J. Antibiot. 59(1): 18–28, 2006

ORIGINAL ARTICLE

THE JOURNAL OF ANTIBIOTICS

DNA Sequencing and Transcriptional Analysis of the Kasugamycin Biosynthetic Gene Cluster from *Streptomyces kasugaensis* M338-M1

Souichi Ikeno, Daisuke Aoki, Masa Hamada, Makoto Hori, Kayoko S. Tsuchiya

Received: August 23, 2005 / Accepted: December 4, 2005

Abstract *Streptomyces kasugaensis* M338-M1 produces the aminoglycoside antibiotic kasugamycin (KSM). We previously cloned, sequenced and characterized the KSM acetyltransferase, transporter, and some of the biosynthetic genes from this strain. To identify other potential genes in a chromosome walk experiment, a 6.8-kb *Eco*RI-*Pst*I region immediately downstream from the KSM transporter genes was sequenced. Five open reading frames (designated as *kasN*, *kasO*, *kasP*, *kasQ*, *kasR*) and the 5′ region of *kasA* were found in this region. The genes are apparently co-transcribed as bicistrons, all of which are co-directional except for the *kasPQ* transcript. Homology analysis of the deduced products of *kasN*, *kasP*, *kasQ* and *kasR* revealed similarities with known enzymes: KasN, D-amino acid oxidase from *Pseudomonas aeruginosa* (35% identity); KasP, F420-dependent H4MPT reductase from *Streptomyces lavendulae* (33% identity); KasQ, UDP-*N*-acetylglucosamine 2-epimerase from *Streptomyces verticillus* (45% identity); and KasR, NDP-hexose 3,4-dehydratase from *Streptomyces cyanogenus* (38% identity); respectively. A gel retardation assay showed that KasT, a putative pathway-specific regulator for this gene cluster, bound to the upstream region of *kasN* and to the intergenic region of *kasQ-kasR*, suggesting that the expression of these operons is under the control of the regulator protein.

Keywords kasugamycin, biosynthetic gene cluster, *Streptomyces kasugaensis*

Introduction

The genus *Streptomyces* is a group of Gram-positive bacteria that produce a wide variety of secondary metabolites including antibiotics and other pharmacologically active agents. The biosynthetic pathways and genetic regulation that result in the production of these compounds are poorly understood and need to be studied to develop the ability to increase antibiotic productivity and to design and produce new antibiotics, potentially with hybrid structures that would result in more effective properties, particularly against resistant infective agents.

Kasugamycin (KSM) is an aminoglycoside antibiotic produced by *Streptomyces kasugaensis* M338-M1 [1] that is effective against *Pyricularia oryzae* Cavara and is widely used in agriculture in Japan to prevent rice blast disease. A gene encoding an enzyme that inactivates the antibiotic by acetylation of the 2′-NH_2 of KSM was cloned from *Streptomyces kasugaensis* MB273-C4, another KSM producing strain, and named *kac* (JP. A-05-23187, Hirasawa *et al.* 1993). In *Streptomycetes*, genes required for biosynthesis of an antibiotic are usually clustered and linked to a gene or genes for self-protection from its own antibiotic. Genes associated with regulation of antibiotic biosynthesis are also often found in clusters. Starting with identification of KSM acetyltransferase (kac^{338}) [2], we found three KSM biosynthetic genes (*kasC*, *kasD*, and *kasJ*) [3, 4], three KSM transporter genes (*kasK*, *kasL*, and

S. Ikeno (Corresponding author), **D. Aoki, M. Hori, K. S. Tsuchiya:** Showa Pharmaceutical University, 3-3165 Higashi Tamagawagakuen, Machida-shi, Tokyo 194-8543, Japan.
E-mail: ikeno@ac.shoyaku.ac.jp

M. Hamada: Institute of Microbial Chemistry, 3-14-23 Kamiosaki, Shinagawa-ku, Tokyo 141-0021, Japan

Table 1 Strains and plasmids

Strains and plasmids	Genotype and genetic construct	Source or reference
Strains		
S. kasugaensis		
M338-M1	Kasugamycin producing strain	1)
E. coli		
DH5α	ϕ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *deoR recA*1 *endA*1 *hidR*17(r_K^-, m_K^+) *phoA supE*44 λ-*thi*-1 *gyrA*96 *relA*1	TOYOBO
TH2	*supE*44 *hsdS*20(r_B^-, m_B^+) *recA*13 *ara*-14 *proA*2 *lacY*1 *galK*2 *rpsL*20 *xyl*-5 *mtl*-1 *thi*-1 *trpR*624	TaKaRa
BL21(DE3)	F′ *ompT* (r_B^-, m_B^+)	Novagen
Plasmids		
pUC118	Cloning vector. Amp^r. 3.1-kb.	TaKaRa
pKF 3	Cloning vector. Sm^s, Cm^r. 2.2-kb.	TaKaRa
pET-32a(+)	Expression vector derived from pBR322, containing *trxA* [*1-109*] and His-tag expressed from T7 promoter for construction of Trx hybrid proteins. Amp^r. 5.7-kb.	Novagen
pSKE 4	pKF 3 derivative containing 8.4-kb *Kpn*I-*Kpn*I fragment. 11-kb.	5)
pET-KasT	pET-32a(+) derivative containing 1.0-kb *Nco*I-*Bam*HI fragment. Trx-KasT expression plasmid. 6.7-kb.	4)

kasM) [5] and a putative regulatory gene (*kasT*) [4] in the upstream region of *kac*[338].

In the present paper, we report the nucleotide sequence of a 6.8-kb region downstream from the KSM transporter genes and propose the existence of KSM biosynthetic genes in that segment of the chromosome. We suggest that these genes are transcribed as three bicistronic transcripts and that these operons are under the control of KasT.

Materials and Methods

Strains, Growth Conditions, and Plasmids

Characteristics of bacterial strains and plasmids used in this work are summarized in Table 1. *S. kasugaensis* M338-M1 is maintained at the Institute of Microbial Chemistry, Tokyo, Japan. *E. coli* TH2 (TaKaRa), *E. coli* DH5α (TOYOBO), *E. coli* BL21(DE3) (STRATAGENE), pKF 3 (TaKaRa), pUC118 (TaKaRa) and pET-32a(+) (Novagen) were of commercial origin. All other plasmids were produced in the present study.

S. kasugaensis M338-M1 was grown in MR medium (KSM-producing medium) [5] with shaking at 27°C for 72 hours. *E. coli* TH2 transformants were grown at 37°C in L-broth containing 12 μg/ml chloramphenicol and 50 μg/ml streptomycin. *E. coli* DH5α and BL21(DE3) transformants were grown at 37°C in Luria-Bertani (LB) medium containing 50 μg/ml ampicillin.

DNA Sequencing and Sequence Analysis

Isolation of genomic DNA from *S. kasugaensis* M338-M1 and cloning of the region downstream from KSM transporter genes were described previously [2, 5]. The nucleotide sequence of both strands was determined using an ALF*red*™ DNA sequencer (Amersham Biosciences). Sequencing reactions were carried out with Cy5™ AutoCycle™ Sequencing Kit (Amersham Biosciences) according to the supplier's instructions. M13-40 universal primer, M13-reverse primer and some synthesized oligonucleotide primers (labeled with Cy5, purchased from Amersham Biosciences) were used as sequencing primers. Amino acid sequences of the corresponding proteins were deduced from the nucleotide sequences of appropriate regions of DNA using DNASIS-Mac version 3.7 (Hitachi Software Engineering). FramePlot 2.3.1 [6] was used to search for open reading frames (ORF). Amino acid sequences of potential gene products were compared with the SWISS-PROT and PIR databases using FASTA [7] and BLAST [8]. Nucleotide sequence data reported in this paper have been deposited in the DDBJ, EMBL and GenBank nucleotide sequence databases with the accession number AB120043.

