Mol Gen Genet (1995) 248:610–620

ORIGINAL PAPER

Andreas Bechthold · Jae Kyung Sohng
Todd M. Smith · Xin Chu · Heinz G. Floss

Identification of *Streptomyces violaceoruber* Tü22 genes involved in the biosynthesis of granaticin

Received: 3 November 1994 / Accepted: 7 February 1995

Abstract A 50 kb region of DNA from *Streptomyces violaceoruber* Tü22, containing genes encoding proteins involved in the biosynthesis of granaticin, was isolated. The DNA sequence of a 7.3 kb fragment from this region, located approximately 10 kb from the genes that encode the polyketide synthetase responsible for formation of the benzoisochromane quinone skeleton, revealed five open reading frames (ORF1–ORF5). The deduced amino acid sequence of GraE, encoded by ORF2, shows 60.8% identity (75.2% similarity) to a dTDP-glucose dehydratase (StrE) from *Streptomyces griseus*. Cultures of *Escherichia coli* containing plasmids with ORF2, on a 2.1 kb *Bam*HI fragment, were able to catalyze the formation of dTDP-4-keto-6-deoxy-D-glucose from dTDP-glucose at 5 times the rate of control cultures, confirming that ORF2 encodes a dTDP-glucose dehydratase. The amino acid sequence encoded by ORF3 (GraD) is 51.4% identical (69.9% similar) to that of StrD, a dTDP-glucose synthase from *Streptomyces griseus*. The amino acid sequence encoded by ORF4 shares similarities with proteins that confer resistance to tetracycline and methylenomycin, and is suggested to be involved in transporting granaticin out of the cells by an active efflux mechanism.

Communicated by W. Goebel

A. Bechthold[1] · J. K. Sohng[2] · X. Chu[3] · H. G. Floss (✉)
Department of Chemistry, University of Washington, Seattle, WA 98195 USA

T. M. Smith[4] · H. G. Floss
Department of Medicinal Chemistry, University of Washington, Seattle, WA 98195 USA

Present addresses:
[1] Institut für Pharmazeutische Biologie, Eberhard-Karls-Universität Tübingen, 72076 Tübingen, Germany
[2] Department of Chemistry, Sunmoon University, Chunahm, Sambyongdong 381-7, Chungnam 330-150, Korea
[3] Department of Microbiology, 241 E. University Village North Dakota State University Fargo, ND 58102 USA
[4] Department of Molecular Biotechnology FJ-20, University of Washnigton, Seattle, WA 98195 USA

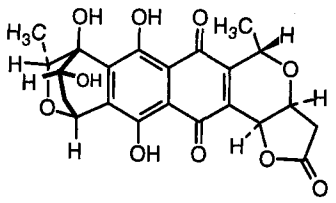
Dedicated to Professor Satoshi Ōmura, a pioneer in the field of antibiotics, on the occasion of his 60th birthday

Introduction

Streptomyces violaceoruber Tü22 produces granaticin, a benzoisochromane quinone antibiotic which shows antibacterial activity against gram-positive bacteria and weak antitumor activity (cf. Snipes et al. 1979). Granaticin differs from other benzoisochromane quinones in that it contains a 2,6-dideoxyhexose moiety uniquely annealed to the aromatic ring system through two carbon-carbon linkages at C-1 and C-4 (Fig. 1). Feeding experiments have established that granaticin is synthesized from eight molecules of acetate and one molecule of glucose (Snipes et al. 1979; He et al. 1986). Genes encoding the polyketide synthetase (PKS) involved in the formation of the benzoisochromane quinone moiety of granaticin have been cloned, sequenced (Sherman et al. 1989) and expressed in a heterologous *Streptomyces* host (Sherman et al. 1992; Khosla et al. 1993). However, there is currently little information about the genes encoding formation of the carbohydrate moiety of this antibiotic.

Feeding experiments had also shown that the biosynthesis of the granaticin sugar moiety from glucose involves intramolecular tansfer of 4-H to C-6, where it displaces the 6-OH group with inversion of configuration (Snipes et al. 1979); this mechanistic feature is characteristic of the dTDP-glucose 4,6-dehydratase reaction (Snipes et al. 1977). It was thus likely that a dTDP-glucose 4,6-dehydratase is one of the first enzymes involved in the biosynthesis of the 2,6-dideoxyhexose portion of granaticin. In other organisms that produce antibiotics containing

Fig. 1 Structure of granaticin



6-deoxy-hexose moieties, the genes that encode 4,6-dehydratases are found within gene clusters that contain the biosynthetic genes (Pissowotzki et al. 1991; Thorson et al. 1993a, b). These genes are highly conserved in actinomycetes; DNA probes derived from the genes encoding the streptomycin biosynthetic enzymes, StrD, E, L, M in *S. griseus*, have been used to detect DNA fragments that presumably contain genes encoding enzymes responsible for formation of 6-deoxyhexoses in other actinomycetes (Stockmann and Piepersberg 1992). In this paper we report the cloning and partial analysis of a DNA fragment from *S. violaceoruber* Tü 22 encoding, a dTDP-glucose 4,6-dehydratase, using DNA probes derived from the N-terminal amino acid sequence of the purified *S. violaceoruber* enzyme and from amino acid consensus sequences found within other dTDP-glucose dehydratases.

Materials and methods

Bacterial strains, bacteriophages, plasmids and growth conditions

Streptomyces violaceoruber Tü22 was obtained from Prof. H. Zähner, Tübingen, and grown as described (Snipes et al. 1979). *Escherichia coli* strain XL-1 Blue MRF′ (Stratagene, La Jolla, Calif.) was grown as described (Sambrook et al. 1989). Replicative form M13mp18 and M13mp19 phage DNA (Yanisch-Perron et al. 1985) was obtained from United States Biochemicals (Cleveland, Ohio). Plasmid pBluescript SK^- (pSK^-) was obtained from Stratagene, and cosmid pOJ446 (Bierman et al. 1992) was a gift from Eli Lilly Co. (Indianapolis, Ind.). Other cosmids and plasmids constructed in this work are listed in Table 1.

Buffers and reagents

All buffers were prepared as described (Sambrook et al. 1989). Apralan® (apramycin sulfate) was obtained from SmithKline Beecham (Animal Health Division), restriction enzymes from New England Biolabs (Beverly, Mass) or from Stratagene and T_4 DNA Ligase from Boehringer Mannheim (Indianapolis, Ind.). Gigapack XL (Stratagene) was used to package DNA for cosmid libraries. Chemicals were obtained from Sigma (St. Louis, Mo.), and radiolabeled nucleotides from New England Nuclear (Boston, Mass.).

DNA isolations

Genomic DNA was isolated from *S. violaceoruber* Tü22 as described (Hopwood et al. 1985). Plasmids were propagated in *E. coli* XL-1 Blue MRF′ and purified according to standard procedures (Sambrook et al. 1989).

