Volume 15 Number 19 1987 Nucleic Acids Research

Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: nucleotide sequence of three genes and analysis of transcriptional activity

Jürgen Distler, Andrea Ebert, Kambiz Mansouri, Klaus Pissowotzki, Michael Stockmann and Wolfgang Piepersberg*

Institut für Organische Chemie und Biochemie der Technischen Hochschule Darmstadt, Petersenstrasse 22, D-6100 Darmstadt, FRG

Received July 15, 1987; Revised and Accepted September 4, 1987 Accession no. Y00459

ABSTRACT

Three streptomycin (SM) production genes from Streptomyces griseus clustered around aphD, the major resistance gene, have been sequenced: strB, coding for an aminocyclitol amidinotransferase, ORF5 (strR), a putative regulatory gene, and ORF1 (strD), possibly coding for a hexose nucleotidylating enzyme. Three promoters and at least five, partially overlapping, transcripts have been identified by S1 mapping and Northern blot experiments. aphD, the resistance gene, is transcribed from two promoters. One of them, located inside the strR gene, seems to be constitutive and the other is switched on later in the growth phase. The late transcripts cover the resistance gene (aphD) and a regulatory gene (strR) which controls the expression of strB.

INTRODUCTION

The production of secondary metabolites such as aminoglycoside antibiotics in streptomycetes is only little understood, both biochemically and genetically. Of special interest are the organization and the regulation of the genes involved in biosynthesis. As a model system we are studying the genetics of streptomycin (SM) production in Streptomyces griseus, which has been analysed in part on the enzymatic level only (1,2). Genes for production of SM and hydroxy-SM as well as two different resistance genes were cloned from various strains of S. griseus, S. bikiniensis and S. glaucescens by direct cloning into another host or by complementation of mutants (3-9). Recently the nucleotide sequences of the major resistance determinant, the genes for SM-6-phosphotransferase (APH(6)), aphD (sph), have been analysed from both S. griseus and S. glaucescens (10-12). Downstream and upstream of the aphD gene at least four further genes involved in biosynthesis of various intermediates of the

© IRL Press Limited, Oxford, England.

SM pathway were identified in S. griseus (4,7). Divergent orientation of transcription upstream of the aphD gene was suggested by preliminary sequence analysis (10) in an area which could be involved in regulation of SM production. Confirmation of this phenomenon by transcriptional studies are reported herein. Three further putative genes have been identified clustered around the aphD resistance gene.

MATERIALS AND METHODS

Bacteria, Plasmids, Growth Conditions, and Transformation.

Strains of S. lividans, S. griseus, Escherichia coli and conditions for their cultivation have been published previously (4,10, 13,14). Streptomycin non-producing (smi) mutants M66 and M67 (15) as well as an A-factor negative mutant M852 have been obtained from S. griseus N2-3-11 after nitrosoguanidine or UV light mutagenesis, respectively. The recombinant plasmids used were pIJ424 and pIJ702 (13), pJDM10 and pJDM40 (4,10), pUC18 and pUC19 (16). Minimal inhibitory concentrations (MIC) for kanamycin were tested on SMA medium (3). Transformation conditions for S. lividans were as given (13), whereas for S. griseus N2-3-11 and its derivatives the method described earlier (4) was used.

Preparation, Manipulation, and Sequencing of DNA.

Plasmid DNA was prepared according to published methods (13,14). Restriction, ligation, nick-translation, and end-labelling of DNA was carried out as proposed by the suppliers of the enzymes. DNA was sequenced basically by use of the enzymatic method (17) with the modifications described earlier (10). In some cases of ambiguous results and for S1 mapping the chemical method (18) was employed.

Preparation and Analysis of RNA.

Total cellular RNA was prepared from mycelia (19). Northern blotting experiments were carried out as described (11,14) after separation of 20 µg RNA of each preparation on formaldehyde gels (14). As molecular weight standards bacterial and yeast ribosomal RNAs were used. S1 mapping was done exactly as given in a standard protocol (20).

Enzymatic Tests.

The kanamycin-3'-phosphotransferase expressed from promoter probe

plasmid pIJ424 and the APH(6) enzyme were tested by the radioisotope assay (21). The aminocyclitol amidinotransferase was assayed by use of the hydroxylamine according to a published procedure (1). Enzymes were generally tested in ribosome-free extracts from cells disintegrated by a French pressure cell press (Aminco) in a standard buffer (50 mM Tris/HCl, 10 mM $MgCl_2$, 7 mM mercaptoethanol, pH 8.0).

RESULTS

Analysis of DNA Sequence. Up to now a segment of more than 5.5 kb length has been sequenced from two genomic clones, pJDM10 and pJDM40, of S. griseus DNA by use of the strategy outlined in Fig.1. The sequence shown in Fig. 2 includes the region common to pJDM10 and pJDM40 with aphD, the major SM resistance gene (10). Three large open reading frames which show the codon bias typical for Streptomyces DNA (22) are indicated in Fig. 2. Two of the open reading frames are upstream and one is downstream of aphD (see Fig. 7 for a summary diagram). The downstream gene

