Identification of the midecamycin biosynthetic gene cluster in *Streptomyces mycarofaciens* UC189B (ATCC 21454) and analysis of the enzymes for dTDP-D-mycaminose biosynthesis

A dissertation submitted to the
Fachbereich Naturwissenschaften II (Chemie/Biologie)
Bergische Universität – Gesamthochschule Wuppertal
for the degree of Doctor of Natural Science

presented by
Lina Cong

Wuppertal
February 2000

To all my teachers

To my parents and John

ACKNOWLEDGEMENTS

This study was carried out during the period between July, 1996 to February, 2000, am Lehrstuhl für Chemische Mikrobiologie, Fachbereich Chemie, der Bergischen Universität GH Wuppertal. I wish to express my sincere gratitude to:

My supervisor, Professor Piepersberg, for introducing me to this interesting project and for his invaluable support and guidance throughout the years. I especially appreciate his help for constructing the phylogenetic trees. His scientific ideas and vision will benefit me in my future career.

Professor Vogel for his review of my thesis.

Dr. Wehmeier for his intensive help in solving the experimental problems and fruitful discussion during these years. Special thanks for his criticism of part of my thesis.

The "Hyglide" project co-operators, Dr. Raynal, Professor Leadlay, Professor Salas, Professor Elling and all colleagues from their laboratories, for fruitful collaboration and mutual experimental support.

Dr. Beyer for his support and help during the initial phase of my work.

Christoph for his skilful help to manipulate the HPLC analysis system.

Petra and Dr. Blüggel for their help and co-operation in performing the LC-MS analysis.

Elisabeth and Yvonne for their assistance with the sequencing work.

All my former and current colleagues for their help and for creating a pleasant and friendly atmosphere in the laboratory during these years.

Professor Mattiasson (University of Lund, Sweden) for providing me the opportunity to work in his group for a year. The most important lesson I learnt from him is to understand how hard-working and active-thinking a scientist should be.

Professor Mäntsälä (University of Turku, Finland) for introducing me into the field of molecular biology.

Last but not least, my parents, and my boy friend, John, for their love, support and patience.

Contents

Abbreviations vi

Abstract viii

Zusammenfassung ix

1. **Introduction** 1

1.1. Secondary metabolites and macrolide antibiotics 1

1.1.1 Roles of secondary metabolites in *Streptomyces* 1

1.1.2. Character, structure and classification of macrolides 1

1.2. Cloning of erythromycin and tylosin biosynthesis genes 4

1.3. Biosynthesis of macrolide lactones 8

1.3.1 Basis of fatty acid and polyketide biosynthesis 8

1.3.2. Synthesis of macrolide lactone via type I PKSs 10

1.4. Biosynthesis of 6-deoxysugars 11

1.4.1. 6-Deoxysugar genes found in antibiotic biosynthesis 11

1.4.2. Genetics of 6-deoxysugars in erythromycin and tylosin biosynthesis 12

1.4.3. Mechanism elucidation of some deoxysugar biosynthetic enzymes 13

1.5. Aims of the present study 16

2. Materials 17

2.1. Chemicals and enzymes 17

2.2. Media and Buffers 18

2.2.1 Media for cultivation of *E. coli* 18

2.2.2. Media for cultivation of *Streptomyces* 19

2.2.3. Buffers 21

2.3 Bacteria, cosmids and plasmids 27

2.3.1 Bacteria 27

2.3.2 Vectors 28

2.3.3 Recombinant plasmids 29

2.4	Oligonucleotides	37
3.	**Methods**	39
3.1.	Cultivation and maintenance of bacterial strains	39
3.1.1.	Cultivation and maintenance of *E. coli*	39
3.1.2.	Cultivation and maintenance of *Streptomyces*	39
3.2.	*S. mycarofaciens* fermentation and product extraction	40
3.3.	General manipulation of DNA	40
3.3.1.	Isolation of genomic DNA from *Streptomyces*	40
3.3.2.	Preparation of plasmid DNA from *E. coli*	41
3.3.3.	Preparation of plasmid DNA from *Streptomyces* sp.	41
3.3.4.	*In vitro* manipulation of DNA	42
3.3.5.	Polymerase chain reactions (PCR)	43
3.3.6.	Separation of DNA fragment	44
3.3.7.	Radioactive labelling of DNA fragments	44
3.3.8.	DNA-DNA hybridization	45
3.3.9.	DNA sequencing	45
3.3.10.	Determination of DNA concentration	45
3.3.11.	Transformation of *E. coli* and *S. lividans*	46
3.3.12.	Heterologous gene expression in *E. coli* BL21(DE3)pLysS	46
3.3.13.	Heterologous gene expression in *S. lividans* 1326	46
3.4.	Manipulation of the cosmid genomic library of *S. mycarofaciens*	47
3.4.1.	Construction of the cosmid genomic library	47
3.4.1.1.	Partial digestion of *S. mycarofaciens* genomic DNA	47
3.4.1.2.	Size fractionation of DNA fragments	48
3.4.1.3.	Digestion of the cosmid vector pKU206	49
3.4.1.4.	Ligation of insert DNA to vector DNA	49
3.4.1.5.	Packaging of ligated DNA	50
3.4.1.6.	Adsorption of packaged recombinant cosmids to *E. coli*	50
3.4.1.7.	Plating of packaged recombinant cosmids	50
3.4.2.	Plating out, amlification and maintenance of the cosmid genomic library	51
3.4.2.1.	Plating on agar for the library amplification	51

3.4.2.2.	Picking and streaking colonies for screening and maintenance of the library	51
3.4.3.	Screening of the cosmid genomic library	52
3.4.3.1.	Preparing duplicate membranes of cosmid colonies	52
3.4.3.2.	Cosmid DNA binding to the membranes	52
3.4.3.3.	Screening the cosmid membranes	52
3.5.	Biochemical methods	53
3.5.1.	Cell-free extracts of *E. coli* and *Streptomyces*	53
3.5.2.	Determination of protein concentration	53
3.5.3.	SDS polyacrylamide gel electrophoresis (PAGE)	53
3.5.4.	Western blotting and immuno-detection of proteins	54
3.5.5.	Analysis of the activity of dTDP-D-glucose 4,6-dehydratase by RmlB	54
3.5.6.	Analysis of the activity of enzymes participating in dTDP-D-mycaminose biosynthesis via a coupling reaction of RmlB, MidH and MidC	55
3.6.	Chromatography methods	55
3.6.1.	Thin-layer chromatography (TLC)	55
3.6.2.	High pressure liguid chromatography (HPLC)	56
3.6.3.	HPLC coupled to mass spectrometry (LC-MS)	56
3.7.	Computer programs	56

4. Results 57

4.1.	Detection of pathway-specific genes from *S. mycarofaciens* by PCR	57
4.1.1.	Identification and isolation of a gene, *midB*, encoding a dTDP-glucose 4, 6- dehydratase from *S. mycarofaciens*	58
4.1.2.	Identification and isolation of a gene, *midI*, encoding a glycosyltransferase from *S. mycarofaciens*	58
4.2.	Construction and screening of a genomic library from *S. mycarofaciens*	60
4.2.1.	Construction of the genomic cosmid library	60
4.2.1.1.	Selection of 20-35 kb DNA fragments in *S. mycarofaciens* partial digestion	60
4.2.1.2.	Ligation, packaging and plating the library	63
4.2.2.	Screening of the cosmid genomic library probed by the *midB* and *midI* DNA fragments	64
4.3.	Restriction analysis and sequencing of partial regions of the recombinant cosmids	65

4.4.	Chromosomal walking by using a 2.3 kb DNA fragment close to one end of Smyc-LC1 as a probe to screen the *S. mycarofaciens* genomic library	68
4.5.	Open reading frames found in the gene sub-cluster analysed from Smyc-LC1	68
4.6.	Open reading frames found in the gene sub-cluster analysed from Smyc-LC3	71
4.7.	The PKS genes (*midG*)	71
4.8.	Regulation genes (*midF* and *midR*)	72
4.9.	Resistance genes (*mdmA* and *mirB*)	73
4.10.	Identification of *midE*, *mdmB* and *mdmC*	74
4.11.	The *midC* gene	74
4.12.	The *midA* and *midB* genes	76
4.13.	The *midK* gene	77
4.14.	The *midH* gene	79
4.15.	The *midI* gene	79
4.16.	Identification of genes encoding dTDP-D-mycaminose biosynthetic pathway	80
4.17.	About strains of *S. mycarofaciens* UC189B (ATCC 21454) and *S. mycarofaciens* ATCC 21454	82
4.18.	Heterologous complementation of *midK* and *midI* into *Sac. erythraea* mutants	82
4.19.	Heterologous expression of the MidH, MidC, MidK and MidI proteins	83
4.19.1.	Overexpression of the MidH protein	85
4.19.2.	Overexpression of the MidC protein	86
4.19.3.	Expression and detection of the MidK protein	87
4.19.4.	Expression and detection of the MidI protein	88
4.20.	Assay for activity of a coupling reaction of the RlmB, MidH and MidC protein	89
4.20.1.	Activity assay of dTDP-D-glucose 4,6-dehydratase by RmlB	89
4.20.2.	Reaction products from combined catalysis by RmlB, MidH and MidC	90

5. Discussion 92

5.1.	The *S. mycarofaciens* cosmid genomic library	92
5.2.	Architecture of the *mid* gene cluster in *S. mycarofaciens*	93

5.3. Hybrid antibiotic production by heterologous complementation 98

5.4. Heterologous enzyme expression and combined enzymatic analysis 99

5.5. Phylogenetic analysis of some enzymes involved in midecamycin biosynthesis 101

6. References 109

7. Appendices 122

7.1 The nucleotide and encoded amino acid sequences of the *midM* region of Smyc-LC1 122

7.2. The nucleotide and encoded amino acid sequences of the *midCA* region of Smyc-LC1 123

7.3. The nucleotide and encoded amino acid sequences of the *midDBL* region of Smyc-LC1 125

7.4. The nucleotide and encoded amino acid sequences of the *midE* region of Smyc-LC1 128

7.5. The nucleotide and encoded amino acid sequences of the *midGKHIJ* region of Smyc-LC3 129

7.6. Sequence alignment of the MidD protein with other thioesterases for construction of MidD-tree in Fig. 5.2 133

7.7. Sequence alignment of the MidA protein with other dTDP-glucose synthases for construction of MidA-tree in Fig. 5.2 134

7.8. Sequence alignment of the MidB protein with other dTDP-glucose 4,6-dehydratases for construction of MidB-tree in Fig. 5.2 135

7.9. Sequence alignment of the MidH protein with other putative isomerases for construction of MidH-tree in Fig. 5.2 137

7.10. Sequence alignment of the MidC protein with other aminotransferases for construction of MidC-tree in Fig. 5.2 138

7.11. Sequence alignment of the MidK protein with other methyltransferases for construction of MidK-tree in Fig. 5.2 139

7.12. Sequence alignment of the MidI protein with other glycosyltransferases for construction of MidI-tree in Fig. 5.2 141

Abbreviations

A	adenine
aa	amino acid
acc. no.	accession number
ALF	automatic laser fluorescens DNA sequencer
APS	ammonium persulfate
ATCC	American Type Culture Collection
bp	base pair
BSA	bovine serum albumin
C	cytosine
ca	*circa*
Ci	curie
d	day
dATP	2´-deoxyadenosine triphosphate
dCTP	2´-deoxycytidine triphosphate
dGTP	2´-deoxyguanosine triphosphate
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
6DOH	6-deoxyhexose
dTDP	2´-deoxythymine diphosphate
DTT	dithiothreitol
dTTP	2´-deoxythymidine triphosphate
E. coli	*Escherichia coli*
EDTA	ethylendiaminotetraacetic acid
ery genes	erythromycin biosynthetic genes
G	guanine
g	gram
g	gravity
h	hour
HPLC	high performance liquid chromatography
IPTG	isopropyl-1-thio-β-D-galactoside
kb	kilobase
kDa	kilodalton
λ	wave length
nt	nucleotide
OD_x	optical density at x nm and 1 cm depth
orf	open reading frame
ori	origin
mid genes	midecamycin biosynthetic genes
min	minute
Mr	relative molecule weight
MW	molecule weight
No.	number
nt	nucleotide
p	promoter
PAGE	polyacrylamide gel electrophoresis

PAA	polyacrylamide
PCR	polymerase chain reaction
PEG	polyethylenglycol
psi	'pounds per square inch'
PVDF	polyvinyl difluoride
RBS	ribosome binding site
PLP	pyridoxal phosphate
PMP	pyridoxamine phosphate
Rnase	ribonuclease
rpm	rounds per minute
RT	room temperature
Sac	*Saccharoployspora*
sec	second
S.	*Streptomyces*
SDS	sodium dodecyl sulphate
ssp.	subspecies
T	thymine
Tab.	Table
TEMED	N, N, N´, N´-tetramethylethylendiamine
TES	N-tris(hydroxymethyl)methyl-2-aminoethanesulphonic acid
TLC	thin-layer chromotography
Tris	trishydroxymethylaminomethane
tyl genes	tylosin biosynthetic genes
U	enzyme unit
Uni.	University
UV/VIS	ultraviolet/visible light
V	volt
X-gal	5-bromo-4-chloro-3-indolyl-β-D- galactoside

ABSTRACT

In this study the following aspects have been worked out by studying the genetics of midecamycin biosynthesis as a model system to elucidate the biosynthesis of sugar components in 16-membered macrolides and to acquire tools for the production of new hybrid macrolide antibiotics.

1. By sequentially screening a genomic library in the cosmid vector pKU206 via two homologous gene probes which were detected by PCR, the midecamycin biosynthetic (*mid*) gene cluster has been identified in the isolated ca. 74 kb DNA in the genome of *S. mycarofaciens*, falling into two regions.

2. About 10 kb DNA from the cosmid Smyc-LC1 and ca. 7 kb DNA from the cosmid Smyc-LC3 were sequenced contiguously. The remaining regions of these two cosmids were partially sequenced to get further information on extension and informational contents of the cluster. Among them, 10 genes with complete reading frames and 14 genes incomplete were identified, all of which are necessary for midecamycin biosynthesis. The region of genes responsible for mycarose biosynthesis was found in Smyc-LC1. A complete set of genes for the biosynthesis and transfer of mycaminose was identified. The organisation of the overall *mid* gene cluster revealed that, similar to other macrolide gene clusters, the PKS genes are flanked by two regions containing genes encoding enzymes for sugar biosynthesis, with others for resistance, regulatory, and sugar or lactone modification.

3. The genes, *midC*, *midH*, *midK* and *midI*, were characterised by heterologous expression of these enzymes. MidC and MidH were over-produced as soluble proteins in *E. coli* both in native form and as His-tag fusion. Soluble proteins His-tag-MidK and His-tag-MidI were detected only by Western blotting in low quantities.

4. To characterise the postulated function of MidH (3,4-isomerase) and MidC (3-aminotransferase), the products of a coupling enzymatic reaction to convert dTDP-D-glucose by RmlB (4,6-dehydratase), MidH and MidC were analysed by HPLC and LC-MS. It has been confirmed that the MidC protein is responsible for transamination to form dTDP-amino-6-deoxy-D-glucose.

ZUSAMMENFASSUNG

In dieser Arbeit wurden die folgenden Aspekte der Genetik und Biosynthese des Makrolids Midecamycin ausgearbeitet, um die Mechanismen der Makrolidzuckersynthese zu klären und Handwerkszeuge für die Herstellung hybrider Makrolide zu liefern.

1. Durch sequentielles Screening einer genomischen DNA-Bank im Cosmidvektor pKU206 durch zwei homologe PCR-amplifizierte Genproben wurden zwei Regionen von insgesamt ca. 75 kb DNA isoliert, die die gewünschten Anteile des Midecamycin (*mid*) Genclusters enthielten.

2. Ca. 10 kb DNA aus Cosmid Smyc-LC1 und ca. 7 kb DNA aus Smyc-LC3 wurden komplett sequenziert. Die übrigen Regionen wurden durch partielle Sequenzierung charakterisiert, um weitere Informationen über Struktur und Genverteilung zu erhalten. Unter den Genen wurden 10 komplette und 14 unvollständige Leserahmen identifiziert, die für die Midecamycin-Biosynthese notwendig sind. Die Region der Mycarose-Biosynthesegene wurde auf Smyc-LC1 gefunden. Das komplette Set der Gene für Biosynthese und Transfer der Mycaminose wurde kloniert. Die Organisation des *mid* Genclusters ist ähnlich wie in andern Makroliden zu beiden Seiten der zentralen PKS Gene mit zwei gemischten Gruppen von Zucker-, Acylierungs-, Resistenz- und Regulatorgenen strukturiert.

3. Die Genen, *midC*, *midH*, *midK* und *midI*, wurden durch heterologe Expression charakterisiert. MidC und MidH wurden als lösliche Proteine in *E. coli* überproduziert, jeweils in nativer und His-tag fusionierter Primärstruktur. Die ebenfalls löslichen His-tag Proteine MidK und MidI wurden nur in geringer Menge gebildet und über Westernblot nachgewiesen.

4. Um die postulierten Funktionen der MidH (3,4-Isomerase) und MidC (3-Aminotransferase) nachzuweisen, wurden die Produkte eines gekoppelten Enzymtests zur Umsetzung von dTDP-D-Glucose mit RmlB (4,6-Dehydratase), MidH und MidC mittles HPLC und LC-MS Techniken ausgewertet. Es wurde bestätigt, daß das MidC Protein verantwortlich ist für die Transaminierungsreaktion bei der Bildung des dTDP-Amino-6-Deoxy-D-Glucose.

1. Introduction

1.1. Secondary metabolites and macrolide antibiotics

1.1.1. Roles of secondary metabolites in *Streptomyces*

Secondary metabolites are synthesized via multistep pathways leading from precursors (usually intermediates of primary metabolism) to the specific moieties of these metabolites (Drew *et al*., 1978). The majority of bioactive products (for instance, antibiotics, pigments, alkaloids, enzyme inhibitors, and so on) of microorganisms, is generated by secondary metabolism. This part of the metabolic machinery of microorganisms may play no essential role in the vegetative development of the producing organisms, but seems to convey advantages to the pertinent species with regard to its long-term survival in the biological community and environment (Vining, 1992).

Actinomycetes, to which the streptomycetes belong, are well known for the synthesis of a broad range of biologically active secondary metabolites with antibacterial, anti-viral, anti-tumor and immuno-suppressive activities. Tab. 1.1 lists a few of these compounds with their respective activities. Understanding roles of secondary metabolites in *Streptomyces* has been increased by cloning numerous genes involved in the biosynthesis of antibiotics, pigments and other secondary metabolites (Beppu, 1996; Minas, 1997). A comprehensive picture of their organization in clusters is emerging. Knowledge about expression of these genes, which are not essential for growth, is of utmost importance for establishing how they have evolved as compared to operons of primary biosynthetic genes. Furthermore, understanding of the regulatory mechanisms controlling gene expression has relevance for industrial over-production of these metabolites.

1.1.2. Character, structure and classification of macrolides

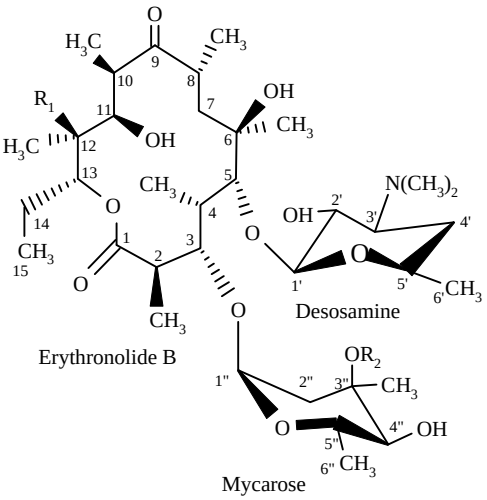
Macrolides are a class of antibiotics that contain a macrocyclic lactone ring (Omura, 1984) composed of a polyketide-derived backbone to which one, two or three sugars are

Tab. 1.1. Some medically important *Streptomyces* secondary metabolites and their mode of action

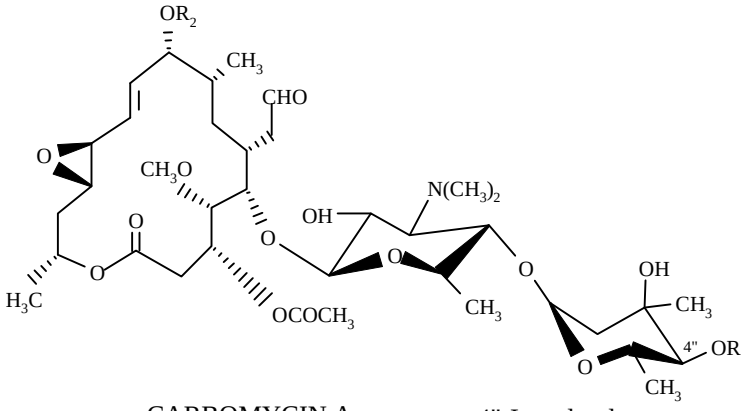
Type	Antibiotic	Producer	Activity	Target
β-Lactams (peptides)	Cephalosporin Cephamycin	*S. clavuligerus*	Gram +/- bacteria	D-Ala metabolism in cell wall formation
Chorismic acid	Chloramphenicol	*S. venevuelae* Chemical Synthesis	Bacteria, Mycoplasma	Ribosome 50S subunit
Aromatic polyketides	Mithramycin	*S. argillaceus*	Antitumor	DNA intercalating or binding agents
	Tetracyclines	*S. rimosus*	Gram +/- bacteria	Ribosome 30S subunit
Macroildes	Erythromycin	*Sac. erythraea*	Bacteria	Ribosome 50s subunit
	Rapamycin, FK506	*S. hygroscopicus*	Immunosuppressants	T-cell differentiation

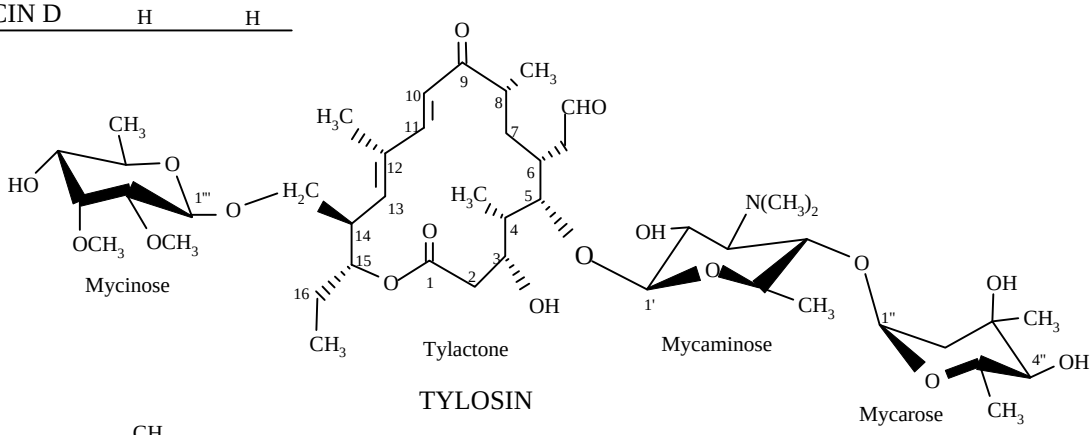
commonly attached. They are produced as secondary metabolites by mycelium-forming soil bacteria from the order *Actinomycetales*; the majority are from members of the genera *Streptomyces*, *Micromonospora* and *Saccharopolyspora* (Nakagawa *et al*., 1984). Macrolide antibiotics are widely used as anti-infective, immuno-supressive, insecticidal, and parasiticidal agents in the clinic or for agricultural purposes. The known mechanism of the biological function of the main group of classical macrolides (erythromycin, tylosin, etc.) is to bind to the peptidyltransferase center of the 50S subunit of the bacterial ribosome, thereby inhibiting bacterial protein synthesis (Gale *et al*., 1981). Other macrocyclic lactones, like polyenes and avermectin, have different modes of action.

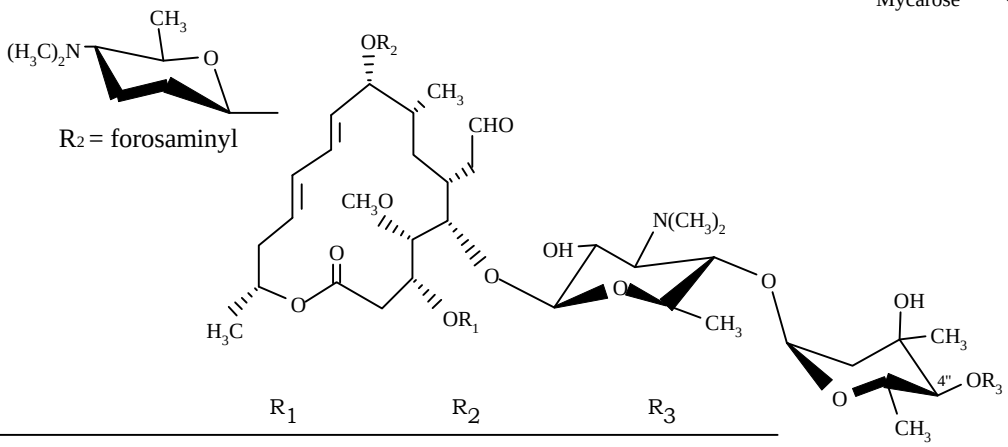
Macrolides are conveniently divided into different groups according to the atom numbers of the formation of lactone rings (Omura, 1984). They are composed of 12 to 16 atoms, so-called 12-membered macrolides (e.g., methymycin); 14-membered macrolides (e.g., erythromycin and oleandomycin); and 16-membered macrolides (e.g., tylosin, spiramycin, carbomycin and midecamycin) (Fig. 1.1).



	R_1	R_2
ERYTHROMYCIN A	OH	CH_3
ERYTHROMYCIN B	H	CH_3
ERYTHROMYCIN C	OH	H
ERYTHROMYCIN D	H	H







	R_1	R_2	R_3
MIDECAMYCIN A_1	$COCH_2CH_3$	H	$COCH_2CH_3$
MIDECAMYCIN A_2	$COCH_2CH_3$	H	$COCH_2CH_2CH_3$
MIDECAMYCIN M_1	$COCH_2CH_3$	H	H
SPIRAMYCIN I	H	forosaminyl	H
SPIRAMYCIN II	$COCH_3$	forosaminyl	H
SPIRAMYCINIII	$COCH_2CH_3$	forsaminyl	H

Fig. 1.1. Examples of macrolides. Structures of one 14-membered (erythromycin) and four 16-membered macrolides (tylosin, midecamycin, spiramycin and carbomycin).

1.2. Cloning of erythromycin and tylosin biosynthesis genes

Large numbers of macrolides have been structurally isolated and some of them have been in practical use for many years, but cloning and genetic analysis of biosynthesis genes have only been extensively studied for erythromycin and tylosin. The molecular biology of picromycin, carbomycin, spiramycin and midecamycin has also been studied but at a much lower level. In the cases of erythromycin and tylosin, the biosynthesis genes are demonstrated to be clustered among one or more resistance-determining genes in the chromosome. Therefore, cloning of the resistance genes has allowed the isolation of large segments of adjacent DNA.

For cloning of erythromycin biosynthesis genes, the *ermE* gene, that determines macrolide-lincosamide-streptogramin B (MLS) resistance in the erythromycin-producing strain *Sac. erythraea*, was first used as a hybridization probe to identify linked genomic fragments from *Sac. erythraea* (Stanzak *et al*., 1986). The first biosynthesis gene identified in the cluster was *eryG*, which encodes the *O*-methyltransferase catalyzing the last step of the pathway (Weber *et al*., 1989). The isolation of the erythromycin gene cluster (ca. 54 kb; Fig. 1.2) has been completed (Dhillon *et al*., 1989; Donadio *et al*., 1991; 1992; Vara *et al*., 1989; Weber *et al*., 1990), and the nucleotide sequence of all the genes in the cluster has been revealed (Cortes *et al*., 1990; Dhillon *et al*., 1989; Donadio *et al*., 1991; Gaisser *et al*., 1997; 1998; Haydock *et al*., 1991; Salah-Bey *et al*., 1998; Summers *et al*., 1997; Weber *et al*., 1991). This macrolide, according to the present knowledge, requires approximately 30 enzymatic steps for its construction.

The first gene in tylosin biosynthesis, designated as *tylF*, encoding macrosin *O*-methyltransferase that acts at the last step in the pathway in *S. fradiae* (Fishman *et al*., 1987), was isolated and cloned by using a reverse genetics approach, which employed a partially degenerated oligonucleotide probe derived from a portion of the N-terminal sequence obtained from the purified enzyme (Bauer *et al*., 1988). Evidence shows tylosin biosynthetic and self-resistance genes are closely linked in the genome of *Streptomyces* (Beckmann *et al*., 1989). Four different resistance genes, *tlrA*, *tlrB*, *tlrC* and *tlrD*, have been cloned from *S. fradiae* (Baltz & Seno, 1988; Birmingham *et al*., 1989; Zalacain &

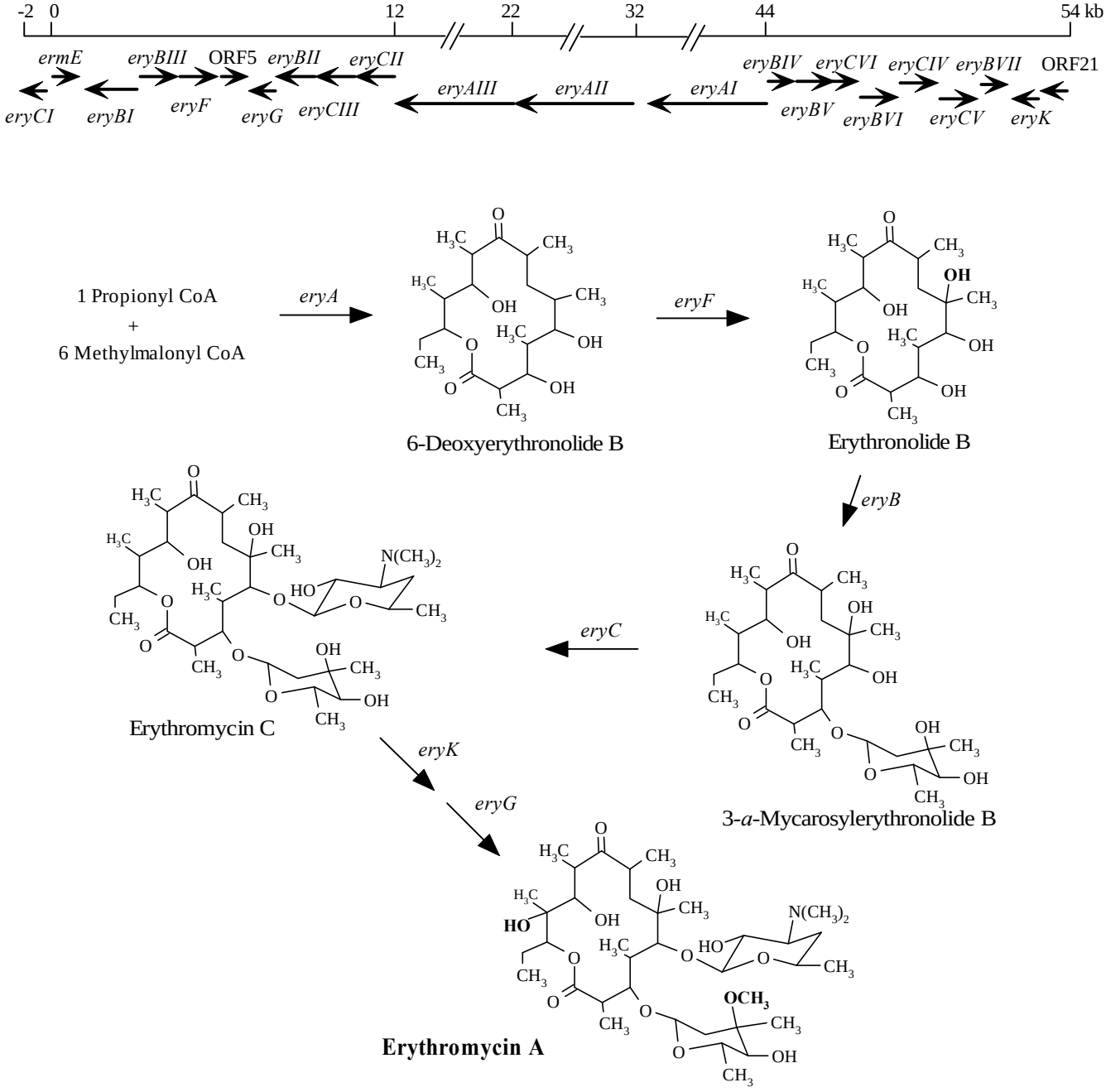


Fig. 1.2. Organisation of the *ery* cluster of *Sac. erythraea* and pathway for the formation of erythromycin. Top: *arrows*, each aligned along the direction of the transcription, indicate the open reading frames of the cluster (not drawn to scale). Bottom: only the macrolide portion of the pathway is illustrated. The proposed pathways for 6-deoxysugar formation see Section 1.4.

Cundliffe, 1991) in which *tlrB* and *tlrC* are located at the two ends of the cluster (Beckmann *et al*., 1989) and *tlrD* is within a region of the cluster (Gandecha *et al*., 1997). More recent analysis shows that the tylosin gene cluster, which covers about 85 kb in the genome of *S. fradiae,* contains over 40 genes, including and surrounding the *tylG* (PKS-encoding) sub-cluster. This large interval segment is flanked by *tlrB* and *tlrC*, and has originally been mapped into 13 different loci (*tylA* to *tylM*) according to the results of co-synthesis studies with idiotrophic mutants blocked at different steps in tylosin biosynthesis (Baltz *et al*., 1983) (Fig. 1.3). The five *tylG* genes occupy ca. 41 kb (acc. no. U78289). The *tylIBA* region, in which 11 genes cover about 14.8 kb, is located between *tylG* and *tlrC* (Merson-Davies & Cundliffe, 1994; Butler *et al*., 1999). The *tylLM* region is located downstream of *tylG*, showing five genes which are mainly involved in mycaminose biosynthesis and attachment (Gandecha *et al*., 1997). The *tylCK* region, in which five genes have recently been found in a 7 kb region, is adjacent to the left side of *tylLM,*. These genes are involved in mycarose biosynthesis and attachment (acc. no. AF147704). The remaining region, *tylEDHFJ*, is located at the left side of the cluster, and recently, the sequence of 11.9 kb of this region has been shown to include 11 orfs, 10 of them belonging to the biosynthetic cluster, which is involved in mycinose biosynthesis and attachment (Fouces *et al*., 1999). Of particular interest in the tylosin gene cluster is the presence of at least five candidate regulatory genes: *tylP* encodes a γ-butyrolactone signal receptor for which *tylQ* is the probable target; *tylQ* is a transcriptional regulator; *tylS* and *tylT* encode pathway-specific regulatory proteins of the *Streptomyces* antibiotic regulatory protein (SARP) family (Wietzorrek & Bibb, 1997); *tylR* has been shown by mutational analysis to control various aspects of tylosin production, thereby designated as a global regulator (Bate *et al*., 1999). The regulatory genes identified here probably control tylosin biosynthesis in cascade fashion and might form a link to the control of sporulation. In contrast, other antibiotic biosynthetic gene clusters are not known to contain multiple pathway-specific regulators, and no regulatory genes are present in the erythromycin gene cluster.

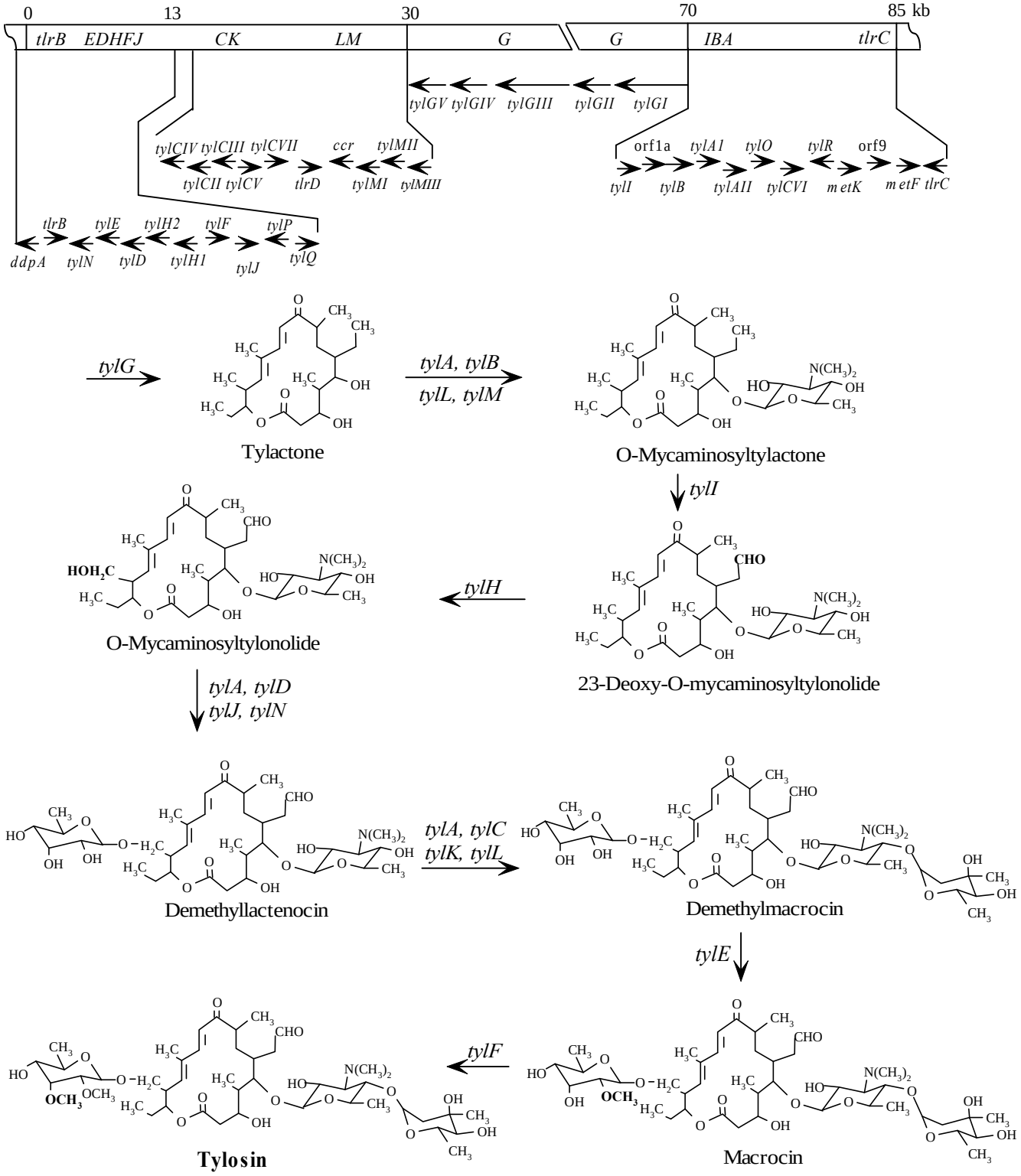


Fig. 1.3. Organisation of the *tyl* cluster of *S. fradiae* and pathway for the formation of tylosin. Top: arrows, denote the open reading frames of the cluster (not drawn to scale). Bottom: only the macrolide portion of the pathway is illustrated. The proposed pathways for 6-deoxysugar formation see Section 1.4.

1.3. Biosynthesis of macrolide lactones

1.3.1 Basis of fatty acid and polyketide biosynthesis

The early stages in both fatty acid and polyketide biosynthesis are catalyzed by enzymes, referred to collectively as fatty acid synthases (FASs) and polyketide synthases (PKSs), that operate on identical principles (Fig. 1.4). Simple carboxylic acids are activated as thioesters (CoA). The acetyl unit which acts as a 'starter' is transferred via an acyl carrier protein (ACP) to a cysteine thiol group at the active site of a β-ketoacyl synthase (KS: the condensing enzyme). A malonate ('extender') unit is then transferred onto the thiol of ACP. Thereafter, the KS catalyzes the condensation of the two acid residues to give a β-ketoacyl C_4 intermediate attached to ACP. At this point the pathway to fatty acids and polyketides usually diverges. In the fatty acid pathway, the β-ketoacyl C_4 intermediate is acted upon successively by a β-ketoacyl reductase (KR), a dehydrase (DH) and an enoyl reductase (ER) to generate a fully reduced acyl C_4 intermediate that will normally be returned to the active-site cysteine of KS to serve as the starter unit for another round of extension and reduction. In the pathway usually followed by PKSs, the β-ketoacyl C_4 intermediate is not automatically reduced; instead it is transferred directly to the active-site cysteine of KS, where it serves as the starter unit for condensation with a second ACP-borne extender unit. The condensation yields a β,δ-diketoacyl C_6 intermediate attached to the ACP, and then this intermediate can in turn be transferred to the vacant active site on the condensing enzyme for further extension of the polyketide chain. Evolutionarily it is more logical to regard FASs as a special case of PKSs, since the basic mechanisms are all identical and catalysed by enzymes all in common families and the PKSs already can contain modules with full β-ketoreduction capacity.

PKSs usually can be subdivided into two types: type I, a modular PKS of complex polyketide, consists of several large multifunctional proteins carrying a separate active site for each individual enzyme-catalyzed reaction in the pathway; type II, the aromatic PKS, consists of a group of separate proteins with iterative active sites for enzyme-catalyzed reactions (Katz & Donadio, 1993). Information about type I PKSs has been obtained from sequence analysis of the cloned genes from *Sac. erythraea* (Cortes *et al*., 1990; Donadio

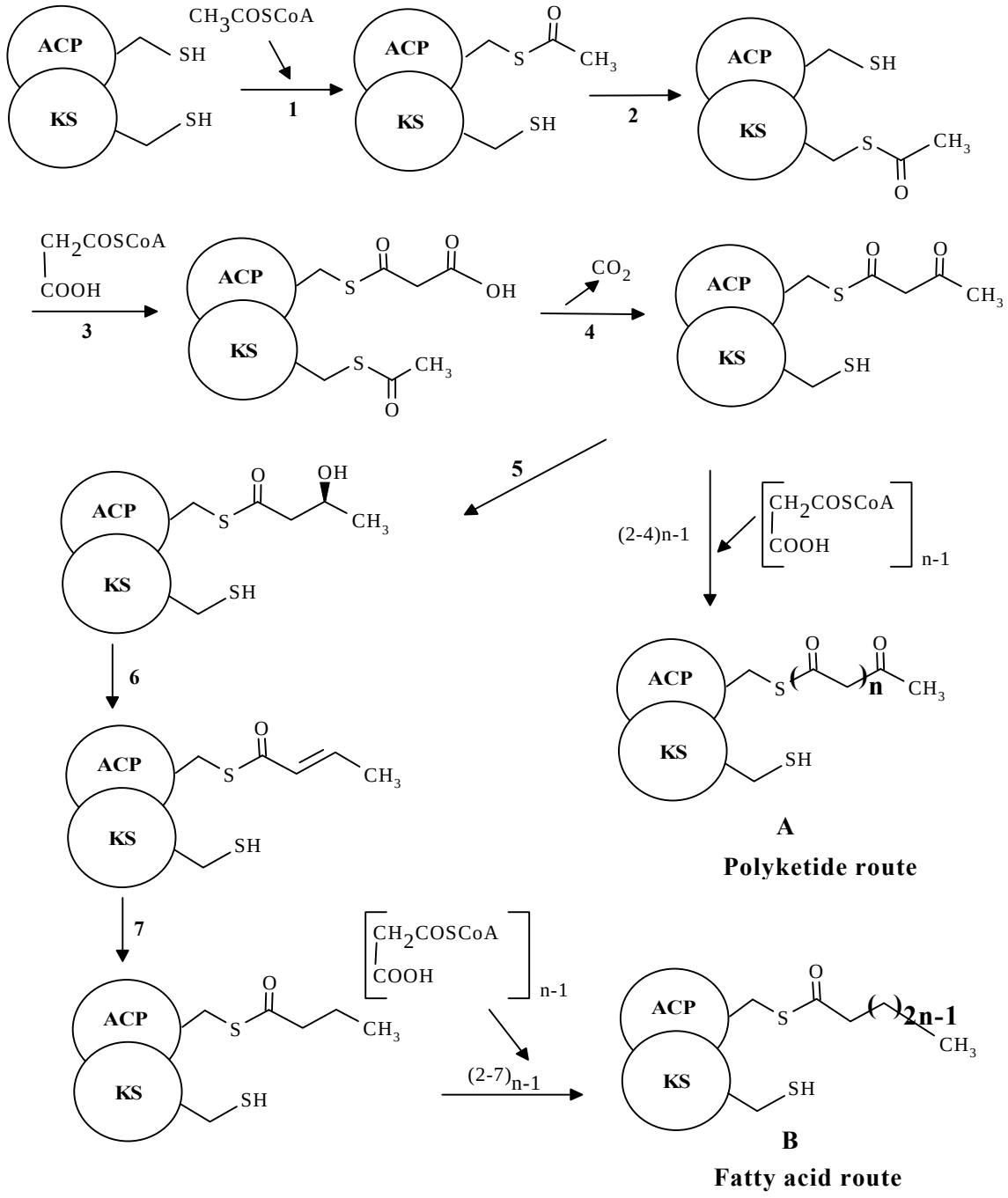


Fig. 1.4. A schematic diagram of fatty acid and polyketide biosynthesis. The first four reactions are common to both pathways; the subsequent reactions in route A lead to polyketides, and in route B to fatty acids. *Circle* represents the synthase complex; ACP is the acyl carrier protein, on which SH is the functional thiol on the 4'-phosphopantetheinyl prosthetic group of the protein. KS is the condensing enzyme, on which SH is the cysteine thiol group at the active site of the enzyme.
Enzymes: 1, acetyl transferase; 2, acyl transferase; 3, malonyl transferase; 4, β-ketoacyl synthase; 5, ketoreductase; 6, dehydrase; 7, enoylreductase.

et al., 1991) and *S. avermitilis* (MacNeil *et al*., 1992b), which reveals that there is a separate set of active sites (a 'module'; Donadio *et al*., 1991) for each round of carbon chain assembly and appropriate reduction. The genetic organization and enzyme function in the biosynthesis of the polyketide aglycone of erythromycin is reviewed below.

1.3.2. Synthesis of macrolide lactone via type I PKSs

As described above, complex polyketides, which include the aglycone components of macrolides, are synthesized through the successive condensation of activated acetate, propionate or butyrate units which are either directly extended or followed by reduction at the β-keto carbon to varying degrees. In the synthesis of the macrocyclic lactones of the well-known macrolides, starter units are commonly acetate (oleandomycin, midecamycin, spiramycin) or propionate (erythromycin and tylosin), and extender units for C_2, C_3 and C_4 chains are likely to arise from malonyl-, methymalonyl- and ethylmalonyl-CoA, respectively. Classical lactone rings contain an even number of atoms: 12- (e.g., methymycin), 14- (e.g., erythromycin, oleandomycin and pikromycin) and 16-membered rings (e.g., midecamycin, spiramycin and tylosin), resulting from the successive condensation of 5, 6 and 7 extender units, respectively (Omura *et al*., 1984).

The nucleotide sequence of the genes for the lactone of erythromycin supports the organization of the modular type I PKSs that is composed of a few very large multifunctional enzymes, each composed of modules containing all the activities required in a single round of synthesis (Cortes *et al*., 1991; Donadio *et al*., 1991; Donadio & Katz, 1992). Corresponding to the sequence of the 6-deoxyerythronolide B (6dEB) PKS genes (*eryA*), the EryA polypeptides include three proteins (EryAI, EryAII and EryAIII) containing domains typical of fatty acid synthases: ACP (acyl carrier protein), AT (acyltransferase), DH (dehydratase), ER (enoylreductase), KR (ketoreductase), KS (ketosynthase) and AT (thioesterase). These are grouped into six modules where modules 1 and 2 in *eryAI* encode a 6-deoxyerythronolide B synthase (DEBS) subunit, DEBS1, believed to participate in the first two steps of 6dEB synthesis. Similarly, *eryAII* and *eryAIII* contain modules 3 and 4, and 5 and 6, respectively, encoding DEBS2 and DEBS3, which catalyze the third through sixth cycles. For the detail processes of 6dEB synthesis

through sequential action of the six modules, the reader is directed to Katz and Donadio (1993).

The genes involved in the synthesis of the 16-membered lactone rings of spiramycin and tylosin have been designated *srmG* and *tylG* in *S. ambofaciens* and *S. fradiae*. Like *eryA*, the genetic organization of the PKS genes for platenolide (*srmG*) and tylactone (*tylG*) synthesis appears to correspond to the order of the chemical reactions, both containing five genes which are grouped into 7 modules, respectively (Kuhstoss *et al*., 1996). Platenolide and tylactone undergo an identical set of post-condensation processing steps; however, the substrates chosen for incorporation in the two pathways differ in some steps. For example, the first reaction in platenolide synthesis involves the condensation of an acetyl-CoA and a malonyl-CoA, while in tylactone synthesis, the corresponding substrates are propionyl-CoA and methylmalonyl-CoA (Omura *et al*., 1975a,b; 1979; Marsden *et al*., 1994). In other type I polyketides, amino acids, aromatics or short branched-chain fatty acids can be used as starter units, e.g., branched chains, an isobutyryl-CoA or 2-methylbutyryl-CoA used in avermectin (MacNeil *et al*., 1992a, b).

1.4. Biosynthesis of 6-deoxysugars

1.4.1. 6-Deoxysugar genes found in antibiotic biosynthesis

In nature, a wide variety of deoxysugars are found in lipopolysaccharides, glycoproteins, glycolipids and many secondary metabolites (Liu & Thorson, 1994; Piepersberg, 1994). The deoxysugars belong to an important class of carbohydrates and exhibit various potent and interesting biological activities (Williams & Wander, 1980). These compounds are formed from common monosaccharides by replacement of one or more hydroxyl group(s) with hydrogen(s). Such a substitute generally induces dramatic variation in biological function, and in many cases, these unusual sugars have been shown to be indispensable for the activity of the parent molecule (Weymouth-Wilson, 1997).

In the past few years, sequencing and analysis of various antibiotic biosynthetic gene clusters and the correlation of blocked mutants with phenotypes have provided critical evidence allowing the identification of a number of complete sugar biosynthetic gene

clusters, for instance, L-streptose biosynthetic cluster in streptomycin (Distler *et al*., 1992); L-daunosamine biosynthetic cluster in daunorubicin biosynthesis (Otten *et al*., 1997); D-desosamine and L-mycarose biosynthetic clusters in erythromycin biosynthesis (Gaisser *et al*., 1997; Summers *et al*., 1997; Salah-Bey *et al*., 1998); D-desosamine and L-oleandrose biosynthetic clusters in oleandomycin biosynthesis (Salas, personal communication; Olano *et al*., 1998); D-mycaminose, D-mycinose and L-mycarose biosynthetic clusters in tylosin biosynthesis (Merson-Davies & Cundliffe, 1994; Gandecha *et al*., 1997; Fouces *et al*., 1999). The recent advances also include the identification of some genes needed for the production of deoxysugar components in avilamycin of *S. viridochromogenes* Tü57 (Gaisser *et al*., 1997), mithramycin of *S. argillaceus* (Lombó *et al*., 1997), nogalamycin of *S. nogalater* (Torkkel *et al*., 1997). The structure of these sugars is shown in Fig. 1.5.

1.4.2. Genetics of 6-deoxysugars in erythromycin and tylosin biosynthesis

Two extensively studied examples, erythromycin and tylosin, have received a leading position in the genetic analysis of 14-membered and 16-membered macrolactone antibiotics, not only because of their well-studied PKSs and PKS-encoding sub-clusters but also because of the identification of almost all pathway-specific genes of appended sugars (Merson-Davies & Cundliffe, 1994; Gandecha *et al*., 1997; Gaisser *et al*., 1997; Summers *et al*., 1997; Salah-Bey *et al*., 1998; Fouces *et al*., 1999).

Erythromycin contains two 6-deoxyhexoses, L-mycarose first attached to the C3-OH of the aglycone 6-deoxyerythronolide B, and D-desosamine then attached to C5-OH of the aglycone. Tylosin contains three 6-deoxyhexoses, D-mycaminose as the first sugar added to C5-OH of tylactone, D-mycinose then added to C23-OH of tylactone, and L-mycarose finally substituted onto C4-OH of mycaminose. According to the pathway analysis and protein similarity of several identified 6-deoxysugar genes in antibiotic biosynthesis, a common pathway of these sugars is deduced as follows: the first two steps are believed to be the synthesis of dTDP-D-glucose and its conversion into dTDP-4-keto-6-deoxyglucose, catalysed respectively by dTDP-glucose synthase and dTDP-glucose dehydratase; thus, dTDP-4-keto-6-deoxyglucose is a common intermediate before the pathway divergance (Piepersberg, 1994). For the L-mycarose production in erythromycin biosynthesis, the *eryB*

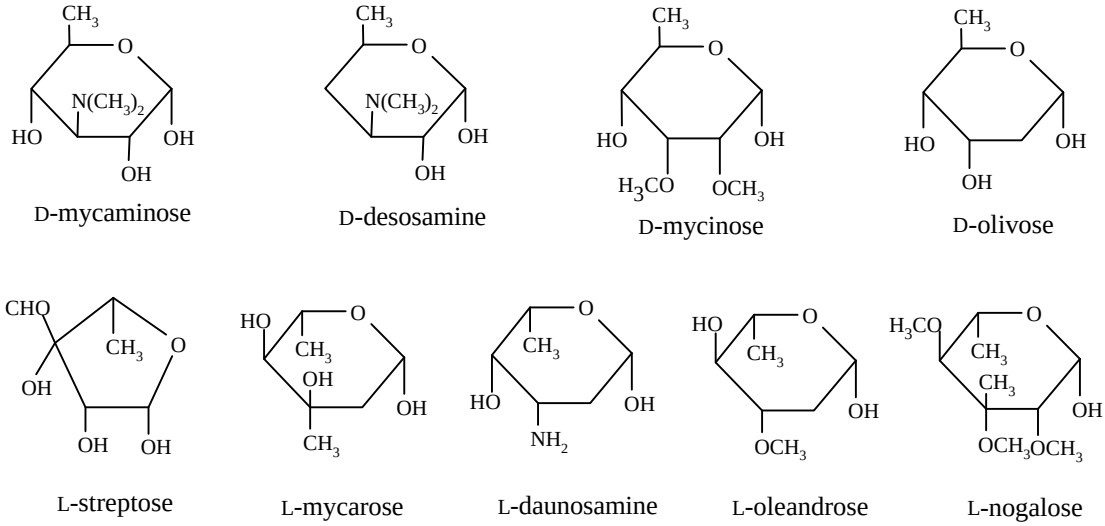


Fig. 1.5. Structure of selected deoxysugars. These sugars participate in the biosynthesis of antibiotics described in the text.

genes contribute to the pathway (see next Section 1.4.3). The genes involved in the D-desosamine pathway have been identified as the *eryC* genes: the *eryCII* product may be responsible for 3,4-isomerase of dTDP-4-keto-6-deoxyglucose, then *eryCI* for C-3 transamination, and *eryCIV* for C-4 dehydration, *eryCV* for C-4 reduction, *eryCVI* for N-methyl transfer and *eryCIII* for glycosyl transfer (Gaisser *et al*., 1997; Salah-Bey *et al*., 1998) (Fig. 1.6). For the production of D-mycaminose involved in tylosin biosynthesis, *tylMIII* was recently identified as 3,4-isomerase (acc. no. X81885), then *tylB* for C-3 transamination, *tylMI* for N-methyl transfer, and *tylMII* for glycosyl transfer (Gandecha *et al*., 1997) (Fig. 1.6).