Table 1 Plasmids and cosmids prepared in this study

Cosmid/plasmid	Relevant properties
pOJ446-22-24	pOJ446 with 40 kb insert (DNA from *S. violaceoruber* Tü22)
pB24a, pB24b	pSK^- with 2.0 kb *Bam*HI insert; GraE (both orientations)
pB32	pSK^- with 5.3 kb *Bam*HI insert; GraR
pB13	pSK^- with 4.6 kb *Bam*HI insert; granaticin PKS
pB8	pSK^- with 2.4 kb *Bam*HI insert; granaticin PKS
pSV501	pSK^- with 4.7 kb *Bgl*II insert; GraE fragment and GraR
pSTü22-PCR	pSK^- with 0.5 kb insert (insert obtained by PCR)

DNA amplification by polymerase chain reaction (PCR)

Oligodeoxynucleotide primers were synthesized on an Applied Biosystems DNA synthesizer (Department of Chemistry, University of Washington). PCR amplification was performed using a Genetic Thermal Cycler system GTC-1 (Precision Scientific). Each reaction mixture contained 0.17 µg genomic DNA, 100 pmol of each primer, 20 pmol of each dNTP, 1 µmol TRIS-HCl pH 8.3, 5 µmol KCl, 2 µmol $MgCl_2$ and 0.1% (w/v) gelatin in a final volume of 100 µl. After addition of 2.5 U AmpliTaq DNA polymerase (Perkin Elmer, Norwalk, Conn.) the DNA template was denatured at 95° C for 3 min. Amplification was carried out by 35 cycles of annealing, extension (2.5 min. at 72° C) and denaturation (1.5 min. at 95° C). The resulting PCR product was cloned into pSK^- to create pSTü22-PCR.

Southern hybridizations

Southern blots of DNA from *S. violaceoruber* Tü22 were probed with pSTü22-PCR labeled with ^{32}P using the Random Primed DNA labeling kit (United States Biochemicals), and [α^{-32}P] dCTP (New England Nuclear). Hybridizations were carried out for 16–24 h at 50° C in a solution containing 50% (v/v) formamide (Sambrook et al. 1989). Following hybridization, the filters were washed twice in 6 × SSC (90 mM NaCl, 90 mM sodium citrate, pH 7.0) containing 0.1% SDS at 20° C for 20 min each, and then once in 2 × SSC (30 mM NaCl, 30 mM sodium citrate pH 7.0) containing 0.1% SDS at 60° C for 60 min.

DNA sequencing

DNA sequencing was performed on single-stranded or double-stranded templates (subclones in either M13mp18 or in pSK^-) by the dideoxynucleotide chain-termination method using the Sequenase version 2.0 DNA sequencing kit (United States Biochemicals) and [α^{-35}S]dCTP. Both strands were sequenced with standard primers (T_7, T_3, KS, SK, supplied in the kit) and custom 18 nucleotide primers (DNA International, Lake Oswego, Ore.).

Computer-assisted sequence analysis

Computer-aided sequence analysis was carried out using the GCG software package (Devereux et al. 1984) (University of Wisconsin Genetics Computer Group, Madison, Wis.), on a VAX mainframe computer (University of Washington, Seattle).

TDP-glucose 4,6-dehydratase expression

TDP-glucose 4,6-dehydratase activity was determined as described (Vara and Hutchinson 1988), with slight modifications. Plasmids B24a, B24b and pSK$^-$ were transformed into *E. coli* XL-1 Blue MRF′ (Sambrook et al. 1989), and single colonies were used to inoculate 2.0 ml LB medium containing carbenicillin (50 μg/ml). After 16 h of growth at 37° C, 100 μl of each culture was used to inoculate 50 ml LB medium containing 0.2% glucose, and growth was continued at 37° C. At an OD_{600} of 0.3 isopropylthiol-β-galactoside (IPTG) was added to give a final concentration of 1.0 mM, and growth was continued for an additional 5 h. The cells were harvested by centrifugation and resuspended in 0.1 M TRIS-HCl buffer pH 8.0 containing 1.0 mM dithiothreitol, (DTT) and disrupted by sonication. Cell debris was removed by centrifugation to give a cell-free extract. Enzyme activity was measured in reactions containing NAD (1.5 mM), TRIS-HCl pH 7.5 (0.15 M), 15 μg protein, and the substrate dTDP-glucose (1.8 mM), in a final volume of 100 μl. After a 30 min incubation at 37° C, 0.9 ml NaOH (0.1 M) was added and the mixtures were incubated at 37° C for an additional 30 min. The amount of product formed was determined by measuring the absorbance at 320 nm and comparing reactions containing dTDP-glucose to ones lacking substrate.

Results

Cloning of the dTDP-glucose dehydratase gene from *S. violaceoruber* Tü22

dTDP-glucose 4,6-dehydratase from *S. violaceoruber* Tü22 was purified to homogeneity using the procedure of Vara and Hutchinson (1988) through the Mono Q column step, followed by preparative SDS gel electrophoresis (data not shown). Gas phase microsequencing (Matsudaira 1993) of the 34 kDa protein gave an N-terminal amino acid sequence VTGAAGFIIS, which showed strong homology to the N-terminus of the *Saccharopolyspora* enzyme. Based on this sequence and on an internal consensus sequence of known dTDP-glucose 4,6-dehydratases, amino acids 207–215 of StrE and 206–214 of the *Sac. erythrea* enzyme (Pissowotzki et al. 1991), two oligodeoxynucleotide primers (JKS-II-36-4, 5′-GGSGCSGCSTTCATSGGSTCSCAC-3′ and JKS-II-42-3, 5′GGGAACTCGTASGGSCCGTAGTT-GTT-3′, S = G or C) were synthesized and used to amplify a 520 bp fragment from *S. violaceoruber* Tü22 genomic DNA by PCR. The resulting fragment was used as a probe in Southern blots of *S. violaceoruber* Tü22 DNA digested with a variety of restriction enzymes. In each digest, only one fragment was observed to hybridize with the probe (Fig. 2). Based on these data, *S. violaceoruber* Tü22 chromosomal DNA was digested with *Bgl*II; fragments of 5-6 kb in length were isolated and ligated into pSK$^-$ previously digested with *Bam*HI. The resulting ligation products were used to transform *E. coli* and colonies containing DNA that hybridized to the PCR derived probe were identified by colony hybridization. Plasmid pSV501 was isolated and found to contain, by DNA sequence analysis, a portion of the dTDP-glucose dehydratase gene.

