BERGISCHE UNIVERSITÄT WUPPERTAL

Genetic studies on the biosynthesis of the major aminoglycoside antibiotics:
Isolation, analysis and comparison of the biosynthetic gene clusters for 12 aminoglycoside antibiotics

A dissertation submitted to the Department of Natural Science
(*Chemical Microbiology*)
Bergische Wuppertal University

For the degree of

Doctor of Natural Sciences (Doctor rerum naturalium)

Presented by

MSc. Khaled Mohamed Anwar Aboshanab
Assistant Lecturer in the Microbiology and Immunology Department
Faculty of Pharmacy-Ain Shams University
Cairo-Egypt

Wuppertal, July 2005

Acknowledgment

Firstly, I thank "***Allah***" for granting me the power to accomplish this work.

I am beholden to a number of people and organisations, who supported me to carry out this work. First I would like to express my deepest thanks, gratitude and appreciation to my advisor ***Prof. Dr. Wolfgang Piepersberg*** for giving me the opportunity to pursue my doctoral degree under his supervision as well as suggesting the point, his encouragement, help and kind support. His invaluable advice, suggestions, discussions and guidance were a real support to complete this dissertation.

I am also indebted and grateful to ***Prof. Dr. Hans-Josef Altenbach*** for his acceptance and kindness to co-examine this thesis.

It is my great pleasure to express my deep everlasting gratitude and thanks to my co-advisor ***Dr. Udo Wehmeier*** for his valuable discussion through out the work, critical reading and discussion and also the useful suggestions to my dissertation.

I am also very grateful and thankful to ***Dr. Heike-Schmidt Beißner*** for her unfailing help, valuable discussion through out the work as well as her help in final revision of the thesis.

Furthermore, I would like to express my deepest gratitude and thanks to all staff members of "Arbeitskreis Chemische Mikrobiologie" for contributing to such an inspiring and pleasant atmosphere.

I am also indebted and thankful to the staff members of Combinature Biopharm Company in Berlin for construction and screening the cosmid banks.

As well, my thanks to *Dr. M. Podeschwa* from the organic chemistry department for his kind supply of the chemically synthesized 2-deoxy-*scyllo*-inosose.

Very special thanks to my family, especially my parents and my wife, for their patience, understanding and encouragement during the different phases of my work. They spared no effort until this work comes to existence.

Last but not least, I wish to acknowledge the financial support of the Missions Department-Egypt and the Faculty of Pharmacy, Ain Shams University for giving me the opportunity to pursue my doctoral degree in Germany.

Khaled M. A. Aboshanab

Wuppertal, 2005

Contents

		Page
Contents		i
Abbreviations		vi
List of tables		ix
List of figures		x
Summary		xii
Zusammenfassung		xiii
1 Introduction		1
1.1	Aminoglycoside antibiotics (AGAs)	1
1.1.1	Definition	1
1.1.2	Classes of aminoglycoside antibiotics	1
1.1.3	Mode of action	8
1.1.4	AGAs in medical treatments	8
1.1.5	Biosynthesis and regulation of ACAGAs	9
1.1.5.1	Biosynthesis of 2DOS-containing ACAGAs	10
1.1.5.2	Biosynthesis of fortamine and 2-deoxyfortamine-containing ACAGAs	12
1.1.6	Resistance mechanisms	14
1.1.7	Toxicity	15
1.2	ACAGAs producers	15
1.2.1	*Actinomycetales*	16
1.2.1	*Bacillus*	18
1.3	Aim of the work	20
2 Materials and Methods		21
2.1	Chemicals and enzymes	21
2.2	Media	22
2.2.1	Media for culturing *Escherichia coli* and *Bacillus* strains	22
2.2.2	Media for culturing *Streptomyces* and *Micromonospora* strains	23
2.2.3	Aminoglycoside production medium	24

2.3	Buffers and solutions	24
2.3.1	For the preparation of plasmid DNA from *E. coli*	24
2.3.2	For the preparation of competent *E. coli* cells	25
2.3.3	For the preparation of genomic DNA	25
2.3.4	For the preparation of protoplasts from *Streptomyces*	25
2.3.5	For the transformation of *Streptomyces*	26
2.3.6	For agarose gel electrophoresis of DNA fragments	26
2.3.7	For harvesting and sonification of cells	26
2.3.8	For Western blotting and detection of proteins	26
2.3.9	For SDS polyacrylamide gel electrophoresis	27
2.3.10	For visualization of spots on TLC	28
2.3.11.	X-Gal solution	29
2.3.12	IPTG stock solution	29
2.4	Bacterial strains, vectors and recombinant plasmids	29
2.4.1	Bacterial strains	29
2.4.2	Vectors	30
2.4.3	Recombinant plasmids	31
2.5	Oligonucleotides	34
2.6	Probes and cosmid banks	37
2.7	Antibiotics	39
2.8	Growth and maintenance of bacterial strains	39
2.8.1	Growth and maintenance of *E. coli* and *Bacillus* strains	39
2.8.2	Growth and maintenance of *Streptomyces* sp. and *Micromonospora* sp. strains	39
2.9	In vitro manipulation of DNA	40
2.9.1	Preparation of plasmid DNA from *E. coli*	40
2.9.2	Preparation of cosmid DNA from *E. coli*	40
2.9.3	Extraction of genomic DNA from actinomycetes and bacilli	40
2.9.4	Restriction endonuclease digestion of DNA	41
2.9.5	Blunt end generation using the Klenow fragment of DNA polymerase I	41
2.9.6	Recovery of DNA fragments from agarose gels	42
2.9.7	Removal of phosphate groups from DNA	42
2.9.8	Ligation of DNA fragments	42
2.9.9	Quantitation of DNA	42

2.9.10	Construction of cosmid banks	42
2.9.11	Preparation and screening of the cosmid banks	43
2.9.12	DNA sequencing	43
2.10	Preparation of competent *E. coli* cells	43
2.11	Transformation of *E. coli* strains	44
2.12	Preparation of protoplasts from *Streptomyces* strains	44
2.13	Transformation of streptomycete protoplasts with plasmid DNA	44
2.14	Agarose gel electrophoresis	45
2.15	Polymerase chain reaction (PCR)	45
2.16	SDS polyacrylamide gel electrophoresis (SDS-PAGE)	45
2.17	Heterologous gene expression	46
2.17.1	Protein production in *E. coli* using the *T7* RNA polymerase system	46
2.17.2	Gene expression in *S. lividans* TK23 under the control of *ermE*-up promoter	47
2.18	Cell-free extracts of *E. coli* and *Streptomyces*	47
2.19	Determination of protein concentration	47
2.20	Western blotting and immuno-detection of proteins	47
2.21	Thin-layer chromatography (TLC)	48
2.22	High performance liquid chromatography (HPLC)	48
2.23	Determination of the activity of KanC	48
2.24	Determination of the activity of KanS1	49
2.25	Determination of the activity of KanE	50
2.26	Computer programs	51

3 Results 52

A	**Gene clusters for the production of 2-deoxystreptamine (2DOS) and related ACAGAs**	52
3.1	Isolation and sequence analysis of the biosynthetic gene clusters	52
3.2	Gene clusters for the individual classes of 2DOS- and related ACAGAs	54
3.2.1	4,5-glycosylated 2DOS-ACAGAs (NM group)	54
3.2.2	4,6-glycosylated 2DOS-ACAGAs (KM group)	58
3.2.3	4-glycosylated 2DOS-ACAGAs (Apr group)	61
3.2.4	5-glycosylated 2DOS-ACAGAs (HM-B group)	63
3.2.5	ACAGAs related to 2DOS-ACAGAs (FTM and IM group)	64

3.3	Newly developed heterologous primers	67
B	**Studies on some selected ACAGAs biosynthetic enzymes**	75
3.4	Identification of the enzymes involved in the biosynthesis of 2DOS	75
3.5	Heterologous expression of the KanC, KanS1 and KanE proteins	76
3.5.1	The general strategy for the overexpression of the selected Kan-proteins	76
3.5.2	Heterologous expression in *E. coli* under *T7* promoter	77
3.5.3	Heterologous expression of KanE *S. lividans* TK23	79
3.6	Biochemical characterization of the KanC, KanS1 and KanE proteins	80
3.6.1	Biochemical analysis of the KanC protein	80
3.6.2	Biochemical analysis of the KanS1 protein	83
3.6.3	Biochemical analysis of the KanE protein	87

4 Discussion 89

4.1	General structure and genomic location of the ACAGA biosynthetic gene clusters	89
4.2	Proposed biosynthetic pathways for the major groups of 2DOS and related ACAGAs	90
4.2.1	Biosynthesis of 2DOS and related diaminocyclitols	90
4.2.2	Biosynthesis of paromamine	91
4.2.3	4,5-glycosylated 2DOS-ACAGAs (NM group)	94
4.2.3.1	Biosynthesis of the NM-related ACAGAs	94
4.2.3.2	Resistance mechanisms for the NM-related ACAGAs	98
4.2.3.3	Possible sensor/response regulatory system	99
4.2.4	4,6-glycosylated 2DOS-ACAGAs (KM group)	100
4.2.5	4-glycosylated 2DOS-ACAGAs (Apr group)	104
4.2.6	5-glycosylated 2DOS-ACAGAs (HM-B group)	106
4.2.7	Related 2DOS-ACAGAs (FTM and IM group)	108
4.3	Reconstitution of the 2DOS pathway in vitro	112
4.3.1	Heterologous expression of KanC, KanS1 and KanE proteins	112
4.3.2	Biochemical characterization of KanC, KanS1 and KanE proteins	112
4.4	Prospective of this work	114

5 References 116

6 Appendix 131

6.1 Tables of proteins encoded in the genomic region covering the biosynthetic gene clusters of the ACAGAs analysed 131

A.1 *neo*-cluster of *S. fradiae* DSM 40063 (accession code: AJ629247) 131

A.2 *rib*-cluster of *S. ribosidificus* NRRL B-11466 (accession code: AJ744850) 133

A.3 *par*-cluster of *S. rimosus* NRRL 2455 (accession code: AJ749845) 135

A.4 *aacC7* of *S. rimosus* NRRL 2455 (accession code: AJ628955) 137

A.5 *liv*-cluster of *S. lividus* ATCC 31603 (accession code: AJ748832) 138

A.6 *btr*-cluster of *B. circulans* ATCC 21558 (accession code: AJ781030) 140

A.7 *kan*-cluster of *S. kanamyceticus* DSM 40500 (accession code: AJ628422) 141

A.8 *tob*-cluster of *S. tenebrarius* DSM 40477 (accession code: AJ810851) 143

A.9 *gen*-cluster of *M. echinospora* DSM 43036 (accession code: AJ628149) 145

A.10 *for*-cluster of *M. olivasterospora* DSM 43868 (accession code: AJ628421) 148

A.11 *apr*-cluster of *S. tenebrarius* DSM 40477 (accession code: AJ629123) 150

A.12 *apr*-cluster of *St. hindustanus* DSM 44523 (accession code AJ875019) 152

A.13 *hyg*-cluster of *S. hygroscopicus* DSM 40578 (accession code: AJ628642) 154

A.14 *ist*-cluster of *S. tenjimariensis* ATCC 31603 (accession code: AJ845083) 156

6.2 Figures for alignments and phylograms of various proteins encoded by conserved genes among the ACAGA clusters analysed 159

A.1 Alignment of BtrC homologous 160

A.2 Phylogram of BtrC homologous 160

A.3 Alignment of BtrS homologous 162

A.4 Phylogram of BtrS homologous 163

A.5 Alignment of BtrE homologous 164

A.6 Phylogram of BtrE homologous 165

A.7 Phylogram of BtrB homologous 165

A.8 Alignment of BtrB homologous 167

A.9 Alignment of LivY homologous 168

A.10 Alignment of LivW and AprD4 proteins 169

A.11 Alignment of NeoI homologous 169

A.12 Alignment of NeoH homologous 170

A.13 Alignment of NeoG homologous 171

A.14 Structural comparison of GM-C1, KM-B and FTM-A 172

Abbreviations

A	adenine
aa	amino acid
AAC	aminoglycoside 3-acetyltransferase
ACAGA(s)	aminocyclitol-aminoglycoside antibiotic(s)
AHB	α-hydroxy-γ-aminobutyryl residue
AHP	α-hydroxy-β-aminopropionyl residue
ALF	automatic laser fluorescence DNA sequencer
APH	aminoglycoside phosphotransferase
apr-cluster	apramycin gene cluster
Apr(s)	apramycin(s)
APS	ammonium persulfate
AT	aminotransferase
ATCC	American Type Culture Collection, Maryland
ATP	adenosine 5'-triphosphate
B.	*Bacillus*
bp	base pair
BSA	bovine serum albumin
btr-cluster	butirosin gene cluster
BU(s)	butirosin(s)
C	cytosine
CB	cosmid bank
cf.	confers to
(c)HP	(conserved) hypothetical protein
COG	cluster of orthologous groups
CTAB	cetyltrimethylammonium bromide
DH	dehydrogenase
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonuclic acid
dNTP	deoxyribonucleosidetriphosphate
2DOS	2-deoxystreptamine
2DOS-ACAGAs	2-deoxystreptamine containing aminocyclitol aminoglycoside antibiotics
DSM	Deutsche Stammsammlung vom Mikroorganismen und Zellkulturen
DTT	dithiotheritol
E. coli	*Escherichia coli*
EDTA	ethylene diaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
FAD	flavin adenine dinucleotide
Fig.	Figure
for-cluster	fortimicin gene cluster
FTM(s)	fortimicin(s)
G	guanidine
g	gram
gen-cluster	gentamicin gene cluster
GM(s)	gentamicin(s)
G-6-P	D-glucose 6-phosphate
GT	gylcosyltransferase
HM-B	hygromycin B
HPLC	high performance liquid chromatography
hr(s)	hour(s)
hyg-cluster	hygromycin B gene cluster

IM(s)	istamycin(s)
IPTG	isopropyl-β-D-thiogalactopyranoside
ist-cluster	istamycin gene cluster
kan-cluter	kanamycin gene cluster
kb	kilobase
kDa	kilodalton
KM(s)	kanamycin(s)
l	liter
liv-cluster	lividomycin gene cluster
LM	lividomycin
M.	*Micromonospora*
MBP2	maltose binding protein 2
min	minute
Mr	relative molecular weight
mRNA	messenger RNA
MT	methyltransferase
NAD(H)	nicotinamide adenine dinucleotide (reduced form)
NADP(H)	nicotinamide adenine dinucleotide phosphate (reduced form)
neo-cluster	neomycin gene cluster
NM(s)	neomycin(s)
NRPS	none ribosomal peptide synthetase
NRRL	Northern Regional Research Laboratory
Nt	nucleotide
OD_x	optical density at x nm and 1 cm depth
ORF(s)	open reading frame(s)
P	phosphate
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
par-cluster	paromomycin gene cluster
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfam	protein family
PKS	polyketide synthase
PLP	pyrodixal phosphate
PM(s)	paromomycin(s)
PMSF	phenylmethylsulfonyl fluoride
PP	inorganic pyrophosphate
PPP	pentose phosphate pathway
PVDF	polyvinyl difluoride
RBS	ribosome binding site
rib-cluster	ribostamycin gene cluster
RM	ribostamycin
RNA	ribonuclic acid
RNase	ribonuclease
rpm	revolution per minute
rRNA	ribosomal ribonuclic acid
RT	room temperature
S	Svedberg unit (centrifugation unit)
S.	*Streptomyces*
sec	second
Sect.	section
SDS	sodium dodecylsulfate
SMART	simple modular architecture research tool
Spc	spectinomycin

spc-cluster	spectinomycin gene cluster
St.	*Streptoalloteichus*
(sub)sp.	(sub)species
T	thymidine
Tab.	table
TB	tubercle bacilli
TEMED	N, N, N`, N`-tetramethylethylenediamine
TES	N-[tris(hydroxymethyl)methyl]-2-aminoethansulfonic acid
THB	tetrahydrobiopterin
TLC	thin layer chromatography
TM	tobramycin
tob-cluster	tobramycin gene cluster
Tris	trishydroxymethylaminomethane
tRNA	transfer ribonuclic acid
U	uridine
UDP	uridine diphosphate
(UDP)-GlcNAc	(uridine diphosphate)-N-acetyl-D-glucosamine
UV/VIS	ultraviolet/visible light
V	volt
X-Gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
λ	wavelength

List of tables

Tab.		Page
1.1	Widely used ACAGAs and their applications	17
2.1	List of strains used in this study	29
2.2	Vectors used in this study	30
2.3	Newly constructed plasmids created in this study	31
2.4	Homologous primers and PCR conditions	34
2.5	Newly developed heterologous primers	36
2.6	Probes used for screening of the different cosmid banks	37
2.7	Cosmids positively screened and used for sequence determinations	38
2.8	Antibiotics	39
3.1	Sequence data of fully sequenced ACAGA gene clusters which have been submitted to the EMBL data base and their accession codes	53
3.2	Detection of biosynthetic genes using newly developed heterologous primers	70
3.3	Biosynthetic proteins encoded by genes conserved among the ACAGA gene clusters analysed	71
3.4	Transporter proteins encoded by genes conserved among the ACAGA gene clusters analysed	72
3.5	Regulatory proteins encoded by genes conserved among the ACAGA gene clusters analysed	73
3.6	Resistance proteins encoded in the ACAGA gene clusters analysed	73
3.7	Transposase proteins encoded in the ACAGA gene clusters analysed	74
3.8	Summary of expression of KanC, KanS1 and KanE proteins	77

