FEMS Microbiology Letters 115 (1994) 273–278

Published by Elsevier

FEMSLE 05782

Identification and sequencing of *pyrG*, the CTP synthetase gene of *Azospirillum brasilense* Sp7

Wolfgang Zimmer * and Barbara Hundeshagen

Botanisches Institut der Universität zu Köln, Gyrhofstraße 15, D-50923 Köln, FRG

(Received 19 October 1993; revision received 5 November 1993; accepted 8 November 1993)

Abstract: An 18.5-kb DNA fragment carrying the *trpGDC* cluster of *Azospirillum brasilense* Sp7 was previously cloned, yielding cosmid pAB1005. Attempts to identify *trpA* in the vicinity of *trpGDC* failed but led to the detection of a locus strongly homologous to *pyrG*, the structural gene for the CTP synthetase. The function of the *A. brasilense pyrG* gene was verified by complementation of the cytidine-requiring PyrG-deficient mutant JF646 of *Escherichia coli*. A second open reading frame was identified downstream of *pyrG*. The deduced amino acid sequence showed homology to dienelactone hydrolases of *Pseudomonas* and *Alcaligenes*, enzymes involved in utilization of halogenated aromatic compounds.

Key words: CTP synthetase gene; *pyrG*; Dienelactone hydrolase gene homologue; *Azospirillum brasilense*

Introduction

Cosmid pAB1005 of a gene library of the root-associated bacterium *Azospirillum brasilense* Sp7 [1] had led to the interesting phenotype of an enhanced tryptophan dependent indole-3-acetic acid production after transfer into *A. irakense* KA3 [2,3]. This phenotype was due to *trpD*, which was identified as a part of the *trpGDC* gene cluster on the 18.5-kb insert of pAB1005 [2,3]. Thus, a regulatory coupling of the indole-3-acetic acid- and tryptophan-synthesis was proposed [3]. As the organization of *trpG*, *trpD* and *trpC* in *Azospirillum* was similar to that found in *Pseudomonas* [4], where other Trp-genes are located near the *trpGDC* cluster [5], we tried to identify such genes in the neighborhood of the *trpGDC* cluster carried by pAB1005.

Materials and Methods

Strains, plasmids, bacteriophages, growth conditions

Derivatives of bacteriophage M13 were propagated in *Escherichia coli* TG1 (see [6]). The cytidine and uracil requiring PyrG$^-$ strain JF646 of

* Corresponding author. Tel.: (0221) 470 2480; Fax: (0221) 470 5181; e-mail: WZIMMER@BIOLAN.UNI-KOELN.DE.

SSDI 0378-1097(93)E0474-Q

E. coli [7] was kindly provided by James Friesen, University of Toronto, Canada and Howard Zalkin, Purdue University, USA. *E. coli*, carrying plasmid pAB1005 or its Tn*5* derivatives [3], were grown on Luria-Bertani (LB) medium [6] with tetracycline, 5 μg ml^{-1}.

Oligonucleotides

The Tn*5* oligonucleotide was a sequence complementary to the first 30 bases of the Tn*5* repeats: 5′ TTC AGG ACG CTA CTT GTG TAT AAG AGT CAG 3′. The *trpA* oligonucleotide corresponded to a highly conserved region in TrpA polypeptides [8] at amino acid positions 49–58 and was adapted to the codon usage in *Azospirillum*: 5′ CTG GAG CTG GGC GTC CCC TTC TCC GAC CCG 3′. The oligonucleotides, synthesized by a Pharmacia LKB Gene Assembler Plus, were purified by separation on a 1.5 ml NAP^{TM}-10 column (Pharmacia LKB, Uppsala, Sweden).

Polymerase chain reaction (PCR)

The reaction was performed in 30 circles (30 s denaturation at 92°C, 30 s annealing at 57°C, 60 s polymerization at 72°C) in a volume of 50 μl containing: Taq-polymerase (Boehringer Mannheim, FRG), 2 U; 10 × Taq-polymerase buffer (Boehringer Mannheim), 5 μl; deoxynucleotides, 2.5 nmol of each; oligomers, 50 pmol of each; DNA of Tn*5*-derivatives of pAB1005, 1 ng.

Plasmid and phage construction

The amplified PCR fragments were directly cloned into M13mp18 as recently described [9]. Plasmid pAB10055 (see Fig. 1) was constructed ligating the 4.2-kb *Hin*dIII–*Eco*RI fragment from pAB1005 in the polylinker of pBluescript. Transformation, DNA isolation and restriction analysis were performed by conventional techniques [6].