1.4.3. Mechanism elucidation of some deoxysugar biosynthetic enzymes

It is evident that our knowledge of the biosynthesis of deoxysugars, especially their genetics and pathways, has accumulated rapidly over the past few years (Kirschning *et al*., 1997; Liu & Thorson, 1994). A number of important deoxysugar biosynthetic genes have been cloned and identified, and mechanistic studies of the expressed enzymes have provided fresh insights into deoxysugar biosynthesis.

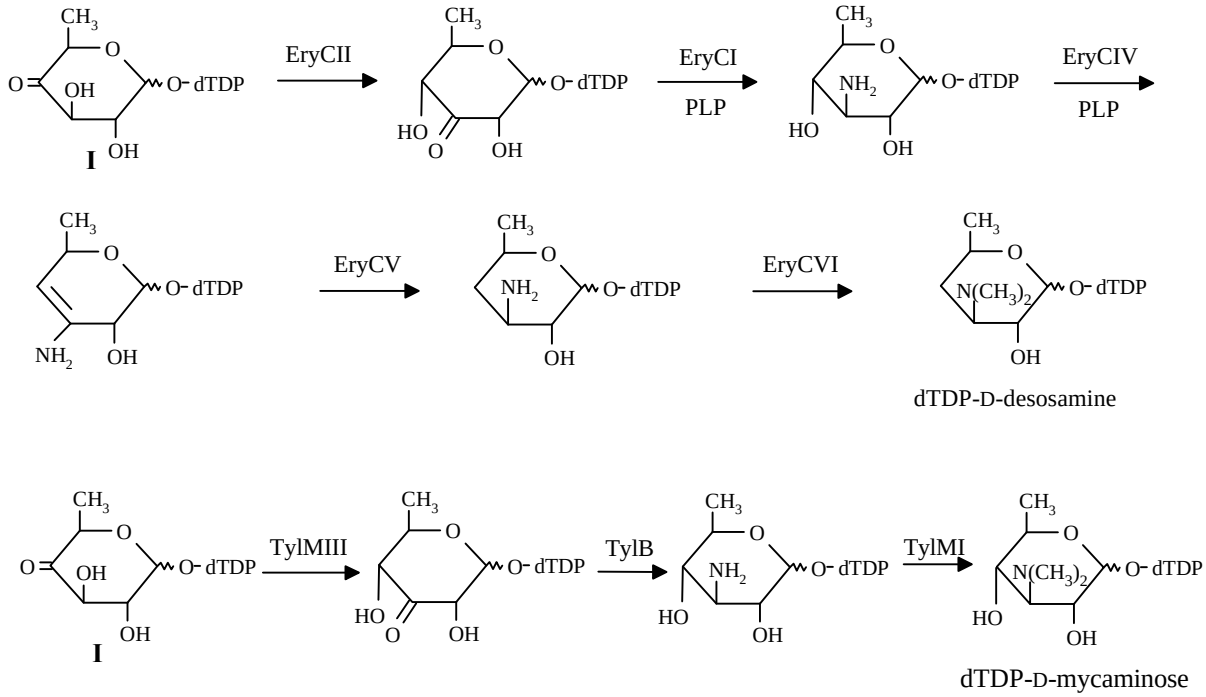


Fig. 1.6. Proposed pathways for dTDP-desosamine synthesis in the erythromycin gene cluster and dTDP-mycaminose synthesis in the tylosin gene cluster. They all start from the common intermediate, dTDP-4-keto-6-deoxyglucose (**I**).

A earlier example is that *in vitro* characterization of several TDP-D-glucose 4,6-dehydratases (TDPGDH; EC 4.2.1.46) from the erythromycin-producing strain *Sac. erythraea* (Vara & Hutchinson, 1988), from the daunorubicin- and baumycin-producing organisms *Streptomyces* sp. C5, and from the daunorubicin-producing strain *S. peucetius* ATCC 29050 (Thompson *et al*., 1992). These TDPGDHs were purified to homogeneity or near to homogeneity and showed requiring NAD^+ as a cofactor. The N-terminal amino acid sequences of the TDPGDHs from *Sac. erythraea* and *S. peucetius* were similar, whereas the enzyme from *Streptomyces* sp. C5 contained a different N-terminal amino acid sequence from either of the other two enzymes (Thompson *et al*., 1992).

A recent notable case is the identification of a set of genes (the cluster *strO-stsABCDEFG*) that encode proteins for streptomycin production in *S. griseus* (Ahlert *et al*., 1997). Sequence analysis revealed that StsA and StsC proteins are members of a new class of aminotransferases that are used mainly in carbohydrate biosynthetic pathways. Purification

of the StsC protein permitted unambiguous assignment of StsC as the L-glutamine:*scyllo*-inosose aminotransferase, which catalyzes the first cyclitol transamination reaction in the biosynthesis of the streptidine subunit of streptomycin. Because genes related to *stsA* and *stsC* also occur in actinomycete producers of other diaminocyclitol aminoglycosides such as neomycins, kanamycins and hygromycin B, the StsA and StsC proteins may be considered as representatives of aminoglycoside-specific aminotransferases.

Another significant example is the recent determination of the function of the *eryBVII* gene in erythromycin biosynthesis (Kim *et al*., 1999). The *eryBVII* was overexpressed in *E. coli* strain and the crude enzyme was able to convert TDP-6-deoxy-L-threo-D-glycero-4-hexulose into the epimerized product. It is not clear at this point whether the epimerization occurs at C-5 and /or C-3. However, since the configuration of the hydroxyl group at C-3 in TDP-6-deoxy-L-threo-D-glycero-4-hexulose is not changed in TDP-L-mycarose, and thus the EryBVII enzyme may catalyze the epimerization at C-5 only. According to this experiment, the biosynthetic route of TDP-L-mycarose from TDP-6-deoxy-L-threo-D-glycero-4-hexulose, or of dTDP-L-mycarose from dTDP-4-keto-6-deoxyglucose, requires at least four chemical transformations, as proposed by Gaisser *et al*. (1998) and Summers *et al*.(1997). They would be the epimerization at C-5 (EryBVII), the deoxygenation at C-2 (?), the methylation at C-3 (EryBIII) and the reduction at C-4 (EryBIV) (Fig. 1.7).

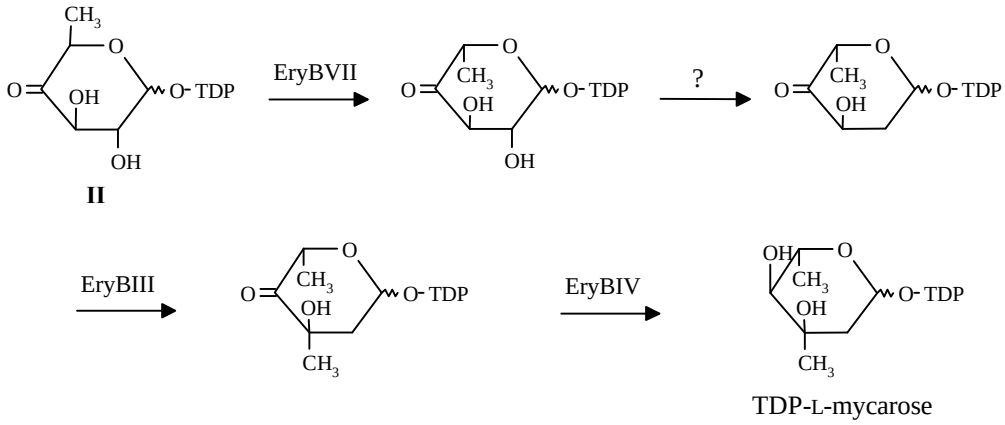


Fig. 1.7. Possible biosynthetic pathway of TDP-L-mycarose in *Sac. erythraea* from TDP-6-deoxy-L-threo-D-glycero-4-hexulose (II).

1.5. Aims of the present study

This study belongs to a project designed to first elucidate basic mechanisms of the biosynthesis and transfer of sugar residues present on macrolide compounds. Secondly, this knowledge should be used to produce new hybrid macrolide antibiotics by re-targeting sugar residues on heterologous lactones. The particular target of this work was cloning of the gene cluster from a 16-membered macrolide producing strain. The aims of the research were:

(1) to clone the gene cluster of midecamycin biosynthesis from *S. mycarofaciens.*
(2) to identify genes involved in mycaminose and mycarose biosynthesis from the midecamycin biosynthetic gene cluster.
(3) to over-express the 6-deoxysugar genes in *E. coli* and *S. lividans.*
(4) to *in vitro* analyse the expressed protein function.

2. Materials

2.1. Chemicals and enzymes

Antibiotics	
Ampicillin	Roche, Mannheim
Chloramphenical	Serva, Heidelberg
Kanamycin	Serva, Heidelberg
Midecamycin	Sigma, Deisenhofen
Thiostrepton	Squibb and Sons, Princeton, USA
Chemicals p.A. quality	
	Fluka, Buchs, CH Merck, Darmstadt Roth, Karlsruhe Serva, Heidelberg Sigma, Deisenhofen
Enzymes	
Alkaline Phosphatase (calfs intestine)	Roche, Mannheim
DNA-Polymerase I Klenow-Fragment	Life Technologies, Eggenstein
Lysozyme	Serva, Heidelberg
Protease Inhibitor Set	Roche, Mannheim
Restriction enzymes	New England Biolabs, Schwalbach Roche, Mannheim Life Technologies, Eggenstein Promega-Serva, Heidelberg
Ribonuclease A	Sigma, Deisenhofen
Taq DNA ploymerase	Life Technologies, Eggenstein
T4-DNA ligase	Life Technologies, Eggenstein
Vent DNA polymerase	New England Biolabs, Schwalbach
Media Components	
	Difco, Detroit, USA Merck, Darmstadt Oxoid, Wesel

	Roth, Karlsruhe

Kits

Bio-Rad Protein Assay Kit	Bio-Rad, München
BM Chromogenic Western Blotting Kit	Roche, Mannheim
QIAquick PCR Purification Kit	Qiagen, Hilden
QIAprep Spin Miniprep Kit	Qiagen, Hilden
NucleoSpin Extract	Macherey-Nagel, Düren
Rediprime Random Primer Labeling Kit	Amersham Buchler, Braunschweig
Thermosequenase Cycle-Sequencing Kit	Pharmacia, Freiburg

Other materials

α-^{32}P-dCTP	Amersham Buchler, Braunschweig
Hybond-N^{+} Membrane	Amersham Buchler, Braunschweig
Hybond-P Membrane	Amersham Buchler, Braunschweig
Membrane filters BA 85 (0.45 μm)	Schleicher & Schuell, Dassel
3MM Whatman	Biometra, Göttingen
X-ray film Hyperfilm-MP	Amersham Buchler, Braunschweig
X-ray film Hyperfilm-β-max	Amersham Buchler, Braunschweig

2.2. Media and Buffers

2.2.1. Media for cultivation of *E. coli*

LB medium (Miller, 1972)

Difco Tryptone	10	g/l
Difco Yeast extract	5	g/l
NaCl	5	g/l

LB agar

LB medium		
Agar	15	g/l

SNA: Soft Nutrient Agar (Hopwood *et al*., 1985)

Difco Nutrient Broth	8	g/l
Difco Bacto Agar	3	g/l

SOB (Hanahan,1983)

Difco Tryptone	20	g/l
Difco Yeast extract	5	g/l
NaCl	0.58	g/l
KCl	0.19	g/l
After autoclaving, supplemented with:		
$MgCl_2$ x 6 H_2O (1M)	10	ml/l
MgSO4 x 7 H_2O (1M)	10	ml/l

SOC (Hanahan, 1983)

Glucose in SOB	3.6	g/l

2.2.2. Media for cultivation of *Streptomyc*es

EP1 (erythromycin production) (Salah-Bey *et al*., 1998)

Solulys L corn steep liquor	5	g/l
Defatted soya flour	10	g/l
$CaCO_3$	2	g/l
NaCl	5	g/l
pH	6.8	
After autoclaving, supplemented with:		
Glucose	15	g/l

EP2 (erythromycin production) (Salah-Bey *et al*., 1998)

Defatted soya flour	10	g/l
$CaCO_3$	2	g/l
$CoCl_2$ x 6 H_2O	1	mg/l
pH	6.8-7.0	
After autoclaving, supplemented with:		
Glucose	20	g/l

ISP2 (according to the International Streptomyces Project)

Difco Yeast Extract	4	g/l
Difco Malt Extract	10	g/l
Difco Dextrose	4	g/l
Agar	20	g/l

SMA (Distler *et al*., 1985)

Soybean powder	20	g/l

Mannitol	20	g/l
Agar	20	g/l
Tap water used		

SPMR (Babcock & Kendrick, 1988)

Sucrose	103	g/l
$MgCl_2$ x 6 H_2O	10	g/l
Glucose	5	g/l
Difco yeast extract	5	g/l
TES buffer, pH7.6 (1M)	20	ml/l
Trace elements solution*	2	ml/l
Difco-Bacto agar	22	g/l
After autoclaving, supplemented with:		
$CaCl_2$ x 2 H_2O (5M)	2	ml/l

*Trace elements solution (Hopwood *et al*., 1985)

$ZnCl_2$	0.04	g/l
$FeCl_2$ x 6 H_2O	0.2	g/l
$CuCl_2$ x 2 H_2O	0.01	g/l
$MnCl_2$ x 4 H_2O	0.001	g/l
$Na_2B_4O_7$ x 10 H_2O	0.01	g/l
$(NH_4)_6Mo_7O_{24}$ x 4 H_2O	0.01	g/l
Sterile filtration		

TSB (Hopwood *et al*., 1985)

Tryptone Soja Broth (Oxoid)	30	g/l

TSB-PEG 8000 (Babcock & Kendrick, 1988)

TSB	30	g/l
PEG 8000	50	g/l

YEME (Hopwood *et al*., 1985)

Difco yeast extract	3	g/l
Difco peptone	5	g/l
Difco malt extract	3	g/l
Glucose	10	g/l
Sucrose	340	g/l
After autoclaving, supplemented with:		
$MgCl_2$ x 6H_2O (2.5M)	2	ml/l

SGYEME

A modified YEME medium.		
Sucrose	110	g/l

2.2.3. Buffers

-for isolation of plasmid DNA

Alkaline Lysis Solution (Birnboim & Doly, 1979)

Solution I:		
Glucose	50	mM
Tris/HCl, pH 8.0	50	mM
EDTA, pH 8.0	10	mM
Solution II:		
SDS	20	ml/l
NaOH	0.2	M
Solution III:		
Potassium acetate	3	M
Formic acid	1.8	M

STET buffer (Sambrook *et al*., 1989)

Sucrose	80	g/l
Triton X-100	50	ml/l
Tris/HCl, pH 8.0	50	mM
EDTA, pH 8.0	50	mM

Lysozyme solution (Hopwood *et al*., 1985)

Sucrose	0.3	M
Tris/HCl, pH 8.0	25	mM
EDTA, pH 8.0	25	mM

-for preparation of competent *E. coli* cells (Hanahan, 1983)

FSB-buffer

Potassium acetate, pH7.0	10	mM
KCl	100	mM
$MnCl_2$ x 4 H_2O	45	mM
$CaCl_2$ x 2 H_2O	10	mM
Hexaminecobaltchloride	3	mM
Glycerol	100	g/l
pH	6.4	
Filtration		

-for preparation of genomic DNA from *Streptomyces* (Pospiech & Neumann, 1995)

SET buffer

NaCl	75	mM
EDTA	25	mM
Tris	20	mM
pH	7.5	

TE buffer (Sambrook *et al*., 1989)

Tris	10	mM
EDTA	1	mM
pH	8.0	

- for preparation of protoplasts from *Streptomyces* (Chater *et al*., 1982)

P-buffer

Sucrose	103	g
K_2SO4	0.25	g
$MgCl_2$ x 6 H_2O	2.02	g
Trace elements	2	ml
Add H_2O to 800 ml volume		
Autoclave in 80 ml aliquots and supplement with:		
K_2HPO4 (0.5%)	1	ml
$CaCl_2$ x 2 H_2O (3.68%)	10	ml
TES (5.73%), pH7.2	10	ml

- for transformation in *Streptomyces* (Babcock & Kendrick, 1988)

T-buffer

Sucrose (10.3%)	25	ml
K_2SO4 (2.5%)	1	ml
H_2O	75	ml
Trace elements	0.2	ml
CaCl2 (0.25 M)	0.2	ml
Tris/maleicacid (1M); pH 8.0	0.5	ml

T-buffer/PEG

Polyethyleneglycol 1000	1	g
T-buffer	900	µl

- for agarose gel electrophoresis of DNA fragments (Sambrook *et al*., 1989)

DNA loading buffer (10 x)

Sucrose	0.5	g/ml
EDTA; pH 8.0	160	mM
Bromophenol Blue	0.5	mg/ml
Xylene cyanol	0.5	mg/ml

TAE (50 x)

Tris	242	g/l
Glacial acetic acid	57.1	ml/l
EDTA	0.4	g/l
pH	8.3	

TBE (10 x)

Tris	108	g/l
Boricacid	61	g/l
EDTA	0.4	g/l

-for constructure of genomic libraries

Sucrose gradients (Weis, 1987)

Sucrose	100	g/l
Tris/HCl, pH 8.0	10	mM
EDTA	1	mM
NaCl	1	M
autoclaved		

SM phage dilution buffer (Boehringer's DNA packaging protocol)

Tris/HCl pH 7.5	50	mM
MgSO4	10	mM
NaCl	100	mM
gelatine	0.1	g/l
autoclaved		

- for DNA-DNA hybridization (Southern, 1975)

Hybridization solution

SSC (20 x)	330	ml/l
Sodium phosphate buffer (1M), pH6.8	10	ml/l
EDTA (0.5 M), pH 8.0	2	ml/l
SDS (10%)	50	ml/l ???
Blocking reagent	1	g/l

Wash solution (2 x)

SSC (20 x)	200	ml/l
SDS (10%)	10	ml/l

Wash solution ((0.5 x)

SSC (20 x)	25	ml/l
SDS (10%)	10	ml/l

SSC (20 x)

NaCl	3	M
Sodium citrate	0.3	M
pH 7.2 adjust with citric acid		

- for denatured polyacrylamide gel electrophoresis (Laemmli, 1970)

30% PAA solution

Acrylamide	290	g/l
N,N-methylene bisacrylamid	10	g/l

Resolving gel buffer

Tris/HCl; pH 8.9	1.5	M
SDS	4	g/l

Stacking gel buffer

Tris/HCl; pH 6.7	0.5	M
SDS	1	g/l

SDS electrophoresis running buffer (10 x)

Tris	30	g/l
Glycine	144	g/l
SDS	10	g/l
pH	8.3	

Sample loding buffer (5 x)

SDS (20 g/l)	2	ml
β-mercaptoethanol	4	ml
Glycerol	2	ml
Bromophenol Blue (1 g/l in H_2O)	2	ml

Gel staning solution

Coomassie Brilliant Blue R250	1.5	g/l
Methanol	450	ml/l
Acetic acid	100	ml/l

Destaining solution

Methanol	250	ml/l
Acetic acid	100	ml/l

-for Western blotting and detection of proteins

Western blotting buffer

Tris	98	mM
Glycine	39	mM
10% SDS	3.75	ml/l
Methanol	200	ml/l

- for sonification of expression proteins

Buffer 1

Tris/HCl, pH 7.5	50	mM
DTT	1	mM
$MgCl_2$	10	mM
EDTA, pH 8.0	1	mM

Buffer 2

Potassium phosphate buffer, pH 7.2	50	mM
β-mercaptoethanol	10	mM
EDTA, pH 8.0	0.1	mM

Buffer 3

Buffer 2		
PLP	50	µM

- for HPLC analysis

Reversed phase chromatography

Running buffer A:

Potassium phosphate buffer, pH 6.0	30	mM
Tetrabutylammoniumhydrogen	5	mM

sulphate		
Acetonitril	20	ml/l
Running buffer B:		
Acetonitril	100	%

2.3 Bacteria, cosmids and plasmids

2.3.1. Bacteria

Tab. 2.1. Bacteria used in this study

Strain	Genotype/Product	Reference/Origin
E. coli BL21(DE3) pLysS	*ompT*, *hsdSB* ($r_B^- m_B^-$), λ prophage with polymerase gene, pLysS, *cat*	Studier *et al*., 1990
E. coli DH5α	F[JS1], ϕ80d*lac*ZΔM15, *recA*1, *endA*1, *gyrA*96, *thi*-1, *hsd R*17(r_k^-, m_k^+), *supE*44, *relA*1, *deoR*, Δ(*lac*ZYA-*arg*F)U169	D. Hanahan, 1983
E. coli JM108	*endA*1, *recA*1, *gyrA*96, *thi*, *hsd R*17 ($r_k^- m_k^+$), *relA*1, *supE*44, Δ(*lac-proAB*),	Yanisch-Perron *et al.,* 1985
E. coli JM109	*endA*1, *recA*1, *gyrA*96, *thi*, *hsd R*17 ($r_k^- m_k^+$), *relA*1, *supE*44, Δ(*lac-proAB*), [F', *traD*36, *proAB*, *lacI*^{q}ZΔM15]	Yanisch-Perron *et al.,* 1985
E. coli JM109 (DE3)	*endA*1, *recA*1, *gyrA*96, *thi*, *hsd R*17 ($r_k^- m_k^+$), *relA*1, *supE*44, Δ(*lac-proAB*), [F', *traD*36, *proAB*, *lacI*^{q}ZΔM15], λ(DE3)	Yanisch-Perron *et al.,* 1985
E. coli XL1-Blue	*rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17, *sup*E44, *rel*A1, *lac*, [F' *pro*AB, *lac*I^{q}ZΔM15, Tn10 (*tet*^{R})]	Loenen & Blattner, 1983
ET12567	F$^-$, *dam*13::Tn9, *dcm*6, *hsdM*, *hsdR*, *recF*143, *zjj*201::Tn10, *galK*2, *galT*22, *ara*-14, *lacY*1, *xyl*-5, *leuB*6, *thi*-1, *tonA*31, *rpsL*136, *hisG*4, *tsx*-78, *mtl*-1, *glnV*44	MacNeil *et al.,* 1992
S. lividans 66 1326	actinorhodin, prodigiosin	John Innes Institute, Norwich, UK
S. lividans 66 TK23	actinorhodin, *spc-1*	John Innes Institute, Norwich, UK
S. mycarofaciens	non-producing	UC189B (ATCC 21454) Heochst
S. mycarofaciens	midecamycin	ATCC 21454

2.3.2. Vectors

Tab. 2.2. Plasmid and cosmid vectors used in this study

Plasmid/Cosmid	Genetype, Properties and Use	Reference/Origin
pAL201	*bla, tsr*, pUC18 ori, pJV1 ori, *Streptomyces-E. coli* shuttle vector, for transformation in *Saccharopolyspora erythraea*	Hoechst Marion Roussel
pDNW16RBSY	derived from pUWL201 with *lmbY* RBS, His-tag, expression of *lmbY* gene	Neußer, Uni. Wuppertal
pEFBA	*bla*, aparamycin resistance cassette, for mutagenesis	Fernandez, Uni. Oviedo
pET11a	*bla, lacI*T7-ϕ10, for heterologous expression in *E. coli*	Studier *et al.,* 1990
pET16b	*bla, lacI*T7-ϕ10, His-tag, for heterologous expression in *E. coli*	Novagen, Heidelberg
pHM8a	*hyg*, mini-circle, integrative into *Streptomyces*	Motamedi *et al*., 1995
pIJ4123	*kan*, *tsr*, *redD*, His-tag, for heterologous expression in *S. lividans*	Takano *et al.,* 1995
pSL1180	*bla*, polylinker containing 78 recognition sites, for subcloning	Pharmacia
pSVW701	bla, T7-ϕ10, for expression of *rmlB* gene	Verseck, 1997
pUWL201	*bla*, *tsr*, pUC18 ori, pIJ101 ori, *ermE* up promoter, *Streptomyces-E. coli* shuttle plasmid, for heterologous expression in *S. lividans*	Wehmeier, Uni. Wuppertal
pUWL219	*bla*, *tsr*, pUC18 ori, pIJ101 ori, *Streptomyces-E. coli* shuttle plasmid for subcloning	Wehmeier, 1995
pUC18	*bla, lacZ-a*, for subcloning	Vieira & Messing, 1982
pKU206	SCP2* ori and stability, pMB1 ori, *tsr*, *bla*, *cos*, *Streptomyces-E. coli* shuttle cosmid for gene library	Kakinuma *et al*., 1991

2.3.3. Recombinant plasmids

Tab. 2.3. New constructed plasmids for screening *S. mycarofaciens* genomic library

Plasmid	Description	Host strain
pLCW1	301 bp PCR fragment by primers AS2 and AS5 (identified as *midB* gene) from *S. mycarofaciens* in pUC18 *Sma*I	*E. coli* DH5α
pLCW17	420 bp PCR fragment by primers LC6 and EryCIII/I (identified as *midI* gene) from *S. mycarofaciens* in pUC18 *Sma*I	*E. coli* DH5α

Tab. 2.4. Recombinant cosmid plasmids constructed in this study

Cosmid plasmid	Description	Host strain
Smyc-LC1	ca. 30.2 kb fragment from partially digested *S. mycarofaciens* gemomic DNA with *Sau*3AI in pKU206 *Bam*HI	*E. coli* JM108 *S. lividans* TK23
Smyc-LC2	ca. 22.6 kb fragment from partially digested *S. mycarofaciens* gemomic DNA with *Sau*3AI in pKU206 *Bam*HI	*E. coli* JM108 *S. lividans* TK23
Smyc-LC3	ca. 26.3 kb fragment from partially digested *S. mycarofaciens* gemomic DNA with *Sau*3AI in pKU206 *Bam*HI	*E. coli* JM108 *S. lividans* TK23
Smyc-LC4	ca. 29.3 kb fragment from partially digested *S. mycarofaciens* gemomic DNA with *Sau*3AI in pKU206 *Bam*HI	*E. coli* JM108 *S. lividans* TK23

Tab. 2.5. New constructed plasmids for sequencing

Plasmid	Description	Host strain
pLC1-1	0.8 kb *Bam*HI/*Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-2	1.5 kb *Bam*HI/*Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-3	2.3 kb *Bam*HI/*Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-4	4.2 kb *Bam*HI/*Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-5	4.2 kb *Bam*HI/*Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-6	4.8 kb *Bam*HI/*Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-7	8.2 kb *Bam*HI/*Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-8	0.8 kb *Sst*I fragment from pLC1-2 in pUC18 *Sst*I	*E. coli* DH5α
pLC1-9	0.7 kb *Sst*I fragment from pLC1-2 in pUC18 *Sst*I	*E. coli* DH5α
pLC1-10	4.0 kb *Eco*RI/*Hin*dIII fragment from pLC1-7 in pUC18 *Eco*RI/*Hin*dIII	*E. coli* DH5α
pLC1-11	2.0 kb *Eco*RI fragment from pLC1-7 in pUC18 *Eco*RI	*E. coli* DH5α
pLC1-12	1.0 kb *Eco*RI fragment from pLC1-7 in pUC18 *Eco*RI	*E. coli* DH5α
pLC1-13	0.9 kb *Eco*RI fragment from pLC1-7 in pUC18 *Eco*RI	*E. coli* DH5α
pLC1-14	4.2 kb *Bam*HI/*Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-15	2.1 kb *Sph*I/*Bam*HI fragment from pLC1-4 in pUC18 *Sph*I/*Bam*HI	*E. coli* DH5α
pLC1-16	1.3 kb *Sph*I fragment from pLC1-4 in pUC18 *Sph*I	*E. coli* DH5α
pLC1-17	0.7 kb *Sph*I fragment from pLC1-4 in pUC18 *Sph*I	*E. coli* DH5α

pLC1-18	1.8 kb *Pst*I/*Eco*RI fragment from pLC1-10 in pUC18 *Pst*I/*Eco*RI	*E. coli* DH5α
pLC1-19	1.5 kb *Pst*I fragment from pLC1-10 in pUC18 *Pst*I	*E. coli* DH5α
pLC1-20	0.6 kb *Pst*I/*Sst*I fragment from pLC1-18 in pUC18 *Pst*I/*Sst*I	*E. coli* DH5α
pLC1-21	1.2 kb *Eco*RI/*Sst*I fragment from pLC1-18 in pUC18 *Eco*RI/*Sst*I	*E. coli* DH5α
pLC1-22	1.5 kb *Bam*HI fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* DH5α
pLC1-23	0.2 kb *Pst*I/*Bam*HI fragment from pLC1-10 in pUC18 *Pst*I/*Bam*HI	*E. coli* DH5α
pLC1-24	0.4 kb *Pst*I fragment from pLC1-10 in pUC18 *Pst*I	*E. coli* DH5α
pLC1-25	1.6 kb *Eco*RI fragment from pLC1-4 in pUC18 *Eco*RI	*E. coli* DH5α
pLC1-26	1.4 kb *Eco*RI fragment from pLC1-4 in pUC18 *Eco*RI	*E. coli* DH5α
pLC1-27	0.75 kb *Eco*RI fragment from pLC1-4 in pUC18 *Eco*RI	*E. coli* DH5α
pLC1-28	0.55 kb *Eco*RI fragment from pLC1-4 in pUC18 *Eco*RI	*E. coli* DH5α
pLC1-29	0.7 kb *Sst*I fragment from pLC1-19 in pUC18 *Sst*I	*E. coli* DH5α
pLC1-30	0.9 kb *Sst*I fragment from pLC1-19 in pUC18 *Sst*I	*E. coli* DH5α
pLC1-31	14 kb *Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* XL1-Blue
pLC1-32	5.1 kb *Bgl*II fragment from Smyc-LC1 in pUC18 *Bam*HI	*E. coli* XL1-Blue
pLC2-1	8.2 kb *Bam*HI/*Bgl*II fragment from Smyc-LC2 in pUC18 *Bam*HI	*E. coli* DH5α
pLC2-2	0.7 kb *Bam*HI/*Bgl*II fragment from Smyc-LC2 in pUC18 *Bam*HI	*E. coli* DH5α
pLC2-3	1.5 kb *Bam*HI/*Bgl*II fragment from Smyc-LC2 in pUC18 *Bam*HI	*E. coli* DH5α

pLC2-4	5.0 kb *Bam*HI/*Bgl*II fragment from Smyc-LC2 in pUC18 *Bam*HI	*E. coli* DH5α
pLC2-5	0.8 kb *Bam*HI/*Bgl*II fragment from Smyc-LC2 in pUC18 *Bam*HI	*E. coli* DH5α
pLC2-6	2.0 kb *Bam*HI/*Bgl*II fragment from Smyc-LC2 in pUC18 *Bam*HI	*E. coli* DH5α
pLC2-7	4.2 kb *Bam*HI/*Bgl*II fragment from Smyc-LC2 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-1	6.7 kb *Bam*HI fragment from Smyc-LC3 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-2	5.9 kb *Bam*HI fragment from Smyc-LC3 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-3	4.3 kb *Bam*HI fragment from Smyc-LC3 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-4	2.8 kb *Bam*HI fragment from Smyc-LC3 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-5	1.8 kb *Bam*HI fragment from Smyc-LC3 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-6	1.5 kb *Bam*HI fragment from Smyc-LC3 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-7	0.9 kb *Bam*HI fragment from Smyc-LC3 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-8A	1.6 kb *Bam*HI/*Eco*47III fragment from pLC3-3 in pUC18 BamHI/*Hin*cII	*E. coli* DH5α
pLC3-9A	1.6 kb *Bam*HI/*Eco*47III fragment from pLC3-3 in pUC18 BamHI/*Hin*cII	*E. coli* DH5α
pLC3-10	0.7 kb *Bam*HI/*Sal*I fragment from pLC3-3 in pUC18 *Bam*HI/*Sal*I	*E. coli* DH5α
pLC3-11	1.5 kb *Bam*HI fragment from Smyc-LC3 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-12	1.2 kb *Sal*I fragment from pLC3-3 in pUC18 *Sal*I	*E. coli* DH5α
pLC3-13	0.9 kb *Bam*HI/*Hin*cII fragment from pLC3-3 in pUC18 *Bam*HI/*Hin*cII	*E. coli* DH5α
pLC3-14	0.7 kb *Sal*I fragment from pLC3-3 in pUC18 *Sal*I	*E. coli* DH5α

pLC3-15	0.7 kb *Hin*cII fragment from pLC3-3 in pUC18 *Hin*cII	*E. coli* DH5α
pLC3-16	1.2 kb *Eco*47III fragment from pLC3-3 in pUC18 *Hin*cII	*E. coli* DH5α
pLC3-17A	0.2 kb *Bam*HI/*Sst*I fragment from pLC3-6 in pUC18 *Bam*HI/*Sst*I	*E. coli* DH5α
pLC3-18A	0.6 kb *Bam*HI/*Sst*I fragment from pLC3-6 in pUC18 *Bam*HI/*Sst*I	*E. coli* DH5α
pLC3-18D	0.6 kb *Sst*I fragment from pLC3-6 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-19A	1.4 kb *Bam*HI/*Bgl*II fragment from pLC3-4 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-19B	1.4 kb *Bam*HI/*Bgl*II fragment from pLC3-4 in pUC18 *Bam*HI	*E. coli* DH5α
pLC3-20	2.3 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-21	1.7 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-22	1.7 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-23	1.4 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-24	1.4 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-25	1.2 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-26	0.8 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-27	0.6 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-28	0.6 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-29	0.4 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-30	0.4 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-31	6.0 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-32	0.45 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-33	0.45 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-34	0.4 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α
pLC3-35	0.2 kb *Sst*I fragment from Smyc-LC3 in pUC18 *Sst*I	*E. coli* DH5α

Tab. 2.6. New constructed plasmids for expression

Plasmid	Description	Host strain
pLCW1-22	1.3 kb PCR fragment by primers LC9 and LC10 (contaning *midI* gene) from Smyc-LC3 in pUC18 *Sma*I	*E. coli* XL1-Blue
pLCW1-23	1.3 kb *Nde*I/*Bam*HI fragment from pLCW1-22 in pET16b *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW1-33	1.3 kb *Nde*I/*Bam*HI fragment from pLCW1-22 in pET11a *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW1-41a	1.7 kb *Xba*I/*Hin*dIII-blunt fragment from pLCW1-33 in pUC18 *Xba*I/*Sma*I	*E. coli* XL1-Blue
pLCW1-41b	1.7 kb *Eco*RI/*Hin*dIII fragment from pLCW1-41b in pUWL201	*E. coli* XL1-Blue *S. lividans* TK23
pLCW1-51a	1.75 kb *Xba*I/*Hin*dIII-blunt fragment from pLCW1-23 in pUC18 *Xba*I/*Sma*I	*E. coli* XL1-Blue
pLCW1-51b	1.75 kb *Eco*RI/*Hin*dIII fragment from pLCW1-51b in pUWL201	*E. coli* XL1-Blue *S. lividans* TK23
pLCW1-61	1.3 kb *Nde*I/*Bam*HI fragment from pLCW1-22 in pIJ4123	*S. lividans* 66 1326
pLCW2-1	1.2 kb PCR fragment by primers LC13 and LC14 (containing *midC* gene) from Smyc-LC1 in pUC18 *Sma*I	*E. coli* XL1-Blue
pLCW2-20	1.2 kb *Nde*I/*Bgl*II fragment from pLCW2-1 in pET16b *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW2-8	1.2 kb PCR fragment by primers LC17 and LC14 (containing *midC* gene) from Smyc-LC1 in pUC18 *Sma*I	*E. coli* XL1-Blue
pLCW2-21	1.2 kb *Nde*I/*Bgl*II fragment from pLCW2-8 in pET16b *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW2-23	1.2 kb *Nde*I/*Bgl*II fragment from pLCW2-8 in pET11a *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW3-1	0.8 kb PCR fragment by primers LC11 and LC12 (containing *midK* gene) from Smyc-LC3 in pUC18 *Sma*I	*E. coli* XL1-Blue

pLCW3-2	0.8 kb *Nde*I/*Bam*HI fragment from pLCW3-1 in pET11a *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW3-3	0.8 kb *Nde*I/*Bam*HI fragment from pLCW3-1 in pET16b *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW3-6	1.2 kb *Xba*I/*Hin*dIII-blunt fragment from pLCW3-2 in pUC18 *Xba*I/*Sma*I	*E. coli* XL1-Blue
pLCW3-7	1.25 kb *Xba*I/*Hin*dIII-blunt fragment from pLCW3-3 in pUC18 *Xba*I/*Sma*I	*E. coli* XL1-Blue
pLCW3-8	1.2 kb *Eco*RI/*Hin*dIII fragment from pLCW3-6 in pUWL201	*E. coli* XL1-Blue *S. lividans* TK23
pLCW3-9	1.25 kb *Eco*RI/*Hin*dIII fragment from pLCW3-7 in pUWL201	*E. coli* XL1-Blue *S. lividans* TK23
pLCW3-12	0.8 kb *Nde*I/*Bam*HI fragment from pLCW3-1 in pIJ4123	*S. lividans* 66 1326
pLCW4-1	1.3 kb PCR fragment by primers LC15 and LC16 (containing *midH* gene) from Smyc-LC3 in pUC18 *Sma*I	*E. coli* XL1-Blue
pLCW4-2	1.3 kb *Nde*I/*Bgl*II fragment from pLCW4-1 in pET11a *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW4-3	1.3 kb *Nde*I/*Bgl*II fragment from pLCW4-1 in pET16b *Nde*I/*Bam*HI	*E. coli* BL21(DE3) pLysS; XL1-Blue
pLCW4-4	1.7 kb *Xba*I/*Hin*dIII-blunt fragment from pLCW4-2 in pUC18 *Xba*I/*Sma*I	*E. coli* XL1-Blue
pLCW4-5	1.75 kb *Xba*I/*Hin*dIII-blunt fragment from pLCW4-3 in pUC18 *Xba*I/*Sma*I	*E. coli* XL1-Blue
pLCW4-6	1.7 kb *Eco*RI/*Hin*dIII fragment from pLCW4-4 in pUWL201	*E. coli* XL1-Blue *S. lividans* TK23
pLCW4-7	1.75 kb *Eco*RI/*Hin*dIII fragment from pLCW4-5 in pUWL201	*E. coli* XL1-Blue *S. lividans* TK23
pLCW4-10	1.3 kb *Nde*I/*Bgl*II fragment from pLCW4-1 in pIJ4123	*S. lividans* 66 1326

Tab. 2.7. New constructed plasmids for heterologous complementation

Plasmid	Description	Host strain
pRBS201	pJV1 ori from pAL201 instead of pIJ101 ori in pDNW16RBSY	*E. coli* XL1-blue
pLCW1-21	4.3 kb *Bam*HI fragment from pLC3-3 (containing *midHI* region) in pAL201 *Bam*HI	*E. coli* XL1-blue ET12567
pLCW1-25	1.3 kb PCR fragment by primers LC29 and LC19 (containing *midI* gene) from Smyc-LC3 in pUC18 *Sma*I	*E. coli* XL1-blue
pLCW1-26	1.3 kb *NcoI*/*Bam*HI fragment from pLC1-25 in pRBSY201 *Nco*I/*Bam*HI	*E. coli* XL1-blue ET12567
pLCW1-28	4.3 kb *Bam*HI fragment from pLC3-3 (containing *midHI* region) ligated with pLCW3-20 *Bam*HI (select the orientation and combine *midKHI* genes)	*E. coli* XL1-blue
pLCW1-29	5.7 kb *Hin*dIII/*Eco*RI fragment from pLCW1-28 in pAL201 *Hin*dIII/*Eco*RI	*E. coli* XL1-blue ET12567
pLCW3-15	1.4 kb *Bam*HI fragment from pLC3-6 in pAL201 *Bam*HI	*E. coli* XL1-blue ET12567
pLCW3-16	0.85 kb PCR fragment by primers LC24 and LC25 (containing 64 bp upstream sequence of *midK*, and *midK*) from Smyc-LC3 in pUC18 *Sma*I	*E. coli* XL1-blue
pLCW3-17	0.8 kb PCR fragment by primers LC28 and LC25 (containing *midK*) from Smyc-LC3 in pUC18 *Sma*I	*E. coli* XL1-blue
pLCW3-18	0.85 kb *Hin*dIII/*Bam*HI fragment from pLCW3-16 in pAL201 *Hin*dIII/*Bam*HI	*E. coli* XL1-blue ET12567
pLCW3-19	0.8 kb *Nco*I/*Bam*HI fragment from pLCW3-17 in pRBSY201 *Nco*I/*Bam*HI	*E. coli* XL1-blue ET12567
pLCW3-20	1.4 kb *Bam*HI fragment from pLC3-6 (containing *midK* region) in pSL1180 *Bam*HI/*Bg*lII (select the orientation of *Bam*HI/*Bgl*II site of recombinant plasmid located upstream *midK* gene)	*E. coli* XL1-blue

2.4. Oligonucleotides

Tab. 2.9. Oligonucleotides used in this study

Primer	Nucleotide sequence*	Restriction site	Gene
AS2	5´ GCCGCCGCGTCCCATGTCGAC 3´		*strE*
AS5	5´ CCCGTAGTTGTTGGAGCAGCGGGT 3´		*strE*
EryCIII/H	5´ CACGCGCGGCTGCTGTGGGGACCCGAC 3´		*eryCIII*
EryCIII/I	5´ CGCCGCGCAGGTCGGCAGCAGCGCGTG CAT 3´		*eryCIII*
LC1	5´ GCCGCCGAATCCCATGTGGAC 3´		*midB*
LC2	5´ CCCGTAGTTGTTGGAGCAGCGGGT 3´		*midB*
LC6	5´ GAGGAGCCCCGGGAGGACCCGGTCGCCGA 3´		*tylM2*
LC9	5´ GTCCATATGCGCGTCCTG 3´	*Nde*I	*midI*
LC10	5´ TCGGGATCCGCATGACTG 3´	*Bam*HI	*midI*
LC11	5´ACACATATGTACGCCAACG 3´	*Nde*I	*midK*
LC12	5´ CGGATCCGGTCAGTTGAA 3´	*Bam*HI	*midK*
LC13	5´ GAGCATCATATGAACGTGCCCTTTCC 3´	*Nde*I	*midC*
LC14	5´ CGGCGAAGATCTTTCCCTTCATTCC 3´	*Bgl*II	*mid*C
LC15	5´ GAGGAACATATGCCAATCCCTGCCA 3´	*Nde*I	*midH*
LC16	5´ ACTCCAGATCTGGCCGGGGTGTAC 3´	*Bgl*II	*midH*
LC17	5´ GCGCGACATATGGAGCATAAAGTGA 3´	*Nde*I	*mid*C
LC19	5´CATGGTGGATCCGCTCCGTTCGAACG 3´	*Bam*HI	*midI*
LC24	5´ ACGACAAGCTTGGTGCCGACC 3´	*Hin*dIII	*midK*
LC25	5´ AAGCCGGATCCCGAGCTCCC 3´	*Bam*HI	*midK*
LC28	5´ AAACACCCATGGACGCCAACG 3´	*Nco*I	*midK*
LC29	5´AGTCACCATGGGCGTCCTGCTGACCT 3´	*Nco*I	*midI*
LC30	5´ AGGGCAGCGCTGTCGTTG 3´	*Eco*47III	*midG*
LC31	5´ TCCTTCAGCGCTGGCTACGTC 3´	*Eco*47III	*midKH*

LC33	5´ GCGCATATGGACTCCAGTTCG 3´	*Nde*I	*midKH*
LC35	5´CGGAGATCTCCATCGCCG 3´	*Bgl*II	*midG*

The underlined sequences are the corresponding restriction site.

3. Methods

3.1. Cultivation and maintenance of bacterial strains

3.1.1. Cultivation and maintenance of *E. coli*

E. coli strains were generally cultivated at 37°C on LB plates or in LB liquid medium (Miller, 1972), except for some *E. coli* strains containing the expressed genes that were cultivated in LB liquid supplemented with 1 M sorbitol and 2.5 mM betaine (Blackwell *et al*., 1991). Strains harboring plasmids were grown on solid or in liquid media supplemented with the appropriate antibiotics. Unless otherwise stated, the final concentrations of used antibiotics were: ampicillin 80 μg/ml; chloramphenicol 25 μg/ml; hygromycin B 200 μg/ml. In the case of pUC-derived recombinant plasmids, 40 μg/ml X-gal was added into LB plates for the blue-white selection. Cultures were stored at –20°C

3.1.2. Cultivation and maintenance of *Streptomyces*

Wild type strains of *Streptomyces* sp. were grown on SMA plates and cultivated in TSB, YEME or SGYEME liquid media at 28-30°C for 3-5 d. For the cultivation of strains containing plasmids, a final concentration of thiostrepton in 25 μg/ml was added to agar plates and 15 μg/ml in liquid media. Kanamycin and hygromycin B were supplemented into solid and liquid media with concentrations of 50 μg/ml and 200 μg/ml, respectively.

Spore suspensions were prepared according to Hopwood *et al*. (1985) by adding 20% glycerol onto a well-sporulated plate and scraping off the spores from the surface of the plate. The suspension was then filtrated through sterile non-absorbant cotton wool to remove mycelial remnants. The spore suspensions were stored at –20°C.

3.2. *S. mycarofaciens* fermentation and product extraction

S. mycarofaciens was grown on ISP2 plates and sporulated at 28°C for 5 d. To produce a seed culture, 10 ml of EP1 medium (Salah-Bey *et al*., 1998) in a 10 × 22 cm glass tube with a short metal spring was inoculated with spore suspensions of *S. mycarofaciens*. Cultivation was carried out at 28°C for 3 d in a rotary shaker at 230 r.p.m. Then 4.2 ml of the seed culture were inoculated into 60 ml of EP2 medium (Salah-Bey *et al.*, 1998) in a 500 ml Erlenmeyer flask with a round metal spring. The production culture was grown for 5-7 d by shaking (230 r.p.m.) at 28°C.

The extraction of the fermenting broth of *S. mycarofaciens* was followed as in Salah-Bey *et al.* (1998). The broth was separated by centrifugation and the cells were washed once with water. The supernatant was adjusted to pH 9-10 with 1 M NaOH and extracted three times with an equal volume of ethyl acetate. The extracts were dried by evaporation. The residue was dissolved in an appropriate volume of dichloroform for assay.

3.3. General manipulation of DNA

3.3.1. Isolation of genomic DNA from *Streptomyces*

Chromosomal DNA was isolated from mycelia of *Streptomyces* sp. according to the protocol of Pospiech *et al.* (1995) with minor modifications. The following steps were used.

(1) Strains were inoculated in 25 ml of TSB liquid medium in 250 ml Erlemeyer flasks, each flask containing a short metal spring. The culture was heavily aerated by rotary shaking and grown at 28°C for 2 d to the late logarithmic phase.
(2) The mycelia were pelleted and washed twice with 10.3% sucrose.
(3) Approximately 0.5 g of cells were resuspended in 5 ml SET buffer (see Section 2.2.3), lysozyme was added to a concentration of 1 mg/ml and incubated at 37°C for 30-60 min.
(4) 1/10 volumes of 10% SDS and 0.5 mg/ml of proteinase K were added and incubated at 55°C with occasional inversion for 60 min.

(5) The lysates were extracted by the addition of 1/3 volumes of 5 M NaCl and 1 volume of chlorofrom and incubated at room temprature for 30 min with frequent inversion.
(6) The flocculant precipitate was removed by centrifugation at 4°C , 4500 rpm for 15 min. The aqueous phase was transferred to a new tube using a blunt-ended pipette tip.
(7) The chromosomal DNA was precipitated by adding 1 volume of isopropanol, centrifuged, washed with 70% ethanol, dried at room temperature and dissolved in 200 µl TE buffer with 2 µl of 10 mg/ml RNase.

In the case of isolation of chromosomal DNA from *S. mycarofaciens* for the preparation of the cosmid genomic library, in step (3) it was found to be better to incubate for 30 min with 2-3 gentle inversion. After step (6), the aqueous phase was extracted again by adding 1 volume chlorofrom and incubated as in step (5).

3.3.2. Preparation of plasmid DNA from *E. coli*

Plasmid DNA from *E. coli* was prepared via three methods depending on the different application of the DNA.

(1) Plasmid DNA to be used for cloning and sequencing was isolated from cells by using the Miniprep Plasmid Kit (Qiagen).
(2) The boiling preparation method was used for rapid screening of transformants (Sambrook *et al*., 1989).
(3) In this work , the alkaline lysis method (Birnboim & Doly, 1979) was especially adapted for the isolation of recombinant cosmid DNA. After the cells were lysed and neutralized by solution I, II and III (see Section 2.2.3.), the lysates were extracted twice with phenol/chloroform/isoamylalcohol (25:24:1) (Carl Roth, Karlruhe) and once with chloroform/isoamyl alcohol (24:1).

3.3.3. Preparation of plasmid DNA from *Streptomyces* sp.

Two methods were used to isolate plasmid DNA from *Streptomyces* mycelia.

(1) For generally checking the recombinant transformants, the Birnbiom & Doly (1979) method was adapted with minor modification. The cells were suspended in solution I,

then solution II was added in the presence of 5 mg/ml lysozyme and incubated at 37°C for 30 min.

(2) For the preparation of *Streptomyces* plasmid vectors, an alkaline lysis method (Hopwood *et al*., 1985) was introduced with modifications (this study). An aliquot of *Streptomyces* culture (1.5 ml) was harvested and the pellet was resuspended in 1 ml lysozyme solution (see Section 2.2.3) in a 2 ml Eppendorf tube. The suspension was incubated for 30 min at 37°C and gently mixed after the incubation. Then 500 µl of alkaline SDS solution (0.3 M NaOH, 2% SDS) was added and, the mixture was agitated by vortexing and incubated at 70°C for 15 min. After incubation, the mixture was placed at room temperature for 3 min. Then 160 µl of acid phenol/chloroform (5 g phenol, 5 ml chloroform and 1 ml H_2O) was added and mixed vigorously. The mixture was centrifuged for 2 min. The supernatant was transferred into a new Eppendorf tube containing 140 µl of 3 M sodium acetate and 1.4 ml of isopropanol. The suspension was mixed and left at room temperature for 5 min, and then centrifuged. The supernatant was decanted and the pellet was dried. The pellet was then re-dissolved in 100 µl TE buffer, and 10 µl of 3 M sodium acetate and 50 µl of neutral phenol/chloroform (saturated with TE buffer; Carl Roth, Karlsruhe) were added. The mixture was vortexed and centrifuged. The supernatant was transferred into a new Eppendorf tube and 100 µl isopropanol was added. The solution was mixed, left at room temperature for 10 min and centrifuged. The pellet was dried and re-dissolved in 50 µl TE buffer with addition of 2 µl of 10 mg/ml RNase.

3.3.4. *In vitro* manipulation of DNA

Restriction endonucleases for DNA hydrolysis, alkaline phosphatase for removing 5'-phosphate group from linearized DNA, Klenow fragment of DNA polymerase I for generation of blunt end at 3'- or 5'- overhang of linearized DNA, and T4-DNA ligase for ligation of DNA fragments were routinely used according to the manufacturer's specifications. For the first three enzyme reactions, the enzymes were inactivated by incubation for 10 minutes at 70°C. The resulting DNA fragments were isolated either by phenol/chloroform extraction or by agarose gel separation.

3.3.5. Polymerase chain reactions (PCR)

The PCR amplification was done according to Innis *et al*. (1992). In this study, two polymerases (Vent- and *Taq*-) were used depend of the further uses of genes found by PCR. Therefore, two PCR reactions were performed as follows:

(1)	Vent polymerase buffer (10 ×)	10	µl
	dNTP's	200	µM
	DMSO	10	µl
	$MgSO_4$ (100 mM)	1	µl
	Forward primer	50	pmol
	Reverse primer	50	pmol
	DNA template	10-100	ng
	Vent-DNA polymerase	2.5	U
	H_2O	to 100	µl
	Mineral oil	50	µl
(2)	*Taq* polymerase buffer (10 ×)	10	µl
	dNTP's	200	µM
	DMSO	5	µl
	$MgCl_2$ (50 mM)	3	µl
	Forward primer	50	pmol
	Reverse primer	50	pmol
	DNA template	10-100	ng
	Taq-DNA polymerase	2.5	U
	H_2O	to 100	µl
	Mineral oil	50	µl

The amplification profile was performed in two ways.

(1) Usually the amplification was done as follows. After an initial hot start at 98°C for 2 min, the polymerase was immediately added into the above mixture tubes. The mixture was subjected to 30 cycles of amplification, each consisting of 1 min at 95°C, 40 s at an annealing temperature (variable) and 1-2 min at 72°C. The thermal cycle was

(2) To detect the genes containing high GC contents from chromosomal DNA, 6 special cycles were introduced before starting the 30 cycles of method (1). The first cycle was 1 min at 95°C, 40 s at an annealing temperature of T°C, and 1-2 min at 72°C. In every subsequent cycle the annealing temperature was decreased by 1°C until the sixth cycle

at (T-5)°C. Then the annealing temperature was returned to T°C, performing amplication as (1).

The PCR products were analysed and purified by agarose gels or using the NucleoSpin Extract kit.

3.3.6. Separation of DNA fragment

Agarose gel electrophoresis was performed essentially as described by Sambrook *et al*., 1989. DNA fragments were separated on horizontal 0.7-1.5% agrose gels containing 0.1 μg/ml ethidium bromide. One tenth volume of 10 × loading buffer was added to the DNA solution. The agarose was dissolved in 1 × TAE buffer and electrophoresis was carried out at 5-10 V/cm. DNA was visualized by illumination with long wavelength UV-light (366 nm).

The separation of DNA fragments for sequencing was carried out on a vertical denaturing PAA gel. The gel contained 5.5% polyacrylamide (the FMC Long-Ranger Solution, Biozym, Hameln), 7 M urea in 1 × TBE buffer. Then 350 μl APS (10%) and 42 μl TEMED were added into 80 ml above solution to start the polymerization. The gel solution was sandwiched between the glass plates (50 × 50 × 0.05 cm) and polymerized for 2 h. The gel was run with 1 × TBE buffer at 45°C at 1500-2000 V in an A.L.F. DNA Sequencer (Amersham).

