Complete sequence and organization of the 
*Serratia marcescens* biotin operon

Naoki Sakurai,† Hiroyuki Akatsuka, Eri Kawai, Yuji Imai and Saburo Komatsubara†

Author for correspondence: Naoki Sakurai. Tel: +81 48 433 2737. Fax: +81 48 433 2734. 
e-mail: nsakurai@tanabe.co.jp

Research Laboratory of 
Applied Biochemistry, 
Tanabe Seiyaku Co. Ltd, 
16-89 Kashima-3-chome, 
Yodogawa-ku, Osaka 532, 
Japan

The nucleotide sequence of the biotin (*bio*) operon of wild-type *Serratia marcescens* Sr41 was determined. Five ORFs were identified to encode BioA (7,8-diaminopelargonic acid aminotransferase), BioB (biotin synthase), BioF (7-keto-8-aminopelargonic acid synthase), BioC (an enzyme catalysing the synthesis of pimeloyl-CoA) and BioD (dethiobiotin synthase), in this order. The operon was deduced to be transcribed divergently to the left into *bioA* and to the right into the *bioBFCD* genes. The promoters and a common predicted operator for both *bioA* and *bioBFCD* genes were located between the *bioA* and *bioB* genes. The predicted amino acid sequences of these enzymes were similar to the sequences of the corresponding enzymes of *Escherichia coli*. Analysis of expression of the *lacZ* structural gene fused with the *bioA* and *bioB* promoters revealed that the biotin operon was subject to biotin-mediated feedback repression.

**Keywords**: biotin biosynthesis, biotin operon, *Serratia marcescens*

---

**INTRODUCTION**

In micro-organisms, d-biotin is synthesized from pimeloyl-CoA through four enzymic steps (Fig. 1) (Eisenberg, 1987; Bachmann, 1990). The precursors of pimeloyl-CoA are still unknown although some micro-organisms have been reported to convert pimelic acid to pimeloyl-CoA (Izumi et al., 1974; Ploux et al., 1992; Sanyal et al., 1994b; Ifuku et al., 1994). In *Escherichia coli*, the biosynthetic pathway from pimeloyl-CoA to d-biotin includes 7-keto-8-aminopelargonic acid (KAPA), 7,8-diaminopelargonic acid (DAPA) and dethiobiotin, and the five biotin biosynthetic enzymes are encoded by the *bioABFCD* genes, which together form an operon, the biotin (*bio*) operon (Fig. 1) (Eisenberg, 1987; Bachmann, 1990). In addition, the *bioH* gene, unlabeled to this *bio* cluster, is involved in an unknown and early step before pimeloyl-CoA synthesis. The final step in d-biotin biosynthesis remains un-elucidated because of the unusual nature of the transformation reaction and an unknown source of the sulfur atom (Demoll & Shieve, 1983; Parry, 1983; Ifuku et al., 1992).

In *E. coli*, transcription of the *bio* operon is repressed by biotin, which is converted by the *birA* product into a co-repressor, biotinyl 5'-adenylate (Barker & Campbell, 1981; Prakash & Eisenberg, 1979). The *bioA* gene is transcribed leftward and the *bioBFCD* genes are transcribed rightward by the action of the promoters located between the *bioA* and *bioB* genes and controlled by a common operator (Guha, 1971).

*Serratia marcescens* is an enteric bacterium. Classical mutagenesis and molecular genetic techniques have been employed to construct various L-amino acid-hyper-producing strains of this bacterium, which have been used commercially for L-amino acid production (Komatsubara, 1994). Recently, d-biotin-producing strains of *S. marcescens* Sr41 have been constructed by selection of mutants resistant to acidomycin, a biotin analogue (Sakurai et al., 1993a). Among these strains, SB304 and SB412 produced d-biotin at 5 and 20 mg L⁻¹, respectively, in a fermentation medium containing sucrose and urea. Subsequently, we have reported the cloning of the biotin biosynthesis gene cluster of SB304 and SB412 on a 7.2 kb *EcoR*I-*Hind*III fragment and the recombinant strains constructed from the above two mutants showed a marked d-biotin...
production (200 mg l⁻¹) (Sakurai et al., 1993b). Complementation of the E. coli biotin auxotrophs with deletion plasmids and analysis of Tn1000 insertion derivatives from these plasmids have revealed that the S. marcescens bio genes are clustered in the order bio,ABFCD, and constitute an operon (Sakurai et al., 1993b).