Isolation of Total RNA and Northern Blot Analysis

Isolation of total RNA from *S. kasugaensis* M338-M1 was as described previously [5]. The RNA sample was quantified by absorbance at 260 nm and a portion was

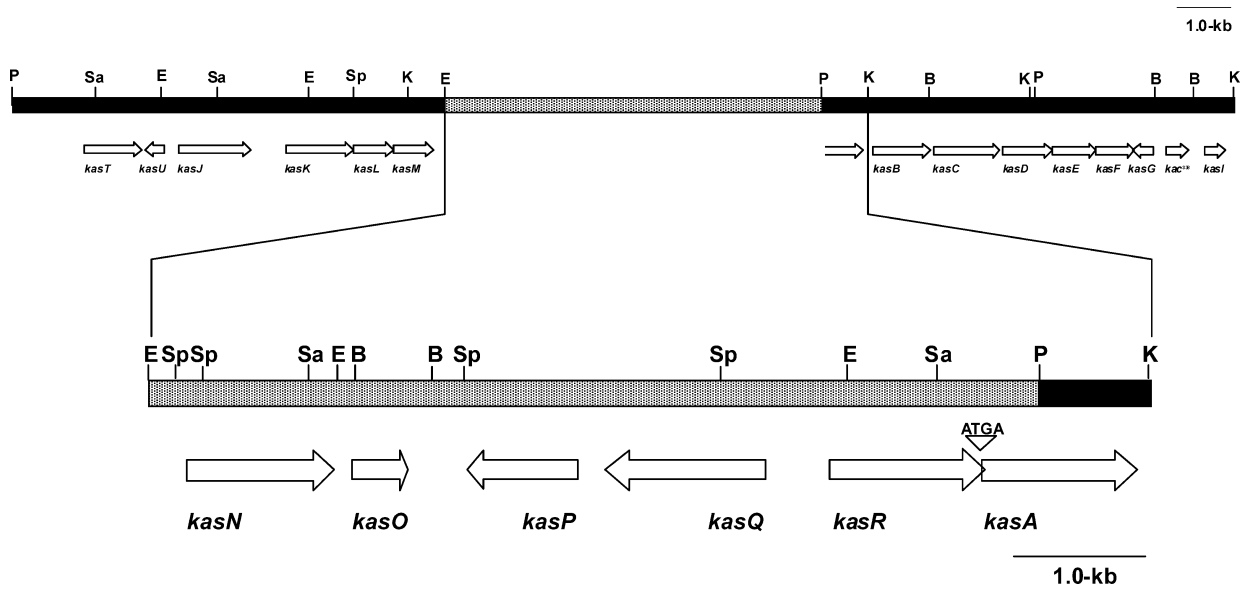

Fig. 1 Restriction map of the kasugamycin biosynthetic gene cluster from *Streptomyces kasugaensis* M338-M1.

The gray region of the restriction map is the subject of the present paper. The black region has been previously reported [4, 5]. The open arrows indicate the deduced genes and direction of transcription. The start codon (ATG) of *kasA* overlaps the stop codon (TGA) of the adjacent *kasR*. Abbreviations: B, *Bam*HI; E, *Eco*RI; K, *Kpn*I; P, *Pst*I; Sa, *Sac*I; Sp, *Sph*I.

electrophoresed (10 μg/lane) on a 1.2% agarose gel and transferred to a cellulose nitrate membrane (Schleicher & Schuell). Hybridization was carried out at 42°C for 16 hours in hybridization buffer (50% formamide, 900 mM NaCl, 6 mM EDTA, 90 mM Tris-HCl [pH 7.5], 10×Denhardt's, 0.1% SDS, 100 μg/ml Salmon testis DNA), followed by washing at 68°C for 1 hours with 2×SSC (300 mM NaCl, 30 mM trisodium citrate). The 811 bp *Sph*I-*Sac*I fragment, which corresponds to the internal region of the *kasN* structural gene, was used as probe (Fig. 1). The DNA fragment was labeled with [α-^{32}P]dCTP using a Random Primer DNA Labeling Kit (TaKaRa).

RT-PCR

RT-PCR was conducted as described previously [5]. 3′-pKS17 (5′-CAGTCGGACTCCCTGATGCCCAGGA-3′), 3′-Q1 (5′-ACGGAACTCGCCCCGAAGCCATCAA-3′) and 3′-ABAM (5′-CCGGATCCTCACCGCTGGCCACA-CGCCACTTCGTA-3′) were used as cDNA synthesis primers for *kasNO*, *kasPQ* and *kasRA* transcripts, respectively. Three sets of primers [5′-pKS16b (5′-AGAACAGGTGGCTCAGCCCCGCGGA-3′)/3′-pKS17, 5′-kasP (5′-CGGCAGTGCCAAGTCGCCGGTCAGC-3′)/3′-Q1 and 5′-RT20 (5′-ATCGTGTCCGGCGCGACCT-TCGTCA-3′)/3′-ABAM] were used in subsequent PCR amplification with 30 cycles of 30 seconds denaturation at 98°C and 2 minutes annealing/extension at 72°C. RT-PCR products were electrophoresed on a 0.6% agarose gel and visualized using ethidium bromide staining.

Determination of Transcription Start Sites

The 5′ end of *kasNO* mRNA was determined by the 5′ RACE method with a 5′-Full RACE Core Set (TaKaRa) according to the manufacturer's instructions [9]. RT reaction was performed in a 15 μl solution containing 200 pmol of 5′-phosphorylated primer 3′-PKasN (5′-ATCGCCCACCACTGCCACGTCCACA-3′), 5 U of AMV reverse transcriptase XL and 5 μg of total RNA, and incubated at 50°C for 1 hour. A negative control lacked reverse transcriptase. The first PCR mixture contained 20 pmol each of 5′-RACE15a (5′-CTCCGCCCCGCAC-CACCCCT-3′) and 3′-RACE15a (5′-GGAGGTGTCACG-GGACGGTCTT-3′). The reaction was carried out for 15 cycles with denaturation for 30 seconds at 98°C and annealing and extension for 2 minutes at 72°C. The resulting PCR product (1 μl) was used as a template for a second PCR amplification with primers 5′-RACE15b (5′-CCCTGCAGAAAGGAAGACGGCCCGATGA-3′) and 3′-RACE15b (5′-CCGTCGACCCGGCCACCGCATGCGA-AGA-3′) under the same conditions as described above. The major 230 bp product was digested with *Sal*I and *Pst*I (recognition sites are underlined in the primers), cloned into pUC118, and the DNA sequence of the product was determined.

Overexpression and Purification of Trx-KasT

Overexpression and purification of the fusion protein was conducted as described previously [4] using the expression plasmid (pET-KasT) for Trx-KasT which was KasT protein preceded by a thioredoxin (Trx) peptide and six histidine residues.