Sucrose gradients were used for the size fractionation of DNA fragments for the cosmid genomic library (see Section 3.4.1.2).

3.3.7. Radioactive labelling of DNA fragments

Double strand DNA fragments were labelled with [^{32}P] using the *Redi*prime DNA labelling system (Amersham) according to the manufacturer's specifications. In general, 5 μl of α-[^{32}P]-dCTP with a specific activity of 3000 Ci/mmol were used in the labelling reaction.

3.3.8. DNA-DNA hybridization

The DNA fragments were separated on 0.8-1.0% agarose gels and transferred by capillary stock onto a Hybond N^+ membrane (Amersham) using 0.4 M NaOH for 12 h (Reed & Mann, 1985). The membranes were shortly washed with 2 × SSC, air-dried and fixed at 80°C for 10 min or at UV light (360 nm) for 3 min. Membranes were prehybridized in 200-400 ml prehybridization solution in a shaking water bath at 68°C for 3-4 h. The [^{32}P] labelled DNA probe was denatured by boiling for 10 min and immediately placed on ice. Membranes were transferred to 100 ml hybridization solution and the denatured radioactive probe was then added. The hybridization was carried out at 68°C for 12 h. A series of washing steps followed. Membranes were washed twice with 6 × SSC / 0.1% SDS, each for 15 min. Stringency washes were carried out with 0.5-2 × SSC / 0.1% SDS at 68°C for 10 min with gentle shaking.

The cosmid colony hybridization for screening the gene library was performed as described in Section 3.4.3.3.

3.3.9. DNA sequencing

Sequencing of double-stranded DNA was performed using the dideoxy nucleotide chain termination method (Sanger *et al*., 1977). 6-8 μg of pUC18-derived plasmid DNA, mp13 (universal) and pUC18 (reverse) primers labelled with fluorescent dye, and the Thermosequenase Cycle-Sequencing Kit were used in a sequencing reaction according to the manufacturer's specifications in an A.L.F. DNA sequencer.

3.3.10. Determination of DNA concentration

The concentration of double-stranded DNA was photometrically quantified at a wavelength of 260 nm as described by Sambrook *et al*. (1989). An OD_{260} of 1.0 is equivalent to approximately 50 μg/ml of double-stranded DNA. The value of OD in the DNA solution was also measured at 280 nm to examine the contamination with phenol or proteins. The ratio of OD_{260} to OD_{280} in pure DNA solution should be in a range of 1.8-2.0.

3.3.11. Transformation of *E. coli* and *S. lividans*

Competent cells of *E. coli* strains were prepared as described by Hanahan (1983). The transformation of *E. coli* cells was followed according to the protocol of Sambrook *et al.* (1989). In general, 0.01-1 μg DNA was added to 100 μl competent cells, SOC or LB medium was used to regenerate the cells.

Protoplasting and transformation of *Streptomyces* were carried out based on the method of Babcock and Kendrick (1988). TSB-PEG 8000 medium was used to cultivate the cells. PEG 1000 was used to mediate the transformation.

3.3.12. Heterologous gene expression in *E. coli* BL21 (DE3) pLysS

The method used was based on the description of Studier *et al*. (1990). A single colony of *E. coli* BL21 (DE3) pLysS harboring a pET-derived plasmid was inoculated in 3 ml of LB medium containing 50 μg/ml ampicillin and 25 μg/ml chloramphenicol and the cells were grown overnight at 37°C. The overnight culture (0.5 ml) was re-inoculated in 50 ml of LB medium (or LB supplemented with 1M sorbitol and 2.5 mM betaine) with the same antibiotics, and the cells were grown at 28°C in a rotary shaker (100 r.p.m.) to an OD_{600nm} of 0.5-0.7. Then the final concentration of 0.4 mM IPTG was added to the culture to induce the T7-RNA polymerase production. In LB culture, the other 4 h cultivation was done after induction, and 12 h was required in the culture of LB containing sorbitol and betaine. Samples (1 ml) were taken prior to induction and at intervals after induction. The cells were harvested by centrifugation and re-suspended in 50–100 μl of the loading buffer. The proteins were subjected to electrophoresis by SDS-PAGE gel (see Section 3.5.3). The remaining cultivated cells were extracted as described in Section 3.5.1.

3.3.13. Heterologous gene expression in *S. lividans* 1326

The expression of the specific genes inserted in pIJ4123 in *S. lividans* 1326 was followed as the method of Takano *et al*. (1995) but using SGYEME medium. A single colony was

inoculated into 10 ml of SGYEME medium with 50 μg/ml kanamycin and the cells was grown at 28°C for 2 d as pre-culture. The pre-culture (0.5 ml) was then inoculated in 50 ml of SGYEME medium containing the same antibiotic. After 12 h of cultivation, 10 μg/ml thiostrepton was added to induce *tipA*p promoter. The cultivation was continued for another 24-36 h. The cells were harvested and extracted as described in section 3.5.1.

3.4. Manipulation of the cosmid genomic library of *S. mycarofaciens*

3.4.1. Construction of the cosmid genomic library

The genomic library of *S. mycarofaciens* UC189B (ATCC 21454) was constructed in this work by using the pKU206 vector (Omura, personal communication; Kakinuma *et al*., 1991) (Fig. 3.1) and *S. mycarofaciens* chromosomal DNA. The following manipulation was based on Weis *et al.* (1987) with modifications.

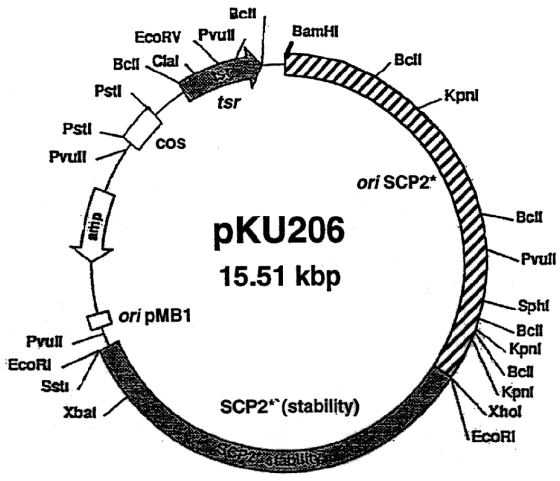


Fig. 3.1. The bifunctional cosmid vector pKU206.

3.4.1.1. Partial digestion of *S. mycarofaciens* genomic DNA

The chromosomal DNA from *S. mycarofaciens* was prepared as described in Section 3.3.1. 0.3% agarose gel (using 0.5 cm depth of 1% agarose gel as a supporter under the bottom of the gel lid) was used to examine the size of DNA (Little, 1987). The molecular length of DNA must be larger than λ DNA.

Partial digestion of the genomic DNA with *Sau*3AI was done in a Eppendorf tube, including 215 µl of DNA (~20 µg), 25 µl of 10 × React 4 buffer (Gibco BRL) and 10 µl diluted *Sau*3AI (1U) and mixed carefully. The digestion was preformed at 37°C, and a 50 µl aliquot was removed at 5 min, 10 min, 15 min, 20 min, and 25 min and inactivated at 68°C for 10 min. Samples were combined when they showed the DNA in the size range of 25-45 kb by electrophoresis through 0.5% agarose gel.

3.4.1.2. Size fractionation of DNA fragments

Sucrose gradients were used for size fractionation of DNA molecules (Hadfield, 1987). The new SW-41 ultracentrifuge tubes (Beckman, München) were filled with 10 ml of sucrose gradient solution (see Section 2.2.3), sealed with Parafilm and frozen overnight at -20°C. The frozen tubes were thawed at room temperature for 4-5 h. The freeze-thaw process was repeated once to form the gradient.

The combined partially digested DNA was heated at 65°C for 5 min in order to dissociate any DNA aggregates. 166 µl of the DNA solution was carefully layered on the top of the sucrose gradient. The tubes were then centrifuged in an SW-28 rotor (Beckman) at 20°C, 25,000 r.p.m., for 18 h. After centrifugation, the bottom of the tubes were gently cleaned with ethanol. The gradients were fractionated by carefully piercing the bottom of the tube with a sterile needle. The small hole was rapidly blocked with the thumb and drops of the gradient solution were allowed by slightly moving the thumb. Samples (each 0.2-0.3 ml) were dropped into the new Eppendorf tubes and around 50 aliquots were collected from 10 ml of the sucrose gradient. Samples (15 µl) from every fifth aliquot were taken and checked by electrophoresis on 0.5% gels, using the DNA marker diluted by the sucrose gradient solution because of the high salt concentration in the solution affects DNA mobility.

The correctly sized DNA (25-35 kb) was precipitated by adding two volumes of ethanol. To avoid precipitation of sucrose, a further 1.5 ml of 70% ethanol was added and mixed thoroughly and placed at –20°C overnight. The DNA was pelleted by centrifugation, and the pellets were washed twice with 70% ethanol and re-suspended in an appropriate TE

buffer. The aliquots of the DNA solution were combined to give a final concentration of 0.3-0.5 μg/μl.

3.4.1.3. Digestion of the cosmid vector pKU206

The cosmid vector pKU206 contains a *Bam*HI site cloning site, (Kakinuma *et al*., 1991). Hydrolysis of the vector was performed including 200 μl of pKU206 DNA (~20 μg), 1 × React 3 buffer (Life Technologie), 100 U of *Bam*HI enzyme and distilled water in a 400 μl volume. The incubation was carried out at 37°C for 3-4 h. An aliquot (3 μl) was removed and analyzed by electrophoresis through a 0.6% agarose gel and compared to undigested pKU206. The completely hydrolised sample was then extracted with phenol/chloroform and the DNA was precipitated with ethanol. The DNA was re-suspended in 200 μl of TE buffer. A 5 μl aliquot was taken and stored at –20°C.

The remainder of the sample was treated with 50 units (ATP hydrolysis units) of calf intestinal phosphatase by an incubation at 37°C for 30 min. The dephosphorylation was terminated by adding 1 μl of 0.5 M EDTA. The DNA was extracted twice with phenol and once with chloroform and precipitated with ethanol. The pellet DNA was re-suspended in TE buffer to give a final concentration of 0.3-0.5 μg/μl. A test re-ligation was carried out to determine the effectiveness of the phosphatase treatment of the linearized vector. The reaction was set up in a 10 μl volume including *Bam*HI-cleaved pKU206 either phosphatase-treated DNA or untreated (i.e. the above aliquot stored at –20°C) and using a standard ligation buffer and conditions (see below). The test ligations were examined using agarose gels. The phosphatase-treated sample should not be re-ligated.

3.4.1.4. Ligation of insert DNA to vector DNA

The following reaction mixture was set up in a 20 μl volume:

25-35 kb fragments of *S. mycarofaciens*	5 μl (~2.5 μg)
phosphatase-treated pKU206	10 μl (~5 μg)
10 × ligation buffer* (Roche)	2 μl
H_2O	3 μl

* the ligation buffer should be PEG-free quality.

An aliquot (1 μl) was removed and stored at 4°C. Then 1 μl (5 Weiss units) of T4 DNA ligase (Roche) was added to the remainder of the mixture. The reaction was incubated overnight at 12°C. At the end of the ligation, another 1 μl aliquot was taken and analyzed by electrophoresis using 0.5% agarose gel together with the above aliquot stored at 4°C. If the ligation was successful, the *S. mycarofaciens* DNA should have been converted to high-molecular-weight concatemers.

3.4.1.5. Packaging of ligated DNA

The ligated DNA was packaged *in vitro* into bacteriophage λ particles following the protocol of DNA Packaging Kit (Roche). 4 μl of the ligation mixture was used, and the rest of the ligation was stored at –20°C. After packaging was complete, 0.5 ml of SM buffer was added to the reaction, and this packaging reaction was then mixed with 20 μl of chloroform and stored at 4°C for two weeks.

3.4.1.6. Adsorption of packaged recombinant cosmids to *E. coli*

The host strain *E. coli* JM108 was prepared as recommended by the protocol of DNA Packaging Kit (Roche). 10 μl aliquot of the packaging reaction was mixed with 0.1 ml of SM buffer and 0.2 ml of the host strain. The mixture was incubated for 30 min at 30°C to allow the adsorption of the bacteriophage particles to the *E. coli* cells. Then 1 ml of SOC medium was added and the incubation was continued for a further 60 min.

3.4.1.7. Plating of packaged recombinant cosmids

Samples (0.1 ml and 0.5 ml) of the bacterial culture were spread onto LB agar plates containing 50 μg/ml of ampicillin. After incubating the plates at 30°C for 14-16 h, the number of bacterial colonies was counted. Usually each microgram of ligated DNA should yield at least 5×10^4 bacterial colonies.

A number of individual colonies were picked and grown in 4 ml of overnight LB culture at 37°C. The plasmid DNA was isolated by using the alkaline lysis method (Birnboim & Doly, 1979). The plasmid DNA was digested with *Bam*HI or other restriction enzyme(s) and the size of the resulting fragments was analysed by gel electrophoresis.

3.4.2. Plating out, amplification and maintenance of the cosmid genomic library

This work was done with reference to Sambrook *et al*. (1989) and Weis *et al*. (1987).

3.4.2.1. Plating on agar for the library amplification

200 µl of the cells prepared as in Section 3.4.1.7 was spread on a LB plate containing 50 µg/ml ampicillin. The plates were incubated overnight at 30°C and usually over 2000 colonies were produced on a plate. Then 2 ml of LB liquid containing 15% glycerol was added to the grown-colony plates. The colony suspension was carefully mixed, scraped, pooled and stored at –70°C.

3.4.2.2. Picking and streaking colonies for screening and maintenance of the library

100 µl of the cells prepared as in Section 3.4.1.7 was spread on a LB plate containing 50 µg/ml ampicillin. The plates were incubated overnight at 30°C and usually over 1000 colonies were produced on a plate. In order to absolutely ensure the authentic result of screening the library, ten thousand colonies were picked and streaked one by one onto new LB plates containing 50 µg/ml ampicillin, in a matrix of 47 numbered points (each point size ca. 2 mm). Each colony was picked and streaked onto two plates in the same position, which one plate was used for screening the library and the other for maintenance of the library.

Microtiter plates were used to store the cosmid colony cultures. A 96-well microtiter plate was labelled and divided into two parts which were used for transferring two plates of colonies. 300 µl of LB liquid containing 15% glycerol were first titred into each well. A plate containing 47 colonies grown overnight was gently duplicated using a sterile stainless steel stamp containing 47 feet which is exactly the same as the matrix mentioned above.

The stamp with the adhering bacteria was then dipped into the LB and glycerol medium at the correct position in the microtiter plate and gently shaked for 1 min. The plates were incubated overnight at 30°C and stored at –20°C.

3.4.3. Screening of the cosmid genomic library

The following experiments were based on the protocols described as Hanahan & Meselson (1983) and Weis *et al*. (1987) with modifications.

3.4.3.1. Preparing duplicate membranes of cosmid colonies

Bacterial plates each containing 47 colonies grown overnight were first stored at 9°C for 1 h. Numbered disc membranes were then placed on the surface of the colony plates for 1 min, peeled off and air-dried for 10 min. The master plates were incubated for 3-4 h and stored at 4°C.

3.4.3.2. Cosmid DNA binding to the membranes

A Whatman 3MM filter paper in a tray was soaked in 0.5 M sodium hydroxide. The disc membranes, colony side uppermost , were placed onto this wet filter and lysed for 5 min to denature the DNA. This step was repeated with a new 3MM paper with 0.5 M NaOH. The membranes were then transferred onto a 3MM paper soaked with a solution of 1 M Tris/HCl, pH 7.5 for neutralization. After 3 min the membranes were neutralized once again by placing them onto a new 3 MM paper with the same solution. The membranes were then washed with 2 × SSC solution for 5 min and air-dried. The dried membranes were baked at 80°C for 90 min.

3.4.3.3. Screening the cosmid membranes

The baked membranes were placed in a plastic box with 300 ml of prehybridization. The box was shaked slowly in a water bath at 65°C. After 1.5 h, the bacterial debris on the membranes were wiped using soft tissues and the new prehybridization solution was added into the box. The process was continued for another 1.5 h. The probe DNA was labelled as

described in Section 3.3.7. The conditions of hybridization for the disc membranes were the same as described in Section 3.3.8. After hybridization, the membranes were washed at 65°C twice with 2 × SSC/0.5% SDS, each time for 10 min, twice with 0.5 × SSC/0.5% SDS, each time for 15 min, and once with 0.1 × SSC/0.5% SDS for 15 min. The autoradiography of the membranes was followed by the standard method (Sambrook *et al*., 1989). The positive spots were identified by comparing precisely the same position on the membranes and the master plates. Finally the positive cosmid clones were picked from the master plates and the plasmid DNA was prepared for further genetical analysis.

3.5. Biochemical methods

3.5.1. Cell-free extracts of *E. coli* and *Streptomyces*

The *E. coli* or *Streptomyces* cells containing the expression-plasmids were harvested by centrifugation and washed twice with a cold buffer of 25 mM Tris/HCl, pH 7.5 or 25 mM potassium phosphate buffer, pH 7.2. The pellet was suspended with an appropriate sonification buffer (see Section 2.2.3) according to the further use of the extracts in a ratio of 1.0 ml buffer/0.2 g for *E. coli* cells and 1.5 ml /g for *Streptomyces*. The cells were disrupted by treating 2 times for *E. coli* and 4 times for *Streptomyces* (each time for 20 sec following 20 sec break) with the French-Press under a pressure of 1300 psi. The cell suspension was centrifuged for 30 min at 30 000 *g* and the crude cell-free extracts were stored at –80°C for 1-2 weeks.

3.5.2. Determination of protein concentration

The protein concentration of the cell-free extracts was measured according to Bradford (1976) using Bio-Rad Protein Assay Kit (Bio-Rad) and BSA as a standard at a wavelength of 595 nm.

3.5.3. SDS polyacrylamide gel electrophoresis (PAGE)

SDS-PAGE was performed according to Laemmli (1970) using 10-12% of polyacrylamide in vertical slab gels. The protein standard VII-L (Sigma), BSA (Mr = 66 kDa), ovalbumin

(Mr = 45 kDa), glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle (Mr = 36 kDa), carbonic anhydrase from bovine erythrocytes (Mr = 29 kDa), trypsinogen from bovine pancreas (Mr = 24 kDa), trypsin inhibitor from soybean (Mr = 20 kDa) and α-lactalbumin (Mr = 14 kDa), was applied to the gels in order to determine the Mr of the proteins.

3.5.4. Western blotting and immuno-detection of proteins

The proteins produced in *E. coli* and *Streptomyces* were separated by SDS-PAGE. When the expected protein was not detected on the gels, Western blotting was used for the detection. The protein gels were transferred onto PVDF membrane ('Hybond P', Amersham) using the Semi-Dry Transfer Cell (Bio-Rad). The Western blotting was followed by a standard protocol (Winston *et al*., 1987) and the method of immuno-detection of proteins was adapted here as recommended by Qiagen's protocol using an Anti-His-tag antibody.

3.5.5. Analysis of the activity of dTDP-D-glucose 4, 6-dehydratase by RmlB

The activity of dTDP-D-glucose 4,6-dehydratase was determined according to Wang & Gabriel (1969) and Okazaki *et al.* (1962) with minor modifications. The *rmlB* gene (formerly called as *rfbB*) from *Salmonella enterica* B was cloned on the plasmid pSVW701 by Verseck (1997) and overexpressed in *E. coli* BL21(DE3). The reaction for analysis of RlmB protein was performed as follows.

Potassium phosphate buffer, pH 7.2	50 mM
dTDP-D-glucose	4 mM
RlmB crude extract	0.1-0.2 mg/ml
Final volume	200 µl

The mixture was incubated at 37°C for 1 h. The formation of dTDP-4-keto-6-deoxy-glucose was determined by taking 40 µl of the reaction sample and adding 960 µl of 0.1 N NaOH and the mixture was re-incubated at 37°C for 20 min. The absorption was measured

at 320 nm (ε_{320nm} = 4800 l/mol × cm). Blank experiments were carried out in the same way but omitting the substrate.

3.5.6. Analysis of the activity of enzymes participating in dTDP-D-mycaminose biosynthesis via a coupling reaction of RmlB, MidH and MidC

The genes encoded in the biosynthetic pathway of dTDP-D-mycaminose in macrolide antibiotics were proposed as described as Gandecha *et al*. (1997). In this study, the determination of the activities of the 3,4-isomerase (MidH) and the aminotransferase (MidC) was performed with reference to the methods of Walker & Walker (1969). The conversion of dTDP-glucose to dTDP-D-mycaninose by the coupling reaction of RlmB, MidH and MidC was carried out as follows.

Potassium phosphate buffer, pH 7.2	50 mM
dTDP-glucose	4 mM
$MgCl_2$	10 mM
L-Alanine or L-Glutamic acid	2.5 mM
PLP	1 mM
RlmB crude extract	0.1-0.2 mg/ml
MidH crude extract	0.2-0.3 mg/ml
MidC crude extract	0.2-0.3 mg/ml

The mixture was incubated at 37°C for 1 h and the reaction was terminated by heating at 95°C for 2 min. The mixture was centrifuged at 4°C for 20 min to remove the debris and the supernatant was stored at –20°C for the product assay.

3.6. Chromatography methods

3.6.1. Thin-layer chromatography (TLC)

Antibiotic metabolites from cultures of *Streptomyces* were preliminarily identified by TLC as described by Salah-Bey *et al.* (1998). The supernatants of cultures were adjusted to pH 9.0-10.0 with sodium hydroxide and extracted with ethyl acetate. Chromatogarphy was

performed using dichloromethane-methanol (90/10, v/v). The compounds were visualized by spraying the TLC plates with *p*-anisaldehyde/sulphuric acid/ethanol (1/1/9, v/v) with subsequent heating for a few minutes at 80°C.

3.6.2. High pressure liquid chromatography (HPLC)

Analysis of nucleotide-activated sugars from the enzymatic reaction was performed by HPLC according to Payne & Ames (1982) with modifications. A reversed phase chromatography was used as separation system equipped with a Eurospher 100 C18 (corn: 5 μ, 250 × 4.6 mm) (Knauer, Berlin). The mobile phase and operation of this system were recommended by Verseck (1997).

3.6.3. HPLC coupled to mass spectrometry (LC-MS)

HPLC coupled to mass spectrometry (LC-MS) was carried out on a chromatography system (Waters) equipped with a Finnigan TSQ 7000 mass spectrometer to identify the novel products. This assay was performed at the University of Bochum by Dr. Blüggel.

3.7. Computer programs

DNA sequences were analysed using the program DNA-Strider 1.1 (Mark, 1988). The programs Brugene II and MacFrame 1.3 were used to determine open reading frames and G+C contents. Multiple sequence alignments were performed using the Clustal V 3.0 program (Higgins, 1991). Comparisons of DNA- and protein-sequences to GenBank were done using FASTA (Pearson & Lipman, 1988) and BLAST (Altschul, 1990). Phylogenetic trees were construcetd via the neighbor joining method (Saitou & Nei, 1987) using the PAUP 3.1 program (D.L. Swofford, Center of Biodiversity, Champaign, III., USA). Amplification temperatures for primers were calculated by the PrimFind V3.0 program (Fröbel Labor-Geräte, Lindau).

4. Results

4.1. Detection of pathway-specific genes from *S. mycarofaciens* by PCR

As mentioned in the introduction, most of the 6-deoxysugar biosynthetic genes are clustered on both sides of a large portion of polyketide synthases (PKSs) in the macrolide antibiotic biosynthesis, such as erythromycin (Gaisser *et al*., 1997; 1998; Salah-Bey *et al*., 1998), oleandomycin (Olano *et al*., 1998), tylosin (Merson-Davies *et al*., 1994; Gandecha *et al*., 1997; Fouces *et al*., 1999) and so on. Experimental evidence has suggested that dTDP-4-keto-6-deoxyglucose is the precusor of most of the deoxysugars, for instance, found from macrolide-producing strains (Liu & Thorson, 1994; Piepersberg, 1994) as well as from anthracycline-producing strains (Thompson *et al*., 1992). Thus, because of the similarity to tylosin, the detection of the gene encoding dTDP-glucose 4,6-dehydratase was hoped to be the first and rational choice in order to probe the midecamycin biosynthetic (*mid*) gene cluster from *S. mycarofaciens*. Based on mechanisms and pathways from deoxysugar biosynthesis research, a glycosyltransferase is responsible for transferring the sugar to the antibiotic aglycone when a D- or L-6-deoxysugar is formed (Otten *et al*., 1995; Liu *et al*., 1993). Since midecamycin is structurally closely related to tylosin, the organisation of the gene cluster of tylosin biosnythesis (acc. no. U08223; Merson-Davies *et al*., 1994; acc. no. X81885; Gandecha *et al*., 1997) gives a hint that at least some of the glycosyltransferase genes are located in the gene cluster on the side opposite to the dTDP-glucose 4,6-dehydratase gene. Therefore, as a prediction, the detection of a glycosyltransferase gene is the second choice in order to probe the gene cluster of midecamycin biosynthesis in *S. mycarofaciens*.

In this study, PCR amplifications were applied to directly detect the genes encoding the biosynthetic pathway of midecamycin from *S. mycarofaciens* UC189B (ATCC21454). The identified genes were then used as homologous probes to hybridize the cosmid library of *S. mycarofaciens* in order to identify the *mid* gene cluster.

4.1.1. Identification and isolation of a gene, *midB*, encoding a dTDP-glucose 4, 6-dehydratase from *S. mycarofaciens*

To identify the specific gene in midecamycin biosynthesis, the genomic DNA of *S. mycarofaciens* was amplified by PCR using two oligodeoxynucleotide primers, AS2 and AS5 (Stratmann, 1997), which were derived from the *strE* gene, encoding dTDP-glucose 4,6-dehydratase in the gene cluster of streptomycin biosynthesis from *S. griseus* (Pissowotzki *et al*., 1991) and showed the highly conserved sequences in the other dTDP-glucose 4,6-dehydratase genes, such as *tylAII* from *S. fradiae* (Merson-Davies *et al*., 1994) and *rfbB* from *Salmonella typhimurium* LT2 (Jiang *et al.,* 1991). As a result of PCR amplification, a 303 bp DNA fragment from *S. mycarofaciens* genome was detected, and the gel-purified DNA fragment was cloned into the *Sma*I site of pUC18 (Arnold, 1996). This recombinant plasmid was named pLCW1, and the gene in the corresponding insert DNA was designated *midB*. Sequencing analysis indicated that *midB* encodes a dTDP-glucose 4,6-dehydratase and the product of *midB* in a 101 amino acid (aa) overlap is 58% identical to the products of *strE* and *tylAII*, respectively, and 44% identical to the product of *rfbB*. The sequence comparison between these proteins is shown in Fig. 4.1.

4.1.2. Identification and isolation of a gene, *midI*, encoding a glycosyltransferase from *S. mycarofaciens*

The oligodeoxynucleotide primers, EryCIII/H, EryCIII/I and LC6, were derived from the highly conserved sequence regions of macrolide glycosyltransferse genes, *eryCIII* and *tylMII* in the gene clusters of erythromycin and tylosin biosynthesis (Salah-Bey *et al*., 1998; Gandecha *et al*., 1997). The genomic DNA of *S. mycarofaciens* was first amplified using primers EryCIII/H and EryCIII/I by PCR method 2 (see Section 3.3.5). The reaction was carried out at an annealing temperature of 63°C and for an enlongation time of 60 sec. An expected DNA fragment (500 bp) was observed in the PCR products by the use of 1.5% agarose gel. However, there existed unspecific bands mixed with or close to this expected band, which caused difficulties for purification and cloning. Therefore, a pair of internal primers, LC6 and EryCIII/I, were used to re-amplify the above product mixture by using PCR method 1 (see Section 3.3.5) at an annealing temperature of 65°C and for an enlongation time of 45 sec. As expected, a 420 bp DNA fragment was detected, and the resulting fragment was cloned into the *Hin*cII site of pUC18. This recombinant plasmid was named pLCW17, and the gene in the corresponding insert DNA was designated *midI*.

Sequencing analysis confirmed that *midI* encodes a glycosyltransferase. The comparison of the *midI* product to other known glycosyltransferases revealed that MidI in a 142 amino acid overlap is 56% identical to TylMII and 53% identical to EryCIII, respectively (for sequence alignment see Section 4.15).

```
TylAII          ----MRVLVTGGAGFIGSHFTGQLLT-GAYPDLGATRTVVLDKLTYAGNPANLEVA-GHP
StrE            MALTTHLLVTGAAGFIGSQYVRTLLGPGGPPDVVVT---ALDALTYAGNPDNLAAVRGHP
RfbB            ----VKILITGGAGFIGSAVVRHIIK--NTQDTVVN----IDKLTYAGNLESLSDISESN
MidB
                                                AS2
                                              ------>
TylAII          DLEFVRGDIADHGWWRRLMEGVGL--VVHFAAESHVDRSIESSEAFVRTNVEGTRVLLQA
StrE            RYRFERGDICDAPG-RRVMAGQDQ--VVHLAAESHVDRSLLDASVFVRTNVHGTQTLLDA
RfbB            RYNFEHADICDSAEITRIFEQYQPDAVMHLAAESHVDRSITGPAAFIETNIVGTYALLEV
MidB                                        AAESHVDRSIDDADAFVRTNYLGTHVLLTE
                                            *********      *  **  **  **

TylAII          AVDAG-VG--------RFVHISTDEVYGSIAE----------GSWPEDHPVAPNSPYAAT
StrE            ATRHG-VA--------SFVQVSTDEVYGSLEH----------GSWTEDEPLRPNSPYSAS
RfbB            ARKYWSALGEDKKNNFRFHHISTDEVYGDLPHPDEVENSVTLPLFTETTAYAPSSPYSAS
MidB            ALAVRRPG--------RFVHVSTDEVYGSIPE----------GSWSEDHPLSPNSPYAAS
                *                *   *******                 *     * *** *

                                    AS5
                                  <------
TylAII          KAASDLLALAYHRTYGLDVRVTRCSNNYGPRQYPEKAVPLFTTNLLDGLPVPLYGDGGNT
StrE            KASGDLLALAHHVSHGLDVRVTRCSNNYGPRQFPEKLIPRFITLLMDGHRVPLYGDGLNV
RfbB            KASSDHLVRAWRRTYGLPTIVTNCSNNYGPYHFPEKLIPLVILNALEGKPLPIYGKGDQI
MidB            KAASDQLALAFHRTHGLPVCVTRCSNNYG
                **  * *  *     **   ** ******

TylAII          REWLHVDDHCRGVALVGAGGRPGVIYNIGGG-------TELTNAELTDRILELCGADRSA
StrE            REWLHVDDHVRGIEAVRTRGRAGRVYNIGGG-------ATLSNKELVGLLLEAAGADWGS
RfbB            RDWLYVEDHARALHMVVTEGKAGETYNIGGHNEKKNLDVVFTICDLLDEIVPKATSYREQ
MidB

TylAII          LRRVADRPGHDRRYSVDTTKIREELGYAPRTGITEGLAGTVAWYRDNRAWWEPLKRSPGG
StrE            VEYVEDRKGHDRRYAVDSTRIQRELGFAPAVDLADGLAATVAWYHKHRSWWEPLVPAGSL
RfbB            ITYVADRPGHDRRYAIDAGKISRELGWKPLETFESGIRKTVEWYLANTQWVNNVKSGAYQ
MidB

TylAII          RELERA-----
StrE            PA---------
RfbB            SWIEQNYEGRQ
MidB
```

Fig. 4.1. Amino acid sequence comparison of MidB with other dTDP-glucose dehydratases from different bacteria. The proteins TylAII from *S. fradiae*, StrE from *S. griseus* and RfbB from *Salmonella typhimurium* LT2 are aligned with MidB. The highly conserved sequence regions used for designing the primers, AS2 and AS5, are marked and indicated by *arrows*. Sequence identities between these proteins are indicated by *asterisks*.

4.2. Construction and screening of a genomic library from *S. mycarofaciens*

A genomic library of *S. mycarofaciens* UC189B (ATCC 21454) was constructed during this work by using the cosmid vector pKU206 and *S. mycarofaciens* chromosomal DNA. The DNA fragments of *midB* and *midI* obtained by heterologous PCR (see Section 4.1.1-2) were used as probes to sequentially screen the cosmid library.

4.2.1. Construction of the genomic cosmid library

To clone large DNA fragments (20-35 kb) in *E. coli* strains, and to be able at the same time to directly introduce recombinant plasmids into *Streptomyces* strains, a *Sreptomyces-E. coli* bifunctional cosmid vector pKU206 (15.5 kb) was used (Omura, personal communication) (see Fig. 3.1). The shuttle cosmid pKU206 was derived from pKU205 (Kakinuma *et al*., 1991), containing the replication regions SCP2* and pMB1 and the lambda *cos* region, by introduction of the 5 kb SCP2* stability region. Therefore, it was expected that the vector and its recombinant derivatives should be relatively stable in *S. lividans*. The steps taken to construct the cosmid library from the genomic DNA of *S. mycarofaciens* are shown in Fig. 4.2.

4.2.1.1. Selection of 20-35 kb DNA fragments in *S. mycarofaciens* partial digestion

The genomic DNA of *S. mycarofaciens* was prepared as described in Section 3.3.1. Since maximally non-sheared DNA fragments (usually in the size of 200-500 kb) are required for obtaining a statistically useful cosmid library, the chromosomal DNA was tested for its size before digestion. It is shown in Fig. 4.3 that the prepared *S. mycarofaciens* DNA (*track* 3) was much larger than the λ DNA (48 kb) (*track* 1).

The partial digestion of the genomic DNA of *S. mycarofaciens* was tested first in two trial digestions in order to establish the optimal conditions to employ (see Section 3.4.1.1). About 20 μg of the genomic DNA was used for partial digestion and for the subsequent size fractionation. The digestion was controlled by time limitation during an incubation at

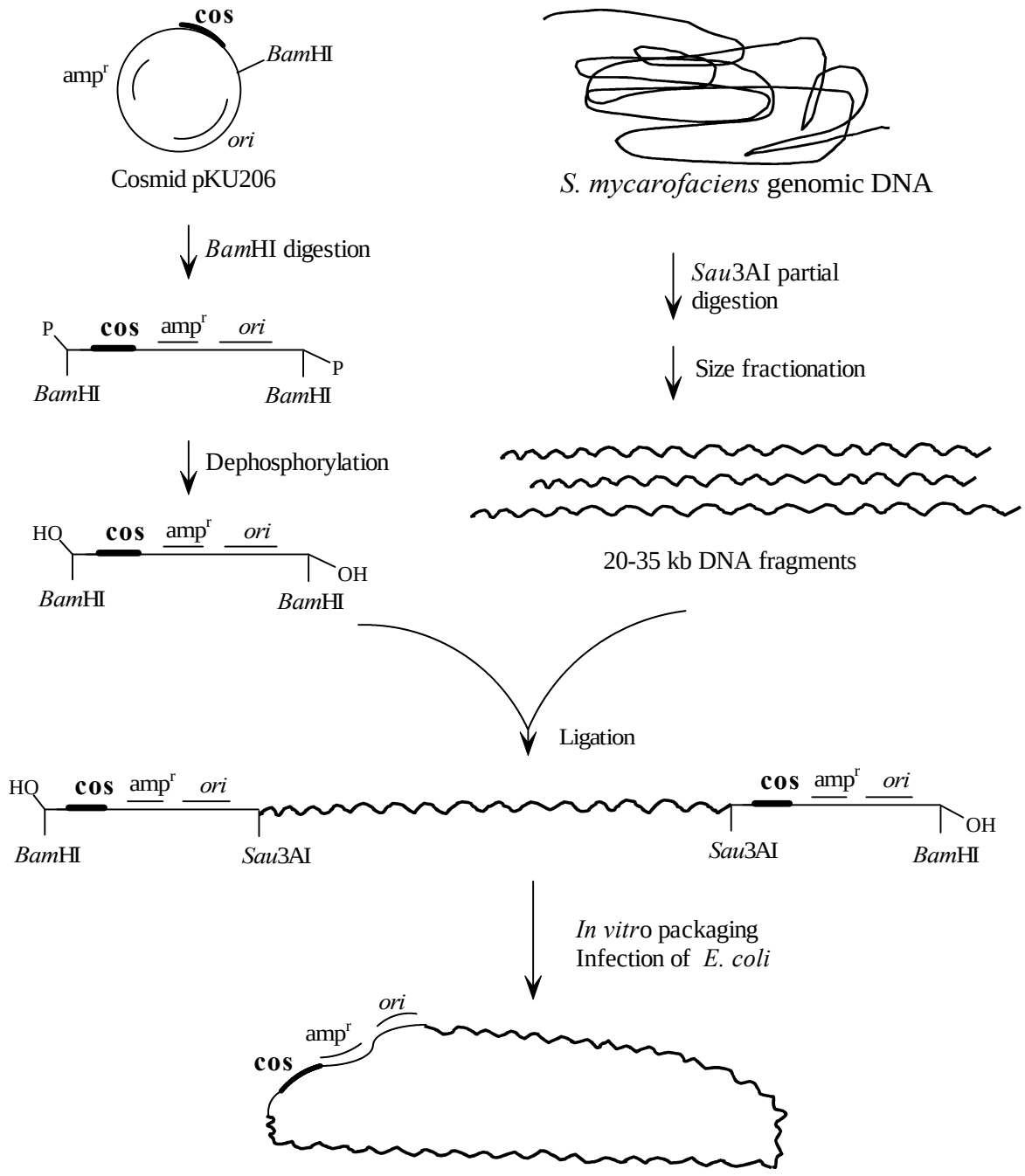


Fig. 4.2. Construction of the genomic library of *S. mycarofaciens* in the cosmid pKU206. 20-35 kb fragments of *S. mycarofaciens* genomic DNA generated by partial digestion with *Sau*3AI were ligated to DNA of the cosmid pKU206 digested with *Bam*HI and dephosphorylated with alkaline phosphatase. The resultant concatemers served as substrates for *in vitro* packaging of bacteriophage λ particles. Following introduction into *E. coli*, the cosmid DNA re-circularized and replicated in the form of a large plasmid. The plasmid contains a β-lactamase gene that confers resistance to ampicillin on the host bacterium.

37°C. and the samples were removed at 5 min intervals, up to 25 min (Fig. 4.3). In this experiment, samples from 5 min to 20 min digestions were combined for further use since they seem to be the most suitable for providing high molecular weight genomic fragments (≥ 20 kb).

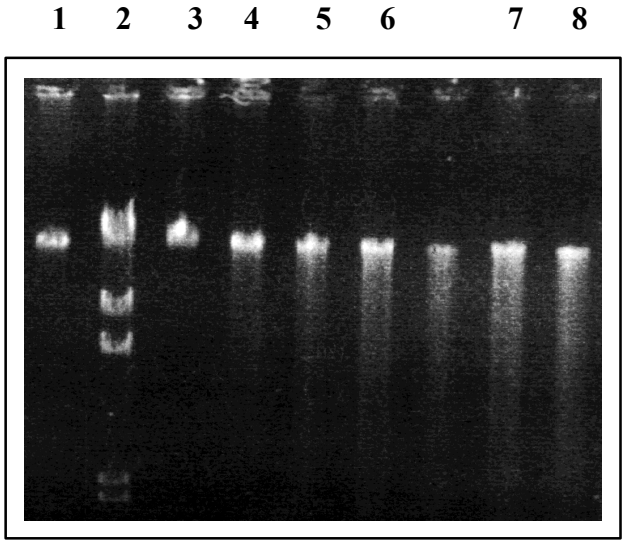


Fig. 4.3. Analysis of *Sau*3AI partial digestion of genomic DNA by agarose-gel electrophoresis. A 0.5% gel was used. *Tracks* (1) λ DNA; (2) λ DNA digested with *Hin*dIII as size markers. *Tracks* containing identical samples of genomic DNA digested with equal unit amounts of *Sau*3AI for different time periods (as described in Section 3.4.1.1) (3) 0 min; (4) 5 min; (5) 10 min; (6) 15 min; (7) 20 min; (8) 25 min.

Sucrose gradients were used for size fractionation of DNA molecules (see Section 3.4.1.2). The fractionated DNA solutions were analysed and the DNA samples were selected in a range of 20-35 kb (Fig. 4.4). Since it turned out that the sucrose concentration affected migration of DNA fragments, the size marker of DNA had to be diluted 1:2 with the 40% sucrose solution. Finally, fractions 1-20 in Fig. 4.4 were combined for further use.

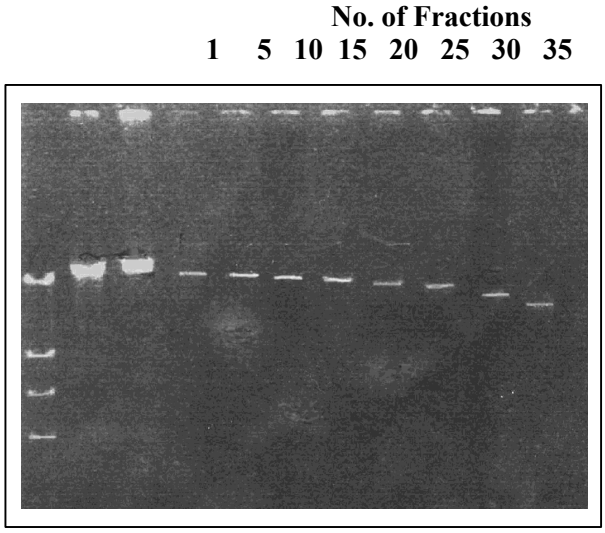


Fig. 4.4. Analysis of sucrose gradient fractions by agarose gel electrophoresis. The first three tracks on the left contain λ DNA digested with *Hin*dIII as size markers, λ DNA and *S. mycarofaciens* genomic DNA. A 0.5% agarose gel was used. 15 µl of every fifth sample was loaded.

4.2.1.2. Ligation, packaging and plating the library

According to their relative sizes (Tab. 4.1), approximately 5 μg of cosmid DNA and 2.5 μg of genomic DNA fragments were used for ligation as described in Section 3.4.1.4. An aliquot sample was used to examine the results of ligation, and it proved that the ligation was successful (Fig. 4.5) .

Tab. 4.1. Relative sizes of DNA molecules

DNA molecule	Size (kb)	Relative size
Cosmid pKU206	15.5	1.0
S. mycarofaciens genomic fragment	20-35	1.6-2.2

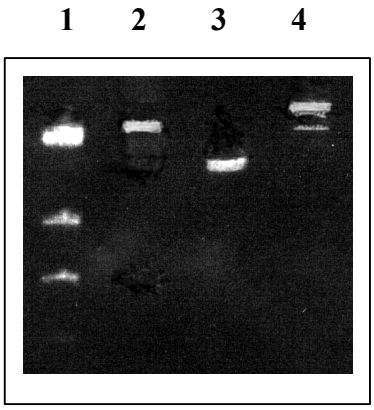


Fig. 4.5. Analysis of ligation reaction by agarose-gel electrophoresis. A 0.5% gel was used. Tracks (1) λ DNA digested with *Hin*dIII as size markers; (2) *S. mycarofaciens* 20-35 kb fragments; (3) pKU206 digested with *Bam*HI and treated with phosphatase; (4) ligation of DNA fragments from (2) and (3) (1 μl ligation mixture).

The adsorption of packaged recombinant cosmids to *E. coli* and the plating of packaged recombinant cosmids were followed as in Section 3.4.1.6-7. The result was that around 5 x 10^4 to 5 x 10^5 transfected bacterial colonies per microgram of ligated DNA were counted. About 5 x 10^3 ampicillin-resistant colonies were collected and used as the cosmid genomic library for screening.

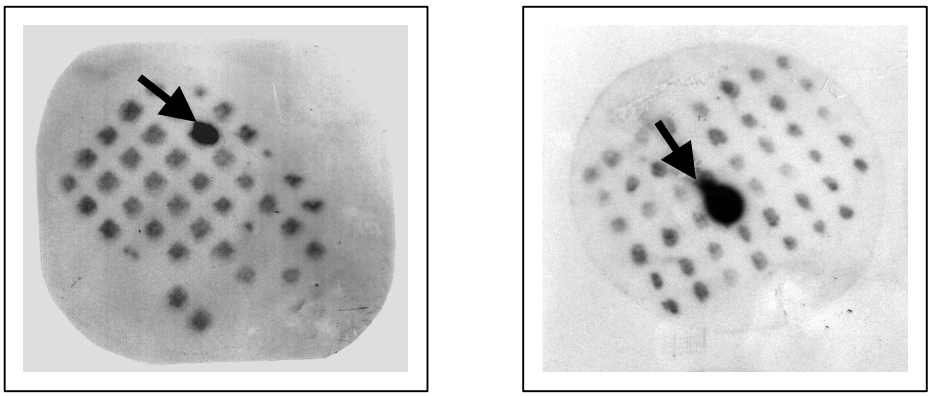
Twenty of the cosmid clones from the library were picked at random and the plasmid DNA was isolated as described in Section 3.4.1.7. The DNA of plasmids was digested by *Bam*HI to examine the inserted DNA size of the recombinant cosmid clones. As a result,

eighteen of the cosmid plasmids were observed to contain the insert DNA, the size being in a range of 20-35 kb (data not shown).

4.2.2. Screening of the cosmid genomic library probed by the *midB* and *midI* DNA fragments

As described in Section 4.1.1-2, the *midB* DNA fragment (303 bp) from pLCW1 and *midI* DNA fragment (420 bp) from pLCW17 were identified as dTDP-glucose 4,6-dehydratase and gylcosyltransferase, respectively. They were used as probes to hybridize the above-established *S. mycarofaciens* genomic library. The probe fragments were purified twice by agarose gel electrophoresis before radiolabelling in order to avoid non-specific hybridization with vector DNA. Resulting from hybridization experiments, 2 out of 3000 colonies on the disc membranes showed significant signals against the *midB* probe (Fig. 4.6). The cosmid DNA was isolated from these two putative positive clones and digested with *Bam*HI and *Bgl*II. The resulting DNA was transferred onto nylon filters and hybridized with the *midB* probe again. As expected, the DNA of these two clones showed signal bands in which the location was the same as the positive control from *S. mycarofaciens* genomic DNA (data not shown). Also, both cosmids were used as templates and amplified by primers AS2 and AS5, which could produce the *midB* DNA fragment as described in Section 4.1.1. These two positive recombinant cosmids were designated as Smyc-LC1 and Smyc-LC2.

One significant signal from 2000 colonies was obtained by the *midI* probe when screening the same genomic library for a second time (data not shown). This putative positive cosmid was further examined again by DNA hybridization and PCR amplification (data not shown). All results supported that the insert DNA of this cosmid contained the *midI* DNA fragment. This recombinant cosmid was designated as Smyc-LC3.



A **B**

Fig. 4. 6. Autoradiographic analysis of screening the cosmid genomic library probed by the *midB* DNA fragment through Southern hybridization. Sets of each 47 single cosmid colonies were blotted on Hybond N^+ disc membranes. Screening colonies against the probe was carried out as described in Section 3.4.3.3. Two strong signals from discs A and B hybridized with the radiolabelled *midB* DNA fragment are visible (see *arrows*).

4.3. Restriction analysis and sequencing of partial regions of the recombinant cosmids

In order to analyse the insert DNA in the three recombinant cosmids, Smyc-LC1, Smyc-LC2 and Smyc-LC3 were digested with *Bam*HI, *Bgl*II and *Bam*HI/*Bgl*II. The calculated results showed that the insert DNA of Smyc-LC1, Smyc-LC2 and Smyc-LC3 were ca. 30.2 kb, 22.6 kb and 26.3 kb, respectively. The DNA fragments of Smyc-LC1 and Smyc-LC2 overlap each other in a region of ca. 14.7 kb. This overlap was also confirmed through hybridization of Smyc-LCl DNA with the DNA fragment probes from both terminal sides of Smyc-LC2. Therefore, Smyc-LC1 and Smyc-LC3 were used for further restriction enzyme analysis. A set of single or double enzymes were applied in the digestion reactions. From the results of the restriction enzyme analysis, restriction maps of Smyc-LC1, Smyc-LC2 and Smyc-LC3 were constructed (Fig. 4.7 and Fig.4.8).

To quickly localize pathway-specific genes in Smyc-LC1 and Smyc-LC3, a set of subclones in *E. coli* cloning vectors were also made as given in Fig. 4.7 and Fig. 4.8. The insert DNA fragments of all subclones were sequenced at least at their ends. The regions of Smyc-LC1 and Smyc-LC3 surrounding the already identified genes *midB* and *midI*, were fully sequenced in about 10 kb and 6 kb, respectively.

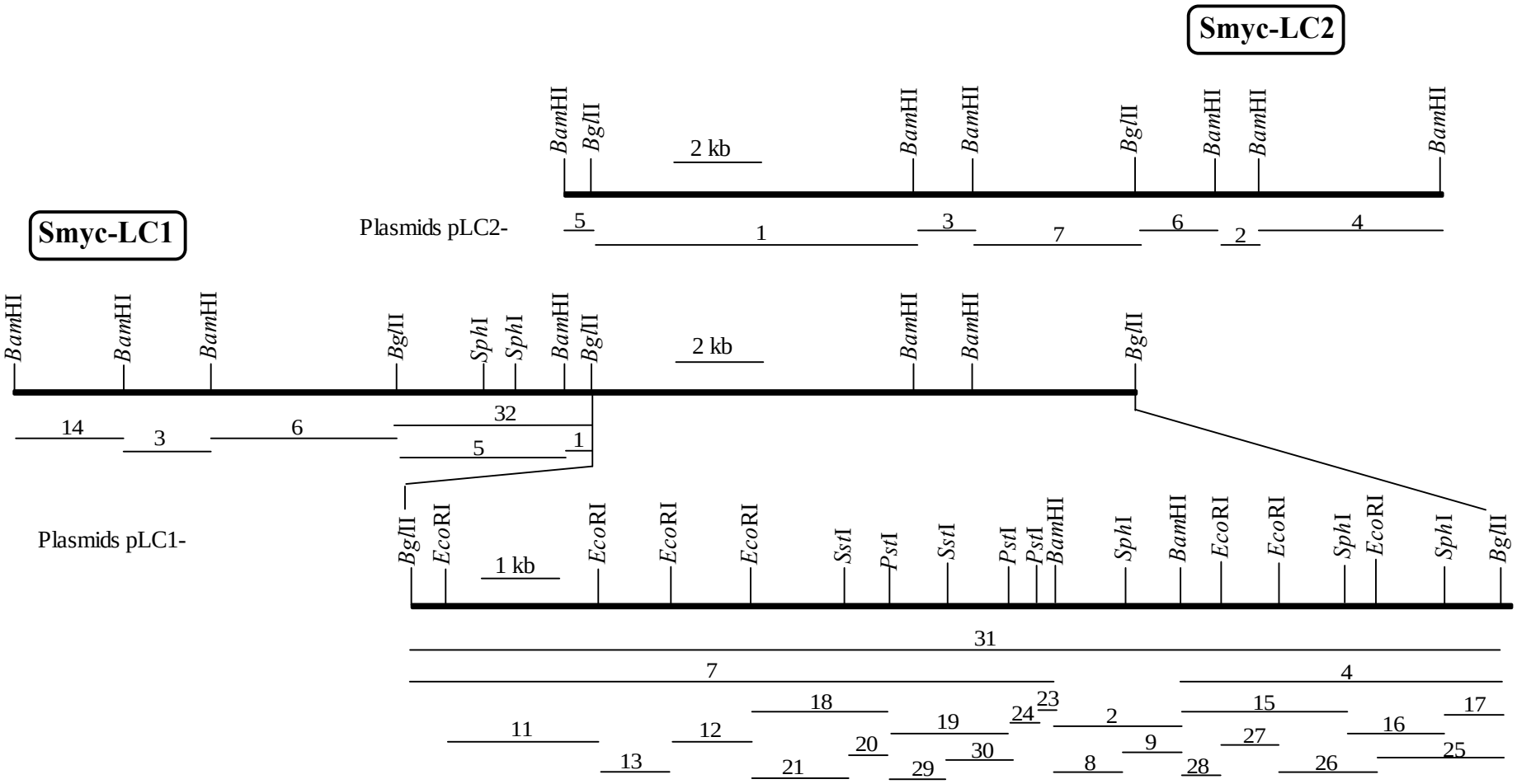


Fig. 4.7. Restriction map of Smyc-LC1 and Smyc-LC2 and sequencing strategy. The plasmids for sequencing were designated as pLC1- and pLC2- with a number. The region of ca.15 kb *Bgl*II fragment is shown in greater detail.

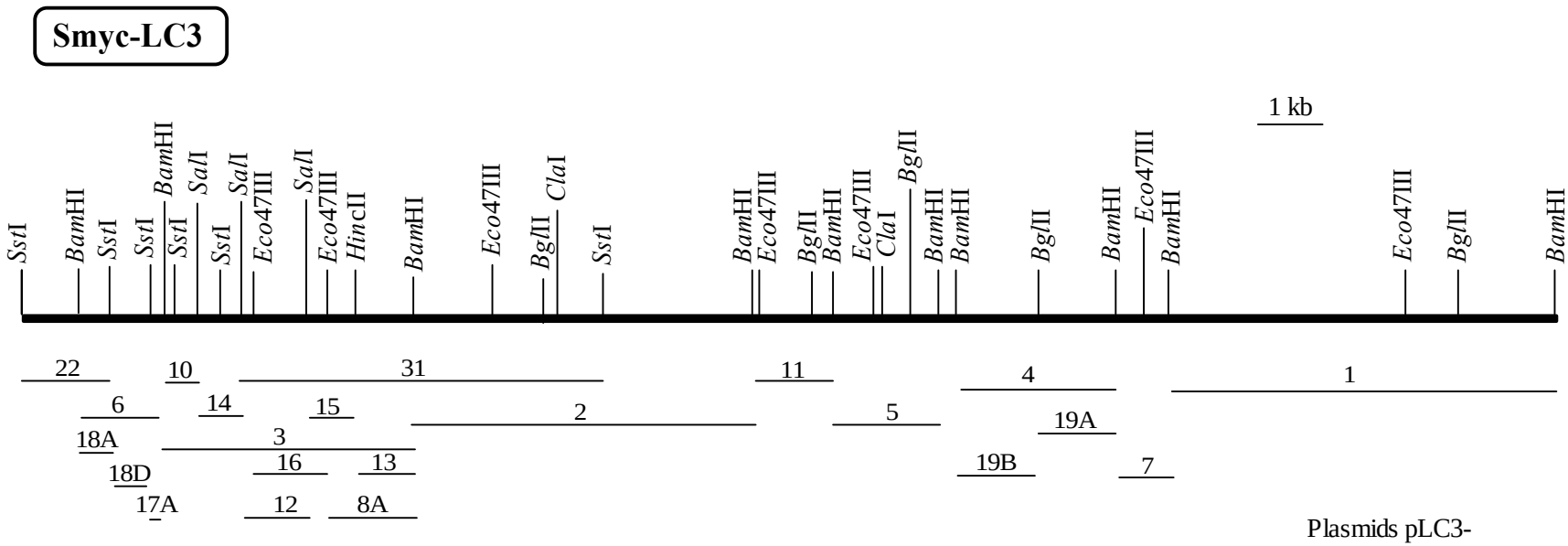


Fig. 4.8. Restriction map of Smyc-LC3 and sequencing strategy. Plasmids for sequencing were named as pLC3- with a number.