There have been few genetic and enzymic findings on the regulation of d-biotin biosynthesis in S. marcescens. This paper deals with the gene structure and expression of the wild-type bio operon of S. marcescens.

### METHODS

#### Strains, plasmids and media.

Bacterial strains and plasmids used are shown in Table 1. Nutrient medium containing 0.5% (w/v) glucose, 10% (w/v) peptone, 0.3% (w/v) meat extract, 1.0% (w/v) yeast extract and 0.5% (w/v) NaCl was used for routine colony isolation and growth study of S. marcescens strains. Luria–Bertani (LB) medium was used as a rich medium for E. coli strains. The minimal medium of Davis & Mingioli (1950) was modified by omitting sodium citrate and increasing the glucose concentration to 0.5% (w/v). Antibiotics were added at the following concentrations for E. coli: kanamycin (Km), 100 µg ml⁻¹; ampicillin (Ap), 100 µg ml⁻¹; streptomycin (Sm), 50 µg ml⁻¹. For S. marcescens, Km and Ap were added at concentrations of 200 and 500 µg ml⁻¹, respectively.

#### Primer-extension analysis.

The 5' ends of the bioA and the bioB transcripts were mapped by primer-extension analysis with primer PbioA1 (5'-GGGTTGCGATATGCGG-3') complementary to a region from position nt 2577 to 2594 of the mRNA transcript for bioA, and primer PbioB1 (5'-ATCAAACAGGGCCCTGGTCTT-3') complementary to a region from position nt 2759 to 2739 of the mRNA transcript for bioB. To facilitate the detection of extended products, pUC19-derived high-copy-number plasmids were used. Blunt-ended 1·1 kb ScaI–PvuII fragments containing the bioA and bioB promoters and the truncated bioA and bioB structural genes excised from pLGW101. This fragment was ligated with the 2·4 kb PvuII fragment of pUC19 of which the tac promoter region was removed, producing plasmid pABPI01. Total S. marcescens RNA was prepared from mid-exponential-phase cells of 8000(pABPI01) grown in minimal medium with ampicillin.

RNA was purified using the hot-phenol extraction procedure (Hinnebusch & Fink, 1983) and 20 µg RNA were mixed with 20 pmol 5' end α-³²P-labelled primer. The mixture was denatured at 80 °C in 50 mM PIPES/KOH (pH 6.4), 0·4 M NaCl containing 1 mM EDTA and 50% formamide for 15 min, and chilled on ice. Primers were hybridized with the RNA by incubating at 20 °C for 16 h. After ethanol precipitation, DNA–RNA hybrids were extended at 42 °C for 60 min in 50 mM Tris/HCl (pH 8·3), 100 mM KCl, 10 mM MgCl₂, 1 mM DTT, 1 µM each of dGTP, dATP, dCTP and dTTP, 0·5 units human placental ribonuclease inhibitor (Promega) and 22 units Rous-associated virus (RAV-2) reverse transcriptase (Takara Shuzo). The products were separated on an 8% polyacrylamide gel. A dideoxy sequencing ladder was run in parallel for size comparison.

### Subcloning and construction of lacZ fusions.

For monitoring the expression of bioA and bioB, lacZ transcriptional/ translational fusions were constructed. Since amplification of the regulatory site of the genes in a high-copy-number plasmid might cause abnormal expression because of the absence of its repressor, we used pLGLacZ339, a low-copy-number promoter analysis vector, consisting of the pSC101 replicon and the promoterless lacZ structural gene (Omoti et al., 1991).

The 0·59 kb DNA fragments containing the bio promotor/operator region and 5' ends of both bioA and bioB structural genes (at 2278–2834) were amplified by PCR from pLGW101. The primers used for this PCR were 5'-TTGGATCC(BamHI)-CACATCT(BglII)-ACGGCCACCGAACCAGAGTCG-3' and 5'-TATAGATCT(BgIII)-CGGATCC(BamHI)-GCTGACGCGGGTCGAAGTC-3', both of which involve BglII and BamHI restriction sites in their 5' ends (underlined sequences) to facilitate directional cloning. To construct the bioA–lacZ fusion plasmids, the 0·59 kb PCR products were digested with BglII and ligated into the BamHI site of pLGLacZ339 to render the direction matching the bioA transcription in-frame. The resultant lacZ fusion, pZPA101, was fused with the lacZ gene of pLGLacZ339 in-frame to the 5' regions of the bioA gene.