In order to clone the entire dTDP-glucose dehydratase gene and to expand the region of cloned DNA surrounding the gene, a cosmid library of *S. violaceoruber* Tü22 DNA was prepared. *S. violaceoruber* Tü22 DNA was partially digested with *Sau3*A, dephosphorylated, and ligated into pOJ446, previously digested with *Hpa*I and *Bam*HI. The ligation products were packaged and transduced into *E. coli* XL-1 Blue MRF′. Approximately 5000 apramycin-resistant colonies were probed by colony hybridization with a DNA fragment subcloned from pSV501. Several colonies were identified

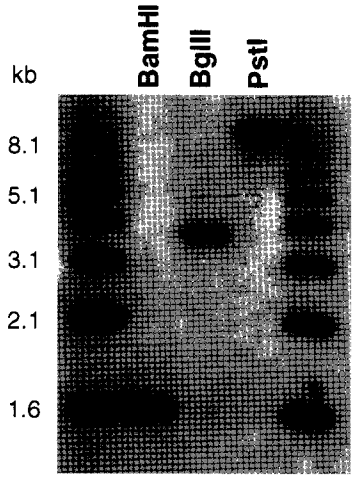
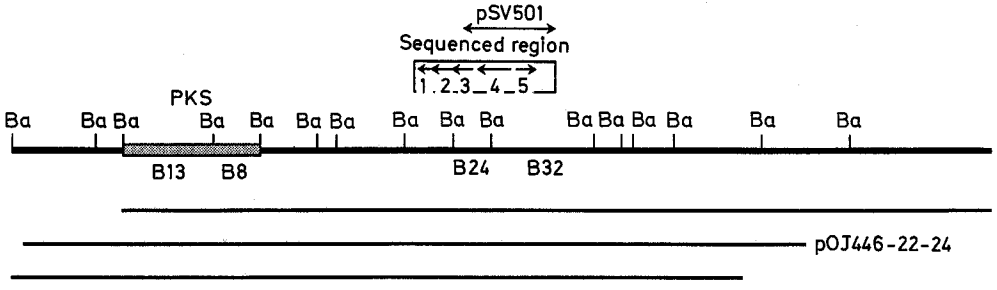


Fig. 2 Southern blot of *Streptomyces violaceoruber* Tü22 genomic DNA digests probed with the dTDP-glucose dehydratase gene fragment from pSU501

Fig. 3 *Bam*HI restriction map of the cloned region of *S. violaceoruber* Tü22 DNA containing the dTDP-glucose dehydratase gene. The cosmid clones isolated are shown below the map. The region containing the polyketide synthetase genes is represented as a *shaded box* (PKS), and the region sequenced is shown as an *open box*. The five open reading frames are indicated