4.4. Chromosomal walking by using a 2.3 kb DNA fragment close to one end of Smyc-LC1 as a probe to screen the *S. mycarofaciens* genomic library

To extend the existing gene sub-cluster region covered by Smyc-LC1, a 2.3 kb *Bam*HI fragment was isolated from pLC1-3 (see Fig. 4.7). This DNA fragment, when used as a probe to again screen the cosmid genomic library, resulted in isolation of a positively hybridizing cosmid, which was designated Smyc-LC4. The restriction analysis of this recombinant cosmid with *Bam*HI and *Bam*HI/*Bgl*II showed the size of the inserted DNA to be ca. 29.3 kb. The DNA fragment of Smyc-LC4 overlaps with Smyc-LC1 in a region of ca. 11.3 kb and extends away from Smyc-LC1 in a region of ca. 18 kb, containing a unique *Bam*HI fragment. The summary of the suggested map location of the four cosmids, Smyc-LC1, Smyc-LC2, Smyc-LC3 and Smyc-LC4, is shown in Fig. 4.9.

4.5. Open reading frames found in the gene sub-cluster analysed from Smyc-LC1

The sequence of a ca. 14 kb *Bgl*II fragment (site 1-15) of Smyc-LC1 which includes the *midB* region was determined from the respective segments subcloned, in which some regions were completely sequenced (ca. 10 kb) and others were partially sequenced (ca. 4 kb) (Fig. 4.10). The arrangement of the identified open reading frames and their encoded proteins, e.g., *midC* encodes a TylB-related aminotransferase, was very similar to the tylosin gene cluster of *S. fradiae* (see below). Therefore, the previous proposal was confirmed that the sub-cluster covered by Smyc-LC1 is part of the gene cluster in midecamycin biosynthesis and that these genes are essential for the 6-deoxysugar biosynthetic pathway.

Upstream of *midC* there was a 500 bp non-coding sequence gap. Adjacent to this region, one open reading frame was found to be orf1 (*midM*), showing the same reading direction as *midC*. Downstream of *mid*C, five further complete open reading frames, orf3 (*midA*), orf4 (*midB*), orf5 (*midD*), orf6 (*midE*) and orf8 (*midL*) were identified. The locations and the reading directions are given in Fig. 4.10. An uncomplete reading frame, orf7, identified as *midF* from the sequence of two small segments, was also found downstream of *midE*. All of the reading frames display a typical *Streptomyces* codon bias (Bibb *et al*., 1984).

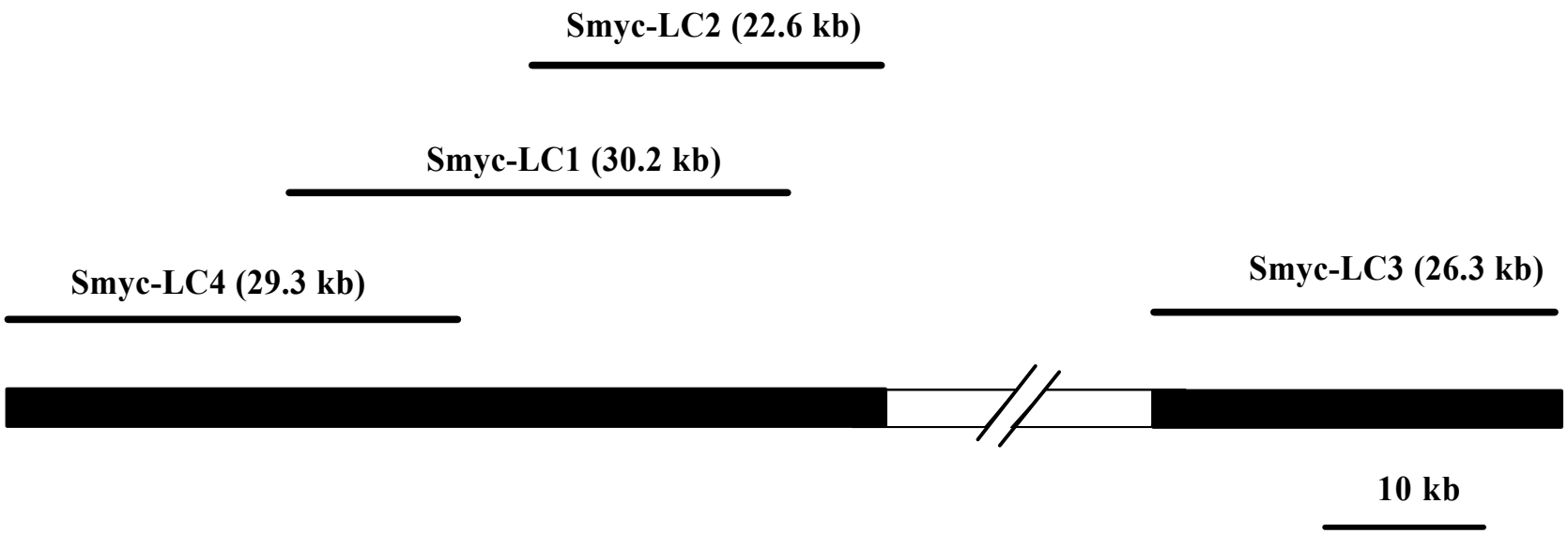


Fig. 4.9. Schematic representation of cosmid clone location in the *mid* gene cluster. The black bars represent the region of *S. mycarofaciens* chromosomal DNA found in existing cosmids. The white bar indicates the unresearched region of the chromosome with a break shown as a double slant.

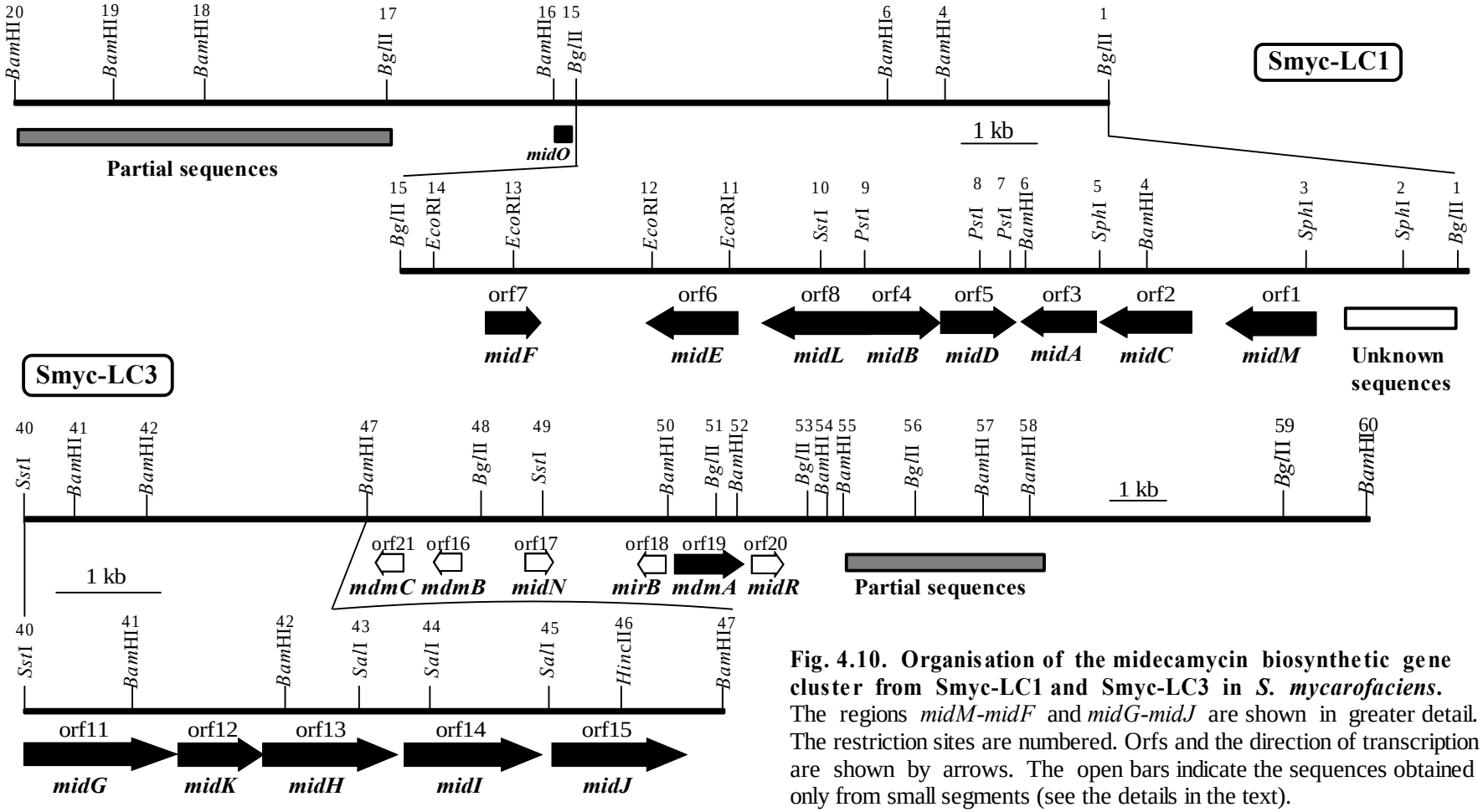


Fig. 4.10. Organisation of the midecamycin biosynthetic gene cluster from Smyc-LC1 and Smyc-LC3 in *S. mycarofaciens*. The regions *midM-midF* and *midG-midJ* are shown in greater detail. The restriction sites are numbered. Orfs and the direction of transcription are shown by arrows. The open bars indicate the sequences obtained only from small segments (see the details in the text).

The further analysis of some pathway-specific genes postulated to be involved in midecamycin sugar biosynthesis will be described in the following sections.

4.6. Open reading frames found in the gene sub-cluster analysed from Smyc-LC3

The complete sequence of a 4.3 kb *Bam*HI fragment (sites 42-47) of Smyc-LC3 which includes the *midI* region was determined from the suitable plasmids of the pLC3- series. Adjacent to this region, a 1.38 kb *Bam*HI fragment (sites 41-42) was also completely sequenced, and another ca. 1.0 kb *Sst*I-*Bam*HI fragment (sites 40-41) was partially sequenced. As a result, five open reading frames (orf11-15) were found in the region between sites 40 and 47, four of which represented the complete reading frames, and all read in the same direction (Fig. 4.10). Although the first open reading frame, orf11 (*midG*), had no complete sequence, it was found to encode PKS at the 3´ end. Orf12 (*midK*) is located immediately downstream of *midG*. Orf13 (*midH*) was found to share the sequence with the 3´ end of the *midK* gene in a region of 31 base pairs. Orf14 (*midI*) is located downstream of the *midH* gene, and following the *midI* gene comes orf15 (*midJ*). The detail analysis of the first four genes will be done in the following sections.

The sequence of an approximately 9.2 kb DNA fragment (sites 47-54 in Fig. 4.10) was partially performed using respective plasmids from the pLC3- series. In this region, only orf19 (*mirA*) was identified to be a complete open reading frame. The organization of the partially sequenced orf16 (*mdmB*), orf17 (*midN*), orf18 (*mirB*), orf20 (*midR*) and orf21(*mdmC*) is shown in Fig. 4.10.

4.7. The PKS genes (*midG*)

Although finding the midecamycin PKS genes was not the target of this study, they could facilitate the identification of the overall organisation and orientation of the gene cluster and the genes encoding 6-deoxysugar biosynthetic enzymes. A little information about the PKS genes was obtained from two adjacently sequenced fragments in Smyc-LC3. The first sequence (404 bp) derived from plasmid pLC3-6, named as orf11 (*midG*), is located immediately upstream of the 5´ end of orf12 (*midK*) in the *Bam*HI fragment (site 41-42)

(see Fig. 4.10). The deduced 133 aa sequence encoded by *midG* was found to be the 3´ end sequence of the midecamycin PKS, since the product of *midG* shows 60% identity to PKS module 7, NidA5, in 16-membered niddamycin biosynthesis from *S. caelestis* (acc. no. AF016585; Stephan *et al*., 1997), and 48% identity to tylactone PKS module 7, TylGV, in tylosin biosynthesis (acc. no. U78289). The high sequence identity between the deduced C-terminal amino acids derived from MidG and NidA5 has revealed that the current found sequence is possibly encoded as a thioesterase (TE) which is responsible for the release and cyclization of the macrolide ring, identifying the end of the PKS sub-cluster (Donadio & Katz, 1992) (for sequence see Appendices 7.5). Another DNA sequence of about 1.0 kb was obtained from part of the insert fragment of the plasmid pLC3-22 extending from the *Bam*HI (site 41) to the *Sst*I (site 40) at the end of the Smyc-LC3 insert. Although this fragment was not completely sequenced, information obtained from partial sequencing has strongly supported that it is encoded as a PKS near the C-terminus. The deduced 264 aa sequence derived from pLC3-22 shows end-to-end similarity to niddamycin PKS module 7, NidA5 (58% identity) in which the nucleotide sequence in this region is encoded as an acyl carrier protein (ACP) and a TE.

4.8. Regulation genes (*midF* and *midR*)

Sequencing two DNA segments, from pLC1-11 and pLC1-13 in Smyc-LC1, revealed that the deduced product encoded by the nucleotide sequence is similar to the SrmR protein. Therefore, this gene was identified as orf7 (*midF*) (see Fig. 4.10). The deduced 228 aa sequence close to the N-terminus of MidF was found to have 49% identity to the product of *srmR* (acc. no. X63451). SrmR has been identified as a transcriptional activator of the expression of the PKS genes in the spiramycin biosynthetic gene cluster from *S. ambofaciens* (Geistlich *et al*., 1992).

A second putative regulator gene, orf20 (*midR*) was found on Smyc-LC3 cloned in pLC3-26, which is located downstream of orf19 (see Fig. 4.10). This partial sequence revealed that the encoded 140 aa sequence, close to the N-terminus of MidR, was similar to OrfY (36% identity), a regulator of antibiotic transport complexes in the rapamycin producing strain *S. hygroscopicus* (acc. no. X86780; Schwecke *et al*., 1995), and SC7H1.21 (34% identity), a putative transcriptional regulator in *S. coelicolor* (acc. no. AL021411).

4.9. Resistance genes (*mdmA* and *mirB*)

The 813 bp nucleotide sequence of orf19 located in Smyc-LC3 was found to be identical to the previously identified gene, *mdmA*, whose sequence had been published elsewhere (acc. no. A60725; Hara & Hutchinson, 1990). The deduced product of orf19 (271 aa) exhibits 54% sequence identity to that of TlrD from *S. fradiae* (acc. no. X97721), which confers ribosomal resistance to macrolides, lincosamides and streptogramin B type (MLS) antibiotics, and methylates 23S rRNA (Gandecha & Cundliffe, 1996). So Orf19 (MdmA) is identified as a putative midecamycin-resistance protein.

The second resistance gene, orf 18 (*mirB*), was found to be located upstream of *mdmA* (see Fig. 4.10). Only a small segment encoding the N-terminal region (135 aa) of *mirB* was sequenced from the plasmid pLC3-2. This showed strikingly high sequence similarity to three resistance proteins from the products of other 16-membered macrolides: CarA (67% identity) from *S. thermotolerans*, a carbomycin producer (Schoner *et al*., 1992), SrmB (63% identity) from *S. ambofaciens* in the spiramycin gene cluster (Geistlich *et al*., 1992), and TlrC (54% identity) from *S. fradiae* in the tylosin gene cluster (Rosteck *et al*., 1991). These proteins possess significant sequence similarity to ABC-transporters. The sequence alignment (Fig. 4.11) demonstrates that all these proteins contain highly conserved regions in the cytoplasmic ATP-binding domains typically present within this superfamily of transport proteins.

```
MirB    ---MLTAQLALHDITKRYNDRVVLDRVGFTIKPGEKVGIIGHNGSGKSTLLKLIAGREQA
CarA    ---MSTAQLALHDITKRYQDHVVLDRIGFTIKPGEKVGVIGDNGSGKSTLIKLIAGREQP
SrmB    ---MSIAQYALHDITKRYHDCVVLDRVGFSIKPGEKVGVIGDNGSGKSTLLKILAGRVEP
TlrC    MRTSPSSQLSLHGVTKRYDDRVVLSQVSLAISPGEKAGIIGDNGAGKSTLLRLLAGEERP
               *  **  **** * ***      * **** * ** ** *****    **

MirB    DNGAVTMVAPGGTGYLAQTLELAPEATVQDAVDLAMVELREIEAGVRRAEAELGRTAL
CarA    DNGAVTVVAPGGVGYLAQTLELPLEATVQDAVDLALADLRELEEGMRRTEAELAER--
SrmB    DNGALTVVAPGGVGYLAQTLELPLDATVQDAVDLALSDLRELEAAMREAEAELGES--
TlrC    DAGEVTVIAPGGVGYLPQTLGLPPRATVQDAIDLAMTELRVLEAELRRTEAALAEAA-
        * *  *  **** *** *** *   ****** ***   **  *   *  ** *
```

Fig. 4.11. Comparison of the N-terminal amino acid sequence of MirB with those of other ABC-transporters from products of 16-membered macroildes. The following resistance proteins are compared: CarA protein from *S. thermotolerans* (M80346); SrmB protein from *S. ambofaciens* (X63451); TlrC protein from *S. fradiae* (M57437). GenBank accession numbers are given in *parentheses*. A consensus ATP-binding motif is marked. Sequence identities between these proteins are indicated by *asterisks*.

4.10. Identification of *midE*, *mdmB* and *mdmC*

In the sequence of a complete open reading frame, orf6 (*midE*), located in Smyc-LC1, it was found that the deduced product (388 aa) encoded by *midE* shows 51% sequence identity to that of *acyB1* from *S. thermotolerans* (acc. no. D31821; Arisawa *et al*., 1993). AcyB1 is identified as 4"-*O*-isovaleryl transferase to convert the isovaleryl group to the C4-OH of mycarose in carbomycin biosynthesis. Therefore, MidE is deduced as a 4"-*O*-propionyl transferase to convert the propionyl group to the C4-OH of mycarose in midecamycin biosynthesis (see the structures in Fig. 1.1; for sequence see Appendices 7.4).

Two adjacent genes, orf16 and orf21, are located downstream of *midGKHIJ* in Smyc-LC3 and both read in the opposite direction to *midGKHIJ* (see Fig. 4.10). Based on the nucleotide sequences from small segments, orf16 and orf21 were found to be identical to the genes, *mdmB* and *mdmC*, whose sequences have earlier been published by Hara and Hutchinson (1992; acc. no. M93958). The *mdmB* gene encodes a 3-*O*-acyltransferase which has the ability to catalyse the addition of a propional group to C3-OH of the lactone ring in midecamycin biosynthesis. The *mdmC* gene encoding a *O*-methyltransferase is located immediately downstream of *mdmB*, and the deduced product catalyzes the addition of a methyl group to C4-OH of the lactone ring in the production of midecamycin.

4.11. The *midC* gene

Orf2 (*midC*), was identified in Smyc-LC1, encoding a polypeptide of 398 aa with an estimated Mr of 42 323 and an average G+C content of 72.8%. At the stage of sequence analysis, orf2 shows three putative start codons: an ATG codon which contains no satisfied potential ribosomal binding site, a GTG codon located 12 base pairs downstream of the first putative start codon, and another GTG codon appeared 6 base pairs downstream of the second putative start codon, of which the first GTG codon is preceded by a sequence (GGAGC) that possibly represents a ribosomal binding site, since it shows a degree of complementarity to a region close to the 3' end of the 16S rRNA of *S. lividans* (Bibb & Cohen, 1982) (Fig. 4.12). The definitive start codon will be determined through gene expression by cloning these possible start codons into the expression vectors and

examining the expression patterns (see Section 4.19.2). The orf2 ends in a TGA codon which is one nucleotide before starting orf3 (*midA*) (for sequence see Appendices 7.2).

```
CAC ATT CCT GCG CGA CGC ATG GAG CAT AAA GTG AAC GTG CCC TTT CCT
H   I   P   A   R   R   M   E   H   K   V   N   V   P   F   P
```

Fig. 4.12. The sequence around the starting region of *midC*. Three putative start codons of *midC* are printed in *bold*. A possible ribosomal binding site is *underlined*.

The deduced product of *midC* shows significant sequence similarity with putative aminotransferases from several deduced gene products in different antibiotic biosynthetic pathways. The highest identity scores were found to OleN2 in the oleandomycin-producing strain, *S. antibioticus* (49% identity); TylB in *S. fradiae* (47% identity); EryCI in *Sac. erythraea* (47% identity; Dhillon *et al*., 1989); DesV in *S. venezuelae* (45% identity), which is notable for its ability to produce two distinct groups of macrolide antibiotics, 12-membered methymycin and neomethymycin, and 14-membered narbomycin and pikromycin (Xue *et al*., 1998). The *tylB* product might catalyse the conversion of dTDP-3-keto-6-deoxyglucose to dTDP-3-amino-6-deoxyglucose during dTDP-D-mycaminose biosynthesis in the tylosin biosynthetic gene cluster. The EryCI, OleN2 and DesV proteins might be involved in transamination during dTDP-D-desosamine biosynthesis in the biosynthetic gene clusters of erythromycin, oleandomycin and macrolides, respectively. The aminotransferase enzyme is thought to be dependent on pyridoxal phosphate as a cofactor (Thorson *et al*., 1993; Piepersberg, 1994; Pascarella & Bossa, 1994). The sequence similarities between these proteins are most covered in a region of the protein that contains the conserved lysine residue which is supposed to be the attachment site for pyridoxal phosphate and the conserved aspartate residue which makes a hydrogen bond with N1 of the pyridoxal ring (Fig. 4.13).

```
MidC     PVHLYGHPADLDPLLAIAERHGLAVVEDAGSAR-RPLPGRRIGS
TylB     PVHLYGHPVDLDPVGAFAEPHGLAVVEDAAQAT-ARYRGRRIGS
OleN2    PVHLYGHPADLAALSEVAERHGVRILEDAAQAHGAQAYGRRVGA
EryCI    PVHLYGHPADLDALRAIADRHGLALVEDVAQAVGARHRGHRVGA
DesV     PVHLYGHPADMDALRELADRHGLHIVEDAAQAHGARYRGRRIGA
         ******** *        *  **    **   *      * * *

MidC     GH-VVAFSFYPGKNLGAMGDGGAVVTGDAALADRIRLLRKCG
TylB     GH-RTAFSFYPGKNLGALGDGGAVVTSDPELADRLRLLRNYG
OleN2    WS-TTAFSFYPGKNLGGFGDGGAVVTDDAELAERVRLLRNYG
EryCI    GSNAAAFSFYPGKNLGALGDGGAVVTTDPALAERIRLLRNYG
DesV     GSSVAAFSFYPGKNLGCFGDGGAVVTGDPELAERLRMLRNYG
           ***********  ******** *  ** * * **  *
```

Fig. 4.13. Partial sequence alignment of MidC with other macrolide aminotransferases. TylB from *S. fradiae* (U08223), OleN2 from *S. antibioticus* (AF055579), EryCI from *Sac. erythraea* (X155541) and DesV from *S. venezuelae* (AF079762). GenBank accession numbers are given in *parentheses*. The marked letters indicate the proposed consensus sequence for putative pyridoxal phosphate-binding site (Piepersberg, 1994).

4.12. The *midA* and *midB* genes

Two complete open reading frames, orf3 and orf4, designated as *midA* and *midB*, were found in Smyc-LC1, respectively. Orf3 is translationally coupled to orf2 (*midC*) using an ATG start codon by overlapping one nucleotide with the 3' end of orf2, and terminates in a TGA codon. Orf4 overlaps the 5' end of orf8 (*midL*) by two nucleotides, but reads away from orf8. It starts with an ATG and ends six nucleotides upstream of orf5 (*midD*) in a TGA codon (see Fig. 4.10; for sequence see Appendics 7.2 and 7.3).

The deduced product of *midA* (a protein of 303 aa residues) shows significant sequence identity to putative dTDP-glucose syntheses from several *Streptomycetes*. These proteins include DesIII from *S. venezuelae* (62% identity; acc. no. AF079762), TylAI from *S. fradiae* (61% identity; acc. no. U08223), AveBIII from the avermectin-producing strain, *S. avermitilis* (61% identity; acc. no. AB032523; Ikeda *et al*., 1999), SnogJ from the nogalamycin-producing strain, *S. nogalater* (56% identity; acc. no. AF187532; Torkkell, *et al*., 1997), and StrD from *S. griseus* (54% identity; acc. no. AF128273; Pissowotzki *et al*., 1991). They all share a highly conserved motif close to the N-terminus (data not shown).

The deduced product of *midB* (a protein of 326 aa residues) is highly similar in sequence to dTDP-glucose-4,6-dehydratases from several deduced gene products in macrolide antibiotics. These proteins include TylAII from *S. fardiae* (64% identity; acc. no.

U08223), AveBII from *S. avermitilis* (64% identity; acc. no. AB032523), Gdh from *Sac. erythraea* (60% identity; acc. no. L37354; Vara & Hutchinson, 1988), and DesIV from *S. venezuelae* (60% identity; acc. no. AF079762). The dTDP-glucose-4,6-dehydratase purified from *Sac. erythraea* showed that the enzyme required NAD^+ as a cofactor (Vara & Hutchinson, 1988). There is a putative binding site for the cofactor NAD^+ located within the N-terminus 30 aa residues of these proteins (data not shown) (Wierenga *et al*., 1985; Scrutton *et al*., 1990). The enzyme activity of dTDP-glucose-4,6-dehydratase from *S. griseus* in the gene cluster of streptomycin biosynthesis was also stimulated by addition of NAD^+ (Verseck, 1997).

4.13. The *midK* gene

Orf12 (*midK*) located in Smyc-LC3 encodes a polypeptide of 249 aa (Mr = 28 899 and average G+C = 69%). Orf12 is immediately dowmstream of and reads in the same direction as the genes encoding PKS, with an ATG start condon. Analysis of the open reading frame showed that orf12 ends with a TGA codon and overlaps the 5' end of orf13 (*midH*) by 31 base pairs (Fig. 4. 14(a), (b); for sequence see Appendices 7.5).

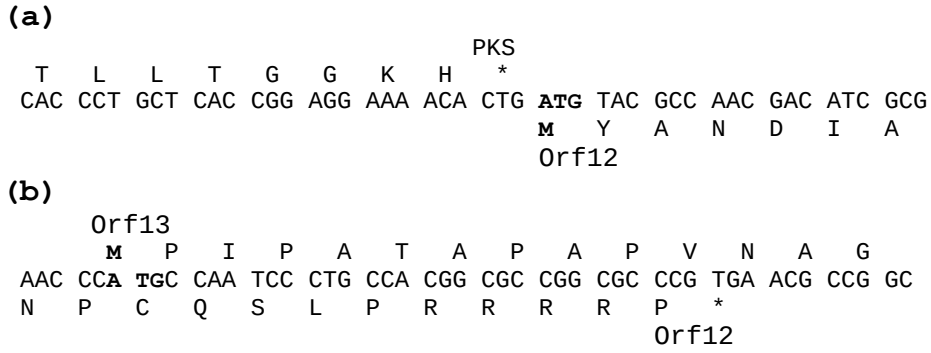


Fig. 4. 14 (a), (b). The location of start and stop sequences of orf12 (*midK*).

The deduced product of *midK* shows significant sequence similarity to a family of enzymes which has recently been proposed to function as S-adenosylmethionine (SAM)-dependent methyltransferases (Gandecha *et al*., 1997). Members of this family include TylMI from *S. fradiae* (60% identity), which acts on the amino group of dTDP-3-amino-6-deoxyglucose during mycaminose biosynthesis, perhaps by introducing two methyl groups at that site (Gandecha *et al*., 1997); OleM1 from *S. antibioticus* (57% identity), which is

involved in oleandomycin biosynthesis (Olano *et al*., 1998), DesVI from *S. venezuelae* (53% identity), which is deduced as a N, N-dimethyltransferase involved in macrolide biosynthesis (Xue *et al*., 1998), and EryCVI from *Sac. erythraea* (51% identity), which participates in erythromycin biosynthesis (Gaisser *et al*., 1997). A sequence alignment of these proteins shows that they each possess all three of the consensus sequence motifs typical of methyltransferases that use SAM as co-substrate (Kagan & Clarke, 1994) (Fig. 4.15). Motif I, near their N-terminus in these proteins, containing the consensus sequence LLDVACGTG, is thought to be a variant of canonical methyltransferase motif G×G×G (Ingrosso *et al*., 1989; Haydock *et al*., 1991) which is found in glycine N-methyltransferases, and forms part of the SAM-binding pocket according to crystallographic data relating to the HhaI DNA m^5C methyltransferase (Cheng *et al*., 1993).

```
MidK     YDLVHEGKGKDYRQEAEEIAQLVRAHRPATRSLLDVACGTGQHLRHLDGLFDHVEGLE
TylMI    YDLVHQGKGKDYHREAADLAALVRRHSPKAASLLDVACGTGMHLRHLADSFGTVEGLE
OleM1    YDAVYRGRGKDYAGEAKDVADLVRDRVPDASSLLDVACGTGAHLRHFATLFDDARGLE
DesVI    YDLFYLGRGKDYAAEASDIADLVRSRTPEASSLLDVACGTGTHLEHFTKEFGDTAGLE
EryCVI   YDRFYRGRGKDYAAEAAQVARLVRDRLPSASSLLDVACGTGTHLRRFADLFDDVTGLE
         **    * ****  **   * ***   *   *oooooooooo **      *    **o
                                        motif I

MidK     LSQDMLAIAIGRNPDVTLHEGDMRSFALGRRFDAVICMFSSIGHLRTTDELDSTLRCF
TylMI    LSADMLAIPRRRNPDAVLHHGDMRDFSLGRRFSAVTCMFSSIGHLAGQAELDAALERF
OleM1    LSASMLDIARSRMPGVPLHQGDMRSFDLGPRVSAVTCMFSSVGHLATTAELDATLRCF
DesVI    LSEDMLTHARKRLPDATLHQGDMRDFRLGRKFSAVVSMFSSVGYLKTTEELGAAVASF
EryCVI   LSAAMIEVARPQLGGIPVLQGDMRDFALDREFDAVTCMFSSIGHMRDGAELDQALASF
         **  *               **** * * oooooooo*****       **      *
                                     motif II

MidK     AGHLEPGGAIVIEPWWFPDSFTPGYVGATSPRRANGTICRVSDSVREGDATRIEVH
TylMI    AAHVLPDGVVVVEPWWFPENFTPGYVAAGTVEAGGTTVTRVSHSSREGEATRIEVH
OleM1    ARHTRPGGVAVIEPWWFPETFTDGYVAGDIVRVDGRTISRVSHSVRDGGATRMEIH
DesVI    AEHLEPGGVVVVEPWWFPETFADGWVSADVVRRDGRTVARVSHSVREGNATRMEVH
EryCVI   ARHLAPGGVVVVEPWWFPEDFLDGYVAGDVVRDGDLTISRVSHSVRAGGATRMEIH
         * *oooooooooo******  *  * *          *  ***** * *** * * *
            motif III
```

Fig. 4.15. Partial comparison of the amino acid sequence of MidK with other macrolide methyltransferases. The following proteins are compared: TylMI from *S. fradiae* (X81885); OleM1 from *S. antibioticus* (AJ002638); DesVI from *S. venezuelae* (AF079762); EryCVI protein from *Sac. erythraea* (U77459). GenBank accession numbers are given in *parentheses*. Sequence identities between these proteins are indicated by *asterisks*. Motifs I to III are marked and shown by *circles*.

4.14. The *midH* gene

Orf13 (*midH*) found in Smyc-LC3 encodes a 414 aa polypeptide (Mr = 44 275 and average G+C = 75.7%). Based on sequence alignment and open reading frame analysis, orf13 is translationally coupled to orf12, since the putative ATG start codon of orf13 begins 31 base pairs upstream of the TGA stop codon of orf12 (see Fig. 4. 14(b)). The overlapping genes, like orf12 and orf13, which are involved in the same metabolic process, are quite common in bacteria (Normark *et al*., 1983). Orf13 ends 42 base pairs upstream of orf14 with a TGA codon (for sequence see Appedix 7.5) .

The deduced product of *midH* shows sequence similarity to TylMIII from the tylosin biosynthetic gene cluster of *S. fardiae* (35% identity; acc. no. X81885), which was recently identified as a hypothetical NDP-hexose 3,4-isomerase and may be involved in mycaminose biosynthesis, and to DnrQ from the daunorubicin biosynthetic gene cluster of *S. peucetius* (32% identity; acc. no. L47164; Otten *et al*., 1995), which might participate in daunosamine biosynthesis. The *midH* gene product also shows a moderate overall sequence similarity (28-30%) to OleP1 from *S. antibioticus* (acc. no. AJ002638), DesVIII from *S. venezuelae* (acc. no. AF079762), SnogN from *S. nogalater* (acc. no. AF187532), and EryCII from *Sac. erythraea* (acc. no. Y14332). All these proteins also show end-to-end sequence similarity to cytochrome P450 enzymes from a variety of sources, but none of them have the characteristic P450 motif (Bairoch, 1992) which contains the highly-conserved heme binding site and includes the cysteine residue that serves as the heme iron ligand (Nelson *et al*., 1993). In the case of the EryCII protein, experimental evidence demonstrated that it possibly acts as a 3,4-isomerase to form dTDP-3-keto-6-deoxy-hexose during dTDP-D-desosamine biosynthesis (Summers *et al*., 1997; Salah-Bey *et al*., 1998).

4.15. The *midI* gene

Orf14 (*midI*) found in Smyc-LC3 encodes a 421 aa polypeptide (Mr = 45 993 and average G+C = 80%). Orf14 starts 42 base pairs downstream of orf13 (*midH*) using a GTG codon with a putative ribosomal binding site (GGAGT) and ends 31 base pairs upstream of orf15 (*midJ*) using a TGA codon (for sequence see Appendices 7.5).

The deduced product of *midI* shows convincing end-to-end sequence similarity to several macrolide antibiotics. The highest identity scores were found to TylMII from *S. fradiae* (59% identity), which encodes a glycosyltransferase that adds mycaminose to the 5-OH of tylactone, the polyketide aglycone of tylosin (Gandecha *et al*., 1997), DesVII from *S. venezuelae* (55% identity), which has the ability to catalyze glycosylation of both the 12-membered and 14-membered ring macrolactones in macrolide biosynthesis (Xue, *et al*., 1998), OleG2 from *S. antibioticus* (54% identity), which catalyses glycosylation of oleandolide in the oleandomycin biosynthetic pathway (Olano *et al*., 1998), and EryCIII from *Sac. erythraea* (51% identity) (Summers *et al*., 1997; Salah-Bey *et al*., 1998), which encodes the desosaminyl glycosyltransferase that adds dTDP-D-desosamine to the 5-OH of 3-α-mycarosyl erythronolide B, an intermediate after the first sugar, mycarose, is attached to erythronolide B that is catalysed by another glycosyltransferase, EryBV (46% sequence identity to MidI) (Gaisser *et al*., 1997; Summers *et al*., 1997). A sequence alignment of these proteins is shown in Fig. 4.16. All these glycosyltransferases retain a characteristic motif, P-NVR-VDFVPL-ALLP-C---VHHGG-GT--TA--HG-P, present in UDP-glycosyl transferases(Jenkins & Cundliffe, 1991), which is localised close to the C-terminus of these enzymes.

4.16. Identification of genes encoding dTDP-D-mycaminose biosynthetic pathway

As described in the introduction, various 6-deoxyhexoses, present in a range of antibiotic molecules, are made from D-glucose-1-phosphate via dTDP-glucose and dTDP-4-keto-6-deoxy-glucose before the pathways diverge (Liu & Thorson, 1994; Piepersberg, 1994). Based on the above sequence analysis and comparison, the *midA* and *midB* genes are supposed to be responsible for the early steps in the midecamycin biosynthetic gene cluster. Three genes, *midC*, *midH* and *midK*, might participate in the synthesis route to dTDP-D-mycaminose, and the *midI* gene is believed to be responsible for attachment of mycaminose to midecamycin lactone. The detail discussion about these genes and their deduced functions involved in the mycaminose biosynthetic pathway and transfer will be commented and concluded in Section 5.

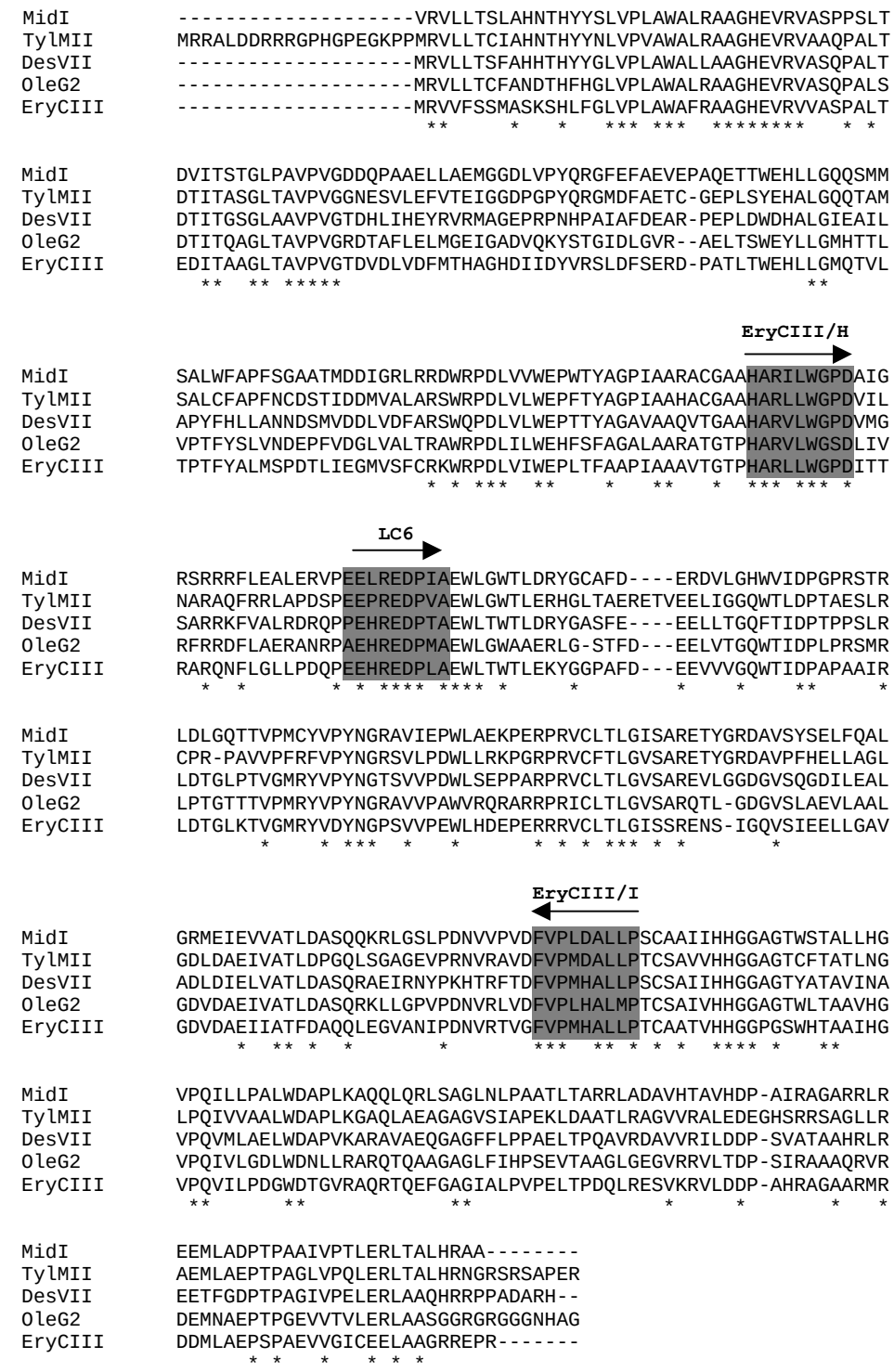


Fig. 4.16. Comparison of amino acid sequence of MidI with other macrolide glycosyltransferases. The following proteins are compared: TylMII from *S. fradiae* (X81885); DesVII from *S. venezuelae* (AF079762); OleG2 from *S. antibioticus* (AJ002638); EryCIII protein from *Sac. erythraea* (Y14332). GenBank accession numbers are given in *parentheses*. Sequence similarity between these proteins are indicated by *asterisks*. The highly conserved sequence regions used for designing the primers, EryCIII/H, EryCIII/I and LC6, are marked and indicated by *arrows* (see Section 4.1.2).

4.17. About strains of *S. mycarofaciens* UC189B (ATCC 21454) and *S. mycarofaciens* ATCC 21454

During the course of examining the production pattern of *S. mycarofaciens* UC189B (ATCC 21454), it was proved that it is a non-producing strain. In order to re-evaluate the two currently identified gene sub-clusters, a series of PCR experiments to compare *S. mycarofaciens* UC189B (ATCC 21454) with *S. mycarofaciens* ATCC 21454, which is a midecamycin-producing strain, were carried out by using primers designed from the identified genes, *midB*, *midC*, *midK* and *midI*, as previously described. As a result, it is confirmed that the two existed sub-clusters encoding deoxysugar genes do contain the same pattern in *S. mycarofaciens* UC189B (ATCC 21454) and ATCC 21454 (data not shown). Further sequence analysis found that downstream of the *midG* (PKS) starting region is a large segment gap (about 4 kb) of unknown coding sequence, that is unusual when compared to those of other macrolide PKS genes, e.g., tylactone PKS *tylG* from *S. fradiae* (acc. no. U78289) and platenolide PKS *SrmG* from *S. ambofaciens* (acc. no. Z46913), which are themselves clustered. Therefore, it is concluded that a mutation in the early steps of midecamycin lactone biosynthesis pre-existed in *S. mycarofaciens* UC189B (ATCC 21454).

4.18. Heterologous complementation of *midK* and *midI* into *Sac. erythraea* mutants

To characterise some pathway-specific genes involved in the mycaminose biosynthetic pathway and for the sugar attachment to the midecamycin lactone, plasmids containing either *midK*, or *midI* or both genes were constructed to transform *Sac. erythrea* mutants CIII68 and CIV89 (Fig. 4.17). The mutants CIII68 and CIV89 were generated by Salah-Bey *et al.* (1998) and identified as the deficient of desosamine biosynthesis and its attachment to macrolactone in the erythromycin gene cluster. When cell supernatants were analysed by TLC from individual transformants of CIII68/pLCW1-29, CIII68/pLC1-31 and CIII68/pLCW1-21, there were some new spots available compared to several control samples. However, none of them showed antibiotic activities via bioassay (data not shown).

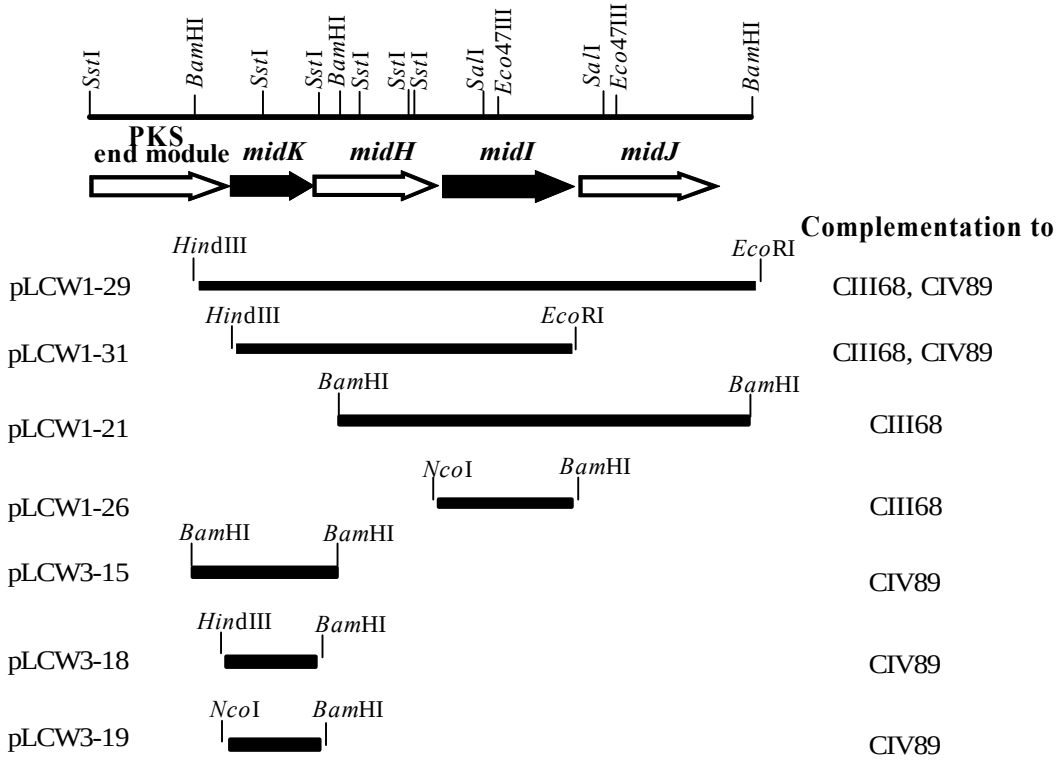


Fig. 4.17. Plasmids used to transform into *Sac. erythraea* mutants strains CIII68 and CIV89

4.19. Heterologous expression of the MidH, MidC, MidK and MidI proteins

In order to characterise enzyme proteins encoded by the proposed genes for the pathway of mycaminose biosynthesis and transfer of the sugar to the midecamycin lactone, the over-expression of individual genes in suitable hosts was investigated. The genes, *midH, midC*, *midK* and *midI*, were amplified by PCR from the cosmid DNA of Smyc-LC1 and Smyc-LC3. The forward primers were designed for introduction of an *Nde*I site, changing the sequence at the natural start codon for the ability to create start codon fusion of these genes into the promoter/ribosome-binding-site cassettes of expression vectors. The reverse primers were designed for introduction of a *Bam*HI or *Bgl*II site located immediately downstream of the stop codon of the gene. The amplified DNA fragments were cloned into (i) two expression vectors of *E. coli*, pET11a, and pET16b (a His-tag fusion protein), under the control of the T7 promoter; (ii) a *E. coli-Stretomyces* shuttle vector, pUWL201 under the control of *ermE* up promoter; (iii) a *S. lividans* vector, pIJ4213 (a His-tag fusion protein) under the control of *tipAp* promoter (Tab. 4.2). The expression conditions and the host strains used will be given in detail in the following sections.

Tab. 4.2. The structure of recombinant plasmids for the heterologous expression of proteins MidH, MidC, MidK and MidI in *E. coli* and *S. lividans*

Protein	Recombinant Plasmid	Vector	Mr. of recombinant proteins (kDa)
MidH	pLCW4-2	pET11a	44.3
His-tag-MidH	pLCW4-3	pET16b	46.3
MidH	pLCW4-6	pUWL201	44.3
His-tag-MidH	pLCW4-7	pUWL201	46.3
His-tag-MidH	pLCW4-10	pIJ4123	45.1
His-tag-MidC	pLCW2-20	pET16b	43.5
His-tag-MidC	pLCW2-21	pET16b	44.3
MidC	pLCW2-23	pET11a	42.3
MidK	pLCW3-2	pET11a	28.9
His-tag-MidK	pLCW3-3	pET16b	30.9
MidK	pLCW3-8	pUWL201	28.9
His-tag-MidK	pLCW3-9	pUWL201	30.9
His-tag-MidK	pLCW3-12	pIJ4123	29.7
MidI	pLCW1-33	pET11a	46.0
His-tag-MidI	pLCW1-23	pET16b	48.0
MidI	pLCW1-41b	pUWL201	46.0
His-tag-MidI	pLCW1-51b	pUWL201	48.0
His-tag-MidI	pLCW1-61	pIJ4123	46.8

For the plasmid construction see Section 2.3.3. The amplification was performed by using PCR method 2 as described in Section 3.3.5.

4.19.1. Over-expression of the MidH protein

Two recombinant expression-plasmids, pLCW4-2 and pLCW4-3 were constructed to introduce the *midH* gene and transformed into *E. coli* BL21(DE3)pLysS. When the respective transformants were cultivated in LB liquid medium at 28°C or 37°C at 220 r.p.m. in a rotary shaker, the over-expressing MidH proteins in both expression-clones were visible on SDS-PAGE gels, but they formed insoluble inclusion bodies. However, when the shaking rate was decreased from 200 r.p.m. to 100 r.p.m., surprisingly, a production of soluble native MidH protein was obtained in large quantities (Fig. 4.18A). A further effort to produce soluble MidH protein was made by using LB liquid supplemented with 1 M sorbitol and 2.5 mM betaine. Because of enhanced osmotic stress of the medium, the culture grew much more slowly (see Section 3.3.12). The resulting expression pattern of *E. coli* /pLCW4-2 or /pLCW4-3 (Fig. 4.18B) showed that the proteins of MidH and its His-tag derivative were produced in large quantities in soluble form. The attempt to express the *midH* gene in *S. lividans* TK23 using the recombinant plasmids pLCW4-6 and pLCW4-7 failed, since neither the SDS-PAGE gels nor Western blotting could observe any visible expected bands.

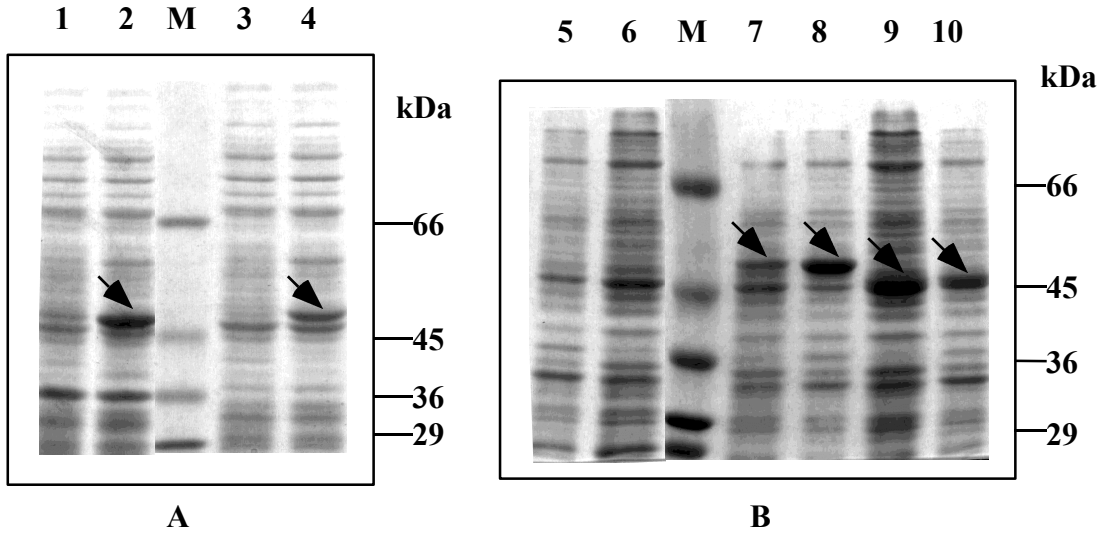


Fig. 4.18. Expression of MidH in *E. coli* BL21 (DE3) pLysS. The cultivation of transformants was performed by (A) agitation at 100 r.p.m, or (B) using LB medium supplemented with sorbitol and betaine. For the other expression conditions see text. Electorphoresis was run in 10% SDS-PAGE gel. About 5 μg of protein was loaded in each lane. Extracts were obtained from *E. coli* transformants containing the following plasmids: pET11a, from the intact cells (*lanes* 1 and 5) and the cell-interrupted supernatants (*lanes* 3 and 6); pLCW4-2, from the intact cells (*lanes* 2 and 9) and the cell-interrupted supernatants (*lanes* 4 and 10); pLCW4-3, from the intact cells (*lane* 7) and the cell-interrupted supernatants (*lane* 8), respectively. The molecular masses (kDa) of the marker proteins (*lane* M) are given. The overproduction of the MidH protein is indicated by *arrows*.

4.19.2. Overexpression of the MidC protein

As described in Section 4.11, the sequence analysis of *midC* shows three putative start codons: the first is an ATG codon, the second is a GTG codon located 12 bp downstream of the first putative start codon, and the third is another GTG codon, which is located 6 bp downstream of the second putative start codon (see Fig. 4.12). In order to be able to express an active protein, three recombinant expression-plasmids were constructed in *E. coli* expression vectors (see Tab. 4.2). In pLCW2-21 and pLCW2-23 the first start codon (ATG) regions, and in pLCW2-20 the second start codon (GTG) region, were replaced by an *Nde*I site. When transformed into *E. coli* BL21 (DE3)pLysS, the expression of MidC proteins was tested in LB liquid medium at 100 r.p.m., 28°C. From the results of SDS-PAGE separation of the respective cell extracts (Fig. 4.19), the overproduction of large amounts of the soluble MidC protein was achieved from the cell-free extracts of the transformants *E. coli*/pLCW2-21 or /pLCW2-23. However, *E. coli*/pLCW2-20 didn't show the expected expression band. The expression could not even be detected by Western blotting analysis (data not shown). Based on this experimental evidence, it is concluded that the translation of *midC* starts with the first ATG codon.

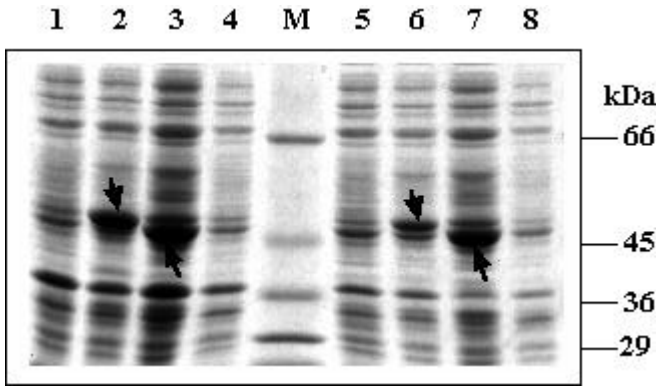


Fig. 4.19. Expression of MidC in *E. coli* BL21 (DE3) pLysS. For the expression conditions see text. Electrophoresis was run in 10% SDS-PAGE gel. About 5 µg of protein was loaded in each lane. Extracts were obtained from *E. coli* containing the following plasmids: pET11a, from the intact cells and the cell-interrupted supernatants (*lanes* 1 and 5); pLCW2-21 from the intact cells and the cell-interrupted supernatants (*lanes* 2 and 6); pLCW2-23 from the intact cells and the cell-interrupted supernatants (*lanes* 3 and 7); pLCW2-20 from the intact cells and the cell-interrupted supernatants (*lanes* 4 and 8), respectively. The molecular masses (kDa) of the marker proteins (*lane* M) are given. The overproduction of soluble MidC protein is indicated by *arrows*.