To construct the bioB–lacZ fusion plasmids, the above 0·59 kb PCR products were digested with BamHI and ligated into the BamHI site to render the direction matching the bioB transcription in-frame. The resultant lacZ fusion, pZPB101, was fused with the lacZ structural gene of pLGLacZ339 in-frame to the 5' region of the bioB gene. The direction of the insert was
confirmed by restriction analysis and their junctions were verified by nucleotide sequencing.

Monitoring β-galactosidase activities in cells containing the bio-lacZ fusion plasmids. The bio-lacZ translational fusion plasmids described above were introduced into S. marcescens TT392, a wild-type strain for biotin biosynthesis, using electroporation (Sakurai et al., 1996). Transformants were grown at 30 °C to the mid-exponential phase in 3 ml minimal medium containing kanamycin in the presence or absence of 1 μM p-biotin. Cells were collected by centrifugation at 12000 g for 10 min at 4 °C. Cell pellets were resuspended in 20 mM Tris/HCl (pH 8.0) buffer and disrupted by sonication on an ice bath. The resultant cell lysates were used for the assay of β-galactosidase activities, using ONPG as substrate (Miller, 1972). One unit β-galactosidase activity formed 1 nmol products min⁻¹ (mg protein)⁻¹.

Sequence analysis. The DNA sequence of the plasmid containing the wild-type S. marcescens bio operon was determined by the dideoxy chain-termination method (Sanger et al., 1977) with templates prepared from DNA subcloned in pHSG298 and pHSG299 using universal primers. Ordered deletion subclones were obtained by unidirectional digestion (Yanisch-Perron et al., 1985) using the Kilo-Sequence Deletion Kit (Takara Shuzo). Synthetic primers were used to complete the sequence with the same templates as the double-stranded template.

Computer analysis of the sequence was done with the GENEWORK software (Software Development) and CLUSTAL W (Higgins et al., 1988). For homology studies on amino acid sequences, the international protein and nucleotide databases (GenBank, EMBL and DDBJ) were searched on-line using the TFASTA program on the host computer of DDBJ (Pearson & Lipman, 1988).

In vitro transcription/translation analysis. In vitro transcription/translation reactions were carried out with S-30 extracts prepared from S. marcescens 8000 according to DeVries & Zubay (1967). Reaction mixture contained 0.4 μg purified pLGW101, 10 μg S-30 extract and 1000 μCi (37 MBq) L-[³⁵S]methionine in a total volume of 5 μl. The resulting labelled products were separated by SDS-PAGE (12.5% ; Laemmli, 1970) and visualized by autoradiography.

<table>
<thead>
<tr>
<th>Table 1. Strains and plasmids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain/plasmid</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
</tr>
<tr>
<td>MG1063</td>
</tr>
<tr>
<td>3104ST</td>
</tr>
<tr>
<td>3105ST</td>
</tr>
<tr>
<td>3106ST</td>
</tr>
<tr>
<td>3107ST</td>
</tr>
<tr>
<td>3108ST</td>
</tr>
<tr>
<td><strong>S. marcescens</strong></td>
</tr>
<tr>
<td>8000</td>
</tr>
<tr>
<td>TT392</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
</tr>
<tr>
<td>pBR322</td>
</tr>
<tr>
<td>pUC19</td>
</tr>
<tr>
<td>plLG339</td>
</tr>
<tr>
<td>pHSG298</td>
</tr>
<tr>
<td>pHSG299</td>
</tr>
<tr>
<td>pLGlacZ339</td>
</tr>
<tr>
<td>pLGW101</td>
</tr>
<tr>
<td>pLGW101dA3</td>
</tr>
<tr>
<td>pLGW101dB7</td>
</tr>
<tr>
<td>pLGW101dF26</td>
</tr>
<tr>
<td>pLGW101dC26</td>
</tr>
<tr>
<td>pLGW101d1D8</td>
</tr>
<tr>
<td>pABP101</td>
</tr>
<tr>
<td>pZPA101</td>
</tr>
<tr>
<td>pZPB101</td>
</tr>
</tbody>
</table>