Gel Retardation Assay

The upstream region of *kasN* was amplified by PCR using 5′-pKS15CP (5′-CGTTCACCGCGCAAGCCGCCCTATT-3′) and 3′-pKS15B (5′-ATCGCCCACCACTGCCACGTC-CACA-3′) as primers. The intergenic region of *kasQ-kasR* was also amplified in three separate segments (19A, 19B and 19C). PCR primer pairs were 5′-pKS19B (5′-GAGCT-TGATGGCTTCGGGGCGAGTT-3′) and 3′-pKS19 (5′-GCGCAATCAACCCTGCGACGGGGAA-3′) for 19A, 5′-pKS19C (5′-TTGACATGCAAGGGAAGTCGAAA-3′) and 3′-pKS19C (5′-CGCCGCACGCTAACACAGCGAC-C-3′) for 19B, and 5′-pKS19 (5′-CACTCCGGCTGCG-GTACATGAGTAA-3′) and 3′-pKS19B (5′-GTACCGTC-CCCACCAGCGTGAATGA-3′) for 19C, respectively. These PCR fragments were labeled at the 5′ ends with [γ-^{32}P]ATP and T4 polynucleotide kinase. For the gel retardation assay, 0.3 to 0.5 ng of the ^{32}P-labeled DNA fragments (10,000 to 15,000 cpm) were incubated with 3 to 5 μg of Trx-KasT at 30°C for 1 hours in a buffer containing 10 mM Tris-HCl (pH 8.0), 50 mM KCl, 5 mM DTT, 10% (v/v) glycerol, 12.5 mM spermidine and 1 μg of poly(dI-dC)·poly(dI-dC), in a total volume of 20 μl. Protein-DNA complexes were separated from free DNA on 4% polyacrylamide gels by electrophoresis at 150 V for about 1 hour.

Results

Sequence of a 6.8-kb Region Downstream of the KSM Transporter Genes (*kasKLM*)

We previously cloned the genomic DNA region (*Sac*I-*Eco*RI) including KSM transporter genes (*kasKLM*) and its downstream region (*Eco*RI-*Pst*I) [5]. To search for genes involved in KSM biosynthesis, we sequenced the 6,861 bp *Eco*RI-*Pst*I DNA fragment which had a G+C content of 69.5%. Open reading frames (ORFs) were located based on codon usage and the third codon position bias which are characteristic of *Streptomyces* genes [6, 10], resulting in identification of five ORFs, designated as *kasN*, *kasO*, *kasP*, *kasQ*, and *kasR*, and the 5′ region of *kasA* in this region. The *kasP* and *kasQ* genes are encoded on the opposite strand from the others (Fig. 1).

Characterization of the ORFs and Their Putative Proteins

kasN

The *kasN* gene is located 520 bp downstream of *kasM* in the same direction of transcription. The *kasM-kasN* intergenic region contains a relatively low G+C content region for *Streptomyces* DNA (G+C%=50%, −263∼−134 nucleotides upstream of the *kasN* start codon), suggesting that the intergenic region is where the DNA unwinds to initiate transcription of the downstream genes. The deduced amino acid sequence of KasN (383 amino acids, M_r of 41,699, pI of 5.12) included a glycine box motif (GDGLIG) at positions 13∼18 that could allow binding of an FAD. Homology analysis of KasN showed similarity to D-amino acid oxidase (Protein ID: AAG07936.1) [11] from *Pseudomonas aeruginosa* (35% identity) and glycine oxidase (Protein ID: BAB05153.1) [12] from *Bacillus halodurans* (32% identity). In addition, KasN had some similarity to N-formimidoyl fortimicin A (FI-FTM A) synthase [13] from *Micromonospora olivasterospora* (31% identity). These enzymes are responsible for the oxidation of glycine.

kasO

The *kasO* gene is located 131 bp downstream of *kasN* with the same orientation of transcription. No typical promoter sequences were found in the *kasN-kasO* intergenic region. A palindromic sequence (ΔG=−46.10 kcal/mol) which could act as a putative transcriptional terminator for *kasO*, was found 119∼154 nucleotides downstream of the *kasO* stop codon. This observation suggested that *kasO* was transcribed with *kasN* as a single transcript. The deduced KasO protein (149 amino acids, M_r of 15,856, pI of 4.96) showed no significant similarity with known enzymes.

kasP

The *kasP* gene is located 443 bp downstream of *kasO* in the opposite orientation. No typical promoter sequences were found in the *kasO-kasP* intergenic region. A palindromic sequence (ΔG=−39.20 kcal/mol) which could serve as a transcriptional terminator for *kasP*, was found 90∼133 nucleotides downstream of the *kasP* stop codon. The deduced KasP protein (281 amino acids, M_r of 30,233, pI of 6.66) had similarity with an F420 dependent H4MPT reductase (MmcI) [14] involved in mitomycin C biosynthesis in *Streptomyces lavendulae* (33% identity). In addition, KasP showed similarity to the following enzymes: Rif17 [15], an alkanal monooxygenase involved in rifamycin biosynthesis in *Amycolatopsis mediterranei* (32% identity); MitK [14], an F420 dependent H4MPT dehydratase involved in mitomycin C biosynthesis in

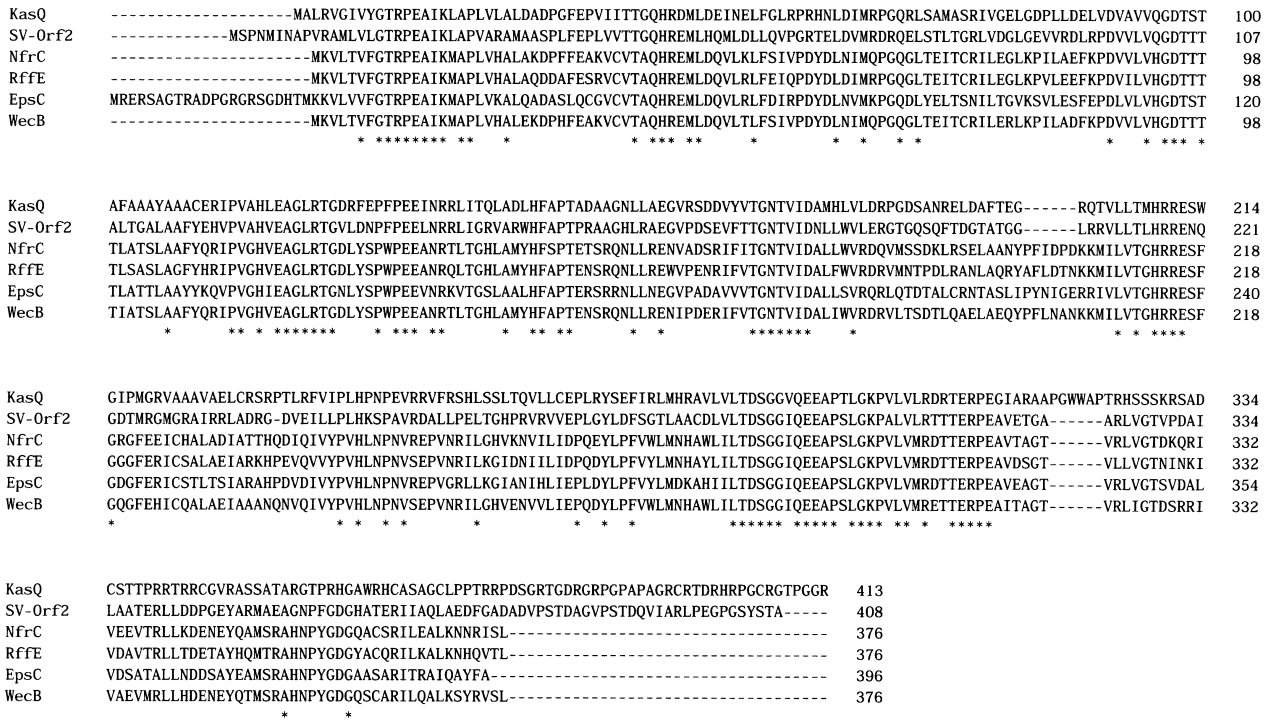

Fig. 2 Alignment of the amino acid sequences of KasQ and other UDP-*N*-acetylglucosamine 2-epimerases.