4.19.3. Expression and detection of the MidK protein

An attempt was made to express the MidK protein from two recombinant plasmids, pLCW3-2 and pLCW3-3. These were again transformed into *E. coli* BL21 (DE3)pLysS. The expression of both corresponding clones resulted in the overproduction of large amounts of the MidK protein. However, the protein was formed as insoluble inclusion bodies although various experimental conditions, such as cultivating in LB medium with or without sorbitol and betaine, or in mineral medium; also, different incubation temperatures at 37°C, 28°C or 20°C with high or slow shaking rates, were tried, but to no avail (data not shown).

Attempts were made to clone the *midK* gene into the *E. coli-S. lividans* shuttle expression vector, pUWL201. The respective plasmids, pLCW3-8 and pLC3-9, were transformed into *S. lividans* TK23. Unfortunately, no visible specific-expression bands were visible either on SDS-PAGE gels from the extracts of both expressing clones or by Western blotting detection from the extracts of TK23/pLCW3-9 (data not shown). Next, the *midK* gene was cloned into the *S. lividans* expression vector, pIJ4123, and the resulting recombinant plasmid pLCW3-12 was transformed into *S. lividans* 1326. Expression of MidK was tested in SGYEME liquid medium at 220 r.p.m., 28°C by thiostrepton induction (see Section 3.3.13). These transformants didn't express the MidK protein in a visible form on SDS-PAGE gels. Therefore, Western blotting and immuno-detection were used to detect the His-tag MidK protein. The result of the expression (Fig. 4.20) showed that soluble His-tag-MidK protein was indeed produced by the recombinant clone, but in low quantities.

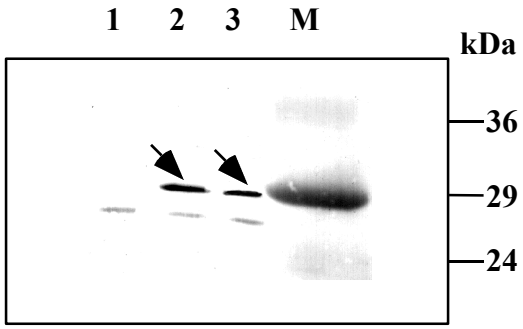


Fig. 4.20. Analysis of the expression of MidK in *S. lividans* 1326 by Western blotting. For the expression conditions see text. Western blotting and immuno-detection see Section 3.5.4. Cell-free extracts were obtained from *S. lividans* containing the following plasmids: pIJ6021 (*lane* 1) and pLCW3-12 (*lanes* 2 and 3). The expression of MidK is indicated by *arrows*.

4.19.4. Expression and detection of the MidI protein

For the expression of the putative glycosyltransferase, MidI, two recombinant expression-plasmids, pLCW1-23 and pLCW1-33 were constructed. These were transformed into *E. coli* BL21 (DE3) pLysS and *E. coli* JM109 (DE3). Overproduction of the MidI protein was observed in the intact cell extracts of both expression-clones in both *E. coli* host strains. However, the products were not visible in the cell-disrupted supernatant (cell-free extracts) (data not shown). By using Western blotting and immuno-detection, the soluble His-tag-MidI protein was detected from the clone *E. coli* BL21 (DE3) pLysS/pLCW1-23 (Fig. 4.21).

Other efforts were made in order to produce the soluble MidI protein, such as testing different media and temperatures, but all failed. The *midI* gene was also cloned into pUWL201, resulting in the plasmids pLCW1-41b and pLCW1-51b, and into pIJ4123, producing the respective plasmid pLCW1-61. These were transformed into *S. lividans* TK23 and *S. lividans* 1326, respectively. Unfortunately, the expression of MidI was not visible on SDS-PAGE gels in any of the recombinant clones. Further analysis by Western blotting revealed that small quantities of the soluble MidI protein were obtained from the cell-free extracts of *S. lividans* 1326/pLCW1-61, however, the expression band was much weaker than the expression shown in *lane 2* of Fig. 4.21 (data not shown).

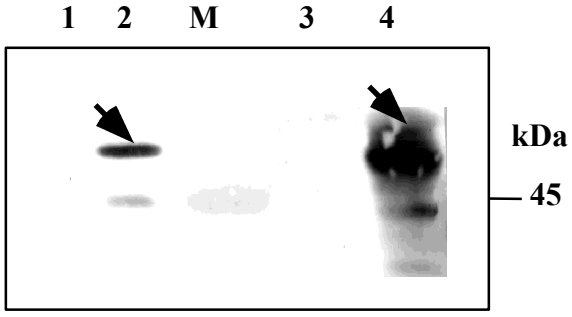


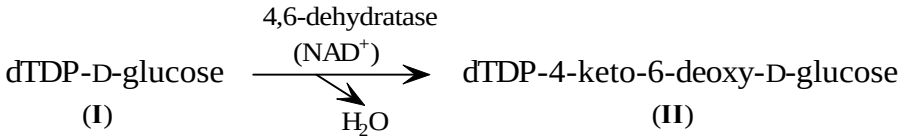
Fig. 4.21. Analysis of the expression of MidI in *E. coli* BL21 (DE3) pLysS by Western blotting. The expression strain contains the following plasmids: pET16b from the cell-disrupted supernatants (*lane* 1) and the intact cell extracts (*lane* 3); pLCW1-23 from the cell-disrupted supernatants (*lane* 2) and the intact cell extracts (*lane* 4). The expression of MidI is indicated by *arrows*.

4.20. Assay for activity of a coupling reaction of the RlmB, MidH and MidC proteins

The proposed mycaminose biosynthetic pathway starts from D-glucose-1-phosphate, via dTDP-glucose and dTDP-4-keto-6-deoxy-glucose. In course of their further conversion to dTDP-D-mycaminose, the following two intermediates, dTDP-3-keto-6-deoxy-D-glucose and dTDP-3-amino-6-deoxy-D-glucose, are postulated (see Fig. 1.6). In order to confirm this hypothesis by experimental data, the over-expressed proteins, MidH and MidC, putatively involved in the biosynthesis of mycaminose, were tested by coupled enzyme reactions in *vitro*.

4.20.1. Activity assay of dTDP-D-glucose 4,6-dehydratase by RmlB

To prepare the first intermediate, dTDP-4-keto-6-deoxy-D-glucose (**II**), the expressed RmlB protein (Verseck, 1997) was used to catalyse the following reaction using dTDP-D-glucose (**I**) as a substrate.



The crude extract of soluble RmlB protein was obtained from the expression-clone *E. coli* BL21(DE3)/pSVW701, thereby the enzymatic reaction doesn't need NAD^+ as a cofactor since NAD^+ is present in tightly bound dehydratases in *E. coli* (Wang & Gabriel, 1969). The analysis of this enzyme reaction (see Section 3.5.5) was performed by HPLC. A typical wide peak appeared at 33.37 min, which was identified as the converted prodcut **II**, compared to a peak at 34.49 min for the substrate **I** (Fig. 4.22A, B). Some small peaks came out at the early stage in Fig. 4.22B because the RmlB was extracted directly from the expression culture without further purification.

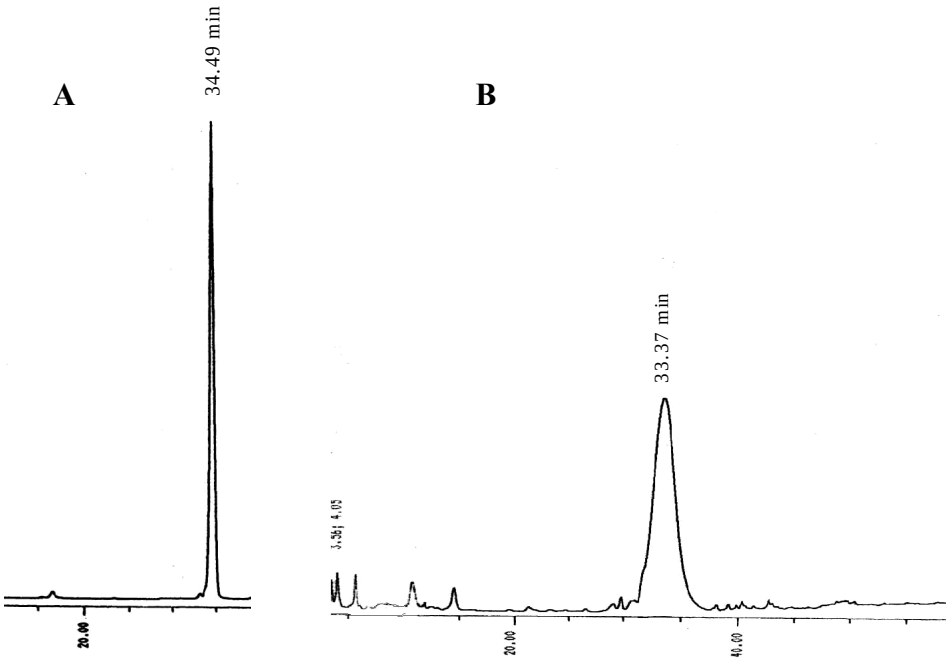


Fig. 4.22. HPLC analysis of the RmlB reaction product. (A) a peak at 34.49 min is dTDP-D-glucose (**I**) as a control. (B) a peak at 33.37 min is dTDP-4-keto-6-deoxy-D-glucose (**II**) converted by the crude RmlB extracts.

4.20.2. Reaction products from combined catalysis by RmlB, MidH and MidC

Two enzymes are postulated to be involved in the following reactions in the pathway of mycaminose biosynthesis. Based on the above study, the over-expressed MidH and MidC proteins were used to test their deduced functions as an isomerase and an aminotransferase, respectively.

dTDP-4-keto-6-deoxy-D-glucose (**II**) $\xrightarrow{\text{3,4-isomerase}}$ dTDP-3-keto-6-deoxy-D-glucose (**III**) $\xrightarrow{\text{3-amino transferase}}$ dTDP-3-amino-6-deoxy-D-glucose (**IV**)

Since the intermediates **II** and **III** were produced as an unstable form, a coupling enzymatic reaction to combine RmlB, MidH and MidC was done to detect the conversion of the substrate **I** to the product **IV,** by using L-alanine or L-glutamic acid as an amino acid

donor and PLP as a cofactor. The crude extracts of MidH and MidC were obtained from the expression clones *E. coli* BL21(DE3)/pLCW4-2 and /pLCW2-23, respectively. The result of HPLC assay showed that there was a specific peak at 4.6 min when compared to all other controls. Unfortunately, because all used enzymes, RmlB, MidH and MidC, were not purified, the high background appeared at the beginning of retention time on HPLC (data not shown). Therefore, LC-MS was used to further analyse this new peak. It showed a peak at m/z = 546 ($M\text{-}H^+$) which is the expected dTDP-amino-6-deoxy-D-glucose product, compared to a peak at m/z = 545 ($M\text{-}H^+$) which is the compound **II** and a peak at m/z = 547 ($M\text{-}H^+$) which is dTDP-L-rhamnose that was produced as a by-product due to the crude enzymes used. ESI-MS/MS spectra by single ion monitoring confirmed this result to compare a peak at m/z = 562.8 ($M\text{-}H^+$) which is the substrate **I** (Fig. 4.23).

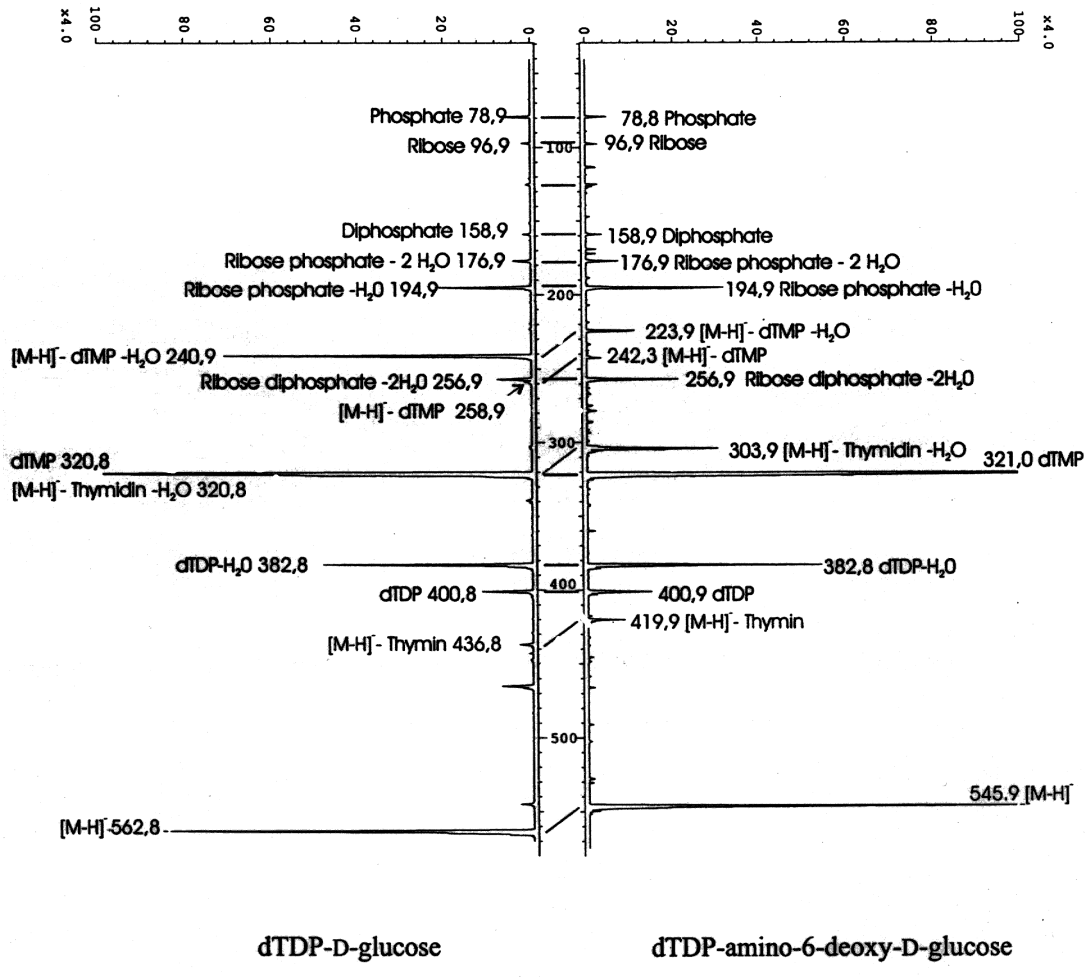


Fig. 4. 23. ESI-MS/MS spectra of dTDP-D-glucose and dTDP-amino-6-deoxy-D-glucose.

5. Discussion

Midecamycin biosynthesis proceeds via the production of a 16-atom polyketide lactone, to which two deoxyhexose moieties are subsequently added (see Fig. 1.1). For a better understanding and application of this process, cloning of the midecamycin biosynthetic (*mid*) gene cluster was suggested as a useful tool to step-wise analyse the functions encoded by these genes in the production of midecamycin. The individual tools delivered by the *mid* gene cluster could also be used as part of a system for combinatorial biosynthesis, which involves the genetic manipulation of mutistep biosynthetic pathways to create molecular diversity in natural products for use in drug discovery. The main goals of this work, therefore, were to identify the sub-clusters for 6-deoxyhexose formation and transfer as well as the expression and enzymological analysis of the dTDP-D-mycaminose pathway.

5.1. The *S. mycarofaciens* cosmid genomic library

To quickly isolate the *mid* gene cluster and identify the sugar sub-clusters in it, a cosmid genomic library was first constructed using the *S. mycarofaciens* chromosomal DNA and the cosmid vector pKU206. This turned out to be a straightforward process for the following reasons. (1) Large and expected DNA fragments were cloned in a range of 20-35 kb. (2) The recombinant cosmid clones can directly replicate both in *E. coli* and in *S. lividans*, which could facilate gene function analysis in *Streptomyces*. (3) It is more stable when the positive recombinant cosmid clones are kept in *S. lividans* for maintenance.

There were two major problems encountered during the manipulation of this gene library. First, how to maintain all cosmid clones as long as possible? Three alternative methods for storing different stages of the gene library were used in this study. They were (i) as ligation stored at -20°C for an indefinite period; (ii) as frozen transformed *E. coli* clone mixture at -70°C for a few years; (iii) as frozen transformed *E. coli* single clones in microtiter plates stored at -70°C for a few years (see Section 3.4.1.5 and 3.4.2.1-2). We had used the two-year-old ligation for packaging and plating a new gene library. The efficiency of producing

cosmid clones turned out to be almost the same as the first gene library. The second method was easier to handle but it implicated the amplification of the whole library during repeated uses. The third method can keep an authentic gene library, but it requires some skill to handle the process of storing. Secondly, cosmid clones have been noted to have a problem of instability (Weis *et al*., 1987). Deletions arose if the cloned DNA contained larger repeated sequences, especially PKS genes (Omura & Ikeda, personal communication). In this study, *E. coli* JM108 was chosen as the host strain, which is *recA* deficient, and transductions and transformations were performed at 30°C to minimize the potential problems. It was also suggested that plating and screening the gene library be done immediately, because a proportion of *E. coli* cells containing cosmids lose viability.

5.2. Architecture of the *mid* gene cluster in *S. mycarofaciens*

DNA sequencing of two cosmid plasmids, Smyc-LC1 and Smyc-LC3, allowed the identification of PKS genes, lactone and sugar modification genes, resistant and regulatory genes, and sugar biosynthetic genes, which are all necessary for midecamycin biosynthesis (Tab. 5.1 and Tab. 5.2). Up to now, the *mid* gene cluster has been obtained from about 74 kb in the genome of *S. mycarofaciens*. The gene organisation in the midecamycin producer, *S. mycarofaciens*, at the beginning seemed to be quite similar to that of the tylosin-producing *S. fradiae* in the large central region. Both gene clusters demonstrate that the large PKS genes (*midG* or *tylG*) are flanked by two regions containing genes encoding enzymes for the biosynthetic deoxysugars, as well as resistance and regulatory enzymes (for midecamycin biosynthesis: Hara & Hutchinson, 1990; 1992; this study; for tylosin biosynthesis: Baltz & Seno, 1988; Birmingham *et al*., 1989; Zalacain & Cundliffe, 1991; Merson-Davies *et al*., 1994; Gandecha *et al*., 1997; Bate *et al*., 1999; Buter *et al*., 1999). As shown in Fig. 5.1, the gene order for the mycaminose biosynthetic genes and some others for lactone ring modification and regulation was found to be conserved between both clusters, but the transcription patterns were different in part.

However, the location of genes involved in the mycarose biosynthetic pathway turned out to be different between the *mid* and *tyl* (tylosin biosynthetic) gene clusters. The five genes (*tylCII*, *tylCIII*, *tylCIV*, *tylCV* and *tylCVII*) are located downstream of *tylG*. In contrast,

Tab. 5.1. Summary of the genes found in Smyc-LC1

Orf in Fig. 4.10	Gene	Coding capacity		
		aa	Putative function	Proteins with identity score[3]
orf1	*midM*[1]	{388}	cytochrome P450	TylI: 57%; PicK: 32%
orf2	*midC*	398	aminotransferase	OleN2: 49%; TylB: 47%; EryCI: 47%; DesV: 45%
orf3	*midA*	303	dTDP-glucose synthase	DesIII: 62%; TylAI: 61%; AveBIII: 61%; SnogJ: 56%
orf4	*midB*	326	dTDP-glucose 4,6-dehydratase	TylAII: 64%; AveBII: 62%; Gdh: 60%; DesIV: 60%
orf5	*midD*	264	thioesterase	TylO: 44%; PikAV: 39%; Rif-ORF12: 43%
orf6	*midE*	388	4"-*O*-propionyl transferase	AcyB1: 51%
orf7	*midF*[2]	{228}	regulator	SrmR: 49%
orf8	*midL*[1]	{372}	2,3-dehydratase	TylCVI: 50%; OleV: 48%; EryBVI: 42%

Location at site No. in Fig. 4.10	Temporary gene name	Coding capacity		
		DNA length (kb)	Putative function of deduced aa seq.	Proteins with identity score
1-3	*mid1-4*	1.8	unknown	
15-16	*midO*	0.6	4-ketoreductase	TylCIV: 50% {in 54 aa}
17-18	*mid1-6rp*	0.55	dehydratase	OleN2: 72%; EryCIV: 67% {in 132 aa}
17-18	*mid1-6up*	0.55	acyl-CoA dehydrogenase	SCI30A.22c: 44% {in 95 aa}
18-19	*mid1-3up*	0.6	pyruvate dehydrogenase	PdhA: 41% {in 175 aa}
18-19	*mid1-3rp*	0.8	unknown	

(1) Partial sequence of orfs, incomplete at the 3' end. (2) Partial sequence of orf. (3) AcyB1 from *S. thermotolerans* (D31821); AveBII and AveBIII from *S. avermitilis* (AB032523); DesIII, DesIV and DesV (AF079762), and PicK (AF087022) from *S. venezuelae*; EryCI (X155541), EryBVI (Y11199) and Gdh (L37354) from *Sac. erythraea*; OleN2 and OleV from *S. antibioticus* (AF055579); PdhA (L31844) from *Thiobacillus ferrooxidans*; Rif-ORF12 (AF040571) from *Amycolatopsis mediterranei*; SCI30A.22c (AL096811) from *S. coelicolor*; SnogJ from *S. nogalater* (AF187532); SrmR from *S. ambofaciens* (X63451); TylI, TylAI, TylAII, TylB and TylO (U08223), TylCIV (U77459) and TylCVI (AF210634) from *S. fradiae*. Acc. no. are given in *parentheses*. The aa numbers in partial protein sequence are indicated in *braces*.

Tab. 5.2. Summary of the genes found in Smyc-LC3

Orf in Fig. 4.10	Gene	Coding capacity		
		aa	Putative function	Proteins with identity score[3]
orf11	*midG*[1]	{133}	PKS module 7?(3'-end)	NidA5: 60%; TylGV: 48%
orf12	*midK*	249	N-methyltransferase	TylMI: 60%; OleM1: 57% DesVI: 53%; EryCVI; 51%
orf13	*midH*	414	3,4-isomerase	TylMIII: 35%; DnrQ: 32%;
orf14	*midI*	421	glycosyltransferase	TylMII: 59%; DesVII: 55%; OleG2: 54%; EryCIII: 51%
orf15	*midJ*	448	crotonyl-CoA reductase	Ccr: 75%; SC9C7.09c: 75%
orf16	*mdmB*[2]	387	3-*O*-acyltransferase	AcyA: 59%
orf17	*midN*[1]	{109}	cytochrome P450	Car-orfP450: 56%
orf18	*mirB*[1]	{135}	ABC-transporter	CarA: 67%; SrmB: 63%
orf19	*mdmA*[2]	271	resistance	TlrD: 54%
orf20	*midR*[1]	{140}	regulator	OrfY:36%; SC7H1.21:34%
orf21	*mdmC*[2]	221	*O*-methyltransferase	Car-orfB: 51%

Location at site No. in Fig. 4.10	Temporary gene name	Coding capacity		
		aa	Putative function	Proteins with identity score
40-41	*midG*[1]	{265}	PKS module7?	NidA5: 55%; TylGV: 43%
55-56	*mid3-19B*[1]	{243}	3-oxoadipate enol-lactone hydrolase/4-carboxymucono-lactone decarboxylase	SC3A7.07: 61%
56-57	*mid3-19A*[1]	{246}	hypothetical protein putative exonuclease	SC3A7.08: 72% {in 124 aa} SC3A7.09: 78% {in 101 aa}
57-58	*mid3-7*[1]	{250}	glutamate uptake system ATP-binding protein	SC4H8. 16c (GluA): 73% {in 147 aa}

(1) Partial sequence of orfs. (2) The genes, *mdmA* (A60725), *mdmB* and *mdmC* (M93958) were identified elsewhere (see Section 4.9-10). (3) AcyA, Car-orfB and Car-orfP450 (D30759), and CarA (M80346) from *S. thermotolerans*; DesVI and DesVII from *S. venezuelae* (AF079762); DnrQ from *S. peucetius* (L47164); EryCIII (Y14332) and EryCVI (U77459) from *Sac. erythraea*; NidA from *S. caelestis* (U78289); OleG2 and OleM1 from *S. antibioticus* (AJ002638); OrfY from *S. hygroscopicus* (X86780); SC3A7.07, SC3A7.08 and SC3A7.09 (AL031155), SC4H8.16c (AL020958), and SC9C7.09c (AL035161) from *S. coelicolor*; TylG (U78289), TylMI, TylMII, TylMIII and ccr (X81885), and TlrD (X97721) from *S. fradiae*. Acc. no. are given in *parentheses*. The aa numbers in partial protein sequence are indicated in *braces*.

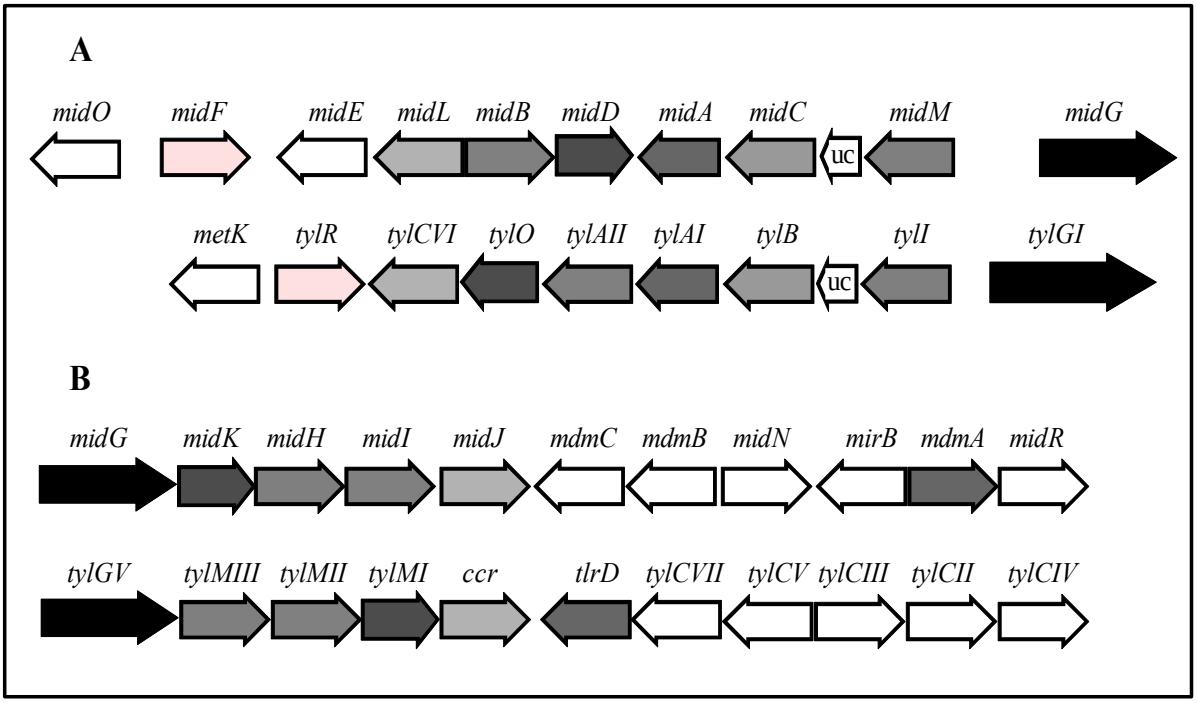


Fig. 5.1. Comparison of part of the gene clusters of midecamycin and tylosin. (A) The region of upstream of PKSs (*midG* and *tylGI*). (B) The region of downstream of PKSs (*midG* and *tylGV*). Genes are not drawn to scale. Genes having the same function from the *mid* and *tyl* clusters are drawn by arrow bars with the same dark colour or pattern, in which uc means uncoding sequences, and the deduced gene functions from the *mid* cluster were listed in Tab. 5.1 and Tab. 5.2. The white arrow bars indicate various genes which are specific in the *mid* or *tyl* clusters, in which *tyl*-orf9 and *metK* are not published the sequence data, *tylCII* encodes a 2,3-enoyl reductase, *tylCIII* encodes a 3-C-methyltransferase, *tylCIV* encodes a 4-ketoreductase, *tylCV* encodes a mycarosyl transferase and *tylCVII* encodes a 3,5- (or 5-) epimerase (acc. no. AF147704), and the *mid* genes see Tab. 5.1 and Tab. 5.2.

a partial sequence *midO* (*tylCIV*-like) was found on the cosmid Smyc-LC1, upstream of *midG* (Fig. 5.1). Analogous to *tylC*-gene organisation, the further genes encoding mycarose biosynthetic enzymes are expected to locate in the region of the 4.2 kb *Bam*HI-*Bgl*II DNA fragment (site16-17 in Fig. 4.10), next to the *midO* gene. Very recently, this fragment was partially sequenced by a Cambridge group (Leadlay, personal communication), confirming that it contains a set of genes for mycarose biosynthesis.

As seen in Tab. 5.1, a putative dehydratase gene, temporarily named as *mid1-6rp*, was found in the Smyc-LC1 sub-cluster at the site 17-18 of Fig. 4.10. The deduced product of this gene showed high identities with OleN2 and EryCIV, in which EryCIV might act as a

pyridoxal phosphate-dependent dehydratase to catalyse the removal of a water molecule at C3-C4 during the biosynthesis of desosamine in *Sac. erythraea* (Gaisser *et al*., 1997), and OleN2 might be involved in the course of transamination for the biosynthesis of the desosamine precursor of oleandomycin in *S. antibioticus*. However, since the structure of mycaminose differs from that of desosamine in that no C-4 hydroxyl elimination has to take place, therefore, the finding of *mid1-6rp* gene in the *mid* gene cluster is currently beyond a useful explanation.

In the sub-cluster contained on Smyc-LC3, the region (4 kb or larger) downstream of *midR* was found to host genes for which the deduced polypeptide products are highly similar in both structure and order to those of so-called house-keeping genes from the *S. coelicolor* genome (see Tab. 5.2). These genes in the vicinity of the *mid* gene cluster might play an important role in the primary metabolism for the synthesis of the sugar or lactone precusors before or during midecamycin biosynthesis. Another, more likely, explanation would be that the whole *mid*-cluster has been inserted into a control region of the chromosome containing essential genes and, therefore, might be more stable than the genes at the chromosome ends (Lin & Chen, 1997; Volff *et al*., 1997).

The *mid* gene cluster was successfully probed by two homologous genes, *midB* (encoding a dTDP-glucose 4,6-dehydratase) and *midI* (encoding a glycosyltransferase). These two genes were first amplified by PCR and then chosen as probes to screen the cosmid library of *S. mycarofaciens*. This basic strategy was considered mainly on the basis of the published data of the *tyl* gene cluster, since the two genes, *tylB* (as *midB*) and *tylMII* (as *midI*) are located at the two sides of the PKS sub-cluster. The data obtained have confirmed that the *midB* and *midI* genes in the *mid* cluster have the same order as in the *tyl* cluster, and this could be understood as both antibiotics are structurally closely-related 16-membered macrolides. This study could suggest that the *midB*- or *tylB*-like genes, and the *midI*- and *tylMII*-like genes can now be used as further hybridization probes with much higher safety to quickly isolate other 16-membered macrolide biosynthetic clusters. Since the dTDP-glucose 4,6-dehydratase is a common enzyme catalysing in the biosynthetic pathways of mycaminose, mycarose and other 6-deoxysugars, and glycosyltransferase for the macrolide aminosugar is necessary for sugar transfer to the aglycone directly, i.e., the specific macrolactone ring.

5.3. Hybrid antibiotic production by heterologous complementation

Recombinant DNA techniques now make it possible to produce new antibiotic structures by engineering the biosnythetic pathways of antibiotics. This experimental approach is called combinatorial biosynthesis and the substances thus produced are called hybrid antibiotics. One of the methods to reach this goal is to transform the specific genes into heterologous host strains, express and recombine them there in new combinations. Although there were no antibiotic activities detected when the *midK* and *midI* genes were transformed into the *Sac. erythraea* mutants CIII68 and CIV89 (see Section 4.18), it is a preliminary trial towards hybrid antibiotic production. Moreover, the finding of the *mid* gene cluster and the analysis of gene functions carried out in this study have opened several opportunities for further efforts to produce novel antibiotics. (1) Three acylation genes, *mdmB*, *mdmC* and *midE*, could be used to modify the activity of macrolide antibiotics by acylating hydroxyl groups on both lactone and sugar moieties, as was demonstrated earlier engineering of *carE* from *S. thermotolerans* into *S. thermotolerans* and *S. lividans* (Epp *et al*., 1989). For example, if *mdmB* (3-*O*-acyltransferase), or/and *midE* (4"-*O*-propionyltransferase) could be introduced into the tylosin-producing strain, *S. fradiae*, which does not acylate tylosin, it is expected that the recombinants would produce 3-*O*-acetyltylosin, or 4"-*O*-propionyltylosin or 3-*O*-acetyl 4"-*O*-propionyltylosin, respectively. (2) The *midI* gene encodes a glycosyltransferase possibly responsible for transfer of mycaminose to the midecamycin aglycone. Recently, several experiments demonstrated remarkably relaxed specificities of other glycosyltransferases towards both their sugar and aglycone substrates. For instance, Xue *et al*. (1998) found that the *desVII* gene could accept both 10-deoxymethynolide, the macrolactone of the 12-membered macrolides methymycin and neomethymycin, and narbonolide, the macrolactone of the 14-membered macrolides narbomycin and pikromycin, as substrates to add the desosamine moiety to their aglycones in *S. venezuelae*. Moreover, Borisova *et al*. (1999) revealed that the glycosyltransferase (DesVII) of the desosamine biosynthetic pathway is capable of recognizing the keto sugar intermediate, the macrolide product produced by a *des* mutant, to attach it to the macrolactone of methymycin. Therefore, it is conceivable that the MidI protein as a component of the condensing machinery in macrolide biosynthesis could further explore its application, including carrying sets of diverse nucleotide sugars, and attaching them to different targets, such as 14- and 16-membered macrolide aglycones, or even an already

glycosylated macrolide. Protein engineering even could widen the specificity of MidI, or hybrids thereof, for the acceptance of completely different aglycones and/or NDP-sugars.

5.4. Heterologous enzyme expression and combined enzymatic analysis

In order to achieve the soluble proteins of MidH, MidC, MidK and MidI, each in their active forms, different expression systems in *E. coli* and *S. lividans* were tested and the interpreted results are summarised in Tab. 5.3. To optimize the expression conditions, lower temperature (28°C), and slow shaking rate (100 r.p.m.) and different media were compared for expression of MidH, MidC, MidK and MidI in *E. coli* BL21(DE3)pLysS. As a successful example the native soluble protein of MidH could be obtained when the shaking rate was decreased from 200 r.p.m. to 100 r.p.m. during expression cultivation (cf. Fig. 4.18A). This possibly causes change in the cytoplasmic environment for expression and especially could reduce the rate of protein synthesis and, therefore, could facilitate its correct folding. When the medium of LB/sorbitol/betaine was used instead of LB medium for expression of the four proteins, the production in all cases was significantly increased (see Tab. 5.3). This osmotic stress culture system could decrease growth rates because of reduction of the carbon dioxide level, however, it can facilitate the uptake of the 'compatible solute', glycyl betaine, which is believed to cause thermodynamically unfavourable 'preferential hydration' if excluded from the immediate domains of proteins and, thus minimisation of solvent-protein contact and stabilisation of protein structure could be the result (Arakawa & Timasheff, 1985; Blackwell & Horgan, 1991). However, the expression of MidK and MidI in *E. coli* resulted in strong production of inclusion bodies under various conditions of manipulation tested. Further efforts will have to be made to recover the active proteins of MidK and MidI in *E. coli*, e.g., by applying other expression strategies, including co-expression of molecular chaperones (Martin & Hartl, 1997), using thioredoxin-deficient host strains to maintain a favorable redox potential and facilitate disulfide bond formation in the cytoplasmic environment, and/or using soluble polypetide products of the host as fusion partners (Makrides, 1996; Hannig & Makrides, 1998). Expression of the four proteins failed by using the pUWL201 vector and insertion at the 3' end of the *ermE* up promoter in *S. lividans* TK23. One of the possible reasons could be the structure of the ribosomal binding sites (RBS) of these recombinant expression-

plasmids which were taken from that of the pET-cassette. This RBS might not be suitable for the translation of mRNA to protein in *Streptomyces*.

Tab. 5.3. Summary of protein expression in different expression systems

Protein	Expression plasmids						
	pET11a		pET16b		pUWL201		pIJ4123 (His-tag)
	M1	M2	M1	M2	(native)	(His-tag)	
MidH	++/s	+++/s	+++/is	++++/s	-	- w/-	
MidC	++++/s		+++/s		-	- w/-	
MidK	+++/is	++++/is	+++/is w/+	++++/is w/+	-	- w/-	- w/+
MidI	+++/is	++++/is	+++/is w/+	++++/is w/+	-	- w/-	- w/+

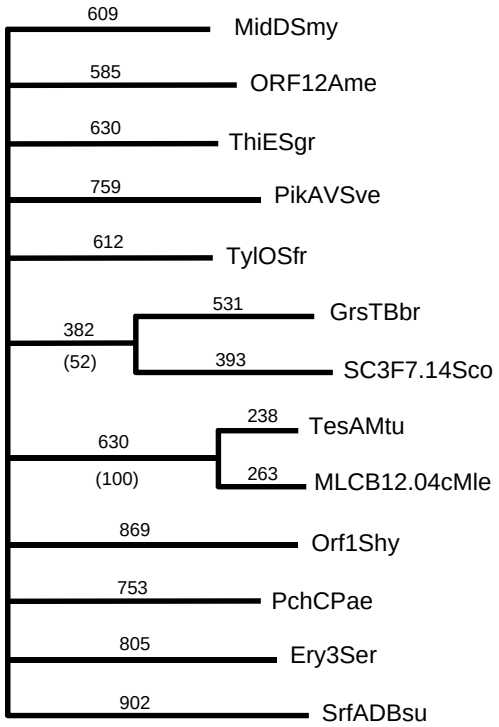
++ = protein overexpression, the production level is increased with more plus; - = no protein expression on SDS-PAGE gels; ++/s or ++/is = soluble or insoluble proteins were observed on SDS-PAGE gels, respectively; w/+ = soluble proteins were detected by Western boltting; w/- = no expected proteins were detected by Western boltting. M1: LB liquid medium; M2: LB with sorbitol and betaine.

To characterise the postulated function of MidH as an isomerase and MidC as an aminotransferase in the pathway of mycaminose biosynthesis, a coupling enzymatic reaction of RmlB, MidH and MidC was analysed by HPLC and followed by LC-MS. It was confirmed that: (i) RmlB as a 4,6-dehydratase is to convert the substrate dTDP-D-glucose to the intermediate dTDP-4-keto-6-deoxy-D-glucose (Verseck, 1997); and (ii) MidC is indeed responsible for transamination to form dTDP-amino-6-deoxy-D-glucose (see Section 4.20.2). But it is not clear at this point whether this transamination occurs at C-3 or C-4 since the function of MidH was not individually elucidated. To further characterise the specific function of MidC in the future, it is suggested that all three enzymes have to be purified to reduce non-specific products produced. For instance, dTDP-L-rhamnose was detected as a by-product in the reaction catalysed by the mixture of the crude extracts containing the above three enzymes. The MidK (N-methyltransferase) protein was also combined with RmlB, MidH and MidC to assay the reaction products. However, there was no measurable transfer of methyl groups to the products. The possible reasons were either because the soluble MidK protein was produced at very low levels (see Fig. 4.20), or the protein is inactived.

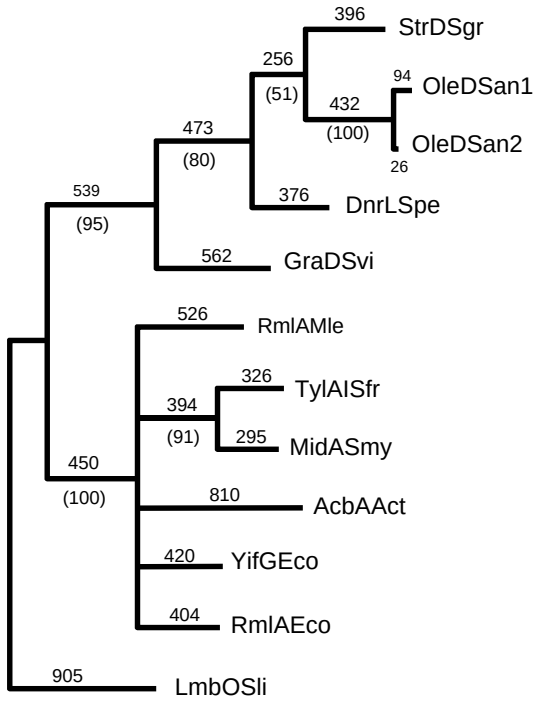
5.5. Phylogenetic analysis of some enzymes involved in midecamycin biosynthesis

A detailed analysis of the phylogenetic relationships of several of the *mid*-encoded proteins within the same family of proteins was carried out in order to obtain further information about how the secondary metabolic gene clusters, especially for macrolides, have been evolved. In Fig. 5.2, seven phylograms are re-produced in which the MidD-tree deals with thioesterases, affecting the product accumulation in the producing strains, and the other six trees concern proteins which are involved in *de novo* mycaminose biosynthetic enzymes.

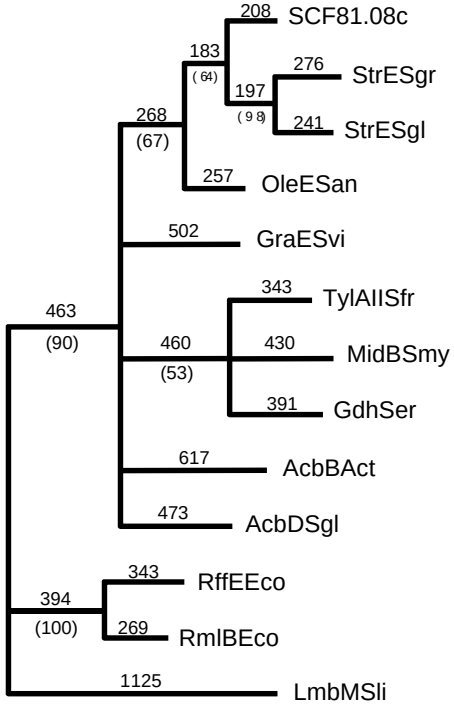
Fig. 5.2. Phylogenetic trees (non-rooted) of the proteins of MidD, MidA, MidB, MidC, MidH, MidK and MidI aligned with the protein sequences of the corresponding family. The phylograms were calculated by the program PAUP 3.1. The *numbers* denote the calculated lengths of branches; The *numbers* in *parentheses* give the percent score of nodes found in bootstrapping (500 times). For the amino acid sequence alignments see Appendices 7.6 to 7.12. The following sequences were used and identified by their accession numbers in *parentheses*. **MidD-tree:** ORF12Ame: *Amycolatopsis mediterranei* (AF040571); ThiESgr: *S. griseus* (M93058); PikAVSve: *S. venezuelae* (AF079138); TylOSfr: *S. fradiae* (U08223); GrsTBbr: *Bacillus brevis* (M29703); SC3F7.14Sco: *S. coelicolor* (AL021409); TesAMtu: *Mycobacterium tuberculosis* (Z74697); MLCB12.04cMle: *M. leprae* (AL035480); Orf1Shy: *S. hygroscopicus* (M64783); PchCPae: *Pseudomonas aeruginosa* (X82644); Ery3Ser: *Sac. erythraea* (X60379); SrfADBsu: *Bacillus subtilis* (X70356). **MidA-tree:** StrDSgr: *S. griseus* (X62567); OleDSan1: *S. antibioticus* (AF055579); DnrLSpe: *S. peucetius* (L47163); GraDSvi: *S. violaceoruber* Tü22 (AJ011500); RmlAMle: *M. leprae* (L78815); TylAISfr: *S. fradiae* (U08223); AcbAAct: *Actinoplanes sp.* (Y18523); YifGEco: *E. coli* (U00096); RmlAEco: *E. coli* (AF125322); LmbOSli: *S. lincolnensis* (X79146). **MidB-tree:** SCF81.08c: *S. coelicolor* (AL133171); StrESgr: *S. griseus* (X62567); StrESgl: *S. glaucescens* (AJ006985); OleESan: *S. antibioticus* (AF055579); GraESvi: *S. violaceoruber* Tü22 (AJ011500); TylAIISfr: *S. fradiae* (U08223); GdhSer: *Sac. erythraea* (L37354); AcbBact: *Actinoplanes sp.* (Y18523); AcbDSgl: *S. glaucescens* (Fa. Hoechst. Dt. Patentanmeldung Nr. 19622783 [1996]); RffEEco: *E. coli* (M87049); RmlBEco: *E. cloi* (AF125322); LmbMSli: *S. lincolnensis* (X79146). **MidH-tree:** DnrQSpe: *S. peucetius* (L47164); TylMIIISfr: *S. fradiae* (X81885); OleP1San: *S. antibioticus* (AJ002638); DesVIIISve: *S. venezuelae* (AF079762); SnogNSno: *S. nogalater* (AF187532); Pra10Ahi: *Actinomadura hibisca* (D87924); **MidC-tree:** StsASgr: *S. griseus* (Y08763); StsCSgr: *S. griseus* (Y08763); StrSSgr: *S. griseus* (Y00459); DnrJSpe: *S. peucetius* (M80237); TylBSfr: *S. fradiae* (U08223); OleN2San: *S. antibioticus* (AF055579); DesVSve: *S. venezuelae* (AF079762); EryCISer: *Sac. erythraea* (X155541); EryCIVSer: *Sac. erythraea* (U77459); OleN1San: *S. antibioticus* (unpublished data); PerSVch: *Vibrio cholerae* (X59554); LamSSli: *S. lincolnensis* (X79146). **MidK-tree:** OleM1San: *S. antibioticus* (AJ002638); EryCVISer: *Sac. erythraea* (U77459); StsGSgr: *S. griseus* (Y08763); GlyMRa: glycine methyltransferase from rat (X07833); LmbJSli: *S. lincolnesis* (X79146). **MidI-tree:** OleG1San: *S. antibioticus* (AJ002638); OleG2San: *S. antibioticus* (AJ002638); EryCIIISer: *Sac. erythraea* (Y14332); TylMIISfr: *S. fradiae* (X81885); EryBVSer: *Sac. erythraea* (U77459); TylNSfr: *S. fradiae* (AJ005397).



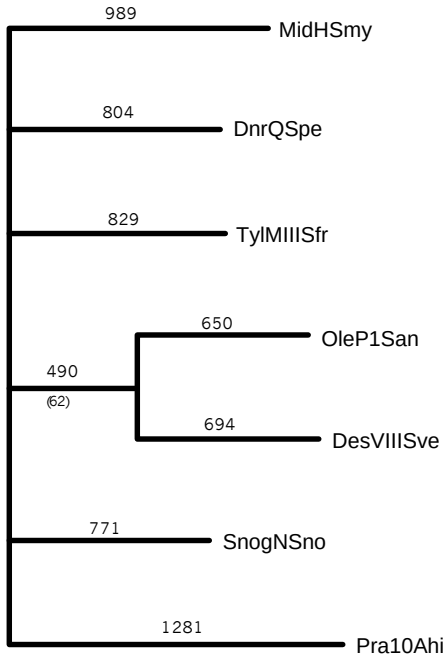
MidD-tree



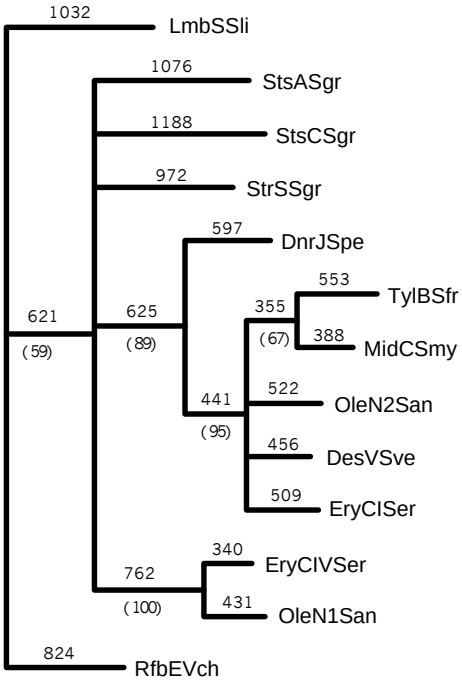
MidA-tree



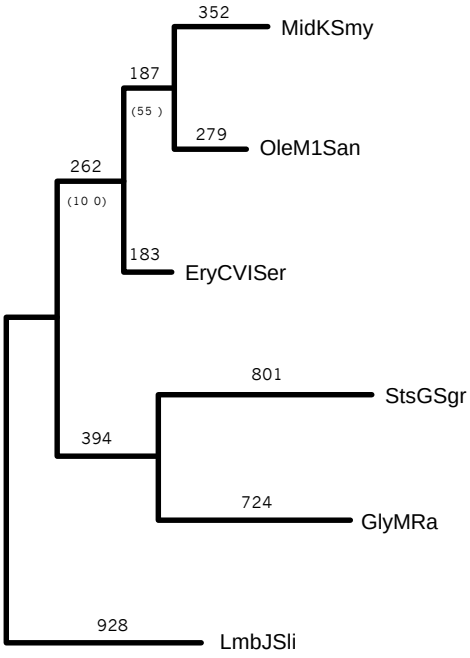
MidB-tree



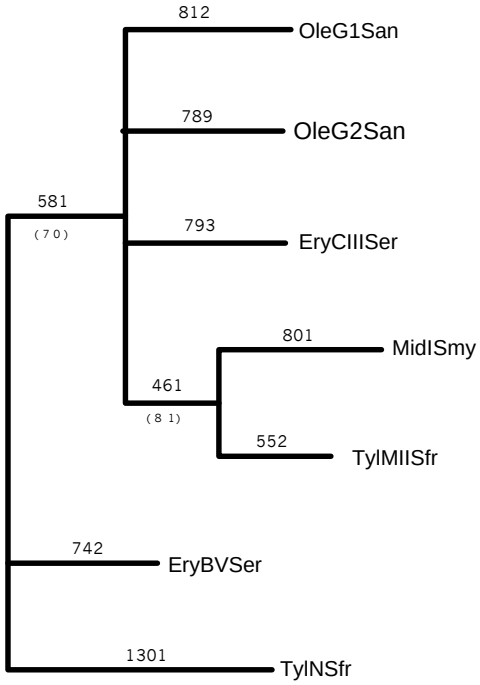
MidH-tree



MidC-tree



MidK-tree



MidI-tree

The second thioesterase (MidD): The polypeptide product of *midD* was shown to be highly identical to the second thioesterase encoded by the *tylO* gene in tylosin biosynthesis of *S. fradiae* (Merson-Davis *et al*., 1994) (see Tab. 5.1), and both genes are located shortly upstream of *midG* and *tylG*, respectively. As with other type I PKS, the enzyme catalysing the final enlongation step (TylGV) in *S. fradiae* contains an integral C-terminal thioesterase domain that is believed to be responsible for chain termination and ring closure to produce tylactone; also, the aa sequence is highly identical to the C-terminal segment of the midecamycin PKS, MidG (see Section 4.7). The question then arises what is the function of MidD or TylO, these second (or free) thioesterases in macrolide biosynthesis. Bulter *et al*. (1999) recently demonstrated that this free thioesterase activity of TylO is necessary for bulk accumulation of tylosin in *S. fradiae* and that at least 85% of antibiotic production is lost when TylO is inactivated. Therefore, they proposed that the TylO protein carries out an editing function, whereby aberrant polyketide precusors blocking the synthase complex can be removed to allow continued synthesis of normal products. Such free thioesterase genes also have been found in the gene clusters of other macrolide biosynthesis, such as *orf5* in *Sac. erythraea* (Haydock *et al*., 1991), *pikV* in *S. venezuelae* (Xue *et al*., 1998), as well as the free thioesterase gene in *Bacillus subtilis* during non-ribosomal peptide biosynthesis (Schneider & Marahiel, 1998). However, suprisingly, the results of the phylogram showed a close correlation between several sources of organisms (MidD-tree in Fig. 5.2). This could be explained in that either, all of the genes have had a high rate of evolution, or, the DNA fragments have horizontally been transferred between species which probably could reduce the enzyme specificity. In part recent horizontal gene transfer and sometimes recent gene duplication and divergence possibly resulted in two branches, GrsTBbr and SC3F7.14Sco, and TesAMtu and MLCB12.04cMle, respectively.

The dTDP-D-glucose synthase (MidA): This enzyme has been studied in several cases and catalyses the activation of nucleotidyl glucoside in the biosynthetic sequence for 6-deoxysugars. In the comparison of the aa sequences of 12 proteins in this family, the MidA-tree in Fig. 5.2 revealed that overall there are two principal subgroups, when LmbO, with clearly having a different function, was used as an outgroup. It is indicated in the first group that MidA and TylAI which are the enzymes for 16-membered macrolide sugar biosynthesis in *Streptomyces*, and AcbA which is the enzyme for the biosynthesis of the α–glucosidase inhibitor acarbose in *Actinoplanes*, are closer to three other proteins from *E.*

coli and *Mycobacteria*. However, MidA and TylAI are distinct from the second group, i.e., enzymes for the same function in the biosynthesis of 14-membered macrolides (OleD), aminoglycosides (StrE), and anthracyclines (DnrL) in *Streptomyces* species. This indicates that all of these genes, though being of common ancestry, have undergone divergence in at least the *Actinomycetes* after horizontal gene migration. Therefore, the relationships seem to be distributed non-taxonomically. In contrast to all other proteins in the MidA-tree, the LmbO protein was proposed to use dTDP-octose instead of dTDP-glucose as substrate for lincomycin biosynthesis in *S. lincolnensis* (Peschke *et al*., 1995; Piepersberg, personal communication). Thus, the changed function to accept different substrate probably led to fast evolution of the gene in the producers of lincosamids.