*Symbols used for relevant genotypes and phenotypes are as follows: Nuc⁰, no production of extracellular nuclease; r⁻, defect of host restriction enzymes; Ap⁰, resistance to ampicillin; Km', resistance to kanamycin; Sm', resistance to streptomycin; Tc', resistance to tetracycline; Ap', enhanced sensitivity to ampicillin; Km', enhanced sensitivity to kanamycin.
Disruption of the bio genes by Tn1000 insertion. Plasmid pLGW101, a low-copy-number plasmid carrying the entire bio operon, was used in this experiment to avoid segregational and structural instability. Each of the bio genes on pLGW101 was disrupted by transposon Tn1000 insertion as described by Sakurai et al. (1993b). Strain MG1063(pLGW101) was mated with E. coli 3105ST, 3105ST, 3105ST, 3105ST or 3108ST. Km', Sm' and Te' transconjugants were selected on plates and tested for each of the Bio+ phenotypes. The plasmids were extracted from the transconjugants and the sites of Tn1000 insertions were determined by restriction analysis.

Determination of the N-terminal amino acid sequence. Cells of TT392(pZPA101) and TT392(pZPB101) were harvested from exponentially growing cultures in minimal medium. Cell extracts were purified as described above. LacZ fusion proteins were purified using ProtoSorb lacZ Immunoaffinity Adsorbent kit (Promega) according to the manufacturer's instructions. Purified proteins were subjected to SDS-PAGE and transferred electrophoretically to Immobilon-P membranes (Millipore) in 25 mM Tris/HCl (pH 8.3)/1920 mM glycine/10 mM EDTA. The transferred proteins corresponding to BioA- and BioB-LacZ fusions were excised from the membrane and analysed on a gas-phase sequenator (Hewlett Packard model HP G1005A).

Other genetic procedures and recombinant DNA techniques. Standard procedures were used for plasmid preparation, restriction enzyme digestion, ligation and agarose gel electrophoresis (Maniatis et al., 1982). Cells of E. coli and S. marcescens were transformed using the method of Maniatis et al. (1982) and by electroporation as described by Sakurai & Komatsubara (1996), respectively.

Protein determinations. Protein concentration was determined using a Bio-Rad protein assay kit and BSA as the standard (Bradford, 1976).

RESULTS
Sequencing of the S. marcescens bio operon
We have reported that the S. marcescens bio genes form an operonic structure (Sakurai et al., 1993b). The 7-2 kb EcoRI–HindIII fragment of pLGW101 contains the entire wild-type bio operon coding region (nt 1353–6331). Analysis of this nucleotide sequence has revealed one ORF and four tandem ORFs (GenBank/EMBL/DDBJ accession no. D17468). The deduced molecular mass of the five gene products are 46986 (425 amino acid), 38569 (382 aa), 27776 (255 aa) and 25615 Da (227 aa), respectively. The proximity and common orientation of the products of the S. marcescens bio genes have suggested that these genes constitute a typical bacterial operon. Subsequently, we constructed deletion derivatives of the 7-2 kb DNA fragment and examined them for complementation test using E. coli bio auxotrophs, confirming that the above five ORFs correspond to the bioA, -B, -F, -C and -D genes, respectively (data not shown). Although the bioA transcription is independent of the bioBFCD transcription, both bioA and bioB transcriptions are coordinately repressed by a common operator sequence as shown below. Hence, we defined bioA as a member of the S. marcescens bio operon as in E. coli. The nucleotide sequence confirmed that the gene encoding the first ORF corresponding to that of the E. coli BioA (DAPA aminotransferase) is transcribed in the opposite direction to the other four genes. This structure is very similar to that of the E. coli bio operon (Otsuka et al., 1988). The remaining four genes encoding BioB (biotin synthase), BioF (KAPA synthase), BioC (an enzyme responsible for pimeloyl-CoA formation) and BioD (dethiobiotin synthase) are coordinately transcribed as a polycistron from a single promoter located between the bioA and bioB genes. The first ORF corresponding to BioA extends from an ATG initiation codon at position 2630 to an ATT termination codon at position 1353. The second ORF corresponding to BioB starts from an ATG codon at nt 2716 and is terminated by a TGA stop codon at nt 3756. The initiation site of the third ORF (BioF, KAPA synthase) is ATG of nt 3756 and the termination is TGA at nt 4904. The fourth ORF corresponding to BioC starts from ATG at nt 4888 and to TAA at nt 5655. The last, encoding BioD (dethiobiotin synthase), extends from a GTG (Val) codon at nt 5648 and is terminated by TGA codon at nt 6331.

