Analysis of a *Streptomyces coelicolor* A3(2) locus containing the nucleoside diphosphate kinase (*ndk*) and folylpolyglutamate synthetase (*folC*) genes

Annette Burger [a], Birgit Brandt [a], Urs Süsstrunk [b], Charles J. Thompson [b], Wolfgang Wohlleben [a],*

[a] *Maikrobiologie/Biotechnologie, Eberhard-Karls-Universität Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany*
[b] *Department of Microbiology, Biozentrum, University of Basel, Klingelbergstraße 70, CH-4056 Basel, Switzerland*

Abstract

A 3.6-kb DNA fragment from *Streptomyces coelicolor* A3(2) with the genes *valS* probably encoding a valyl-tRNA synthetase, *folC* encoding folylpolyglutamate synthetase, and *ndk* encoding a nucleoside diphosphate kinase was analysed. *folC* and *ndk* are separated by a small open reading frame of unknown function, *orfX*. The deduced *folC* gene product is a protein of 46 677 Da whose sequence is similar to other folylpolyglutamate synthetases and folylpolyglutamate synthetase-dihydrofolate synthetases from both Gram-positive and Gram-negative bacteria. After cloning *folC* behind the *lacZ* promoter, the *Streptomyces folC* complemented a *folC* mutant of *Escherichia coli*. An essential function for *Streptomyces folC* was suggested by the fact that it could not be mutated using a conventional gene disruption technique. © 1998 Federation of European Microbiological Societies. Published by Elsevier Science B.V.

1. Introduction

Folic acid derivatives serve as one-carbon donors in a wide variety of cellular reactions and transfer methyl, methylen, formyl, formimino, and methenyl groups. They are essential cofactors in the biosynthesis of purines, thymidilate, glycine, methionine, formyl-methionyl-tRNA, and pantothenate. The predominant intracellular form of folate coenzymes in both bacteria and eucaryotes are the poly-γ-glutamate derivatives. The enzyme folylpolyglutamate synthetase (Fpgs) catalyzes the addition of two to seven glutamyl residues to the monoglutamate tetrahydrofolate. The resulting folylpolyglutamates are preferentially retained and help to concentrate the folates intracellularly. The polyglutamate tail is required for intracellular retention and plays an important role in the specific recognition for the enzymes of one-carbon metabolism (for review see [1]).

The *Escherichia coli* [2], *Corynebacterium spec.* [3], and *Neisseria gonorrhoeae* [4] FolC gene products have been shown to be bifunctional in that they pos-

sess a second activity, dihydrofolate synthetase (DHFS). DHFS adds one glutamate residue to dihydropteroate to form dihydrofolate which is reduced in the next step by dihydrofolate reductase to tetrahydrofolate, the substrate for FolC.

We here present a genetic and functional analysis of the *S. coelicolor folC* gene which was identified upstream of the formerly characterized *ndk* gene [5].

2. Materials and methods

2.1. Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are listed in Table 1.

2.2. Molecular cloning

Isolation of plasmid DNA from *Streptomyces* and *E. coli*, cleavage of the DNA, ligation and transformation of the respective strains were performed using standard protocols [6,10]

2.3. DNA sequencing and analysis

The DNA sequence of the 3.6-kb *Bam*HI fragment was determined by standard techniques modified for double stranded DNA using the Autoread Sequencing Kit (Pharmacia) for sequencing with the Automated Laser Fluorescence Sequencer (A.L.F., Pharmacia). The DNA fragment was examined for open reading frames by applying the codon usage program of Staden and McLachlan [11]. The programs

BLAST [12,13] and FASTA [14] were used for homology searches.

2.4. Investigation of growth behaviour

Growth curves from *E. coli* strains were measured in minimal medium [6]. One ml of o/n cultures of SF4, SF4(pAB12), K12, and K12(pAB12) grown in LB medium were centrifuged, the cells were washed with 3×1 ml minimal medium and resuspended in 1 ml minimal medium. Hundred µl of these suspensions were used for inoculation of 50 ml MM. Cultures were incubated at 37°C on a rotary shaker (180 rpm). At appropriate times 1 ml samples were taken and the O.D. was measured over 37 h.

2.5. Nucleotide accession number

The nucleotide sequence data reported have been assigned the accession no. Y13070 in the EMBL data library.

3. Results

3.1. Cloning of a 3.6-kb BamHI fragment including the S. coelicolor A3(2) genes valS, folC, orfX and ndk

A phage gene bank of *S. coelicolor* A3(2) was constructed by cloning partially *Sau*3AI digested *S. coelicolor* M145 genomic DNA into the vector λ-FixII (STRATAGENE GmbH, Heidelberg, Germany). One phage containing a 3.6-kb *Bam*HI fragment was identified which specifically hybridized against

Table 1
Bacterial strains and plasmids used in this study

Strain	Relevant genotype/phenotype	Source or reference
S. coelicolor M145	Pgl$^+$, SCP1$^-$, SCP2$^-$	[6]
E. coli XL1-Blue	*recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac* (F'*proAB lacI*^{q} *Z*ΔM15)	[7]
E. coli SF4	F$^-$, *folC, strA, recA, srlC*::Tn*10*	[2]
Plasmid		
pK18/pK19	*aphII*	[8]
pGM9	*tsr* (thiostrepton resistance gene), temperature-sensitive *Streptomyces* vector	[9]
pAB12/pAB21	pK19/pK18 carrying the complete *folC* gene on a 1694-bp *Xho*I/*Nco*I fragment*	this work
pAB3	pGM9 derivative with the internal 680-bp *Fsp*I fragment for disruption of *folC**	this work