Protein names are indicated to the left of the amino acid sequences. *Streptomyces verticillus* SV-Orf2 (Protein ID: AAF68965.1), *Escherichia coli* NfrC (Protein ID: AAC36847.1), *Yersinia pestis* RffE (Protein ID: CAC93332.1) *Ralstonia solanacearum* EpsC (Protein ID: AAA91626.1), *Salmonella enterica* WecB (Protein ID: CAD09396.1). Asterisks mark the conserved residues.

Streptomyces lavendulae (31% identity).

kasQ

The *kasQ* gene is located 206 bp upstream from *kasP* in the same orientation. The deduced KasQ protein (413 amino acids, M_r of 45,121, pI of 9.45) had significant similarity (41∼45% identity) to UDP-*N*-acetylglucosamine 2-epimerases from *Streptomyces verticillus* (Protein ID: AAF68965.1), *Escherichia coli* (Protein ID: AAC36847.1) [16], *Yersinia pestis* (Protein ID: CAC93332.1) [17], *Ralstonia solanacearum* (Protein ID: AAA91626.1) [18] and *Salmonella enterica* (Protein ID: CAD09396.1) [19]. Bacterial UDP-*N*-acetylglucosamine 2-epimerase catalyzes the reversible epimerization at C-2 of UDP-*N*-acetylglucosamine (UDP-GlcNAc). Recently, the X-ray structure of UDP-GlcNAc 2-epimerase from *E. coli* was determined and amino acid residues that are involved in interactions between this enzyme and UDP-GlcNAc were identified [20]. Three amino acid residues (Arg10, Ser290, and Glu296) are associated with binding to UDP, and six amino acid residues (Lys15, Asp95, Glu117, Glu131, Arg135, and His213) have been assigned to GlcNAc binding. These amino acid residues were conserved in KasQ (Fig. 2), suggesting that *kasQ* encodes a UDP-GlcNAc 2-epimerase for KSM biosynthesis.

kasR

The *kasR* gene is located 493 bp upstream from *kasQ* in the opposite direction of transcription. The deduced KasR protein (399 amino acids, M_r of 42,562, pI of 6.06) showed significant similarity to hexose C-3 dehydrases which are involved in 3-deoxygenation of deoxysugar moieties of some secondary metabolites. Homology analysis for KasR revealed similarity with LanQ [21], an NDP-hexose 3,4-dehydratase involved in landomycin biosynthesis in *Streptomyces cyanogenus* (38% identity); RdmI [22], a hexose C-3 dehydratase involved in rhodomycin biosynthesis in *Streptomyces purpurascens* (37% identity); UrdQ [23], an NDP-hexose 3,4-dehydratase involved in the formation of the L-rhodinose moiety of urdamycin in *Streptomyces fradiae* (37% identity); AscC [24], and a CDP-4-keto-6-deoxy-D-glucose-3-dehydrase involved in ascarylose biosynthesis in *Yersinia pseudotuberculosis* (36% identity). These enzymes are dependent on pyridoxamine 5′-phosphate (PMP) and contain an iron-sulfur cluster. Alignment of these proteins suggested that some amino acid residues were conserved (Fig. 3). *Yersinia pseudotuberculosis* AscC catalyzing the C-3 deoxygenation

```
KasR   ---------------------MEIIHAGGDGTLLSFGRPTYDEREIEAVVAALRSGELATGVTTRKFEAEFAESFGFAHALAVTSGSTANLLACAAMLELGR----ARPGDRVIVSGATF  95
LanQ   --MSVHRKLVLDEVRKYHQDTQGDGVFVPGTTEIWPSGAVLDEDDRIALVEAALEMRIAAGRSSRKFESSFARKMKRRKAHLTNSGSSANLLATTALTSHLLEDRRLRPGDEVITVAAGF 118
RdmI   --MTDTKALVLELARKYHKEQETRG-FEPGVTPILSSGAVLDEEDRGALVEAALEMRIAAGVRTRKFESEFARYFRLRKAHMVNSGSSANLLAVSALTSPKLKDLRLRPGDEAITVGAGF 117
UrdQ   --MGDRKELVLEEVRTYHRDVSPDQEFIPGTTEIWPSGAVLDESDRVALVEAALDMRIAAGTSSRKFESAFARRLKRRKAHLTNSGSSANLLAVSALTSHVLEDRRLRPGDEVITVAAGF 118
AscC   MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTKELQLMVEASLDGWLTTGRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKLGVRALKPGDEVITVAAGF 120
                                                  * *       *     **               *** *****  *            ***  *   * *

KasR   VSAVTPVVQLGLVPVFVDVAAGHVNVDLDLVEQAVVEHGARGVLLPHTLGQALPMDRLAEIKRRHGVFVIEDCCESLGAADGSTPVGSAADVATFSFYAGHHLTMGEGGVAAGHSAEIDS 215
LanQ   PTTVNPILQNGLIPVFVDVELGTYKTTADRVAQAIGPK-TRAIIIAHTLGNPFEVAEIAQLAVDHDLFLIEDNCDAVGSTYDGKLTGTFGELTTVSFYPAHHLTMGEGGCVLTADLALAR 237
RdmI   PTTINPLVQNGLTPVFVDVELGTYNASLAQIEAAISPR-TRLIMIAHALGNPFPVAEVAELARRHDLFFIEDNCDAVGSTYQGRLTGTFGDLSTVSFYPAHHITAGEGGCVLTSSLELAR 236
UrdQ   PTTVNPILQNGLVPVFVDVDLPTYNATAERVAQAIGPK-TRAIIIAHALGNPFEVAEMAQLAEEHDLFLIEDNCDAVGSTYDGQLTGTFGDLTTVSFYPAHHLTMGEGGCVLTSNLSLAR 237
AscC   PTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDK-TKAIMIAHTLGNLFDLAEVRRVADKYNLWLIEDCCDALGSTYDGKMAGTFGDIGTVSFYPAHHITMGEGGAVFTQSAELKS 239
          *  * ** ******             *           * **                  *** *   *       *      * ***  ** * ****

KasR   VLRSLRAFGRNPDYRLG-------RFEHPVDDRPLAPEERYIHLRLGYNAKITDFQAAFGRVQLTRHAELARQRRQLAQELVPVLREFGWGVLGDPVSPGASP----FAVATLLPEGLPL 324
LanQ   IVESLRDWGRDCWCEPGESDKCLKRFKYQMGTLPAGYDHKYIFSHVGYNLKATDLQAALGLTQLAKLDDFVEARKRNWRRLRDGLDGVP-HLLLPEATPRRHPSWFGFVLTVSPDAPFKR 356
UrdQ   IVESLRDWGRDCWCEPGENDRCLKRFKYQMGTLPAGYDHKYIFSHVGYNLKATDIQAALGLTQLAKLDDFIEARQRNWRRLREGLDGVP-GLLLPEPTPRSQPSWFGFVITVAPDAPFSR 356
RdmI   IVESMRDWGRDCWCEPGTDNTCLKRFDYQLGTLPAGYDHKYIFSHVGYNLKATDLQGALALSQLRKVDDFGAARRHNWQRLRDGLADVP-GLLLPRATPGSDPSWFGFALTVLPDAGFTR 355
AscC   IIESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGYDHKYTYSHLGYNLKITDMQAACGLAQLEPIEEFVEKRKANFKYLKDALQSCADFIELPEATENSDPSWFGFPITLKEDSGVSR 359
          * *  **      *       **       *      *     *** * ** * *    **         *      *   *                    *    *

KasR   TRAVGVLIEHGIDPRGFLGASQPHQPCFDGVTKVVHEPYLHTRTLAERGLLLGCPPRTDRAAAVKALRRALESLS--- 399
LanQ   AELVDFLEGRKIGTRRLFAGNLTRHPAYIDQPHRVVGNLDNSDIITEQTFWVGVYPGLTDEMLDYVISSVKEFVEARG 434
RdmI   RDLVDFLEGRRIGTRRLFGGNITRHPAYEDVRYRIVDELVNCDTVTEDTFWIGVYPGLTKDMLDYVVESITEFCTRKG 433
UrdQ   AELVDFLEDRKIGTRRLFAGNLTRHPAYIDQPHRIVGELTNSDLVTEQTFWIGVYPALTDEMLDYVTASIKEFVAARG 434
AscC   IDLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKYRVVGELTNTDRIMNQTFWIGIYPGLTHDHLDYVVSKFEEFFGLNF 437
          *  *       *          *                                 *  *               *
```