DNA sequencing

Overlapping restriction fragments from the 2.6-kb *Sph*I–*Eco*RI region of pAB1005 were cloned into M13mp19 and M13mp18 to facilitate sequencing of both DNA strands. The sequences on both strands of the 2.6-kb region and of the PCR fragments in M13mp18 were determined using the TAQuence^{TM} Sequencing kit (United States Biochemical Corporation, Cleveland, OH).

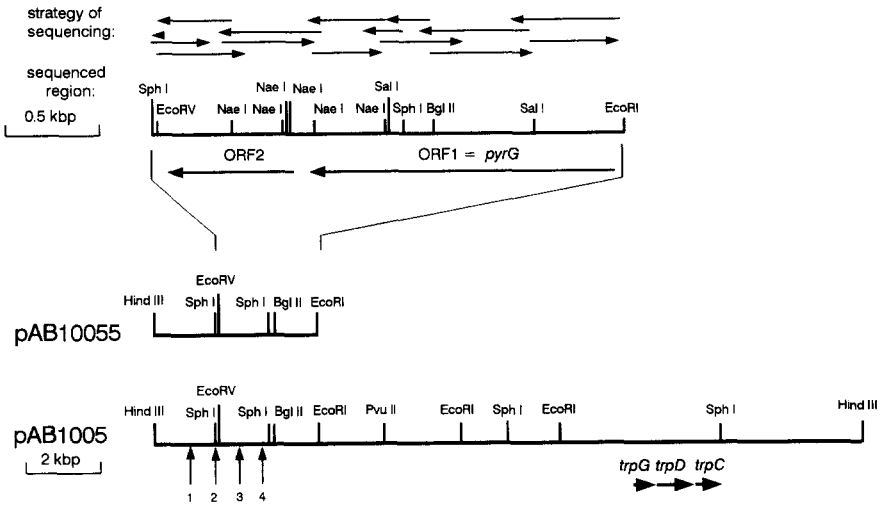

Fig. 1. Location of the Tn*5* insertions, the amplified PCR fragments and the sequenced region on the physical map of pAB1005 and pAB10055. The Tn*5* insertions are indicated by the vertical arrows below the restriction map of pAB1005. The positions of the previously sequenced *trpGDC* [3] and of the identified open reading frames *pyrG* and ORF2 are shown by horizontal arrows. Above the enlarged map of the *Sph*I–*Eco*RI fragment, the sequencing strategy is indicated by horizontal arrows.

Results and Discussion

Screening of pAB1005 for the presence of trpA in the vicinity of trpGDC

As a set of pAB1005 derivatives was available, each of them containing a Tn*5* insertion at a different position in the insert [3] (Fig. 1), there was the possibility to localize additional genes by the use of only one gene specific primer for PCR amplification. Thus, a series of PCR experiments was performed with these pAB1005-Tn*5* derivatives using an oligomer for a highly conserved region of *trpA* [8] and an oligomer complementary to the terminal part of the Tn*5* repeats. The PCR led to the amplification of fragments from pAB1005-Tn*5* derivatives 1, 2 and 3. The decreasing sizes of the amplified fragments of Tn*5* insertions 1, 2 and 3 of 1.7 kb, 1.0 kb and 0.5 kb (not shown), respectively, suggested that the *trpA* oligonucleotide had annealed between Tn*5* insertion 3 and 4 (Fig. 1).