*For cloning in the pAB vectors, the *folC* fragments were provided with suitable restriction sites by subcloning in different multiple cloning sites.

a *ndk* gene probe. The *ndk* gene is part of a 11.2-kb gene region (to be published elsewhere) involved in cell division and differentiation and was characterized as nucleoside diphosphate kinase [5]. After having localized the *ndk* gene on a 800-bp *Pst*I-*Bam*HI fragment at one end of the 3.6-kb *Bam*HI fragment, the preceding region upstream from *ndk* was analysed. Three other functional open reading frames were predicted using the Staden programs [11] to identify regions which display *Streptomyces* codon preference [15]. These orfs are orientated in the same direction and are colinear with the downstream *ndk*. A gene bank similarity search identified the open reading frames as the C-terminus of a *valS* (valyl-tRNA synthetase) gene, a *folC* (folylpolyglutamate synthetase) gene and an *orfX* with no significant similarity to any gene previously described.

3.2. Sequence analysis of the 3.6-kb BamHI fragment

The DNA sequence of the 3.6-kb *Bam*HI fragment of the cloned region is shown in Fig. 1.

The first 779-nt (nucleotides) showed high coding probability up to a TAG stop codon at nt position 777–779. The start codon of this orf is missing on the 3.6-kb *Bam*HI fragment. The overall GC content is 71.5%, typical for *Streptomyces* genes [15]. Comparison of the deduced amino acid sequence with the proteins from peptide sequence databases using the BLAST server revealed high similarity (29–33% identity) with the C-terminal part of several valyl-tRNA synthetases (ValS) found in *Bacillus subtilis*, *Haemophilus influenzae* and *E. coli*.

The second open reading frame, *folC*, probably begins 475 nt downstream of the end of *valS* with a GTG codon at position 1254 and ends with a stop codon TGA at position 2586. It would encode a protein of 444 aa with a deduced molecular mass of 46 677 Da. Immediately upstream from *folC* there is a sequence GGAGGA (position 1241–1246) with strong complementarity to the 3′ end of *S. coelicolor* A3(2) 16S rRNA [16] and with a distance of 7 nt to

the translational start codon properly positioned to serve as a ribosome binding site for the *folC* gene. This orf possesses a GC content of 71.5%.

Immediately downstream from *folC* there is a region (*orfX*) with a high coding probability in the same reading frame as *folC*. There are several possible putative start codons for this *orfX*: GTG at position 2595, GTG at position 2637, ATG at position 2661, ATG at position 2682, or ATG at position 2715. Putative ribosome binding sites can be identified 13 nt in front of the first GTG codon (GAGGAGG; position 2576–2582), 2 nt in front of the first GTG codon (GAGAAA; position 2587–2592) or 13 nt in front of the fourth start codon ATG (GAAGGA, position 2663–2668), indicating a possible function of the first or fourth start codon. The first putative ribosome binding site would be located completely within the 3′ part of the *folC* gene. *orfX* extends up to the stop codon TGA at position 2946 and consists, depending on the start codon, of 354 or 267 nt. It shows a GC content of 69.5 or 71.9%. *orfX* encodes a putative protein of 118 or 89 aa with a predicted molecular mass of 12 286 or 9190 Da. Gene bank search did not show any significant similarity to any previously described gene. However, the order of genes in this region resembles that of *E. coli*, *N. gonorrhoeae*, or *Lactobacillus casei*. In all those cases the *folC* gene is immediately followed by orfs of unknown function which are not similar to each other.

124 nt downstream from the end of *orfX* begins the *ndk* gene which has previously been identified and characterized [5].

3.3. Comparison of the FolC protein with homologs of other organisms

Data bank search using the BLAST network service showed that the deduced *Streptomyces* FolC sequence had significant similarity to all known folylpolyglutamate synthetase enzymes described so far.

The *S. coelicolor folC* gene product (444 aa) is

→

Fig. 1. (Pages 286 and 287) DNA sequence of a 3.1-kb segment of the 3.6-kb *Bam*HI fragment (up to the *ndk* gene) containing *valS*, *folC*, *orfX* and the beginning of *ndk*. The nucleotide sequence was determined on both strands. The predicted aa sequence is given in a single letter code below the DNA sequence (* = stop codon). The beginning of the reading frames is indicated by arrows at the end of the lines. Start codons are marked in bold type. Important restriction sites are given above the sequence. Proposed ribosome binding sites (rbs) are underlined.