The dTDP-glucose 4,6-dehydratase (MidB): This enzyme is characterised as for conversion of dTDP-D-glucose into dTDP-4-keto-6-deoxyglucose, a common intermediate for most 6-deoxysugars. This is a catalytically NAD^+-dependent redox reaction for water elimination (Wierenga *et al*., 1985; Scrutton *et al*., 1990). As shown in MidB-tree of Fig. 5.2, it is obvious that the enzymes in the first large group including 10 members of *Actinomycetes* sequences are distinct from the second group, two enzymes from *E. coli*. Multiple branches in the first group may be of recent origin or the genes have recently evolved at different rates. As a result, the MidB, TylAII and Gdh proteins are closely-related to each other which are all likely to be involved in macrolide biosynthesis. However, it is important to note, (1) the *gdh* gene was not found within the erythromycin gene cluster of *Sac. erythraea*, which lacks the dTDP-glucose 4,6-dehydratase gene, (2) the *midB* gene is separated by another gene, *midD*, from *midA*, which is in contrast to the organisation of the equivalent genes in the *tyl* cluster (cf. Fig. 5.1; the *midBD* unit might have been inverted later after branching off from a common ancestor in which the gene order was as in the *tyl* cluster, *tylAI,AII,O*), (3) the OleE protein is more related to those proteins in another branch of actinomycete members, though being used in an erythromycin-like macrolide pathway. This allows the following general conclusions: (i) the genes for the more frequently used first- and second-step enzymes (e.g., MidA, MidB) in dTDP-6-deoxysugar pathways are not necessarily needed as inside components (strongly co-regulated) of the gene clusters (as are more specific genes); (ii) these genes, when present inside the biosynthetic clusters, do not necessarily co-evolve with the cluster. Rather they can be easily gathered from various different sources, probably by horizontal

gene transfer in most cases; (iii) their presence and conservation of these genes are kept under selective pressure only by the end product(s) which are formed from the pathway(s).

The differently functioning LmbM protein was used as an outgroup again in the MidB-tree and confirmed that in the lincomycin biosynthesis, a biosynthetic unit, α-methylthiolincosaminide (MTL), utilized a starting sugar other than D-glucose.

The 3,4-isomerase (MidH): The aa sequence comparison of *midH* to other proteins in the GenBank showed that the top scores of identities to TylMIII and DnrQ are only 35% and 32%, respectively. Although these proteins showed end-to-end sequence similarity to cytochrome P450 enzymes from a variety of organismal sources, they all lack the critical cysteine residue which provides the characteristic ligand to the heme iron in cytochrome P450 (Bairoch, 1992; Nelson *et al*., 1993). This fact suggests that they represent a new family of proteins with unknown enzymatic mechanism though they are likely to be derived from cytochrome P450 proteins. The MidH-tree in Fig. 5.2 revealed that these proteins from *Actinomycetes* seem not to need high sequence conservation. Therefore, rapid divergence could cover to form a new family of enzymes, probably because very few or none of the aa-motifs needed to be conserved for the mechanism. These genes may have had a long period of evolution and during this period they have acquired their species specificity by rapid horizontal gene exchanges to add up to the overall picture of a family of proteins with the low conservation level and no heme-binding motif. It is clear from the recently reported data (Gandecha *et al*., 1997; Salah-Bey *et al*., 1998; Madduri *et al*., 1998) that these proteins are 6-deoxy-hexose 3,4-isomerases involved in 3-aminosugar biosynthesis in several gene clusters. Therefore, the still unknown mechanism of action and the high divergence rate will have to be further analysed in order to interpret this result.

The aminotransferase (MidC): The polypeptide product of *midC* highly resembles the putative aminotransferases from several deduced gene products in antibiotic biosynthesis, so-called secondary metabolic aminotransferases (SMAT). The phylogenetic tree of 13 members of proteins in the SMAT family (MidC-tree in Fig. 5.2) demonstrated the following. (1) The proteins MidC, TylB, EryCI, OleN2 and DesV for transamination in macrolide biosynthesis are more closely-related than the proteins from other *Streptomyces* species. In addition, the MidC and TylB proteins of 16-membered macrolides are much

closer than the proteins of 12- or 14-membered macrolides. (2) The StsA, StsC and StsS proteins are more similar to one another, where StsC acts as the L-glutamine:*scyllo*-inosose aminotransferase to catalyse the first amino transfer in the biosynthesis of the streptidine subunit of streptomycin, and StsA might catalyse the second step of the cyclitol transamination reaction in *S. griseus* (Ahlert *et al*., 1997). The EryCIV and OleN1 proteins are distinct from the proteins of the above macrolide group because they might act as pyridoxal-phosphate-dependent dehydratases during desosamine biosynthesis in the macrolide production as explained in Section 5.2. (3) The branch of the LmbS protein in this phylogram was again used as an outgroup because it is a putative ketooctose aminotransferase (Peschke *et al*., 1995; Piepersberg, personal communication). Taken together, the macrolide sugar aminotransferases here behave as would be expected from a common brand of PKS and sugar genes (in contrast to the other trees above).

The methyltransferase (MidK): The MidK protein shows significant sequence similarity to a family of SMA-dependent N-methyltransferases from several *Streptomces* species. A mini-phylogram (MidK-tree in Fig. 5.2) showed that there are two subgroups, a macrolide group and the other group in addition to one outgroup (LmbJ). In the first group, the relationship between MidK and OleM1 is surprisingly closer than that between EryCVI and OleM1. The StsG protein from *S. griseus* is close to GlyM, a glycine methyltransferase from rat (Ogawa *et al*., 1987). This reflects the common ancestry on one hand, but also the random and non-predictable horizontal gene transfer between various organisms.

The glycosyltransferase (MidI): The MidI protein highly resembles various glycosyltransferases that are involved in the transfer of various deoxysugars to different aglycones during antibiotic biosynthesis. The MidI-tree (see Fig. 5.2) revealed that, as is expected, the relationship between MidI and TylMII, which are the putative mycaminosyltransferases, is closer than that between EryCIII and OleG2, which are desosaminyltransferases. It is understood that the EryBV protein is distinct from the above four proteins because it is a putative mycarosyltransferase. The OleG1 protein is responsible for oleandrose transfer in the biosynthesis of oleandomycin in *S. antibioticus* (Olano *et al*., 1998). However, OleG1 is closer to OleG2 and EryCIII in the MidI-tree. This could be explained in that the *oleG1* and *oleG2* genes are adjacent in the oleandomycin gene cluster of *S. antibioticus*, and one gene was derived from the other via recent gene

duplication and subsequent functional divergence. The TylN protein is responsible for 6-deoxyallose transfer in tylosin biosynthesis of *S. fradiae* (Wilson & Cundliffe, 1998), therefore, it is less-closely related to the TylMII and MidI proteins.

Based on analysis of the deduced functions of the above proteins and mechanistic considerations, the biosynthetic route to dTDP-D-mycaminose, a precursor of midecamycin in *S. mycarofaciens* is proposed as given in Fig. 5.3, in which the proposed action of the MidH protein (3,4-isomerase) is still hypothetical, since it is the only remaining essential gene product for aminosugar biosynthesis, which could have that function.

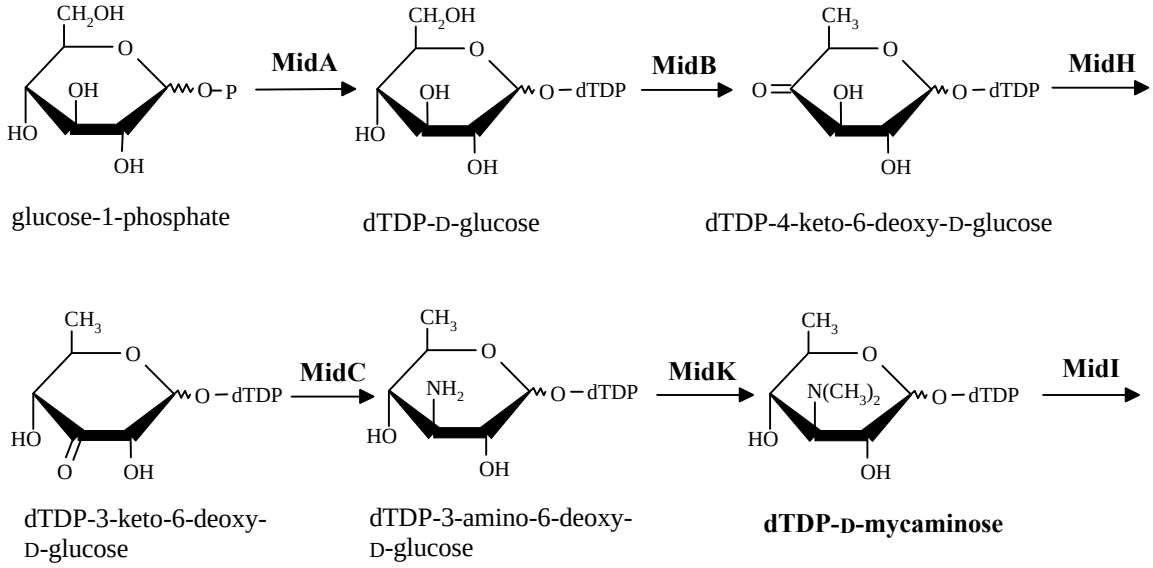


Fig. 5.3. The proposed biosynthetic route to mycaminose and the *mid*-encoding proteins deduced to catalyse the various steps.

6. References

Ahlert J, Distler J, Mansouri K, Piepersberg W (1997) Indentification of *stsC,* the gene encoding the L-glutamine:*scyllo*-inosose aminotransferase from streptomycin-producing Streptomycetes. Arch. Microbiol. 168: 102-113.

Altschul SF, Gish W, Miller W, Myers EW, Lipmann DJ (1990) Basic logical alignment search tool. J. Mol. Biol. 215: 403-410.

Arakawa T, Timasheff SN (1985) The stabilization of proteins by osmolytes. J. Biophys. 47: 411-414.

Arisawa A, Kawamura N, Tsunekawa H, Okamura K, Tone H (1993) Cloning and nucleotide sequences of two genes involved in the 4''-*O*-acylation of macrolide antibiotics from *Streptomyces thermotolerans*. Biosci. Biotechnol. Biochem. 57: 2020-2025.

Arisawa A, Kawamura N, Takeda K, Tsunekawa H, Okamura K, Okamoto R (1994) Cloning of the macrolide antibiotic biosynthesis gene *acyA,* which encodes 3-*O*-acyltransferase, from *Streptomyces thermotolerans* and its use for direct fermentative production of a hybrid macrolide antibiotic. Appl. Environm. Microbiol. 60: 2657-2660.

Arnold A (1996) Klonierung und Screening von 6-Desoxyhexose-Genen in Makrolid-produzierenden Streptomyceten. Diploma thesis, Uni. Wuppertal.

August PR, Tang L, Yoon YJ, Ning S, Müller R, Yu TW, Taylor M, Hoffmann D, Kim CG, Zhang X, Hutchinson CR, Floss HG (1998) Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. Chemistry & Biology 5: 69-79.

Babcock MJ, Kendrick KE (1988) Cloning of DNA involved in sporulation of *Streptomyces griseus*. J. Bacteriol. 170: 2802-2808.

Baltz RH (1998) Genetic manipulation of antibiotic-producing *Streptomyces*. Trends in Microbiology 6: 76-83.

Baltz RH, Seno ET (1981) Properties of *Streptomyces fradiae* mutants blocked in biosynthesis of the macrolide antibiotic tylosin. Antimicrob. Agents Chemother. 20: 214-225.

Balz RH, Seno ET (1988) Genetics of *Streptomyces fradiae* and tylosin biosynthesis. Annu. Rev. Microbiol. 42: 547-574.

Bairoch A (1992) PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. 20: 2013-2018.

Bauer NJ, Kreuzman AJ, Dotzlaf JE, Yeh WK (1988) Purification, characterization, and kinetic mechanism of S-adenosyl-L-methionine:macrocin *O*-methyltransferase from *Streptomyces fradiae*. J. Biol. Chem. 263: 15619-15625.

Beckmann RJ, Cox K, Seno ET (1989) A cluster of tylosin biosynthetic genes is interrupted by a structurally unstable segment containing four repeated sequences. In: Hershberger CL, Queener SW, Hegeman G (eds). Genetics and Molecular Biology of Industrial Microorganisms. Washington, DC: American Society for Microbiology, p176-186.

Beppu T (1996) Genes, enzymes and secondary metabolites in industrial microorganisms. The 1995 Thom Award Lecture. J. Ind. Microbiol. 16: 360-363.

Bibb MJ, Cohen SN (1982) Gene expression in *Streptomyces*: construction and application of promoter-probe plasmid vectors in *Streptomyces*. Mol. Gen. Genet. 187: 265-277.

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

Birmingham VA, Cox KL, Larson JL, Fishman SE, Hershberger CL, Seno ET (1986) Cloning and expression of a tylosin resistance gene from a tylosin-producing strain of *Streptomyces fradiae*. Mol. Gen. Genet. 204: 532-539.

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7: 1513-1523.

Blackwell JR, Horgan, R (1991) A novel strategy for production of highly expressed recombinant protein in an active form. FESB Lett. 295: 10-12.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgramm quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72: 248-254.

Butler AR, Bate N, Cundliffe E (1999) Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*. Chem. Biology, 6: 287-292.

Chater K, Hopwood DA, Kieser T, Thompson C (1982) Gene cloning in *Steptomyces*. Curr. Top. Microbiol. Immunol. 96: 69-95.

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ (1993) Crystal structure of the Hha DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74: 299-307.

Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF (1990) An unusually large multifunctional polyketide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. Nature 348: 176-178.

Cortes J, Weissman K, Roberts G, Brown M, Staunton J, Leadlay P (1995) Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. Science 268:1487-1489.

Dhillon N, Hale RS, Cortes J, Leadlay PF (1989) Molecular characterization of a gene from *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) which is involved in erythromycin biosynthesis. Mol. Microbiol. 3: 1405-1414.

Dickens ML, Ye J, Strohl WR (1996) Cloning, sequencing, and analysis of aklaviketone reductase from *Streptomyces* sp. strain C5. J. Bacterial. 178: 3384-3388.

Distler J, **Manouri K, Piepersberg W (1985)** Streptomyces biosynthesis in *Streptomyces griseus* II: Adjacent genomic location of biosynthetic genes and one of two streptomycin resistance genes. FEMS Microbiol. Lett. 30: 151-154.

Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. Science 252: 675-679.

Donadio S, Katz L (1992) Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. Gene 111: 51-60.

Drew S, Demain AL (1978) Effect of primary metabolites on secondary metabolism. Annu. Rev. Microbiol. 31: 343-356.

Epp J, Huber M, Turner J, Goodson T, Schoner B (1989) Production of a hybrid macrolide antibiotic in *Streptomyces ambofaciens* and *Streptomyces lividans* by introduction of a cloned carbomycin gene from *Streptomycins thermotolerans*. Gene 85:293-301.

Fernandez E, Weißbach U, Reillo C, Brana A, Mendez C, Rohr J, Salas J (1998) Identification of two genes from *Streptomyces argillaceus* encoding glycosyltransferases involved in transfer of a disaccharide during biosynthesis of the antitumor drug mithramycin. J. Bacteriol. 180: 4929-4937.

Fishman SE, Cox K, Larson JL, Reynolds PA, Seno ET, Yeh WK, Van Frank R, Hershberger CL (1987) Cloning genes for the biosynthesis of a macrolide antibiotic. Proc. Natl. Acad. Sci. USA 84: 8248-8252.

Fouces R, Mellado E, Diez B, Barredo (1999) The tylosin biosynthetic cluster from *Streptomyces fradiae*: genetic organisation of the left region. Microbiology, 145: 855-868.

Gaisser S, Böhm G, Cortes J, Leadlay P (1997) Analysis of seven genes from the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*. Mol. Gen. Genet. 256: 239-251.

Gaisser S, Trefzer A, Stockert S, Kirschning A, Bechthold A (1997) Cloning of an avilamycin biosynthetic gene cluster from *Streptomyces viridochromogenes* Tü57. J. Bacteriol. 1997, 179: 6271-6278.

Gaisser S, Böhm G, Doumith M, Raynal M, Dhillon N, Cortes J, Leadlay P (1998) Analysis of *eryBI, eryBIII, eryBVII* from the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*. Mol. Gen. Genet. 258: 78-88.

Gandecha AR, Cundliffe E (1996) Molecular analysis of *tlrD*, a MLS resistance determinant from the tylosin producer, *Streptomyces fradiae*. Gene 180: 173-176.

Gandecha AR, Large S, Cundliffe E (1997) Analysis of four tylosin biosynthetic genes from the *tylLM* region of the *Streptomyces fradiae* genome. Gene 184: 197-203.

Geistlich M, Losick R, Turner JR, Rao RN (1992) Characterization of a novel regulatory gene governing the expression of polyketide synthese gene in *Streptomyces ambofaciens*. Mol. Microbiol. 6: 2019-2029.

Gray PP, Bhuwapathanapun S (1980) Production of the macrolide antibiotic tylosin in batch and chemostat culture. Biotechnol. Bioeng. 22: 1785-1804.

Hadfield C (1987) Genomic library cloning – with plasmid, lambda and cosmid vectors. In: Boulnois GJ (eds). Gene Cloning Analysis, a laboratory guide. Blackwell Scientific Publications. p61-105.

Hanahan D (1983) Studies on transformation of *E. coli* with plasmids. J. Mol. Biol. 166: 557-580.

Hanahan D, Meselson M (1983) Plasmid screening at high density. Methods Enzymol. 100: 333-342.

Hannig G, Makrides SC (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*. TIBTECH 16: 54-60.

Hara O, Hutchinson CR (1992) A macrolide 3-*O*-acyltransferase gene from the midecamycin-producing species *Streptomyces mycarofaciens*. J. Bacteriol. 174: 5141-5144.

Hara O, Hutchinson CR (1990) Cloning of midecamycin (MLS)-resistance genes from *Streptomyces mycarofaciens*, *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). J. Antibiotics 43: 977-991.

Haydock SF, Dowson JA, Dhillon N, Roberts GA, Cortes J, Leadlay PF (1991) Cloning and sequence analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythraea*: sequence similarities between EryG and a family of S-adenosylmethionine-dependent methyltransferases. Mol. Gen. Genet. 230: 120-128.

Higgins DG, Bleasby AJ, Fuchs R (1991) CLUSTAL V: improved software for multiple sequence alignment.

Hohn B, Collins J (1980) A small cosmid for efficient cloning of large DNA fragments. Gene 11: 291-298.

Hopwood DA, Bibb MJ, Chater KF, Kieser T, Bruton CJ, Kieser HM, Lydiate DJ, Smith CP, Ward JM, Schrempf H (1985) Genetic manipulation of *Streptomycetes*: A laboratory manual. The John Innes Foundation, Norwich.

Hopwood DA (1988) Towards an understanding of gene switching in *Streptomyces*, the basis of sporulation and antibiotic production. Proc. Roy. Soc. London B 235: 121-138.

Hopwood DA, Khosla C (1992) in Secondary Metabolites: Their Function and Evolution. Ciba Foundation Symposium 171, Wiley, Chichester, UK, p 88-112.

Ikeda H, Nonomiya T, Usami M, Ohta T, Omura S (1999) Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. Proc. Natl. Acad. Sci. USA 96: 9509-9514.

Ingrosso D, Fowler AV, Bleibaum J, Clarke S (1989) Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. common sequence motifs for protein, DNA, RNA, and small molecule S-adenosylmethionine-dependent methyltransferases. J. Biol. Chem. 264: 20130-20139.

Innis MA, Gelfand DH, Sninsky JJ, White TJ (1990) PCR-Protocols: a guide to methods and applications. Academic Press, INC.

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

Jiang XM, Neal B, Santiago F, Lee SJ, Romana LK, Reeves PR (1991) Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *thyphimurium* (strain LT2). Mol. Microbiol. 5: 695-713.

Kagan RM, Clarke S (1994) Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch. Biochem. Biophys. 310: 417-427.

Kakinuma S, Takada Y, Ikeda H, Tanaka H, Ömura S, Hopwood DA (1991) Cloning of large DNA fragments, which hybridize with actinorhodin biosynthesis genes, from kalafungin and nanaomycin A methyl ester producers and identification of genes for kalafungin biosynthesis of the kalafungin producer. J. Antibiotics 44: 995-1005.

Katz L, Donadio, S (1993) Polyketide synthesis: prospects for hybrid antibiotics. Annu. Rev. Microbiol. 47: 875-912.

Kim W, Kim C, Han O (1999) The function of *eryBVII* gene is to epimerize TDP-6-deoxy-L-threo-D-glycero-4-hexulose in the biosythesis of erythromycim A. J. Biochemistry & Mol.Biology 32: 72-75.

Kirschning A, Bechthold AFW, Rohr J (1997) Chemical and biochemical aspects of deoxysugars and deoxysugar oligosaccharides. Top. Curr. Chem. 1997, 188: 1-84.

Kuhstoss S, Huber M, Turner J, Paschal J, Rao R (1996) Production of a novel polyketide through the construction of a hybrid polyketide synthase. Gene 183: 231-236.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

Lin YS, Chen CW (1997) Instability of artificially circularized chromosomes of *Streptomyces lividans*. Mol. Microbiol. 26: 709-719.

Linton KJ, Jarvis BW, Hutchinson CR (1995) Cloning of the genes encoding thymidine diphospho-glucose 4,6-dehydratase and thymidine diphospho-4-keto-6-deoxyglucose 3,5-epimerase from the erythromycin-producing *Saccharopolyspora erythraea*. Gene 153: 33-40.

Little P (1987) Choice and use of cosmid vectors. in : Glover DM (eds). DNA cloning Volume III, a practical approach. IRL Press Limited, England, p19-42.

Liu D, Haase AM, Lindqvist L, Lindberg AA, Reeves PR (1993) Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: Identification and characterisation of transferase genes of groups B, C2 and E1. J. Bacteriol. 175: 3408-3413.

Lombó F, Siems K, Brana A, Mendez C, Bindseil K, Salas J (1997) Cloning and insertional inactivation of *Streptomyces argillaceus* genes involved in the earliest steps of biosynthesis of the sugar moieties of the antitumor polyketide mithramycin. J. Bacteriol. 179: 3354-3357.

MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T (1992a) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene 111: 61-68.

MacNeil DJ, Occi JL, Gewain KM, MacNeil T, Gibbons PH, Ruby CL, Danis SJ (1992b) Complex organization of the *Streptomyves avermitilis* genes encoding avermectin polyketide synthase. Gene 115: 119-125.

Madduri K, Kennedy K, Rivola G, Inventi-Solari A, Filippini S, Zanuso G, Colombo A, Gewain K, Occi J, MacNeil D, Hutchinson C (1997) Production of the antitumor drug epirubicin (4´-epidoxorubicin) and its precursor by a genetically engineered strain of *Streptomyces peucetius*. Nature Biotech. 16: 69-74.

Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol. Rev. 60: 512-538.

Marck C (1988) "DNA-Strider": a 'C' programm for the fast analysis of DNA and protein sequences on the Macintosh family of computers. Nucleic Acids Res. 16.

Marsden AFA, Caffrey P, Aparico JF, Loughran MS, Staunton J, Leadlay PF (1994) Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. Science 263: 378-380.

Martin JF, Liras P (1981) Biosynthetic pathway of secondary metabolites in industrial microorganisms. In: Rehm HJ, Reed G (eds). Biotechnology. Vol.1, *Microbial Fundaments,* Verlag Chemie, Weinheim, p211-233.

Martin J, Hartl FU (1997) Chaperone-assisted protein folding. Curr. Opin. Struct. Biol. 7: 41-52.

Merson-Davies LA, Cundliffe E (1994) Analysis of five biosynthetic genes from the *tylIBA* region of the *Streptomyces fradiae* genome. Mol. Microbiol. 13(2): 349-355.

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, p433.

Minas W (1997) Fascination *Streptomycetes*. Bio World 5: 14-17.

Motamedi H, Shafiee A, Cai SJ (1995) Integrative vectors for heterologous gene expression in *Streptomyces* spp. Gene 160: 25-31.

Nakagawa A, Omura S (1984) Structure and stereochemistry of macrolides. In: Omura S (eds). Macrolide antibiotics. Chemistry, biology and practice. Academic Press, Inc., New York, p37-84.

Nelson DR, Kamataki T, Wasman DJ, Guengueric FP, Estrabrook RW, Feyerensen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K, Nebert DW (1993) The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes and nomenclature. DNA Cell Biol. 12: 1-51.

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

Ogawa H, Konishi K, Takata Y, Nakashima H, Fujioka M (1987) Rat glycine methyltransferase. Complete amino acid sequence deduced from a cDNA clone and characterization of the genome DNA. Eur. J. Biochem. 168: 141-151.

Okazaki R, Okazaki T, Strominger JT, Michelson AM (1962) Thymidine-diphosphate-4-keto-6-deoxyglucose, an intermediate in thymidine-diphosphate-L-rhamnose synthesis in *Escherichia coli* strains. J. Biol. Chem. 237: 3014-3026.

Olano C, Rodriguez A, Michel J, Mendez C, Raynal M, Salas J (1998) Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring. Mol. Gen. Genet. 259: 299-308.

Omura S, Nakagawa A, Neszmeli A, Gero SD, Sepulchre AM, Piriou F, Lukacs G (1975a) Carbon-13 nuclear magnetic resonance spectral analysis of 16-membered macrolide antibiotics. J. Am. Chem. Soc. 97: 4001-4009.

Omura S, Nakagawa A, Takeshima H, Mlyazawa J, Kitao C (1975b) A C13 nuclear magnetic resonance study of the biosynthesis of the 16-membered macrolide antibiotic tylosin. Tetrahedon Lett. 50: 4503-4506.

Omura S, Kitao C, Hamada H, Ikeda H (1979) Bioconversion and biosynthesis of 16-membered macrolide antibiotics. X. final steps in the biosynthesis of spiramycin, using enzyme inhibitor: cerulenin. Chem. Pharm. Bull. 27: 176-182.

Omura S (1984) In: Omura S (eds). Macrolide Antibiotics: Chemistry, Biology, and Practice. Academic Press, New York.

Otten SL, Liu X, Ferguson F, Hutchinson CR (1995) Cloning and characterisation of the *Streptomyces peucetius dnrQS* genes encoding a daunosamine biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis. J. Bacteriol. 177: 6688-6692.

Otten S, Gallo MA, Madduri K, Liu X, Hutchinson CR (1997) Cloning and characterization of *Streptomyces peutetius dnmZUV* genes encoding three enzymes required for biosynthesis of the daunorubicin precursor thymidine diphospho-L-daunosamine. J. Bacteriol. 179: 4446-4450.

Pascarella S, Bossa F (1994) Similarity between pyridoxal/pyridoxamine phosphate-dependent enzymes involved in dideoxy and deoxyaminosugar biosynthesis and other pyridoxal phosphate enzymes. Protein Sci. 3: 701-705.

Payne SM, Ames BN (1982) A procedure for rapid extraction and high pressure liquid chromatographic separation of the nucleotides and other small molecules from bacterial cells. Anal. Biochem. 123: 151-161.

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. USA 85: 2444-2448.

Peschke U, Schmidt H, Zhang HZ, Piepersberg W (1995) Molecular characterisation of the lincomycin production gene cluster of *Streptomyces lincolnensis* 78-11. Mol. Microbiol. 16: 1137-1156.

Piepersberg W (1994) Pathway engineering in secondary metabolite-producing actinomycetes. Crit. Rev. in Biotech. 14(3): 251.285.

Piepersberg W (1995) Streptomycin and related aminoglycoside. In: Vining LC, Stuttard C (eds). Genetics and biochemistry of antibiotic production. Butterworth-Heinemann, Biotechnology series, Vol. 28, Newton, MA, p531-570.

Piepersberg W, Distler J (1997) Aminoglycosides and sugar components in other secondary metabolites. In: Rehm HJ, Reed G (eds.). Biotechnology. second, completely revised edition, Vol. 7, Verlag Chemie, Weinheim, p397-488.

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.

Pospiech A, Neumann B (1995) A versatile quick-prep of genomic DNA from Gram-positive bacteria. Technical Tips 11: 217-218.

Redenbach M, Kieser HM, Denapaite D, Eichner A, Cullum J, Kinashi H, Hopwood DA (1996). A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. Mol. Microbiol. 21: 77-96.

Reed KC, Mann DA (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. 13: 7207-7221.

Reeves A, Post D, Vanden Boom T (1998) Physical-genetic map of the erythromycin-producing organism *Saccharopolyspora erythraea*. Microbiology 144: 2151-2159.

Robinson JA (1988) Enzymes of secondary metabolism in microorganisms. Chem. Soc. Rev. 17: 383-452.

Rosteck P Jr, Reynolds P, Hershberger C (1991) Homology between proteins controlling *Streptomyces fradiae* tylosin resistance and ATP-binding transport. Gene 102: 27-32.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.

Sakakibara H, Omura S (1984) Chemical modification and structure-activity relationship of macrolides. In: Omura S (eds). Macrolide antibiotics. Chemistry, biology and practice. Academic Press, Inc., New York, p85-126.

Salah-Bey K, Doumith M, Michel JM, Haydock S, Cortes J, Leadlay PF, Raynal MC (1998) Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer *Saccharopolyspora erythraea*. Mol. Gen. Genet. 257: 542-553.

Sales J, Mendez C (1998) Genetic manipulation of antitumor-agent biosynthesis to produce novel drugs. Tibtech 16: 475-482.

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

Sanger F, Nicklan S, Coulson AR (1977) DNA sequencing with chain-determination inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Schneider A, Marahiel MA (1998) Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. Arch. Microbiol. 169: 404-410.

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

Schwecke T, Aparicio JF, Molnár I, König A, Khaw LE, Haydock SF, Oliynyk M, Caffrey P, Cortés J, Lester JB, Böhm GA, Staunton J, Leadlay PF (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proc. Natl. Acad. Sci. USA 92: 7839-7843.

Scrutton NS, Berry A, Perham RN (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature 343: 38-43.

Solenberg P, Matsushima P, Stack D, Wilkie S, Thompson R, Baltz R (1997) Production of hybrid glycopeptide antibiotics in vitro and in *Streptomyces toyocaensis*. Chem. Biology 4: 195-202.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-521.

Stanzak R, Matsushima P, Baltz RH, Rao RN (1986) Cloning and expression in *Streptomyces lividans* of clustered erythromycin biosynthesis genes from *Streptomyces erythraeus*. Bio/Technology 4: 229-232.

Stassi D, Donadio S, Staver MJ, Katz L (1993) Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. J. Bacteriol. 175: 182-189.

Steven AF, Cundliffe E (1997) Stimulation of polyketide metabolism in *Streptomyces fradiae* by tylosin and its glycosylated precursors. Microbiology 143: 3871-3876.

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.

Stratmann A (1997) Identifizierung eines Acarbose-Biosynthesegenclusters in *Actinoplanes* sp. und Charakterisierung ausgewählter Enzyme des Acarbose-Stoffwechsels. PhD thesis, Uni. Wuppertal.

Studier WF, Rosenberg AH, Dunn JJ, Dubendroff JW (1990) Use of the T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185: 61-89.

Summers RG, Donadio S, Staver MJ, Wendt-Pienkowski E, Hutchinson CR, Katz L (1997) Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production. Microbiology 143:3251-3262.

Takano E, White J, Thompson CJ, Bibb MJ (1995) Construction of thiostrepton-inducible, high-copy-number expression vectors for use in *Streptomyces* spp. Gene 166: 133-137.

Thompson MW, Strohl WR, Floss HG (1992) Purification and characterisation of TDP-D-glucose 4,6-dehydratase from anthracycline-producing streptomycetes. J. Gen. Microbiol. 138: 779-786.

Thorson JS, Lo SF, Liu HW, Hutchinson CR (1993) Biosynthesis of 3,6-dideoxyhexoses: new mechanistic reflections upon 2,6-dideoxy, 4,6-dideoxy, and amino sugar construction. J. Amer. Chem. Soc. 115: 6993-6994.

Torkkell S, Ylihonko K, Hakala J, Skurnik M (1997) Characterization of *Streptomyces nogalater* genes encoding enzymes involved in glycosylation steps in nogalamycin biosynthesis. Mol. Gen. Genet. 256: 203-209.

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

Vara J, Lewandowska-Skarbek M, Wang YG, Donadio S, Hutchinson CR (1989) Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). J. Bacteriol. 171: 5872-5881.

Verseck S (1997) Klonierung und Expression von bakteriellen Enzymen der Biosynthese von 6-Desoxyhexose-Derivaten und deren Verwendung zur präparativen Herstellung von dTDP-aktivierten Hexosen. PhD thesis, Uni. Wuppertal.

Vieira J, Messing JR (1982) The pUC plasmids and M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259-268.

Vining LC (1992) Roles of secondary metabolites from microbes. In: Wiley J, Sons (eds). Secondary Metabolites: Their Function and Evolution. Ciba Foundation Symposium 171, a Wiley-interscience Publication, p184-198.

Volff JN, Viell P, Altenbuchner J (1997) Artifical circularization of the chromosome with concomitant deletion of its terminal inverted repeats enhances genetic instability and genome rearrangement in *Streptomyces lividans*. Mol. Gen. Genet. 253: 753-760.

Wakil SJ (1989) Fatty acid synthase, a proficient mutifunctional enzyme. Biochemistry 28: 4523-4530.

Walker JB (1975) Pathway of biosynthesis of the guanidinated inositol moieties of streptomycin and bluensomycin. Methods Enzymol. 43: 429-470.

Walker JB, Walker MS (1969) Streptomycin biosynthesis. Transamination reactions involving inosamines and inosadiamines. Biochemistry 8: 763-770.

Wang SF, Gabriel O (1969) Biological mechanisms involved in the formation of deoxy sugars. V. Isolation and crystallization of thymidine diphosphate-D-glucose oxidoreductase from *Escherichia coli* B. J. Biol. Chem. 244: 3430-3437.

Weber JM, Wierman CK, Hutchinson R (1985) Genetic analysis of erythromycin production in *Streptomyces erythraeus*. J. Bacteriol. 164: 425-433.

Weber JM, Schoner B, Losick R (1989) Identification of a gene required for the terminal step in erythromycin A biosynthesis in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). Gene 75: 235-241.

Weber JM, Leung JO, Swanson SJ, Idler KB, McAlpine JB (1991) An erythromycin derivative produced by trageted gene disruption in *Saccharopolyspora erythraea*. Science 252: 114-117.

Wehmeier UF (1995) New multifunctional *Escherichia coil-Streptomyces* shuttle vectors allowing blue-white screening on Xgal plates. Gene 165: 149-150.

Weis JH, Quertermous T (1987) Construction of recombinant DNA libraries. In : Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, Struhl K (eds.). Current Protocols in Molecular Biology. Supplement 4. New York: John Wiley and Sons.

Weymouth-Wilson AC (1997) The role of carbohydrates in biologically active natural products. Nat. Prod. Rep. 14: 99-110.

Wierenga RK, De Maeyer MCH, Hol WGJ (1985) Interaction of pyrophosphate moieties with α-helices in dinucleotide binding proteins. J. Biochem. 24: 1346-1357.

Wietzorrek A, Bibb M (1997) A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. Mol. Microbiol. 25: 1177-1184.

Wilkinson DL, Harrison RG (1991) Predicting the solubility of recombinant proteins in *Eschericha coli*. Biotechnology 9: 443-448.

Williams N, Wander J (1980) Deoxy and branched-chain sugars. In: Pigman W, Horton D (eds). The Carbohydrates: Chemistry and Biochemistry. New York: Academic, p761-798.

Wilson V, Cundliffe E (1998) Characterization and targeted disruption of a glycosyltransferase gene in the tylosin producer *Streptomyces fradiae*. Gene 214: 95-100.

Winston SE, Fuller SA, Hurrell JGR (1987) Western blotting. In : Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, Struhl K (eds.). Current Protocols in Molecular Biology. Supplement 4. New York: John Wiley and Sons.

Woodward RB (1957) Angrew. Chem. 69: 50-58.

Wu JC, Santi DV (1987) Kinetic and catalytic mechanism of *Hha*I methyltransferase. J. Biol. Chem. 262: 4778-4786.

Wu JC, Santi DV (1988) High level expression and purification of HhaI methyltransferase. Nucleic Acids Res. 16: 703-717.

Xue Y, Zhao L, Liu HW, Sherman DH (1998) A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. Proc. Natl. Acad. Sci. USA 95: 12111-12116.

Ylihonko K, Tuikkanen J, Jussila S, Cong L, Mäntsälä P (1996) A gene cluster involved in nogalamycin biosynthesis from *Streptomyces nogalater*: sequence analysis and

complementation of early-block mutations in the anthracycline pathway. Mol. Gen. Genet. 251: 113-120.

Zalacain M, Cundliffe E (1991) Cloning of *tlrD*, a fourth resistance gene, from the tylosin producer, *Streptomyces fradiae*. Gene 97: 137-142.

7. Appendices

7.1 The nucleotide and encoded amino acid sequences of the *midM* region of Smyc-LC1

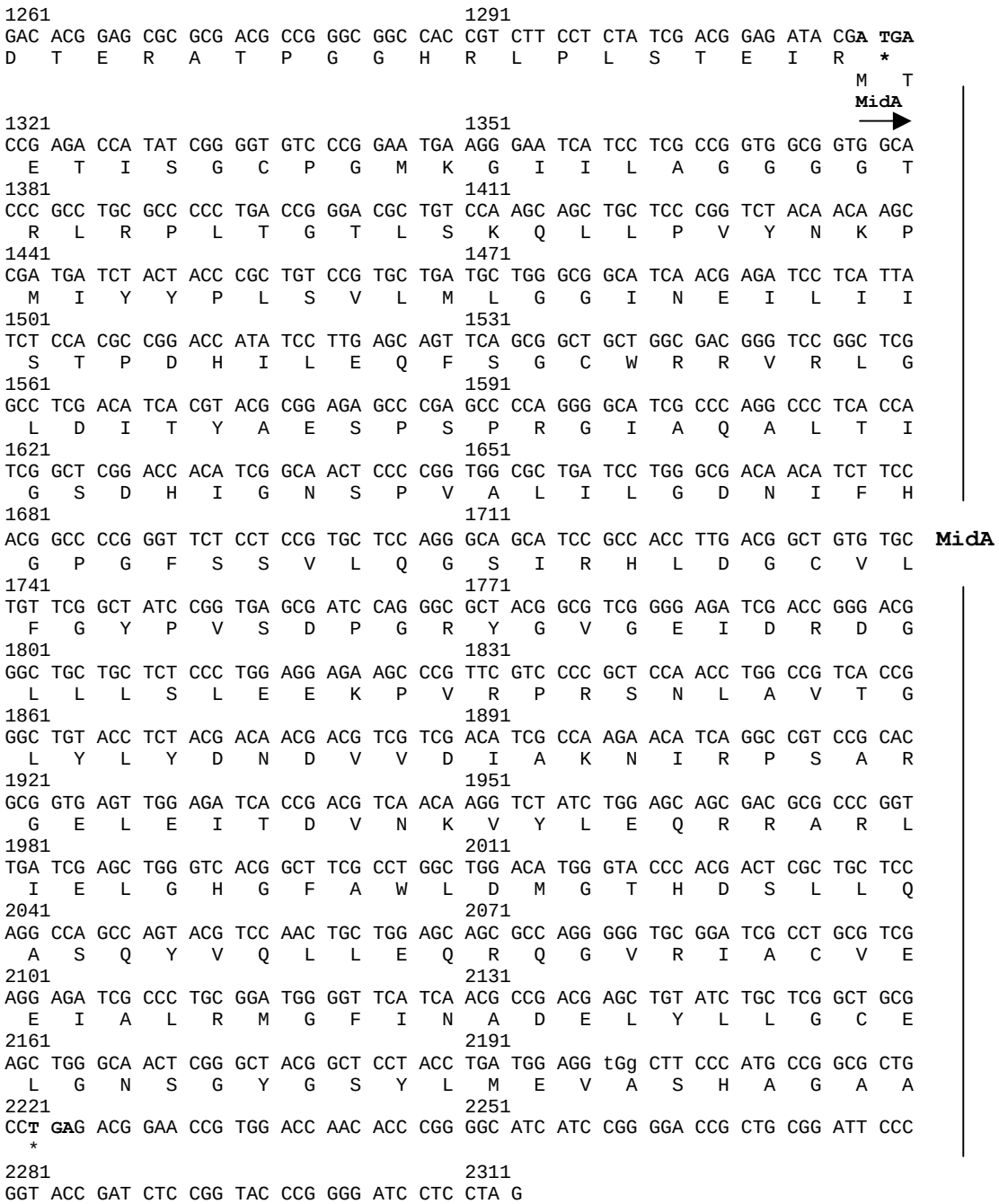
```
1                                                31
TTT CTC CGT GCA TCC CTC TCG CAC GTG AAC AGT GAA GTG TTC AGC ACC GGG ACG GCG AGG
61                                               91
ACA ACC GGC GAG CGG TGA ATG TGT CAG GGT GCC GGG GGC GCT GCG GCA GCC CCG GCG GTT
121                                              151
TCC CGT CCG CGG GGG CGG ACG GGG AGG TCA TTG TGT CAG GAC TCC CTT GCT GGC GAT CTT
181                                              211
GAG CGA TTC ATA GGC TCA CCG GCG TCG AGA GAC CGA AGT GCC CGC GTC GCG TGC TCG GTC
241                                              271
CGC CGT ACC ACT TCC CAA GGA GAA TGG CCG ATG TCC GAG GCT CCG ACA GTG CCA CTC GAA
                                          M   S   E   A   P   T   V   P   L   E
                                                 MidM ──►
301                                              331
CTC AGC AAG GAG GCG AAC GCC CAG GAA CTC CTG GGA CTG GGT TCG CCT TCA ACC GGG ACC
L   S   K   E   A   N   A   Q   E   L   L   G   L   G   S   P   S   T   G   T
361                                              391
CAT CAC CCG GTC TTC TGG GAC GAG AAT TGG GCA TGC CTG GGC AAG GTC TTC CGT TAC GAC
H   H   P   V   F   W   D   E   N   W   A   C   L   G   K   V   F   R   Y   D
421                                              451
GAC TAC CTG ACG GTC TCC AAC AAT CCG CAG TTC TTC TCC TCC GAC TTC AAC GAG GTC ATG
D   Y   L   T   V   S   N   N   P   Q   F   F   S   S   D   F   N   E   V   M
481                                              511
CCG ACC CCG CCC GAG CTG GAG AtG GTC ATC GGG CCG GGG ACC ATC GGC GCC CTC GAC CCG
P   T   P   P   E   L   E   M   V   I   G   P   G   T   I   G   A   L   D   P
541                                              571
CCC GCC CAC GGC CCG ATG CGC AAG CTG GTC AGC CAG GCC CTC ACC CCG CGG CGG ATG GCC
P   A   H   G   P   M   R   K   L   V   S   Q   A   L   T   P   R   R   M   A
601                                              631
CGT CTG GGA CCC CGC ATC CGG GCC GTC ACT CAG GGG CTC CTC GAT GCG GTG CGC GGC CAG
R   L   G   P   R   I   R   A   V   T   Q   G   L   L   D   A   V   R   G   Q
661                                              691
GAG ACC ATC GAC GTC GTC GGC GAC CTC TCC TAC GCC CTG CCC GTG ATC GTC ATC GCC GAG
E   T   I   D   V   V   G   D   L   S   Y   A   L   P   V   I   V   I   A   E
721                                              751
CTG CTG GGC ATA CCG TCC GGT GAC CGC GAT GTG TTC CGC GGG TGG GTC GAC ACC CTG CTC
L   L   G   I   P   S   G   D   R   D   V   F   R   G   W   V   D   T   L   L
781                                              811
ACC AAC GAG GGC CTG GGG TAC CCG AGC CTC CCG GAC AAC TTC AGC GAG ACG ATC GCC CCC
T   N   E   G   L   G   Y   P   S   L   P   D   N   F   S   E   T   I   A   P
841                                              871
GCC CTC AAG GAG ATG ACC GAC TAT CTC CTG CAC CAG ATT CAC GCC AAG CGC GAG GCC CCG
A   L   K   E   M   T   D   Y   L   L   H   Q   I   H   A   K   R   E   A   P
901                                              931
GTC GAC GAC CTG ATC TGC GGC CTG GTC CAG GCC GAG CAG GAC GGC CGC AAG CTC ACC GAC
V   D   D   L   I   C   G   L   V   Q   A   E   Q   D   G   R   K   L   T   D
961                                              991
GTG GAG ATC GTC AAC ATC GTG GCC CTG CTG CTC ACC GCC GGG CAC GTC TCG TCA AGC ACG
V   E   I   V   N   I   V   A   L   L   L   T   A   G   H   V   S   S   S   T
1021                                             1051
CTG CTC AGC AAC CTC TCC CTG GTA CTG GAG GAG AAC CCG CAG GCG CTG GCG GAC CTG CGC
L   L   S   N   L   S   L   V   L   E   E   N   P   Q   A   L   A   D   L   R
1081                                             1111
GCC GAC CGC GAG CTG GTG ACC GGT GCG GTG GAG GAG ACA CTG CGC TAC CGC AGC CCC TTC
A   D   R   E   L   V   T   G   A   V   E   E   T   L   R   Y   R   S   P   F
1141                                             1171
AAC AAC ATC TTC CGC TTC CTC AAG GAG GAC ACC GAC ATC CTC GGC CCG GAA ATG AAG AAG
N   N   I   F   R   F   L   K   E   D   T   D   I   L   G   P   E   M   K   K
1201                                             1231
GGC CAG ATG GTC ATC GCC TGG AGC CAG TCG GCC AAC CGC GAC CCC GAA CAC TTC CCG GAG
G   Q   M   V   I   A   W   S   Q   S   A   N   R   D   P   E   H   F   P   E
1261                                             1291
CCC GAC ACC TTC GAC ATC CGA CGC TCA AGC AGC TCC CGC CAT ATG GCG TTC GGC ATC GGT
P   D   T   F   D   I   R   R   S   S   S   S   R   H   M   A   F   G   I   G
1321                                             1351
ATT CAC CAC TGC CTG GGC GCC TTC CTG GCA CGT CAG GAA GGC AAG GTG GTA CTG GAA CTG
I   H   H   C   L   G   A   F   L   A   R   Q   E   G   K   V   V   L   E   L
```

```
1381                                                1411
ATG CTG GAT CAG GTC AGG GAA TTC GAG CTC GGT ACC CGG GAT CCT CTA GAG TCG AC*
M   L   D   Q   V   R   E   F   E   L   G   T   R   D   P   L   E   S
```

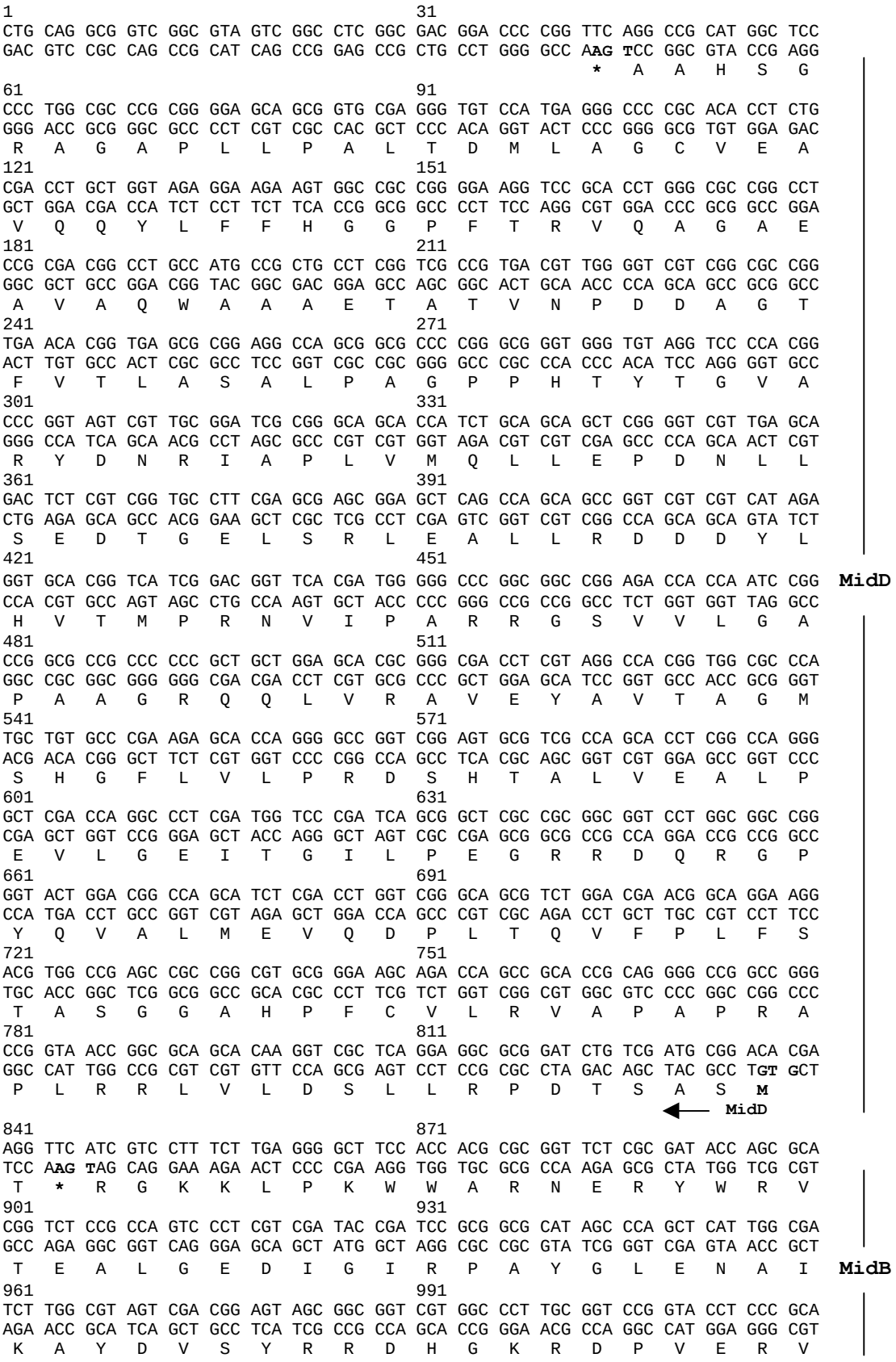
* The sequence is not complete at the 3´end.