List of figures

Fig.		Page
1.1	Chemical structures of streptomycin and spectinomycin	2
1.2	Chemical structures of the some selected 4,5-glycosylated ACAGAs	3
1.3	Chemical structures of some selected 4,6-glycosylated ACAGAs	4
1.4	Chemical structures of apramycin and hygromycin B	5
1.5	Chemical structures of fortimicin A and istamycin A	6
1.6	Chemical structures of some selected semisynthetic ACAGAs	7
1.7	General pathway design for the biosynthesis of the major classes of 2DOS-ACAGAs	11
1.8	The proposed biosynthetic pathway of fortimicins (astromicin)	13
3.1	Restriction maps and arrangement of the selected cosmids containing gene clusters of the NM group	55
3.2	Organization of ORFs for the gene clusters of the 4,5-glycosylated 2DOS-ACAGAs (NM group)	56
3.3	Genetic map around the gene for the aminocyclitol acetytransferase VII (*aacC7*) from *S. rimosus* NRRL 2455	58
3.4	Restriction maps and arrangement of the selected cosmids containing gene clusters of the KM group	59
3.5	Organization of ORFs for the gene clusters of the 4,6-glycosylated 2DOS-ACAGAs (KM group)	60
3.6	Restriction maps and arrangement of the selected cosmids containing the *apr*-gene clusters	62
3.7	Organization of ORFs for the two *apr*-gene clusters derived from SteO08 and ShinN01 cosmids	62
3.8	Restriction map of the insert of cosmid ShyG17 and location of the *hyg*-cluster	64
3.9	Organization of ORFs for the *hyg*-cluster	64
3.10	Restriction maps and arrangement of the selected cosmids for the *for*- and *ist*-clusters	65
3.11	Organization of the ORFs for the *for*- and *ist*-clusters	66
3.12	Partial alignment of 2-deoxy-*scyllo*-inosose synthase proteins (“NeoC”-family) for the design of heterologous primers	67
3.13	Partial alignment of 2-deoxy-*scyllo*-inosose aminotransferase (“NeoS”-family) and other aminotransferase proteins for the design of heterologous primers	68
3.14	Partial alignment of the putative 6’-aminotransferase proteins (“NeoB”-family) for the design of heterologous primers	69
3.15	Partial alignment of the putative ribosyltransferase proteins (“NeoA”-family) for the design of heterologous primers	69
3.16	Partial alignment of the putative 6’-C-methyltransferase proteins (“GenK”-family) for the design of heterologous primers	70

3.17	Biosynthetic pathways for the formation of 2DOS in *S. kanamyceticus* DSM 40500	76
3.18	SDS-PAGE analysis of the production of the proteins KanC, KanS1 and KanE in different *E. coli* strains	78
3.19	Detection of His-tag proteins KanC, KanS1 and KanE by Western Blotting	78
3.20	SDS-PAGE analysis of the production of KanE in *S. lividans* TK23	80
3.21	TLC analysis of KanC catalysed formation of 2-deoxy-*scyllo*-inosose	81
3.22	HPLC analysis of 2-deoxy-*scyllo*-inosose formed by KanC	82
3.23	First transamination reaction catalysed by the KanS1 enzyme using L-glutamine as aminodonor and 2-deoxy-*scyllo*-inosose as aminoacceptor	83
3.24	TLC detection of 2-deoxy-*scyllo*-inosamine formed from G-6-P and L-glutamine in the coupled reaction catalysed by KanC and KanS1	84
3.25	TLC detection of α-keto-glutaramate formed in the coupled reaction catalysed by KanC and KanS1	85
3.26	HPLC analysis of the KanC/KanS1 assays for the production of 2-deoxy--*scyllo*-inosamine	86
3.27	Spectrophotometric assay (NADH consumption) for measurement of the KanS1 activity	87
3.28	TLC detection of 2-deoxystreptamine (2DOS) formed by the 3-enzymes catalysed reaction	88
4.1	A proposal for the biosynthetic route to paromamine in the producers of the major classes of 2DOS-ACAGAs	93
4.2	Proposal of a general pathway design for the biosynthesis of the 4,5-glycosylated 2DOS-ACAGAs starting from paromamine	97
4.3	Proposal of a general pathway design for the biosynthesis of the 4,6-glycosylated 2DOS-ACAGAs starting from paromamine	102
4.4	A proposal for a biosynthetic pathway for the 4-glycosylated 2DOS-ACAGAs (apramycins) starting from 2DOS	105
4.5	Proposal of a biosynthetic pathway for the 5-glycosylated 2DOS-ACAGAs (HM-B) starting from 2DOS	107
4.6	Proposal of a pathway for the biosynthesis of FTM and IM group starting from G-6-P	111

Summary

This dissertation focuses mainly on the elucidation of the genetics and the proposed pathways for the production of 2DOS-containing and related aminocyclitol aminoglycoside antibiotics (ACAGAs). The following results using different strategies had been achieved:

1) Homologous and heterologous probes were constructed for screening the cosmid banks and the positively hybridized cosmids were identified. Analysis of the respective cosmids (restriction analysis, PCR, cosmid mapping) was carried out to select one or more cosmids harboring the ACAGAs biosynthetic gene clusters (or part thereof) for sequencing;

2) The insert sequences of a single or of overlapping cosmid clones were determined, analysed, sequenced and submitted to the EMBL gene bank and received the accession codes;

3) The biosynthetic gene clusters for NM, RM, PM, LM, KM, TM, GM, FTM, IM, Apr and HM-B have been fully sequenced and analyzed except the one for BU which was still incomplete;

4) Further proposals of a general pathway design for the biosynthesis of the respective ACAGAs was made;

5) A characteristic feature of the *gen*-, *kan*- and *tob*-clusters is that some of the centrally important genes, also conserved in the other ACAGAs gene clusters had been duplicated or even multiplied in the past, most probably in order to acquire new biosynthetic functions. Interestingly, analysis of the *for*-, *gen*- and *kan*-clusters gave a good evidence that GMs are products formed via a hybrid of KM/FTM pathway;

6) Newly developed heterologous primers have been designed and tested for their efficiency via detection of various homologous biosynthetic genes in different ACAGAs producing strains;

7) The three *kanC*, *kanS1* and *kanE* genes were selected from the KM biosynthetic gene cluster for biochemical characterization of their encoded proteins. In order to achieve this goal, the three genes were cloned, heterologous expressed in *E. coli* and/or *S. lividans* TK23 followed by biochemical analysis of the expressed proteins. TLC, HPLC and spectrophotometric assays were the tools used for testing the respective enzyme activities;

8) Results showed that *kanC* encodes a 2-deoxy-*scyllo*-inosose synthase, *kanS1* encodes a 2-deoxy-*scyllo*-inosose aminotransferase as well as, 1-keto-2,3-deoxy-3-amino-*scyllo* inositol aminotransferase [bifunctional enzyme] and *kanE* was high likely to encode a 2-deoxy-*scyllo*-inosamine 1-dehydrogenase (in the presence of Zn^{++} ion and NAD).

Zusammenfassung

Die Dissertation befaßt sich hauptsächlich mit der Aufklärung der Genetik und der postulierten Biosynthesewege der 2DOS- und verwandten Aminocyclitol-Aminoglykosid-Antibiotika (ACAGAs).

Folgende Ergebnisse konnten mittels der beschriebenen Vorgehensweisen erzielt werden:

1) Durch die Konstruktion homologer und heterologer Sonden (Primer) konnten aus den entsprechenden Cosmidbänken positiv hybridisierende Cosmide identifiziert werden. Die anschließenden Analysen (Restriktion, PCR, Cosmidkarte) wurde durchgeführt um die Cosmide für die Sequenzierung auszuwählen, die ACAGAs-Gencluster (oder Teile davon) enthielten;

2) Die Insertsequenzen einzelner bzw. überlappender Cosmid-Klone wurde sequenziert, analysiert, und bei der EMBL Genbank eingereicht;

3) Die Biosynthese-Gencluster für Neomycin, Ribostamycin, Paromomycin, Lividomycin, Kanamycin, Tobramycin, Gentamicin, Fortimicin, Istamycin, Apramycin und Hygromycin B wurden analysiert und vollständig sequenziert. Das Butirosin-Gencluster konnte nur teilweise charakterisiert werden;

4) Weiterführende Aussagen über einen allgemeinen Biosyntheseweg für die entsprechenden ACAGAs konnten getroffen werden;

5) Ein Charakteristikum der *gen*-, *kan*- und *tob*-Cluster ist, daß einige der zentral wichtigen Gene, ebenfalls konserviert in anderen ACAGAs Genclustern sind, in der Vergangenheit dupliziert oder sogar multipliziert wurden, und so neue biosynthetische Funktionen erworben wurden. Interessanterweise, ergab die Analyse der *for*-, *gen*- und *kan*-Cluster hohe Evidenz dafür, daß die Gentamicine Produkte sind, die aus einem Hybrid der Kanamycin/Fortimicin Biosynthesewege entstanden sind.

6) Es wurden verschiedene heterologe Primer entwickelt um Biosynthesegene aus unterschiedlichen ACAGA produzierenden Stämmen zu detektieren und ihre Anwendbarkeit getestet;

7) Aus dem Kanamycin Biosynthese-Gencluster wurden die drei Gene *kanC*, *kanS1* und *kanE* ausgewählt, um die Genprodukte biochemisch zu charakterisieren. Um dies zu erreichen wurden die drei Gene kloniert und in *E. coli* und/oder *S. lividans* TK23 expremiert. Die entsprechenden Enzymaktivitäten wurden mit DC, HPLC und spektrophotometrischen Analysen belegt;

8) *kanC* codiert für eine 2-Desoxy-*scyllo*-Inosose Synthase, *kanS1* für eine 2-Desoxy-*scyllo*-Inosose Aminotransferase, sowie für eine 1-keto-2,3-Desoxy-3-Amino-*scyllo*-Inositol- Aminotransferase (bifunktionelles Enzym) und *kanE* für eine 2-Desoxy-*scyllo*-Inosamine-1-Dehydrogenase (in Anwesenheit von Zn^{++} Ionen und NAD).

1 Introduction

1.1 Aminoglycoside antibiotics (AGAs)

1.1.1 Definition

AGAs are a heterogeneous chemical class of natural products composed of most strongly modified sugar units bound glycosidically (mono- to oligosaccharides) and cyclitol derivatives (aglycone) containing amino nitrogen (Umezawa *et al*., 1986; Rinehart *et al*., 1992; Piepersberg, 1995 and 1997; Piepersberg and Distler, 1997). AGAs are idiolities (nonessential, low molecular weight compounds formed enzymatically by a limited number of species) synthesized by certain soil bacteria (Walker, 1995). The most important group of AGAs are produced by actinomycetes, bacilli and pseudomonades. However, the most currently known members are those produced by actinomycetes. AGAs are considered one of the most important classes of antibiotics both on the biological and industrial levels. Aminoglycosides including the non-antibiotic members constitute a group of natural products which is heterogeneous with respect to their putative evolution and ecology as well as the biochemistry of the molecules and their action, presumed biosynthetic routes, primary metabolic sources, resistance of producers against their end products and regulation of their production (Umezawa and Hooper, 1982; Piepersberg *et al*., 2002). AGAs are both indispensable chemotherapeutics and interesting targets of basic research (Piepersberg, 1997). The most attractive aspects of aminoglycoside research are: i- the emergent genetic and biochemical data on their biosynthesis and regulation (Walker, 1971 and 1995; Rinehart and Stroshane, 1976; Grisebach, 1978; Piepersberg, 1997; Piepersberg *et al*., 2002); ii- the molecular mechanisms of resistance development in both producers and clinically relevant pathogens (Cundliffe, 1989; Shaw *et al.*, 1993; Piepersberg, 1997); iii- the molecular aspects of interaction with cellular components in both prokaryotes and eukaryotes (Fong and Bergheim, 2002); iv- new fields of application and new pharmacologically relevant targets such as the successful glycosidase inhibitors (Piepersberg *et al*., 2002). In this study, the biosynthetic gene clusters of 2DOS-containing and related AGAs having antibacterial activity were studied.

1.1.2 Classes of aminoglycoside antibiotics

Aminocyclitol aminoglycoside antibiotics (ACAGAs) are considered the predominant representatives of the aminoglycoside group. They are classified according to the chemical nature

of the aminocyclitol moiety, the basic aglycone unit in all the ACAGAs into the following:

A- Streptamine containing ACAGAs

The streptomycins, historically marking the entrance to aminoglycoside research, are basically composed of a *scyllo*-inositol-derived aminocyclitol (streptidine or bluensidine) which formally are derived from streptamine, the 1,3-cis-diamino-cyclitol and sugar units glycosidically attached to this aglycone. Several other ACAGAs probably are biosynthetically related to the streptomycins as judged from the structure of their cyclitol moieties and other motifs: bluensomycin; ashimycin A and B; and spectinomycin are members of this group which contain a modified cyclitol unit called actinamine (Piepersberg, 1997). The chemical structures of streptomycin and spectinomycin are shown in Fig. 1.1.

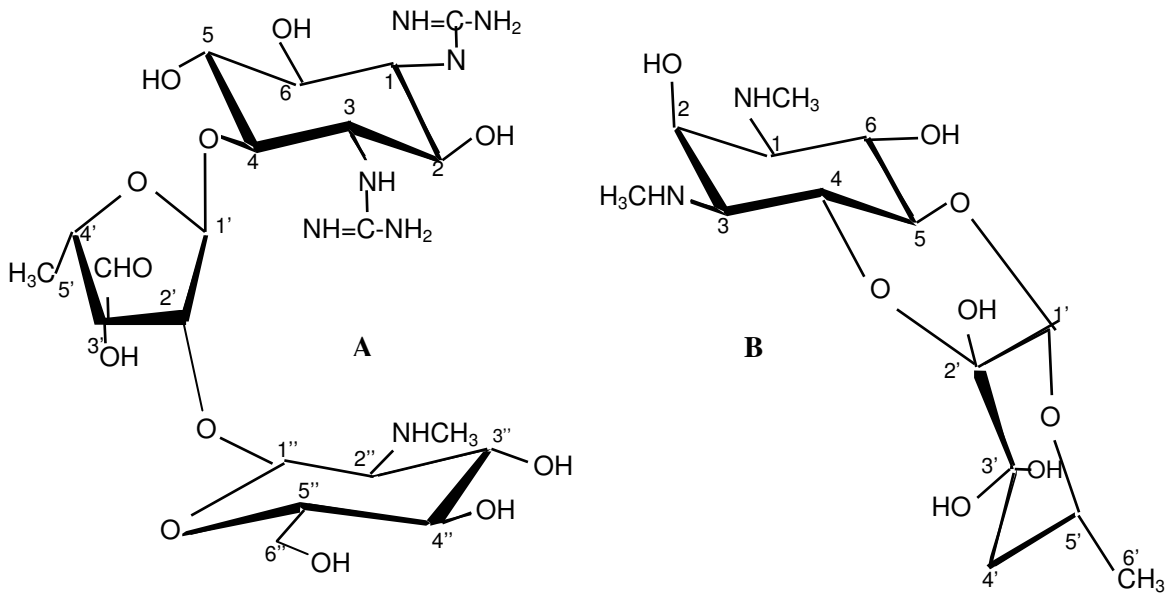

Fig. 1.1 Chemical structures of streptomycin (A) and spectinomycin (B). Both antibiotics represent examples of streptamine-containing ACAGAs.