The above five ORFs are preceded by sequences with homology to the 3' end of 16S rRNA and presumably represent RBSs (Ringoquist et al., 1992). At the 3' ends of the bioA and bioD genes, regions of dyad symmetries with calculated AD (Gibbs free energy) values of −18.3 kcal mol⁻¹ (−76.9 kJ mol⁻¹) and −29.8 kcal mol⁻¹ (−125.2 kJ mol⁻¹) are formed between positions at nt 1341–1313 (5'-TCTGAGCTCGATAGGGGCGA-3') and at nt 6351–6380 (5'-CAACCCGGGTTATCAACCAGGTT-3'). The later sequence is followed by four thymidine residues and resembles a ρ-independent transcription termination signal (Rosenberg & Court, 1984).

Comparison of S. marcescens bio operon products with other proteins
The deduced amino acid sequences of the bio operons were used for searching the GenBank, EMBL and DDBJ databases using the TFASTA program. The S. marcescens BioA is very similar to that of E. coli (71.5% identity in 424 aa overlap) and Bacillus sphaericus BioA (35.8% identity in 436 aa overlap). BioA has a high sequence similarity (up to 20% identity) to the extensive family of eukaryotic ornithine aminotransferase.

The S. marcescens BioB shows significant similarity with the E. coli BioB (86.5% identity in 346 aa overlap), Saccharomyces cerevisiae biotin synthase (49.3% identity in 294 aa overlap), Brevibacterium flavum biotin synthase (34.8% identity in 299 aa overlap) and B. sphaericus (33.6% identity in 304 aa overlap). The BioB protein has a local homology with the E. coli lipA product (16.0% identity in 238 aa overlap), which catalyses the sulfur incorporation to octanoic acid involved in lipoic acid synthesis. BioB (biotin synthase) is involved in the insertion of sulfur into dethiobiotin. The sequence similarity between the lipA and bioB products suggests that a common sulfur donor and/or a common reaction mechanism are involved in both biotin and lipoic acid biosynthesis.

3298
The scan of the amino acid sequences of *S. marcescens* BioF (KAPA synthase) against the protein databases has revealed that *S. marcescens* BioF has a similarity with the *E. coli* (64.8% identity in 383 aa overlap) and *B. sphaericzis* (30.5% identity in 383 aa overlap) BioF proteins.

Significant similarity of *S. marcescens* BioC, involved in the formation of pimeloyl-CoA, was observed with the *bioC* product of *E. coli* (52.2% identity in 248 aa overlap).

BioD catalyses the formation of ureide ring by trans-carboxylation of DAPA into dethiobiotin. Computer searches did not find proteins having significant homology with *S. marcescens* BioD except *E. coli* and *B. sphaericzis* BioDs (742 and 35.1% identities in 213 and 191 aa overlaps, respectively).

**Structure and transcription initiation sites of the bio promoters**

A computer search for the *bioA* promoter sequence has revealed two conserved sequences: TCGCCA (nt 2692–2687) and TATACT (nt 2670–2665), corresponding to −35 and −10 (Pribnow box) regions, respectively. For the *bioB* promoter, TTGACA (nt 2660–2665) and TAAACT (nt 2682–2687), were found in the complementary sequence, respectively (Fig. 2). It is possible that these sequences, which are similar to the consensus sequence of *E. coli* promoters, constitute the divergently transcribed promoters of the *S. marcescens* bio operon. The *bioA* and *bioB* genes are apparently controlled by a single regulatory region: a putative 32 bp operator (nt 2634–2665), common to both *bioA* and *bioB* promoters, containing an imperfect palindromic sequence that partially overlaps the *bioB* promoter (Fig. 2).