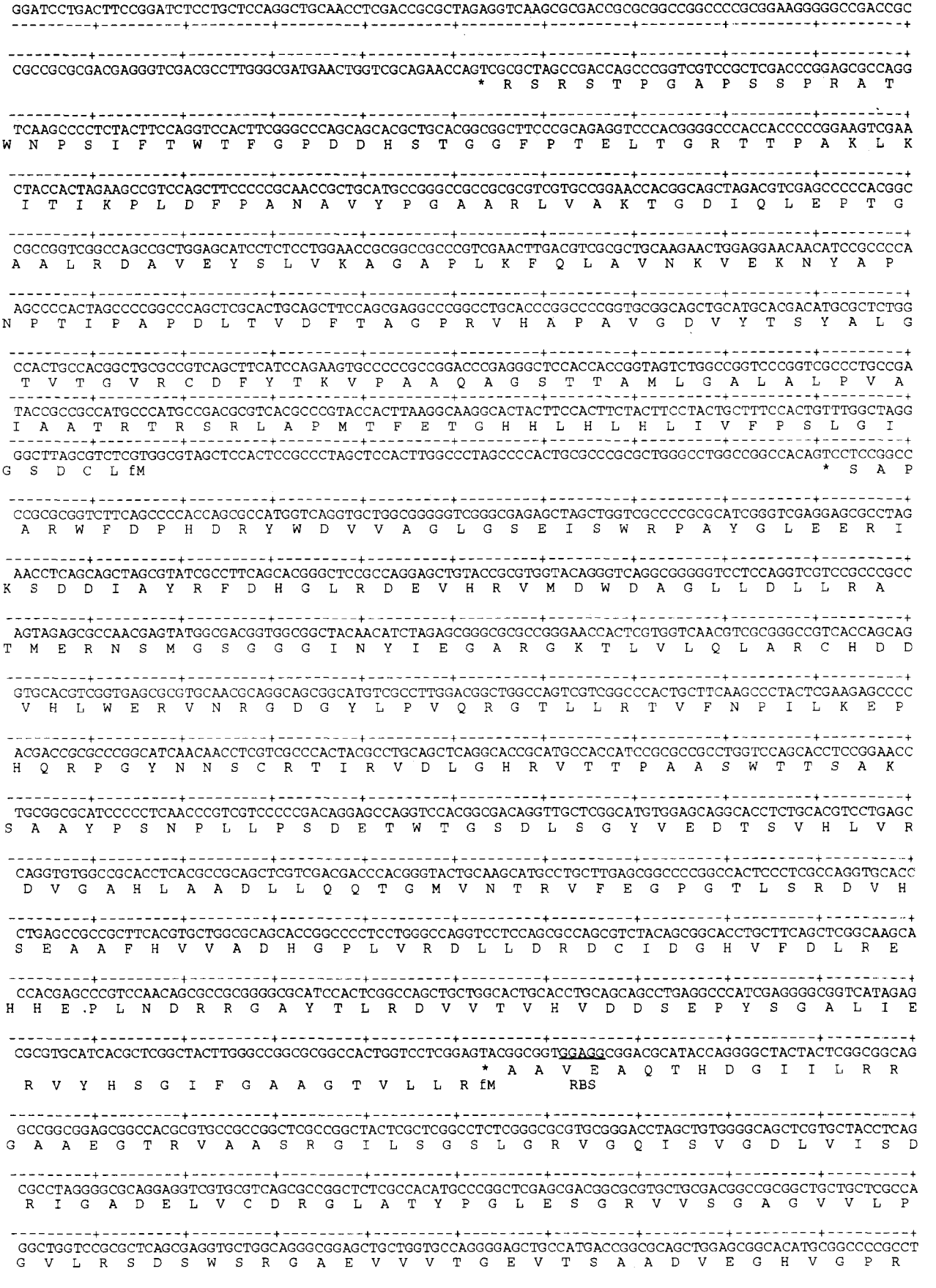


Fig. 4 Continued

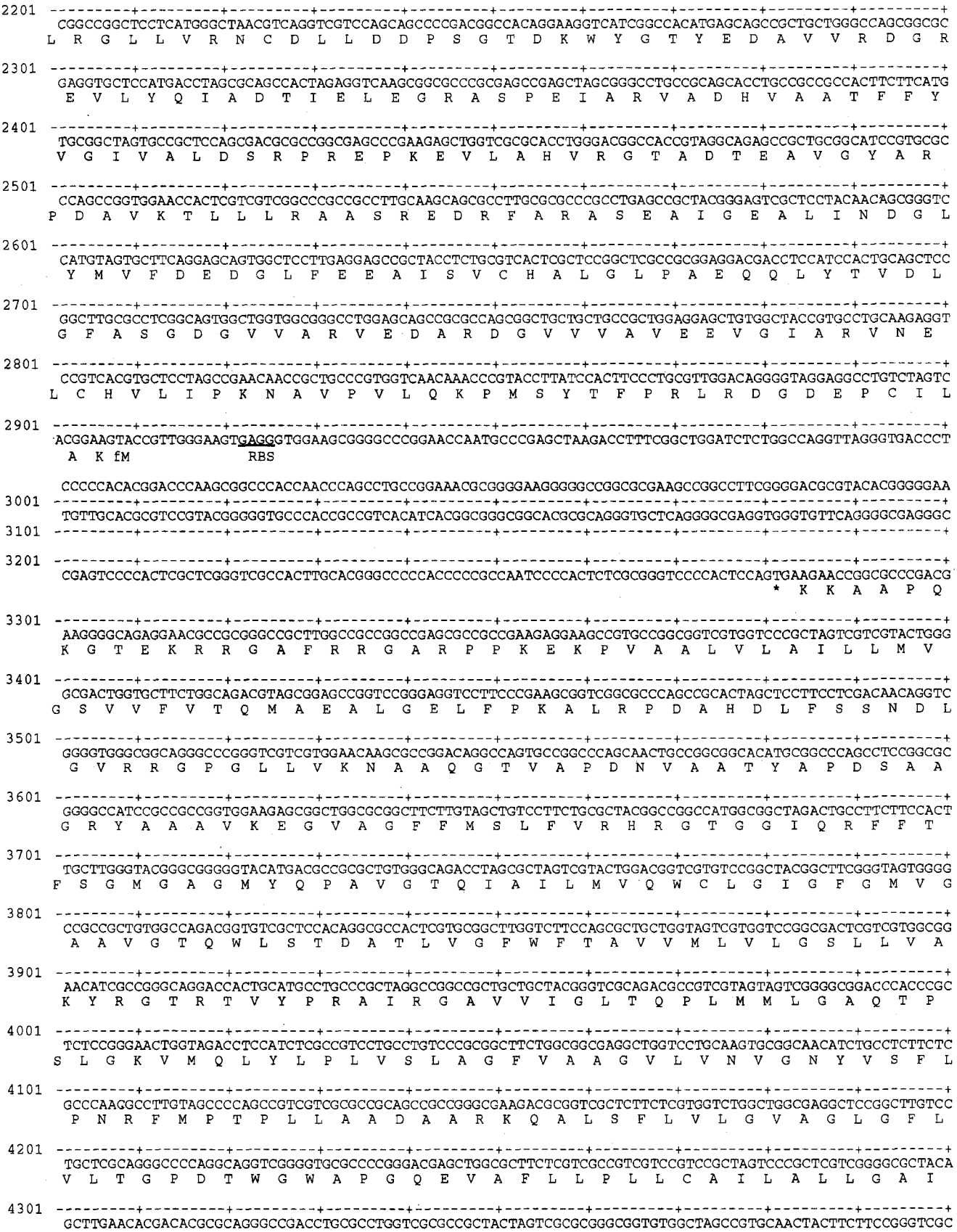


Fig. 4 Continued

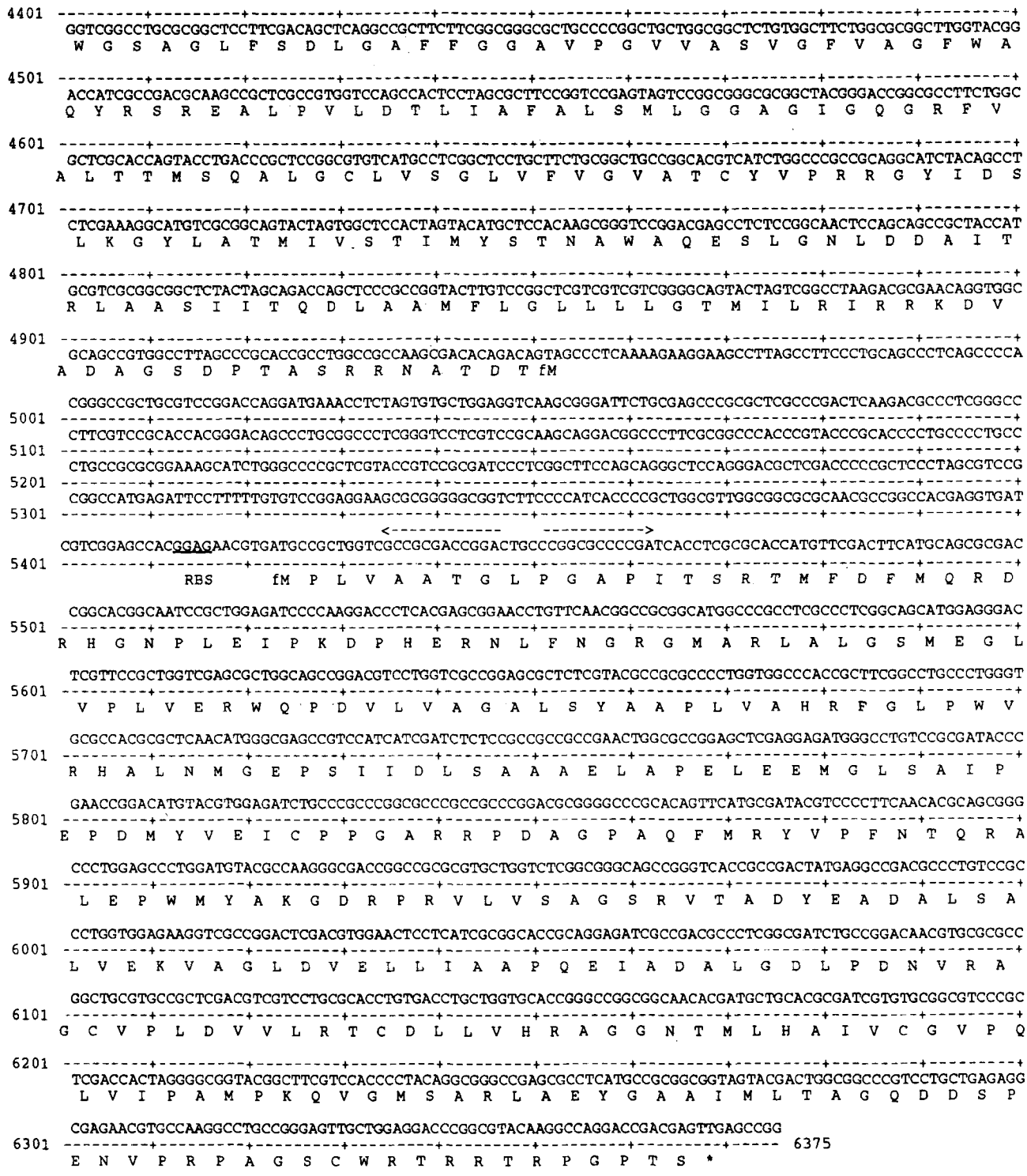


Fig. 4 Nucleotide sequence for the analyzed region (nucleotides 1–6375) of the 7365 nt fragment (GSDB/GenBank Accession L37334). ORF1, positions 736 (ATG) to 151 (TGA) (612 nt or 203 amino acids) or positions 814 (GTG) to 151 (TGA) (663 nt or 220 amino acids); ORF2, 1854 (ATG) to 891 (TGA) (963 nt or 320 amino acids); ORF3, 2909 (ATG) to 1850 (TGA) (1059 nt or 352 amino acids); ORF4, 4952 (ATG) to 3278 (TGA) (1674 nt or 557 amino acids); ORF5, 5423 (ATG) to 6370 (TGA) (948 nt or 314 amino acids). Possible ribosome binding sites are *underlined*. A possible inverted repeat between ORFs 4 and 5 is denoted by the *arrows*

which contained DNA that hybridized to the probe. The cosmid DNA isolated from these colonies was analyzed by Southern blot and restriction mapping. Three overlapping clones were identified that contained the hybridizing fragment (Fig. 3). These clones encompassed approximately 50 kb of the *S. violaceoruber* Tü22 genome. One cosmid, pOJ446-22-24, was used for further restriction mapping, hybridization, and sequencing experiments.