B- 2-deoxystreptamine containing ACAGAs (2DOS-ACAGAs)

The major class of the 2-deoxystreptamine (2DOS)-containing ACAGAs represent a relatively homogeneous biosynthetic group, since most of them share the common pseudodisaccharidic intermediate paromamine (Pearce and Rinehart, 1981; Kase *et al*., 1982; Okuda and Ito, 1982; Umezawa *et al*., 1986; Piepersberg, 1997). The most important members of the 2DOS class are

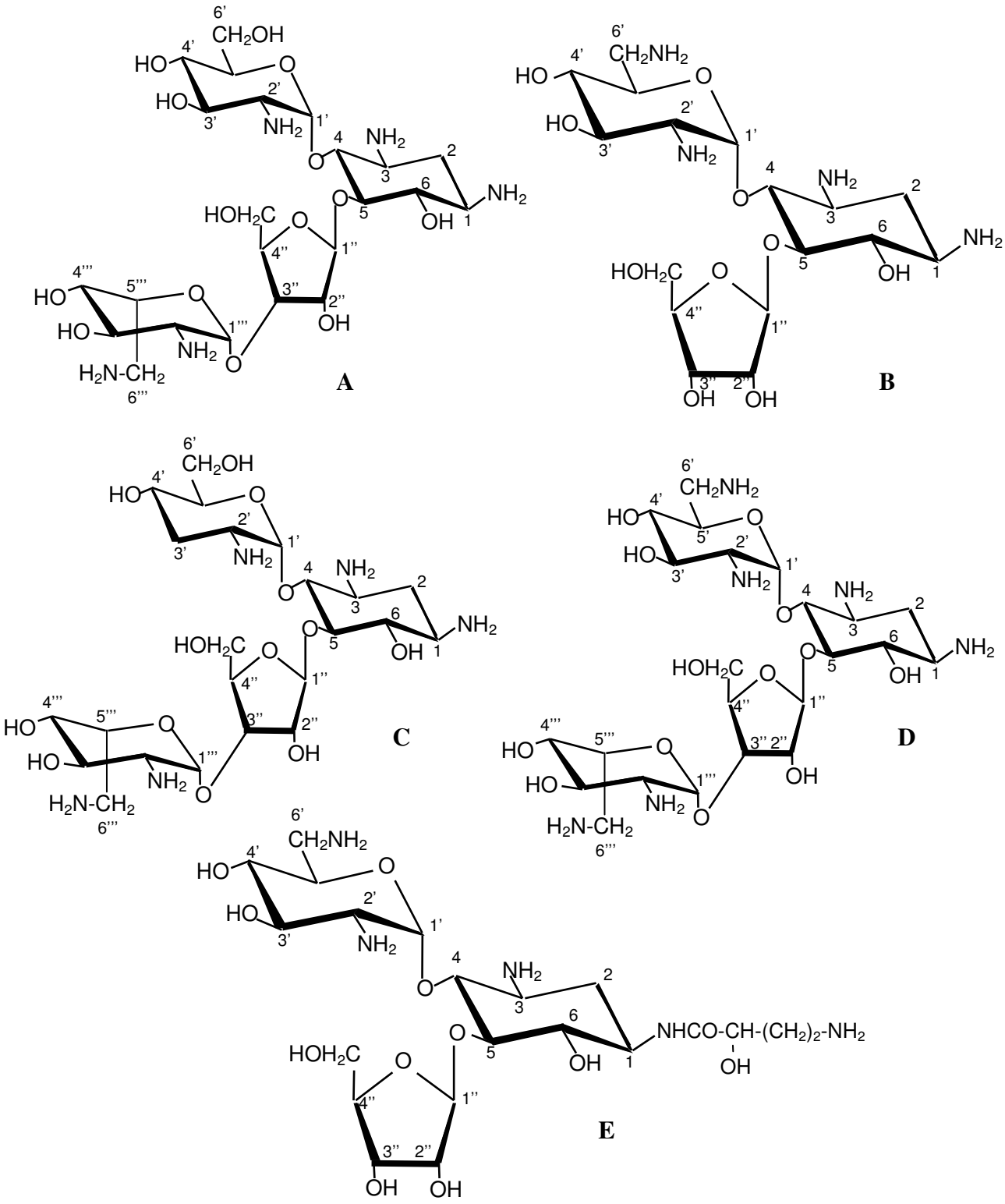

Fig. 1.2 Chemical structures of some selected 4,5-glycosylated ACAGAs (NM family). The letters in bold represent: A = paromomycin I; B = ribostamycin; C = lividomycin B; D = neomycin B; E = butirosin B.

the NMs (including PM, LM, RM and BUs), KMs and GMs. More distantly related groups are the 2DOS containing non-paromamine ACAGAs, members of which are Apr and HM-B (Fig. 1.4; Piepersberg, 1997). In order to focus more on chemistry and biosynthetic origin of the 2DOS-ACAGAs class, further classification based on the substitution of the 2DOS unit with the sugar side chains was attempted as follows: i- ACAGAs containing 4,5-di-glycosylated 2DOS-ACAGAs, this subclass includes both pseudotetrasaccharidic and pseudotrisaccharidic antibiotics, i.e. basically the NM family (Fig. 1.2). ii- ACAGAs containing 4,6-di-glycosylated 2DOS-ACAGAs, this subclass includes the family of KMs, and GMs (Fig. 1.3); iii- ACAGAs containing 4-mono-glycosylated 2DOS-ACAGAs (e.g. Apr; Fig. 1.4); iv- ACAGAs containing 5-mono-glycosylated 2DOS-ACAGAs (e.g. HM-B; Fig. 1.4).

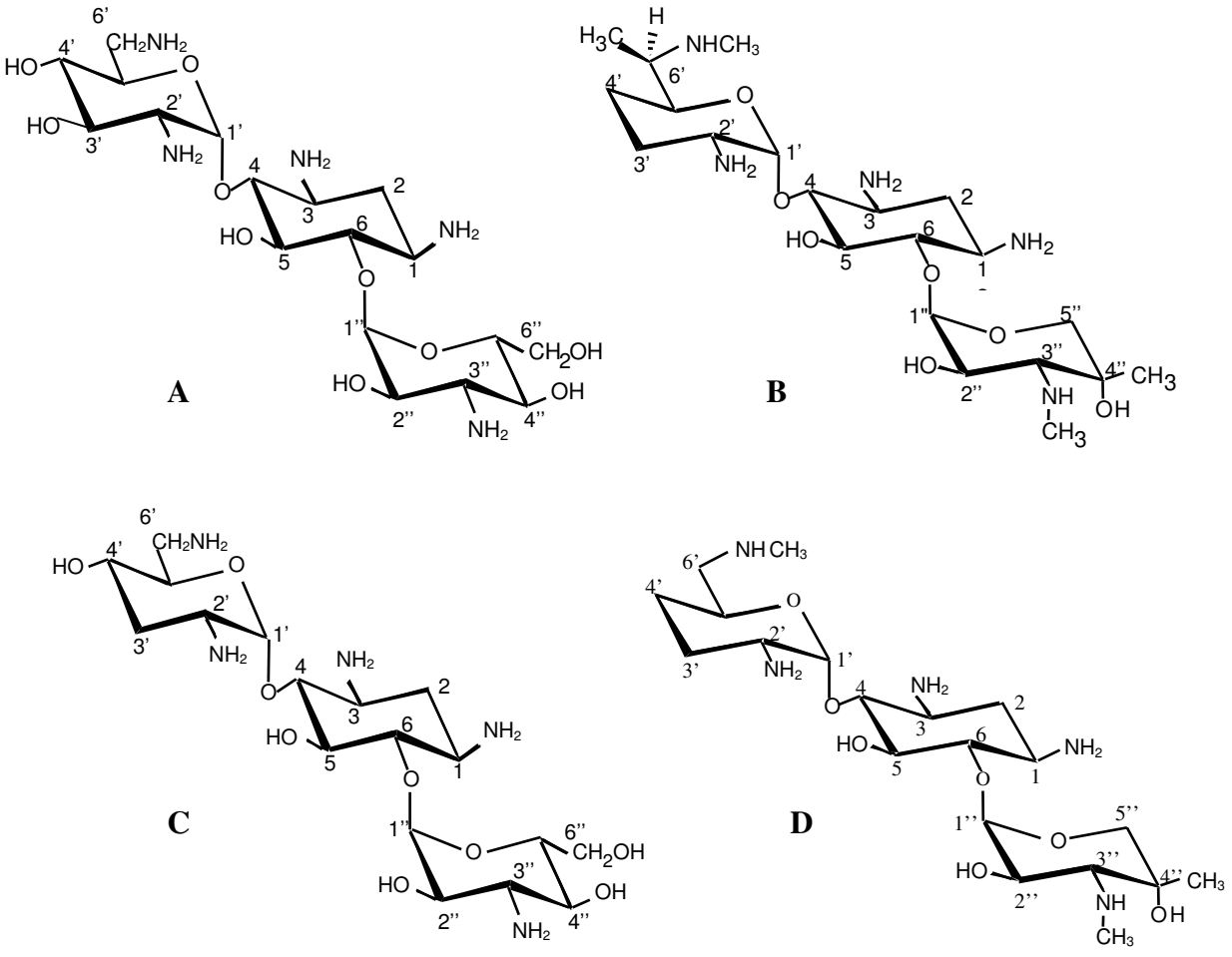

Fig. 1.3 Chemical structures of some selected 4,6-glycosylated ACAGAs. The letters in bold represent: A = kanamycin B; B = gentamicin C1; C = tobramycin; D = gentamicin C1a.

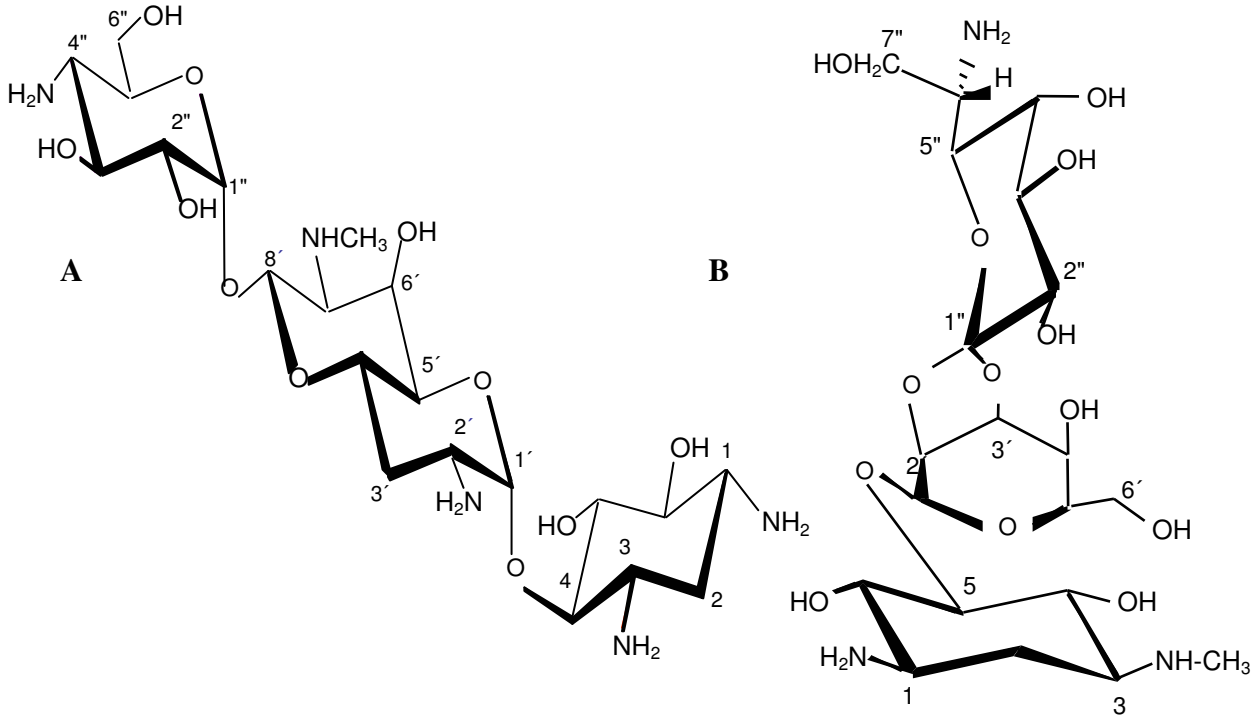

Fig. 1.4 Chemical structures of apramycin (A) and hygromycin B (B). Both apramycin and hygromycin B represent examples of 4- and 5-mono-glycosylated 2DOS-ACAGAs, respectively.

C- Fortamine and 2-deoxyfortamine-containing ACAGAs

Both FTMs (astromicins) and IMs form a group of compounds strongly related to the 2DOS-ACAGAs, containing fortamine and 2-deoxyfortamine as a basic cylitol unit, respectively. The more recently detected FTM/IM group of aminoglycosides has been intensively studied throughout the 1980s and several new members or producers have been described: the 2''-N-formimidoyl derivative of FTM (identical with dactimicin) produced by *M. olivasterospora* (Hotta *et al*., 1989b); derivatives of IM-A and -B produced by *S. tenjimariensis* (Kondo *et al*., 1982); derivatives of sporaricin A produced by *Saccharopolyspora hirsuta* (Umezawa *et al.*, 1987); the new variant lysinomicin produced by *M. pilosospora*. Interestingly, the members of this group produced by *Micromonospora* sp. (FTM) and *Dactylosporangium* sp. (dactimicin) are based on fortamine, whereas the related compounds produced by *Streptomyces* sp. (IMs) and *Saccharopolyspora* sp. (sporaricins) are based on 2-deoxyfortamine (Piepersberg, 1997). The chemical structures of both FTM-A and IM-A are outlined in Fig. 1.5.

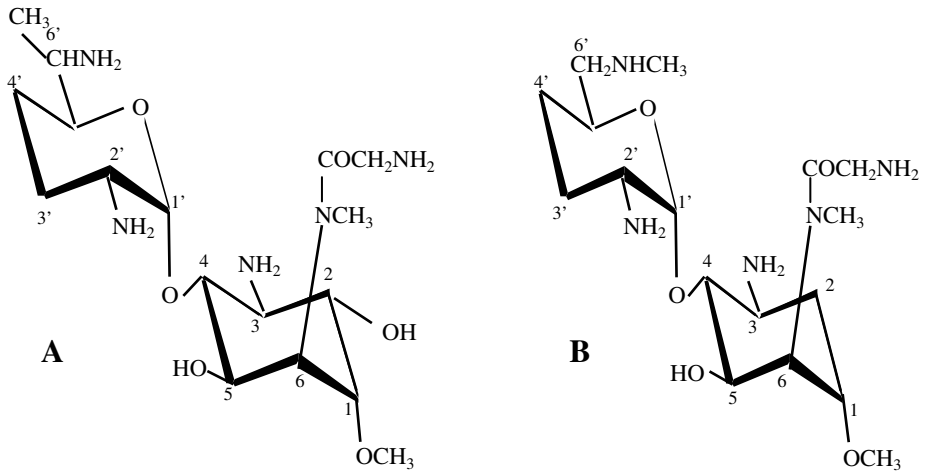

Fig. 1.5 Chemical structures of fortimicin A (A) and istamycin A (B). Both fortimicin A and istamycin A represent examples of fortamine and 2-deoxyfortamine-containing ACAGAs, respectively.

E- C_7-Aminocyclitol-Aminoglycoside (C_7-ACAGA)

Another group of ACAGAs containing a C_7-cyclitol moiety that is not derived from glucose was firstly described in the early 1970s. This class includes two variably glucosylated compounds: the validamycins and their nonglucosylated relatives, the validoxylamines; and the acarbose-related compounds (amylostatins and others). Also, a monomeric C_7-aminocyclitol, valiolamine which is assumed to be an intermediate or a side product of validamycin biosynthesis has been described as an independent end product of *S. hygroscopicus* subsp. *limoneus* (Asano *et al*., 1990b). The acarbose related metabolites of various actinomycetes act as inhibitors of various glycosidases, for example α-glycosidases and trehalases (Truscheit *et al*., 1981; Müller, 1989). Several research groups have recently done much work to investigate the basic pathways leading to such unusual compounds (Takeuchi *et al*., 1990; Asano *et al*., 1990a; Stratmann *et al*., 1999; Zhang *et al*., 2002).

E- Semisynthetic ACAGAs

Semisynthetic derivatives of chemotherapeutically successful ACAGAs have been obtained in a limited number by preparative organic chemistry. Interestingly, the only compounds that were chemotherapeutically successful so far were those into which structural alterations had been introduced that mimicked naturally occurring modifications, thereby resulting in activity against

clinically relevant (nosocomial) pathogens conferring ACAGAs resistance determinants. The most important examples are: amikacin, dibekacin (and its derivative arbekacin); and netilmicin and isepamicin which were derived from the members of the KM and GM families, respectively (Kawaguchi *et al.*, 1972; Kondo *et al.*, 1973; Nagabhushan *et al.*, 1978; Holm *et al.*, 1983; Tanaka *et al.*, 1983; Inoue *et al.*, 1994). The major alteration leading to potent antibiotics such as 3,4-dehydroxylation of a hexosamine moiety and introduction of an N-1-α-hydroxy-γ-aminobutyryl residue (= AHB; or its propionyl analog, AHP), also similarity occur in the naturally produced FTM (cf. Fig. 1.5) and the BUs (cf. Fig. 1.2), respectively. Some selected semisynthetic ACAGAs are structurally outlined in Fig. 1.6.