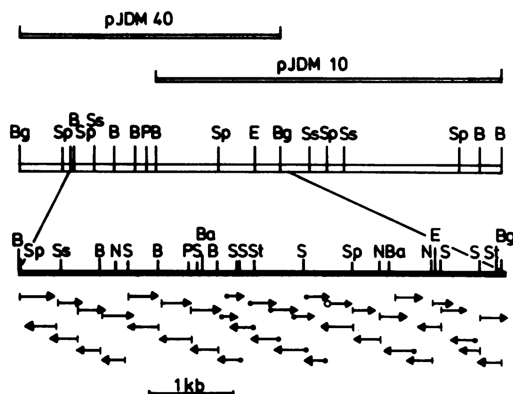


Fig.1. Sequencing strategy for a cluster of SM genes from S. griseus. Nucleotide sequences were determined from plasmids pJDM10 and pJDM40 (4,10) by use of the dideoxy/double strand-method from given restriction sites or on Bal31-shortened fragments cloned in pUC vectors, symbolized by arrows starting from cross-lines or dots, respectively. The arrow starting from an open circle marks a fragment sequenced from an internal oligonucleotide primer. Restriction sites are labelled B = BamHI, Ba = BalI, Bg = BglII, E = EcoRI, N = NruI, P = PstI, S = SalI, Sp = SphI, Ss = SstI, St = StuI.

```
                                                                                4                                                          63
                                                                             GGGA GGGATCGGATCGGTTCGACGGAGGGGGACCGGTGGTGTCAGGTGGGGGTTCGGCGGTG
                                                                             CCCT CCCTAGCCTAGCCAAGCTGCCCTCCCCCTGGCCACCACAGTCCACCCCCAAGCCGCCAC

Stop Ser Ser Ile Gln Ala Arg Ser His Asp Gly Leu Val Leu Arg His Ala His Pro Val Glu Pro Ala Ser Thr Val Gln Val His Arg
 AGT CCT CCT CTA GAC GCG GGA CGA CAC CAG CGG CTC GTG CTC GGC CAC ACG CAC TCC GTG GAG CCC GCG CCT CCA GTG GAC CTG CAC GGA
          Bgl II                                         108                                                         153
326
Gly Ile Leu Ser Ala Glu Ile Arg Arg Val Gly Ser Ile Ser Ala Gly Arg Leu Val Ile Ser Phe Glu Val Glu Ser Asp Glu Val Val
CGG CTA CTC ACT CCG GAG CTA CGC TGC GTG GGG CCT CTA GCT GCG TGG CGC GTC GTG CTA CCT CTT GAG CTG GAG CGA CAG GAG CTG CTG
             Stu I                                       198                                                         243
296
Cys Asp Glu Ala Leu Ser Thr Phe Pro Gly Val Tyr Ser Asn Thr Val Arg Thr Gly Ala Gly Ile Val Thr Pro Gly Met Val Arg Ser
TGT CAG GAG CCG CTC CCT TCA CTT TCC AGG GTG TAT CCT CAA CCA CTG CGC CCA CGG CCG CGG CTA GTG CCA GCC CGG GTA GTG GGC CCT
                                                         288                                                   Sma I 333
266
Asn Arg Val Glu Ala Gly Glu Glu Val Leu Val Arg Gly Ile Leu Asp Ser Arg Glu Asp Val Leu Gly Asp Cys Arg Pro Glu Thr Thr
CAA CGC GTG GAG CCG CCG GAG GAG CTG GTC CTG GGC CGG CTA CTC TAG CGA GGC AAG CAG GTG GTC CGG CAG TGT CGC GCC GAG CCA GCA
                                                         378                                                         423
236
Glu Leu Val Leu Arg Asn Val Glu Leu Met Asp Thr Val Asn Gly Thr Asp Lys Trp Tyr Gly Ser Ile Val Thr Ser Arg Val Asp Arg
GAG GTC GTG CTC CGC CAA GTG GAG GTC GTA CAG CCA GTG TAA CGG CCA CAG GAA GGT CAT CGG CCT CTA GTG CCA CCT CGC GTG CAG GGC
                                                         468                                                         513
206
Gly Ala Asp Ile Leu Trp Gln Val Ala Asp Thr Ile Glu Leu Glu Gly Arg Trp Ser Pro Thr Ile Ala Ala Val Ala Glu His Ile Ala
CGG CCG CAG CTA GTC GGT GAC CTG CCG CAG CCA CTA GAG GTC AAG GGG CGC GGT GCT CCC GCA CTA GCG GCG GTG CCG GAG CAC CTA GCG
                                                         558                                                         603
176
Pro Ser Phe Leu Tyr Val Gly Val Leu Ala Leu Asp Ser Lys Pro His Ala Pro Lys Glu Glu Leu Gly Leu Val Gln Gly Ser Asp Ser
CCC ACT CTT CTC CAT CTG CGG CTG GTC TCG CTC CAG CGA GAA CCC CAC CCG GCC GAA GAG GAG GTC CGG CTC GTG GAC GGG GCT CAG CGA
                                                         648                                                         693
146
Leu Glu Ala Val Gly Phe Ser Arg Pro Glu Pro Val Arg Thr Leu Met Leu His Ala Asp Pro Arg Ala Ala Arg Phe Glu Arg Val Ser
GTC GAG CCG GTG CGG CTT CCT GGC CCC GAG CCC GTG GGC CCA CTC GTA GTC CAC GCG CAG GCC GGC CCG CCG GGC CTT AAG GGA GTG GCT
                                                 Sma I   738                                           EcoR I        783
116
Asp Glu Val Val Gly Val Val Phe Asn Asp Gly Leu Tyr Met Ile Phe Asp Asp Glu Gly Leu Phe Asp Arg Ser Ile Leu Val Cys His
CAG GAG CTG CTG CGG GTG GTG TTT TAA CAG CGG CTC TAT GTA CTA CTT CAG CAG GAG CGG GTC CTT TAG CGC TCT TTA GTC ATG CGT CAC
                                                         828                                   Nru I                 873
 86
Ala Leu Gly Leu Pro Lys Ser Gln Pro Ile Tyr Ser Val Lys Leu Gly Phe Arg Ser Gly Asp Gly Val Ala Ala Val Ile Glu Asp Ala
GCG CTC GGG GTC TCC AAA GCT GAC GCC TTA CAT CGA GTG AAA GTC AGG CTT AGA GCT TGG TAG AGG GTG GCG CCG CTG CTA AAG CAG CCG
                                                         918                                                         963
 56
Thr Asp Gly Val Val Ile Gly Val Asp Ile Ile Gly Ala Ala Arg Ile Ala Glu Leu Gly Tyr Phe Leu Val Pro Lys Asn Ala Val Pro
CCA CAG CGG GTG CTG CTA CGG CTG CAG CTA CTA GGG CCG CCG TGC ATA CCG GAG GTC GGG TAT CTT TTC CTG TCC AAA TAA CCG CTG GCC
                                                        1008                                                        1053
 26                                                                                                            1    1146
Val Leu Gln Lys Ala Ser Thr His Thr Ile Pro Arg Leu Arg Thr Gly Thr Gly Gly Ala Leu Val Leu Ala Lys Met GTGGTGCTCCTCGCG
TTG GTT AAC AAA GCG TCT GCA CAC CCA CTA CCC GGA CTC CGC CCA AGG ACA CGG CGG ACG GTC ATG CTC GCG AAA GTA CACCACGAGGAGCGC
                                                        1098                                                   sd