Fig. 3 Alignment of the amino acid sequences of KasR and other C-3 dehydratases.

Protein names are indicated to the left of the amino acid sequences. *Streptomyces cyanogenus* LanQ (Protein ID: AAD13547.1), *Streptomyces purpurascens* RdmI (Protein ID: AAL24451.1), *Streptomyces fradiae* UrdQ (Protein ID: AAF72550.1), *Yersinia pseudotuberculosis* AscC (Protein ID: AAA88701.1). Asterisks mark the conserved residues.

step of ascarylose biosynthesis is well-characterized [24]. The His220 residue would act as an active-site base, abstracting the C-4′ proton from the PMP-substrate complex [25]. This histidine residue is conserved in KasR and other homologs (Fig. 3), suggesting that *kasR* encodes an enzyme responsible for the C-3 deoxygenation step of KSM biosynthesis.

kasA

We determined the nucleotide sequence of the 5′ region (420 bp) of *kasA*. The start codon ATG of *kasA* overlaps the stop codon TGA of *kasR* (ATGA), suggesting translational coupling of these two genes. We previously reported the 3′ region (756 bp) of *kasA* [3]. The deduced KasA protein (391 amino acids, M_r of 42,745, pI of 5.68) shared similarity with a glycosyl transferase (Protein ID: AAK46529.1) from *Mycobacterium tuberculosis* (36% identity), *N*-acetylglucosaminyltransferase (Protein ID: AAB99629.1) from *Methanocaldococcus jannaschii* (34% identity) and sucrose-phosphate synthase (Protein ID: CAC87818.1) from *Nostoc punctiforme* (30% identity). This result suggests that *kasA* encodes an enzyme responsible for the transglycosylation step of KSM biosynthesis.

Transcriptional Analysis of the *kasN*, *kasO*, *kasP*, *kasQ*, *kasR* and *kasA* Genes

Northern blot analysis was performed using an 811 bp *Sph*I-*Sac*I fragment located in the *kasN* structural gene as a probe (Fig. 1). The *kasN* region transcript is approximately 2,000 nucleotides mRNA (Fig. 4A) which was far larger than *kasN* alone (1,152 bp). Analysis of this transcript by RT-PCR suggested that *kasN* and *kasO* were transcribed into a bicistronic mRNA (*NO*-mRNA) (Fig. 4B). This conclusion is consistent with the lack of a consensus promoter sequence in the 5′-untranslated region of *kasO* and with the presence of a putative transcriptional terminator in the 3′-region of *kasO*. The 5′-end of *NO*-mRNA was determined by the RACE method. The transcription start site was located at G, 226 nucleotides upstream of the putative start codon of *kasN* (Fig. 4C). The −35 (ATCCCG) and −10 (TATTAT) hexamer sequences of the putative *kasN* promoter (*kasN*-p) were similar to that of a streptomycete *E. coli*-like promoter (SEP) [26].

Transcriptional analysis of the *kasP*, *kasQ*, *kasR* and *kasA* genes by RT-PCR also suggested that *kasP*-*kasQ* and *kasR*-*kasA* were co-transcribed as bicistronic transcripts (*PQ*-mRNA and *RA*-mRNA, respectively) (Fig. 4B).

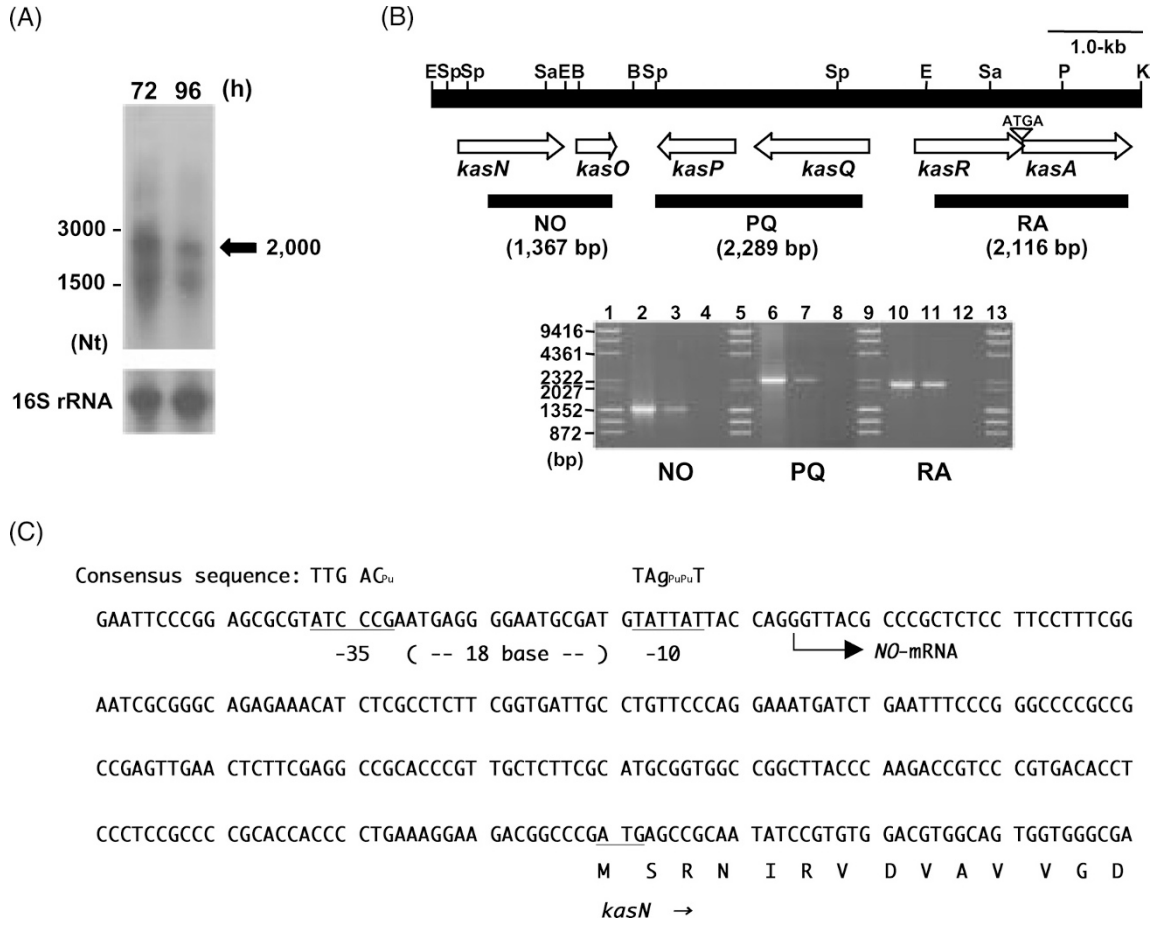

Fig. 4 Transcriptional analysis of *kasN*, *kasO*, *kasP*, *kasQ*, *kasR* and *kasA*.