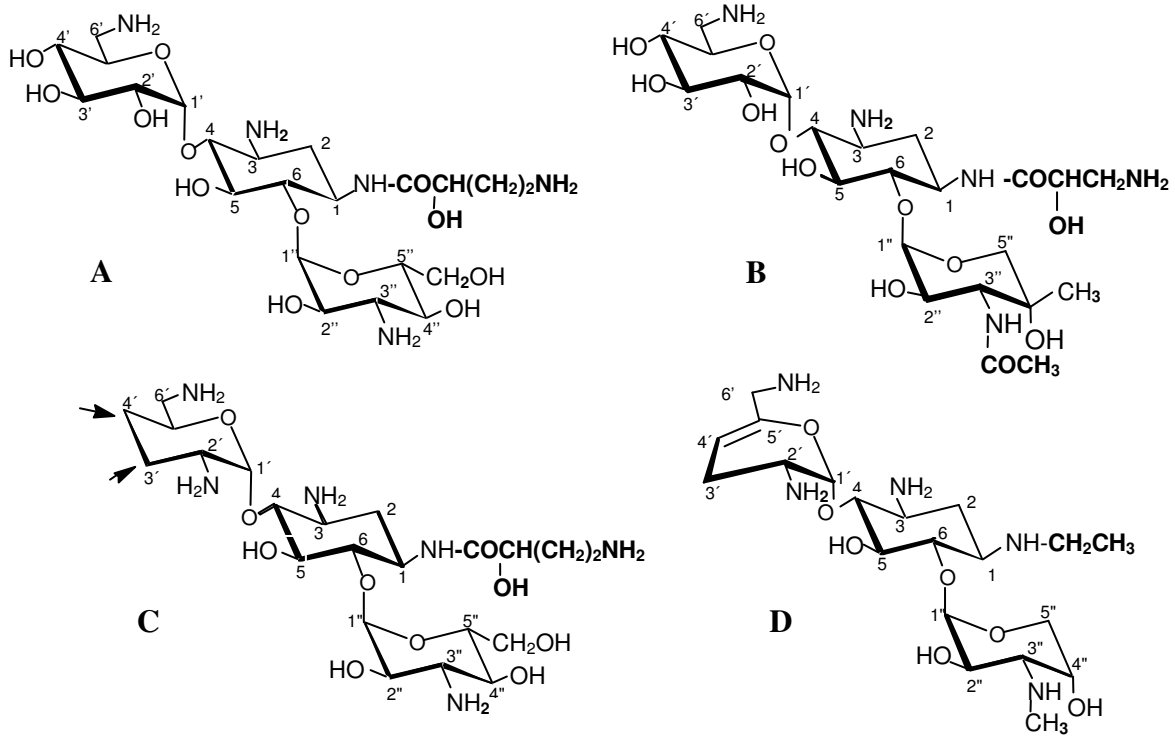

Fig. 1.6 Chemical structures of some selected semisynthetic ACAGAs. A = amikacin (N-1-AHB-kanamycin A); B = isepamicin (N-1-AHP-N-3''-acetylgentamicin B); C = arbekacin (habekacin; N-1-AHB-derivative of dibekacin = 3',4'-dideoxykanamycin B); D = netilmicin (N-1-ethylsisomicin). AHB = α-hydroxy-γ-aminobutyryl residue; AHP = α-hydroxy-β-aminopropionyl residue. The arrows and the substitution groups in bold represent the type of alterations made on the naturally occurring ACAGAs.

E- Other ACAGAs

Structurally more distantly related to the aforementioned ACAGAs classes and of even higher uncertainty in their pathway relationships are a number of old and new aminoglycoside-related secondary metabolites. For instance, the hexitol-containing sorbistins are produced interestingly

in both actinomycetes and Gram-negative bacteria (Piepersberg, 1997). Moreover, the aminoglycoside hygromycin A which is produced in the same strain that produces HM-B, the monosaccharidic glycosidase inhibitors and antibiotics (such as prumycin, galactostatin, sistatin and streptozotocin) and several disaccharidic amino-N-containing secondary or semisynthetic carbohydrates lack an (amino-)cyclitol constituent (such as trehalosamines, mannosyl glucosaminide, etc.) are examples of this group (Numata *et al*., 1986; Tsuno *et al*., 1986; Hardick *et al*., 1992).

Completely new structural classes of ACAGAs have been detected in the past with the isolation of the allosamidins and trehazolin by screening actinomycetes cultures for the formation of glucosidase inhibitors. Allosamidin is the first chitinase inhibitor isolated from streptomycetes (Sakuda *et al*., 1987). Trehazolin (or trehalostatin) is a very specific trehalose inhibitor produced by *Micromonospora* sp. and *Amycolatopsis* sp. (Ando *et al*., 1991; Kobayashi and Shiozaki, 1994).

1.1.3 Mode of action

The major classes of AGAs inhibit cellular protein synthesis through their binding to the smaller 30S ribosomal unit of the bacterial ribosome (Moazed and Noller, 1987; Fourmy *et al*., 1996; Ban *et al*., 1998, 1999 and 2000; Cate *et al*., 1999; Clemons *et al*., 1999; Brodersen *et al*., 2000). The recent elucidation of the three-dimentional structures of the bacterial 70S ribosomes and ribosomal subunits now allow identifying the exact type of molecular interaction with their target sites. Specifically, AGAs bind to the A-site of the 30S ribosome, including conformational changes in key nucleotides in the decoding center of the 16S rRNA. As a result they affect the fidelity of protein translation via wrong codon-anticodon pairing which in turn is expressed as mistranslation of mRNA or premature termination of protein synthesis leading to cell death (Ban *et al*., 2000; Carter *et al*., 2000; Cech, 2000; Nissen *et al*., 2000; Pape *et al*., 2000; Schlünzen *et al*., 2001; Fong and Berghuis, 2002; Brodersen *et al*., 2002 and 2003).

1.1.4 AGAs in medical treatments

Aminoglycosides are considered one of the most important classes of antibiotics which have a broad spectrum of action (mostly bactericidal) against most of the Gram-positive and Gram-negative bacteria. AGAs are not stable in the gut and therefore, they are often administered intravenously or intramuscularly to treat serious systemically bacterial infections (e.g. GMs,

KMs, amikacin, TM etc.). Streptomycin is administered intramuscularly for the treatment of tubercle bacilli (TB) and leprosy caused by *Mycobacterium tuberculosis* and *Mycobacterium leprae*, respectively. Some aminoglycosides are also used orally to treat intestinal infections (e.g. NM, PM, LM), or topically to treat eye infections (e.g. NM, TM, KMs) or wounds and burns (NM, TM). The AGAs acting as glycosidase inhibitors, e.g. amylostatins (acarbose) are used in the treatment of diabetes patients. Validamycins are used as a plant-protecting antibiotic that acts against plant-pathogenic fungi which attack roots. They are also effective in protecting seedlings and clone cuttings against damping-off disease (Frommer *et al*., 1981; Kameda *et al*., 1984; Robson *et al.*, 1988).

1.1.5 Biosynthesis and regulation of ACAGAs

The evolutionary origins of bacterial and fungal secondary metabolites in general and of ecological meanings and their functions as well as of their biosynthetic pathways have started to be discussed several times in the past two decades (Ban *et al*., 1998 and 1999; Chadwick and Whelan, 1992). Principally, we do not know what their ecological meaning are and which selective pressures created their biosynthetic pathways and regulatory adaptations to the metabolic networking of the producing cells. Moreover, the sugar-based aminoglycosidic microbial products are found in two forms, as cell wall-attached polymers and as excreted low molecular weight compounds (Piepersberg, 1992; Piepersberg and Distler, 1997). Only the latter are regarded as secondary metabolites in the traditional sense. However both types of variable uses of rare sugars and cyclitols are based on the parallel and largely alternative use of very similar pools of genes/enzymes (Piepersberg *et al*., 2002). Therefore, their distribution in microbial genomes and use in individual cell systems might be of more general relevance for both the ecological understanding of secondary sugar metabolism and the biotechnological use in future pathway engineering (combinatorial biochemistry).

Mycothiol (MSH) which represents an example of a cell wall attached polymer is the major cytoplasmic thiol in actinomycetes, where it occurs instead of glutathione (in other bacteria and eukaryotes) as an antioxidant and redox buffering substance. MSH is an aminoglycoside-related substance involved in the reactions catalysed by NAD/coenzyme dependent enzyme and hence oxidized to mycothione which is later reduced by an NADPH-dependent reductase (Newton *et al*., 1995 2000; Norin *et al*., 1997). The occurrence of *myo*-inositol as an essential metabolite so far only found in actinomycetes, explains the stable existence of genes for a D-*myo*-inositol-3-P

synthase in actinomycete genomes. Several aspects in the design of a productive pathway for a strain specific and extracellulary targeted metabolite are of immediate interest for characterization and comparison of its natural function and evolutionary origin: i- The individual branch points between primary and secondary metabolism, i.e. the genes/enzymes starting the flow of intermediates into the branches of the biosynthetic pathway of the final product condensation will inform us about the use of primary precursors and possibly about general substrate flows inside the producing cell; ii- The adaptation of genes/enzymes involved in the biosynthetic steps of cytoplasmic activation, modification and condensation of oligomeric end products; iii- The mechanism of avoidance of accumulation of toxic intermediates and/or end products of the cytoplasmic and extraceullar biosynthesis will also reflect either principal or individual use of common strategies; iv- In general, it becomes apparent that the regulatory control of the secondary pathway couples the production of such metabolites not only to a distinct phase of the cell cycle, but also makes it responding to various environmental conditions; v- The mechanism and other biochemical features of the specific transport (in general export, but also import maybe relevant in some cases) of the product through the cytoplasmic membrane (Piepersberg *et al*., 2002).

Furthermore, studies on the biosynthesis of AGAs have utilized three main approaches: i- isotopic competition techniques with ^{14}C- and ^{13}C- labelled putative precursors; ii- idiotrophic mutants; iii- enzyme characterization and assays to detect the enzymatic reactions postulated to be involved (Demain and Inamine, 1970; Reinhart and Stroshane, 1976; Grisebach, 1978, Lucher *et al*., 1989; Ota *et al*., 2000; Walker, 2002). In general, the basic design of the biosynthetic pathways for AGAs differ considerably. Also, the sources of precursors, before all the (amino-) cyclitol moieties are quite variable. In the present study, a focus will be made on the biosynthesis of the more closely related 2DOS-, fortamine- and 2-deoxyfortamine-containing ACAGAs.

1.1.5.1 Biosynthesis of 2DOS-containing ACAGAs

Unfortunately, the biosynthesis of none of the classical 2DOS-ACAGAs such as the KMs, NMs or GMs has been studied biochemically in detail as yet. All these major 2DOS families are produced via a common intermediate, the pseudodiasaccharide paromamine condensed from the preformed 2DOS and D-glucosamine units (Piepersberg and Distler, 1997). Recently, new molecular data have been reported on the genetics and the biosynthesis of BU in the producer *B. circulans* including a description of the partial *btr*-gene cluster comprising at least 17 genes

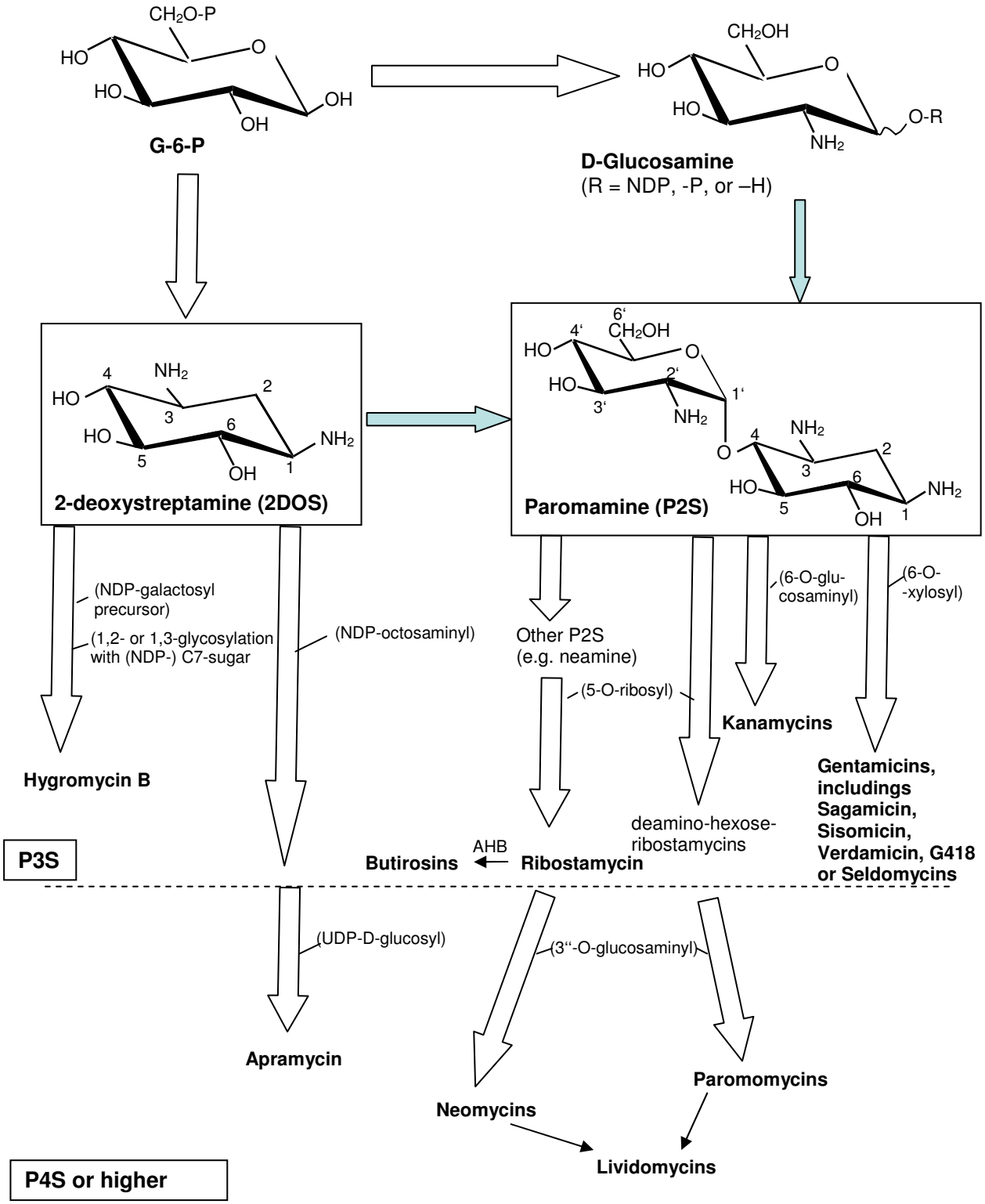

Fig. 1.7 General pathway design for the biosynthesis of the major classes of 2DOS-ACAGAs. The central role of 2DOS and paromamine intermediates are emphasized. The route leading to seldomycin, wether related to that of KMs or GMs, is unknown. The distribution of modifying phases are indicated with open arrows. AHB = α-hydroxy-γ-aminobutyryl residue; G-6-P = D-glucose-6-phosphate. P2S, P3S, P4S = pseudodi-, tri-, or tetrasaccharides, respectively.

(Kudo *et al*., 1999a and b; Ota *et al*., 2000). Gene disruption studies proved the function of the cluster in BU production. An extensive study of the first-step enzyme 2-deoxy-*scyllo*-inosose synthase, BtrC had already started before and as predicted, BtrC follows the same enzymatic mechanism as dihydroquinic acid synthase in cyclization of G-6-P via an intramolecular phosphate lyase which had classified as a Cb cyclitol pathway (dihydroquinate-synthase like 2DOS synthase; Yamauchi and Kakinuma, 1993, Piepersberg and Distler, 1997; Piepersberg, 1997). As well, the second-step enzyme, 2-deoxy-*scyllo*-inosose aminotransferase was later proven to be catalyzed by BtrS (Tamegai *et al*., 2002b). In fact, the protein sequence of BtrS demonstrated it to be a close relative of the *scyllo*-inosose aminotransferase, StsC in the streptomycin biosynthetic pathway (Ahlert *et al*., 1997).

Furthermore, the gene product BtrM of *B. circulans* is a glycosyltransferase related to others postulated to participate in bacterial cell wall polysaccharide synthases. Other gene products, BtrI, BtrK and maybe others could be involved in the multi-step AHB formation (Yukita *et al.*, 2003). In addition, genes for other steps such as modification of the 2-deoxy-*scyllo*-inosamine, glucosamine subunits, resistance and transport have not yet been identified.

For the 2DOS-containing non-paromamine ACAGAs (e.g. Apr and HM-B), the biosynthetic pathways have also not been characterized since the respective biosynthetic gene clusters have not yet been identified. However, certain genes encoding for resistance to the respective antibiotics such as *kamB* (Holmes *et al*., 1991) and HM-B *aph* (Zalacain *et al*., 1986) as well as some anticipated individual biosynthetic genes (*aprA*, AY129957; Walker, 2002) had been identified. The general pathway design for the biosynthesis of the major classes of 2DOS-ACAGAs is illustrated in Fig. 1.7.

1.1.5.2 Biosynthesis of fortamine and 2-deoxyfortamine-containing ACAGAs

Most information on the genetics and biochemistry of the biosynthesis of this group of ACAGAs originates from the investigation of the FTM (astromicin) producer *M. olivasterospora* and the IM/sannamycin producers *S. tenjimariensis* and *S. sannaensis* (Hasegawa, 1992; Hotta *et al*., 1994; Piepersberg and Distler, 1997). However, as pointed out earlier, the origin of the cyclitol moieties in this family of closely related compounds might follow either a Ca (*myo*-inositol phosphate synthase) in case of fortamine-containing ACAGAs (i.e. FTM, dactimicin) or a Cb (dihydroquinate synthase like 2DOS synthase; AroB proteins) in case of 2-deoxyfortamine containing ACAGAs (i.e. IM, sananamycin, sporaricins).