                                                        1205                                  <—— ORF 1             1264
TGGTCTTGGGCCGGGCCGCCGTGGGCGAGTGGTTCTTCCGAACAGTTATCGAAGAGAAT CAGCCGCCGTGGCGCGATCCTGTGCATCCGTGTAAGGGGCGCTGCAAGTTGCATGTATT
ACCAGAACCCGGCCCGGCGGCACCCGCTCACCAAGAAGGCTTGTCAATAGCTTCTCTTA GTCGGCGGCACCGCGCTAGGACACGTAGGCACATTCCCCGCGACGTTCAACGTACATAA

                             Bal I                      1323                                                        1382
CGGCAATCAAACTGCGGTTTATTTTGATGGCCAGAATGAAATCGCTGACAGGCGGTGCG TGATCGGGCGCAGGCGCTGGCTGGAGGGGGCCGTTCCGGTCCTCTCCGCCCTGGAGGAT
GCCGTTAGTTTGACGCCAAATAAAACTACCGGTCTTACTTTAGCGACTGTCCGCCACGC ACTAGCCCGCGTCCGCGACCGACCTCCCCCGGCAAGGCCAGGAGAGGCGGGACCTCCTA

                                                        1441       -35        aphD P1    -10 Nru I                  1500
GGGCCGGGCGGGGTGGCCCTCCTATTTCGAGGTCGCTTTTCGGTCATGCGGACAGCTTT ACTTGGCCGTTGCCCGGATGTCCGGGTGCTACTATTCGCGAAGTGCGAAAGCATTGCTC
CCCGGCCCGCCCCACCGGGAGGATAAAGCTCCAGCGAAAAGCCAGTACGCCTGTCGAAA TGAACCGGCAACGGGCCTACAGGCCCACGATGATAAGCGCTTCACGCTTTCGTAACGAG

                             sd              ORF 5 ——> 1552                                                         1597
TTTATCGCCGTGAAAGCCCGTGATTGAAGGGTGTGTA        ATG GAG CAT ATT TCA GGG AAC AGC CCC GAG CAA GTC CGT GAG AGG TCG GCG GCG GTG ACG
AAATAGCGGCACTTTCGGGCACTAACTTCCCACACAT        Met Glu His Ile Ser Gly Asn Ser Pro Glu Gln Val Arg Glu Arg Ser Ala Ala Val Thr
                                             1
                                                        1642                                                        1687
GGT GCG GTC GAG GAA AGT GAA CTC AAA TTA TCC GCC GTG ACA ATG GTC CCG GTC GAA TCA TTG CTT CCC TCC GAT TCC CCG CGC AGT GCC
Gly Ala Val Glu Glu Ser Glu Leu Lys Leu Ser Ala Val Thr Met Val Pro Val Glu Ser Leu Leu Pro Ser Asp Ser Pro Arg Ser Ala
 21
                                                        1732                                                        1777
GGC GAG GAT GTC GAG CAC ATC CGC ACG CTC GCC GCA TCC GGA GCG GAA TTG CCC GCT ATC GTG GTC ATG CCC ACC ACG AAG CGC GTC ATC
Gly Glu Asp Val Glu His Ile Arg Thr Leu Ala Ala Ser Gly Ala Glu Leu Pro Ala Ile Val Val Met Pro Thr Thr Lys Arg Val Ile
 51
     Sph I                                              1822                                                        1867
GAC GGC ATG CAC CGA CTG CGC GCC ACC AAA ATG CGC GGA GCC ACC GAG ATC GCC GTG CGG TAT TTC GAA GGC GGA GAG GAA GAG GCG TTC
Asp Gly Met His Arg Leu Arg Ala Thr Lys Met Arg Gly Ala Thr Glu Ile Ala Val Arg Tyr Phe Glu Gly Gly Glu Glu Glu Ala Phe
 81
                                                        1912                                                        1957
ATC TTC GCG GTG AAG TCC AAC GTC ACC CAC GGA CTG CCG CTC TCC CTC GAC GAC CGC AAG GCC GCG GCG ACC CGT GTC CTG GAG ACC CAT
Ile Phe Ala Val Lys Ser Asn Val Thr His Gly Leu Pro Leu Ser Leu Asp Asp Arg Lys Ala Ala Ala Thr Arg Val Leu Glu Thr His
111
                                                        2002                                                        2047
CCG TCC TGG TCC GAC CGG GCC ATC GGC CTG GCG ACC GGA CTG TCG GCG AAG ACG GTG GGG ACC CTC AGG TCC TGT TCG ACT GCC GGG GTT
Pro Ser Trp Ser Asp Arg Ala Ile Gly Leu Ala Thr Gly Leu Ser Ala Lys Thr Val Gly Thr Leu Arg Ser Cys Ser Thr Ala Gly Val
141
                                                        2092                                                        2137
CCG CAG TCG AAC GTG AGG ATC GGG AGG GAC GGG CGG GCC CGG CCG CTG GAC CCC ACC GAG GGG CGG AAG CTG GCG AGC CGG CTG CTC CAG
Pro Gln Ser Asn Val Arg Ile Gly Arg Asp Gly Arg Ala Arg Pro Leu Asp Pro Thr Glu Gly Arg Lys Leu Ala Ser Arg Leu Leu Gln
171
                                                        2182                                                        2227
GAG AAC CCC TCG GCG TCG CTG CGT CAG ATC GCC GCA CAG GCC GGC GTC TCC CCG AGC ACC GCC TCC GAC GTC CGC AAG CGG CTG AGC CGT
Glu Asn Pro Ser Ala Ser Leu Arg Gln Ile Ala Ala Gln Ala Gly Val Ser Pro Ser Thr Ala Ser Asp Val Arg Lys Arg Leu Ser Arg
201
                                                        2272                                           Sma I        2317
GGT GAG AGC CCG CTG CCC GAA CGC GAT CGT CAA CAG GAA GTG CCG GCC GTC GCC CGG ACT CCG GCC CGT GTC TCC CGG GCC GAC GGG AGT
Gly Glu Ser Pro Leu Pro Glu Arg Asp Arg Gln Gln Glu Val Pro Ala Val Ala Arg Thr Pro Ala Arg Val Ser Arg Ala Asp Gly Ser
231
                                                        2362                 -35          aphD P2        -10        2407
TGG GCG CCG CAC ACC GTG GCG CTG CGC CAC CTC AGC CGG GAC CCG TCC GTG CGG CTC ACC GAG GAC GGT CGG GCG CTG CTA CGT TGG CTG
Trp Ala Pro His Thr Val Ala Leu Arg His Leu Ser Arg Asp Pro Ser Val Arg Leu Thr Glu Asp Gly Arg Ala Leu Leu Arg Trp Leu
261
                                                        2452                                                        2497
AAC GTG GTG GCC GTG CGC AAC CAG GAC TGG GAC CGC CTC CTG GGC AAC GTC CCT CCG CAC TGC GTC AAG GTC ATA GCC GAG CTG GCC CGT
Asn Val Val Ala Val Arg Asn Gln Asp Trp Asp Arg Leu Leu Gly Asn Val Pro Pro His Cys Val Lys Val Ile Ala Glu Leu Ala Arg
291
                                                        2542                                                        2587
GGC TGT GCC GAC ATC TGG CAT CGG GTG GCG GAG GAA CTG GAC CAG GCC GGC ATC GAC GAG GCG GCG GGC CGG TCC TTG AGC GAT GTC GGA
Gly Cys Ala Asp Ile Trp His Arg Val Ala Glu Glu Leu Asp Gln Ala Gly Ile Asp Glu Ala Ala Gly Arg Ser Leu Ser Asp Val Gly
321
```

