Characterization of the gene for factor C, an extracellular signal protein involved in morphological differentiation of *Streptomyces griseus*

Zsuzsa Birkó,¹ Andrea Sümegi,¹ Andrea Vinnai,¹ Gilles van Wezel,³ Ferenc Szeszák,¹ Sándor Vitális,¹ Pál T. Szabó,² Zoltán Kele,² Tamás Janáky² and Sándor Bíró¹

Author for correspondence: Sándor Bíró. Tel: +36 52 416 531. Fax: +36 52 416 531.
e-mail: sbiro@jaguar.dote.hu

The gene encoding factor C (*facC*), an extracellular signal protein involved in cellular differentiation, was cloned from *Streptomyces griseus* 45H, and the complete nucleotide sequence was determined. The deduced amino acid sequence was confirmed by HPLC/electrospray ionization-mass spectrometry analysis. The full-length protein consists of 324 amino acids and has a predicted molecular mass of 34523 Da. The mature extracellular 286 amino acid protein (31038 Da) is probably produced by cleaving off a 38 amino acid secretion signal sequence. Southern hybridization detected *facC* in several other *Streptomyces* strains, but database searches failed to identify a protein with significant homology to factor C. Expression of *facC* from a low-copy-number vector in *S. griseus* 52-1 resulted in a phenotypic effect similar to that given by exogenously added factor C protein.

Keywords: extracellular signalling, cell differentiation, sporulation, autoregulator, *Streptomyces*

INTRODUCTION

Streptomycetes are Gram-positive soil bacteria. Their study is particularly attractive because their mycelial form of growth develops into spores, and the concomitant production of secondary metabolites includes many compounds of great medical and/or industrial importance. Extracellular regulatory molecules – also called autoregulators (Khokhlov, 1991) – play a key role in controlling cellular differentiation and secondary metabolism in streptomycetes (Horinouchi & Beppu, 1992).

Studies on morphological differentiation and its intimate connection to secondary metabolism in *Streptomyces coelicolor*, the prototype and most extensively studied *Streptomyces* strain, have recently been reviewed by Chater (1998). In this strain, continuously increasing collections of developmental mutants and the corresponding cloned genes are available.

In *Streptomyces griseus* the loss of aerial mycelium formation, streptomycin production and ability to produce the low-molecular-mass γ-butyrolactone A-factor are frequently associated. Aerial mycelium formation and antibiotic synthesis can be restored by the addition of A-factor at nanomolar concentrations (Khokhlov, 1991). Studies of the genetics of A-factor biosynthesis have shown that A-factor-dependent initiation of sporulation involves phosphorylation of two regulatory proteins by cognate membrane-bound protein kinases (Horinouchi, 1996) and have also led to an understanding of the regulatory cascades involved in antibiotic biosynthesis (Horinouchi & Beppu, 1992).

Factor C was detected and later isolated as a protein from the culture fluid of *S. griseus* 45H (Szabó et al., 1962), a strain that readily sporulates in liquid medium (Szabó et al., 1961). Factor C induced the formation of preconidia in liquid cultures of the susceptible *S. griseus*.
strain S2-1 (Vitális & Szabó, 1969), at concentrations as low as 0.5–1 ng ml⁻¹ (Szeszák et al., 1991). This strain is otherwise blocked in submerged sporulation (Szabó et al., 1961). Factor C also stimulated sporulation on solid medium (Szabó et al., 1995). The pattern of proteins present in S. griseus S2-1 mycelium showed a specific change after the administration of factor C (Vitális et al., 1988). The majority or all of the factor-C-like antigen was released into the fermentation liquid of Streptomyces cultures (Szeszák et al., 1990, 1991).

Factor C has been purified to electrophoretic homogeneity by phosphocellulose and DNA-agarose chromatography. Its molecular mass was estimated to be about 34 kDa (Biró et al., 1980), and its isoelectric point is around 9.9 (Szeszák et al., 1991). By using polyclonal and monoclonal antibodies raised against factor C in conjunction with ELISA and immunoblotting techniques, the factor-C-like antigen was detected in 23 Streptomyces strains, in Bacillus subtilis, Escherichia coli and an archaean, as well as in mammalian cells. On this evidence, factor C was assumed to be evolutionarily conserved.

