthermore, genetic transformation data of strain SR265, carrying the *strB47* mutation (Staal & Hoch, 1972) with respect to a chloramphenicol-resistance plasmid (pLE1.3) integrated 1.3 kb downstream of the *ampS* gene (strain BG2872), indicates that *strB* is 80% linked to *ampS* (data not shown). Likewise, PBS1 transduction data generated using a 168 strain carrying pLE1.3 as donor and JH503 (*ura strB3*) as recipient confirms the very tight link between *strB3* and chloramphenicol-resistance. To date, although genetic mapping data places the *strB* gene in this region, we have not yet identified the ORF responsible for this phenotype. Work toward this identification is in progress.

Because of the presence of many ORFs without any assignable function, we prepared plasmids carrying deletions in the regions between *nprE* and *cad* as well as between *pdhA* and *kinC* in *B. subtilis* 168. The deletion of the *nprE*–*cad* fragment was easily obtained, and the deleted

strain has been analysed for a number of phenotypes, including sporulation deficiency, auxotrophic requirements other than tryptophan, temperature sensitivity and protease deficiency. No differences were observed in any of the phenotypes analysed. The most obvious conclusion is that this deleted region carries genes whose functions are not easily revealed under laboratory growth conditions. Alternatively, additional mutations may be needed to detect a possible phenotype. Work to delete the *pdh*–*kinC* region is still in progress.

ACKNOWLEDGEMENTS

We would like to thank Dr James A. Hoch for providing the streptomycin resistance strains.

REFERENCES

- Anagnostopoulos, C. & Spizizen, J. (1961). Requirements for transformation of *Bacillus subtilis*. *J. Bacteriol.* **81**, 741–746.
- Glaser, P., Kunst, F., Arnaud, M., Coudart, M.-P., Gonzales, W., Hullo, M.-F., Ionescu, M., Lubochinsky, B., Marcelino, L., Moszer, I., Presecan, E., Santana, M., Schneider, E., Schweizer, J., Vertes, A., Rapoport, G. & Danchin, A. (1993). *Bacillus subtilis* genome

project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol Microbiol* **10**, 371–384.

Hemilä, H. (1991). Sequence of a PAL-related lipoprotein from *Bacillus subtilis*. *FEMS Microbiol Lett* **82**, 37–42.

Hemilä, H., Palva, A., Paulin, L., Arvidson, S. & Palva, I. (1990). Secretory S complex of *Bacillus subtilis*: sequence analysis and identity to pyruvate dehydrogenase. *J Bacteriol* **172**, 5052–5063.

Hoch, J. A., Barat, M. & Anagnostopoulos, C. (1967). Transformation and transduction in recombination-defective mutants of *Bacillus subtilis*. *J Bacteriol* **93**, 1925–1937.

Kobayashi, K., Shoji, K., Shimizu, T., Nakano, K., Sato, T. & Kobayashi, Y. (1995). Analysis of a suppressor mutation *ssb* (*kinC*) of *surOB20* (*spo0A*) mutation in *Bacillus subtilis* reveals that *kinC* encodes a histidine protein kinase. *J Bacteriol* **177**, 176–182.

Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *J Mol Biol* **3**, 208–218.

Nygaard, P., Duckert, P. & Saxild, H. H. (1988). Purine gene organization and regulation in *Bacillus subtilis*. In *Genetics and Biotechnology of Bacilli*, pp. 57–61. Edited by A. T. Ganesan & J. A. Hoch. San Diego: Academic Press.

Perego, M. (1993). Integrational vectors for genetic manipulations in *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 615–624. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.

Perego, M., Higgins, C. F., Pearce, S. R., Gallagher, M. P. & Hoch, J. A. (1991). The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. *Mol Microbiol* **5**, 173–185.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Scotti, C., Piatti, M., Cuzzoni, A., Perani, P., Tognoni, A., Grandi, G., Galizzi, A. & Albertini, A. M. (1993). A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases. *Gene* **130**, 65–71.

Sorokin, A., Zumstein, E., Azevedo, V., Ehrlich S. D. & Serron, P. (1993). The organization of the *Bacillus subtilis* 168 chromosome region between the *spoVA* and *serA* genetic loci, based on sequence data. *Mol Microbiol* **10**, 385–395.

Staal, S. P. & Hoch, J. A. (1972). Conditional dihydrostreptomycin resistance in *Bacillus subtilis*. *J Bacteriol* **110**, 202–207.

Yang, M. Y., Ferrari, E. & Henner, D. J. (1984). Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an *in vitro*-derived deletion mutation. *J Bacteriol* **160**, 15–21.

Received 18 April 1996; revised 29 June 1996; accepted 8 July 1996.