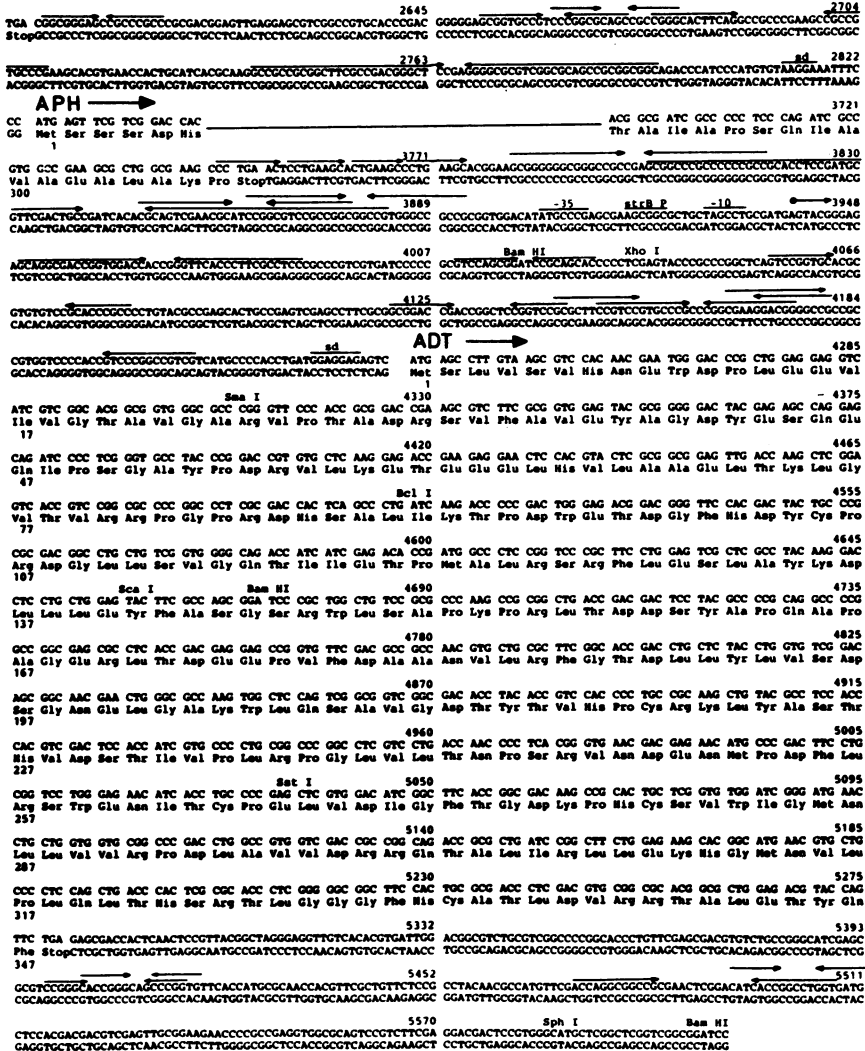


Fig.2. Nucleotide sequences adjacent to the aphD gene (APH) comprising three putative genes involved in SM production in S. griseus. Only the terminal portions of the already published aphD gene (10; positions 2825 to 3754) are shown. The arrows starting from a filled circle indicate start sites of transcription (cf. Fig.6). Other arrows represent significant direct or inverted repeat structures in the intercistronic regions or indicate the orientation of reading frames. Promoters are marked at their -35 and -10 regions (23); sd = possible translational initiation sites.

(strB) codes for a 38kDa aminocyclitol amidinotransferase (ADT), as was confirmed by N-terminal sequencing of the purified protein by Tohyama et al. (12). The upstream genes strR (ORF5 37.7 kDa) and strD (ORF1 32.6 kDa) are in a region where Ohnuki et al. (7) postulated a regulatory gene. Our experiments (see below) indicate that strR is the regulatory gene and sequence comparisons suggest that strD could be a nucleotidylating enzyme involved in late steps of the SM biosynthetic pathway. The 4.8 kb left BamHI-SstI fragment of pJDM10 (Fig. 1), when subcloned in pIJ702, was able to complement the defect in mutant M67 blocked in the later SM pathway, beyond the streptidin biosynthetic route. This suggests that strD could be the mutated gene. Several inverted or direct repeats with potential signal functions have been identified in the intercistronic regions (Fig. 2). Recently, the sequence of an overlapping DNA fragment starting with the SphI site in position 1782 (Fig. 2), derived from the same gene cluster in an obviously related strain, S. griseus ISP2536, has been reported (12). The DNA sequences derived from the two strains differ in 56 out of 3838 bases. Surprisingly, some of the differences affect the reading frames of APH(6) and ADT mainly as a result of pairs of frameshifts compensating each other within a few triplets distance. Because of this three peptides in APH(6) and one peptide in ADT are differing. Also, for ORF5 a completely different reading frame was postulated and the strong inverted repeat preceding APH(6) was not found by Tohyama et al. (12).