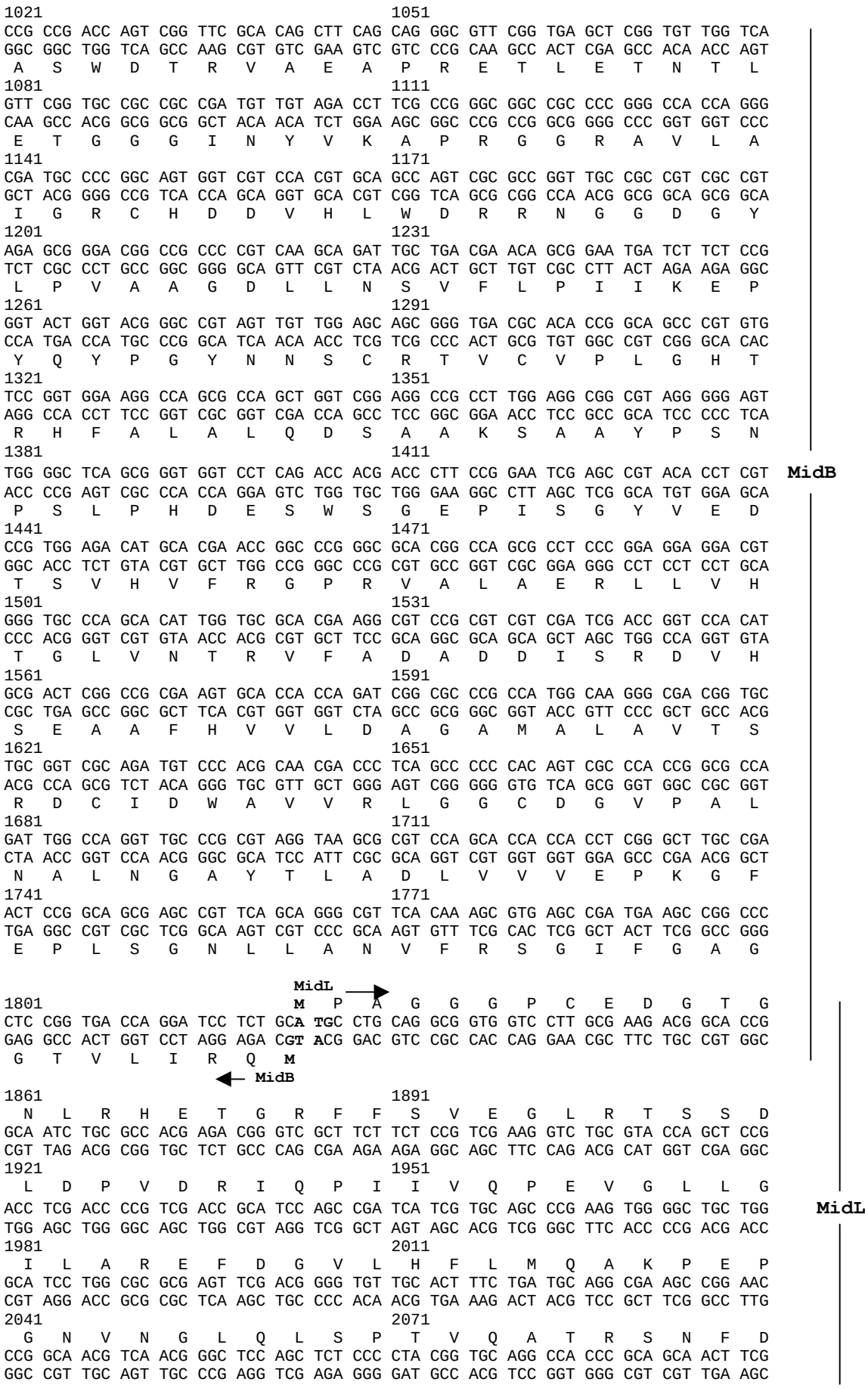
7.2. The nucleotide and encoded amino acid sequences of the *midCA* region of Smyc-LC1

```
1                                                   31
GGC CCA ATG GTT TTG CCG AGT GCT CAA GGA ATT CGC ACC GGA ATC GTT CGC CCT TAG TCC
61                                                  91
TTC GCT CGC GGC ATA CGA GGA GTC CGA TTA TTA CCG CGA CTA CGA CAC ATT CCT GCG CGA
121                                                 151
CGC ATG GAG CAT AAA GTG AAC GTG CCC TTT CCT CGA CGC GGG TGC GGC TAT CGG GAG CTG
    M   E   H   K   V   N   V   P   F   P   R   R   G   C   G   Y   R   E   L
    MidC→
181                                                 211
CGG GCC GAC ATC GAC GGG GCC CTT CGG CGG GTG TCC GCC TCC GGA CGC TAT CTG CTG GGT
R   A   D   I   D   G   A   L   R   R   V   S   A   S   G   R   Y   L   L   G
241                                                 271
GCG GAA CTG GCG GGG TTC GAG GCG GAA TTC GCC GCG TAC TGC GAC AAC GAC TGC GTG GCG
A   E   L   A   G   F   E   A   E   F   A   A   Y   C   D   N   D   C   V   A
301                                                 331
GTC GGC AGC GGC TGC GAC GCC CTG GAG CTG GTG CGC GCG CTC GGC ATC GGC CCC GGT GAC
V   G   S   G   C   D   A   L   E   L   V   R   A   L   G   I   G   P   G   D
361                                                 391
GAG GTG GTG GTG CCC GCG CAC ACC TTC ATC GGC ACC TGG CTG GCG GTG TCC GCC GCC GGG
E   V   V   V   P   A   H   T   F   I   G   T   W   L   A   V   S   A   A   G
421                                                 451
GCC CGG CCG GTG GGT GTC GAC CCG ACG CCG GAC GGG CTG TCC ATG GAC CCG GCG CAG GTG
A   R   P   V   G   V   D   P   T   P   D   G   L   S   M   D   P   A   Q   V
481                                                 511
GAG GCG GCG ATC ACT CCC CGG ACC AGG GCC GTG ATG CCG GTG CAT CTG TAC GGG CAT CCG
E   A   A   I   T   P   R   T   R   A   V   M   P   V   H   L   Y   G   H   P
541                                                 571
GCC GAT CTG GAC CCG CTT CTG GCG ATC GCC GAA CGG CAC GGC CTG GCC GTG GTG GAG GAC
A   D   L   D   P   L   L   A   I   A   E   R   H   G   L   A   V   V   E   D
601                                                 631
GCC GAG CAG GCG CAC GGC GCC CGC TAC CGG GGC CGC CGG ATC GGC TCG GGC CAT GTG GTC  MidC
A   E   Q   A   H   G   A   R   Y   R   G   R   R   I   G   S   G   H   V   V
661                                                 691
GCG TTC AGC TTC TAC CCC GGC AAG AAC CTC GGC GCC ATG GGA GAC GGC GGG GCG GTG GTC
A   F   S   F   Y   P   G   K   N   L   G   A   M   G   D   G   G   A   V   V
721                                                 751
ACC GGC GAC GCC GCC CTG GCC GAC CGG ATC CGG CTG CTG CGC AAA TGC GGC TCC CGT GAG
T   G   D   A   A   L   A   D   R   I   R   L   L   R   K   C   G   S   R   E
781                                                 811
AAG TAC CGG CAC GAG GTA CAG GCC ACC AAC TCG CGG CTC GAC GAG TTC CAG GCC GCC GTG
K   Y   R   H   E   V   Q   A   T   N   S   R   L   D   E   F   Q   A   A   V
841                                                 871
CTG CGC GCC AAG CTG CCA CGG GTG CCG GCC TGG AAC GCC CTC CGA GTG CGC ACG GCC GAG
L   R   A   K   L   P   R   V   P   A   W   N   A   L   R   V   R   T   A   E
901                                                 931
CGC TAC AGC CAG GTC TTG GGT GCC CTC CCG CAG ATT GCC GTC CCC GCC GCC GCC CCC TGG
R   Y   S   Q   V   L   G   A   L   P   Q   I   A   V   P   A   A   A   P   W
961                                                 991
GCC GAT CCG GTG TGG CAC CTT TAT GTG ATC CTG CGC GCG AAC CGC GAC GAA CTG CGC CGC
A   D   P   V   W   H   L   Y   V   I   L   R   A   N   R   D   E   L   R   R
1021                                                1051
CGA ATC GAA CGG GCC GGG GTG GAG ACC CTG ATC CAC TAC CCC GTA CCG CCC CAC CGG ACC
R   I   E   R   A   G   V   E   T   L   I   H   Y   P   V   P   P   H   R   T
1081                                                1111
CCG GCC TAC GCC GAC GAT TCC GGC CGG CGC TCC GGC CGC ACC CAC CCG CTC AGC GAA CGC
P   A   Y   A   D   D   S   G   R   R   S   G   R   T   H   P   L   S   E   R
1141                                                1171
CGT GCG GCG GAG AGC CTC AGC CTT CCC TCG GGC CCC CAC CTC GGG GAC GAC GCG TTC CAG
R   A   A   E   S   L   S   L   P   S   G   P   H   L   G   D   D   A   F   Q
1201                                                1231
ACC GTC GTA GCG GCG GTC CGC GCG GCG GCC GTA GGA CTG CCG GCG TAT CCG GCG CCG GAC
T   V   V   A   A   V   R   A   A   A   V   G   L   P   A   Y   P   A   P   D
```


1261
GAC ACG GAG CGC GCG ACG CCG GGC GGC CAC
D T E R A T P G G H
1291
CGT CTT CCT CTA TCG ACG GAG ATA CGA TGA
R L P L S T E I R *
M T
MidA
1321
CCG AGA CCA TAT CGG GGT GTC CCG GAA TGA
E T I S G C P G M K
1351
AGG GAA TCA TCC TCG CCG GTG GCG GTG GCA
G I I L A G G G G T
1381
CCC GCC TGC GCC CCC TGA CCG GGA CGC TGT
R L R P L T G T L S
1411
CCA AGC AGC TGC TCC CGG TCT ACA ACA AGC
K Q L L P V Y N K P
1441
CGA TGA TCT ACT ACC CGC TGT CCG TGC TGA
M I Y Y P L S V L M
1471
TGC TGG GCG GCA TCA ACG AGA TCC TCA TTA
L G G I N E I L I I
1501
TCT CCA CGC CGG ACC ATA TCC TTG AGC AGT
S T P D H I L E Q F
1531
TCA GCG GCT GCT GGC GAC GGG TCC GGC TCG
S G C W R R V R L G
1561
GCC TCG ACA TCA CGT ACG CGG AGA GCC CGA
L D I T Y A E S P S
1591
GCC CCA GGG GCA TCG CCC AGG CCC TCA CCA
P R G I A Q A L T I
1621
TCG GCT CGG ACC ACA TCG GCA ACT CCC CGG
G S D H I G N S P V
1651
TGG CGC TGA TCC TGG GCG ACA ACA TCT TCC
A L I L G D N I F H
1681
ACG GCC CCG GGT TCT CCT CCG TGC TCC AGG
G P G F S S V L Q G
1711
GCA GCA TCC GCC ACC TTG ACG GCT GTG TGC
S I R H L D G C V L
MidA
1741
TGT TCG GCT ATC CGG TGA GCG ATC CAG GGC
F G Y P V S D P G R
1771
GCT ACG GCG TCG GGG AGA TCG ACC GGG ACG
Y G V G E I D R D G
1801
GGC TGC TGC TCT CCC TGG AGG AGA AGC CCG
L L L S L E E K P V
1831
TTC GTC CCC GCT CCA ACC TGG CCG TCA CCG
R P R S N L A V T G
1861
GGC TGT ACC TCT ACG ACA ACG ACG TCG TCG
L Y L Y D N D V V D
1891
ACA TCG CCA AGA ACA TCA GGC CGT CCG CAC
I A K N I R P S A R
1921
GCG GTG AGT TGG AGA TCA CCG ACG TCA ACA
G E L E I T D V N K
1951
AGG TCT ATC TGG AGC AGC GAC GCG CCC GGT
V Y L E Q R R A R L
1981
TGA TCG AGC TGG GTC ACG GCT TCG CCT GGC
I E L G H G F A W L
2011
TGG ACA TGG GTA CCC ACG ACT CGC TGC TCC
D M G T H D S L L Q
2041
AGG CCA GCC AGT ACG TCC AAC TGC TGG AGC
A S Q Y V Q L L E Q
2071
AGC GCC AGG GGG TGC GGA TCG CCT GCG TCG
R Q G V R I A C V E
2101
AGG AGA TCG CCC TGC GGA TGG GGT TCA TCA
E I A L R M G F I N
2131
ACG CCG ACG AGC TGT ATC TGC TCG GCT GCG
A D E L Y L L G C E
2161
AGC TGG GCA ACT CGG GCT ACG GCT CCT ACC
L G N S G Y G S Y L
2191
TGA TGG AGG tGg CTT CCC ATG CCG GCG CTG
M E V A S H A G A A
2221
CCT GAG ACG GAA CCG TGG ACC AAC ACC CGG
*
2251
GGC ATC ATC CGG GGA CCG CTG CGG ATT CCC
2281
GGT ACC GAT CTC CGG TAC CCG GGG ATC CTC
2311
CTA G

7.3. The nucleotide and encoded amino acid sequences of the *midDBL* region of Smyc-LC1




1021
1051
CCG CCG ACC AGT CGG TTC GCA CAG CTT CAG CAG GGC GTT CGG TGA GCT CGG TGT TGG TCA
GGC GGC TGG TCA GCC AAG CGT GTC GAA GTC GTC CCG CAA GCC ACT CGA GCC ACA ACC AGT
A S W D T R V A E A P R E T L E T N T L
1081
1111
GTT CGG TGC CGC CGC CGA TGT TGT AGA CCT TCG CCG GGC GGC CGC CCC GGG CCA CCA GGG
CAA GCC ACG GCG GCG GCT ACA ACA TCT GGA AGC GGC CCG CCG GCG GGG CCC GGT GGT CCC
E T G G G I N Y V K A P R G G R A V L A
1141
1171
CGA TGC CCC GGC AGT GGT CGT CCA CGT GCA GCC AGT CGC GCC GGT TGC CGC CGT CGC CGT
GCT ACG GGG CCG TCA CCA GCA GGT GCA CGT CGG TCA GCG CGG CCA ACG GCG GCA GCG GCA
I G R C H D D V H L W D R R N G G D G Y
1201
1231
AGA GCG GGA CGG CCG CCC CGT CAA GCA GAT TGC TGA CGA ACA GCG GAA TGA TCT TCT CCG
TCT CGC CCT GCC GGC GGG GCA GTT CGT CTA ACG ACT GCT TGT CGC CTT ACT AGA AGA GGC
L P V A A G D L L N S V F L P I I K E P
1261
1291
GGT ACT GGT ACG GGC CGT AGT TGT TGG AGC AGC GGG TGA CGC ACA CCG GCA GCC CGT GTG
CCA TGA CCA TGC CCG GCA TCA ACA ACC TCG TCG CCC ACT GCG TGT GGC CGT CGG GCA CAC
Y Q Y P G Y N N S C R T V C V P L G H T
1321
1351
TCC GGT GGA AGG CCA GCG CCA GCT GGT CGG AGG CCG CCT TGG AGG CGG CGT AGG GGG AGT
AGG CCA CCT TCC GGT CGC GGT CGA CCA GCC TCC GGC GGA ACC TCC GCC GCA TCC CCC TCA
R H F A L A L Q D S A A K S A A Y P S N
1381
1411
TGG GGC TCA GCG GGT GGT CCT CAG ACC ACG ACC CTT CCG GAA TCG AGC CGT ACA CCT CGT
ACC CCG AGT CGC CCA CCA GGA GTC TGG TGC TGG GAA GGC CTT AGC TCG GCA TGT GGA GCA
P S L P H D E S W S G E P I S G Y V E D
1441
1471
CCG TGG AGA CAT GCA CGA ACC GGC CCG GGC GCA CGG CCA GCG CCT CCC GGA GGA GGA CGT
GGC ACC TCT GTA CGT GCT TGG CCG GGC CCG CGT GCC GGT CGC GGA GGG CCT CCT CCT GCA
T S V H V F R G P R V A L A E R L L V H
1501
1531
GGG TGC CCA GCA CAT TGG TGC GCA CGA AGG CGT CCG CGT CGT CGA TCG ACC GGT CCA CAT
CCC ACG GGT CGT GTA ACC ACG CGT GCT TCC GCA GGC GCA GCA GCT AGC TGG CCA GGT GTA
T G L V N T R V F A D A D D I S R D V H
1561
1591
GCG ACT CGG CCG CGA AGT GCA CCA CCA GAT CGG CGC CCG CCA TGG CAA GGG CGA CGG TGC
CGC TGA GCC GGC GCT TCA CGT GGT GGT CTA GCC GCG GGC GGT ACC GTT CCC GCT GCC ACG
S E A A F H V V L D A G A M A L A V T S
1621
1651
TGC GGT CGC AGA TGT CCC ACG CAA CGA CCC TCA GCC CCC CAC AGT CGC CCA CCG GCG CCA
ACG CCA GCG TCT ACA GGG TGC GTT GCT GGG AGT CGG GGG GTG TCA GCG GGT GGC CGC GGT
R D C I D W A V V R L G G C D G V P A L
1681
1711
GAT TGG CCA GGT TGC CCG CGT AGG TAA GCG CGT CCA GCA CCA CCA CCT CGG GCT TGC CGA
CTA ACC GGT CCA ACG GGC GCA TCC ATT CGC GCA GGT CGT GGT GGT GGA GCC CGA ACG GCT
N A L N G A Y T L A D L V V V E P K G F
1741
1771
ACT CCG GCA GCG AGC CGT TCA GCA GGG CGT TCA CAA AGC GTG AGC CGA TGA AGC CGG CCC
TGA GGC CGT CGC TCG GCA AGT CGT CCC GCA AGT GTT TCG CAC TCG GCT ACT TCG GCC GGG
E P L S G N L L A N V F R S G I F G A G
MidL
1801
M P A G G G P C E D G T G
CTC CGG TGA CCA GGA TCC TCT GCA TGC CTG CAG GCG GTG GTC CTT GCG AAG ACG GCA CCG
GAG GCC ACT GGT CCT AGG AGA CGT ACG GAC GTC CGC CAC CAG GAA CGC TTC TGC CGT GGC
G T V L I R Q M
MidB
1861
1891
N L R H E T G R F F S V E G L R T S S D
GCA ATC TGC GCC ACG AGA CGG GTC GCT TCT TCT CCG TCG AAG GTC TGC GTA CCA GCT CCG
CGT TAG ACG CGG TGC TCT GCC CAG CGA AGA AGA GGC AGC TTC CAG ACG CAT GGT CGA GGC
1921
1951
L D P V D R I Q P I I V Q P E V G L L G
ACC TCG ACC CCG TCG ACC GCA TCC AGC CGA TCA TCG TGC AGC CCG AAG TGG GGC TGC TGG
TGG AGC TGG GGC AGC TGG CGT AGG TCG GCT AGT AGC ACG TCG GGC TTC ACC CCG ACG ACC
1981
2011
I L A R E F D G V L H F L M Q A K P E P
GCA TCC TGG CGC GCG AGT TCG ACG GGG TGT TGC ACT TTC TGA TGC AGG CGA AGC CGG AAC
CGT AGG ACC GCG CGC TCA AGC TGC CCC ACA ACG TGA AAG ACT ACG TCC GCT TCG GCC TTG
2041
2071
G N V N G L Q L S P T V Q A T R S N F D
CCG GCA ACG TCA ACG GGC TCC AGC TCT CCC CTA CGG TGC AGG CCA CCC GCA GCA ACT TCG
GGC CGT TGC AGT TGC CCG AGG TCG AGA GGG GAT GCC ACG TCC GGT GGG CGT CGT TGA AGC
MidB
MidL

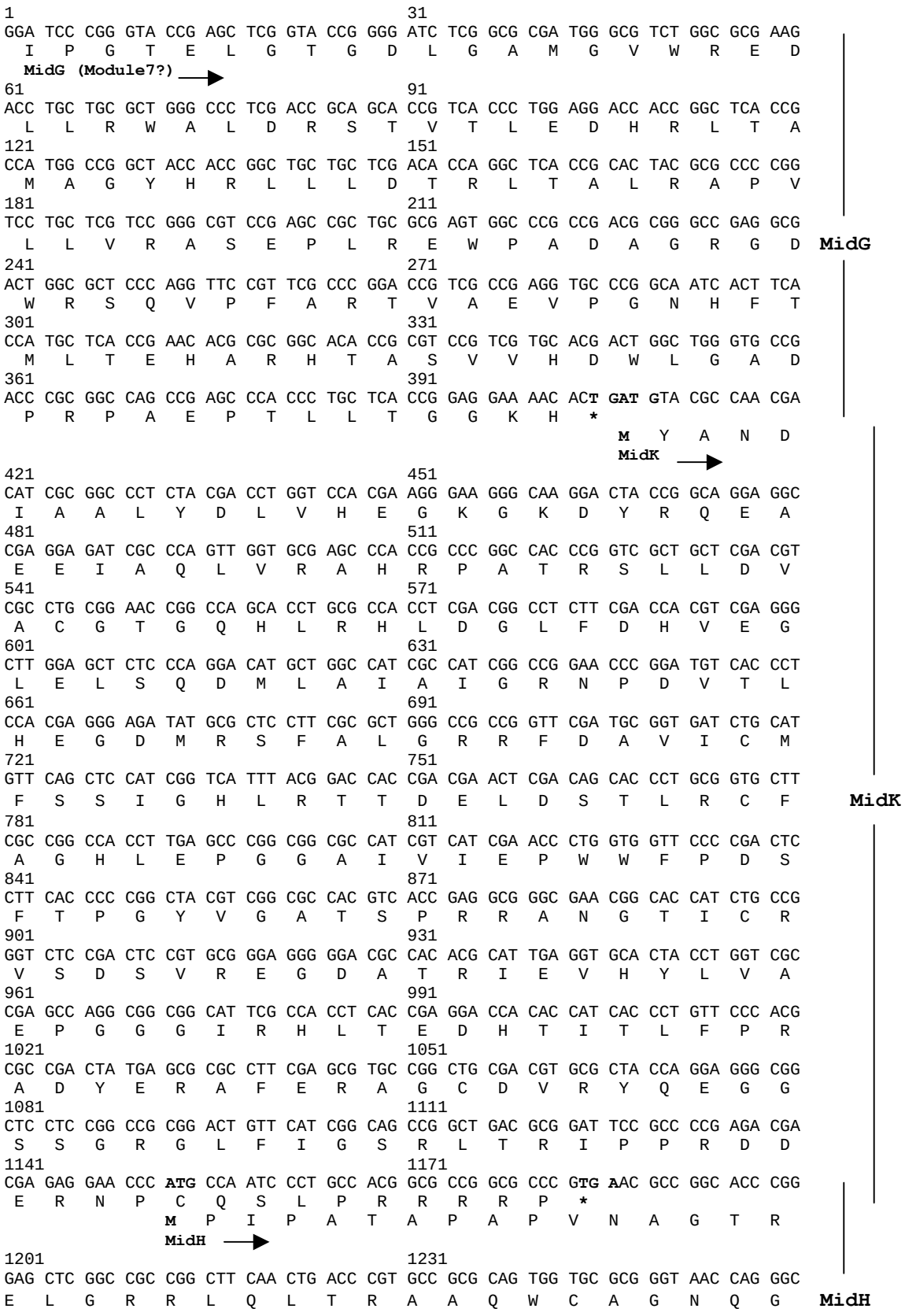
```
2101                                        2131
  E   V   H   R   G   R   S   T   P   F   L   D   R   F   I   Q   R   P   G   R
ACG AGG TGC ATC GCG GCC GCT CGA CGC CGT TCC TGG ACC GCT TCA TCC AGC GAC CGG GGC
TGC TCC ACG TAG CGC CGG CGA GCT GCG GCA AGG ACC TGG CGA AGT AGG TCG CTG GCC CCG
2161                                        2191
  R   V   L   V   D   A   I   Q   S   E   Q   A   D   W   F   L   H   K   R   N
GCC GGG TGC TGG TCG ATG CCA TTC AGT CGG AGC AAG CCG ACT GGT TCC TGC ACA AAC GCA
CGG CCC ACG ACC AGC TAC GGT AAG TCA GCC TCG TTC GGC TGA CCA AGG ACG TGT TTG CGT
2221                                        2251
  R   N   M   V   V   E   I   D   S   G   V   A   E   H   C   S   F   R   W   L
ACC GCA ACA TGG TCG TCG AGA TCG ACT CGG GCG TGG CGG AGC ACT GCT CGT TCC GCT GGC
TGG CGT TGT ACC AGC AGC TCT AGC TGA GCC CGC ACC GCC TCG TGA CGA GCA AGG CGA CCG
2281                                        2311
  T   L   G   Q   I   R   R   L   L   L   R   D   D   L   V   N   M   D   T   R
TGA CGC TCG GCC AGA TCC GTC GCC TGC TGC TCC GGG ACG ACC TCG TCA ATA TGG ACA CCC
ACT GCG AGC CGG TCT AGG CAG CGG ACG ACG AGG CCC TGC TGG AGC AGT TAT ACC TGT GGG
2341                                        2371
  S   V   L   A   C   L   P   N   C   A   R   R   T   R   R   R   R   R   R   F
GCA GCG TGC TGG CCT GCC TGC CGA ACT GCG CAC GGC GCA CCC GGC GAC GAC GAC GAA GGT
CGT CGC ACG ACC GGA CGG ACG GCT TGA CGC GTG CCG CGT GGG CCG CTG CTG CTG CTT CCA
2401                                        2431
  P   A   A   L   R   R   S   F   Y   G   E   T   E   L   N   A   I   T   G   C
TCC CGG CGG CGC TGA GGC GCT CCT TCT ACG GGG AGA CCG AGC TCA ACG CGA TCA CCG GCT
AGG GCC GCC GCG ACT CCG CGA GGA AGA TGC CCC TCT GGC TCG AGT TGC GCT AGT GGC CGA
2461                                        2491
  L   I   D   V   Q   A   L   R   V   L   R   Q   Q   K   V   P   L   N   Q   V
GTC TCA TCG ACG TCC AGG CGC TGC GTG TGC TGC GCC AGC AGA AGG TCC CGC TCA ACC AGG
CAG AGT AGC TGC AGG TCC GCG ACG CAC ACG ACG CGG TCG TCT TCC AGG GCG AGT TGG TCC  MidL
2521                                        2551
  Y   E   D   G   W   Q   R   I   G   A   T   I   R   H   R   S   G   E   G   L
TGT ACG AGG ACG GCT GGC AAC GGA TCG GGG CCA CCA TCC GGC ACC GCA GCG GCG AGG GCT
ACA TGC TCC TGC CGA CCG TTG CCT AGC CCC GGT GGT AGG CCG TGG CGT CGC CGC TCC CGA
2581                                        2611
  P   I   M   A   V   E   V   T   A   E   Q   R   E   V   A   S   W   T   Q   P
TGC CCA TCA TGG CGG TCG AGG TCA CCG CGG AGC AGC GCG AGG TGG CGT CCT GGA CCC AGC
ACG GGT AGT ACC GCC AGC TCC AGT GGC GCC TCG TCG CGC TCC ACC GCA GGA CCT GGG TCG
2641                                        2671
  L   L   A   P   V   S   Q   G   L   M   A   L   V   V   R   R   I   N   G   A
CGC TGC TGG CGC CCG TGT CCC AGG GGC TGA TGG CCC TGG TCG TCC GGC GGA TCA ACG GGG
GCG ACG ACC GCG GGC ACA GGG TCC CCG ACT ACC GGG ACC AGC AGG CCG CCT AGT TGC CCC
2701                                        2731
  L   H   A   L   V   A   A   R   S   D   V   G   T   L   N   F   A   E   F   G
CGT TGC ACG CCC TGG TGG CGG CCC GGT CGG ACG TGG GCA CGC TGA ACT TCG CCG AGT TCG
GCA ACG TGC GGG ACC ACC GCC GGG CCA GCC TGC ACC CGT GCG ACT TGA AGC GGC TCA AGC
2761                                        2791
  P   T   V   Q   L   R   S   A   W   P   R   G   K   G   N   P   P   P   Y   L
GCC CCA CCG TGC AGC TCA GGT CGG CGT GGC CGC GCG GCA AGG GCA ACC CGC CGC CGT ATC
CGG GGT GGC ACG TCG AGT CCA GCC GCA CCG GCG CGC CGT TCC CGT TGG GCG GCG GCA TAG
2821                                        2851
  E   Y   V   Q   S   A   A   P   G   R   V   R   Y   D   A   V   A   L   V   R
TAG AGT ACG TGC AGT CCG CTG CTC CGG GCC GCG TAC GGT ACG ACG CGG TGG CTC TCG TAA
ATC TCA TGC ACG TCA GGC GAC GAG GCC CGG CGC ATG CCA TGC TGC GCC ACC GAG AGC ATT
2881                                        2911
  G   W   A   L   L   S   T   R   A   T   G   S   Q   V   V   E   A   G   P   *
GAG GGT GGG CGC TTC TAT CCA CGC GCG CAA CCG GGT CAC AGG TCG TCG AGG CCG GCC CTT
CTC CCA CCC GCG AAG ATA GGT GCG CGC GTT GGC CCA GTG TCC AGC AGC TCC GGC CGG GAA
2941
  L   P   G   G   L   P   A   R   L
AAC TTC CCG GTG GAC TTC CCG CCC GGC TT*
TTG AAG GGC CAC CTG AAG GGC GGG CCG AA
```

* The sequence of *midL* is not complete at the 3’ end.

7.4. The nucleotide and encoded amino acid sequences of the *midE* region of Smyc-LC1

```
1                                                   31
GGC GGC CGG CAG CGC ACC CGT ACC TGA TGG CCA ACT CAC CTG TAC GGA CCG CTG GTT GGT
61                                                  91
GTC GGG ACA CCT CAT CGA ATG GCG CTA CGG AAC GAC GCC GCT ACG TCC GGT GAT TGC GAA
121                                                 151
ATC CAT TCT TCC TGA CGT TTT CCG GAC GCT GAC ACC ACT GTG TCA GCT GCC ACT TGC CGG
181                                                 211
GCT CAG CGG CCA TGC CCT AGA AAT CCC CTC TCA TCC ACG CCC ATT TAC CTG CGA GGT ACT
241                                                 271
GCT ATG CCC TTG CCG AAA CAC CTG CCG TCG CTC GGC GGC ATG CGG GCC ATC GCC GCA CTG
    M   P   L   P   K   H   L   P   S   L   G   G   M   R   A   I   A   A   L
    MidE ──►
301                                                 331
GTG GTG TTC TGC TCT CAT ATC GCT TCC CAG CCG TTT TTC CGC AAC GCC AAG ATA TAC TCC
V   V   F   C   S   H   I   A   S   Q   P   F   F   R   N   A   K   I   Y   S
361                                                 391
ACC GCA CAG GTC CCG CTG GAC GTC CTG GGG CCG CTG GCG GTC TCG TTC TTC TTC ATG CTC
T   A   Q   V   P   L   D   V   L   G   P   L   A   V   S   F   F   F   M   L
421                                                 451
AGC GGA TTC GTC CTC ACC TGG GCG GGC ATG CCC GAC CCG TCC AAG CCT GCC TTC TGG CGC
S   G   F   V   L   T   W   A   G   M   P   D   P   S   K   P   A   F   W   R
481                                                 511
CGC CGT TGG GTT CGG GTC TAC TCG CTG CAC CTG CCG GTC CTG CTG CTC ACG CTG GCG ATC
R   R   W   V   R   V   Y   S   L   H   L   P   V   L   L   L   T   L   A   I
541                                                 571
GTG CTG TGG CTG AAG GAA CCC AAT ATG GGC GGG TCG GTG TGG GAC GGC TTC CTC AGC AAC
V   L   W   L   K   E   P   N   M   G   G   S   V   W   D   G   F   L   S   N
601                                                 631
CTG CTG CTC GTC CAG TCG TGG TGC CCC GAC TAC CAC CAG TAC GGC AGC ATG AAC CCG GTG
L   L   L   V   Q   S   W   C   P   D   Y   H   Q   Y   G   S   M   N   P   V
661                                                 691
GCG TGG TCC CTC TCC TGC GAG ATG CTG TTC TAC GCC GCC TTC CCG TTC CTG TTC GCC TTC
A   W   S   L   S   C   E   M   L   F   Y   A   A   F   P   F   L   F   A   F
721                                                 751
TTC TCC AAG ATG CGT GCC GAG CGG CTG TGG TCC TGG GTC CTG GGC ATC TCC GTC GTC GCC
F   S   K   M   R   A   E   R   L   W   S   W   V   L   G   I   S   V   V   A
781                                                 811
GCG GCC GTG CCC GCC CTC GCC CTG CTG CTC CCC TCG GCC CCC ACG CTG CCC TGG GAC CCG
A   A   V   P   A   L   A   L   L   L   P   S   A   P   T   L   P   W   D   P
841                                                 871
AAC ATG CCG GAG CTC CAA TAC TGG TTC ATC TAC ATG CTT CGC GCG GTG CGG CTG CTG GAA
N   M   P   E   L   Q   Y   W   F   I   Y   M   L   R   A   V   R   L   L   E
901                                                 931
TTC GCG CTC GGC GGC GTC CTG ATG GCG CAG ATC GTC AGG CGC GGC CGC TGG ATC GGC CCG
F   A   L   G   G   V   L   M   A   Q   I   V   R   R   G   R   W   I   G   P
961                                                 991
ACC CCG GGG GTG TGC GCG CTG CTG TTC GCC GGC GCG TTC GCG CTG TCC TTC GCC CTG CCG
T   P   G   V   C   A   L   L   F   A   G   A   F   A   L   S   F   A   L   P
1021                                                1051
TCC TAT CTG GCT CGC GAT GCG CCG ACG GTC CCG CTG ATC GCG CTG CTG CTC GGC TCC CTG
S   Y   L   A   R   D   A   P   T   V   P   L   I   A   L   L   L   G   S   L
1081                                                1111
GCA GCT GGC GAC ATA CGC GGT ACC CGG TCG TGG CTG GGC ACC CGG ACG ATG GTG CTG CTG
A   A   G   D   I   R   G   T   R   S   W   L   G   T   R   T   M   V   L   L
1141                                                1171
GGT GAA CTC ACC TTC GCC TTC TAC GTC ATC CAC TAC CTC GTC ATC CAG TAC GGG CAC CGC
G   E   L   T   F   A   F   Y   V   I   H   Y   L   V   I   Q   Y   G   H   R
1201/401                                            1231/411
TTC CTC GGC GGT GAG CTG AGC TAC TAC CGA CAG TGG GAC ACC CCG GCC GCG ATC GGC CTC
F   L   G   G   E   L   S   Y   Y   R   Q   W   D   T   P   A   A   I   G   L
1261                                                1291
ACC GTT CTC GCC CTC GGG CTC AGC GTG GGC CTC GCC GCG CTC CTC CAC TTC TTC GTG GAG
T   V   L   A   L   G   L   S   V   G   L   A   A   L   L   H   F   F   V   E
1321                                                1351
AAG CCG GTC GTC CGG GCC CTC GGC CGC TCC GGC AAG GCG TCC CGC GCG TCC AAG GCC CCG
K   P   V   V   R   A   L   G   R   S   G   K   A   S   R   A   S   K   A   P
1381                                                1411
CAG CCC GAG CCG CCG GCG CCC CTG CTG TCC TGA GCG GGT CCG GCG GCA CAA CAG TGT GCG
Q   P   E   P   P   A   P   L   L   S   *
1441
GGG CGC GCC ACA TG
```

7.5. The nucleotide and encoded amino acid sequences of the *midGKHIJ* region of Smyc-LC3



```
1261                                                1291
GAC CCG AAC GCG CTG ATC CTG CGC GCC ACC GCC GAC CCC GCC CCG CTC GAA CGG GAG ATC
D   P   N   A   L   I   L   R   A   T   A   D   P   A   P   L   E   R   E   I
1321                                                1351
CGC GCC CGC GGA CCA TGG TTC CGC AGC GAG TTG ACC GGC GCT TGG GTG ACC GCG GAT CCG
R   A   R   G   P   W   F   R   S   E   L   T   G   A   W   V   T   A   D   P
1381                                                1411
AGG TGG CGG CAG GCC GCG CTG GCC GAC CGC GCC TTT GCA CGC TCG ACC GCG CCG ACC GTC
R   W   R   Q   A   A   L   A   D   R   A   F   A   R   S   T   A   P   T   V
1441                                                1471
GTC CGG ACG CGG AAC TGC TGC CCC TCG CAG AGG CTT TCC CCT GCC ATG AGC GTG CAG AGC
V   R   T   R   N   C   C   P   S   Q   R   L   S   P   A   M   S   V   Q   S
1501                                                1531
TCG CCC GGC TAC GGG CTG GCC GCC CCG GTG CTG AGT CGC TGC GCC CCG GCC GAG GCG CCC
S   P   G   Y   G   L   A   A   P   V   L   S   R   C   A   P   A   E   A   P
1561                                                1591
TGC GAG GCG CGT ACC GCC GCT CGT CGG TTG CTC CGC CGT CTC CTT CCC TCC GAC GGC GCC
C   E   A   R   T   A   A   R   R   L   L   R   R   L   L   P   S   D   G   A
1621                                                1651
GGG TTC GAC CTC GTC ACC GAG GTC GCC CGG CCG TAC GCC GTC GGG CTG GTG CTC CGG CTT
G   F   D   L   V   T   E   V   A   R   P   Y   A   V   G   L   V   L   R   L
1681                                                1711
CTC GGC GTG CCG GAC TGC GAC CGC GAC ACC ATG GGG CGG GCG CTC GCC GGC TGC GCT CCC
L   G   V   P   D   C   D   R   D   T   M   G   R   A   L   A   G   C   A   P
1741                                                1771
AAC TTt GAC GCC CGG TTG GCC CGC AGA CCC TGG CTG TCG CTC GGG AGT CCA CCG ACG CCG
N   F   D   A   R   L   A   R   R   P   W   L   S   L   G   S   P   P   T   P    MidH
1801                                                1831
TCC AGA CCT TGG CCG ACC ATG TCC CGG AAC TCG TTG CTG AGA AGC AGC GGG CCG TCG AGA
S   R   P   W   P   T   M   S   R   N   S   L   L   R   S   S   G   P   S   R
1861                                                1891
GCG CCG AGC CCC GGC CCG ACG ATG TTC TCG CCC TCC TCC TGC GCG ACG GTG CCG CCC CGC
A   P   S   P   G   P   T   M   F   S   P   S   S   C   A   T   V   P   P   R
1921                                                1951
GAT GTC GAG CGG ATC GCG CTG CTC CTC GCC ATC GGC ACC CCC GAG CCC GCG GCC ACC GCC
D   V   E   R   I   A   L   L   L   A   I   G   T   P   E   P   A   A   T   A
1981                                                2011
GTC GCG AAC ACG GTG CAC CGG CTG CTG AAC CGG CCG GGG GAG TGG GGA CGT GTC CGC CGG
V   A   N   T   V   H   R   L   L   N   R   P   G   E   W   G   R   V   R   R
2041                                                2071
ACC CCG GCC GCC GCG CGG GCC GTC GAC CGG ACC CTG CGC GAC CGG CCC CCG GCC CGA CTG
T   P   A   A   A   R   A   V   D   R   T   L   R   D   R   P   P   A   R   L
2101                                                2131
GAG AGC AGG GTC GCC AGC ACC GAC CTT GAG TCG cGT GGT TGC CGG ATC GCC GCC GAC GAC
E   S   R   V   A   S   T   D   L   E   S   R   G   C   R   I   A   A   D   D
2161                                                2191
CAC GTC GTG GTG CTG GCC GCC GCG GGG CGG GAC GCT CCG GGG CCC GAG CCG CTC GGC GGC
H   V   V   V   L   A   A   A   G   R   D   A   P   G   P   E   P   L   G   G
2221                                                2251
CGG ACG CGA CCG CAC TTG GCC CTC GCC CTC CCG CTC ATC CGG CTG GCC GCC ACC ACC GCT
R   T   R   P   H   L   A   L   A   L   P   L   I   R   L   A   A   T   T   A
2281                                                2311
GTC CAG GTC ATG GCC GGA CGC CTG CCC GGA CTG AGG GTC GAG GAC GAG CCT CTG ACC CGG
V   Q   V   M   A   G   R   L   P   G   L   R   V   E   D   E   P   L   T   R
2341                                                2371
CCG CGC TCC CCG GTC GTA TGC GCC TGT GCC CGC TTC CGG GTC CAC CCG GGA TGA CCC TGC
P   R   S   P   V   V   C   A   C   A   R   F   R   V   H   P   G   *
2401                                                2431
CGC CCG TAC ACC CCG GCC CGA ACT GGA GTC ACC GTG CGC GTC CTG CTG ACC TCC CTA GCC
                                            M   R   V   L   L   T   S   L   A
                                           MidI ⟶
2461                                                2491
CAC AAC ACC CAC TAC TAC AGC CTG GTG CCC TTG GCG TGG GCC CTA CGC GCG GCC GGG CAC
H   N   T   H   Y   Y   S   L   V   P   L   A   W   A   L   R   A   A   G   H
2521                                                2551
GAG GTG CGG GTG GCG AGC CCG CCC TCG CTC ACC GAT GTC ATC ACC TCC ACC GGG CTG CCC
E   V   R   V   A   S   P   P   S   L   T   D   V   I   T   S   T   G   L   P
2581                                                2611                            MidI
GCC GTC CCC GTC GGC GAC GAC CAG CCC GCC GCC GAA CTG CTC GCC GAG ATG GGC GGC GAC
A   V   P   V   G   D   D   Q   P   A   A   E   L   L   A   E   M   G   G   D
2641                                                2671
CTC GTC CCC TAT CAG CGG GGC TTT GAG TTC GCC GAG GTG GAG CCC GCC CAG GAG ACC ACC
L   V   P   Y   Q   R   G   F   E   F   A   E   V   E   P   A   Q   E   T   T
2701                                                2731
TGG GAG CAT CTG CTC GGC CAG CAG AGC ATG ATG TCC GCC TTG TGG TTC GCG CCG TTC AGC
W   E   H   L   L   G   Q   Q   S   M   M   S   A   L   W   F   A   P   F   S
```

```
2761                                                2791
GGC GCC GCC ACG ATG GAC GAC ATC GTC GAC TTC GCC CGC GAC TGG CGT CCC GAC CTC GTC
G   A   A   T   M   D   D   I   V   D   F   A   R   D   W   R   P   D   L   V
2821                                                2851
GTA TGG GAA CCC TGG ACC TAC GCC GGG CCG ATC GCG GCT CGT GCC TGC GGC GCC GCC ACC
V   W   E   P   W   T   Y   A   G   P   I   A   A   R   A   C   G   A   A   T
2881                                                2911
GCG CGT ATC CTC TGG GGC CCC GAC GCC ATC GGA CGG TCC CGG CGG CGC TTC CTC GAA GCG
A   R   I   L   W   G   P   D   A   I   G   R   S   R   R   R   F   L   E   A
2941                                                2971
CTC GAA CGA GTG CCG GAG GAG CTG CGC GAG GAC CCC ATC GCC GAA TGG CTC GGC TGG ACG
L   E   R   V   P   E   E   L   R   E   D   P   I   A   E   W   L   G   W   T
3001                                                3031
CTG GAC CGG TAC GGG TGC GCC TTC GAC GAA CGC GAC GTG CTC GGC CAC TGG GTG ATC GAC
L   D   R   Y   G   C   A   F   D   E   R   D   V   L   G   H   W   V   I   D
3061                                                3091
CCG GGG CCC CGC AGT ACC CGA CTG GAC CTG GGA CAG ACC ACG GTC CCC ATG TGC TAC GTG
P   G   P   R   S   T   R   L   D   L   G   Q   T   T   V   P   M   C   Y   V
3121                                                3151
CCC TAT AAC GGG CGC GCC GTC ATC GAA CCC TGG CTT GCC GAG AAG CCC GAG CGC CCT CGC
P   Y   N   G   R   A   V   I   E   P   W   L   A   E   K   P   E   R   P   R
3181                                                3211
GTC TGC CTC ACT CTC GGG ATC TCC GCC CGC GAG ACC TAC GGC CGC GAC GCG GTC TCC TAC
V   C   L   T   L   G   I   S   A   R   E   T   Y   G   R   D   A   V   S   Y
3241                                                3271                        MidI
TCC GAG TTG CTT CAG GCG CTG GGC CGC ATG GAG ATC GAG GTG GTG GCC ACC CTC GAT GCC
S   E   L   L   Q   A   L   G   R   M   E   I   E   V   V   A   T   L   D   A
3301                                                3331
TCG CAG CAG AAG CGC CTC GGC AGC CTT CCC GAC AAC GTC GTG CCG GTG GAC TTC GTG CCG
S   Q   Q   K   R   L   G   S   L   P   D   N   V   V   P   V   D   F   V   P
3361                                                3391
CTC GAC GCG CTG CTG CCG AGC TGT GCC GCG ATC ATC CAC CAC GGC GGC GCG GGC ACT TGG
L   D   A   L   L   P   S   C   A   A   I   I   H   H   G   G   A   G   T   W
3421                                                3451
TCC ACC GCC CTG CTC CAC GGC GTA CCG CAG ATC CTG CTG CCC GCG CTG TGG GAC GCG CCG
S   T   A   L   L   H   G   V   P   Q   I   L   L   P   A   L   W   D   A   P
3481                                                3511
CTC AAG GCC CAG CAG CTC CAG CGC CTG TCG GCC GGA CTC AAC CTG CCC GCC GCG ACC CTC
L   K   A   Q   Q   L   Q   R   L   S   A   G   L   N   L   P   A   A   T   L
3541                                                3571
ACG GCG CGC CGC TTG GCC GAC GCG GTG CAC ACG GCC GTA CAC GAT CCC GCG ATC CGG GCG
T   A   R   R   L   A   D   A   V   H   T   A   V   H   D   P   A   I   R   A
3601                                                3631
GGC GCG CGG CGG CTG CGC GAG GAG ATG CTC GCC GAC CCC ACG CCC GCG GCA ATC GTC CCC
G   A   R   R   L   R   E   E   M   L   A   D   P   T   P   A   A   I   V   P
3661                                                3691
ACG CTG GAG CGC CTC ACC GCC CTG CAC CGG GCG GCC TGA CGC AAC GTT CGA ACG GAG CCG
T   L   E   R   L   T   A   L   H   R   A   A   *
3721                                                3751
ATC CAC CAT GCC CGA CAG TCA TGC CCT GAG CGA GCT GCT CGC CGC GAG CCG CGC GCC CGA
        M   P   D   S   H   A   L   S   E   L   L   A   A   S   R   A   P   D
        MidJ →
3781                                                3811
CCA CAC CCC CGA GGA CAT CGC CGC GCT GCC CCT GCC CGA ATC CTT CCG GGC CGT GAC CGT
 H   T   P   E   D   I   A   A   L   P   L   P   E   S   F   R   A   V   T   V
3841                                                3871
CCA CAA AGA GGA CAC CGA GAT GTT CCG CGG CAT GAC CAG CGC GGA CAA GGA CCC GCG CAA
 H   K   E   D   T   E   M   F   R   G   M   T   S   A   D   K   D   P   R   K
3901                                                3931
GTC GCT GTG CGT CGA CGA GGT GCC GGT TCC CGA AGT CGG GCC CGG CGA GGC CCT GAT AGC
 S   L   C   V   D   E   V   P   V   P   E   V   G   P   G   E   A   L   I   A
3961                                                3991
GGT GAT GGC CAG CTC GGT CAA CTA CAA CAC CGT GTG GTC GTC CCT CTT CGA GCC GAT GCC  MidJ
 V   M   A   S   S   V   N   Y   N   T   V   W   S   S   L   F   E   P   M   P
4021                                                4051
GAC CTT CGG CTT CCT GGA GCG CTA CCG GCC GCA CCT CGC CGT GGC CGC TCG TCA CGA CCT
 T   F   G   F   L   E   R   Y   R   P   H   L   A   V   A   A   R   H   D   L
4081                                                4111
GCC GTA CCA CAT CCT CGG CTC CGA CCT GGC CGG CGT TGT GCT ACG CAC CGG CCC GGG GGT
 P   Y   H   I   L   G   S   D   L   A   G   V   V   L   R   T   G   P   G   V
4141                                                4171
GAA TGT TTG GGC GCC CGG CGA CGA GAT CGT GGC GCA CTG TCT AAC GGT GGA GCT GGA AAA
 N   V   W   A   P   G   D   E   I   V   A   H   C   L   T   V   E   L   E   N
4201                                                4231
CCC GGA CGG ACA CGA CGA CAC CCT GCT CGA CCC GGC CCA GCG GAT CTG GGG CTT CGA GAC
 P   D   G   H   D   D   T   L   L   D   P   A   Q   R   I   W   G   F   E   T
```

```
4261/1421                                4291
CAA CTT CGG CGG CCT GGC CGA GAT AGC CCT GGT CAA GGC CAA CCA GCT GAT GCC CAA GGC
 N   F   G   G   L   A   E   I   A   L   V   K   A   N   Q   L   M   P   K   A
4321                                     4351
CGC ACA CCT CAC CTG GGA GGA GGC CGC CGC ACC GGG TCT GGT GAA CTC CAC CGC CTA CCG
 A   H   L   T   W   E   E   A   A   A   P   G   L   V   N   S   T   A   Y   R
4381                                     4411
TCA GCT GGT CTC CCG CAA CGG CGC CGG CAT GAA GCA GCA CAA CGT GTT GAT CTG GGG CGC
 Q   L   V   S   R   N   G   A   G   M   K   Q   H   N   V   L   I   W   G   A
4441                                     4471
CAG CGG CGG TCT GGG CTC GTA CGC CAC CCA GCT CGC CCT CGC CGG TGG GGC CAA CCC CGT
 S   G   G   L   G   S   Y   A   T   Q   L   A   L   A   G   G   A   N   P   V
4501                                     4531
CTG TGT GGT CTC CAA CCA GCG CAA GGC CGA GGT GTG CCG GGC CAT GGG CGC GGG GGC GAT
 C   V   V   S   N   Q   R   K   A   E   V   C   R   A   M   G   A   G   A   I
4561                                     4591
CAT CGA CCG CTC GGC CGA GGA CTA CCG CTT CTG GAG CGA CGA GCA GAC CCA GAA TCC GCG
 I   D   R   S   A   E   D   Y   R   F   W   S   D   E   Q   T   Q   N   P   R
4621                                     4651
GGA GTG GAA GCG GTT CGG TGC CCG TAT CCG GGA GTC GAC CGG TGG TGA GGA CGT GGA CAT
 E   W   K   R   F   G   A   R   I   R   E   S   T   G   G   E   D   V   D   I   MidJ
4681                                     4711
CGT CTT CGA GCA TCC TGG CCG GGA GAC GTT CGG GGC GTC TGT CTA CGT CGC CCG CCG GGG
 V   F   E   H   P   G   R   E   T   F   G   A   S   V   Y   V   A   R   R   G
4741                                     4771
CGG CAC CAT CGT CAC CTG CGC CTC CAC TTC CGG CTA CCG TCA CGA GTT CGA CAA CCG CTA
 G   T   I   V   T   C   A   S   T   S   G   Y   R   H   E   F   D   N   R   Y
4801                                     4831
TCT GTG GAT GCA CCT CAA GCG CAT CGT CGG CAC CCA CTT CGC CAA CTA CCG CGA GGC ATG
 L   W   M   H   L   K   R   I   V   G   T   H   F   A   N   Y   R   E   A   W
4861                                     4891
GGA GGC GAA CCG CCT CGT CAC CAA AGG GAA GAT TTA CTC CAC CCT CTC CTG CAC CTA CCC
 E   A   N   R   L   V   T   K   G   K   I   Y   S   T   L   S   C   T   Y   P
4921                                     4951
GCT GGC TTT TTT CGC GCT TGC CGT CCA CGA CGT GCA CCG CAA CGT CCA CCA GGG CAA GGT
 L   A   F   F   A   L   A   V   H   D   V   H   R   N   V   H   Q   G   K   V
4981                                     5011
CGG CGT GCT GTG TCT GGC CCC GAT GGA GGG TCT GGG CGT GCG CGA CGA GGA GAT GCG CGC
 G   V   L   C   L   A   P   M   E   G   L   G   V   R   D   E   E   M   R   A
5041                                     5071
GCA GCA CCT CGA CGC GAT CAA CCG ATT CCG CTG ACC GCT CCT TTG TCC CGA GGC ATA TCC
 Q   H   L   D   A   I   N   R   F   R   *
5101                                     5131
GCC GCT CGT CCC GGA GGT TTT TTC AAA GGA GGG GCC CAC AGT CCG AAA GCG GTT TCA TGC
5161                                     5191
AGG CGC TCG GCT GGG GGT TTC CCA GCC GAG CGT TTG TTG CGT GGG CTT TTG GTC GCG ATG
5221
GCC GGC G
```

7.6. Sequence alignment of the MidD protein with other thioesterases for construction of MidD-tree in Fig. 5.2

```
MidDSmy       LVLRRL-PARPAPAVRLVCFPHAGGSATSF-LPFVQTLPDQVEMLAVQYPGRQDRRGEPLIGTIEG
ORF12Ame      KWLRRF-ERAPDARARLVCLPHAGGSASFF-FPLAKALAPAVEVLAVQYPGRQDRRHEPPVDSIGG
ThiESgr       RWLRRYHP-AEADAVRLVCFPHAGGSASFY-HPVSARFAPGAEVVSLQYPGRQDRRKEPCVPDLGT
PikAVSve       LWIRRFHP-APNSAVRLVCLPHAGGSASYF-FRFSEELHPSVEALSVQYPGRQDRRAEPCLESVEE
TylOSfr       MWLRCYAPVPGTPAHRLVCFPHAGGSARAYRPFALELAAAGVETHAVQYPGRQDRRKEPFARTLEE
GrsTBbr       KWFVNAN-VNSAAKLRLFCIPYAGGGASAF-YEWSHFFPKEIEVCSIQLPGRENRGAEVPLTNLQQ
TesAMtu       GHSDDSSGDAKQAAPTLYIFPHAGGTAKDY-VAFSREFSADVKRIAVQYPGQHDRSGLPPLESIPT
MLCB12.04C    GHSNNGNDDETSTTPTLYIFPHAGGDATYY-VPFSREFSADIKRIAVHYPGQRDGYGLPALTSIPA
Orf1Shy       MTDWIQSVSAPDAVARVVCLSRAGGSARDF-DRWRAPMGEDVELAAVQLPGRLDRFHEPPLSDLHE
PchCPae       AAWVRPFRLTPMPRLRLACFPHAGGSASFF-RSWSERLPPDIDLLALQYPGREDRFNEAPATRLED
SC3F7.14Sco   AAPAAEPPDPAAAPLRLVCFPYAGGTVSAF-RGWQERLGDEVAVVPVQLPGRGLRLRERPYDTMEP
Ery3Ser       STWLRRFGPPVEHRARLVCFPHAGAAADSY-LDLARALAPEIDVDAVQYPGRQDRRDEEPLGTAGE
SrfADBsu      SQLFKSFDA--SEKTQLICFPFAGGYSASF-RPLHAFLQGECEMLAAEPPGHGTNQ-TSAIEDLEE

MidDSmy       LVEPLAEVLATHSDRP----LVLFGHSMGATVAYEVARVLQQRGA-APAGLVVSGRRAPIVNRPMT
ORF12Ame      LTNRLLEVLRPFGDRP----LALFGHSMGAIIGYELALRMPEAGLPAPVHLFASGRRAPSRYRDDD
ThiESgr       LDLITEQLL-PLDERP----TVFFGHSMGAALAFETAWRLEQKGA-GPRTVIASGRRGPSTTRAER
PikAVSve       LAEHVVAATEPWWQEGR---LAFFGHSLGASVAFETARILEQRHGVRPEGLYVSGRRAPSLAPDRL
TylOSfr       LAERVLPELRRLLDAPDGVPVALFGHSMGAVVAYETARLLHRSGAPRPAGLILSGRRAPTADRTET
GrsTBbr       IVEIVAEEIQPLINIPF----AFLGHSMGALISFELARTIRQKSNVNPVHLFVSGRHAPQIPCAKQ
TesAMtu       LADEIFAMMKPSARIDD--PVAFFGHSMGGMLAFEVALRY-QSAGHRVLAFFVSACSAPGHIRYKQ
MLCB12.04C    LADEIFAIMKPSAPPEG--AVAFFGHSMGGMLAFEVALRF-QSAGYRLIALFVSACSAPGYIRYKQ
Orf1Shy       IAEEVAAALTTLPARP----YVLFGDCMGALLAFETACALRRRGAAPPDCLVVASYPAPDRLRTER
PchCPae       LADGAALALRDFADAP----LALFGHSLGAALAYETALRLET-PAALR-HLFVSAHPAPHRQRGGA
SC3F7.14Sco   LAEAVADALEEHRLTHD---YALFGHSMGALLAYEVACVLRRRGAPRPRHLFVSGSRAPHLYGDRA
Ery3Ser       IADEVAAVLRASGGDG---PFALFGHSMGALIAYETARRLEREPGGGPLRLFVSGQTAPRVHERRT
SrfADBsu      LTDLYKQELNLRPDR----PFVLFGHSMGGMITFRLAQKLEREGIFPQAVII-SAIQPPHIQRKKV

MidDSmy       -VHLYDDDRLLAELRSLEGTDESLLNDPELLQMVLPAIRNDYRAVGTYTHPPGAPLASALTVFTG
ORF12Ame      -VRGASDERLVAELRKLGGSDAAMLADPELLAMVLPAIRSDYRAVETYRHEPGRRVDCPVTVFTG
ThiESgr       -VHTRDDDGIVAEMKRLNGTAAGVLGDEEILRMALPALRGDYRAIETYTCPPDRRLACGLT-LTG
PikAVSve       -VHQLDDRAFLAEIRRLSGTDERFLQDDELLRLVLPALRSDYKAAETYLHRPSAKLTCPVMALAG
TylOSfr       -AHLLGDRELLAEIRRLQGTDPGALADEEVLRMVLPAIRGDYAAVGRYRHVPGPRPGCPLTVFTG
GrsTBbr       DYHLLPDEQFIQELRSLNGTPEIVLQDAEMMSILLPRLRADFSVCGSYQYKNDEPFECPITAFGG
TesAMtu       -LQDLSDREMLDLFTRMTGMNPDFFTDDEFFVGALPTLRAV-RAIAGYSCPPETKLSCPIYAFIG
MLCB12.04C    -IKDFSDNDMLDLVVRMTGMNPDFFEDEEFRVGVLPTLRAA-RIIAGYNCPPETTVSCPIYTYIG
Orf1Shy       PYGDGSADDLRQRLREVGGVPPAVLDEDELFELMLPMLRADFAAFEGYRHRPTEPLSVDIHALVG
PchCPae       -LHRG-DEAALLEDVRRQGGASELLEDADLRALFLPILRADYQAIETYRRAQPIALACALDVLLG
SC3F7.14Sco   -DHTLSDTALREVIRDLGGLDDADTLGAAYFDRRLPVLRADLRACERYDWHPRPPLDCPTTAFSA
Ery3Sery      -DLPGDDG-LVDELRRL-GTSEAALADEALLAMSLPVLRADYRVLRSYAWADGPPLRAGITALCG
SrfADBsu      -SHLPDDQFLDHIIQ-LGGMPAELVENKEVMSFFLPSFRSDYRALEQFELYDLAQIQSPVHVFNG

MidDSmy       ADDPNVTATEAAAWQAVAEAGAQVRTFPGGHFFLYQQVAEVCGALMD-TLAPLLPAG
ORF12Ame      DHDPRVSVGEARAWEEHTTGPADLRVLPGGHFFLVDQAAPMIATMTE-KLAG--PAL
ThiESgr       EDDPLTTVEEAERWRDHTTGPFRLRVFTGGHFFLTQHLDAVNTEIAQ-ALHPD-RAA
PikAVSve      DRDPKAPLNEVAEWRRHTSGPFCLRAYSGGHFYLNDQWHEICNDISD-HLLVTRGAP
TylOSfr       DADPNVTLPEAEAWRELTTGAFALRVFPGGHFYLNDQREAVCRTIEE-TLRHGSKSA
GrsTBbr       KNDNGVTYQSLEAWREQTKREFSVCMYPGDHFFLYESKYEMIEFMCK-QLRLVLAPK
TesAMtu       DKDWIATQDDMDPWRDRTTEEFSIRVFPGDHFYLNDNLPELVSDIEDKTLQWHDRA-
MLCB12.04C    DKDWIATQEDMKPWRERTTGAFAIRVFPGDHFYLNGNLSELVCDIEDKTLEWCDRA-
Orf1Shy       ADDPYVTVTDLHGWQRHTTGEFTARALPGGHFFLHESDDAVSRVRSL-ALAGARAAR
PchCPae       EHDEEVSAAEAQAWSDASRTPARLRRFPGGHFYLSEGRDAVIEHLLR-RLAHPDALS
SC3F7.14Sco   AADPIATPEMVEAWRPYTTGSFLRRHLPGNHFFLNGGPSRDRLLAHLGTELDALGTT
Ery3Ser       DADPLTATGDAERWLQHSVIPGRTRTFPGGHFYLGEQVTEVAGAVRRDLLRAGLAG-
SrfADBsu      LDDKK-CIRDAEGWKK-WAKDITFHQFDGGHMFLLSQTEEVAERIFA-ILNQHPIIQ
```