Since the dTDP-glucose dehydratase gene is thought to be involved in the early steps of formation of the carbohydrate moiety that is attached to granaticin, it is likely that the genes encoding these biosynthetic enzymes are closely linked to the granaticin PKS (Martin and Liras 1989). Previous studies have shown that the *actI* and *actIII* DNA fragments (Malpartida and Hopwood 1986) from *S. coelicolor* can hybridize to DNA containing the granaticin PKS from *S. violaceoruber* Tü22 (Sherman et al. 1989). When these fragments were used to probe Southern blots of pOJ446-22-24, strong hybridization was observed, suggesting that the putative dTDP-glucose dehydratase gene and granaticin PKS are closely linked. These data were confirmed by sequencing a small portion of DNA (ca. 300 nt) that hybridized with the *actI* and *actIII* DNA probes. The DNA sequence obtained from this region was identical to a section of the previously published sequence of the granaticin PKS (Sherman et al. 1989) (Fig. 3).

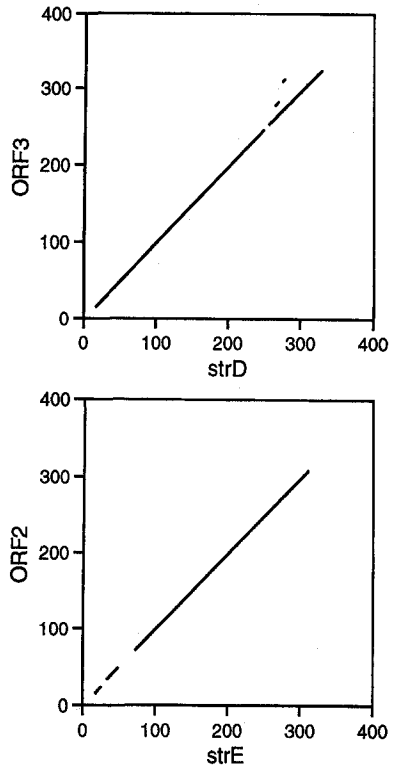


Fig. 5 DOTPLOT (Devereux et al. 1984) comparisons of ORFs 2 and 3 with the deduced StrE and StrD gene products from *S. griseus* (Pissowotzki et al. 1991). The data were compared using a window of 30 and a stringency of 20

Expression of GraE in *E. coli*

To confirm that the putative dTDP-glucose dehydratase (GraE) gene actually encoded a functional enzyme, it was necessary to express this gene in a heterologous system. This was accomplished by subcloning the 2.1 kb *Bam*HI fragment (B24, Fig. 3), containing the putative gene, into pSK$^-$. The fragment was subcloned in both orientations to give pB24a (transcription of ORF3 expected to be under control of the *lacZ* promoter) and pB24b (transcription expected not to be under control of the *lacZ* promoter), and tested for expression in *E. coli* XL-1 Blue MRF′. When assayed for dTDP-glucose 4,6-dehydratase activity, only one cell-free extract was able to catalyze the conversion of dTDP-glucose to dTDP-4,6-deoxyglucose at levels above the background (0.429 μmol/mg protein for *E. coli* strain XL-1 Blue MRF′ pB24a, *vs.* 0.083 μmol/mg protein for the host carrying pSK$^-$ or pB24b) confirming that the cloned fragment contained a functional dTDP-glucose dehydratase gene. Since increased enzyme activity could only be obtained from the DNA fragment cloned in one orientation, it is likely that transcription of the mRNA for this gene product is under the control of the *lacZ* promoter in pSK$^-$.

Sequence analysis of a 7.3 kb region containing GraE

By analogy to the streptomycin biosynthetic cluster (Stockmann and Piepersberg 1992), it is expected that the genes encoding the other biosynthetic enzymes for the carbohydrate moiety of granaticin are clustered around the dTDP-glucose dehydratase gene. Therefore, as a first step towards identifying these genes, a 7.3 kb region of DNA (B24 and B32, Fig. 3) containing the dTDP-glucose dehydratase gene was sequenced (Fig. 4). Five open reading frames (ORFs) were identified by their predicted codon usage (Bibb et al. 1984) (Fig. 4). ORF2 (GraE) encodes the dTDP-glucose dehydratase. The deduced amino acid sequence of GraE shows 75.2% similarity and 60.8% identity to the deduced amino acid sequence of StrE from *S. griseus* (Pissowotzki et al. 1991) (Fig. 5). A comparison between the amino acid sequences encoded by ORF3 and *strD* from *S. griseus* (Pissowotzki et al. 1991) (Fig. 5) reveals 69.9% similarity and 51.4% identity, suggesting that ORF3 codes for a dTDP-glucose synthetase (GraD). When the amino acid sequence encoded by ORF4 was used to search peptide databases (GenPept, release 77, 6/93; and SwissProt, release 25, 4/93), using FASTA (Pearson 1990), similarities to a family of proteins known to confer antibiotic reistance by an active efflux

mechanism were identified. These included the tetracycline resistance factor (Tet347) from *S. rimosus* (Reynes et al. 1988), 33.9% identity; the actinorhodin transporter from *S. coelicolor* A3(2) (Fernández-Moreno et al. 1991), 30.3% identity; a methylenomycin A resistance factor (Mmr) from *S. coelicolor* (Neal and Chater 1987), 29.3% identity; and a putative tetracenomycin transporter (Guilfoile and Hutchinson 1992), 25.0% identity; and others (data not shown) (Fig. 6). When the hydrophilicity plot of ORF4 is compared to those of other transporters (e.g., the putative tetracenomycin tranporter from *S. glaucensens*) a high degree of structural similarity is observed (Fig. 7), suggesting that ORF4 encodes a protein involved with transporting granaticin out of the cells. Database searches with the deduced amino acid sequences encoded by ORF1 and ORF5 failed to identify any significant similarities.