Cloning of the PCR fragments into M13mp18 and sequencing of the inserts

The fragments amplified from Tn*5* derivatives 2 and 3, were directly cloned into M13mp18 [9]. For each fragment, a pair of clones was chosen that contained the insert in opposite direction. As expected, sequencing of the inserts revealed that the *trpA* oligomer had primed at the same position between Tn*5* insertion 3 and 4 in case of the PCR fragments obtained with the Tn*5* derivatives

```
Ab ORF1   MT-RYIFITGGVVSSLGKGLASAALGALLQARGYKVRLAKLDPYLNVDPGTMSPYQHGEVYVTDDGAETDLDLGHYERFT
Ec PyrG   MTTNYIFVTGGVVSSLGKGIAAASLAAILEARGLNVTIMKLDPYINVDPGTMSPIQHGEVFVTEDGAETDLDLGHYERFI
Bs PyrG   MT-KYIFVTGGVVSSLGKGIVAASLGRLLKNRGLNVTIQKFDPYINVDPGTMSPYQHGEVFVTDDGAETDLDLGHYERFI
Hs PyrG   M--KYILVTGGVISGIGKGIIASSVGTILKSCGLHVTSIKIDPYINIDAGTFSPYEHGEVFVLDDGGEVDLDLGNYERFL
          *   **++*****+*++***+ +++++ +*   *  *   *+****+*+*+** ** +*****+* +***+* *****+****

GVAARRGDNITTGRIYSNVIAKERRGDYLGATVQVIPHVTDQIKD--FIGA------ETTDEDFILCEIGGTVGDIESTPFLEAIRQFGN
RTKMSRRNNFTTGRIYSDVLRKERRGDYLGATVQVIPHITNAIKERVLEGG------EGH--DVVLVEIGGTVGDIESLPFLEAIRQMAV
DINLNKFSNVTTGKIYSTVLKKERRGDYLGGTVQVIPHITNELKDRVYRAG------KETNADVVITEIGGTVGDIESLPFLEAIRQMKS
DIRLTKDNNLTTGKIYQYVINKERKGDYLGKTVQVVPHITDAIQEWVMRQALIPVDEDGLEPQVCVIELGGTVGDIESMPFIEAFRQFQF
     +  * ***+**  *+ ***+***** *****+**+*+ +++     +           +  + *+********** **+**+**

EVGPENALFIHLTLLPYIPTAGELKTKPTQHSVKELLGMGIQANILLCRADRPIPENERKKIALFCNIRPERVIAALDVDSIYQVPVSYH
EIGREHTLFMHLTLVPYMAASGEVKTKPTQHSVKELLSIGIQPDILICRSDRAVPANERAKIALFCNVPEKAVISLKDVDSIYKIPGLLK
DIGRENVMYIHCTLVPYIKAAGELKTKPTQHSVKELRSLGIQPNIIVVRTEMPISQDMKDKIALFCDIDTKAVIECEDADNLYSIPLELQ
KVKRENFCNIHVSLVPQPSSTGEQKTKPTQNSVRELRGLGLSPDLVVCRCSNPLDTSVKEKISMFCHVEPEQVICVHDVSSIYRVPLLLE
 +  *+   +* +*+*   ++** ******+**+** ++*+ +++++ *   ++    + **++**++    **   *  +** +*

EEGFDTQVLAYFGLPTEGKPD--LSRWTSIVERVRKPQGEVTIAVVGKYTSLLDSYKSLAEALTHGGIANNVKVKLDWIDSEIFEDESAV
SQGLDDYICKRFSLNCPE-AN--LSEWEQVIFEEANPVSEVTIGMVGKYIELPDAYKSVIEALKHGGLKNRVSVNIKLIDSQDVETRGLE
KQGLDKLVCEHMKLACKE-AE--MSEWKELVNKVSNLSQTITIGLVGKYVELPDAYISVVESLRHAGYAFDTDVKVKWINAEEVTENNIA
EQGVVDYFLRRLDLPIERQPRKMLMKWKEMADRYDRLLETCSIALVAKYTEFSDSYASVIKALEHSALAINHKLEIKYIDSADLEPITSQ
 +*          *     +   +  *              + +*++*+**  + *+* *+  +* *++       + +  *++

Q----------RLENVHGILVPGGFGSRGTEGKIRAAQFARERKVPYFGICFGMQMAVIESARNMAGIVDAGSTELGKPGN-PVVGLLGL
I----------LK-GLDAILVPGGFGYRGVEGMITTARFARENNIPYLGICLGMQVALIDYARHVANMENANSTEFVPDCKYPVVALI--
E----------LTSGTDGIIVPGGFGDRGVEGKIVATKYARENNIPFLGICLGMQVASIEYARNVLGLKGAHSAEIDPSTQYPIIDLL--
EEPVRYHEAWQKLCSAHGVLVPGGFGVRGTEGKIQAIAWARNQKKPFLGVCLGMQLAVVEFSRNVLGWQDANSTEFDPTTSHPVVVDMP-
                +++++*****  **  **  * +   **+ + *+++*+****+* ++ +*++      * *+*+       *++  +

MTEWMRGNS-LEKRTEGTDVGGTMRLGTYPAKLVP-GSKVAEVYGTTD-ITERHRHRYEVNVYYKDRLEKVGLLFSGLSPT-QLPEIVEI
-TEWRDENGNVEVRSEKSDLGGTMRLGAQQCQLVD-DSLVRQLYNAPT-IVERHRHRYEVNNSLLKQIEDAGLRVRARSGDDQLVEIIEV
-PE----------QKDVEDLGGTLRLGLYPCKLEE-GTKAFEVYQDEV-VYERHRHRYEFNNEFRQQMEEQGFVFSGTSPDGRLVEIIEL
------------EHNPGQMGGTMRLGKRRTLFQTKNSVMRKLYGDADYLEERHRHRFEVNPVWKKCLEEQGLKFVGQDVEGERMEIVEL
            +  ++****+***     +    +    +*     + *******+* *      +*  *+   +         **+*+

PDHPWFIGVQFHPELKSKPFDPHPLFTSFIKAAIEQSRLV
PNHPWFVACQFHPEFTSTPRDGHPLFAGFVKAASEFQKRQAK
KDHPWFVASQFHPEFKSRPTRPQPLFKGFIGASVEAANQK
EDHPFFVGVQYHPEFLSRPIKPSPPYFGLLLASVGRLSHYLQKGCRLSPRDTYSDRSGSSSPDSEITELKFPSINHD
 +** *++ *+****+ * *    * + +++ *+
```