In Vivo Transcriptional Activities. The influence of various subfragments of pJDM40 on expression in S. lividans of the kanamycin resistance gene, aphA, when subcloned in promoter probe vector pIJ424 was tested. The results (Fig. 3) suggested presence of two promoter activities of intermediate strengths and with divergent orientation on the right 3.3 kb BamHI-BglII fragment, and of a strong promoter on the 0.68 kb BamHI fragment containing part of the aphD-strB intercistronic region (position 3749 to 4234 in Fig. 2) oriented towards the ADT gene. No significant transcriptional activity could be found at both ends of the left BglII-StuI fragment of pJDM40 (cf. pJDM43 and pJDM44 in Fig. 3).

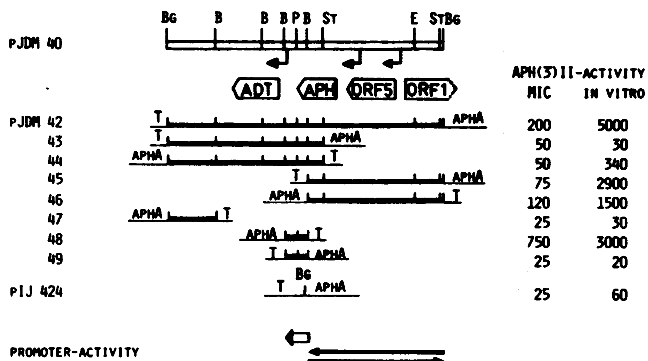
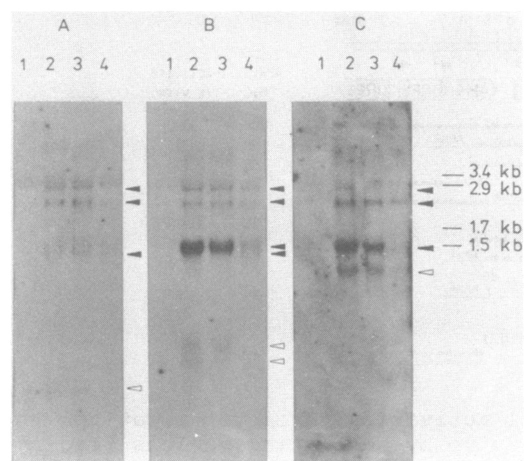


Fig.3. In vivo transcriptional activities. Promoter locations and activities were tested in vivo by subcloning various fragments from pJDM10 and pJDM40 in promoter probe vector pIJ424 between the fd terminator (T) and the kanamycin-3'-phosphotransferase gene (aphA). The expression of kanamycin resistance (MIC) and APH(3') enzymatic activity in S. lividans (CPM/µg protein) is given for each subclone. An interpretation of promoter strength and orientation is represented by arrows below. For orientation reading frames and transcriptional start sites are included (cf. Figs. 2,5).

Mapping of Transcripts and Transcriptional Start Points. The transcripts which could be identified by northern hybridization experiments with both restriction fragments and synthetic oligonucleotides as probes are shown in Fig. 4. Supposed that no site-specific processing occurred, this indicated a rather complicated pattern of RNA synthesis in the area of the genes coding for ORF (strR) and APH(6) (aphD). The results suggested that both genes could be transcribed independently on 1.4 kb and 1.5 kb messengers, respectively. The 2.4 and 2.8 kb RNA molecules contain both open reading frames and can presumably give rise to both proteins, APH(6) and the hypothetical strR gene product. All four transcripts go from left to right, as shown with the single-stranded oligonucleotide probes, none of which fell into an inverted repeat. Also, this implicated the existence of two promoters from which the aphD gene could be read and which both could contribute to the expression measured with plasmid pJDM46 (Fig. 3). S1 mapping of the 5' ends of RNAs in three areas confirmed this conclusion (Fig. 5) and enabled us to identify three



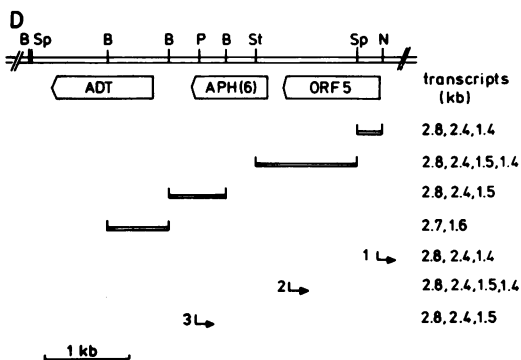


Fig.4. Mapping of transcripts. (A-C) Northern blot experiments with oligonucleotides 1 (A), 2 (B), and 3 (C) as probes (see below), and total RNA from S. griseus N2-3-11 strains M852 (1), M66 grown in absence (2) or presence (3) of 30 µg/ml of streptomycin, and wild-type (4). RNA size standards = yeast and bacterial rRNA; presumed primary transcripts = closed arrowheads; presumed degradation products in M66 = open arrowheads. (D) The lengths of the presumed primary transcripts found in all DNA/RNA hybridization experiments are summarized (cf. also Fig.7). Probes used were either restriction fragments (double lines), or synthetic oligonucleotides (30-mers), which started at the following sequence positions (cf. Fig.2): (1) 1594, (2) 2581, (3) 3735; arrows indicate the 5' to 3' direction.

transcriptional start sites (cf. Fig. 2), two upstream of aphD and one preceding the strB (ADT) gene. No transcripts were found in an A-factor negative mutant. In contrast strain M66, a mutant blocked in an early step of streptidine biosynthesis, gave the same sizes of transcripts as the wild-type strains but much more of the 1.5 kb transcript (from aphD P2) was produced. The amount of specific mRNA synthesised by strain M66 was independent of the presence or absence of SM. Additionally hybridizing diffusive signals in RNA preparations from M66 were presumed to be degradation products (Fig. 4 A-C), because they were localized in regions where no transcriptional start sites could be identified. Promoter aphD P1 seemed to be the preferred transcriptional start site for the APH(6) expression in the logarithmic growth phase, whereas aphD P2 was used preferentially in the