7.7. Sequence alignment of the MidA protein with other dTDP-glucose synthases for construction of MidA-tree in Fig. 5.2

```
StrDSgr         M----KALVLAGGTGTRLRPITHTSAKQLVPVANKPVLFYGLEAIRAAGIIDVGIVVG-DTAD
OleDSan1        V----KALVLAGGSGTRLRPITHTSAKQLVAVANKPVLFYGLEAIAAAGITDVGLIVG-DTAG
OleDSan2        M----KALVLAGGSGTRLRPITHTSAKQLVAVANKPVLFYGLEAIAAAGITDVGLIVG-DTAG
GraDSvi         M----KALILSGGMGDRLRPFTYSMPKQLVPVANKPILVHCLENVRAIGVEEVAVVVG-DRAD
DnrLSpe         M----KALVLSGGSGTRLRPFTHTSPKQLVPVANKPVLYYVLEDIAQASITEVGIVVG-ETSN
RmlAMle         M----RGIILAGGSGTRLYPITLGISKQLLPVYDKPMIYYPLSTLMMAGIRDILVITTAHDAP
TylAISfr        M----KGIILAGGSGTRLRPLTGTLSKQLLPVYDKPMIYYPLSVLMLAGIREIQIISSKDHLD
MidASmy         M----KGIILAGGGGTRLRPLTGTLSKQLLPVYNKPMIYYPLSVLMLGGINEILIISTPDHIL
AcbAAct         MVGHVRGILLAGGTGSRLRPVTWAVSKQLMPVYDKPMIYYPLATLVSCGIREILVITTETEAA
YifGEco         M----KGIMLAGGSGTRLHPITRGVSKQLLPIYDKPMIYYPLSVLMLAGIREILIITTPEDKG
RmlAEco         MKTR-KGIILAGGSGTRLYPVTMAVSKQLLPIYDKPMIYYPLSTLMLAGIRDILIISTPQDTP
LmbOSli         M-T---LVVLAGGRGTRLGAYA-TTAKALLPVDGEPFLVRALRRYRAGAASPMSCCAPGHHAA

StrDSgr         EIVAAVGDGSRFGLKVSYIPQSKPLGLAHCVLISRDFLGEDDFIMYLGDNFVVGVVEDSVREF
OleDSan1        EVRAPVGDGAKFGLDITYIEQSRPLGLAHAVLIAHTYLGDDDFVMYLGDNFIVGGIDDLVRTF
OleDSan2        EVRAAVGDGAKFGLDITYIEQSRPLGLAHAVLIAHTYLGDDDFVMYLGDNFIVGGIDDLVRTF
GraDSvi         EVRAVVGDGSAFGLDVTYLQQEAPLGLAHCVSIAEEFLGDEDFVMYLGDNILAEGIAESARAF
DnrLSpe         EIRKAVGSGDRFGLRVTYLPQEAPLGLAHAVLIARDYLGEEDFVMYLGDNFVVGGIAGNSSTF
RmlAMle         GFKRLLGDGTQFGVNISYATQDHPDGLAQAFVIGANHIGADTVALVLGDNIFYGPGLGTSLR-
TylAISfr        LFRSLLGEGDRLGLSISYAEQREPRGIAEAFLIGARHIGGDDAALILGDNVFHGPGFSSVLTG
MidASmy         EQFSGCWRRVRLGLDITYAESPSPRGIAQALTIGSDHIGNSPVALILGDNIFHGPGFSSVLQG
AcbAAct         QFQRLLGDGSQWGLRLEFAVQQRPGGIAEAFLIGEEFLAGGPVALMLGDNLLHGVDFRPCVQR
YifGEco         YFQRLLGDVGEFGIQLEYAEQPSPDGLAQAFIIGETFLNGEPSCLVLGDNIFFGQGFSPKLRH
RmlAEco         RFQQLLGDGSQWGLNLQYKVQPSPDGLAQAFIIGEDFIGGDDCALVLGDNIFYGHDLPKLMEA
LmbOSli         QVQEVIGDGSALGLRVTHSAEPEPLGPIGALRHALPLLPET----YLLTYCDVVPTIDV----

StrDSgr         RAAR-P--DAHLMLTRVPEPRSFGVAELS-DSGQVLGLEEKPAHPKSDLALVGVYLFSPAIHE
OleDSan1        RDGRRP--AARI-----------------------------------GMVGVYFFTPPSTR
OleDSan2        RDGRPP--AARILLTHVSDPSAFGVAELD-DDGRVVGLEEKPRHPKSDLALVGVYFFTPAIHE
GraDSvi         RDER-S--AARLLLTKVADPRAYGVAETD-ATGRVHALVEKPERPRSDLAVIGVYFFTAAVHD
DnrLSpe         RAER-P--DAQILLTRVSDPSSFGVAEIG-CDGRVVALEEKPRHPRSDLAVVGIYLFTPVVHE
RmlAMle         RFQYVS--GGAIFAYCVANPSSYGIVELGID-GIALSLEEKPATPKSQYAVPGLYFYDNDVVE
TylAISfr        TVARLD--GCELFGYPVKDAHRYGVGEIDSG-GRLLSLEEKPRRPLEP-GRHRLYLYTNDVVE
MidASmy         SIRHLD--GCVLFGYPVSDPGRYGVGEIDRD-GLLLSLEEKPVRPRSNLAVTGLYLYDNDVVD
AcbAAct         ARET-A--GGHVFGVAVADPSAYGVVEFDAA-GRVLSIEEKPVRPRSPYAVPGFYLYDADVVE
YifGEco         VAARTE--GP-VFGYQVMDPERFGXVEFD-DNFRAISLEEKPKQPKSNWAVTGLYFYDSKVVE
RmlAEco         AVNKES--GATVFAYHVNDPERYGVVEFD-NNGTAISLEEKPLEPKSNYAVTGLYFYDNDVVE
LmbOSli         R--------------------SFTAAARASDCPAVMAVGTAPTPPEANVLLSG-----QQVSS

StrDSgr         AVAA-ITPSWRGELEITDAVQWLIDAGRDVRSTVISGY-WKDTGNVTDMLEVNRLVLE
OleDSan1        PSAP-IEPSWRGELEITHAIQHLIDNGADIQSMVIEGY-WKDTGNVADMLEVNRTVLE
OleDSan2        AVRA-IEPSWRGELEITHAIQHLIDNGADIQSMVIEGY-WKDTGNVADMLEVNRTVLE
GraDSvi         AVRA-IEPSARGELEITDAIQYLVERGDRVVADEYTGY-WKDTGSPDDLLDCNRVLLG
DnrLSpe         AVRA-LTPSRRGELEITDALQWLLDGPYDVRYTTISGY-WKDTGNVADMLEVNRAVLD
RmlAMle         IAR-GLTKSARGEYEITEVNQIYLNQGRLTVEVLARGTAWLDTGTFDSLLDASDFV--
TylAISfr        IAR-TISPSARGELEITDVNKVYLEQGR-AAHGAGAVVAWLDMGTHDSLLQAGQYV--
MidASmy         IAK-NIRPSARGELEITDVNKVYLEQRRARLIELGHGFAWLDMGTHDSLLQASQYV--
AcbAAct         TAR-SLRPSARGELEITEVNQAYLRRGALSVTLLGRGAG-LARHRHPGRLHARGRL--
YifGEco         YAKQ-VKPSERGELEITSINQMYLEAGNLTVELLGRGFAWLDTGTHDSLIEASTFV--
RmlAEco         MARKNLKPSARGELEITDINRIYMEQGRLSVAMMGRGYAWLDTGTHQSLIEASNFI--
LmbOSli         YAKSPPPGATHCDRGLLALERRLLDRHP----------GRTEADFYGALARRGELGAV
```

```
StrDSgr         TT-EPRCDGLVDERSGLIGRVLVEEGAEVRNSRVMGPTVIGAGTRVTNSYVGPFTSLAEDC
OleDSan1        DL-EPRIEGTVDEHTVVIGRVVVGEGARVTNSRIMGPAIIGAGPEISDSYIGPFTSVGDNC
OleDSan2        DL-EPRIEGTVDEHTVVIGRVVVGEGARVTNSRIMGPAIIGAGAEISDSYIGPFTSVGDNC
GraDSvi         RL-RPGVHGEVDAASTVEGTVVVEAGRSWSDSRLVGPLVVGAGSVVRGSELGPYTALGRDC
DnrLSpe         GI-EPGMEGQADAASELVGRVRIEAGAQIRASRIVARRHRAG--RVTDR-TSALHVDRGDC
RmlAMle         --------RTLERRQGL--KVSVPEEVSWRMGWIDDEQLALRAHSLAKSGYGCY-------
TylAISfr        --------QLLEQRQGE--RIACIEEIAMRMGFISAEQCYRLGQELRSSSYGSY-------
MidASmy         --------QLLEQRQGV--RIACVEEIALRMGFINADELYLLGCELGNSGYGSY-------
AcbAAct         --------RARHRRGP-----GHQDRLCGGGGLAGRFPRHRARACPRRAVDEQR-------
YifGEco         --------QTVEKRQGF--KIACLEEIAWRNGWLDDEGVKRAASSLAKTGYGQY-------
RmlAEco         --------ATIEERQGL--KVSCPEEIAYRKGFIDAEQVKVLAEPLKKNAYGQY-------
LmbOSli         RIGAPGADIGTAHR-----------------------------------YERYLRTGEK-

StrDSgr         VVEDSEVEFSIVLRGASISGVR-RIEASLIGRHVQVTSAPEVPHAHRLVLGDHSRAQISS
OleDSan1        RITGSEMEFSIMLAESAITGVR-RIEGSLIGRNVQVTQSLHAPNAHRFVLGDHSKVEIQS
OleDSan2        RITGSEMEFSIMLAESAITGVR-RIEGSLIGRNVQVTQSLHAPNAHRFVLGDHSKVEIQS
GraDSvi         VLEDAGIRDSIVLDGVSIQGVR-GLSGSLIGRSAAV-RTGEAAGRR-LIIGDHTQAEVAA
DnrLSpe         SIETARSSLHHA-GRLPAHGTR-RVQHSLLGRNVTVAPAPRVPAGSRLILGDDSRVEISS
RmlAMle         ----------------------------LSELLERGXFRQAXPTPRRLLWSTGLTEQALC
TylAISfr        ----------------------------IIDVAMRGAAA-----------D-SRAQ---
MidASmy         ----------------------------LMEVASHAGAA---------------------
AcbAAct         ----------------------------LRTVPAGSDRRRAQPYPQ-WPA---LTAAAG-
YifGEco         ----------------------------LLELL---RARPRQ-Y----------------
RmlAEco         ----------------------------LLKMI---K--G---Y----------------
LmbOSli         ------------------------------------------------------------
```

7.8. Sequence alignment of the MidB protein with other dTDP-glucose 4,6-dehydratases for construction of MidB-tree in Fig. 5.2

```
SCF81.08c       -NILVTGAAGFIGSRYVRGLLASDAPG---APRVTVLDALTYAGSTANFTLELGHPRLEF
StrESgr         -HLLVTGAAGFIGSQYVRTLL-G--PGGPPDVVVTALDALTYAGNPDNLAAVRGHPRYRF
StrESgl         TRLLVTGGAGFIGSHYVRTLL-G--PDGPPDAVVTVLDALSYAGNLANLDPVRDHPRLRF
OleESan         -NLLVTGAAGFIGSRYVHHLLEATRRGREPAPVITVLDKLTYAGV---LGNVPDDPAVTF
GraESvi         -RLLVTGAAGFIGSHYVREILAGSYPESDDVH-VTVVDRLTYAGRRDNLPE—-HHERLDF
TylAIISfr       -RVLVTGGAGFIGSHFTGQLLTGAYPDLGATRTV-VLDKLTYAGNPANLEHVAGHPDLEF
MidBSmy         QRILVTGGAGFIGSRFVNALLNGSLPEFGKPEVV-VLDALTYAGNLANLAPVGDCGGLRV
GdhSer          -RVLVTGGAGFIGSHYVRQLLGGAYPAFAGADVV-VLDKLTYAGNEENLRPVADDPRFRF
AcbBAct         -KILVTGGAGFIGSHFVTSLISGDIATPQPVTQVTVVDKLGYGGNLRNLAEASADPRFSF
AcbDSgl         TTILVTGGAGFIGSAYVRRLLSPGAP--GGV-AVTVLDKLTYAGSLARLHAVRDHPGLTF
RffEEco         RKILITGGAGFIGSALVRYIIN------ETSDAVVVVDKLTYAGNLMSLAPVAQSERFAF
RmlBEco         -KILVTGGAGFIGSAVVRHIIN------NTQDSVVNVDKLTYAGNLESLAEISDSERYSF
LmbMSli         GRYCVHRGAGFIGSPTSSKRLLAE----EGTRGVVAYDDLSNTTT-RWIEPLLADERLRF

SCF81.08c       VHGDIRDAALVDRLTAGA--DQVVHFAAESHVDRSIHAASDFVLTNVVGTQFTNLLDAAL
StrESgr         ERGDICDAPG-RRVMAG-Q-DQVVHLAAESHVDRSLLDASVFVRTNVHGTQ-T-LLDAAT
StrESgl         VHGDICDADLVDRVMAG-Q-DQVVHLAAESHVDRSLLDAAAFVRTNAGGTQ-T-LLDAAL
OleESan         VRGDIADAPLVDSLMAEA--DQVVHFAAETHVDRSITSPGTFVRTNVLGTQV--LLDAAL
GraESvi         VHGDICDRDLLDRVLPGH--DAVVHFAAESHVDRSLTGPGEFVRTNVMGTQ-Q-LLDAAL
TylAIISfr       VRGDIADHGWWRRLMEGV--GLVVHFAAESHVDRSIESSEAFVRTNVEGTRV--LLQAAV
MidBSmy         VAWDICDRSTVALAMAGA--DLVVHFAAESHVDRSIDDADAFVRTNVLGTHV--LLREAL
GdhSer          VRGDICEWDVVSEVMREV--DVVVHFAAETHVDRSILGASDFVVTNVVGTN-T-LLQGAL
AcbBAct         VRGDICDEGLIEGLMARH--DTVAHFAAETHVDRSVVASGPFVASNLVGTQV--LLDAAL
AcbDSgl         VQGDVCDTALVDTLAARH--DDIVHFAAESHVDRSITDSGAFTRTNVLGTQV--LLDAAL
RffEEco         EKVDICDRAELARVFTEHQPDCVMHLAAESHVDRSIDGPAAFIETNIVGTY-T-LLEAAR
RmlBEco         ENADICDAEGDGLYFGQHQLDAVMHLAAESHVDRSITGPAAFIETNIVGTYV--LLEAAR
LmbMSli         VRADVLDTA---RL-TEELADDVCHLASSVDMRKGYHDRGFDLRQCAEGTL-S-VLNAMR
```

```
SCF81.08c       ---------RHGVGPFVHVSTDEVYGSVDAGS----------ATEEHPLRPSSPYSASKA-
StrESgr         ---------RHGVASFVQVSTDEVYGSLEHGS----------WTEDEPLRPNSPYSASKA-
StrESgl         ---------RHGVAPFVQVSTDEVYGSLETGS----------WTEDEPLRPNSPYATSKA-
OleESan         ---------RHGVGPFVHVSTDEVYGSIEHGS----------WPEHQPLCPNSPYSASKA-
GraESvi         ---------HAGVDRVLHVSTDEVYGSLDSGT----------WTEDSPLLPNSPYAASKA-
TylAIISfr       ---------DAGVGRFVHISTDEVYGSIAEGS----------WPEDHPVAPNSPYAATKA-
MidBSmy         ---------AVRPGRFVHVSTDEVYGSIPEGS----------WSEDHPLSPNSPYAASKA-
GdhSer          ---------AANVSKFVHVSTDEVYGTIEHGS----------WPEDHLLEPNSPYSAAKA-
AcbBAct         ---------RHHIGRFLHVSTDEVYGSIDTGS----------WAEGHPLAPNSPYAASKA-
AcbDSgl         ---------RHGVRTFVHVSTDEVYGSLPHGA----------AAESDPLLPTSPYAASKA-
RffEEco         AYWNALTEDKKSAFRFHHISTDEVYGDLHSTDD--------FFTETTPYAPSSPYSASKA-
RmlBEco         NYWSGLDDEKKKNFRFHHISTDEVYGDLPHPDEVNSNETLQLFTETTAYAPSSPYSASKA-
LmbMSli         A--SG-P--RTVLFS----SSSTVYGD---------PVTLPTPEHAGPYAPHLDVRGREAL

SCF81.08c       SGDLLALSYHRFTTHGLDVRVTRCSNNYGPHQFPEKLVPLFVTRLLDGHRVPLYGDGRNV
StrESgr         SGDLLALAHHV-S-HGLDVRVTRCSNNYGPRQFPEKLIPRFITLLMDGHRVPLYGDGLNV
StrESgl         SGDLLALAMHV-S-HGLDVRITRCSNNYGPYQFPEKLVPRFVTLLLEGRKVPLYGDGLHV
OleESan         SSDLLALSYHR-T-HGLDVRVTRCSNNYGPHQFPEKIVPLFVTNLLDGLRVPLYGDGLNV
GraESvi         STTWSAAPTTV-R-HGLDVRITRCSNNYGPRQHPEKLIPNFVTRLLTGRQVPLYGDGRNV
TylAIISfr       ASDLLALAYHR-T-YGLDVRVTRCSNNYGPRQYPEKAVPLFTTNLLDGLPVPLYGDGGNT
MidBSmy         ASDQLALAFHR-T-HGLPVCVTRCSNNYGPYQYPEKIIPLFVSNLLDGAAVPLYGDGGNR
GdhSer          GSDLIARAYHR-T-HGLPVCITRCSNNYGPYQFPEKVLPLFITNLMDGRRVPLYGDGLNV
AcbBAct         GSDLLALAYHQ-T-HGMDVVVTRCSNNYGPRQFPEKMIPLFVTRLLDGLDVPVYGDGRNI
AcbDSgl         ASDLMALAHHR-T-HGLDVRVTRCSNNFGPHQHPEKLIPRFLTSLLSGGTVPLYGDGRHV
RffEEco         SSDHLVRAWLR-T-YGLPTLITNCSNNYGPYHFPEKLIPLMILNALAGKSLPVYGNGQQI
RmlBEco         SSDHLVRAWKR-T-YGLPTIVSNCSNNYGPYHFPEKLIPLVILNALEGKALPIYGKGDQI
LmbMSli         AAEGLLSANCHL--DGFTAHVFRFGNVVGGRMNHGVIHDFIEKLDARRVRLQVLGDGRQR

SCF81.08c       RDWLHVDDHCRGFTVDLVRTRGRAGE-VYNIGGGTELSNRDLTGL---LLDACGAG------
StrESgr         REWLHVDDHVRGIEA--VRTRGRAGR-VYNIGGGATLSNKELVGL---LLEAAGAD------
StrESgl         RDWLHVDDHVGGIEA--VRARGRAGR-VYNIGGGTSLANRDLVDL---LLKACGAG------
OleESan         REWLHVDDHCLGVDL--VRTQGRPGE-VYHIGGGTELTNRDLTGL---LLDAFGVG------
GraESvi         REWLHVDDHCRALQL--VLTKGRAGE-IYNIGGGSGMSNREMTAR---LLDLLGAD------
TylAIISfr       REWLHVDDHCRGVAL--VGAGGRPGV-IYNIGGGTELTNAELTDR---ILELCGAD------
MidBSmy         RDWLHVDDHCRGIAL--VARGGRPAK-VYNIGGGTELTNTELTER---PAEAVRTD------
GdhSer          RDWLHVTDHCRGIQL--VAESGRAGE-IYNIGGGTELTNKELTER---VLELMGQD------
AcbBAct         RDWLHVSDHCRGLAL--ALGAGRAGE-VYHIGGGWEATNLELTEI---LLEACGAR------
AcbDSgl         RDWLHVDDHVR--AVELVRVSGRPGE-IYNIGGGTSLPNLELTHR---LLALCGAGP-----
RffEEco         RDWLYVEDHARALYC--VATTGKVGET-YNIGGHNERKNLDVVETICELLEELAPNKPHGVA
RmlBEco         RDWLYVEDHARALYT--VVTEGKAGET-YNIGGHNEKKNIDVVFTICDLLDEIVPKEKS---
LmbMSli         KNYFLVEECVDGI----LTASGKLGPG-FHVLNLGNPGTVSVDEIAAIVIDEMGLKGV-GLE

SCF81.08c       --PDRIVHVEDRKGHDLRYSVDWFTSKAREELGYRPHRDLATGLAETVAWYRDNRAWWEPL
StrESgr         --WGSVEYVEDRKGHDRRYAVD--STRIQRELGFAPAVDLADGLAATVAWYHKHRSWWEPL
StrESgl         --WDRVEHVPDRKGHDRRYSVD--ASRIRRELGHVPATDLSTGLAATVAWYRDNRAWWEPL
OleESan         --WDVVDPVADRKGHDRRYALD--CAKAADELGYRPRRDFAEGIARTIDWYRDNRAWWEPL
GraESvi         --WDMVRHVEDRLGHDFRYAID--DSKIREELGYAPRWSIESGLGAVVDWYRDHPDFWRAP
TylAIISfr       --RSALRRVADRPGHDRRYSVD--TTKIREELGYAPRTGITEGLAGTVAWYRDNRAWWEPL
MidBSmy         --WSAVREVPDRKGHDRRYSVD-Y-AKIANELGYAPRIGIDEGLADTVRWYRENRAWWKPL
GdhSer          --WSMVQPVTDRKGHDRRYSVD--HTKISEELGYEPVVPFERGLAETIEWYRDNRAWWEPL
AcbBAct         --RSRISFVTDRKGHDRRYSLD-Y-SKIAGELGYRPRVDFTDGIAETVAWYRANRSWWTDG
AcbDSgl         ---ERIVHVENRKGHDRRYAVD--HSKITAELGYRPRTDFATALADTAKWYERHEDWWRPL
RffEEco         HYRDLITFVADRPGHDLRYAIDA--SKIARELGCVPQETFESGMRKXVQWYLANESWWKQV
RmlBEco         -YREQITYVADRPGHDRRYAIDA-D-KISRELGWKPQETFESGIRKTVEWYLANTNWVENV
LmbMSli         --HEGG--VRGWPGDVPVVEYDL-T-RV-HELGWSAPTDGRQAIRTCARRLLAERGWQRP-
```

7.9. Sequence alignment of the MidH protein with other putative isomerases for construction of MidH-tree in Fig. 5.2

```
MidHSmy      VNAGTRELGRRLQLTRAAQWCAGIQGDPYALILRATADPAPL-EREIRARGPWFRSELTGAWVT
DnrQSpe      AAPTDSELGRHLLTVRGFHFVFGALGDPYARRLRGEADHLSL-GELVRDRGP-LHGSALGTWVT
TylMIISfr    AHDLSR-AGRRLQLTRAAQWFAGNQGDPYGMILRAGTADPAPYEEEIRERGPLFHSELLGTWVT
OleP1San     MED-S-ELGRRLQMLRGMQWVFGANGDPYARLLCGMEDDPSPFYDAIRTLGE-LHRSRTGAWVT
DesVIISve    RAVADRELGTHLLETRGIHWIHAANGDPYATVLRGQADDPYPAYERVRARGA-LSFSPTGSWVT
SnogNSno     MKLTDSELGRALLSLRGYQWLRGIHHDPYALLLRAESDDPAQLGRLLRERGR-LHRSDTGTWVT
Pra10Ahi     MPSSKDAPTVDPRPDVTPAFPFRPD-DPF----QPPCEHARLRASDPVAKVV-LPTGDH-AWVV

MidHSmy      ADPRWRQAALADRAFFDLVTEVAR-PYAVGLVLRLLGVPDCDRDTMGRALAGCAPNFDARLARR
DnrQSpe      ADGGISARLLDDPLLFDVVSDLAR-PAIAGSLAAVLGLPDEARAELPDLLAACGPVLDSALCPP
TylMIISfr    GSRHVADAVTADDAFLGVPAE--R-----SALDAAHGNPGGPLPRFETALTGCRRALDALLCPQ
OleP1San     ADPGLGGRILADRKAFDLVEEYAG-PVEVLA--RIWGVPEEDRARFGRDCRALAPALDSLLCPQ
DesVIISve    ADHALAASILCSTDFFVRPAVTLV-PYAAAA-AAVLGVPADRRADFADLLERLRPLSDSLLAPQ
SnogNSno     ADHATASRLLADPRFFDLRADYAL--VEAAC--ALLGLPAGQCSLFG-AFS-PAVLLDATVVPP
Pra10Ahi     TRYADVRFVTSDRRF--VICEMLGVPPEDRPRFQDWTDRMLTIGAPALAQADEIKAAVGRLRGY

MidHSmy      PWLSLGSPPTPSRPWPTMSRNNRPGEWGRVRRTPA-AARAVDRTLRDRPPARLESRVASTDLES
DnrQSpe      RLPVARAMTQALRRVRELMAAKHDEQWSLLRADPGRAADAVEETLRWAPPVTLRSLITQGEVQI
TylMIISfr    LLADARAGLAAEEALRAVLGEERPAQWRALTADPGLAGAAITETLLWAPPVRLESRVARETAVL
OleP1San     QLALSKDMASALEDLR-LLFD--TGRVAAGQVAGQALHRAVSYRIATRFA-REDLELAGCEVKS
DesVIISve    SLRTVRAADGALAELTALLADDPPVQLDARVVRGETELAGRRLPAGAHVVVLTAATGRDPEVFT
SnogNSno     RLPEARALIASTAELTALWPRAAPA-VEETLRHAPPARLFTLHATGPER-VADVDLPAGAEVA-
Pra10Ahi     LAELIDAKTAAPAD-DLLSLLGIPAAVEELLRYGQIGGGAGAIRIAVEDVEVGGTLVRAGEAV-

MidHSmy      RGC-RIAADDHVVVLAAA-GRDAPGPE-P--------------LGGRTRPHLALALPLIRLAA
DnrQSpe      GGE-TLEADQHVVVLVDAAQRDPALYEDPDRFRLDRPRSPGFTHMALAGRDHLGLVAPLVRVQC
TylMIISfr    AGR-TLPAGTHLVVLAAAANRDACRNAGPAVTGFDVLRRASD----GGPQPH-GLPEDLHFRLS
OleP1San     GDEVVVLAGAIGRNGPSAAPPAPPGPAAPPAPSV-FGAAAFENALAEPLVRAVTGAALQALAEG
DesVIISve    DPERFDLARPD---AAAHLALHPAGPYGPVASLVRLQAEVALRTLAGRFPGLRQAGDVLRPRRA
SnogNSno     ------------VVVAAA-HRDPSWCPDPDRFDLTRNER----HLALPPDLPLGALAPLLRVCA
Pra10Ahi     -----IPL-----FNAANRDPEVFADPEELDLGRTDNPHIALGHGIHYCLGAPLARLELQVVLE

MidHSmy      TTAVQVMAGRLPGLRVEDEPLTRPRSPVVCACARFRVHPG
DnrQSpe      TAVLRALAERLPGLRAEGEPLRRGRSPVVRAPLSLRLAQK
TylMIISfr    GPLVRRTAEAGLRALAEGPPRLTAAGPVVRRRRGLGRLPV
OleP1San     PPRLTAAGPVVRRRRSP-VVGGLHRAPVA-A-A-------
DesVIISve    PVGRGPLSVPVSSSRQAGDVLRPRRAPVGRGPLSVPVSSS
SnogNSno     TAAVAALAAGLLPLRAVGPPVRRLRAPVTRSVLRFPVAPC
Pra10Ahi     TLVERT--PALRLAIDDADITWR-PGLAFARPDALPIAW-
```

7.10. Sequence alignment of the MidC protein with other aminotransferases for construction of MidC-tree in Fig. 5.2

```
LmbSSli     FDTAEEEAVLRVVR---SG-WGVSTGAEAQSFEEEFAAYIGRRHAGA---LT--SCT--A
StsASgr     WAES-REAFDDP------GE-----GEWVRRFEQAAADRFGAAHCL---GV--NSGTS-A
StsCSgr     WPQPGDRALKSLEDVLTSGRWTISCAYQGRDSYERQFASAFADYCGSAMCVPISTGT--A
StrSSgr     WPQLTDDDIEAAVAALRS-NRLV--GQGNSTVEEFEAALAAGQ--GVEHAVAVSTGTA-A
DnrJSpe     REERADIL-DAVETVFESGQLIL--GTSVRSFEEEFAAYHGLPYCT-G--VD-N-GTN-A
TylBSfr     GVESAIGGGAAAVA--ARGRYLL--GAELAAFEERFAEYCGNAHC-----VAVGSGLDDA
MidCSmy     RELRADIDG-ALRRVSASGRYLL--GAELAGFEAEFAAYC-DNDC-----VAVGSGCD-A
OleN2San    LELKHDIDA-ATGRVLDSGRYLL--GPELAAFETEWAAYCGARHC-----VAVGSGCD-A
DesVSve     EELRAETDA-AIARVLDSGRYLL--GPELEGFEAEFAAYCETDHA-----VGVNSGMD-A
EryCISer    LELRSDIF-QACRRVLGSGWYL--HGPEEEAFEAEFAAYCENAHC-----VTVGSGCD-A
EryCIVSer   TVGDRERFFARLEWALNNNW-LTNGGPLVREFEGRVADLAGVRHC-----VATCNATV-A
OleN1San    NTGDRKRLLDRLEWALDNRW-LTNGGPLVREFEQRIADLAGVRNC-----VATCNATAG-
RfbEVch     LDGNERKYLNDC---IDSGW-VSSRGKYIDRFETEFAEFLKVKHA-----TTVSNGT-VA

LmbSSli     ARHGALRANGIGPGDEVSVPQVPFVPRPPAWSTPGGA-PCSRTS-VP-----STSPFDPD
StsASgr     LV-AALVGLGIGPGDEVIVPGYMFVASIAAV-LHCGADVVLAEV-D------DSLTLDPA
StsCSgr     SLAIALEACGVGAGDEVIVPGLSWVASASAV-LGINAVPVLVDV-DP-----ATYCLDPA
StrSSgr     -VHLALHALDVGPGDEVIVPTHTFIGSASPVTY-LGARPVFADV-TP-----DTHCLDPD
DnrJSpe     LV-LGLRALGIGPGDEVVTVSNTAAPTVVAIDA-VGATPVFVDVHE--E---NY-LMDTG
TylBSfr     -V-WALWALGVGEGDEVIVPSHTFIASWLAVSA-TGATPVPVEPGDPGEPGPGAFLLDPD
MidCSmy     L-ELV-RALGIGPGDEVVVPAHTFIGTWLAVSA-AGARPVGVDP-TP-----DGLSMDPA
OleN2San    L-ELALRAMDIGPGDEVIVPAHTFAATWLAVSA-TGAEPVAVEP-EP-----ATFTLDPE
DesVSve     L-QLALRGLGIGPGDEVIVPSHTYIASWLAVSA-TGATPVPVEPHE------DHPTLDPL
EryCISer    L-ELSLVALGVGQGDEVIVPSHTFIATWLGVP--VGAVPVPVEP----E-G-VSHTLDPA
EryCIVSer   L-QLVLRASDVS-G-EVVMPSMTFAATAHAASW-LGLEPVFCDV-DP-ETG----LLDPE
OleN1San    L-QLLLREAEVT-G-EVIMPSMTFVATAHAVRW-LGLRPVFCDI-DPD-TG----CLDPK
VchPerS     L-HLAMSALGITQGDEVIVPTFTYVASVNTIVQ-CGALPVFAEI-E----G-ESLQVSVE

LmbSSli     QVKSLITERTKAVRARCTFRPHGRHWNPLRFLCDSHGLTLPRTPRTPLPARDGDAVAGRA
StsASgr     DVRARITPRTRAVMPVHMLG-APADMTALRAVADEHGLHLLEDCAQSAGGSYRGRPLGTL
StsCSgr     ATEAAITERTRAITVVHAYS-AVADLDALLDIARRHGLPLIEDCAHAHGAGFRGRPVGAH
StrSSgr     SVKSLIGERTKAIVVVHING-IAADMAALTAVAAEAGVPVIEDAAQALGTEIGGRPIGGF
DnrJSpe     RLRSVIGPRTRCLLPVHLYG-QSVDMTPVLELAAEHDLKVLEDCAQAHGARRHGRLVGTQ
TylBSfr     RLEAALTPRTRAVMPVHLYG-HPVDLDPVGAFAEPHGLAVVEDAAQATARY-RGRRIGSG
MidCSmy     QVEAAITPRTRAVMPVHLYG-HPADLDPLLAIAERHGLAVVEDAGSAR-RPLPGRRIGSG
OleN2San    RVEAAITSRTRVILPVHLYG-HPADLAALSEVAERHGVRILEDAAQAHGAQAYGRRVGAW
DesVSve     LVEKAITPRTRALLPVHLYG-HPADMDALRELADRHGLHIVEDAAQAHGARYRGRRIGAG
EryCISer    LVEQAITPRTAAILPVHLYG-HPADLDALRAIADRHGLALVEDVAQAVGARHRGHRVGAG
EryCIVSer   HVASLVTPRTGAIIGVHLWG-RPAPVEALEKIAAEHQVKLFFDAAHALGCTAGGRPVGAF
OleN1San    LVEAAVTPRTGAILGVHLWG-RPSRVDELAAIAAEHGLKLFYDAAHALGCTSRQRRLGSF
VchPerS     DVKRKINKKTKAVMAVHIYG-QACDIQSLRDLCDEHGLYLIEDCAEAIGTAVNGKKVGTF

LmbSSli     GDASAFSFFATKPITTA-EGGMLCTDTARVADEARRWSLHGLSRGAVV-N-------RY-
StsASgr     GSAGTFSLNHYKMITSL-QGGFVLMDDPLVFQRAYSFHDQGWFPYRQD-RG---------
StsCSgr     GAAGVFSMQGSKLLTC-GEGGALVTDDADVALRAEHLRADGRVVRR-E---PV-GVGEME
StrSSgr     GDLACVSLFEQKVITSGGEGGAVLTDNPVYAERVRRLRSHGEG--------PVSGSPGMI
DnrJSpe     GHAAAFSFYPTKVLGAYGDGGAVVTPDAEVDRRLRRLRYYGMGER-------------YY
TylBSfr     -HRTAFSFYPGKNLGALGDGGAVVTSDPELADRLRLLRNYGAREK-------------YR
MidCSmy     -HVVAFSFYPGKNLGAMGDGGAVVTGDAALADRIRLLRKCGSREK-------------YR
OleN2San    ST-TAFSFYPGKNLGGFGDGGAVVTDDAELAERVRLLRNYGSREK-------------YR
DesVSve     SSVAAFSFYPGKNLGCFGDGGAVVTGDPELAERLRMLRNYGSRQK-------------YS
EryCISer    SNAAAFSFYPGKNLGALGDGGAVVTTDPALAERIRLLRNYGSKQK-------------YV
EryCIVSer   GNAEVFSFHATKAVTSF-EGGAIVTDDGLLADRIRAMHNFGIAPD------------KLV
OleN1San    GDAEVFSFHATKVVNSF-EGGGIVTDDDTRAERLRALHNFGLGHD---------GV----
VchPerS     GDVSTFSFFGNKTITS-GEGGMVVSNSDIIIDKCLRLKNQGVVAGK---R--------YW
```

```
LmbSSli     ---R---PGHKYNMSDLGRRPGPAQLAKAGRLHARRTAIAEVYLRE-LAGLD-RLELPRG
StsASgr     -EGD-PLLGMNLGLGELNAAVALAQLGKLDLILDRIRGVKHR-LVAAIGEL-PGVRRRTL
StsCSgr     LEETGRMMGSNACLSEFHAAVLLDQLELLDGQNARRTR-AADHLTDRLSEL-GMTAQATA
StrSSgr     WAHEV---GYNYRLTSVQAAVGLAQHKRLGDL-EARRRNAA-YLSERLAGVE-GLELPVE
DnrJSpe     VVDTP---GHNSRLDEVQAEILRRKLRRLDAYVEGRRAVARRY-EEGLGDLD-GLVLPTI
TylBSfr     HEER----GTNSRLDELQAAARVRKLPRNLDAWNTRRREIAARYGEALAGL-PGVTVPEA
MidCSmy     HEVQ----ATNSRLDEFQAAVLRAKLPRVPAWNALRVRTAERYS-QVLGAL-PQIAVPAA
OleN2San    HEVR----ATNFRLDELQAAVLRVKLAHLDAWTERRAAVAARYL-DGLAGLD-GIVLPRP
DesVSve     HETK----GTNSRLDEMQAAVLRIRLAHLDSWNGRRSALAAEYL-SGLAGLP-GIGLPVT
EryCISer    HEVR----GTNARLDELQAAVLRVKLRHLDDWNARRTTLAQHY-QTELKDV-PGITLPET
EryCIVSer   TD-----VGTNGKMSECAAAMGLTSLDAFAETRVHNRLNHALY-SDELRDV-RGISVHAF
OleN1San    --G----AGINAKMSEAAAAMGLTSLEAFADAVASNRANYELY-RQELSGL-PGVRLIDY
VchPerS     HD----LVAYNYRMTNLCAAIGVAQLERVDKIIKAKRDIAEIY-RSELAGL-PMQVHKES

LmbSSli     HRHQPSSWYLFPV-RVHGHRRDAFRQR-L-HALGVGTS-VHFEPLHRFTWLRDHVVRTGQ
StsASgr     HDAEGECGTVAVYV-FEDAAHALDVARRLGTRVLL-DSPTHYYGGLPALAAFGRGDRSTV
StsCSgr     PGTTARAYYRYLV-RLPDEVLAVAPVERFAHALTAELGFAVTQTHRPLNDNPLNRPSSRR
StrSSgr     PPGTTHAYWKYAV-RVVPGDGRRSAADIAAHLRSRGVP-VLLRYPYPLHKQPAFA-----
DnrJSpe     AEGNDHVYYVYVV-RHPE--RDRILEALTAY-D-IH-LNISYPW--PVHTMSGFAHL-GY
TylBSfr     ARGGAGVWHQYVL-RQPV--RDR-LRRRL--AEAGVETLVHYPV--AVHASGYAGA--G-
MidCSmy     APWADPVWHLYVI-LRAN--RDE-LRRRI--ERAGVETLIHYPV--PPHRTPAYADDSGR
OleN2San    APWADPVWHLFVI-RSAD--RSA-LRERLA--AAGVETLIHYPV--PVHRSEAYAGSRQA
DesVSve     APDTDPVWHLFTV-RTER--RDELRSHL---DARGIDTLTHYPV--PVHLSPAYAGE--A
EryCISer    HPWADSAWHLFVL-RCEN--RDH-LQRHL--TDAGVQTLIHYPT--PVHLSPAYADL-GL
EryCIVSer   DPGEQNNYQ-YVIISVDSAATGIDRDQLQAILRAEKVVAQPYFSPG-CHQMQPYRTE---
OleN1San    DPAERNNYH-YVIALIDAGVTGLHRDLLLTLLRAENVVAQPYFSPG-CHQREPYRTE---
VchPerS     NGTFHSYWLTSIILD-QE-F--EVHRDGLMTFLENNDIES-RPFFYPAH-T---------

LmbSSli     --G-FPVA------DAAA-----DTLV---SLPVFPAMHDDAAVSRVVAAVR---
StsASgr     PFRAPGG---RPSASFEPGTLPRTDDVLGRSIALATGVSDDYLGPGFGVHADSSA
StsCSgr     RFATDARYLERVDPSRFDLPAAKRAHE---SVVSFSHEVLLAPLDAIDDIARAFR
StrSSgr     V--------------SLPVAERLSQELLALLPSHPGLVEG---------EYHG--
DnrJSpe     ----------------------------------------GPGDLPVT-ERLAG-
TylBSfr     VL-------------SLPIGPHLPDEAVE-VVIAAVQSAA-P-------ERLAGE
MidCSmy     SL-------------SLPSGPHLGDDAFQTVVAAVRAAAVRSGRTHPLSERRAAE
OleN2San    VL-------------SLPIGPHLSDDAVKAVIEAVRGAVAARA--QPVAERLARE
DesVSve     VL-------------SLPIGPHLERPQALRVIDAVREWAEPPEGSLPRAESFARQ
EryCISer    VL-------------SLPIGPHLSREAADH-VIATLKAGAPPGSFPVA-ESLAGE
EryCIVSer   VL-------------ALPTGPAVSSEDIRRVCDIIRLAATPPLRLENT-EQLSDR
OleN1San    VI-------------ALPTGPAVSREDIRRVCDIIRVAAAHPVSLPHT-EHLAEQ
VchPerS     TAFPLSNSYSHRGI-NLPSWPGLCDDQVKE-ICNCIKNYF----LPMY-EHLAEK
```

7.11. Sequence alignment of the MidK protein with other methyltransferases for construction of MidK-tree in Fig. 5.2

```
MidKSmy     M------------YANDIAALYDLVHEGKGKD----YRQEAEEI-A-QLVRAHRPATRSLLDVA
TylMISfr    MAHSSATAGPQADYSGEIAELYDLVHQGKGKD----YHREAADL-A-ALVRRHSPKAASLLDVA
OleM1San    MRADTEP---TTGYEDEFAEIYDAVYRGRGKD----YAGEAKDV-A-DLVRDRVPDASSLLDVA
EryCVISer   M------------YEGGFAELYDRFYRGRGKD----YAAEAAQV-AR-LVRDRLPSASSLLDVA
LmbJSli     MS-DHDFA---KLYGDKIADVYDVWPGDAGPP-P-----DADRA-APFLAALANGRPALELGV-
StsGSgr     MNPRK-P---HVDWDHLFGDDYDYF-DLPDLTP-ELSEKEASNM-V-ELGGFDAGMD--LLDAP
GlyMRa      MDSVYRTRSLGVAAEGIPDQYADGEAARVWQLYIGDTRSRTAEYKAWLLGLLRQHGCHRVLDVA

MidKSmy     CGTGQHLRHLDGLFDHVEGLELSQDMLAIAIGRNPDVTL--H-EGDMRSFAL------G-RRFD
TylMISfr    CGTGMHLRHLADSFGTVEGLELSADMLAIPRRRNPDAVL--H-HGDMRDFSL------G-RRFS
OleM1San    CGTGAHLRHFATLFDDARGLELSASMLDIARSRMPGVPL--H-QGDMRSFDL------GPR-VS
EryCVISer   CGTGTHLRRFADLFDDVTGLELSAAMIEVARPQLGGIPV---LQGDMRDFAL---D----REFD
LmbJSli     -GTGRVAVPLAESGVEVHGVDSSARMLEILKEKSGGA-VHGH-QQDFGRLDL------GERRFG
StsGSgr     CGHGRHANVLASRGYRVVGVDRDERFLSMARKEAESMG----VQVDYRHVDLREMSF--SAEFD
GlyMRa      CGTGVDSIMLVEEGFSVTSVDASDKMLKYALKERWNRRKEPAFDKWVIEEANWLTDVPAGDGFD

MidKSmy     AVICMFSSIGHLRTTDELDSTLRCFAGH----LEPGGAIVIEPWWFPDSFTPGY--V-GA—-TS
TylMISfr    AVTCMFSSIGHLAGQAELDAALERFAAH----VLPDGVVVVEPWWFPENFTPGY--VAAGTV-E
OleM1San    AVTCMFSSVGHLATTAELDATLRCFARHT----RPGGVAVIEPWWFPETFTDGY--VAGDIV-R
EryCVISer   AVTCMFSSIGHMRDGAELDQALASFARH----LAPGGVVVVEPWWFPEDFLDGY--VAGDVV-R
LmbJSli     LVFALFNTLFCLLTQDEQIACLRSAA-NC---LETEGLLVLQCLN-PKSLPDGS-DVAL-VELE
StsGSgr     AAVSWYSSFGYFDDETDRD-ILRR----YRRALRPGGRFLLDMHS-PYRHIPS---VLANHEMH
GlyMRa      AVICLGNSFAHLPDSKGDQSEHRLALKNIASMVRPGGLLVIDHRNYDYILSTGCAPPGKNIYYK

MidKSmy     PRRANGTICRVSDSVR-EGKTTRIEVHYLVAEPGGGI--RHLTEDHTITLF--PRADYERAFE
TylMISfr    AGGTTVT--RVSHSSR-EGEATRIEVHYLVAGPDRGI-THH-EESHRITLF--TREQYERAFT
OleM1San    VDG-R-TISRVSHSVR-DGGATRMEIHYVIADAEHGP--RHLVEHHRITLF--RWHAYTAAYE
EryCVISer   -DG-DLTISRVSHSVR-AGGATRMEIHWVVADAVNGP--RHHVEHYEITLF--ERQQYEKAFT
LmbJSli     HDGVHL--D-VSKHDPVAQTLTAHHIVLSESGARFFPYTLRYSHHTELDLMAARFAGFELRSR
StsGSgr     VDILRRGQD-MA-VDIQELDAEASRYYAEKLTIRDDKVVRARYSVRMFTAPEILEWFRSAGFS
GlyMRa      SDLTKDITTSVLTVNNKAHMVTLDYTVQVPGAGRDGAPGFSKFRLSYYPHCLASFTELVQEAF

MidKSmy     RA-GCDVRYQEGGSSGRG-LF------N-RQ-PA
TylMISfr    AA-GLSVEFMPGGPSGRG-LFTGLPGA--KGETR
OleM1San    KA-GYTVEYLDGGPSGRG-LFV-----GTR--T-
EryCVISer   -AAGCAVQYLEGGPSGRG-LFV-----GVRG---
LmbJSli     HADFDGAAYRPDSRYHVS-VYA---RAH--GD--
StsGSgr     HARVMDETGGTFTVSSRR-LMV-------LG-TA
GlyMRa      GGRCQHSVLGDFKPYRPGQAYVPCYFIHVLKKTG
```

7.12. Sequence alignment of the MidI protein with other glycosyltransferases for construction of MidI-tree in Fig. 5.2

```
OleG1San     MRVLLTCFANDTHFHGLVPLAWALRAAGHEVRVASQPALSDTITQAGLTACPWAGT-RFLELMGE
EryBVSer     MRVLLTSFAHRTHFQGLVPLAWALRTAGHDVRVAAQPALTDAVIGAGLTAVPVGSDHRLFDIVPE
OleG2San     MRVMMTTFAANTHFQPLVPLAWALRTAGHEVRVVSQPSLSDVVTQAGLTSVPVGTEAPVEQFAAT
EryCIIISer   MRVVFSSMASKSHLFGLVPLAWAFRAAGHEVRVVASPALTEDITAAGLTAVPVGTDVDLVDFMTH
TylMIISfr    MRVLLTCIAHNTHYYNLVPVAWALRAAGHEVRVAAQPALTDTITASGLTAVPVGGNESVLEFVTE
MidISmy      MRVLLTSLAHNTHYYSLVPLAWALRAAGHEVRVASPPSLTDVITSTGLPAVPVGDDQPAAELLAE
TylNSfr      M-----------------PEVVETLRRGPAMRDLMKGLPPAPEEYDQEVLDRIERAGEGVDLVVH

OleG1San     IGADVQKYSTGIDL-GVR-AELTSWEYLLGMHTTLVPTFYSLVNDEPFVDGLVALTRAWRPDLIL
EryBVSer     VAAQVHRYSFYLDFYHRE-QELHSWEFLLGMQEATSRWVYPVVNNDSFVAELVDFARDWRPDLVL
OleG2San     WGDDAYIGVNSIDFTGNDP-GLWTWPYLLGMETMLVPAFYELLNNESFVDGVVEFARDWRPDLVI
EryCIIISer   AGHDIIDYVRSLDFSERDPA-TLTWEHLLGMQTVLTPTFYALMSPDTLIEGMVSFCRKWRPDLVI
TylMIISfr    IGGDPGPYQRGMDFAETCG-EPLSYEHALGQQTAMSALCFAPFNCDSTIDDMVALARSWRPDLVL
MidISmy      MGGDLVPYQRGFEFAEVEPAQETTWEHLLGQQSMMSALWFAPFSGAATMDDIGRLRRDWRPDLVV
TylNSfr      ---APLTVTTALGEPSTPWLSVNWWP----------------N--T---------STWT-----

OleG1San     WEHFSFAGALAARATGTPHARVLWGSDLIVRFRRDFLAAAANPARRAPRDPMAAWLGW-RTGWVS
EryBVSer     WEPFTFAGAVAARACGAAHARLLWGSDLTGYFRGRFQAQRLRRPPEDRPDPLGTWLTE-VAGRFG
OleG2San     WEPLTFAGAVAARVTGAAHARLPWGQEITLRGRQAFLAERALQPFEHREDPTAEWLP--HARPVR
EryCIIISer   WEPLTFAAPIAAAVTGTPHARLLWGPDITTRARQNFLGLLPDQPEEHREDPLAEWLTWTLEKYGG
TylMIISfr    WEPFTYAGPIAAHACGAAHARLLWGPDVILNARAQFRRLAPDSPEEPREDPVAEWLGWTLERHGL
MidISmy      WEPWTYAGPIAARACGAAHARILWGPDAIGRSRRRFLEALERVPEELREDPIAEWLGWTLDRYGC
TylNSfr      F-P---AVE-----SG--QRRM--GP-LT-PLYNRL-THW-RA--E-R-D---HW-GWRRAEVNE

OleG1San     TFDEEL-----VTGPWTIDPLPRSMRLPTGTTT-V-PMRYVRTT-AA-VVPAWVRHVRGGPRICL
EryBVSer     VE----FGEDLAVGQWSVDQLPPSFRLDTGMET-VVA-RTLPYNGAS-VVPDWLKKGSATRRICI
OleG2San     CSFDE----EMVTGQWTIDTLPRTMRLELSEELRTLDMRYVPYNGPA-VVPPWVWEPCERPRVCL
EryCIIISer   PA----FDEEVVVGQWTIDPAPAAIRLDTGLKT-V-GMRYVDYNGPS-VVPEWLHDEPERRRVCL
TylMIISfr    TAERE-TVEELIGGQWTLDPTAESLRCPR-P-A-VVPFRFVPYNGRS-VLPDWLLRKPGRPRVCF
MidISmy      AFDERD-----VLGHWVIDPGPRSTRLDLGQ-T-TVPMCYVPYNGRA-VIEPWLAEKPERPRVCL
TylNSfr      FRGRPRL---PPFGKSSPLRRLGHPRHHLYPFSPSVLPKPRDWPGQCHVTGYWFWDQPGW-RPSP

OleG1San     TLGVSARQ-TLGDG-VSLA-EVLAALGDVDAEIVATLDASQRKLLGPVPD-----NVRLVDSPCP
EryBVSer     TGGFSGLGLAADADQFARTLAQLARF---DGEIVVTG-SGPDTSAV-------PDNIRLVD-FVP
OleG2San     TIGTSQRD-S-GRDLVP--STTCSTP-DVAAEIVATSTPPSRRSAA-APG-----NVRAGD-FVP
EryCIIISer   TLGISSRENSIGQ--VSI-EELLGAVGDVDAEIIATFDAQQLEGVA-----NIPDNVRTVG-FVP
TylMIISfr    TLGVSARE-TYGRDAVPF-HELLAGLGDLDAEIVATLDPGQLSGAG-----EVPRNVRAVD-FVP
MidISmy      TLGISARE-TYGRDAVSY-SELFQALGRMEIEVVATLDASQQKRLGSLPD-----NVVPVD-FVP
TylNSfr      ELEDFLADGEPPVLLTLGSTWPLHRQEMVEYPVTTARGARRRLLLVGGPE-----NVVRVPS-AD

OleG1San     CTPVLPTCSAIVHHGGAGTWLTAAVHGVPQIVLGDLWDNPVRARQTQARARGLFIHPSEVRRPG
EryBVSer     MGVLLQNCAAIIHHGGAGTWATALHHGIPQISVAHEWDCMLRGQQTAELGAGIYLRPDEVDA-D
OleG2San     LHALMPTCSAIVHHGGPGTWSTA-LHGVPQIILDTSWDTPVRAQRMQQLGAGLSMPVGELGV-E
EryCIIISer   MHALLPTCAATVHHGGPGSWHTAAIHGVPQVILPDGWDTGVRAQRTQEFGAGIALPVPELTP-D
TylMIISfr    MDALLPTCSAVVHHGGAGTCFTATLNGLPQIVVAALWDAPLKGAQLAEAGAGVSIA-PEKLD-A
MidISmy      LDALLPSCAAIIHHGGAGTWSTALLHGVPQILLPALWDAPLKAQQLQRLSAGLNLPAATLTA-R
TylNSfr      YSWLMPRTAAVVHHGGFGTTADAVRAGVPQVLVPVLRRPPLLGRPAAADGHGDHPVPLARMNRE

OleG1San     SVRACAGVLTGAPSI----RAAAQRVRDEMNAEPTPGEVVTVLERLAAS----
EryBVSer     SLASALTQVVEDPTY----TENAVKLREEALSDPTPQEIVPRLEELTRRH-AG
OleG2San     ALRDRVLRLLGEPEF----RAGAERI--------------------------
EryCIIISer   QLRESVKRVLDDPAH----RAGAARMRDDMLAEPSPAEVVGICEELAAGRREP
TylMIISfr    ATLRAGVVRALEDE-GHSRRSAGLL-RAEMLAEPTPAGLVPQLERLTALHRNG
MidISmy      RLADAVHTAVHDPAI----RAGARRLREEMLADPTPAAIVPTLERLTALHRAA
TylNSfr      ALAASVRTAVTDPAMAVRARWLGEAVAAERGVENACVLIEEWAETRTTAHTPG
```