Fig. 2. Comparison of deduced ORF1 and PyrG amino acid sequences. Sequences are from *Azospirillum brasilense* (Ab ORF1, this work), *Escherichia coli* (Ec PyrG [14]), *Bacillus subtilis* (Bs PyrG [15]) and *Homo sapiens* (Hs PyrG [16]). Identical residues in all four sequences are indicated by an asterisk and homologous residues by a plus below the sequences. The regions of the conserved glutamine amide transfer domain [14] are underlined.

2 and 3. The sequence of the fragments derived from the pAB1005-Tn*5* insertions 2 and 3 did not show any similarity to *trpA* sequences beside the *trpA* oligonucleotide. However, the deduced amino acid sequence downstream of the *trpA* oligomer showed similarities to the C-terminal end of CTP synthetases of *E. coli* and *Bacillus subtilis*.

Sequencing of the locus and identification of two open reading frames

The nucleotide sequence of a 2.6-kb *Sph*I–*Eco*RI fragment of pAB1005 (Fig. 1) was established (EMBL, GenBank and DDBJ Nucleotide Sequence Databases, accession number X67216). Two open reading frames (ORF) were identified with a high G + C content (93%, 87%) in the third codon positions typical for *Azospirillum* genes (see [10]). The first ORF (ORF1) with the start codon at position 43 and the stop codon at position 1676 corresponded to a polypeptide of 544 amino acids (56.3 kDa; Fig. 2). It was followed by two inverted repeats (position 1689–1711 and 1750–1774) which can form a hairpin loop with a ΔG (25°C) of -27.6 kcal mol^{-1} and -30.6 kcal mol^{-1}, respectively, after transcription [11]. The latter repeat showed four uridine residues at the 3′ end of the corresponding RNA, typical for a rho-independent termination site of transcription in prokaryotes [12]. The second ORF (ORF2, start codon at position 1820, stop codon at 2513) possibly encoded a polypeptide of 231 amino acids (23.9 kDa; Fig. 3). Putative ribosome binding sites [13] ('GAAGG', 'AGAGG') were located upstream of the start codons at position 26 and 1804, respectively. Comparison of the sequence of the PCR fragment with the 2.6-kb sequenced fragment revealed that the *trpA* oligonucleotide had annealed inside the open reading frame at position 1591–1621. The Tn*5* insertions 2 and 3 were at position 2070 and 2588. Seven nucleotides at the 3′ end and eight nucleotides near the 5′ end of the *trpA* oligonucleotide were identical to ORF1.

Similarity of the amino acid sequence coded by ORF1 with CTP synthetases

The deduced amino acid sequence encoded by the ORF1 showed 55%, 51% and 45% identical residues with sequences of CTP synthetases coded by *pyrG* in *E. coli* [14], *Bacillus subtilis* [15] and *Homo sapiens* [16], respectively (Fig. 2). These enzymes catalyze the reaction: UTP + ATP + glutamine ⇒ CTP + ADP + P_i + glutamate. As in the case of the other CTP synthetases, the C-terminus of the deduced amino acid sequence of ORF1 showed similarity to other glutamine amide transferases, TrpG and PabA. In particular, the regions of the three conserved segments of the glutamine amide transfer domain [14] were present (Fig. 2).