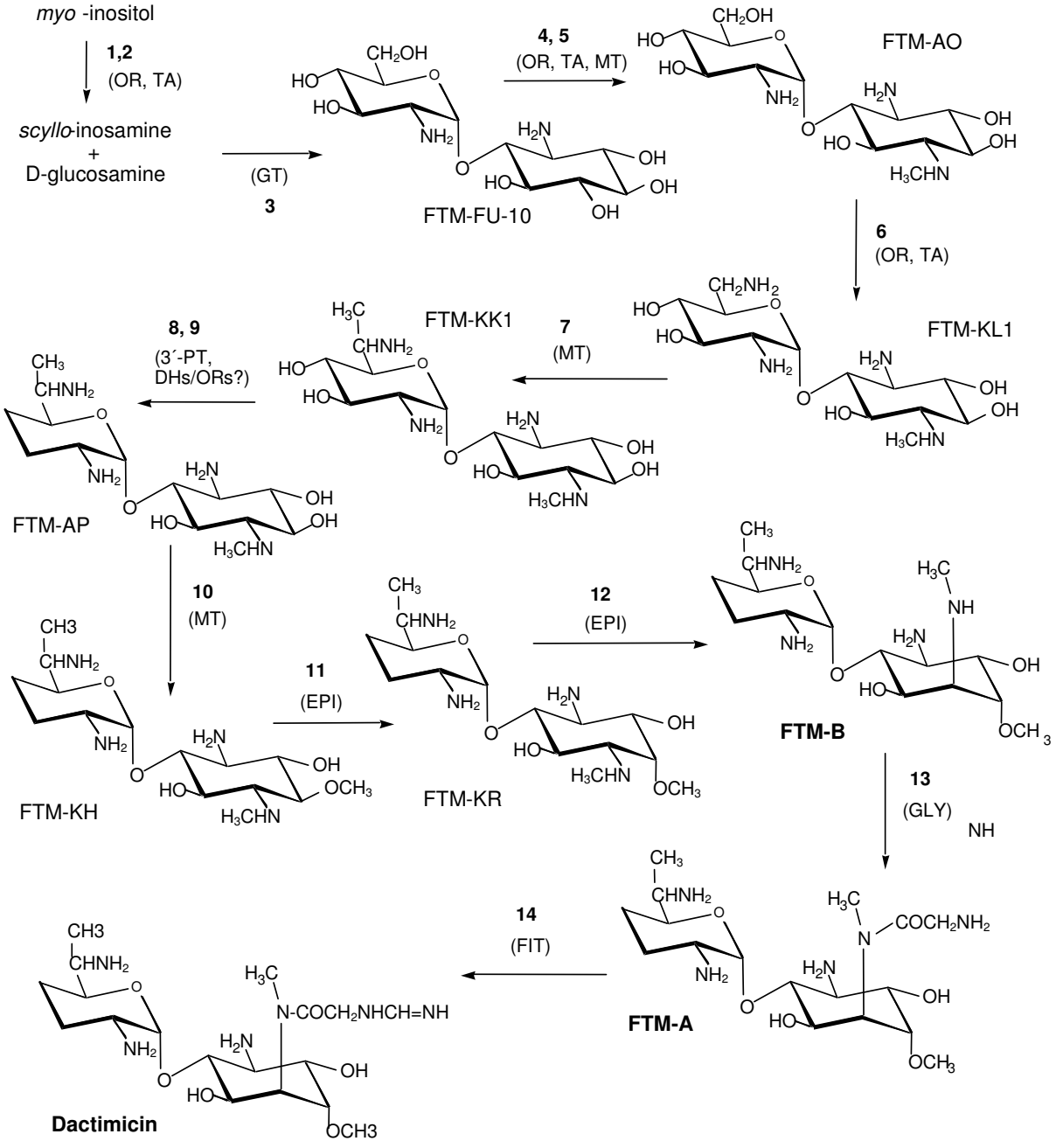

Fig. 1.8 The proposed biosynthetic pathway of fortimicins (astromicin). The same pathway starting from 2-deoxy-*scyllo*-inosamine appears to be established in the producers of IMs, sannamycins and sporaricin. The known intermediates and postulated enzymatic steps are given according to Hasawaga (1992); Hotta *et al.* (1994); Piepersberg (1997). The known or assumed enzymatic conversions involved in the pathway are given: DH = dehydration (dehydroxylation) or dehydratation (e.g. enolase); Epi = epimerization; FIT = forimidoyltransfer; FTM = fortimicin; GLY = glycyltransferase; GT = glycosyltransferase; MT = methyltransferase; OR = oxidoreductase; PT = phosphotransferase (kinase); TA = transaminase.

Note: in FTM producers NO BtrC-like function is required while in IM producers BtrC-like function is required; first two steps similar to that in 2DOS synthesis (cf. Fig. 4.1); Steps no. 8 and 9 occur also in GMs (Fig. 1.3).

The DNA sequences have been reported for a few *fms* genes in the respective cluster for FTM biosynthesis (Hotta *et al*., 1989b; Dairi *et al*., 1992a, b and c; Hasegawa, 1992; Ohta *et al*., 1993; Hotta *et al*., 1994). Most of these genes products are involved either in resistance or in the late modification pathway such as, FmrO, Fms7 and Fms14 (Kuzuyama *et al*., 1995; Dairi *et al*., 1992a). The branch points from primary metabolism can be identified to be the formation of *scyllo*-insose (in case of FTM and related antibiotics) and 2-deoxy-*scyllo*-inososе (in case of IM and related antibiotics) from G-6-P and synthesis of NDP-(N-acetyl-)D-glucosamine. The condensation reaction of the starting pseudodisaccharide (FTM-FU-10) from *scyllo*-inosose and D-glucosamine could be quite similar to that of the paromamine-forming pathway in the production of 2DOS-ACAGAs (Piepersberg *et al*., 2002).

The pathway design for the FTM/IM group is thus in the other extreme as compared to the distribution of modification and condensation reactions in the streptomycin biosynthetic route: here condensation comes first and major modifications second (Piepersberg *et al*., 2002; Fig. 1.8).

1.1.6 Resistance mechanisms

Another and arguably more alarming drawback of the aminoglycosides (and antibiotics in general) is the global development of microbial multi-resistance. In the case of ACAGAs, which are used predominantly against aerobic Gram-negative bacteria, the most common resistance mechanism is structural modification and thereby inactivation of the antibiotic compound by bacterial enzymes: e.g. aminoglycoside phosphotransferases (APH); adenyltransferases (AAD or ANT); and acetyltransferases (AAC). A widely distributed second mechanism of action is alteration of the target site (16S rRNA) via methylation (16S rRNA methyltransferase). A third type of resistance mechanism, mostly occurring via mutation is decreased uptake and/or accumulation of the drug in bacteria (Cundliffe, 1989; Shaw *et al.*, 1993; Piepersberg, 1997; Haddad *et al*., 1999; Mingeot-Leclercq *et al*., 1999; Neonakis *et al.*, 2003).

Aminoglycoside resistance is achieved in 2DOS producers typically by either a combination of aminoglycoside modification (N-1-acetylation and O-3'-phophorylation) in the producers of NMs and RM, HM-B or target site modification (16S rRNA methylation) in the producers of GMs, KMs, Apr (Zalacain *et al*., 1986; Holmes *et al*., 1991; Hotta *et al*., 1994; Piepersberg and Distler, 1997). On the other hand, resistance to the fortamine- (i.e. FTM, dactimicin) and 2-deoxyfortamine- (i.e. IM, sannamycin) containing ACAGAs is achieved by 16S rRNA

methylation of the target site in the producers, where sometimes acetylation of the antibiotic occurs as a second resistance mechanism (Ohta *et al*., 1993; Hotta *et al*., 1996).

1.1.7 Toxicity

Despite the apparent advantages, extensive clinical use of ACAGAs is limited due to the associated toxicities, most notably nephrotoxicity and ototoxicity and to a lesser extent neuromuscular blockade (Walker and Duggin, 1988; Zembower *et al*., 1998). The exact mechanism of toxicity is unknown although aminoglycosides are known to accumulate in renal cortical tubes and then be able to damage the proximal tubules. Nephrotoxicity is dose-dependant and generally reversible in the majority of patients when the drug is discontinued or using effective monitoring procedures (Prins *et al*., 1998). Ototoxicity is the second main adverse effect of aminoglycoside and in contrast to nephrotoxicity is irreversible since it leads to vestibular and/or cochlear damage.

Because ACAGAs are very effective antibiotics especially in the treatment of severe bacterial infections, it seems important to maintain and even develop efforts to improve their therapeutic indices. Clinical research aimed at obtaining intrinsically less toxic compounds has met modest success and a few of other approaches proposed to reduce the toxicities of the available agents have reached clinical practical applications (Piepersberg, 1997; Mingeot-Leclercq and Tulkens, 1999).

1.2 ACAGAs producers

Aminoglycosides are mainly produced by actinomycetes including *Streptomyces*, *Streptoverticillum*, *Nocardia*, *Micromonospora* and *Streptoalloteichus*. Some Bacilli and Pseudomonades species also produce aminoglycosides. These organisms often produce simultaneously several structurally related antibiotics which are mutually convertible by chemical, enzymatic, or microbiologic procedures and give valuable information on the biosynthesis of these aminoglycosides. In general, the Gram positive bacteria include two major branches: the low G+C organisms forming the order *Firmicutes*, containing genera such as *Bacillus*, *Clostridium*, *Staphylococcus* and *Streptococcus*; and the high G+C organisms forming the order *Actinomycetes*. (Kieser *et al.*, 2000). Tab. 1.1 summarizes the most useful and widely used ACAGAs produced by a variety of strains. However, in the present study, a focus was made

for some selected useful ACAGAs produced by some actinomycetes (including *Streptomyces*, *Micromonospora* and *Streptoalloteichus* sp.) and by *B. circulans*.

1.2.1 *Actinomycetales*

Streptomyces, *Micromonospora* and *Nocardia* are genera that belong to the order *Actinomycetales.* These contain Gram-positive bacteria that form branching filaments, all of them form a true mycelium and some are normally bacillary or irregular in shape. With a few exceptions notably the *Streptomyces* members of this order live saprophytically in soil however, most of them produce a variety of biologically active substances, e.g. various antibiotics, pigments, enzymes, amino acids, vitamins, toxins, growth promoting substances and others (Krassilnikov, 1950; Waksman, 1959). Members of the genus *Streptomyces* are filamentous gram positive bacteria (Waksman and Lechevalier, 1962) with a high potential in biotechnology (Piepersberg, 2003). This genus is defined by both chemotaxonomic and phenotypic characters (Miyadoh, 1990). The major taxonomical emphasis lies now on the 16S rRNA homologies, in addition to cell wall analysis, fatty acids and lipid patterns (Wellington *et al*., 1992; Mehling *et al.*, 1995b). One of the quickest methods for preliminary identification to genus level was the precursor of LL isomer of diaminopimelic acid (LL-DAP) as the amino acid in the cell wall peptidoglycan. *Streptomyces* are ubiquitous in nature. Their ability to colonize the soil is greatly facilitated by growth as a vegetative hyphal mass which can differentiate into spores which not only assist in spread and persistence (Mayfield *et al*., 1972; Ensign, 1978), but also impart resistance to low nutrient and water availability (Karagouni *et al*., 1993). Members of the genus *Streptomyces* are well known as producers of many secondary metabolites having some kind of activity (Zaehner and Fielder, 1995). Actinomycetes make two-thirds of the known antibiotics that are produced by microorganisms and among them nearly 80% are made by members of the genus *Streptomyces* (Omura, 1992; Piepersberg, 1993; Kieser *et al*., 2000). A typical property of the genomic DNA of streptomycetes is the high G+C content of 72 – 75 mol% (Enquist and Bradley, 1971; Piepersberg, 1993) which results in a highly biased codon usage (Wright and Bibb, 1992). The G+C content is of intermediate concentration (about 70%) in the first codon position, lowest in the second (about 50%) and highest (about 80-90%) in the third codon position. The family *Micromonsporaceae* comprises nine genera, namely *Micromonospora*, *Actinoplanes*, *Catellatospora*, *Couchioplanes*, *Catenuloplanes*, *Dactylosporangium*, *Pilmelia* (Koch *et al*., 1996b), *Spirilliplanes* (Tamura *et al.*, 1997) and *Verrucosispora* (Rheims *et al*.,

1998). Phylogenetically, this family forms one of the major sublines within the suborder *Micromonosporineae* in the order *Actinomycetales* (Stackebrandt *et al*., 1997). Genus *Micromonospora* (Ørskov, 1923) is the type genus of the family *Micromonosporaceae* and contains many interesting strains such as antibiotic producers i.e. *M. echinospora* (formerly *M. purpurea*), *M. olivasterospora* (Luedemann and Brodsky, 1964) and degraders of natural rubber (Jendrossek *et al.*, 1997).

Tab. 1.1 Widely used ACAGAs and their applications

ACAGAs	Producer(s)	Application
apramycin	*St. hindustanus & S. tenebrarius*	veterinary antibacterial
butirosin	*B. circulans*	antibacterial
fortimicin	*M. olivasterospora*	antibacterial
gentamicin	*M. echinospora*	antibacterial
hygromycin B	*S. hygroscopicus*	antihelminthic
istamycin	*S. tenjimariensis*	antibacterial
kanamycin	*S. kanamyceticus*	antibacterial
lividomycin	*S. lividus*	antibacterial
neomycin	*S. fradiae*	antibacterial
paromomycin	*S. rimosus*	antiamoebal
ribostamycin	*S. ribosidificus*	antibacterial
sisomicin	*M. inyoensis*	antibacterial
spectinomycin	*S. spectabilis*	antibacterial
streptomycin	*S. griseus*	antitubercle (anti TB)
tobramycin	*St. hindustanus & S. tenebrarius*	antibacterial

According to the approval lists of bacterial names (Skerman *et al.*, 1980), 12 species and seven subspecies are listed in the genus *Micromonospora.* A 16S rRNA sequence based study on the taxonomy of *Micromonosporaceae* has been performed by Koch *et al*., (1996a). The 16S rRNA-homologies, based phylogeny of the genus *Micromonospora*, however did not always agree with other taxonomic characteristics. It was assumed that the 16S rRNA homologies are not sufficiently divergent in the genus *Micromonospora* to distinguish the different species.

Yamamoto and Harayama, (1995) proposed that protein encoding genes such as *gyrB* could be suitable for phylogenetic classification and identification of the closely related bacteria.

They succeeded in discriminating among closely related strains of the genus *Pseudomonas* (Yamamoto and Harayama, 1995). In subsequent works (Yamamoto and Harayama, 1996, 1998; Harayama and Yamamoto, 1996; Yamamoto *et al*., 1999), they showed the *gyrB*-based grouping of the genus *Acinetobacter* to be consistent with the results of the DNA-DNA hybridization, indicating the *gyrB* sequence would be useful for resolving bacterial strains at the genomic species level. Moreover, the same method was applied to distinguish different species of *Micromonospora* and it was succeeded to distinguish 15 validly described species and four subspecies. Also this method showed that they form a tight cluster as previously demonstrated by 16S rRNA homologies. However, the intrageneric relationships deduced from the *gyrB*-based phylogeny were different from those based on their 16S rRNA homologies (Kasai *et al.*, 1998).

1.2.2 *Bacillus*

Members of the genus *Bacillus* are aerobic or facultatively anaerobic endspore-forming Gram-positive rods. The ubiquity of *Bacillus* species in nature, the unusual resistance of their endospores to chemical and physical agents, the developmental cycle of endospore formation, the production of antibiotics, the toxicity of their spores and protein crystals for many insects, and the pathogen *B. anthracis*, have attracted ongoing interest in the genus since Koch's time (Sneath *et al*., 1986; Topley and Wilson, 1998). The cell wall peptidoglycan of most species belongs to a type with directly cross-linked *meso*-diaminopimelic acid (Schleifer and Kandler, 1972).

Most *Bacillus* species are versatile chemoheterotrophs capable of respiration using a variety of simple organic compounds (sugars, amino acids, organic acids). In some cases, they also ferment carbohydrates in a mixed reaction that typically produces glycerol and butanediol. A few species, such as *B. megaterium*, require no organic growth factors, others may require amino acids, B-vitamins, or both. The majority are mesophiles, with optimum temperatures between 30 and 45°C, but the genus also contains a number of thermophilic species with optimum growth temperatures as high as 65°C. In the laboratory, under optimum growth conditions, *Bacillus* species exhibit generation times of about 25 minutes (Norris *et al*., 1981; Berkeley *et al.*, 1984; Priest, 1993). Early attempts for setting-up a classification system for *Bacillus* species were based on two characteristics: aerobic growth and endospore formation. This resulted in tethering

together many bacteria possessing different kinds of physiology and occupying a variety of habitats. Hence, the heterogeneity in physiology, ecology, and genetics, makes it difficult to categorize the genus *Bacillus* or to make generalizations about it. There is great diversity in physiology among members of the genus, whose collective features include: degradation of most or all substrates derived from plant and animal sources including cellulose, starch, pectin, proteins, agar, hydrocarbons, and others; antibiotic production; nitrification; denitrification; nitrogen fixation; facultative lithotrophy; autotrophy; acidophily; alkaliphily; psychrophily, thermophily and parasitism. Spore formation universally found in the genus is thought to be a strategy for survival in the soil environment wherein the bacteria predominate. Aerial distribution of the dormant spores probably explains the occurrence of *Bacillus* species in most habitats examined (Hurst and Gould, 1984; Dring *et al*., 1985).