BamHI
```
GGATCCTCTCCCGCCTGAACTCGGTCGTGGCCGAGGTCGACGCGTACTACGAGGACTACC-  60 → valS
  I   L   S   R   L   N   S   V   V   A   E   V   D   A   Y   Y   E   D   Y   Q

AGTTCGCGAACGTCTCCGACGCCCTGTTCCACTTCGCCTGGGACGAGGTCTTCGACTGGT- 120
  F   A   N   V   S   D   A   L   F   H   F   A   W   D   E   V   F   D   W   Y

ACGTCGAGCTGTCCAAGACCACCTTCCAGGCGGGCGGCGAGGCGGCCGAGGTCTCGAAGC- 180
  V   E   L   S   K   T   T   F   Q   A   G   G   E   A   A   E   V   S   K   R

GCGTCCTGGGCGAGGTCCTGGACGTGACGCTGCGTCTCCTCCACCCGGTCGTCCCCTTCG- 240
  V   L   G   E   V   L   D   V   T   L   R   L   L   H   P   V   V   P   F   V

TCACGGAGACGCTGTGGACGACGCTCACGGGCGGCGAGTCGGTCGTCATCGCCGAGTGGC- 300
  T   E   T   L   W   T   T   L   T   G   G   E   S   V   V   I   A   E   W   P

CGACGGACTCCGGTTTCCGGGACGCCGGCGCCGAGCGGGAGATCGAGACCGTGCAGTCGG- 360
  T   D   S   G   F   R   D   A   G   A   E   R   E   I   E   T   V   Q   S   V

TCATCACCGAGGTCCGCCGCTTCCGTGCCGACCAGGGGCTCCAGCCCGGCCAGCGGGTCC- 420
  I   T   E   V   R   R   F   R   A   D   Q   G   L   Q   P   G   Q   R   V   P

CGGCCCGGCTCACCCTCGCGGGCAGCGCGCTCGCGGCCCACGAGGCGGCCGTCCGGCAGC- 480
  A   R   L   T   L   A   G   S   A   L   A   A   H   E   A   A   V   R   Q   L

TGCTGCGGCTCCAGCCGGAGGGCGACGCCTTCACGGCGACGGCCACGCTGCCGGTCGCCG- 540
  L   R   L   Q   P   E   G   D   A   F   T   A   T   A   T   L   P   V   A   G

GCGTCGAGGTGGCCCTCGACCTCTCCGGGGTCATCGACTTCGCCGCCGAGCGCAAGCGGC- 600
  V   E   V   A   L   D   L   S   G   V   I   D   F   A   A   E   R   K   R   L

TGGCGAAGGACCTCGCGGCGGCGGAGAAGGAGAAGGCCCAGGCGAACGCCAAGCTCGGCA- 660
  A   K   D   L   A   A   A   E   K   E   K   A   Q   A   N   A   K   L   G   N

ACGAGGCGTTCCTGGCGAAGGCACCGGACCAGGTCGTGGACAAGATCCGGGGCCGCCTGG- 720
  E   A   F   L   A   K   A   P   D   Q   V   V   D   K   I   R   G   R   L   A

CCAAGGCGGACGAGGACATCACCCGCATCACCGCCCAGCTGGAGAAGCTGCCGGAGTAGT- 780
  K   A   D   E   D   I   T   R   I   T   A   Q   L   E   K   L   P   E   *

CGCCGCCGCCTGCGCCCACCCGTGCCGCTGGGGCGGCACCCCGCAGGCGGCACCCCGCAG- 840

GCGGCGAGCACACCCACCGTGACGGCAACCGAAGGCGCCTCGGGCCCGATCGGCTCCGGG- 900

CACCCTCGGTTGACGGCGGGTGGGTCGCCCGGCCCGGCGCTGACGGGGTGCCGCCGCGCC- 960
```
                    XhoI
```
CACCCTCCCCCACTCTCGGCTTCGCTCGAGCGGGGGGACCCCCATCGCCCCAGGCGGCCC-1020
```
        SacII
```
GACTGCCCGCGGTTACGGGGGACCGGGTACGTAGACTGGTCCCCGTGAGCGACAACCCCG-1080

GCCAGAACGACCAGCCCGACCCCTCCGCCCCCCTCGACTCCTTCGACGAGATCATCGACG-1140

CGGAGACCACCCGCGACCCCGACCTCGCCGTCATCGAGGCCGGCAGCCGCACCCTGCGCA-1200
```
              rbs   rbs
```
CCCAGGGCGGCCCCCCGGAGGCCGACGTCCCCGGGCGCCCGGAGGACCCCGAGGTGGACG-1260→ folC
                                                      fM  D   A

CGGCCCTGCGCGCCGTCGAGACGGAGCTGGCCGGCCGCTGGGGCGAGACCAAGCTGGAGC-1320
  A   L   R   A   V   E   T   E   L   A   G   R   W   G   E   T   K   L   E   P
```
                                        Eco47-III
```
CGTCCGTCACCAGGATCGCCGCGCTGATGGACGTACTGGGCGAGCCGCAGCGCTCGTACC-1380
  S   V   T   R   I   A   A   L   M   D   V   L   G   E   P   Q   R   S   Y   P

CCTCCATCCACATCACCGGGACCAACGGCAAGACCTCCACCGCCCGCATGATCGAGGCCC-1440
  S   I   H   I   T   G   T   N   G   K   T   S   T   A   R   M   I   E   A   L

TGCTCGGCGCCTTCGATCTGCGCACCGGGCGGTACACCTCGCCGCACGTGCAGTCGATCA-1500
  L   G   A   F   D   L   R   T   G   R   Y   T   S   P   H   V   Q   S   I   T

CCGAGCGGATCAGCCTGGACGGGGCGCCGATCTCCGCCGAGCGGTTCATCGAGACCTACG-1560
  E   R   I   S   L   D   G   A   P   I   S   A   E   R   F   I   E   T   Y   E

AGGACATCAAGCCGTACGTCGAGATGGTCGACGCGCAGAGGAGTACCGGCTGTTCCTTCT-1620
  D   I   K   P   Y   V   E   M   V   D   A   Q   R   S   T   G   C   S   F   F
```