Discussion

In bacteria, 6-deoxyhexoses are essential and dTDP-glucose 4,6-dehydratases are involved in the biosynthesis

```
            1                                                    50
Sgltcm      MsTethdeps gvAhtPAsGl RgRPWp.tll AvAvGvmMvA LDsTIVaIAn
Scommr      .mTTvrtggA qtAevPAgGr RdvPsgVkit ALAtGFvMAt LDvTVVnVAg
ORF4        MtdTanrrsA tpdsgadavd krRiRlImtg lLllGLfMAA LDqTIIsaAl
Sritet      .......... .......... .......... .......... ..........
Consensus   M-TT-----A --A--PA-G- R-RP------ ALA-G--MAA LD-TIV--A-

            51                                                   100
Sgltcm      paIQqDLhas LaDvqWItnG YLLaLAVsLi tAGKLgDRFG hRqtFLvGVa
Scommr      aTIQEsLdtt LtqltWIvdG YvLTFAsLLm LAGgLanRiG aktVYLwGmG
ORF4        rTIaDDLng. LsEqaWants YMiTsvIMta LyGKLsDiYG rRpVYctaVG
Sritet      .......... .......... .......... .......... ..........
Consensus   -TIQ-DL--- L----WI--G Y-LT-A--L- LAGKL-DR-G -R-VYL-GVG

            101                                                  150
Sgltcm      gFavtSaAiG LsgSvaaiVV FRvlQGlfgA LMqPSALgLL rvtFPPgkl.
Scommr      VFfLaSlACa LAptaeTLIa aRlvQGaGAA LFMPSsLsLL vfsFPekrqR
ORF4        VFvLgSvlCG LAqSmtTLaV FR.gQGiGAg gLMslAFaiL t.dLvPlaeR
Sritet      .......... .......... .......... .......... ..........
Consensus   VF-L-S-ACG LA-S--TL-V FR--QG-GAA L-MPSAL-LL ---FPP---R

            151                                                  200
Sgltcm      nmaiGiWSgV VGaStaaGPI IGGL...... LVqhvGWEaV FFINVPVGlA
Scommr      tRmlGlWSAI VatSsglGPt VGGL...... MVsAfGWEsI FLlNlPIGai
ORF4        sRyqawFgAV fGvSavvGPV aGGFfaglds FlgAsGWrwa FFINVPIGvA
Sritet      .......... .......... .......... .......... ..........
Consensus   -R--G-WSAV VG-S---GP- -GGL------ -V-A-GWE-- FFINVPIG-A

            201                                                  250
Sgltcm      aLvag..LVi lTDaRAeRap ksFDVsGivl LsgAMFcLvW gLikaPAWGW
Scommr      GMamtyryIa aTEsRAtR.. ..LaVpGhLl wIvALaavsF ALieGPqlGW
ORF4        GaliiaaLVr vqpgRAqh.. .kFDIaGlLa LIacLLPLLF AveQGPAWGW
Sritet      .......... .......... ..MDVlGaaa LalfLvPLLi vaeQGrtWGW
Consensus   G------LV- -T--RA-R-- --FDV-G-L- LI-AL-PLLF AL-QGPAWGW

            251                                                  300
Sgltcm      gDlrTLgFLA aavlaFAgFt LREsRATE.p lMPLaMFRsv pLSagtvlmv
Scommr      tagpvLtaYA vavtaaALLa LREhRvTn.p vMPwqLFRgP gFtgaNlVgf
ORF4        tDPgTLvLFg LgavgLvLFs LaqkRAaDaa lLPtpMFRnP lFSvYNgVNv
Sritet      gsPaaLaLFA LgaagLAvFi pvElRrgDea iLPLgLFRrg sialCsaVNf
Consensus   -DP-TL-LFA L----LALF- LRE-RATD-- --PL--FR-P -FS--N-VN-

            301                                                  350
Sgltcm      LmaFsfiGgL FfvtFYLQnV hGMsPvesGv hLLPLTgmmI VgapVSGiVI
Scommr      LfnFAlFGst FmLgLYFQha rGatPfQAGL eLLPMTiffp VAniVyarIs
ORF4        LVGaAvFGaL svLPLYLQMV kGLsPTQAGL mMLPqTLGiV VAgrIarpyV
Sritet      tIGvgiFGtv ttLPLFLQMV qGrtPTQAGL vviPFmLGtI asqmVSGklI
Consensus   L-GFA-FG-L F-LPLYLQMV -G--PTQAGL -LLP-TLG-I VA--VSG--I

            351                                                  400
Sgltcm      sRFGp.ggpL VVGMLLtAAs LWgMsTLeAD sgMgItslwf vlLGlGlapV
Scommr      ARFsn.gtLL tafLLLagAA sLsMvTiTAs TPyWVvavAV GVaniGaGiI
ORF4        tRtGRYkavL lsGLvLMvvA tFwfgvLTAD TsLWqtgvAa GVMGfGIGLc
Sritet      AssGRFkkLa lVGLgsMAgA LLaMaTtgAt TPMWgivliV lwLGvGIGLs
Consensus   ARFGR---LL -VGLLLMAAA LL-M-TLTAD TPMW----AV GVLG-GIGL-
```

Fig. 6 Continued

```
            401                                                    450
Sgltcm      mvgtTdvivs nAPaElaGVA GGlqqsAmQV GGSlGtAVLg vlM....aSR
Scommr      spgMTaAlvd aAgPEnanVA GsvLnanRQI GslvGIAaMg vvL....hS.
ORF4        wqvMliAiqt gvaPqymGag mGsFtffRQI GGtgrhrVFl smFFgavgek
Sritet      qtviTspmqn sAPksqlGVA nGasacAgQI GGStGIAVLf svMFavalgR
Consensus   ---MT-A--- -APPE--GVA GG----ARQI GGS-GIAVL- --MF----SR

            451                                                    500
Sgltcm      VgDvFpdkwa EanlpRvgpr eAaaiEdAae V..... ..Gavppagt
Scommr      ..... ..... ..... ..... .....
ORF4        VAaaYrgaas DPaYtaavnD PAVTGqaANk VLL.....Gp GrrVgLDnsS
Sritet      lADlLht... .PrYeRlltD PAITGDpANh rFLdmaesGq GaGInLDdtS
Consensus   VAD------- -P-Y-R---D PA-TG--AN- V-L-----G- G-G--LD--S

            501                                                    550
Sgltcm      LpgrhagtLs evVhsSFisG MglaFtvAGa VALvAaaVal FtrkAePDER
Scommr      ..... ...TsdWdHG aaIsF.lAvG lAyLlgglsA WrLiArPErR
ORF4        FLdhaDpRLa kPflegLAea MqtvF.vvsG VmLLiaLVLA ....AvPkEk
Sritet      LLngiDaRLm qPVTdSFAHG fhImF.lpGG VvLLAgFVMt WfLrelqEEt
Consensus   LL---D-RL- -PVT-SFAHG M-I-F--AGG VALLA--V-A W-L-A-PEER

            551                  575
Sgltcm      APEEfpVpAs tAGrg..... .....
Scommr      s....aVtAa t..... .....
ORF4        pPragRrf.a grrketGkqP Aakk*
Sritet      APEEeRpaes gAGaknGplP Asda.
Consensus   APEE-RV-A- -AG---G--P A----
```

of these hexoses (Griesebach 1978). The granaticin producer *S. violaceoruber* Tü22 appears to carry a single copy of the dTDP-glucose 4,6-dehydratase gene, and it is linked to the granaticin polyketide synthase gene. A similar situation is observed in the streptomycin producers *S. griseus* and *S. glaucescens*. These organisms each contain a single copy of the dTDP-glucose 4,6-dehydratase gene located within a gene cluster that encodes enzymes involved in secondary metabolism (Mansouri and Piepersberg 1991). Neither organism expresses the dehydratase during vegetative growth; therefore it has been proposed that 6-deoxyhexose biosynthesis is not required for growth in these organisms (Stockmann and Piepersberg 1992).