Currently, there are 40 recognized species in the genus *Bacillus* listed in Bergey's Manual of Systematic Bacteriology, Vol. 2 (1986). The G+C content of known species of *Bacillus* ranges from 32 to 69% (DeLey, 1978; Golovacheva *et al*., 1975). This observation, as well as DNA hybridization tests, reveals the heterogeneity of the genus. Not only is there variation from species to species, but there are sometimes profound differences in G+C content within strains of a species. For example, the G+C content of the *B. megaterium* group ranges from 36 to 45 mol% and of *B. circulans* ranges from 31.6 to 61.0 mol% (Nakamura and Swezey, 1983; Famy *et al*., 1985). Another approach to *Bacillus* taxonomy has been analysis of 16S rRNA molecules by oligonucleotide sequencing. This technique, of course, also reveals phylogenetic relationships. Surprisingly, *Bacillus* species showed a kinship with certain nonsporeforming species, including *Planococcus, Lactobacillus* and *Staphylococcus* (Woese *et al*., 1975; Joung and Côté, 2002.). In one study, 16S rRNA cataloging showed that *B. subtilis* and other ellipsoidal-sporeforming species, *B. cereus, B. megaterium,* and *B. pumilus,* formed a coherent cluster, but the round-sporeforming species, *B. sphaericus, B. globisporus,* and *B. aminovorans,* did not cluster. In another 16S rRNA sequencing study, three major *Bacillus* taxonomic cluster groups were defined by determining complete or partial sequences of 16S RNA (exceeding 1100 Nt) on 35 recognized reference strains. These cluster groups were quite different from those previously noted. Finally, species of *Bacillus* are attractive microorganisms for industrial application due to several characteristics: i- they synthesize products of commercial interest (e.g. antibiotics, amino acids, etc); ii- they have a large range of metabolic properties (proteolytic, lipolytic, saccharolytic, etc.); iii- most of them are non-pathogenic.

1.3 Aim of the work

ACAGAs are considered one of the most important groups of antibiotics especially in the medical field. However, our knowledge on the genetics and biochemistry for the biosynthesis of such metabolites is still incomplete, especially for the major classes of 2DOS- and 2DOS-related ACAGAs. So, the main goal attempted in this work falls into two main items:

I- Cloning, identification and sequencing the biosynthetic gene clusters from some selected producers of ACAGAs. To fulfil the regarded goal, the following strategies were carried out:

1- Construction of appropriate probes for screening the cosmid banks delivered on a service basis.

2- Identification, mapping and analysis of the positively custom-screened cosmids for selection of those harboring the ACAGAs biosynthetic gene clusters for sequencing.

3- Analysis and evolutionary study of the respective gene clusters.

4- Development of heterologous primers for isolation and characterization of strains producing hybrid ACAGAs.

II- Biochemical characterization of some selected 2DOS biosynthetic genes. In order to achieve the respective aim, the three genes *kanC*, *kanS1* and *kanE* which are putatively involved in the biosynthesis of the 2DOS moiety were selected from the *kan*-gene cluster to fulfil the following points:

1- Cloning and heterologous expression of the selected genes.

2- Biochemical characterization of the respective expressed proteins.

2 Materials and Methods

2.1 Chemicals and enzymes

Antibiotics

Ampicillin	Roth, Karlsruhe
Apramycin	Sigma, Steinheim
Chloramphenicol	Serva, Heidelberg
Kanamycin	Sigma, Steinheim
Nystatin	Merck, Darmstadt
Spectinomycin	Sigma, Steinheim
Thiostrepton	Squibb and Sons, Princeton, USA

Chemicals

All the chemicals used were of high quality standards and were supplied from the following companies:

Fluka Chemie, CH-Buchs
Merck, Darmstadt
Roth, Karlsruhe
Sigma, Steinheim
Roche Diagnostics, Mannheim

Enzymes

DNA polymerase I Klenow fragment	Roche Diagnostics, Mannheim
L-Lactate dehydrogenase	Roche Diagnostics, Mannheim
Proteinase K	Sigma, Steinheim
Lysozyme	Serva, Heidelberg
Restriction endonucleases	Invitrogen, Karlsruhe Roche, Mannheim BioLabs, New England
RNase	Fluka Chemie, CH-Buchs
Taq DNA polymerase	Invitrogen, Karlsruhe
T4-DNA ligase	Roche Diagnostics, Mannheim
T4-polynucleotide kinase	BioLabs, New England

Kits

Thermo-Sequencing kit	Amersham Bioscience, Freiburg
BioRad Protein Assay Kit	BioRad, München
BM Chromagenic Western Blotting Kit	Roche Diagnostics, Mannheim
QIA prep Spin Miniprep kit	Qiagen, Hilden
QIA quick Gel extraction kit	Qiagen, Hilden
QIA quick PCR Purification kit	Qiagen, Hilden

Various materials

Hybond™-C Extra membranes	Amersham Bioscience, Freiburg
Hybond™-N^+ membranes	Amersham Bioscience, Freiburg
Membrane filters BA 85 (0.45 µm)	Schleicher & Schuell, Dassel
TLC plates (ALUGRAM® SIL G/UV_{254})	Merck, Darmstadt
3MM Whatman paper	Biometra, Göttingen
X-ray Hyperfilm- MP	Amersham Bioscience, Freiburg

2.2 Media

2.2.1 Media for culturing *E. coli* and *Bacillus* strains

LB-Medium (Miller, 1972)

Tryptone	10.0	g
Yeast extract	5.0	g
NaCl	5.0	g
ad. distilled water	1000.0	ml
pH adjusted to 7.2 before autoclaving		

LB-Agar (Miller, 1972)

LB-Medium		
Agar	15.0	g/l

LB-Medium with Sorbitol and Betaine (Chen *et al.*, 1972)

LB-Medium		
Sorbitol	1.0	M
Betaine	2.5	mM

2x TY-Medium (Miller, 1972)

Tryptone	16.0	g/l
Yeast extract	10.0	g/l
NaCl	5.0	g/l

SOB- Medium (Hanahan, 1983)

Tryptone	20.0	g/l
Yeast extract	5.0	g/l
NaCl	0.58	g/l
KCl	0.19	g/l
After autoclaving add:		
$MgCl_2$ (1M)	10.0	ml
$MgSO_4$ (1M)	10.0	ml

SOC- Medium (Hanahan, 1983)

Glucose	3.6	g/l
in SOB-Medium		

Slant-Medium (50:50) (Miller, 1972)

Tryptone	10.0	g
Yeast extract	5.0	g
Glycerin 96%	500.0	ml
Ad. distilled water	1000.0	ml

2.2.2 Media for culturing *Streptomyces* and *Micromonospora* strains

Trypticase soy broth (TSB) (Kieser *et al*., 2000)

Trypticase soy broth (Oxoid)	30.0	g/l
(soyabean casein digest broth)		

TSB-PEG 8000 (Babcock and Kendrick, 1988)

Trypticase soy broth	30.0	g/l
PEG 8000	50.0	g/l
After autoclaving add:		
Glycine (20%)	50.0	ml
$MgCl_2$ (1 M)	10.0	ml

SPMR (Babcock and Kendrick, 1988)

Sucrose	103.0	g/l
$MgCl_2$	10.0	g/l
Glucose	5.0	g/l
Yeast extract (Difco)	5.0	g/l
Bacto-agar (Difco)	22.0	g/l
TES, pH 7.2	4.58	g/l
After autoclaving add:		
$CaCl_2$ (5 M)	2.0	ml
Trace element solution	2.0	ml

Trace element solution (Kieser *et al*., 2000)

$ZnCl_2$	0.04	g/l
$FeCl_3 \times 6\ H_2O$	0.2	g/l
$CuCl_2 \times 2\ H_2O$	0.01	g/l
$MnCl_2 \times 4\ H_2O$	0.001	g/l
$Na_2B_4O_7 \times H_2O$	0.01	g/l
$(NH_4)_6Mo_7O_{24} \times 4\ H_2O$	0.01	g/l
Sterilized by filteration		

SMA (Distler *et al*., 1985)

Soy meal	20.0	g/l
Mannit	20.0	g/l
Agar	20.0	g/l

M65 (DSMZ, Braunschweig)

Glucose	4.0	g
Yeast extract	4.0	g
Malt extract	10.0	g
Agar	12.0	g
Distilled water ad.	1000.0	ml
pH adjusted to 7.2		

YEME (Kieser *et al*., 2000)

Yeast extract	3.0	g/l
Peptone	5.0	g/l
Malt extract	3.0	g/l
Glucose	10.0	g/l
Sucrose	340.0	g/l
After autoclaving add:		
$MgCl_2$ (1 M)	10.0	ml

2.2.3 Aminoglycoside production medium (Nam and Ryu, 1985)

Soy meal	30.0	g/l
Ammonium chloride	4.0	g/l
Calcium carbonate	5.0	g/l
Glycerol	40.0	ml
pH adjusted to 7.5		

2.3 Buffers and Solutions

2.3.1 For the preparation of plasmid DNA from *E. coli*

Alkaline lysis solution (Birnboim and Doly, 1979)

Solution I

Glucose	50.0	mM
Tris-HCl, pH 8.0	50.0	mM
EDTA, pH 8.0	10.0	mM

Solution II

SDS	1.0	%
NaOH	200	mM

Solution III (HSS; high salt solution)

Potassium acetate	3.0	M
Formic acid	1.8	M

2.3.2 For the preparation of competent *E. coli* cells (Hanahan, 1983)

TMF-1 buffer

$CaCl_2$ x $2H_2O$	100.0	mM
$RbCl_2$	50.0	mM
$MnCl_2$ x $4H_2O$	40.0	mM

2.3.3 For the preparation of genomic DNA

i- Pospiech & Neumann method (Pospiech and Neumann, 1995)

SET buffer

NaCl	75.0	mM
EDTA pH 8	25.0	mM
Tris-HCl pH 7.5	20.0	mM

TE buffer (Sambrook and Russell, 2001)

Tris	10.0	mM
EDTA	1.0	mM
pH 8.0		

ii- Modified CTAB method (Mehling *et al*., 1995a)

CTAB solution

Cetyltrimethylammonium bromide (CTAB)	10.0	g
NaCl	0.7	g
Distilled water	100.0	ml

2.3.4 For the preparation of protoplasts from *Streptomyces* (Chater *et al*., 1982)

P-buffer

Sucrose	103.0	g
K_2SO_4	0.25	g
$MgCl_2$ x 6 H_2O	2.02	g
Trace element solution	2.0	ml
Distilled water to	800.0	ml
Dispense in 80 ml aliquots and autoclave. Before use, add to each flask in order:		
KH_2PO_4 (0.5%)	1.0	ml
$CaCl_2$ x 2 H_20 (3.68%)	10.0	ml
TES buffer (5.73%, adjusted to pH 7.2)	10.0	ml

2.3.5 For the transformation of *Streptomyces* (Babcock and Kendrick, 1988)

T-buffer

Sucrose	(10.3%)	25.0	ml
K_2SO_4 (2.5%)		1.0	ml
H_2O		75.0	ml
Trace element solution		0.2	ml
$CaCl_2$ x 2 H_20 (0.25 M)		0.2	ml

2.3.6 For agarose gel electrophoresis of DNA fragments

DNA loading buffer (Invitrogen, Karlsruhe)

EDTA pH 8	100.0	mM
Glycerol	43.0	%
Bromophenol blue	0.5	%
Xylene cyanol 0.5%	0.5	%

TAE (50 x) (Sambrook and Russel, 2001)

Tris	242.0	g/l
EDTA	0.4	g/l
Glacial acetic acid	57.1	ml

2.3.7 For harvesting and sonification of cells (Ahlert *et al*., 1997)

Cell washing buffer

Tris-HCl pH 7.5	25.0	mM

Cracking buffer

Tris-HCl pH 7.5	25.0	mM
Dithiothreitol (DTT)	1.0	mM
Phenylmethylsulfonyl fluoride (PMSF)	0.2	mM

2.3.8 For Western blotting and detection of proteins (Qiagen, Hilden, Germany)

Blotting buffer

Tris	48.0	mM
Glycine	39.0	mM
Methanol	20.0	%

TBS buffer

Tris-HCl pH 6.8	10.0	mM
NaCl	150.0	mM

TBS-T buffer

Tris-HCl pH 6.8	20.0	mM
NaCl	500.0	mM
Tween 20	0.05	%
Triton X-100	0.2	%

Blocking buffer

3.0% (w/v) BSA in TBS buffer

2.3.9 For SDS polyacrylamide gel electrophoresis (Laemmli, 1970)

Collecting gel (Focus gel)

Acrylamide 30% (29:1)	1.25	ml
Tris buffer 0.5 M pH 6.8	1.9	ml
SDS 10%	75.0	µl
APS 10%	75.0	µl
TEMED	10.0	µl
Distilled water	3.5	ml

Separation gel

Acrylamide 30% (29:1)	6.1	ml
Tris buffer 1.5 M pH 8.8	3.75	ml
SDS 10%	150.0	µl
APS 10%	150.0	µl
TEMED	10.0	µl
Distilled water	5.0	ml

Running buffer (10 x)

Tris-HCl pH 8.3	30.0	g/l
Glycine	143.0	g/l
SDS	10.0	g/l

Sample buffer

Tris	50.0	mM
Glycerol	12.0%	w/v
SDS	4.0%	w/v
2-mercaptoethanol	2.0%	v/v
Serva Blue G	0.01%	w/v
Adjusted to pH 6.8 with 1 M HCl		

Gel staining solution

Coomassie Brilliant Blue R250	1.5	g/l
Methanol	450.0	ml/l
Acetic acid	100.0	ml/l

Gel destaining solution

Methanol	250.0	ml/l
Acetic acid	100.0	ml/l
Distilled water	650.0	ml/l

2.3.10 For visualization of spots on TLC

Cer reagent (Drepper *et al*., 1996)

Phosphomolybdic acid x H_2O	25.0	g/l
Cer (IV)-sulfate x 4 H_2O	10.0	g/l
H_2SO_4 (conc.)	60.0	ml/l
H_2O	940.0	ml/l

Ehrlich reagent (Cooper, 1978)

p-Dimethylaminobenzaldehyde	1.0	g
Ethanol 95%	95.0	ml
Concentrated HCl	20.0	ml

Ninhydrin solution

Ninhydrin	1.0	g
Ethanol 95%	1000.0	ml

2.3.11 X-Gal solution

X-gal	0.2	g% in DMF

2.3.12 IPTG stock solution

IPTG	0.1	M in sterilized water

2.4 Bacterial strains, vectors and recombinant plasmids

2.4.1 Bacterial strains

Tab. 2.1 List of strains used in this study

Strain	Relevant markers/Product	Source/Reference
1- *Escherichia coli*		
E. coli DH5α	F^-, Φ80Δ *lac*Z ΔM15, *endA1*, *recA1, hsdR17* (r^-, m^+), *gyrA96*, *thi*, *relA1 supE44*, *deoR*	Hanahan, 1983
E. coli ET12567	F^-, *dam13*, *dcm*6, *hsdM*, *hsdR*, *recF143*, *galT22*, *ara14*, *lacY1*, *hisG4*	MacNeil *et al*., 1992
E. coli JM109(DE3)	F^-, *recA1*, *supE44*, *endA1*, *hsdR17*, *gyrA96*, *thi*Δ(lac-proAB), *lacUV5-T7* gene 1	Novagen, Darmstadt
E. coli BL21 (DE3)	F^-, *ompT,* λ-prophage-T7-polymerase *gal, dcm*	Studier *et al*., 1990
E. coli BL21 (DE3) Rosetta	F^-, *ompT,* λ-prophage-T7-polymerase *gal, dcm*, pRARE, *cat*	Novagen, Darmstadt
2- *Bacillus*		
B. circulans	butirosin producer	ATCC 21558
B. subtilis	AGAs sensitive strain	DSM 618
3- *Micromonospora*		
M. echinospora	gentamicin producer	DSM 34036
M. olivasterospora	fortimicin B producer	DSM 43868
3- *Streptomyces*		
S. lividans TK23	*spc-1*, actinorhodin	John Innes Institute, Norwich, UK
S. kanamyceticus	kanamycin producer	DSM 40500
S. fradiae	neomycin producer	DSM 40063
S. rimosus subsp. *paromomycinus*	paromomycin producer	NRRL 2455
S. hygroscopicus subsp. *hygroscopicus*	hygromycin B producer	DSM 40578

Strain	Relevant markers/Product	Source/Reference
S. tenebrarius	apramycin/tobramycin producer	DSM 40477
S. ribosidificus	ribostamycin producer	NRRL B-11466
S. tenjimariensis	istamycin producer	ATCC 31603
S. lividus	lividomycin producer	ATCC 21178
St. hindustanus	apramycin/tobramycin producer	DSM 44523