(A) Northern blot analysis of *kasN* in *S. kasugaensis* M338-M1. RNA was prepared from strain M338-M1 cells grown in KSM-producing medium (MR) at 27°C for 72 or 96 hours. The 16S rRNA panel indicates the result using 16S rDNA fragment (500 bp) from strain M338-M1 as probe. This panel served as internal control. (B) Transcriptional analysis by RT-PCR. Lanes 1, 5, 9, 13: Molecular size markers. Lanes 2, 6, 10: Genomic PCR products (Positive controls). Lanes 3, 7, 11: RT-PCR products. Lanes 4, 8, 12: Negative controls (reactions contain no reverse transcriptase). (C) The *kasN* upstream sequence including the putative promoter. The location of the 5′-end of *NO*-mRNA is indicated by a bent arrow.

Trx-KasT binds to the *kasN* Upstream Region and to the *kasQ-kasR* Intergenic Region

We previously reported that KasT showed a high degree of similarity (50% identity) with StrR, a pathway-specific activator protein of the streptomycin (SM) biosynthetic gene cluster, and that purified Trx-KasT bound to the *kasU-kasJ* intergenic region, rather than to the *kasT* upstream region [4]. Therefore, we propose that KasT is a pathway-specific regulator of the KSM biosynthetic gene cluster. StrR binds to some regions in the SM biosynthetic gene cluster and activates transcription of the *str*/*sts* genes [27]. To demonstrate the DNA-binding ability of Trx-KasT for other regions, we performed gel retardation assay using the *kasN* upstream and the *kasQ-kasR* intergenic DNA fragments. The *kasQ-kasR* intergenic DNA was separated into three parts (19A, 19B and 19C) and used as individual probes. As shown in Fig. 5, Trx-KasT bound to the *kasN* upstream region (−362∼+42, designated 15) and the *kasQ-kasR* intergenic region (19A, 19B and 19C).

Discussion

In this study, we have identified a number of potential genes in the KSM biosynthetic pathway by sequencing the 6.8-kb region downstream of the KSM transporter genes. Altogether, we have cloned and sequenced 22.4-kb of genomic DNA from the KSM region of *S. kasugaensis* M338-M1 (Fig. 1). In this segment, we have identified twenty ORFs, including a KSM acetyltransferase gene (*kac*^{338}), a plausible pathway-specific regulatory gene (*kasT*), three KSM transporter genes (*kasKLM*) and three

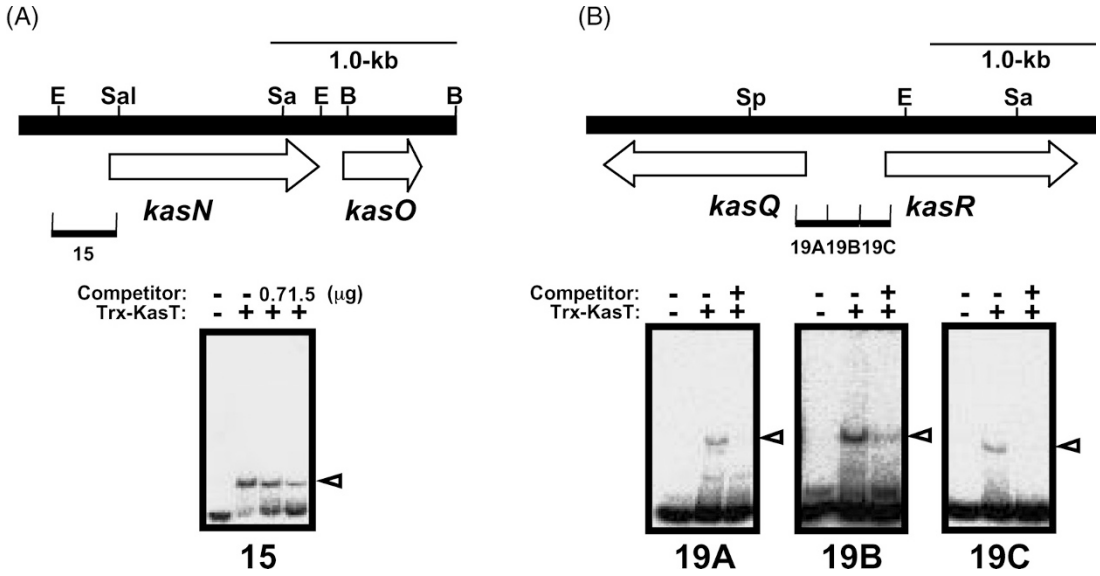

Fig. 5 Gel retardation assays demonstrating the specific binding of Trx-KasT.

Open triangles indicate a shifted band. (A) Gel retardation assay for the upstream region (404 bp, designated 15) of *kasN*. Lane 1, free probe; Lane 2, Trx-KasT; Lane 3, Trx-KasT and competitor (700 ng); Lane 4, Trx-KasT and competitor (1.5 μg). (B) Gel retardation assay for the intergenic region (19A, 19B and 19C) of *kasQ-kasR*. Lane 1, free probes; Lane 2, Trx-KasT; Lane 3, Trx-KasT and competitors (19A=400 ng, 19B=400 ng and 19C=800 ng).

putative KSM biosynthetic genes (*kasJ*, *kasC*, *kasD*).

KSM is constructed with a kasugamine (2,4-diamino-2,3,4,6-tetradeoxy-D-arabinohexopyranose), a carboxyformidoyl group and a D-*chiro*-inositol. Fukagawa *et al.* reported that UDP-*N*-acetylglucosamine (UDP-GlcNAc), glycine and *myo*-inositol are probable precursors of kasugamine, the carboxyformidoyl group and D-*chiro*-inositol, respectively [28～31]. We have proposed a KSM biosynthetic pathway that includes a C-2 epimerization step and C-3 deoxygenation (Fig. 6). KasQ, encoding a homolog of UDP-GlcNAc 2-epimerase, would catalyze the C-2 epimerization step.

C-3 deoxygenation in the ascarylose biosynthesis is catalyzed by the combined action of AscC and AscD [24, 32, 33]. The first half of this transformation is dehydration catalyzed by AscC. The second half of the reaction is an NADH-dependent reduction initiated by AscD, which contains an FAD and an iron-sulfur cluster in its active site [34]. In this study, we found *kasR* encoding an AscC-like protein. Therefore, the C-3 deoxygenation reaction in the biosynthesis of the kasugamine moiety likely occurs by a similar mechanism (Fig. 6). However, no *ascD*-like gene is found in the KSM biosynthetic gene cluster. The lack of the *ascD*-like gene within the C-3 deoxysugar biosynthetic gene cluster was observed in the *gra* cluster of *Streptomyces violaceoruber* Tu22 [35]. Since general electron transfer proteins, such as diaphorase or the reductase component of methane monooxygenase can work with AscC [36], other proteins encoded by a gene located inside or outside the gene cluster may serve as the reductase. In the KSM biosynthetic gene cluster, *kasP* encodes a reductase. Therefore, we speculate that KasP may catalyze the reduction step of C-3 deoxygenation in the KSM biosynthetic pathway (Fig. 6).