Preliminary experiments aimed at examining the possible mechanism of action of factor C detected an effect on transcription (Szeszák & Szabó, 1973, Szabó et al., 1984), on the release of potassium (Szesza et al., 1995). The pattern of proteins present in S. griseus 52-1 protoplasts were carried out by standard methods (Hopwood et al., 1961) or tryptone-soya broth (Szabó et al., 1961) or tryptone-soya broth (Hopwood et al., 1985) medium for 48 or 72 h on a rotary shaker at 28 °C. S. griseus S2-1 transformants were grown on R2YE medium (Hopwood et al., 1985) supplemented with 50 µg thiotreestron ml⁻¹.

In the present paper the cloning, sequence analysis and expression of the cloned factor C gene, facC, are described.

**METHODS**

**Strains and growth conditions.** E. coli strains JM109 (Messing et al., 1981) and XL-1 Blue (Stratagene) were hosts for pUC19 (Yanisch-Perron et al., 1985) and pBluescript II KS+ (Stratagene), respectively, and were grown for 16 h at 37 °C in Luria–Bertani (LB) medium supplemented with 100 µg ampicillin ml⁻¹. Submerged cultures of Streptomyces strains were grown in filtered soybean (Szabó et al., 1961) or tryptone-soya broth (Hopwood et al., 1985) medium for 48 or 72 h on a rotary shaker at 28 °C. S. griseus 52-1 transformants were grown on R2YE medium (Hopwood et al., 1985) supplemented with 50 µg thiotreestron ml⁻¹.

**N-terminal amino acid sequencing.** Factor C protein was blotted electrophoretically to PVDF membrane and sequenced by the Edman degradation method.

**Molecular cloning and DNA manipulations.** Purification of Streptomyces chromosomal DNA and transformation of S. griseus 52-1 protoplasts were carried out by standard methods (Hopwood et al., 1985). Plasmid DNA was isolated from E. coli strains using the techniques described by Sambrook et al. (1989), except if required for DNA sequencing, when Wizard miniprep columns (Promega) were used. Restriction enzymes were used according to the instructions of the suppliers. To construct the mini gene-library of S. griseus 45H, chromosomal DNA was digested with SacII. The DNA fragments were separated electrophoretically on a 1% agarose gel in TAE buffer and transferred to a Hybond-N membrane (Amersham). The membrane was hybridized with the 5'-digoxigenin (DIG)-labelled 39-mer oligonucleotide (TIB MOLBIOL Syntheselabor; Table 1) in 5 × SSC, 2% (w/v) blocking reagent, 0.1% (w/v) N-lauroylsarcosine, 0.02% (w/v) SDS, and 50% (v/v) formamide at 42 °C, overnight. At highest stringency the membrane was washed in 0.5 × SSC, 0.1% (w/v) SDS at 55 °C for 30 min. Hybridizing bands were detected by immunostaining using a DIG nucleic acid detection kit (Boehringer Mannheim). On the basis of the hybridization signal, the SacII fragments in the 2-8 kb region from a parallel gel were purified using a DNA recovery purification kit (Hybaid). The purified DNA fragments were cloned in SacII-digested pBluescript II KS+ (Stratagene) and the recombinant vector was used to transform E. coli XL-1 Blue. Ampicillin-resistant white colonies were screened by colony hybridization with the 39-mer DIG-labelled probe as described above.

**PCR amplification of the 5' end of the gene.** The 5' end of the coding region of the factor C gene was amplified from SacI-digested chromosomal DNA of S. griseus 45H: 100 ng DNA, 2-5 U Taq DNA polymerase (Amersham), 100 pmol degenerate synthetic oligonucleotide primers (Table 1), and 30 µM dNTP were reacted in an amplification protocol consisting of 1 cycle at 94 °C (5 min), 50 °C (2 min), 72 °C (2 min); 8 cycles at 94 °C (1 min), 50 °C (2 min), 72 °C (1 min); 50 cycles at 92 °C (1 min), 50 °C (1 min), 72 °C (1 min); 1 cycle at 94 °C (1 min) 50 °C (2 min), 72 °C (5 min). The amplified 76 bp DNA fragment was cloned into the SmaI site of pUC19 to form p8B and sequenced to verify amplification of the desired DNA fragment.