2.4.2 Vectors

Tab. 2.2 Vectors used in this study

Vector	Description/Scope of use	Reference/Source
pUC18	*bla*, *lac*Z-α / cloning in *E.coli*	Vieira and Messing, 1982
pUCPU21	*bla*, *lac*Z-α / cloning in *E.coli*	U. Wehmeier, Wuppertal
pUCBM21	*bla*, *lac*Z-α / cloning in *E. coli*	Roche-Mannheim
pUWL201	*bla, lac*Z-α, *tsr* / expression in *Streptomyces*	Doumith *et al*., 2001
pUWL218	*bla, lac*Z-α, *tsr* / cloning in *Streptomyecs*	Wehmeier, 1995
pUWL201PW	*bla, lac*Z-α, *tsr* / expression in *Streptomyecs*	Doumith *et al*., 2001
pET11a	*bla*, *lacI*T7 / expression in *E. coli*	Novagen, 1992
pET16b	*bla*, *lacI*T7-Φ10 *lacO* s10 His-tag / expression in *E. coli*	Novagen, Madison, USA
pOJ436	*aac*(3)IV, apramycin resistant, *ori* ColE1 oriT RK2, *attP* øC31, *int* øC31/ cosmid library	Bierman *et al*., 1992

2.4.3 Recombinant plasmids

Tab. 2.3 Newly constructed plasmids created in this study

Plasmid	Target gene (Source)	Description	Receptor strains
pKAW5	*neoB* *(S. fradiae)*	1.356 kb *Nco*I/*Bam*HI PCR fragment (primer KA1-2 & KA3) in pUCBM21	*E. coli* DH5α
pNPW1	*aacC*7 (*S. rimosus* subsp.) *paromomycinus*)	0.88 kb PCR fragment (primer -1 (*Nco*I) & primer-2 (pact-1)) in pUC18 (*Sma*I)	*E. coli* DH5α
pKAC2	*parC* (*S. rimosus* subsp. *paromomycinus*)	0.35 kb PCR fragment (primer HC-F & HC-R) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pKAA1	*parA* (*S. rimosus* subsp.) *paromomycinus*)	0.55 kb PCR fragment (primer HparA-F & HparA-R) in pUC18 (*Sma*I)	*E. coli* DH5α
pKMR1	*kmr* (*S. kanamyceticus*)	0.567 kb *Nde*I/*Bgl*II PCR fragment (primer kmr-1 and kmr-2) in pUCPU21 (*Nde*I/*Bam*HI)	*E. coli* DH5α
pLIVS2	*livS* (*S. lividus*)	0.5 kb PCR fragment (primer HS1-F HS1-R) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pKANS1	*kanS1* (*S. kanamyceticus*)	1.3 kb *Nde*I/*Bam*HI PCR fragment (primer kanS1-F and kanS-2) in pUCPU21 (*Nde*I/*Bam*HI)	*E. coli* DH5α
pKS16b1	*kanS1* (*S. kanamyceticus*)	1.3 kb *Nde*I/*Bam*HI fragment from pKANS1 in pET16b (*Nde*I/*Bam*HI)	*E. coli* JM109 (DE3) *E. coli* DH5α
pS1W41	*kanS1* (*S. kanamyceticus*)	1.3 kb *Nde*I/*Bam*HI fragment from pKANS1 in pUWL201PW (*Nde*I/*Bam*HI)	*E. coli* DH5α *E. coli* ET12567 *S. lividans* TK23
pKANC	*kanC* (*S. kanamyceticus*)	1.1 kb *Nde*I/*Bam*HI PCR fragment (primer kanC-F and kanC-R) in pUCPU21 (*Nde*I/*Bam*HI)	*E. coli* DH5α
pKC16b1	*kanC* (*S. kanamyceticus*)	1.1 kb *Nde*I/*Bam*HI fragment from pKANC in pET16b (*Nde*I/*Bam*HI)	*E. coli* DH5α *E. coli* BL21 (DE3) *E. coli* JM109 (DE3)
pKANE	*kanE* (*S. kanamyceticus*)	1.0 kb *Nde*I/*Bam*HI PCR fragment (primer kanE-F and kanE-R) in pUCPU21 (*Nde*I/*Bam*HI)	*E. coli* DH5α
pKE16b1	*kanE* (*S. kanamyceticus*)	1.0 kb *Nde*I/*Bam*HI from pKANE in pET16b (*Nde*I/*Bam*HI)	*E. coli* DH5α *E. coli* Rosetta

Plasmid	Target gene (Source)	Description	Receiver strains
pKE11a-4	*kanE* (*S. kanamyceticus*)	1.0 kb *Nde*I/*Bam*HI from pKANE in pET11a (*Nde*I/*Bam*HI)	*E. coli* DH5α *E. coli* Rosetta
pKEW1-2	*kanE* (*S. kanamyceticus*)	1.0 kb *Nde*I/*Bam*HI fragment from pKANE in pUWL201PW (*Nde*I/*Bam*HI)	*E. coli* DH5α *E. coli* ET12567 *S. lividans* TK23
pKANS2	*kanS2* (*S. kanamyceticus*)	1.3 kb *Nde*I/*Bam*HI PCR fragment (primer kanS2-F and kanS2-R) in pUCPU21 (*Nde*I/*Bam*HI)	*E. coli* DH5α
pK2S16b1	*kanS2* (*S. kanamyceticus*)	1.3 kb *Nde*I/*Bam*HI fragment from pKANS2 in pET16b (*Nde*I/*Bam*HI)	*E. coli* DH5α *E. coli* JM109 (DE3)
pS2W12	*kanS2* (*S. kanamyceticus*)	1.3 kb *Nde*I/*Bam*HI fragment from pKANS2 in pUWL201PW (*Nde*I/*Bam*HI)	*E. coli* DH5α *E. coli* ET12567 *S. lividans* TK23
pGRM1	*grm* (*M. echinospora*)	0.83 kb *Nde*I/*Bam*HI PCR fragment (primer grm-1 & grm-2) in pUCPU21 (*Nde*I/*Bam*HI)	*E. coli* DH5α
pFMO3	*genI* (*M. echinospora*)	0.9 kb PCR fragment (primer Fm3O-F & Fm3O-R) in pUC18 (*Sma*I)	*E. coli* DH5α
pHYG3	HM-B *pht* (*S. hygroscopicus* subsp. *hygroscopicus*)	0.83 kb PCR fragment (primer hyg-1 & hyg-2) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pHYGC92	*hygC* (*S. hygroscopicus* subsp. *hygroscopicus*)	0.3 kb PCR fragment (primer HC-1 HC-2) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pHYGC-2	*hygC* (*S. hygroscopicus* subsp. *hygroscopicus*)	1.2 kb PCR fragment (primer HygC-F & HygC-R) in pUCPU21 (*Nde*I/*Bam*HI)	*E. coli* DH5α
pBC9	*btrC* (*B. circulans*)	1.1 kb PCR fragment (primer BC-1 BC-2) in pUCPU21 (*Sma*I)	*E. coli* DH5α
pBS1	*btrS* (*B. circulans*)	0.7 kb PCR fragment (primer btrS-1 & btrS-3) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pBB1	*btrB* (*B. circulans*)	1.3 kb *Nco*I/*Bam*HI PCR fragment (primer btrB-1 & btrB-2) in pUCBM21	*E. coli* DH5α
pFMS1419	*fms14* (*M. olivasterospora*)	1.5 kb PCR fragment (primer Fms14-1- & Fms14-2) in pUCPU21 (*Eco*RV)	*E. coli* DH5α

Plasmid	Target gene (Source)	Description	Receptor strains
pFMRO9	*fmrO* (*M. olivasterospora*)	0.9 kb PCR fragment (primer FmrO-1- & FmrO-2) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pRC1	*ribC* (*S. ribosidificus*)	0.3 kb PCR fragment (primer HC-1 HC-2) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pRPH1	*rph* (*S. ribosidificus*)	0.8 kb PCR fragment (primer SribAPH-1 SribAPH-2) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pAPRA5	*aprA* (*S. tenebrarius*)	1.12 kb PCR fragment (primerPHaprA-3 PhaprA-4) in pUCPU21 (*Hin*cII)	*E. coli* DH5α
pKAMB1	*kamB* (*S. tenebrarius*)	0.5 kb *Nde*I/*Bam*HI PCR fragment (primer PhkamB-1 & PhkamB-2) in pUCPU21	*E. coli* DH5α
pTOBS2	*tobS2* (*S. tenebrarius*)	0.5 kb PCR fragment (primer HS1-F HS1-R) in pUCPU21 (*Eco*RV)	*E. coli* DH5α
pKAMA3	*kamA* (*S. tenjimariensis*)	0.5 kb *Nde*I/*Bam*HI PCR fragment (primer PhkamB-1 & PhkamB-2) in pUCPU21	*E. coli* DH5α
pLIVB6	*livB* (*S. lividus*)	0.55 kb PCR fragment (primer HB-1 HB-2) in pUCPU21 (*Eco*RV)	*E. coli* DH5α

2.5 Oligonucleotides

Tab. 2.4 Homologous primers and PCR conditions

Primer designation	Target gene	Primer sequences [1]	Annealing temperature (T), Annealing time (t)
Kmr-1 (*Nde*I)	*kmr*	5' AGAGGATGGCATATGTCGCAGTCC 3'	50°C, 1 min
Kmr-2 (*Bgl*II)		5' GCCCGTCGTCAGATCTTCGTGA 3'	
KanS-F (*Nde*I)	*kanS1*	5' GAGGTATGTCATATGCCCCTGCAA 3'	45°C, 1 min
KanS-R (*Bam*HI)		5' TGTGACGGATCCGGCTCAAG 3'	
KanS2-F (*Nde*I)	*kanS2*	5' CCGACAAGGAGTCCATATGAGCAAG 3'	45°C, 1 min
KanS2-R (*Bam*HI)		5' TGCGGGATCCGTCAGGTCA 3'	
KanC-F (*Nde*I)	*kanC*	5' ATTGGGGACAGCATATGCAGGTCACC 3'	52°C, 1 min
KanC-R (*Bam*HI)		5' CGCGCACGGGATCCCTCACCC 3'	
KanE-F (*Nde*I)	*kanE*	5' GGAGCAGACCATATGAAGGCACTCGT 3'	50°C, 1 min
KanE-R (*Bam*HI)		5' CGGACGGATCCTAGTCGCGCA 3'	
KA1-2 (*Nco*I)	*neoB*	5' CGACTTTCCCATGGGGTGAA 3'	50°C, 1 min
KA-3 (*Bam*HI)		5' CGGGATCCGTCGTCCAGCA 3'	
neoQ-F (*Nde*I)	*neoQ*	5' AAAGGCACCATATGAAGCGCCTTC 3'	55°C, 1 min
neoQ-R (*Bam*HI)		5' CGCGGATCCGCTCAGACGT 3'	
Primer-1 (*Nco*I)	*aacC7*	5' CCGTCCATGGACGAAC 3'	48°C, 45 sec
Pact-I		5' CAGGTGATCATCCGCC 3'	
Phkamb-1 (*Nde*I)	*kamB*	5' ACAAGAGCCATATGGAGAAGATCTCGG 3'	53°C, 1 min
Phkamb-2 (*Bam*HI)		5' CCGGGATCCGGCGTCAC 3'	
PhaprA-3	*aprA*	5' ATGCTGCTGCTGACCTGCCG 3'	53°C, 1 min
PhaprA-4		5' TCAGCCCCGCGACCGGT 3'	
PhkamC-1 (*Nde*I)	*kamC*	5' CAAGGACCATATGCAGAAGATCGC 3'	53°C, 1 min
PhkamC-2 (*Bam*HI)		5' CCTCCGGGATCCTCCAGCATC 3'	
Hyg-1 (*Nde*I)	*pht*	5' GAATAGAGGTCCCATATGACACAAGAATCCC 3'	52°C, 45 sec
Hyg-2 (*Bam*HI)		5' GGGCGGATCCGGGCGGC 3'	
HygC-F (*Nde*I)	*hygC*	5' TTTGGGAACATATGGCGATGGACTTACA 3'	52°C, 45 sec
HygC-R (*Bam*HI)		5' GACGGGATCCCGGTCATGACGGA 3'	
grm-1 (*Nde*I)	*grm*	5' TTCGGAGGACCATATGACGACATCTG 3'	50°C, 45 sec
grm-2 (*Bam*HI)		5' GCCATCGGATCCTTCCGGAA 3'	
Fm3O-F	*genI*	5' GTGGCAGAGGCGGACGGAAC 3'	50°C, 45 sec
Fm3O-R		5' TATCCGCCGTTCCGGTCGC 3'	

Primer designation	Target gene	Primer sequences [1]	Annealing temperature (T), Annealing time (t)
Fms14-1 (*Nde*I) Fms14-2 (*Bam*HI)	*fms14*	5' GATGGTAATTCATATGGTTGATGCTGCCCC 3' 5' TCACGCCGGGATCCGTCCT 3'	50°C, 1 min
FmrO-1 (*Nde*I) FmrO-2 (*Bam*HI)	*fmrO*	5' AGGTGGCCCGACATATGCTCGCCGCGG 3' 5' ACGGCCTGCCGGATCCGCCTACCCCG 3'	60°C, 1 min
BC-1 BC-2	*btrC*	5' CATGACGACTAAACAAATTTGTT 3' 5' TACAGCCCTTCCCGGAT 3'	50°C, 1 min
BtrB-1 (*Nco*I) BtrB-2 (*Bgl*II)	*btrB*	5' GAACGATGCCATGGAACAGGAA 3' 5' TTGTTTAGTCGAGATCTTTAACCTCCA 3'	44°C, 1 min
BtrS-1 (*Nde*I) BtrS-3 (*Bam*HI)	*btrS*	5' TTCATATGACCATTCCATTTGACCA 3' 5' AGCAATAGTTGGATCCTTGAAT 3'	44°C, 1 min
BtrM-F BtrM-R	*btrM*	5' GCGGCATGCAGGTTCAAAT 3' 5' AATTCGTGAAACAAAGAGATCGGA 3'	58°C, 1 min
BtrJ-F BtrJ-R	*btrJ*	5' AAGCCCTATGAAGTTCACTCA 3' 5' TTCATGAGTTAATGAACAGCC 3'	50°C, 1 min
BtrP-F BtrP-R	*btrP*	5' GAGATAGCATGCGATTGATTT 3' 5' TCATATCGAGCGTGCAGTC 3'	51°C, 1 min
ribN-F (*Nde*I) ribN-R (*Bam*HI)	*ribN*	5' GAGAAGGCATATGCCTACATC 3' 5' GTTGTTGGGATCCTGGTTG 3'	40°C, 45 sec
SribAPH-1 (*Nco*I) SribAPH-2 (*Bam*HI)	*rph*	5' ATGCCGACCATGGAAAGCAC 3' 5' GTCTCCGTGGATCCAGAAGAACTCG 3'	57°C, 1 min
nbrB-1 (*Nde*I) nbrB-2 (*Bam*HI)	*nbrB*	5' GAGGACCCCCATATGCCGCACCCG 3' 5' TGCCCCGGGATCCTCAGGCGTT 3'	56°C, 45 sec
StenjkamA-1 (*Nde*I) StenjkamA-2 (*Bam*HI)	*kamA*	5' CAAGGACCATATGCGCAAGGTCGC 3' 5' AAGAGCCTGGATCCGCGGCTCAG 3'	55°C, 1 min
aac4-F aac4-R	*aacC4*	5' GTGCAATACGAATGGCGAAAAGC 3' 5' TGAGCTCAGCCAATCGACTGG -3'	50°C, 1 min
aadA-F aadA-R	*aadA*	5' AACATCATGAGGGAAGCGGTGAT 3' 5' AGACATTCTTTGCCGACTACC 3'	55°C, 1 min

[1] = the recognition sequence for restriction endonucleases are underlined.

Tab. 2.5 Newly developed heterologous primers

Target genes (examples)	Designation and sequence of the heterologous primer	Expected size of PCR product (kb)
2-deoxy-*scyllo*-inosose synthase (*kanC*)	HC- F: 5' CTSTCGATSAAGCAGGCSGTCAACTT 3' HC- R: 5' TGSCCSATSGTGTSGCCGTA 3'	0.35
2-deoxy-*scyllo*-inosose aminotransferase (*kanS1*)	HS1- F: 5' TGYGGSGTCGGYGCSGGIGAC 3' HS1- R: 5' AKGCASCGGTTGCTGCCCATCA 3'	0.5
Aminotransferase (*kanS2*)	HS2- F: 5' GGCGGCATCCCCTTCCCSAACACC 3' HS2- R: 5' GCCCTCGGCRAGGATCGCCIISAC 3'	0.8
Aminotransferase (*tobS2*)	HtobS2- F: 5' TCGGCGGCATCCCCTTCCCGAAC 3' HtobS2- R: 5' CGGCGTGAAGCCGCCGATCTCG 3'	0.7
6'-hexosamine--aminotransferase (*kanB*)	HB- F: 5' CGGCTACCACGGCTACGAC 3' HB- R: 5' CGCCGGTACGAACTGGAAG 3'	0.55
Protein of unknown function (*parA*- like)	HparA- F: 5' CCCTGGATCGTSCGCAGTTCG 3' HparA- R: 5' GAACATCACSGCSGCGTGCTC 3'	0.85
6'-C-methyltransferase (*forK)*	HKL1- F: 5' GGMCGKCACGTGCCGATCGG 3' HKL.1- R: 5' ACGTAGTGCTCGATCTCGTCGACCAC 3'	0.75

The following letters in the primer sequences indicate: I = deoxyinosine; K = (T or G); M = (A or C); R = (G or A); S = (C or G); Y = (C or T).