To confirm the predicted promoter/operator structure, primer-extension analysis was used to locate the transcription initiation sites of the *bioA* and *bioB* (or *bioBFCD*) genes. The promoter regions of the *bioA* and *bioB* genes were subcloned into a pUC-derivative vector, which produces the 5' region of the *bioA* and *bioB* mRNAs but does not affect cell growth because it lacks most of the structural genes. Synthetic oligonucleotides complementary to the 5' region of the *bioA* or *bioB* transcripts were synthesized as primers. The primer-extension analysis using the RNA from *S. marcescens* 8000 containing pABP101 carrying the wild-type *bio* promoter has revealed that the transcript for the *bioB* gene (right transcription) starts primarily 22 bases upstream from the ATG codon or seven bases downstream from the Pribnow box, in agreement with the consensus start point defined by the above predicted promoter (Fig. 3). An additional band was observed at a position located at six bases upstream (G, nt 2710) from the ATG start codon. We could not obtain a clear extended product of the *bioA* gene because of the possible instability.

**Analysis of the regulation of bio expression using bio–lacZ transcriptional–translational fusions**

Transcriptional regulation of the *bio* operon was examined. No quantitative assay method has been reported for biotin synthase (the *bioB* gene product). Thus, to normalize the promoter activity of both directions, we constructed *lacZ* fusion plasmids of the *bioA* and *bioB* genes, pZPA101 and pZPB101, respectively, and analysed the activities of β-galactosidase (LacZ) in the wild-type strain carrying these plasmids. LacZ activities of both *S. marcescens* 8000(pZPA101) and 8000(pZPB101) were strongly repressed by the addition of 1 µM D-biotin to the
Fig. 3. Mapping of the 5' end of the bioB mRNA. Primer extension was performed with RNAs isolated from S. marcescens 8000(pABP101), containing the bio promoter/operator region and the truncated bioA and bioB genes. Primer-extension products and a sequencing ladder generated with the same primer used in the extension mapping were run in parallel on denaturing 5% polyacrylamide gels. Asterisks indicate the positions of the major and minor extension products.

culture medium (Table 2). Note that these lacZ fusions contain the authentic Shine–Dalgarno sequences and the ATG start sites for the S. marcescens bio genes.

Identification of the bio gene products

To determine the sizes of the bio operon products encoded by pLGW101, carrying the 7.2 kb DNA fragment encoding the entire bio operon, we used an in vitro transcription/translation experiment using [³⁵S]methionine. Polypeptides specific to the bio genes were expressed from plasmid pLGW101. Molecular masses calculated by these migrations were approximately 42, 41, 39, 28 and 25 kDa (Fig. 4). Products of 41, 39, 28 and 25 kDa correspond to BioF, BioB, BioC and BioD on the

Table 2. Specific β-galactosidase activities of bio-lacZ fusions

Cells were grown in minimal medium in the presence or absence of 1 μM D-biotin. pZPA101 and pZPB101 are the bioA- and bioB-lacZ translational fusion plasmids, respectively. One unit β-galactosidase activity formed 1 nmol products min⁻¹ (mg protein)⁻¹. Values represent the means of two independent assays.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Promoter</th>
<th>Specific activity (units)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-D-Biotin</td>
</tr>
<tr>
<td>TT392</td>
<td>No plasmid</td>
<td>&lt; 10</td>
</tr>
<tr>
<td>TT392(pLGlacZ339)</td>
<td>Vector plasmid</td>
<td>46</td>
</tr>
<tr>
<td>TT392(pZPA101)</td>
<td>bioA</td>
<td>190</td>
</tr>
<tr>
<td>TT392(pZPB101)</td>
<td>bioB</td>
<td>610</td>
</tr>
</tbody>
</table>
basis of the molecular masses deduced from amino acid sequences. The size of a 42 kDa protein, which is close to BioA, but is somewhat larger than that of the expected molecular mass of BioA (425 aa, 47 kDa).