**DNA sequencing.** Plasmids were sequenced either by the dideoxy chain-termination method (Sanger et al., 1977) using the T7 Sequencing Kit (Pharmacia Biotech) and [32P]dCTP as the labelled nucleotide, or with an ABI 373 automated sequencer (Applied Biosystems) using a dye terminator cycle sequencing kit with AmpliTaq DNA polymerase FS (Applied Biosystems). In all sequencing reactions, double-stranded plasmids were used with universal or specific oligonucleotides (Table 1) as primers. Specific oligonucleotide primers were synthesized by Pharmacia Biotech. All sequence information reported was obtained from the independent analysis of both strands.

**Computer analysis of nucleotide and amino acid sequence.** Sequence data were analysed with the PC/Gene nucleic acid and protein sequence analysis software system (Intelli-Genetics). For coding region analysis, FramePlot Version 2.2.1 (based on Bibb et al., 1984, developed by Ishikawa Jun) was also used. Amino acid sequence homology searches made use of the BLAST network service and non-redundant protein sequence databases (Altschul et al., 1997).

**Molecular mass and sequence determination of tryptic fragments.** Gel bands containing about 13 µg protein from SDS-PAGE were cut into 1 mm³ pieces and soaked in 400 µl 100 mM NH₄HCO₃ (pH 8) with stirring for 10 min. The supernatant was discarded and the gel slices were dehydrated in 800 µl acetone. The protein within the gel slices was reduced with 300 µl 10 mM dithiothreitol at 56 °C for 1 h and repeatedly washed with NH₄HCO₃ and acetone. For alkylation, 300 µl 55 mM iodoacetamide was added at room temperature for 45 min. The washing procedure was repeated three times and the gel slices were dried under vacuum. For tryptic digestion (18 h at 37 °C) the ratio of factor C to trypsin (w/w) was 10:1. The peptide fragments were then washed.
with 400 µl NH₄HCO₃ solution and extracted with a 1:1 (v/v) acetonitrile/water mixture containing 5% (v/v) formic acid, concentrated in a Speedvac Evaporator, and dissolved for HPLC-MS analysis in 1% (v/v) formamide, 2% acetic acid. The peptide fragments were separated using a 0.32 × 200 mm fused silica capillary column (Novotny, 1988) packed with Vydac C₁₈. For mass spectrometric measurements a Finnigan TSQ 7000 tandem mass spectrometer equipped with a microelectrospray ion source (Kele et al., 1998) was used. The amino acid sequence of the most intense peptide fragments was determined by HPLC-MS/MS measurements. Molecular masses obtained in HPLC-MS experiments were compared with theoretical tryptic digestion profiles of candidate proteins generated by a computer program (Mann et al., 1993).

**Southern hybridization.** Chromosomal DNA samples from 12 *Streptomyces* strains were digested to completion with SacII. The DNA fragments were size-fractionated by gel electrophoresis on a 1% (w/v) TAE agarose gel and transferred to a Hybond-N (Amersham) membrane using the manufacturer’s instructions. Hybridization was carried out at 42 °C in a mixture containing 50% (v/v) formamide, 2 × SSC, 5 × Denhardt’s solution, 0.2 mg denatured salmon sperm DNA ml⁻¹ and 0.1% (w/v) SDS. The highest stringency washing was in 0.1 × SSC, 0.1% (w/v) SDS at 65 °C for 30 min. The 860 bp EcoRV–SalI fragment (nucleotide positions 930–1790 in Fig. 1), was used as the probe. The DNA was radio-labelled by random priming with nonamers and the Klenow fragment of *E. coli* DNA polymerase I, using a Megaprime Kit from Amersham to incorporate [α-³²P]dCTP.

**Construction of pSGF4, a low-copy-number vector containing *facC*.** The 2.1 kb *ApaI* fragment from pBZ3, harbouring the entire factor C structural gene and approximately 560 bp of upstream region, was inserted into *SmaI*-digested pJ2925 (Jansen & Bibb, 1993) lacking the BamHI site, after filling in the 5′ protruding ends of the insert with the Klenow fragment of DNA polymerase I and dNTPs. The resulting vector was designated pSGF2. To introduce *facC* into *S. griseus* 52-1 we used pHJL401, an *E. coli–Streptomyces* shuttle vector containing the *E. coli* pUC19 and Streptomyces SCP2⁺ (Lydiate et al., 1985) origins of replication (Larson & Herschberger, 1986) giving 100–200 copies per cell in *E. coli* and approximately 10 copies per chromosome in *Streptomyces*. For this purpose, the 2.1 kb *EcoRI–HindIII* insert of pSGF2 was cloned into pHJL401 digested with the same enzymes, resulting in pSGF4.