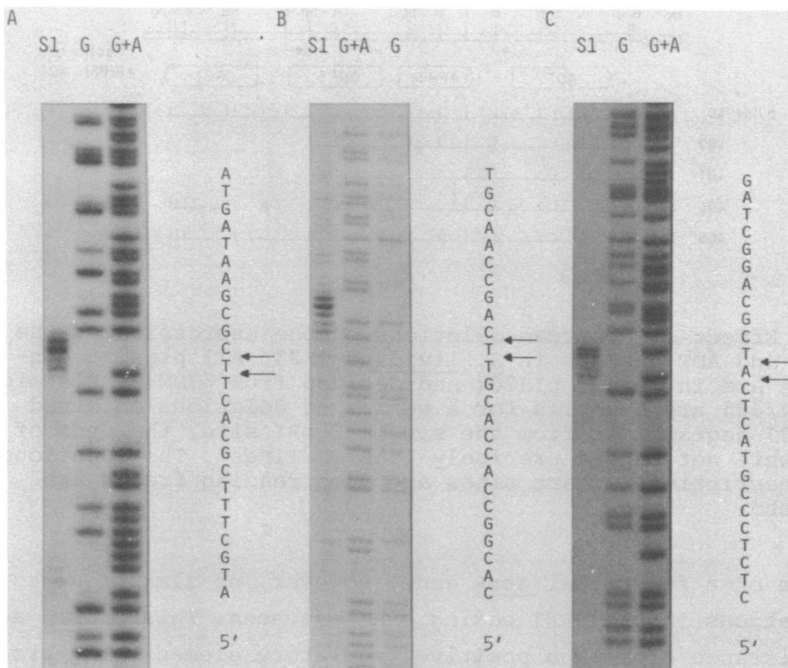


Fig.5. S1 mapping of 5' ends of transcripts started from the following promoters: (A) aphD P1 (B) aphD P2, (C) strB P. The following DNA fragments were used for S1 mapping and for chemical DNA sequencing (18; only 2 lanes are shown) (in brackets: sequence positions in Fig.2): (A) AvaI (1563) to EcoRI (769), (B) TaqI (2550) to TaqI (2054), (C) MluI (4063) to PstI (3660). The first site was the one end-labelled in each case.

stationary phase in wild-type S. griseus N2-3-11 and under all conditions in S. lividans containing a pJDM10 (not shown). Judging from their lengths the 2.4, 1.5 and 1.4 kb transcripts in the strR - aphD area are likely to end in the intercistronic sequence preceding and following the aphD gene, respectively. The 2.8 strR-aphD transcript seemed to end downstream strB P. This means that the termination signals are inefficient.

In Vivo Expression of ADT and APH(6). When enzymatic activities were measured, a dependence of expression in S. lividans of the strB (ADT) gene, but not of the aphD gene, was observed when part of ORF5 upstream the region of aphD P2 (see Fig. 2) was removed (Fig. 6): The ADT activity was completely abolished in

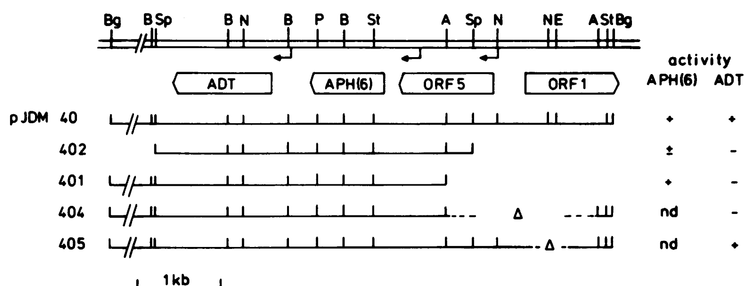


Fig.6. Effect of upstream deletions on the expression of the APH(6) and ADT enzymes in S. lividans TK23. All plasmid constructs are in vector pIJ702 and derived from pJDM40. Plasmids pJDM404/405 are examples for a series of deletions obtained by Bal31 degradation from the single EcoRI site, the ends of which were not mapped precisely (dashed lines). The locations of transcriptional start sites and open reading frames are indicated.

absence of a functional strR gene. However, no similar effect of deletions in the ORF1 coding area was seen. This proved again the existance of both a positive regulatory element, the product of the strR gene (7), which is probably identical with the one expressible from ORF5, and a second independ promoter activity (aphD P2) preceding the aphD gene. Also, the ORF1 gene for further use was identified by the symbol strD in accordance with an unifying nomenclature (7).

DISCUSSION

Out of the five genes and their products involved in SM resistance and production, aphD (strA; APH(6)), strB (ADT), strC, strR, smi-67, and the gene complementing mutant SD-1, which all had been localized by their phenotypes on contiguous segments of DNA in S. griseus and closely related species (4,7,8,10,12), only two have been identified so far by DNA and polypeptide sequencing: the major resistance gene (aphD; 10,12), coding for the APH(6) enzyme, and a biosynthetic gene (strB; 7,12, this paper), coding for an aminocyclitol amidinotransferase (ADT). The use of the two open reading frames, ORF1 and ORF5, additionally found on the determined sequence (Fig. 7), was not yet established by protein chemical analysis. The differences in the sequences re-

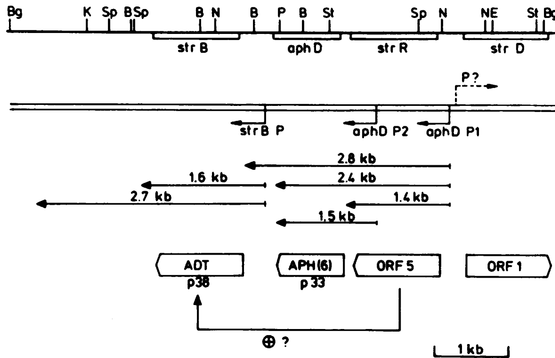


Fig.7. Summary and hypothetical expression scheme for four putative SM production genes. The molecular data were compiled from this work and previous publications (4,7,10,12). Double line = DNA; angled arrows = transcription start sites; thin-lined arrows = transcripts; boxed arrows = presumed proteins. The sizes of identified proteins (p) are given in kDa. A possible positive regulatory circuit is indicated.

ported by Tohyama et al. (12) and by us (10, this paper) could indicate a divergence in strains. However, since some of the differences could also be due to sequencing errors, we have re-sequenced all critical fragments. We believe that our sequence is correct, because the three differing peptides in the APH(6) primary structure are otherwise highly conserved between the far more distant APH(6) enzymes of S. griseus N2-3-11 (10) and S. glaucescens (11). Also, our sequence corresponds exactly to the independently determined partial protein sequence data, where available (10, 12).

