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 Süssstrunk 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

Received 21 August 1997; revised 24 November 1997; accepted 28 November 1997

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.

Keywords: Folate metabolism; Folylpolyglutamate synthetase; Nucleoside diphosphate kinase; *Streptomyces*

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] FocC 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 BamHI 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 Sau3AI digested* S. coelicolor* M145 genomic DNA into the vector λ-FixII (STRATAGENE GmbH, Heidelberg, Germany). One phage containing a 3.6-kb BamHI fragment was identified which specifically hybridized against

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial strains and plasmids used in this study</td>
</tr>
<tr>
<td><strong>Strain</strong></td>
</tr>
<tr>
<td>S. coelicolor M145</td>
</tr>
<tr>
<td>E. coli XL1-Blue</td>
</tr>
<tr>
<td>E. coli SF4</td>
</tr>
<tr>
<td>Plasmid</td>
</tr>
<tr>
<td>pK18/pK19</td>
</tr>
<tr>
<td>pGM9</td>
</tr>
<tr>
<td>pAB12/pAB21</td>
</tr>
<tr>
<td>pAB3</td>
</tr>
</tbody>
</table>

*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 PstI-BamHI fragment at one end of the 3.6-kb BamHI 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 BamHI 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 BamHI 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 GAGGGA (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 (GAGGAG; 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 2637–2642). There are several possible functions 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 (legend on p. 285)**

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.
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 itself 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 plc (pAB12) was able to grow on minimal medium with glucose.

To see the effect of additional copies of *Streptomyces* *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 *FspI* fragment of the *S. coelicolor* A3(2) *folC* gene was cloned into the *Sma*I cut vector pK19 [8] and subsequently inserted as an *EcoR*I/*HindIII* 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 10⁹ 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
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


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.

Acknowledgments

We are grateful to A. Bognar for providing the E.
coli strain SF4. Annette Rohling (Burger) acknowledges the receipt of a scholarship from the Studienstiftung des deutschen Volkes. We thank Günther Muth for supportive discussions and Dirk Franz Brolle for helpful comments on the manuscript. Part of this work was supported by a grant from Fond der Chemischen Industrie (No. 163607).

References