**RESULTS AND DISCUSSION**

**N-terminal amino acid sequence of factor C and PCR amplification of the 5′ end of the gene, *facC***

N-terminal amino acid sequencing of the purified, electrophoretically homogeneous factor C protein (Biró et al., 1980) resulted in the following tentative order for the first 29 amino acids: Ala-Val-Pro-Ala-Thr-Lys-Arg-Phe-Ser-Leu-Thr-Glu-Pro-Ser-His-(Asp/Phe)-Leu-Phe-Arg-His-Ala-Lys-Leu-His-Asp-(Gly/Ala)-Arg-Val-Gln. The residues at positions 16 (Asp/Phe) and 26 Gly/Ala) were not unambiguously identified, and were not used in primer constructions. The underlined amino acids were chosen to design two PCR primers. Since all glycine and alanine codons begin with G, the first base of the amino acid sequence was unambiguously determined as G for the first base of the ambiguous Gly/Ala was also included in the reverse primer.
primer. The forward and reverse primers contained five and two degenerate positions, respectively (Table 1), designed according to the biased codon usage of Streptomyces strains (Bibb et al., 1984). All but one triplet of the chromosomal gene, corresponding to the degenerate positions of the primers, as expected, turned out to contain G or C in the third position. The chromosomal triplet of the second amino acid (valine) of the forward primer is GTA but this mismatch did not affect the usefulness of the primer. In the PCR reaction SacI-digested S. griseus 45H chromosomal DNA was used as the template. Three short PCR products were amplified (approx. 38, 76 and 152 bp; data not shown). Amplification of a 76 bp fragment was expected, based on the N-terminal sequence. That fragment was extracted from the gel, subcloned into pUC19 and sequenced. Sequence analysis verified that the DNA fragment indeed corresponded to the known N-terminal amino acid sequence.

Cloning of facC

Instead of screening a complete chromosomal gene library, we used a ‘mini-genebank’, which was constructed in pBluescript II KS+ by cloning the gel-purified SacII fragments of about 2.8 kb in size, corresponding to the signal obtained by Southern hybridization. The 39-mer oligonucleotide flanked by the two primers in the 76 bp amplified fragment was used as probe. Colony hybridization identified one clone (pBZ3) with a 2–9 kb insert that repeatedly gave a positive signal in high stringency hybridizations. The DNA sequence of this clone was determined.

Sequence analysis of pBZ3

Sequencing was started from primers complementary to the T3 and T7 promoters flanking the cloned DNA fragment in the vector. The 39-mer DIG-labelled oligo-
nucleotide served as an internal primer. Gaps were filled by directed sequencing from ‘walking primers’. Both strands of the DNA were sequenced at least twice. FramePlot and coding region analysis by pc/gene programs identified a 975 bp ORF, typical of Streptomyces, with 96.9% GC in the third letter position and an overall G+C content of 70.7 mol% (Fig. 2). The coding sequence starts with a GTG codon at nt 1007 and extends to the TGA stop codon at nt 1978 (Fig. 1). It encodes a protein of 324 amino acids with a predicted molecular mass of 34,523 Da. Computer analysis of this protein located a sequence Ser-Ala-Ala-Ala/Ala/Val-Pro-Ala that contains two potential secretory signal cleavage sites at the positions indicated by slashes. Since the known N-terminal sequence began with the sequence Ala-Val-Pro it follows that the factor C propeptide contains a 38 amino acid secretion signal conforming to the consensus for prokaryotic signal sequences. It has the usual three domains: a positively charged N domain, a hydrophobic H domain required to initiate translocation across the cytoplasmic membrane, and a C domain preceding the cleavage site (Pugsley, 1993). The signal peptidase cleavage site also conforms to the −3, −1 rules (von Heijne, 1986). Since factor C was identified and isolated from the culture fluid of S. griseus 45H, and most or all of the factor-C-like antigen was found in the fermentation liquid, secretion of the protein was predictable. The mature protein contained 286 amino acids and its molecular mass was calculated to be 31,038 Da, in good agreement with the 34,500 Da estimated by SDS-PAGE (Biró et al., 1980), especially if we consider that factor C is a strongly basic protein. Its calculated isoelectric point was 9.59, which is close to our previous estimate of 9.9 (Szeszák et al., 1991).