A similar order of genetic elements, probably essential for the biosynthesis of hydroxy-SM and only distantly related on the sequence level, was detected in S. glaucescens (10,11; Vögtli and Hütter, personal communication; Mayer and Piepersberg, unpublished). The strikingly long intercistronic regions, however, seem to consist of mostly differing sequences in both organisms. This gave additional evidence that all the genetic elements analysed are in fact engaged in SM biosynthesis in the two organisms. The presently available molecular and in vivo data on the cluster of SM production genes of S. griseus are interpreted in Fig. 7 and revealed a complicated pattern of transcription and its regulation: (i) Divergently oriented transcription, (ii) superimposed transcription units for two genes (strR, aphD),

```
                         -35                            -10
E.COLI     A T           T T G A C A   T                G G T A T A A T G
STREPT.                  T T G A C C                      T A G G A T

aphD P1    T T A C T T G G C C G T T G C C C G G A T G T C C G G G T G C T A C T A T T C G C G A A
aphD P2    G T G C G G C T C A C C G A G G A C G G T C G G G C G G T G C T A C G T T G G C T G A A
strB P     G T G G A C A T A T G C C C G A G C G A A G C G G C G C T G C T A G C C T G C G A T G A
aph  P1    G A A A G G C G C G G A A C G G C G T C T C C G C C T C T G C C A T G A T.G C C G C C C A
afsB P     A G C G A G C G G G G G G C G G C G C T G T C G C G G C T G C T G G A C T T C T A C C T G

CONSENSUS  G T       G C G/T C   G     C G G C G/C G   T   C G C/G C/G G C/G T G C T A     T G/T C       A
```

Fig.8. Comparison of the most highly conserved regions of E. coli (23) and E. coli-like Streptomyces promoters (25) with five promoters of secondary metabolic genes. The three S. griseus promoters identified and the aph P1 and afsB promoters of S. fradiae (26) and S. coelicolor (24) respectively, were aligned for maximal homology and end with the first transcribed nucleotide (underlined). Nucleotides occurring three or more times are marked by an asterisk. A consensus sequence is given below for the five Streptomyces promoters with bigger letters for nucleotides occurring at least four times.

(iii) growth phase dependent transcription from successively used promoters aphD P1 and P2, (iv) absolute dependence of at least one SM biosynthetic gene, strB, from the presence of a functional activator gene (strR), and (v) absence of transcription in an A-factor negative mutant, to mention only some of the puzzling phenomena. Regulation of expression of the SM production genes obviously occurs on various levels and by the mean of several elements. A comparison of the promoters identified with E. coli-like promoters and others found in Streptomycetes (Fig. 8) showed only significant similarity with those of the neomycin-3'phosphotransferase (aphA) gene of S. fradiae (26) and of the afsB gene of S. coelicolor (24), and on the other hand deviate considerably from the E. coli consensus promoter. Since both aphA and afsB also seem to be genes of secondary metabolism, this could mean that their promoters are recognized by a special type of sigma factor of the transcribing RNA polymerase. Surprisingly both aphD P1 and aphD P2 are structurally similar though seemingly being regulated differently. Therefore, their control occur via additional factors recognizing other sites on the DNA.

Another level of regulation of gene expression seems to subdivide the SM production genes in early and late expressed ones. The early and optionally independent transcription from a second promoter of the aphD gene seems reasonable, because it makes the organisms resistant before producing a self-toxic substance. Another product of an early gene, strR, seems to be the activator of later expressed functions, such as the amidinotransferase. It will be interesting to see whether this positive control is also exerted on the upstream genes, i.e. strD. The postulated binding of the putative positive regulator to a DNA fragment containing the strB promoter (7; this work) could be in agreement with the finding that the strB gene product (ADT) is only expressed when a functional strR gene is present. But the presence of a rather strong promoter activity on the same fragment and in the same orientation as strB P (cf. Figs. 2, 3) in absence of the strR gene seems to contradict the above interpretation. Also no similarity to the typical helix-turn-helix domains found in many DNA binding proteins (28) could be detected in the putative StrR (ORF5) protein. However, if the promoter in plasmid pJDM48 is not an artifical one created at the cloning site (which seems unlikely from the sequence), one hypothesis could bring these conflicting observations together: The StrR protein could be an antiterminator of transcription similar to the N and Q proteins of bacteriophage lambda (29). Then, both the existence of a binding site for the antiterminator and of a transcription termination site upstream and downstream, respectively, of the BamHI site in between the aphD und strB genes (position 4017, cf. Fig.2) would have to be postulated. Computer assisted alignment of the ORF5 protein sequence with those of several proteins involved in positive control of transcription in fact revealed some distant similarities to the products of the lamda Q and afsB genes (not shown), but not to others. Interestingly, the afsB gene also seems to be a positive regulator engaged in the production of the A-factor (24).