Fig. 1 (legend on p. 285)

```
TCGAGGTGCTGACCGGCATGGCGTACGCCGCCTTCGCCGACGCACCCGTGGACGTGGCCG-1680
  E   V   L   T   G   M   A   Y   A   A   F   A   D   A   P   V   D   V   A   V
TCGTCGAGGTCGGGATGGGCGGCTCCTGGGACGCCACCAACGTCATCGACGGGGACGTGG-1740
  V   E   V   G   M   G   G   S   W   D   A   T   N   V   I   D   G   D   V   A
CCGTCGTCACCCCCATCGACCTGGACCACACCGACCGGCTCGGCGGGACGCCCGGCGAGA-1800
  V   V   T   P   I   D   L   D   H   T   D   R   L   G   G   T   P   G   E   I
TCGCGGTCGAGAAGGCCGGCATCGTCAAGCAGGACGCCACCGTCATCCTGGCCCAGCAGC-1860
  A   V   E   K   A   G   I   V   K   Q   D   A   T   V   I   L   A   Q   Q   P
CGGTGGACGCGGCGCAGGTGATGCTGAAGAAGGCCGTCGACGTCAACGCCACCGTGGCCC-1920
  V   D   A   A   Q   V   M   L   K   K   A   V   D   V   N   A   T   V   A   R
GCGAGGGGCTGGAGTTCGGCGTCGTGGCCCGCGAGGTCCCGTCGGCGGGCAGCTGGCCAC-1980
  E   G   L   E   F   G   V   V   A   R   E   V   P   S   A   G   S   W   P   P
CCTGCGCGGGACTGGGCGGCGAGTACACCGAGGTCTACCTGCCGCTGCACGGCGCCCACC-2040
  C   A   G   L   G   G   E   Y   T   E   V   Y   L   P   L   H   G   A   H   Q
AGGCGCACAACGCGGCGGTCGCGCTGGCCGCCGTCGAGGCGTTCTTCGGCGTCGGCGCGC-2100
  A   H   N   A   A   V   A   L   A   A   V   E   A   F   F   G   V   G   A   Q
AGCGTCCCGAGCCGCTGGACGTCGACACCGTGCGCAAGGCCTTCGCCTCCGTCGCCTCGC-2160
  R   P   E   P   L   D   V   D   T   V   R   K   A   F   A   S   V   A   S   P
CGGGGCGGATGGAGGTGGTGCGGCGCTCGCCGACCGTGGTCCTCGACGCCGCGCACAACC-2220
  G   R   M   E   V   V   R   R   S   P   T   V   V   L   D   A   A   H   N   P
CGGCCGGCGCCCGGGTCACCGCCGAGGCGGTGGGGGAGGCCTTCCAGTTCAGCCGCCTCA-2280
  A   G   A   R   V   T   A   E   A   V   G   E   A   F   Q   F   S   R   L   I
TCGGGGTGGTCGGCGCCAGCGCCGACAAGAACGTGCGGGGGCTCCTCGAAGCCTTCGAGC-2340
  G   V   V   G   A   S   A   D   K   N   V   R   G   L   L   E   A   F   E   P
CGGTCTTCGCCGAGGTCGTCGTCACCCAGAACTCCAGCCACCGCGCGATGGACGCCGACG-2400
  V   F   A   E   V   V   V   T   Q   N   S   S   H   R   A   M   D   A   D   E
        SacII                                                   SacII
AGCTGGCCGCGGTCGCCGTCGAGGTCTTCGGCGACGACCGCGTCCAGGTCGAGCCGCGGC-2460
  L   A   A   V   A   V   E   V   F   G   D   D   R   V   Q   V   E   P   R   L
TGCCGGACGCCCTGGAGGCCGCGATCACGCTGGCCGAGGAGGAGGGCGAGTTCGCCGGCG-2520
  P   D   A   L   E   A   A   I   T   L   A   E   E   E   G   E   F   A   G   G
                                                            rbs
GCGGTGTGCTCGTCACCGGTTCCGTCATCACCGTCGGCGAGGCCCGACTGCTCCTGAGGA-2580
  G   V   L   V   T   G   S   V   I   T   V   G   E   A   R   L   L   L   R   R
  rbs
GGGGCTGAGAAACCGTGCGTACGCTCTGTGCTTCGACGCTGATCGGCGAGTTCTTCGTGA-2640 →  ***orfX***
  G   *        fM   R   T   L   C   A   S   T   L   I   G   E   F   F  fM   I
                                                    NcoI
TCTTCGCCGGACTGGTCGCCATGAAGGACCCGGACCTGTCCATGGGGACGGTGTGGACGG-2700
  F   A   G   L   V   A   M   K   D   P   D   L   S   M   G   T   V   W   T   V
TCAGCGGCATCGCGATGGTCCTCAGCGTGCTGCTGTGCGGGGTCATCACCCGGCCCGGCG-2760
  S   G   I   A   M   V   L   S   V   L   L   C   G   V   I   T   R   P   G   G
        PstI
GGATCGCCCTCGGCTGGGCGCTGCAGATCGCGCTGATCGCCGCCGGCTTCGTGGTGCCGA-2820
  I   A   L   G   W   A   L   Q   I   A   L   I   A   A   G   F   V   V   P   T
CCATGTTCTTCCTCGGCGCGATCTTCGCCGCCCTGTGGTGGGCCTCCGTGCACTACGGGC-2880
  M   F   F   L   G   A   I   F   A   A   L   W   W   A   S   V   H   Y   G   R
GGAAGATCGACGAGGCGAAGGCGCGGTTCGCGGCCCAGGCCGAGGCCTCCGCACCCGACC-2940
  K   I   D   E   A   K   A   R   F   A   A   Q   A   E   A   S   A   P   D   P
CCGCCTGAGTGCGGAACGCGGCCGACGGGGCCCGGTAGGCCTGACGCTGCGTGACGGACG-3000
  A   *
CCGCGACACGCCCGGTAACCTCAGTCATCACCCGCACGACCGCACATCTCGAACACCTCC-3060
rbs
AAGGGAGCCCTCGTGACCCAGCGCAGCCTCGTCCTCCTCAAGCCCGAAGCCGTCCGTCGT-3120 →  ***ndk***
        fM   T   Q   R   S   L   V   L   L   K   P   E   A   V   R   R   G
```