Two carbons and a nitrogen of the carboxyformidoyl group [–C(NH)COOH] are derived from the two carbons and the nitrogen of the same glycine molecule [29, 30]. FI-FTM A synthase, which showed similarity to KasN, converts FTM A to FI-FTM A by addition of an N-formimidoyl group (FI, –CH=NH). The FI moiety is derived from glycine *via* oxidation of the amino acid in the presence of FTM A and oxygen [13]. Therefore, the carboxyformidoylation in the KSM biosynthetic pathway is probably catalyzed by KasN and is proposed to occur by a similar mechanism (Fig. 6). The *kasN* gene forms an operon with *kasO*. Because the *kasO* product showed no significant similarity with known enzymes, it may play a novel role in carboxyformidoylation or some other step.

We found almost all of the genes responsible for KSM biosynthesis in the 22.4-kb DNA segment (Fig. 1) except for a gene encoding an enzyme for the deacetylation step (Fig. 6-VII). It is likely, then, that this gene cluster constitutes most or all of the KSM biosynthetic gene cluster, though further analyses, such as gene disruption and heterologous expression, would be required for confirmation.

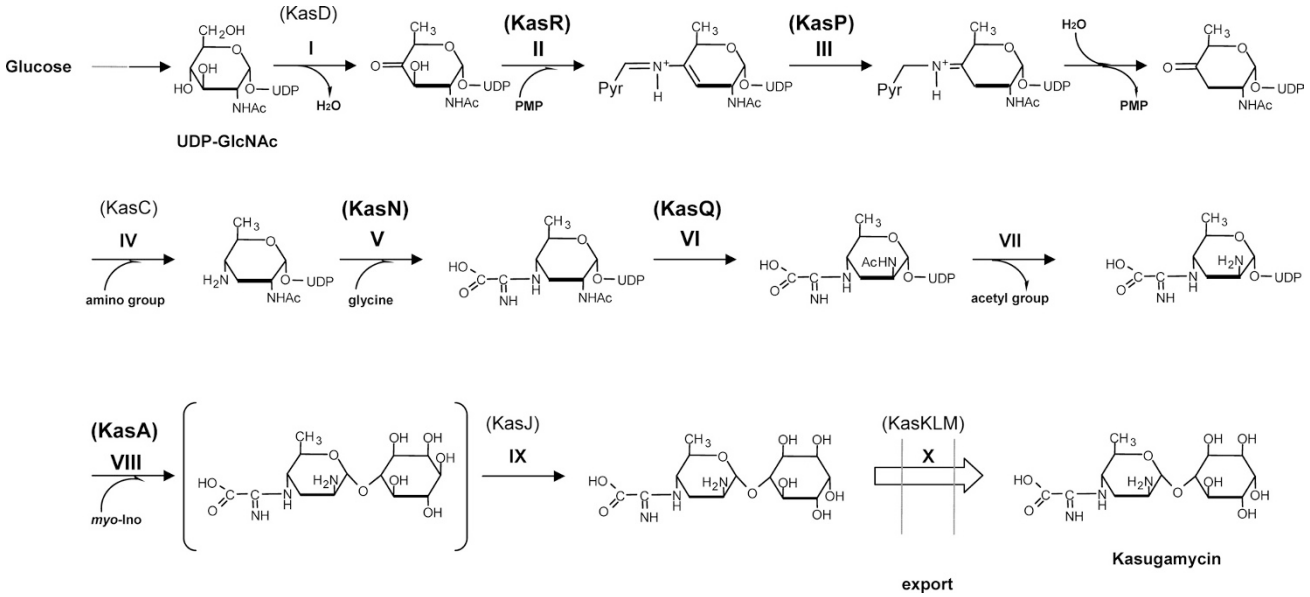

Fig. 6 A tentative pathway of kasugamycin biosynthesis.

Abbreviations: UDP-GlcNAc, UDP-*N*-acetylglucosamine; PMP, Pyridoxamine 5′-phosphate; *myo*-Ino, *myo*-inositol.

Transcriptional analyses of *kasN*, *kasO*, *kasP*, *kasQ*, *kasR* and *kasA* suggested that these genes were transcribed as three bicistronic mRNAs (*NO*-mRNA, *PQ*-mRNA and *RA*-mRNA). The putative pathway-specific regulator KasT bound to the *kasN* upstream region. Although we were not able to determine the transcription start sites of *kasPQ* and *kasRA*, we assumed that the *kasQ-kasR* intergenic region containes two promoters and two KasT-binding sites for these operons. To demonstrate the location of KasT-binding sites, *kasQ-kasR* intergenic region was divided in three fragments and these fragments were used as probes in the gel retardation assay. However, KasT bound to all three fragments of the intergenic region and all of these bindings were inhibited by competitors in the assay. The DNA-binding ability of KasT for three *kasQ-kasR* intergenic fragments suggests that two or three binding sites are located in this region. It is likely that all three of these operons are under the transcriptional control of KasT. In our previous paper [4], we reported that KasT bound to the *kasU-kasJ* intergenic region containing an imperfect inverted repeat. However, there are not any similar inverted repeats in the *kasN* upstream region and the *kasQ-kasR* intergenic region. We are trying to determine the KasT-binding sequences by DNaseI footprinting experiment.

Recently, a pleiotropic mutant was isolated from *Streptomyces kasugaensis* MB273-C4 [37]. The mutant retains an aerial mycelium- and KSM-minus phenotype. Mutation of the RNA polymerase subunit omega gene (*rpoZ*) gives rise to the pleiotropic phenotype, and may affect the expression of other genes or gene clusters. Understanding the mechanisms of deficiency of KSM production in this mutant could provide some genetic information about KSM biosynthesis. We are presently investigating the transcriptional pattern of *kas* genes in this mutant.

Acknowledgements We thank Mr. Yasuhiro Yamane and Mr. Yushi Abe for excellent technical assistance.

References

1. Umezawa H, Hamada M, Suhara Y, Hashimoto T, Ikekawa T. Kasugamycin, a new antibiotic. Antimicrob Agents Chemother 5: 753–757 (1965)
2. Ikeno S, Higashide K, Kinoshita N, Hamada M, Hori M. Correlation between the presence of *kac*, kasugamycin acetyltransferase gene, and the productivity of kasugamycin in *Strepromyces*. Actinomycetologica 10: 73–79 (1996)
3. Ikeno S, Tsuji T, Higashide K, Kinoshita N, Hamada M, Hori M. A 7.6 kb DNA region from *Streptomyces kasugaensis* M338-M1 includes some genes responsible for kasugamycin biosynthesis. J Antibiot 51: 341–352 (1998)
4. Ikeno S, Aoki D, Sato K, Hamada M, Hori M, Tsuchiya KS. *kasT* gene of *Streptomyces kasugaensis* M338-M1 encodes a DNA-binding protein which binds to intergenic region of *kasU-kasJ* in the kasugamycin biosynthesis gene cluster. J Antibiot 55: 1053–1062 (2002)
5. Ikeno S, Yamane Y, Ohishi Y, Kinoshita N, Hamada M,