When the N-terminal 29 amino acid sequence was compared to known sequences in the databases, significant homology to several zinc-finger-type regulatory proteins was detected. It was also known that factor C binds to single- and double-stranded DNA, and that its cytomorphological effect can be increased by low concentrations of zinc (S. Biró and others, unpublished results). This line of evidence was further strengthened by the successful purification of factor C on a zinc affinity column (Szeszák et al., 1997). From these results we believed the factor C to be a zinc-finger-type regulatory protein. However, the deduced complete amino acid sequence of the protein did not support our previous assumption. Interestingly, a hypothetical 1 nt insertion into the sequence at nt 1211 would result in an alternative similarly sized ORF whose product would contain about 20 His residues, most of them positioned regularly, 5–6 amino acids apart. FramePlot analysis of this alternative ORF did not show a typical Streptomyces coding sequence, but the difficulties in sequencing high-G+C-content Streptomyces DNA prompted us to determine the amino acid sequence of factor C by independent methods. Therefore, we determined the molecular mass of the protein and of tryptic digests by HPLC/ESI-MS/MS techniques.

**Molecular mass determination of factor C and amino acid sequencing of its tryptic fragments**

Only a small amount of factor C is produced by S. griseus 45H. Typically a few micrograms of purified protein can be obtained from 10 litres of culture broth. Therefore we used capillary chromatography/microelectrospray mass spectrometry, a method that required only minute amounts of protein for molecular mass determination of factor C and sequence verification of tryptic fragments. The molecular mass obtained for the mature factor C was 31,045 ± 9 Da, very close to the value of 31,038 Da calculated from the sequence.

After tryptic digestion the most intense peptide peaks were selected for sequence analysis in the HPLC-MS/MS measurements. The peptides identified by MS/MS are underlined in Fig. 1. Although the identified peptide fragments do not cover the whole sequence, the fragments are well distributed over the N-terminal, middle and C-terminal regions. Therefore, we concluded that the ORF identified by FramePlot analysis and the deduced amino acid sequence of the factor C protein are correct. One possible explanation for our previous, firmly established finding that factor C binds very specifically to zinc affinity columns, is that binding occurs in association with another protein. To study this possibility and detect any other protein that eluted with, or later than, factor C, we are examining proteins in S. griseus 45H cultures.

**Factor C shows no significant homology to other proteins in the databases**

Comparison of the amino acid sequence of the mature factor C protein with proteins in the databases using BLAST Search (Altschul et al., 1997) showed a low level of
The lane with DNA from expected, a strong hybridization signal was observed in results) and about 80% of the coding region. As digests of 12 strains from 10 species were hybridized Streptomyces strains, Southern blots of chromosomal sporulate in submerged culture to test for the presence of the factor C gene in other strains that was designated pSGF4. It is present at approximately 10 copies per chromosome in Streptomyces. Preliminary transcrip-
tional analysis revealed that the 560 bp upstream region contains the facC promoter (Zs. Birkó and others, unpublished results), allowing expression of facC from its own promoter. In submerged culture, similarity to teichoic acid biosynthesis protein C (tagC) of B. subtilis 168 (29% amino acid identity with 6 gaps). This seemed at first to be an interesting homology, because teichoic acid is a cell wall component of Gram-positive bacteria and expression of the tag genes is sporulation specific (Mauel et al., 1991, 1994). However, recent reports have shown that tagC corresponds to dinC (Cheo et al., 1991, 1993), and thus belongs to the SOS regulon. Therefore, the relevance of the low similarity between factor C and the TagC protein, if any, is unclear.

facC is present in several Streptomyces strains that sporulate in submerged culture