Fig. 1 (*continued*)

```
S.coel   :  VD--------AALRAVETELAGRWGETKLEPSVTRIAALMDVLGEPQRSYPSIHLTGT  :  50
L.case   :  MN----------YTETVAYIHSFPRLAK-TGDHRRILTLLHALCNPQQQGRYIHVTGT  :  47
B.subt   :  MF--------TAYQDARSWIHGRLKFGV-KPGLGRMKQLMARLCHPEKKIRAFHVAGT  :  49
S.pneu   :  MK--------EI--ENNQWIANYRTDQP-HFGLERMVELLALRGNPHLKLKVLHLGGT  :  47
N.gono   :  M---------KTLQDWLSHLETAHSGGLIDMGKERVSEVKKRYN-LTPQCPVVVVAGT  :  48
E.coli   :  MIIKRTPQAASPLASWLSYLENLHSKT-IDLGLERVSLVAARLGVLKPAPFVFTVAGT  :  57


S.coel   :  NGKTSTARMIEALLGAFDLRTGRYTSPHVQSITERISLDGAPTSAERFIETYEDKKPY  :  108
L.case   :  NGKGSAANAIAHVLEASGLTVGLYTSPFTKRFNERIMILHEPIPDAALVNAVAFVRAA  :  105
B.subt   :  NGKGSTVAFIRSMLQEACYTVGTLTSPYIITRNERISVLGIPISDEEWTALVNQCKPH  :  107
S.pneu   :  NGKGSTIAFLKKMLEKLGLRVGVLSSPYLIHTDQISINGESTSEARLEALMAD----  :  101
N.gono   :  NGKGSVCAYLTQIYKQASFKTGTLTSPHLLHYNERIAINAEPVSDDTIIASFERIEA-  :  105
E.coli   :  NGKGTTCRTLESILMAAGYKVGVYSSPHLVRYTERVRVQGQELPESAHTASFAEIES-  :  114


S.coel   :  VE-MVDAQRSTGCSFFEVLTGMAYAAFAD-APVDVAVVEVGMGGSWDATNVIDGDVAV  :  164
L.case   :  LERLQQQQADFNVTEFEFITALGYWYFRQ-RQVDVAVLEVGIGGDTDSTNVITPVVSV  :  162
B.subt   :  VEALDQTEYG-QPTEFEIMTACAFLYFAEFHKVDFVIFETGLGGRFDSTNVVEPLLTV  :  164
S.pneu   :  YQSLLEGEAV-ANTEFEIITALAYDYFAS-EQVDVAIMEVGMGGLFDSTNVCQPILTG  :  157
N.gono   :  ------ARGEISLTYFEFNALAAVDIFIR-EQVDVMIYEVGLGGRLDAVNGFDGDCAV  :  156
E.coli   :  ------ARGDISLTYFEYGTLSALWLFKQ-AQLDVVIIEVGLGGRLDATNIVDADVAV  :  165


S.coel   :  VTPIDLDHTDRLGGTPGETAVEKAGIVKQDATVILAQQPVDAAQVMLKKAVDVNATVA  :  222
L.case   :  LTEVALDHQKLLGHTITAIAKHKAGIIKRGIPVVTGNLVPDAAAVVAAKVATTGSQWL  :  220
B.subt   :  ITSIGHDHMNILGNTIEEILAGEKAGIIKEGIPIVTAVTQPEALQVIRHEAERHAAPFQ  :  222
S.pneu   :  ITTIGLDHVALLGDTLEATAEQKAGIIKQGMPLVTGRIAPEALAVIDRIAEGKDAPRL  :  215
N.gono   :  VTSVDLDHQAFLGDTVEQVGFEKAGVFRSGKPAICGQ--NPTPASLVAHAEAIGAKLL  :  212
E.coli   :  VTSIALDHTDWLGPDRESIGREKAGIFRSEKPAIVGE--PEMPSTIADVAQEKGA--L  :  219


S.coel   :  ----REGLEFGVVAREVPSAGSWPPCAGLGGEYTEVYLPLH-GAHGAHNAAVATAAVE  :  275
L.case   :  ----RFDRDFSVPKAKLHGWGQRFTYEDQDGRISDLEVPLV-GDYCQRNMAIALQTAK  :  273
B.subt   :  ----SLHDACVIFNEEALPAGEQFSFKTEEKCYEDIRTSLICGTHCRQNAALSI-LAA  :  274
S.pneu   :  ----AYGTDYQVRHQESVVTGEVFDYTSAVR-QGRFQTSLIGCLYSIENAGMAIALLD  :  267
N.gono   :  MVQRDFEFH-AMENIQWNYRFRPQHSDGPARNRNAAPFPALRCAYQLSNAACALTVLE  :  269
E.coli   :  LQRRGVEWNYSVTDHDWAFS----DAHGTL---ENLPLPLVPQ----PNAATLLAALR  :  266


S.coel   :  AFFGVGAQRPEPLDVDTVKAFASVASPGRMEVVRRSPTVVLDAAHNPAGARVTAEAV  :  333
L.case   :  VY--AKQTEWPLTPQN-TRQGLAASHWEARLEKISDTPLIYLDGAHNLDGINGLITAL  :  328
B.subt   :  EW--LNKENIAHISDEALRSGLVKAAWPGRLFTVQEHPPVYLLDGAHNEEGVEKLAETM  :  330
S.pneu   :  TF--CQEDGRELASNDFLGQALEETSWPGRLETVSRDPLMYLDGAHNPHAIKALLVTL  :  323
N.gono   :  CL-----DDRLPVDIGALKRGLLLVENPGRFQVLPGRPLTVLDVGHNPHAARALRRNL  :  322
E.coli   :  -------ASGLEVSENAIRDGIASAILPGRFQIVSESPRVIFDVAHNPHAAEYLTGRM  :  317


S.coel   :  GEAFQFSRLIGVVGASADKNVRGLLEAFEPVFAEVVVTQNSSHRAMDADELAAVAVEV  :  391
L.case   :  KQLFSQ-PITVIAGILADKDYAALADRYTA----AFSTVYLVPVPGTPRALPEAGYEA  :  381
B.subt   :  KQRFANSRISVVFSAKDKPYQNVIKRLETIAHA----IHFASFDFPRASLAKDLYDA  :  384
S.pneu   :  QERFADYHKEILFTCKTRALEDYLDLLGAMPDTELTLTHFADSRATDESVLKEAAKS  :  381
N.gono   :  INLAYAQKRTAVFSMLSDKDIDGVSETYKDQFDEWYIAPLDVPRGMTADALKAKLEQ  :  379
E.coli   :  KALPKNGRMLAVIGMLHDKDIAGTLAWLKSVVDDWYCAPLEGPRGATAEQL----LE-  :  370
```