Similarity of the amino acid sequence coded by ORF2 with dienelactone hydrolases

The deduced amino acid sequence of ORF2 showed similarity with dienelactone hydrolases

```
Ab ORF2   M--TDISIPAGDDGSFSAYVAKPAGGGPAPGLVVIQEIFGVNQVMRDLCDAFAAQGWLAVCPDLFWRQEPGVQITDKTQE
Pp ClcD   MLTEGISIQSYDGHTFGALVGSPAKA-PAPVIVIAQEIFGVNAFMRETVSWLVDQGYAAVCPDLYARQAPGTALDPQDER
Ps TcbE   MLTEGLSIDAKGGGRFGAHLQLPARG-RGPVVMVAQEIFGVNPFMTEVLAWLASEGFVGLCPDLYWRHGPGIEFDPNDEV
Ae TfdE   MLSDGVEITSRSGGRFGAYLGKPTTD-SAPIVVIAQEIFGITPFIRETVEWLVGAGFGCVCPDLYWRQAPNIELDANVPS
          *   ++ * +  +  *+* +  *+    +* +++ *****+   + +    +   *   +****+ *+ *   +  +

EWNRAFALMNGMDQDKRWTTSRPPCRGCARIRIAPAKRVGRLL--LGRRLAFMMAARSDSDANVSYYGVGLDGLVGEAASITKPLLMHIA
QREQAYKLWQAFDMEAGVGDLEAAIR-YARHQPYSNGKVGLVGYCLGGALAFLVAAKGYVDRAVGYYGVGLEKQLKKVPEVKHPALFHMG
QRARALGMFRDYKLEDGVADLRATVA-YAASQPFCDGGVAVIGYCLGGALAYEVAAEGFAQCCVGYYGVGFEKRLERARLVKTPSMFHMG
EREQALALFRDFDMEAGVNDLSCAIE-YARALPFSNGRVAVVGYCLGGALAFDVAARSLADCSIGYYGVGLEKKVSLVPAITRPAMFHMG
++ +*  +      +              *          *+ +   **  **+ +** +  +  ++******++  +     +  * + *++

EKDQFVPAEAREKVLAAVKGNPNVTAHVYPGVDHAFARAGGAHFEPEAAELANGRTAAFFKQHLG
GQDHFVPAPSRQLITEGFGANPLLQVHWYEEAGHSFARTSSSGYVASAAALANERRLDFLAPLQSKKP
TNDHFVTAEARQLITNAFEANPAIALHWY-DAGHSFARASSPNFSPEATRTANARTLEMLKRMKPIGTIGQ
TKDHYVTEEARSILEEHFGRNKNLSLHWYP-VGHSFARSSSPNFDQAATTVANARTLELLAMLKDPS
 +*+++*   +*  +      *  +  * *   +*+***+++ +   *+  ** *    +
```

Fig. 3. Comparison of deduced ORF2 and dienelactone hydrolases. Sequences correspond to the deduced amino acid sequences from ORF2 of *Azospirillum brasilense* (Ab ORF2, this work), *clcD* of *Pseudomonas putida* (Pp ClcD [17]), *tcbE* of *Pseudomonas* strain P51 (Ps TcbE [18]) and from *tfdE* of *Alcaligenes eutrophus* JMP134 (Ae TfdE [19]). Identical residues in all four sequences are indicated by an asterisk and homologous residues by a plus below the sequences.

(Fig. 3) which are enzymes catalyzing the conversion of dienelactones to maleylacetate, a reaction of the degradation pathway of halogenated aromatic compounds. The amino acid sequence of ORF2 showed 35%, 28% and 30% identical residues with the dienelactone hydrolases of *Pseudomonas putida* [17], *Pseudomonas* sp. strain P51 [18] and *Alcaligenes eutrophus* JMP134 [19], respectively.

Functional complementation of an E. coli PyrG⁻ mutant by A. brasilense ORF1

In order to check the functional property of ORF1, plasmid pAB10055 was constructed by cloning the 4.2-kb *Hin*dIII–*Eco*RI fragment of pAB1005 with the complete coding region of ORF1 into pBluescript (see Fig. 1). The *E. coli pyrG⁻* mutant JF646 [7] was transformed by pAB10055 and by the pBluescript control plasmid. Only JF646 transformed by pAB10055 was able to grow on minimal medium devoid of cytidine (data not shown). Consequently and because of the strong similarities of the deduced amino acid sequence to other *pyrG* gene products, ORF1 was designated *pyrG*, the CTP synthetase gene of *A. brasilense*. It is the fourth DNA sequence encoding a CTP synthetase available at the Nucleotide Sequence Databases and the first sequenced gene of *Azospirillum* involved in nucleotide biosynthesis.