To locate the bio operon gene products, proteins encoded by pLGW101 and its derivatives mutagenized with transposon Tn1000 were labelled with [35S]methionine, and the synthesized products were analysed by SDS-PAGE, followed by fluorography (Fig. 4b).

pLGW101dA3 (BioA−) produced no 42 kDa protein, whereas other proteins seen in pLGW101 were produced, suggesting that the 42 kDa protein is actually the bioA gene product (Fig. 4b, lane 3). Migrations of other products except for BioA were not affected by bioA disruption, indicating that the bioA gene transcription might be independent of other bio genes. When pLGW101df26, a plasmid in which the Tn1000 insertion inactivated the bioF gene, was used as a template, the 41 kDa protein was not observed, suggesting that this protein is the bioF gene product (Fig. 4b, lane 5). A smaller band of 27 kDa might be the truncated BioF. pLGW101db7 (BioB+) did not produce the band of 39 kDa, but yielded a smaller truncated polypeptide (Fig. 4b, lane 4). In addition, bands of 41, 28 and 25 kDa which correspond to BioF, BioD and BioC, respectively, faded because of the polar effect of Tn1000 insertion.

Tn1000 insertion in pLGW101dc26 (BioC−) yielded the possible BioC truncated form of approximately 27 kDa (Fig. 4b, lane 6). pLGW101d18, in which Tn1000 disrupted bioD allele, did not produce the 28 kDa polypeptide (Fig. 4b, lane 7). The molecular mass of BioD (25 kDa) calculated from the nucleotide sequence is expected to be smaller than that of BioC (28 kDa). However, the band of BioC in this analysis was apparent to be smaller than that of BioD.

In pLGW101db7 (BioB−), pLGW101df26 (BioF+) and pLGW101dc26 (BioC+), the band corresponding to each disrupted gene disappeared with the Tn1000 insertion. In addition, bands of the peptides (BioF, BioC and BioD in pLGW101db7; BioC and BioD in pLGW101df26; BioD in pLGW101dc26) also disappeared because of a polar effect caused from Tn1000 insertions. This result confirmed that the bioB, bioF, bioC and bioD genes might constitute a single cistron in this order.

N-terminal amino acid sequences of BioA and BioB were attempted to be determined. BioA–LacZ and BioB–LacZ fusions were purified by LacZ–immunoaffinity chromatography from TT392 cells carrying pZPA101 and pZPB101, respectively. Elutions containing purified LacZ fusions from the immunoaffinity column contained one major protein band with a molecular mass of about 115 kDa corresponding to that of LacZ (data not shown). N-terminal sequencing of the first five residues of the BioB–LacZ fusion showed that it begins A-D-R-I-H, not in agreement with the predicted start sequence (M-M-A-D-R-I-H). We were unable to determine the N-terminal sequence of native BioA because of apparent N-terminal blockage or a small quantity of the protein obtained.

**DISCUSSION**

The bioABFCD genes constitute an operon in *S. marcescens* as in *E. coli*. Nucleotide sequence analysis has revealed that the structure of this operon in *S. marcescens* is similar to that in *E. coli*. Comparison of the bioABFCD gene products of *S. marcescens* with those of *E. coli* demonstrated 71.5, 86.5, 64.8, 52.2 and 74.2% similarity, respectively. *S. marcescens* differs from *E. coli* in the gene structure downstream of the bioA region. Otsuka *et al.* (1988) have reported that in *E. coli* the bioA gene has no terminator sequence and a short ORF is located after the bioA coding region whose function has been unknown but possibly this ORF and the bioA gene constitutes an operon. Nevertheless, in *S. marcescens*, we have found that no ORF follows the bioA coding region and no significant homology with the ORF gene of *E. coli* in the corresponding nucleotide region of *S. marcescens*. This finding indicates that the unknown ORF of *E. coli* might not take part in the biotin biosynthesis.