Fig. 2. Multiple aa sequence alignment [17] of folylpolyglutamate synthetases and folylpolyglutamate synthetase-dihydrofolate synthetases of *S. coelicolor* A3(2), *L. casei* [20], *Streptococcus pneumoniae* [18], *B. subtilis* [21], *N. gonorrhoeae* [4], and *E. coli* [2]. Groups of conserved aa were summarized as follows: 1. D, N; 2. E, Q; 3. S, T; 4. K, R; 5. F, Y, W; and 6. L, I, V, M. Amino acid residues conserved in all 6 proteins are shaded black, those which are conserved in 5 organisms are shaded grey. Dashes indicate gaps in the alignment. Numbers on the right of each lane correspond to the aa positions relative to the start of each protein.

```
S.coel    :  FGDDRVQVEPRLPDALEAAGTLAEEEGEFAGGGVLVTGSVITVGEARLLLR-------  :  442
L.case    :  LHEGRLK--DSWQEALAASLN-----DV-PDQPIVITGSLYLASAVRQTLLG------  :  425
B.subt    :  SEIS--N--KSWSEDPDDVGKFIESKKG-SNEIGLITGSLYFISDIRKRLK-------  :  430
S.pneu    :  RNLS--Y--QDWHDFLEQNGT--DKKEE-KQTVRIVTGSLYFLSQVRVYLMERKNENG  :  432
N.gono    :  HHIENIQTFAAVRDAYRAAAS-----KAGEDDRVVFGSFHTVADVMSVL--------  :  424
E.coli    :  -HLGNGKSFDSVAQAWDAAMA-----DAKAEDTVLVCGSFHTVAHVMEVIDARRS---  :  419
```

```
S.coel    :  --RG  :  444
L.case    :  -GKS  :  428
B.subt    :  ----  :   -
S.pneu    :  YTKD  :  436
N.gono    :  ----  :   -
E.coli    :  -GGK  :  422
```

Fig. 2 (*continued*)

slightly longer than the other described bacterial FolC enzymes (422–435 aa). By comparing the sequences with the 'lfasta program', highest similarity could be found to the FolC from *L. casei* with 30.7% identity and 74.1% similarity over a range of 391 aa. The *S. coelicolor* FolC protein is slightly more similar to those found in Gram-positive bacteria than to those found in Gram-negative bacteria.

Regions of homology are not distributed all over the sequence but are concentrated over distinct patches as shown in Fig. 2. ATP binding motifs A and B as postulated by Bognar et al. [19] are highly conserved in the *S. coelicolor* enzyme and are located at the positions 47–55 (ITGTNGKTS) and 138–150 (PVDVAVVEVGMGG). Interestingly, the amino acid alanine at *E. coli* position 309 whose exchange causes the defect in the *E. coli* mutant SF4 is not conserved in *S. coelicolor* but replaced by a G, as also shown for the monofunctional *Lactobacillus* FolC [20] and the *Bacillus* FolC [21].

3.4. The S. coelicolor A3(2) folC gene complements the E. coli mutant SF4

To prove that the *folC* orf really encodes a protein with Fpgs activity, plasmids pAB12 and pAB21 were constructed for complementation tests in *E. coli* SF4 auxotrophic for methionine because of its defective folylpolyglutamate synthetase gene *folC*. A fragment containing the entire *folC* gene was cloned in either orientation downstream of the *lac* promoter (pAB12 and pAB21; Fig. 3). Only the construction oriented so as to allow *folC* expression from p*lac* (pAB12) was able to grow on minimal medium with glucose.