To test for the presence of the factor C gene in other Streptomyces strains, Southern blots of chromosomal digests of 12 strains from 10 species were hybridized with a 32P-labelled DNA fragment spanning nt positions 930–1790 (Fig. 1), covering approximately 75 bp of upstream region (Zs. Birkó and others, unpublished results) and about 80% of the coding region. As expected, a strong hybridization signal was observed in the lane with DNA from S. griseus 45H (Fig. 3). The lanes containing DNA from S. albus R-55, S. flavofungini and S. albus 391 each gave a strong signal, suggesting the presence of a gene with high homology to facC. No comparable band was observed with the DNA from any other strains tested, including our factor-C-sensitive test strain S. griseus 52-1. This contrasts with our previous results obtained using monoclonal antibody raised against factor C, which showed the presence of a factor-C-like antigen in all tested Streptomyces strains, including some of those that failed to hybridize in the present experiment (Szeszák et al., 1990). This might be explained by the high stringency of washing. In other experiments when the DIG-labelled 39-mer oligonucleotide (Table 1) was hybridized with chromosomal digests of the same strains, and a lower stringency washing was used, we could detect a hybridizing band or bands in all tested strains (S. Biró, unpublished results). The meaning and significance of these results await further clarification but our findings are reminiscent of those of Onaka et al. (1998), who found two low homologues of the S. griseus A-factor receptor protein, which has a regulatory role in secondary metabolism and morphogenesis, in S. coelicolor. The four Streptomyces species shown to harbour a high homologue of facC are known to sporulate in submerged culture (Kendrick & Ensign, 1983; Vitalis et al., 1981; Daza et al., 1989; S. Biró, unpublished results). This points to a possible involvement of factor C in the onset of submerged sporulation.

Recently ssgA has been recognized as a gene involved in submerged sporulation of S. griseus B2682 (Kawamoto & Ensign, 1995; Kawamoto et al., 1997). Like factor C, the ssgA gene product is a regulatory protein that shows no relevant similarity to any known protein. Factor C and the intracellular ssgA gene product might form parts of a regulatory mechanism unique to those Streptomyces species able to sporulate in submerged culture.

Expression of facC in S. griseus 52-1

To study the phenotypic effect of the cloned gene, an E. coli–Streptomyces SCP2*-based shuttle vector containing the entire structural gene for factor C and approximately 560 bp of upstream region was constructed; it was designated pSGF4. It is present at approximately 10 copies per chromosome in Streptomyces. Preliminary transcriptional analysis revealed that the 560 bp upstream region contains the facC promoter (Zs. Birkó and others, unpublished results), allowing expression of facC from its own promoter. In submerged culture,
under our standard test conditions, \textit{S. griseus} 52-1 forms long, smooth, non-branching vegetative hyphae. It fails to sporulate and lacks the reproductive, frequently branching hyphae containing club-like thickening at their ends (Fig. 4b), characteristic of the factor C producer \textit{S. griseus} 45H (Fig. 4a; Vitális \textit{et al.}, 1963). Addition of purified factor C to submerged cultures of \textit{S. griseus} 52-1 resulted in a morphology very similar to that of \textit{S. griseus} 45H (Fig. 4c). Introducing pSGF4 into \textit{S. griseus} 52-1 resulted in a phenotype (Fig. 4d) almost identical to that of the \textit{facC} strain \textit{S. griseus} 45H. The transformant showed this typical cytomorphology irrespective whether it was grown in the presence or absence of thiostrepton.

In summary, we believe that factor C plays an essential role in the development of \textit{S. griseus} and is probably involved in the onset of sporulation in submerged culture.

**ACKNOWLEDGEMENTS**

We are grateful to Anikó Pay for the automated DNA sequencing, to Wolfgang Piepersberg for the N-terminal protein sequencing and to János Aradi for synthesis of the PCR primers. Andrea Kissné and Katalin Szabó are thanked for technical assistance. This work was supported by research grants T 016796 and T 025553 by the National Scientific and Research Fund (OTKA) and by the Hungarian Academy of Sciences (to S.B.), and also by research grants F 025602 by OTKA (to P.T.Sz.) and T 026624 by OTKA (to T.J.). Zs.B. was the recipient of a short term TEMPUS PHARE JEP student mobility grant.

**REFERENCES**


Janssen, G. R. & Bibb, M. J. (1993). Derivatives of pUC18 that have BglII sites flanking a modified multiple cloning site and that retain the ability to identify recombinant clones by visual screening of Escherichia coli colonies. Gene 124, 133–134.


Factor C from *Streptomyces griseus*


Received 5 March 1999; revised 13 May 1999; accepted 1 June 1999.