*S. marcescens* BioB shows a low similarity to LipA as does *E. coli* BioB (Hayden *et al.*, 1992). LipA catalyses the sulfur introduction into octanoic acid to produce liposate. The catalytic function of LipA is apparently similar to that of BioB which catalyses the insertion of sulfur into dethiobiotin. We have then aligned the sequence of *E. coli* LipA with those of biotin synthases of five different micro-organisms (Fig. 5). We found that there are three highly homologous regions between the five biotin synthases and *E. coli* LipA: namely, motif I [C-x-x-x-C-x-x-x-C, C-P-E-D-C-x-(Y, F)-C] for biotin synthase which contains three cysteine residues; motif II [(Y, F)-N-H-N-(L, I)]; and motif III [S-G-x-(I, M)-(I, V)-G-(L, M)-(G, K)-E-(S, T)]. Since the *E. coli* LipA has regions less homologous with the entire region of biotin synthase except the above three regions, the three motifs were predicted to be of great significance for the sulfur-introduction reaction. Moreover, several hydrophobic amino acid residues are highly conserved between the five BioBs and LipA proteins (boxed in Fig. 5), suggesting that hydrophobicity might be needed for the catalytic function of BioB.

Cysteine residues generally provide ligands for iron atoms in iron–sulfur proteins via a motif as C-x-x-C-x-x-C in bacterial ferredoxins (Loverynberg, 1973). Clusters containing three cysteine residues are known to be iron-binding motifs and are frequently found in oxidoreductases. We have found that the culture of an *S. marcescens* d-biotin-producing strain in the presence of excess amounts of FeCl₂ resulted in the significant increase of the d-biotin production (Masuda *et al.*, 1995). In addition, d-biotin is apparently an oxidized form of dethiobiotin since the oxidation number of d-biotin is greater than that of dethiobiotin. Hence, it is assumed that these three cysteine residues might bind to an iron atom to catalyse an oxidative introduction of sulfur atom into the inactive C=H bonds of dethiobiotin. Recently, the *E. coli* BioB protein has been reported to have two irons and two sulfurs within the molecule: BioB is a [2S–2Fe] protein (Sanyal *et al.*, 1994a).
Fig. 5. Multiple amino acid sequence alignment for S. marcescens BioB (SM; Otsuka et al., 1988). E. coli BioB (EC; B. sphaericus BioB (BS; Osawa et al., 1989), Brev. flavum BioB (BF; Hatakeyama et al., 1992), Sacc. cerevisiae BioB (SC; Zhang et al., 1994) and E. coli LipA (LP; Hayden et al., 1992). Identical residues are indicated by dots. Boxed regions indicate highly conserved motives between BioB and LipA. F, I, M, L, V, Y and W are defined as hydrophobic amino acids.

The promoter/operator region of the S. marcescens bio operon is very similar to those of other enteric bacteria: Citrobacter freundii, Salmonella typhimurium and E. coli (Shiuan & Campbell, 1988). These operons are divergently transcribed leftward to the bioA gene and rightward to the bioBFCD genes. Primer-extension analysis showed multi-extended products for the bioB gene. We do not know whether this additional band represents the real start site or whether it is a simple artefact due to a premature stop during reverse transcription. It should be noted that no possible promoter sequences corresponding to these additional start sites are found by analogy with E. coli and other bacteria. Interestingly, the additional start site for bioB gene was observed in the region four bases downstream of the characteristic T-rich cluster (TTTT-CGTTTT, nt 2696-2706) (Fig. 2).

We tentatively identified an ORF of 1275 bp as a coding region for the bioA gene followed by a potential RBS (AA, nt 2626 to 2627) located 4 bp upstream of the ATG start codon (nt 2480). However, the migration of the expected BioA (42 kDa) on SDS-PAGE analysis was less than that of the predicted ORF sequence (47 kDa). Although estimates of polypeptide size from SDS-PAGE are often inaccurate, the RBS proposed for the bioA gene is not highly complementary to published sequences (Ringquist et al., 1992). Interestingly, a sequence having consensus to a bacterial RBS (GGTG, nt 2487-2490) followed by ATG sequence with a 6 bp space is located 151 bp downstream in which there is a potential ORF starting from nt 2480 to 1356, encoding of 376 aa. If BioA starts from this downstream ATG codon, the calculated molecular mass is 41 378 Da, which is close to that found by SDS-PAGE. Unfortunately, we could not determine the N-terminal amino acid sequence of BioA in the present study. Work is in progress to identify the real transcriptional features and the polypeptide of the bioA gene.

ACKNOWLEDGEMENTS

We thank Drs I. Chibata and T. Tosa for encouragement. We also thank Dr K. Omori for helpful discussion.

REFERENCES


Received 30 January 1996; revised 25 June 1996; accepted 17 July 1996.