To see the effect of additional copies of *Strepto-myces folC* in *E. coli*, growth curves were measured in minimal medium (Fig. 4). SF4 does not grow in minimal medium without methionine. SF4(pAB12) grows in the same medium with a generation time of 164 min. However, this generation time is significantly longer than that of K12 (62 min). The growth rate of K12 is not influenced by additional copies of *Streptomyces folC*. These results show that *Streptomyces folC* can overcome the defect of SF4.

3.5. Disruption of the S. coelicolor A3(2) folC gene results in loss of viability

In *E. coli* [22–24] and in *N. gonorrhoeae* [3] insertional inactivation of the bifunctional *folC* genes is lethal. To show the effect of loss of FolC activity in *S. coelicolor* A3(2) gene disruption experiments were performed using the temperature-sensitive *Streptomyces* vector pGM9 [9]. The internal 680 bp *Fsp*I fragment of the *S. coelicolor* A3(2) *folC* gene was cloned into the *Sma*I cut vector pK19 [8] and subsequently inserted as an *Eco*RI/*Hin*dIII fragment into the vector pGM9, resulting in plasmid pAB3 (Fig. 3). pAB3 was introduced into *S. coelicolor* A3(2). Gene disruption was performed using the optimized method as described by Schwartz et al. [25].

After shifting the temperature to 39°C 10^9 viable cells were spread on thiostrepton-containing LB agar plates to select for bacteria carrying pAB3 integrated in the chromosome thus disrupting the *folC* gene. Further incubation of these plates at 39°C led to the selection of 20 thiostrepton resistant colonies. However, these colonies could not be stably cultivated, neither on solid nor in liquid media under selective pressure. These observations suggest that
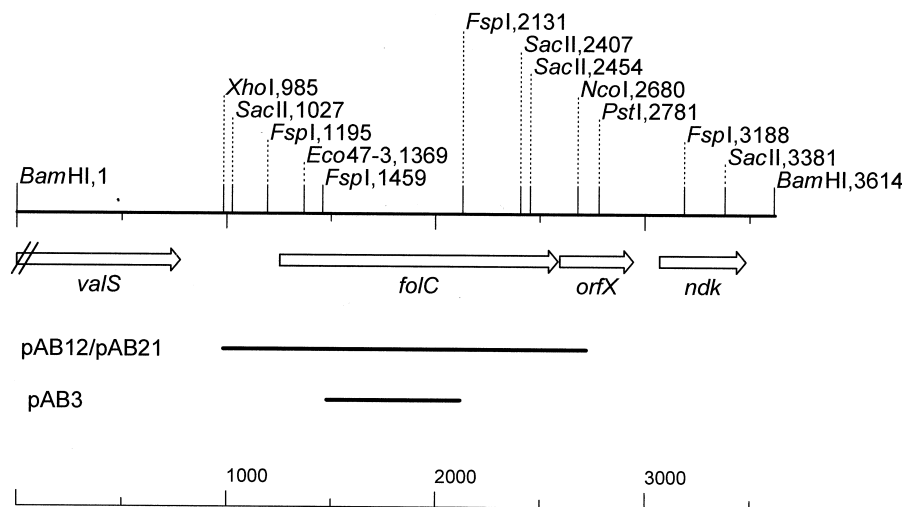
Fig. 3. Restriction map and localization of the genes *valS*, *folC*, *orfX*, and *ndk* on the sequenced *Bam*HI fragment. The lines under the genes show the fragments which were cloned to get plasmids pAB12, pAB21, and pAB3. // indicates that the 5′ terminal part of *valS* is missing on this *Bam*HI fragment.

the *S. coelicolor* A3(2) *folC* gene is necessary for the cells' viability.

4. Discussion

In this work we identify and describe the *S. coelicolor* A3(2) genes *valS*, *folC* and *orfX* which are located upstream of the previously described *ndk* gene
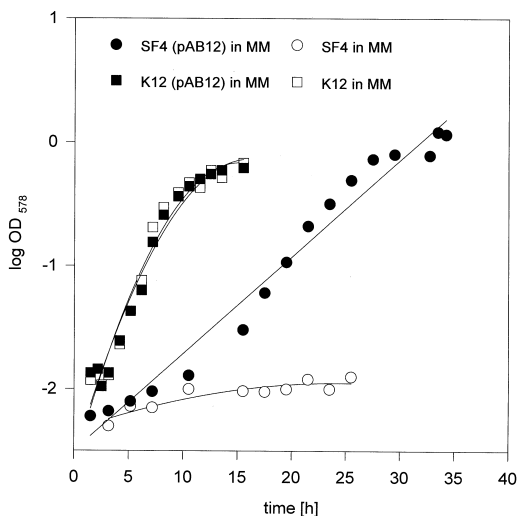
Fig. 4. Growth curves of SF4, SF4(pAB12), K12, and K12(pAB12) in minimal medium.

[5] involved in nucleoside triphosphate formation. The complete *S. coelicolor* A3(2) *folC* gene complemented the methionine auxotrophic *E. coli folC* mutant SF4 [2] to prototrophy when expressed under the control of the *lacZ* promoter.

Comparison of the *S. coelicolor* FolC with homologs from *L. casei*, *B. subtilis*, *S. pneumoniae*, *N. gonorrhoeae*, and *E. coli* shows an identity of about 30% between the species with conservative exchanges increasing the similarity up to 70–74%.