The sequence analysis of a 7.3 kb region of DNA containing the dTDP-glucose 4,6-dehydratase gene (GraE) revealed five potential ORFs. Two of these (GraD and GraE) are highly similar to genes involved in streptomycin biosynthesis (*strD* and *strE*) in *S. griseus*, and show a similar arrangement, in that the TGA codon of GraD overlaps the ATG codon of GraE, suggesting that these genes form a transcription unit. Genes that overlap by a single nucleotide are common in bacteria when the gene products are required in equimolar quantities (Normark et al. 1983). The overlap of a TGA stop codon and an ATG start codon has also been described for genes involved in antibiotic biosynthesis and antibiotic resistance (Fernández-Moreno et al. 1991; Jenkins and Cundliffe 1991). The inducible lincomycin resistance in *S. lividans* is conferred by the expression of two linked genes, *lrm*, encoding a ribosomal methyltransferase, and *mgt*, encoding a glucosyltransferase (Jenkins and Cundliffe 1991). Data obtained from studies of the translation of these proteins have shown that release of the Lrm protein is required to allow a minus-1 frameshift in order to initiate translation of the *mgt* RNA. The mgt coding sequence is not translated independently of the Lrm coding sequence. Expression of the dTDP-glucose 4,6-dehydratases gene in *E. coli*, was obtained from a plasmid (B24) that contained the entire GraE gene but only about 150 nt of the TDP-glucose synthetase

Fig. 6 Alignment of ORF4 with similar antibiotic transporters [Sgltcm, *S. glaucescens* putative tetracenomycin transport protein (Guilfoile and Hutchinson 1992); Scommr, *S. coelicolor* methylenomycin resistance gene product (Neal and Chater 1987); Sritet, *S. rimosus* tetracycline resistance gene product (Reynes et al. 1988)], and the consensus sequence. The amino acid sequences were aligned using PILEUP (Devereux et al. 1984). Capital letters denote conserved amino acids

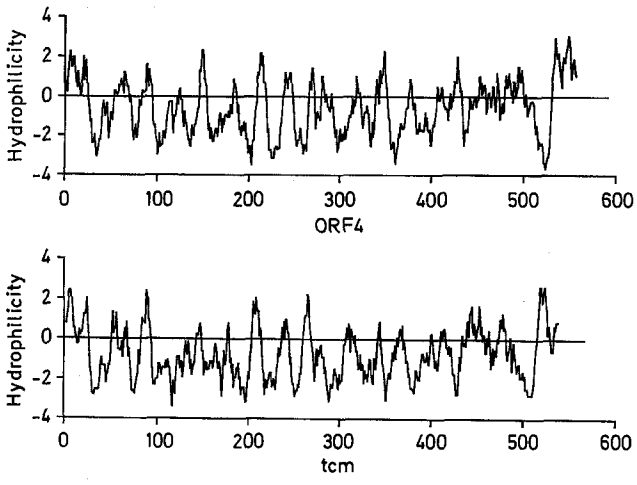


Fig. 7 Hydrophilicity plots for ORF4, and the putative *S. glaucensens* tetracenomycin transport protein (Tcm) (Guilfoile and Hutchinson 1992). The data for each plot were obtained with PEPTIDESTRUCTURE in the GCG software package, and plotted on a Macintosh computer using Cricket Graph (CA Associates)

gene. This indicates that, in *E. coli*, the GraE coding sequence can be translated independently of the GraD coding sequence. There is a strong Shine-Dalgarno sequence (GGAGG) inmmediately (7 nt) preceding the GraE initiation codon.

ORF4 was found to share similarity with a family of transporter proteins that are involved in antibiotic resistance. Two types of these transporters have been described. The first group is represented by proteins which mediate resistance by an ATP-dependent efflux of the antibiotic. Resistance genes belonging to this group have been isolated from macrolide-producing organisms (Schoner et al. 1992) but also from *S. peuceticus*, the producer of daunorubicin and doxorubicin (Guilfoile and Hutchinson 1991), and the bleomycin producer *S. verticillus* (Calcutt and Schmidt 1991). The amino acid sequences deduced from the genes encoding those proteins contain conserved regions that are similar to the ATP-binding domains present in a family of proteins known for their ability to confer multiple drug resistance in human tumor cells (Riordan and Ling 1985). The second group of transporters is represented by membrane proteins that mediate antibiotic efflux in an ATP-independent, probably proton-dependent process. Genes coding for proteins of this type have been isolated from *Streptomyces* species which produce structurally quite different antibiotics [tetracycline (Reynes et al. 1988), methylenomycin (Neal and Chater 1987), lincomycin, (Hui-Zahn et al. 1992) and actinorhodin]. All of these proteins share similarities with sugar transport proteins isolated from *E. coli* (Hui-Zhan et al. 1992). The resemblance between ORF4 and resistance genes of the second group indicates that the protein encoded by ORF4 (GraR) is most likely to be a member of this transport protein family (GraR). Several attempts have been made to align the proteins listed above in order to find consensus sequences (Rouch et al. 1990; Sheridan and Chopra 1991; Hui-Zahn et al. 1992). Based on these results, the proteins have been subdivided into several groups [MmR, QacA, transporter of antiseptic compounds, and tetracycline transporter (Rouch et al. 1990)]. Among the different transport proteins, GraR is most closely related to the family of methylenomycin resistance proteins which may include the actinorhodin and tetracenomycin transporters (Figs. 6, 7). Hydrophilicity plots comparing GraR (Fig. 7) to the tetracenomycin transporter support the idea that GraR is an integral membrane protein.

The possible functions of ORF1 and ORF5 are unknown since they failed to show significant similarities with other known genes in the databases. ORF1 may be expressed as part of an operon that includes GraD and GraE; the DNA sequence between the stop codon of GraE and possible start codons of ORF1 lack a strong inverted repeat structure, a common transcription termination signal found in *Streptomyces* species. The region centered about nt 82-88 can be viewed as an imperfect inverted repeat that could form a 15/16 bp stem loop and possible termination signal. The region separating the diverging ORF3, ORF4 and ORF5 contains a segment (nt 5354–5400) which represents an incomplete inverted repeat centered around nt 5375. This sequence may be involved in regulatory processes. Divergent transcription from the promoter regions of resistance genes in antibiotic producing *Streptomyces* has been described in many cases (Neal and Chater 1987). In one example a gene, transcribed divergently from a resistance gene was found to be a repressor gene (Caballero et al. 1991). However, no significant similarity of ORF5 to any of these divergently transcribed genes was found in our studies.

Acknowledgments We are grateful to Alice Lee and Brian Tillmann for helpful technical assistance. This work has been supported by a postdoctoral scholarship from the Deutsche Forschungsgemeinschaft (Be1389 to A.B.), a National Research Service Award (GM 07750 to T.M.S.), and a grant from the National Institutes of Health (GM 43345 to H.G.F.).