By searching for homology in a data bank the protein expressible from the second gene with unknown function, strD (ORF1) showed significant similarity to the sequence of the ADP-glucose pyrophosphorylase of E. coli (Fig.9). This finding allows to

```
Eco GLGC  MVSLEKNDHLMLARQLPLKSVALILAGGRGTRLKTLTNKRAKPAVHF
SGR ORF1                   MK--ALVLAGGTGTRLRPITHTSAKQLVPV

GGKFRIIDFALSNCINSGIRRMGVITQYQSHTLVQHIQRGWSFF
ANK-PVLFYGLEAIRAAGIIDVGIVVGDTADEIVAAVGDG-SRF

NEEMNEFVDLLPAQQRMKGENWYRGTADAVTQNLDIIRRYKAEY
GLKVS-YIPQSKPLGLAHCVLISR---DFL-GEDDFIMYLGDNF

VVILAGDHIYKQDYSRMLID-HVEKGVRCTVVCMPVPIEEASAF
VVGVVEDSV-R-EFRAARPDAHL------MLTRVPEPR---S-F

GVMAVDENDKTIEFVEKPANPPSMPNDPSKS-LASMGIYVFDAD
GVAELSDSGQVLGLEEKPAHP---------KSDLALVGVYLF-SP

YLYELLEEDDRDENSSHDFGKDLIPKITEAGLAYAHPFPLSCVQ
AIHEAV-------AAITPSWRGELEITDAVQWLIDAGRDVRSTVI

SDPDAEPYWRDVGTLEAYWKAN-LDLASVVPKLDMYDRNNPIRT
S-----GYWKDTGNVTDMLEVNRLVLETTEPRCD-----------

YNESLPPAKFVQDRSGSHGMTLNSLVSGGCVISGSVVVQSVLFS
---------GLVDERSDLIGRVLVEEGAEVRNSRVMGPTVIGAGT

RVRVNSFCNIDSAVLLPEVWVGRSCRLRRCVIDRACVIPEGMVL
RVT-NSY--------VGPFTSLAEDCVVEDSEVEFSIVLRGASIS

GENAEEDARRFYRSEEGIVLVTREMLRK-LG-H-K-QER     429
GVRRIEASLIGRHVQVTSAPEVPHAHRLVLGDHSRAQISS    353
```

Fig.9. Comparison of the ORF1 and *E. coli* ADP-glucose pyrophosphorylase (27) polypeptide sequences. A similarity value of 37 % was obtained, when the number of boxed amino acid residues multiplied by 100 was divided by a figure composed of the total number of residues in ORF 1 and half of the gaps introduced.

postulate a function of the strD gene product in the activation by nucleotidylation of one of the glucose-6-phosphate derived sugar moieties of SM, streptose and N-methyl-L-glucoseamine (1,2). The precursors of both these modified hexoses are activated in the dTDP- and possibly the UDP-forms, respectively (30,8). This would also be in agreement with the result that mutants, i.e. strain M67 (smi-67), blocked in the late SM pathway (15), can be cured from their deficiency in SM biosynthesis by cloned DNA fragments overlapping with strD (4, this paper). Also, the molecular weight of the ORF1 protein calculated from the DNA sequence (38 kDa) corresponds rather well with that of a 41 kDa polypeptide translated from the respective DNA segment in vitro (10). However, further analysis has to clarify the physical and functional nature of the translational products expressed from the strR and strD genes.

ACKNOWLEDGEMENTS

We are grateful to G. Zeising for typing the manuscript. This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

*To whom correspondence should be addressed

REFERENCES

1. Walker,J.B. (1975) Methods Enzymol. 43, 429-470.
2. Grisebach,H (1978) Adv. Carbohydr. Chem. Biochem. 35, 81-126.
3. Distler,J. and Piepersberg,W. (1985) FEMS Microbiol. Lett. 28, 113-117.
4. Distler,J., Mansouri,K. and Piepersberg,W. (1975) FEMS Microbiol. Lett. 30, 151-154.
5. Hintermann,G., Crameri,R., Vögtli,M. and Hütter,R. (1984) Mol. Gen. Genet. 196, 513-520.
6. Tohyama,H., Shigyo,T. and Okami,Y (1984) J. Antibiot. 37, 1736-1737.
7. Ohnuki,T., Imanaka,T. and Aiba,S. (1985) J. Bacteriol. 164, 85-94.
8. Kumada,Y., Horinuchi,S., Uozumi,T. and Beppu,T. (1986) Gene 42, 221-224.
9. Shinkawa,H., Sugiyama,M., Nimi,O. and Nomi,R. (1985) FEBS Lett. 181, 385-389.
10. Distler,J., Braun,C., Ebert,A. and Piepersberg,W. (1987) Mol. Gen. Genet. (in press).
11. Vögtli,M. and Hütter,R. (1987) Mol. Gen. Genet (in press).
12. Thoyama,H., Okami,Y. and Umezawa,H. (1987) Nucl. Acids Res. 15, 1819-1834.
13. 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. The John Innes Foundation, Norwich.
14. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, New York.
15. Distler,J., Klier,K., Piendl,W., Werbitzky,O., Böck,A., Kresze,G. and Piepersberg,W. (1985) FEMS Microbiol. Lett. 30, 145-150.
16. Norrander,J., Kempe,T. and Messing,J. (1983) Gene 26, 101-106.
17. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA. 74. 5463-5467.
18. Maxam.A.M. and Gilbert,W. (1977) Proc. Natl. Acad. Sci. USA. 74, 560-564.
19. Chirgwin,J.M., Przybyla,A.E., McDonald,R.J. and Rutter,W.J. (1979) Biochemistry 18, 5294-5299.
20. Favaloro,J., Treisman,R. and Kamen,R. (1980) Methods Enzymol. 65, 718-749.
21. Haas,M.J. and Dowding,J.E. (1975) Methods Enzymol. 43, 611-633.
22. Bibb,M.J., Findlay,P.R. and Johnson,M.W. (1984) Gene 30, 157-166.
23. Harley,C.B. and Reynolds,R.P. (1987) Nucl. Acids Res. 15, 2343-2361.
24. Horinouchi,S., Suzuki,H. and Beppu,T. (1986) J. Bacteriol. 168, 257-269.
25. Hopwood,D.A., Bibb,M.J., Chater,K.F., Janssen,G.R., Malpartida,F. and Smith,C.P. (1986)in Regulation of Gene Expression (Booth,I. and Higgins,C., eds.) p. 251-276. SGM, Cambridge.
26. Bibb, M.J. and Janssen,G.R. (1986) in Proceedings of the Fifth International Congress on the Genetics of Industrial

Microorganisms (Alacevic,M., ed.). Split, Yugoslavia (in press).
27. Baecker,P.A., Furlong,C.E. and Preiss,J. (1983) J. Biol. Chem. 258, 5084-5088.
28. Pabo,C. and Sauer,R. (1984) Ann. Rev. Biochem. 58, 293-321.
29. Sanger,F., Coulson,A.R. Hong,G.F., Hill,D.F. and Petersen, G.B. (1982) J. Mol. Biol. 162, 729-773.
30. Kniep,B. and Grisebach,H. (1976) FEBS Lett. 65, 44-46.