Tsuchiya KS, Hori M. ABC transporter genes, *kasKLM*, responsible for self-resistance of a kasugamycin producer strain. J Antibiot 53: 373–384 (2000)
6. Ishikawa J, Hotta K. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. FEMS Microbiol Lett 174: 251–253 (1999)
7. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448 (1988)
8. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402 (1997)
9. Maruyama IN, Rakow TL, Maruyama I. cRACE: a simple method for identification of the 5′ end of mRNAs. Nucleic Acids Res 23: 3796–3797 (1995)
10. Wright F, Bibb MJ. Codon usage in the G+C-rich Streptomyces genome. Gene 113: 55–65 (1992)
11. Stover CK, Pham XQT, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FSL, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim RM, Smith KA, Spencer DH, Wong GKS, Wu Z, Paulsen IT. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature 406: 959–964 (2000)
12. Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hirama C, Nakamura Y, Ogasawara N, Kuhara S, Horikoshi K. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. Nucleic Acids Res 28: 4317–4331 (2000)
13. Dairi T, Yamaguchi K, Hasegawa M. N-formimidoyl fortimicin A synthase, a unique oxidase involved in fortimicin A biosynthesis: purification, characterization and gene cloning. Mol Gen Genet 236: 49–59 (1992)
14. Mao Y, Varoglu M, Sherman DH. Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL 2564. Chem Biol 6: 251–263 (1999)
15. Yu TW, Muller R, Muller M, Zhang X, Draeger G, Kim CG, Leistner E, Floss HG. Mutational analysis and reconstituted expression of the biosynthetic genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin biosynthesis in *Amycolatopsis mediterranei* S699. J Biol Chem 276: 12546–12555 (2001)
16. Kiino DR, Rothman-Denes LB. Genetic analysis of bacteriophage N4 adsorption. J Bacteriol 171: 4595–4602 (1989)
17. Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MTG, Prentice MB, Sebaihia M, James KD, Churcher C, Mungall KL, Baker S, Basham D, Bentley SD, Brooks K, Cerdeno-Tarraga AM, Chillingworth T, Cronin A, Davies RM, Davis P, Dougan G, Feltwell T, Hamlin N, Holroyd S, Jagels K, Leather S, Karlyshev AV, Moule S, Oyston PCF, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413: 523–527 (2001)
18. Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, Chandler M, Choisne N, Claudel-Renard C, Cunnac S, Demange N, Gaspin C, Lavie M, Moisan A, Robert C, Saurin W, Schiex T, Siguier P, Thebault P, Whalen M, Wincker P, Levy M, Weissenbach J, Boucher CA. Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature 415: 497–502 (2002)
19. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MTG, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature 413: 848–852 (2001)
20. Campbell RE, Mosimann SC, Tanner ME, Strynadka NCJ. The structure of UDP-*N*-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferase. Biochemistry 39: 14993–15001 (2000)
21. Westrich L, Domann S, Faust B, Bedford D, Hopwood DA, Bechthold A. Cloning and characterization of a gene cluster from *Streptomyces cyanogenus* S136 probably involved in landomycin biosynthesis. FEMS Microbiol Lett 170: 381–387 (1999)
22. Niemi J, Mantsala P. Nucleotide sequences and expression of genes from *Streptomyces purpurascens* that cause the production of new anthracyclines in *Streptomyces galilaeus*. J Bacteriol 177: 2942–2945 (1995)
23. Hoffmeister D, Ichinose K, Domann S, Faust B, Trefzer A, Drager G, Kirschning A, Fischer C, Kunzel E, Bearden DW, Rohr J, Bechthold A. The NDP-sugar co-substrate concentration and the enzyme expression level influence the substrate specificity of glycosyltransferases: cloning and characterization of deoxysugar biosynthetic genes of the urdamycin biosynthetic gene cluster. Chem Biol 7: 821–831 (2000)
24. Weigel TM, Liu LD, Liu HW. Mechanistic studies of the biosynthesis of 3,6-dideoxyhexoses in *Yersinia pseudotuberculosis*: purification and characterization of CDP-4-keto-6-deoxy-D-glucose 3-dehydrase. Biochemistry 31: 2129–2139 (1992)
25. Lei Y, Ploux O, Liu HW. Mechanistic studies on CDP-6-deoxy-L-*threo*-D-*glycero*-4-hexulose 3-dehydrase: Identification of His-220 as the active-site base by chemical modification and site-directed mutagenesis. Biochemistry 34: 4643–4654 (1995)

26. Hopwood DA, Bibb MJ, Chater KF, Janssen GR, Malpartida F, Smith P. *In* Regulation of gene expression. 25 Years On Symp Soc Gen Microbiol. *Ed.*, Booth IR, Higgins CF. pp. 251–276, Cambridge, University of Cambridge Press, (1986)
27. Beyer S, Distler J, Piepersberg W. The *str* gene cluster for the biosynthesis of 5′-hydroxystreptomycin in *Streptomyces glaucescens* GLA.0 (ETH 22794): new operons and evidence for pathway-specific regulation by StrR. Mol Gen Genet 250: 775–784 (1996)
28. Fukagawa Y, Sawa T, Homma I, Takeuchi T, Umezawa H. Studies on biosynthesis of kasugamycin. IV. Biosynthesis of the kasugamine moiety from [1-^{14}C]-glucosamine and [1, 2 or 6-^{14}C]-glucose. J Antibiot 21: 358–360 (1968)
29. Fukagawa Y, Sawa T, Takeuchi T, Umezawa H. Biosynthesis of kasugamycin. II. Biosynthesis of the two-carbon-side chain of kasugamycin. J Antibiot 21: 182–184 (1968)
30) Fukagawa Y, Sawa T, Homma I, Takeuchi T, Umezawa H. Studies on biosynthesis of kasugamycin. V. Biosynthesis of the amidine group. J Antibiot 21: 410–412 (1968)
31. Fukagawa Y, Sawa T, Takeuchi T, Umezawa H. Studies on biosynthesis of kasugamycin. III. Biosynthesis of the D-inositol moiety. J Antibiot 21: 185–188 (1968)
32. Lo SF, Miller VP, Lei Y, Thorson JS, Liu HW, Schottel JL. CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase from *Yersinia pseudotuberculosis*: enzyme purification and characterization of the cloned gene. J Bacteriol 176: 460–468 (1994)
33. Chen XM, Ploux O, Liu HW. Biosynthesis of 3,6-dideoxyhexoses: *in vivo* and *in vitro* evidence for protein-protein interaction between CDP-6-deoxy-L-*threo*-D-*glycero*-4-hexulose 3-dehydrase (E_1) and its reductase (E_3). Biochemistry 35: 16412–16420 (1996)
34. Miller VP, Liu HW. Biosynthesis of 3,6-dideoxyhexoses: new evidence supporting a radical mechanism for C-3 deoxygenation. J Am Chem Soc 114: 1880–1881 (1992)
35. Tornus D, Floss HG. Identification of four genes from the granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tu22 involved in the biosynthesis of L-rhodinose. J Antibiot 54: 91–101 (2001)
36. Miller VP, Thorson JS, Ploux O, Lo SF, Liu HW. Cofactor Characterization and mechanistic studies of CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase: exploration into a novel enzymatic C-O bond cleavage event. Biochemistry 32: 11934–11942 (1993)
37. Kojima I, Kasuga K, Kobayashi M, Fukasawa A, Mizuno S, Arisawa A, Akagawa H. The *rpoZ* gene, encoding the RNA polymerase omega subunit, is required for antibiotic production and morphological differentiation in *Streptomyces kasugaensis*. J Bacteriol 184: 6417–6423 (2002)