References

Bibb MJ, Findlay PR, Johnson MW (1984) The relationship between base composition and codon usage in bacterial genes and its use in the simple and reliable identification of protein codon sequences. Gene 30:157–166

Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schoner BE (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene 116:43–49

Caballero JL, Malpartida FA, Hopwood DA (1991) Transcriptional organization and regulation of an antibiotic export complex in the producing *Streptomyces* culture. Mol Gen Genet 228:372–380

Calcutt MJ, Schmidt FJ (1991) Bleomycin biosynthesis: characterization of the resistance region of the producer streptomycete. International Symposium on Biology of Actinomycetes, University of Wisconsin, Madison wis USA

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Fernández-Moreno MA, Caballero JL, Hopwood DA, Malpartida F (1991) The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. Cell 66:769–780

Griesebach H (1978) Biosynthesis of sugar components of antibiotic substances. Adv Carbohyd Chem Biochem 35:81–126

Guilfoile PG, Hutchinson CR (1991) A bacterial analog of the *mdr* gene of mammalian tumor cells in present in *Streptomyces peuceticus*, the producer of daunorubicin and doxorubicin. Proc Natl Acad Sci USA 88:8553–8557

Guilfoile PG, Hutchinson CR (1992) Sequence and transcriptional analysis of the *Streptomyces glaucescens tcmAR* tetracenomycin C resistance and repressor gene loci. J Bacteriol 174:3651–3658

He X-G, Chang C-C, C-J. C, Vederas JC, McInnes AG, Walter JA, Floss HG (1986) Further studies on the biosynthesis of granaticin. Z Naturforsch 41c:215–221

Hopwood DA, Bibb MJ, Chater KF, Kieser T, Bruton C, Kieser HM, Lydiate DJ, Smith CP, Ward JM (1985) Genetic manipulation of *Streptomyces*: a laboratory manual. John Innes Foundation, Norwich, UK

Hui-Zahn Z, Schmidt H, Piepersberg W (1992) Molecular cloning and characterization of two genes from *Streptomyces lincolnensis* 78–11. Mol Microbiol 6:3147–3157

Jenkins G, Cundliffe E (1991) Cloning and characterization of two genes from *Streptomyces lividans* that confer resistance to lincomycin and macrolide antibiotics. Gene 108:55–62

Khosla C, McDaniel R, Ebert-Khosla S, Torres R, Sherman DH, Bibb MJ, Hopwood DA (1993) Genetic construction and functional analysis of hybrid polyketide synthases containing heterologous acyl carrier proteins. J Bacteriol 175:2197–2204

Malpartida F, Hopwood DA (1986) Physical and genetic characterization of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3 (2). Mol Gen Genet 205:66–73

Mansouri K, Piepersberg W (1991) Genetics of streptomycin production in *Streptomyces griseus*: nucleotide sequence of five genes, *strFGHIK*, including a phosphatase gene. Mol Gen Genet 228:459–469

Martin JF, Liras P (1989) Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. Annu Rev Microbiol 43:173–206

Matsudaira PT (1993). A practical guide to protein and peptide purification for microsequencing. Academic Press, San Diego

Neal RJ, Chater KF (1987) Nucleotide sequence analysis reveals similarities between proteins conferring methylenomycin A resistance in *Streptomyces* and tetracycline resistance in eubacteria. Gene 8:229–241

Normark S, Bergström S, Edlund T, Grundström T, Jaurin B, Lindberg FP, Olsson O (1983) Overlapping genes. Annu Rev Genet 58:229–241

Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 183:63–98

Pissowotzki K, Mansouri K, Piepersberg W (1991) Genetics of streptomycin production in *Streptomyces griseus*: molecular structure and putative function of genes *strELMB2N*. Mol Gen Genet 231:113–123

Reynes LP, Calmels T, Drocourt D, Tibay G (1988) Cloning, expression in *Escherichia coli* and nucleotide sequence of a tetracycline resistance gene from *Streptomyces rimosus*. J Gen Microbiol 134:585–598

Riordan JR, Ling V (1985) Genetic and biochemical characterization of multidrug resistance. Pharmac Ther 28:51–75

Rouch DA, Cram DS, DiBeradino D, Littlejohn TG, Skurray RA (1990) Effluxmediated antiseptic resistance gene *qacA* from *Staphylococcus aureus*: common ancestry with tetracycline- and sugar-transport proteins. Mol Microbiol 4:2051–2062

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Schoner B, Geistlich M, Rosteck Jr P, Rao RN, Seno E, Reynolds P, Cox C, Burgett S, Hershberger C (1992) Sequence similarity between macrolide resistance determinants and ATP-binding transport proteins. Gene 115:93–96

Sheridan RP, Chopra I (1991) Origin of tetracycline efflux proteins: conclusions from nucleotide sequence analysis. Mol Microbiol 5:895–900

Sherman DH, Malpartida F, Bibb MJ, Kieser HM, Bibb MJ, Hopwood DA (1989) Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tü22. EMBO J 8:2717–2725

Sherman DH, Kim E-S Bibb MJ, Hopwood DA (1992) Functional replacement of genes for individual polyketide synthase components in *Streptomyces coelicolor* A3(2) by heterologous genes from a different polyketide pathway. J Bacteriol 174: 6184–6190

Snipes CE, Brillinger G-U, Sellers L, Mascaro L, Floss HG (1977) Sterochemistry of the dTDP-glucose oxidoreductase reaction. J Biol Chem 252:8113–8117

Snipes CE, Chang C-J, Floss HG (1979) Biosynthesis of the antibiotic granaticin. J Amer Chem Soc 101:701–706

Stockmann M, Piepersberg W (1992) Gene probes for the detection of 6-deoxyhexose metabolism in secondary metabolite-producing *Streptomycetes*. FEMS Microbiol Lett 90:185–190

Thorson JS, Lo SF, Liu H-W (1993b) Biosynthesis of 3,6-dideoxyhexoses: new mechanistic reflections upon 2,6-dideoxy, 4,6-dideoxy and amino sugar construction. J Am Chem Soc 115:6993–6994

Thorson JS, Lo SF, Liu H-W, Hutchinson CR (1993a) Molecular basis of 3,6-dideoxyhexose biosynthesis: elucidation of CDP-ascarylose biosynthetic genes and their relationship to other 3,6-dideoxyhexose pathways. J Am Chem Soc 115: 5827–5828

Vara JA, Hutchinson CR (1988) Purification of thymidine-diphospho-D-glucose 4,6-dehydratase from an erythromycin-producing strain of *Saccharopolyspora erythrea* by high resolution liquid chromatography. J Biol Chem 263:14992–14995

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119

Note added in proof

Recent database searches have revealed that *gra* ORF5 is highly similar to glycosyltransferases (Hernandez et al. (1993) Gene 134:139–140), and *gra* ORF1 shows a high degree of similarity to *actVI* ORF3 (Fernández-Moreno et al. (1994) J Biol Chem 269: 24854–24863).