Acknowledgement

This work was kindly supported by a grant of the Deutsche Forschungsgemeinschaft to Professor H. Bothe (No. Bo441-15).

References

1 Bozouklian, H., Fogher, C. and Elmerich, C. (1986) Cloning and characterization of the *glnA* gene of *Azospirillum brasilense* Sp7. Ann. Inst. Pasteur/Microbiol. 137, 3–18.
2 Zimmer, W. and Elmerich, C. (1991) Regulation of the synthesis of indole-3-acetic acid in *Azospirillum*. In: Advances in molecular genetics of plant-microbe interactions (Hennecke, H. and Verma, D.P.S., Eds.), Vol. I, pp. 465–468, Kluwer Academic Publishers, London.
3 Zimmer, W., Aparicio, C. and Elmerich, C. (1991) Relationship between tryptophan biosynthesis and indole-3-acetic acid production in *Azospirillum*: identification and sequencing of a *trpGDC* cluster. Mol. Gen. Genet. 229, 41–51.
4 Zimmer, W. and Elmerich, C. (1992) Organization of *trpG*, *trpD* and *trpC* in *Azospirillum brasilense* Sp7. Symbiosis 13, 133–137.
5 Essar, D.W., Eberly, L., Han, C.-Y. and Crawford, I.P. (1990) DNA sequences and characterization of four early genes of the tryptophan pathway in *Pseudomonas aeruginosa*. J. Bacteriol. 172, 853–866.
6 Sambrook, J., Fritsch, E. and Maniatis, T. (1989) Molecular cloning. Second edition, Vol. 1–3, Cold Spring Harbor Laboratory Press, New York, NY.
7 Friesen, J.D., Parker, J., Watson, R.J., Fiil, N.P., Pedersen, S. and Pedersen, F.S. (1976) Isolation of a Lambda transducing bacteriophage carrying the *relA* gene of *Escherichia coli*. J. Bacteriol. 127, 917–922.
8 Crawford, I.P. (1989) Evolution of a biosynthetic pathway: The tryptophan paradigm. Ann. Rev. Microbiol. 43, 567–600.
9 Zimmer, W. (1993) Efficient cloning of fragments of the polymerase chain reaction directly into the single stranded bacteriophage M13mp18. Nucleic Acids Res. 21, 773–774.
10 Vieille, C. and Elmerich, C. (1992) Characterization of an *Azospirillum brasilense* Sp7 gene homologous to *Alcaligenes eutrophus phbB* and to *Rhizobium meliloti nodG*. Mol. Gen. Genet. 231, 375–384.
11 Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Improved estimation of secondary structure in ribonucleic acids. Nature 246, 40–41.
12 Platt, T. (1986) Transcription termination and the regulation of gene expression. Ann. Rev. Biochem. 55, 339–372.
13 Shine, J. and Dalgarno, L. (1975) Determinant of cistron specificity in bacterial ribosomes. Nature 254, 34–38.
14 Weng, M., Makaroff, C.A. and Zalkin, H. (1986) Nucleotide sequence of *Escherichia coli pyrG* encoding CTP synthetase. J. Biol. Chem. 261, 5568–5574.
15 Trach, K., Chapman, J.W., Piggot, P., LeCoq, D. and Hoch, J.A. (1988) Complete sequence and transcriptional analysis of the *spo0F* region of the *Bacillus subtilis* chromosome. J. Bacteriol. 170, 4194–4208.
16 Yamauchi, M., Yamauchi, N. and Meuth, M. (1990) Molecular cloning of the human CTP synthetase gene by functional complementation with purified human metaphase chromosomes. EMBO J. 9, 2095–2099.
17 Frantz, B. and Chakrabarty, A.M. (1987) Organization and nucleotide sequence determination of a gene cluster involved in 3-chlorocatechol degradation. Proc. Natl. Acad. Sci. USA 84, 4460–4464.
18 Van der Meer, J.R., Eggen, R.I., Zehnder, A.J.B. and De Vos, W.M. (1991) Sequence analysis of the *Pseudomonas* sp. strain P51 *tcb* gene cluster, which encodes metabolism of chlorinated catechols: evidence for specialization of catechol 1,2-dioxygenases for chlorinated substrates. J. Bacteriol. 173, 2425–2434.
19 Perkins, E.J., Gordon, M.P., Caceres, O. and Lurquin, P.F. (1990) Organization and sequence analysis of 2,4-dichlorophenol hydrolase and dichlorocatechol oxidative operons of plasmid pJP4. J. Bacteriol. 172, 2351–2359.