Attempts to stably inactivate the *folC* gene by gene disruption experiments failed as also described for *E. coli* and *N. gonorrhoeae*. Even on rich media mutants were not viable perhaps because of a requirement for formyl-methionyl-tRNA which requires the synthesis of precursors which cannot be taken from the medium. The growth of putative disruption mutants after the first selection step may be attributed to residual cellular folylpolyglutamate pools which allow survival of *S. coelicolor* for several generations. Further cultivation of these mutants possibly requires de novo folylpolyglutamate synthesis and is therefore not possible.

References

[1] Green, J.M., Nichols, B.P. and Matthews, R.G. (1996) Folate biosynthesis, reduction and polyglutamylation. In: *Escherichia coli* and *Salmonella typhimurium* (Neidhardt, F.C., Ed.), pp. 665–673. ASM Press, Washington.

[2] Bognar, A.L., Osborne, C., Shane, B., Singer, S.C. and Ferone, R. (1985) Folylpoly-γ-glutamate synthetase-dihydrofolate synthetase. Cloning and high expression of the *Escherichia coli folC* gene and purification and properties of the gene product. J. Biol. Chem. 260, 5625–5630.

[3] Shane, B. (1980) Pteroylpoly (γ-glutamate) synthesis by *Corynebacterium* species, purification and properties of folylpoly (γ-glutamate) synthetase. J. Biol. Chem. 255, 5655–5662.

[4] Fussenegger, M. and Meyer, T.F. (1996) Cloning and characterization of the *Neisseria gonorrhoeae* MS11 *folC* gene. Mol. Gen. Genet. 250, 277–285.

[5] Brodbeck, M., Rohling, A., Wohlleben, W., Thompson, C.J. and Süsstrunk, U. (1996) Nucleoside diphosphate kinase from *Streptomyces coelicolor*. Eur. J. Biochem. 239, 208–213.

[6] Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. and Schrempf, H. (1985) Genetic Manipulation of *Streptomyces*. A Laboratory Manual. John Innes Foundation, Norwich, United Kingdom.

[7] Bullock, W.O., Fernandez, J.M. and Short, J.M. (1987) XL1-Blue, a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. Focus 5, 376–378.

[8] Pridmore, R.D. (1987) New and versatile cloning vectors with kanamycin-resistance marker. Gene 56, 309–312.

[9] Muth, G., Nußbaumer, B., Wohlleben, W. and Pühler, A. (1989) A vector system with temperature-sensitive replication for gene disruption and mutational cloning in streptomycetes. Mol. Gen. Genet. 219, 341–348.

[10] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[11] Staden, R. and McLachlan, A.D. (1982) Codon preference and its use in identifying protein coding regions in large DNA sequences. Nucleic Acids Res. 10, 141–156.

[12] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

[13] Gish, W. and States, D.J. (1993) Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

[14] Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444–2448.

[15] Wright, F. and Bibb, M.J. (1992) Codon usage in the G+C rich *Streptomyces* genome. Gene 113, 55–65.

[16] Bourn, W.R. and Babb, B. (1995) Computer assisted identification and classification of streptomycete promoters. Nucleic Acids Res. 23, 3696–3703.

[17] Higgins, D.G. and Sharp, P.M. (1989) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73, 237–244.

[18] Lacks, S.A., Greenberg, B. and Lopez, P. (1995) A cluster of four genes encoding enzymes for five steps in the folate biosynthetic pathway of *Streptococcus pneumoniae*. J. Bacteriol. 177, 66–74.

[19] Bognar, A.L., Osborne, C. and Shane, B. (1987) Primary structure of the *Escherichia coli folC* gene and its folylpolyglutamate synthetase-dihydrofolate synthetase product and regulation of expression by an upstream gene. J. Biol. Chem. 262, 12337–12343.

[20] Toy, Y. and Bognar, A.L. (1990) Cloning and expression of the gene encoding *Lactobacillus casei* folyl-γ-glutamate synthetase in *Escherichia coli* and determination of its primary structure. J. Biol. Chem. 265, 2492–2499.

[21] Margolis, P.S., Driks, A. and Losick, R. (1993) Sporulation gene *spoIIB* from *Bacillus subtilis*. J. Bacteriol. 175, 528–540.

[22] Keshavje, K., Pyne, C. and Bognar, A.L. (1991) Characterization of a mutation affecting the function of *Escherichia coli* folylpolyglutamate synthetase-dihydrofolate synthetase and further mutations produced in vitro at the same locus. J. Biol. Chem. 266, 19925–19929.

[23] Ferone, R., Singer, S.C., Hanlon, M.H. and Roland, S. (1983) Isolation and characterization of an *Escherichia coli* mutant affected in dihydrofolate and folylpolyglutamate-synthetase. In: Chemistry and Biology of Pteridines (Blair, J.A., Ed.), pp. 585–589. Walter de Gruyter, Berlin.

[24] Pyne, C. and Bognar, A.L. (1992) Replacement of the *folC* gene, encoding folylpolyglutamate synthetase-dihydrofolate synthetase in *Escherichia coli*, with genes mutagenized in vitro. J. Bacteriol. 174, 1750–1759.

[25] Schwartz, D., Alijah, R., Nußbaumer, B., Pelzer, S. and Wohlleben, W. (1996) The peptide synthetase gene *phsA* from *Streptomyces viridochromogenes* is not juxtaposed with other genes involved in nonribosomal biosynthesis of peptides. Appl. Environ. Microbiol. 61, 570–577.