2.6 Probes and cosmid banks

Tab. 2.6 Probes used for screening of the different cosmid banks

Strain	Cosmid bank (CB)	Template gene and size of the probe (kb)
S. kanamyceticus DSM 40500	2460	*kmr* (16S ribosomal RNA methylase), 0.5
S. fradiae DSM 40063	2458	*neoB* (aminotransferase), 1.2
S. rimosus subsp. *paromomycinus* NRRL 2455	2459	*aacC*7 (aminocyclitol N-acetyltransferase), 0.88 *parC* (2-deoxy-*scyllo*-inosose synthase), 0.35 *parA* (protein of unknown function), 0.8
S. tenebrarius DSM 40477	3730	*kamB* (16S ribosomal RNA methylase), 0.5 *aprA* (apramycin biosynthesis gene), 1.1 *tobS2* (putative 3''-aminotransferase), 0.5
S. hygroscopicus subsp. *hygroscopicus* DSM 40578	2720	*pht* (hygromycin B 7'-phosphotransferase), 1.1 *hygC* (2-deoxy-*scyllo*-inosose synthase), 0.35
M. echinospora DSM 43036	2700	*grm* (16S ribosomal RNA methylase), 0.8 *fmrO*3'-like (protein of unkown function), 0.9
M. olivasterospora DSM 43868	3664	*fms14* (formimidoyl fortimicin A synthase), 1.5 *fmrO* (16S ribosomal RNA methylase), 0.9
B. circulans ATCC 21558	2695 2695.2	*btrC* (2-deoxy-*scyllo*-insose synthase), 1.1 *btrS* (2-deoxy-*scyllo*-inosose aminotransferase), 0.7
S. ribosidificus NRRL B-11466	3753	*ribC* (2-deoxy-*scyllo*-insose synthase), 0.35 *rph* (aminoglycoside-3'-phosphotransferase), 0.8
S. lividus ATCC 21178	3759	*livB* (aminotransferase), 0.55 *livS* (2-deoxy-*scyllo*-inosose aminotransferase), 0.5
S. tenjimariensis ATCC 31603	3758	*kamA* (16S ribosomal RNA methylase), 0.5 *istC* (2-deoxy-*scyllo*-inosose synthase), 1.0
*St. hindustanu*s DSM 44523	2730	*kamB* (16S ribosomal RNA methylase), 0.5 *aprA* (apramycin biosynthesis gene), 1.1 *tobS2* (putative 3''-aminotransferase), 0.5

Tab. 2.7 Cosmids positively screened and used for sequence determinations

Strain	Cosmid banks (CB)	Probe/total number of positively screened cosmids	Selected cosmids for sequencing
S. kanamyceticus DSM 40500	2460	*kmr* / 4	SkaJ19 & SkaJ15
S. fradiae DSM 40063	2458	*neoB* / 10	SfrF04 & SfrA10
S. rimosus subsp. *paromomycinus* NRRL 2455	2459	*aacC7* / 6 *parC* / 6 *parA* / 9	SriA13 SriG07 & SriD03 SriL03
S. tenebrarius DSM 40477	3730	*aprA* / 7 *kamB* / 15 *tobS2* / 11	SteO08 SteO08 SteM07 & SteK17
S. hygroscopicus subsp. *hygroscopicus* DSM 40578	2720	*hyg-pht* / 18	ShyG17
M. echinospora DSM 43036	2700	*grm* /13 *genI* / 9	MecP21, MecE04 & MecG05 MecO02
M. olivasterospora DSM 43868	3664	*fms14* / 5 *fmrO* / 6	MolI14 MolJ05
B. circulans ATCC 21558	2695 2695-2	*btrC* / 19 *btrM* / 5 *btrS* / 2 *btrS* / 10	BciH11 - - -
S. ribosidificus NRRL B-11466	3753	*ribC* / 5 *rph* / 8	SribP10 SribL03
S. lividus ATCC 21178	3759	*livS* / 7 *livB* / 11	SliD01 SliD01
S. tenjimariensis ATCC 31603	3758	*kamA* / 20 *istC* / 5	StenF24 StenO22
*St. hindustanu*s	2730	*kamB* & *aprA* / 15	ShinN01

2.7 Antibiotics

The antibiotics used in this work were normally supplemented to the corresponding growth media at the concentration mentioned in Tab. 2.8.

Tab. 2.8 Antibiotics

Antibiotic	Final concentration in media (µg/ml)	Solvent for dissolving the antibiotic
Ampicillin	100	sterilized water
Apramycin	50	sterilized water
Chloramphenicol	25	ethanol
Gentamicin	5	sterilized water
Hygromycin B	5	sterilized water
Kanamycin	50	sterilized water
Lividomycin	5	sterilized water
Nystatin	50	0.1 N NaOH *
Ribostamycin	5	sterilized water
Spectinomycin	100	sterilized water
Tetracycline	5	ethanol
Thiostrepton	25	DMSO
Tobramycin	50	sterilized water

* = dissolved in a small volume of sterile 0.1 N NaOH and made up to volume with sterilized water

2.8 Growth and maintenance of bacterial strains

2.8.1 Growth and maintenance of *E. coli* and *B.* strains

E. coli and *B.* sp. strains were generally cultivated at 37°C on LB plates, in LB medium or in 2x TY liquid media. Strains harboring plasmids or cosmids were grown on plates or in liquid media supplemented with the appropriate antibiotics. For stock cultures, cells were concentrated and suspended in slant medium 50:50 and stored at - 20°C (Sect. 2.2.1).

2.8.2 Growth and maintenance of *Streptomyces* sp. and *Micromonospora* sp. strains

Most *Streptomyces* sp. strains were grown on solid media such as SMA and SPMR plates or in liquid media such as YEME and TSB whereas *Micromonospora* sp. strains were grown on M65 plates or in TSB liquid media. Both *Streptomyces* sp. and *Micromonospora* sp. strains were incubated at 28 - 30°C for 2 - 7 days. Testing of aminoglycoside antibiotic production by the producing strains was performed through inoculation of each strain in the aminoglycoside antibiotic production medium and incubation for 3 - 7 days at the appropriate temperature (28 -

30°C for *Streptomyces* sp. and *Micromonospora* sp. and 37°C for *B. circulans*). 300 - 400 µl of the supernatants was taken and put into wells made in LB plated previously seeded with 10^7 spores of *B. subtilis* followed by incubation of plates at 37°C for 12 hrs. The resulted inhibition zones were recorded and annotated. Strains harboring plasmids were cultivated in media supplemented with the appropriate antibiotics (Tab. 2.8). Spore suspensions were generated according to Kieser *et al*. (2000) by separating spores from a well sporulated plate by flooding it with 6 ml of 20% (v/v) glycerol and scraping the spores off the substrate mycelia with a sterilized inoculating loop. The spore suspension was filtered through sterile non-absorbent cotton wool to remove mycelial fragments. The spore suspension was then stored at - 80°C.

2.9 In vitro manipulation of DNA

2.9.1 Preparation of plasmid DNA from *E. coli*

Preparation of plasmid DNA from *E. coli* was performed according to two methods: I- Plasmid DNA for DNA sequencing was prepared using the QIA prep Spin Miniprep kit according to the manufacturer's specification. II- The alkaline-SDS method was used for rapid screening of transformants (Birnboim and Doly, 1979).

2.9.2 Preparation of cosmid DNA from *E. coli*

Preparation of cosmid DNA from *E. coli* was carried out using the method of Birnboim and Doly (1979) with some modifications. After addition of solution III and centrifugation for 15 min, the supernatant was extracted with an equal volume of chloroform/isoamylalcohol (24:1). The aqueous phase was further extracted using an equal volume with phenol/chloroform (1:1) and transferred to a new tube where the DNA was precipitated by the addition of 400 µl isopropanol with subsequent centrifugation for 30 min at 13,000 rpm. The supernatant was discarded and the DNA was washed in 70% cold ethanol, centrifuged at 13,000 for 15 min and dried. DNA was then dissolved in 100 µl TE buffer containing RNase 100 µg/ml.

2.9.3 Extraction of genomic DNA from actinomycetes and bacilli

Chromosomal DNA was prepared according to two methods: I- Genomic DNAs of *S. kanamyceticus*, *S. fradiae, S. rimosus* and *M. echinospora* were extracted according to the method of Pospiech and Neumann (1995) with minor modifications. Strains were inoculated in 10 ml TSB or YEME liquid media in a 10 x 22 cm glass tubes with a short metal spring. The

culture was grown at 28°C on a shaker (180 rpm) for 3 - 4 days to the late logarithmic phase. The mycelia was then harvested by centrifugation at 4,000 rpm for 7 min and washed twice with 10.3% sucrose. Approximately 0.5 g of cells was then resuspended in 10 ml SET buffer with 1 mg/ml lysozyme and incubated for 1 - 2 hrs at 37°C. Furthermore, 1/10 volume SDS 10% and proteinase K to the final concentration of 0.5 mg/ml were added and incubated at 55°C for 1 - 2 hrs with frequent gentle inversion. 1/3 volume 5 M NaCl was added and an equal volume phenol/chloroform was added and incubated at room temperature for 20 min with gentle inversion. The mixture was then centrifuged at 4,000 rpm for 10 min and the aqueous phase was further extracted with an equal volume of chloroform/isoamyl alcohol (24:1), incubated at room temperature for 20 min with gentle inversion, centrifuged at 4,000 rpm for 10 min. The DNA was precipitated by the addition of an equal volume of isopropanol, centrifuged at 4,000 rpm for 5 min. DNA was then washed using 70% ice cold ethanol, dried and finally dissolved in 200 - 400 µl TE buffer with RNase 100 µg/ml.

II- Genomic DNAs of *S. hygroscopicus*, *M. olivasterospora, B. circulans, S. tenebrarius*, *S. ribosidificus*, *S. tenjimariensis*, *S. lividus* and *St. hindustanus* were prepared according to the modified cetyltrimethyl ammonium bromide (CTAB) procedure (Mehling *et al*., 1995a).

2.9.4 Restriction endonuclease digestion of DNA

DNA restriction was performed using approximately 2 - 4 units endonuclease enzyme per 1 µg DNA in the reaction mixture. The reaction volume was made up to 20 or 30 µl with sterile water together with the recommended enzyme buffer. The mixture was incubated 1 - 2 hrs at the appropriate temperature according to the recommendations of the manufacturer.

2.9.5 Blunt end generation using the Klenow fragment of DNA polymerase I

PCR products obtained with the *Taq* DNA polymerase were subjected to treatment by Klenow polymerase in order to generate blunt ends prior to ligation. Treatment by Klenow polymerase was performed according to manufacturer's recommendations. The reaction was made in a total volume of 100 µl at 37°C for 30 min using 1 - 2 U of Klenow fragment. For the 5' phosphorylation of both ends of the PCR product, the mixture was treated with 1 U T4-polynucleotide kinase and 5 µl 10 mM ATP and further incubated for 30 min at 37°C. The mixture was then heat inactivated at 75°C for 20 min followed by addition of equal volume of 5

M ammonium acetate. DNA was precipitated with 2.5 fold volume ice-cold absolute ethanol, washed in 70% ethanol and dissolved in an appropriate amount of TE buffer.

2.9.6 Recovery of DNA fragments from agarose gels

DNA fragments were subjected to agarose gel electrophoresis. The DNA fragments less than 10 kb were excised and subsequently purified using QIA quick Gel Extraction Kit (Qiagen, Hilden) according to the manufacture's specifications. On the other hand, DNA fragments bigger than 10 kb were excised from low melting point agarose, melted at 60°C, extracted with phenol/chloroform. The aqueous phase was transferred to a new tube, mixed with equal volume of 5 M ammonium acetate. DNA was precipitated with 2.5 fold volume ice-cold absolute ethanol, washed in 70% ethanol and dissolved in an appropriate amount of TE buffer.

2.9.7 Removal of phosphate groups from DNA

Linearized DNA was treated with 1 - 2 U calf intestine alkaline phosphatase (CIAP) and incubated at 37°C for 20 - 30 min in order to remove 5'-phosphate groups. CIAP was inactivated by heat inactivation at 70°C for 15 min.

2.9.8 Ligation of DNA fragments

Ligation reactions were performed in 20 or 30 µl total volume (4:1 molar ratio insert to vector) with ligation buffer and T4-DNA ligase at 16°C. Furthermore, blunt-end ligation was carried out using 2 U T4-DNA ligase for about 12 hrs, whereas ligation of fragments with cohesive ends were incubated using 1 U T4-DNA ligase for at least 2 - 4 hrs.

2.9.9 Quantitation of DNA

DNA concentrations were quantified photometrically at a wavelength of 260 nm as described by Sambrook and Russell (2001). An OD_{260} of 1.0 is equivalent to approximate 50 µg/ml DNA. The OD of the DNA solution was also measured at 280 nm to check for contamination with phenol or proteins. The ratio of OD_{260}/OD_{280} of pure DNA solutions should range from 1.8 - 2.0.

2.9.10 Construction of cosmid banks

This step was basically done by Combinature Biopharm company (Berlin, Germany) using the protocols described in the literature (Beye *et al.*, 1998; Burgtorf *et al.*, 1998). In short:

homogenized bacterial cultures were embedded in 0.5% low melting point agarose and incubated with lysozyme (2 mg/ml 14 hrs RT) and proteinase K (1 mg/ml 24 hrs 50°C), successively. Embedded DNA was partially digested with *Sau*3A1, the DNA extracted using gelase and dephosphorylated. Ligation-reactions were set up with 750 ng digested vector pOJ436 (Bierman *et al*., 1992), desalted, packged and transfected into *E. coli* DH5α. Colonies were robotically gridded into microtiterplates and transferred onto nylon membranes (Amersham Biosciences, Freiburg). After the colonies were grown, the membranes were processed according to Nizetic *et al*. (1991) and non-radioactively hybridized using standard procedures (Roche Diagnostics).

2.9.11 Preparation and screening of the cosmid banks

This step was also carried out by Combinature Biopharm (Berlin, Germany) using the protocol described by Trefzer *et al*. (2002) under use of roboting devices for the handling of large sets of cosmid libraries at the laboratories of Combinature Biopharm. Cosmids were considered positively tested when giving signals with two different homologous probes or when showing-up in two independent rounds of screening with the same probe. The positively screened cosmids were confirmed by PCR, mapped by restriction and ordered for overlapping genomic areas in order to select the appropriate cosmids for sequencing (Tab. 2.7). Sequence determination for the creation of publishing quality data sets was carried out using customer service via first shotgun-cloning and sequencing of whole cosmids and subsequent primer-walking strategies in order to create single contigs for a given genomic region from overlapping cosmids at GATC Biotech company (Konstanz, Germany), Seqlab (Göttingen, Germany) or at AGOWA company (Berlin, Germany).

2.9.12 DNA sequencing

DNA sequencing was performed according to the method described by Sanger *et al*. (1977) using the Thermosequence Cycle-Sequencing kit (Amersham Bioscience, Freiburg) and primers labelled with Cy5 according to manufacturer's specifications on an automatic sequencer (A.L.F. express; Amersham Bioscience, Freiburg).

2.10 Preparation of competent *E. coli* cells

E. coli competent cells were prepared according to the modified Hanahan (1983) method. The cells were incubated for 12 - 14 hrs at 300 rpm at 37°C in 3 ml LB medium. 200 µl from this

culture were inoculated in 20 ml SOB medium and incubated at 37°C for 2 - 3 hrs at 300 rpm until the OD_{600} reached 0.5 - 0.6. All the following steps were carried out on ice or at 4°C. The cells were spun for 5 min at 3,500 rpm, resuspended in 5 ml ice-cold TMF-1 buffer and kept on ice for 30 min. The cells were harvested again by centrifugation at 3,500 rpm for 10 min and resuspended in TMF-1 buffer containing 20% glycerol. Cells were left on ice until use or 200 µl aliquots were made and stored at - 80°C until use.

2.11 Transformation of *E. coli* strains

E. coli cells were transformed according to Sambrook and Russell (2001). In general 10 µl of a ligation mixture containing 0.01 - 1.0 µg DNA were added to 200 µl competent cells and kept on ice for 30 min. Following a heat shock (90 sec, 42°C), cells were regenerated in 800 µl SOC or LB medium for 1 hr at 37°C, 300 rpm then the cell suspension was plated out on LB agar plates containing the appropriate antibiotic and/or X-Gal as a selective medium.

2.12 Preparation of protoplasts from *Streptomyces* strains

Protoplasts were prepared according to the method of Hopwood and Wright (1978). Mycelium were grown in 30 ml of YEME or TSB supplemented with PEG 8000 and 0.5% glycine in 250 ml buffled flasks for 16 hrs at 30°C at 150 rpm. Media was inoculated with 300 µl of a culture previously grown to the stationary phase (30°C for 36 - 48 hrs). The mycelium was sedimented by centrifugation at 3,000 rpm, washed twice with 10.3% sucrose and suspended in 10 - 15 ml P-buffer containing 1 mg/ml lysozyme. The cell suspension was incubated at 30°C for 30 - 90 min or until the cells were protoplasted. The protoplast suspension was filtered through sterile cotton wool. The protoplasts were sedimented by spinning at 3,000 rpm for 4 min at 4°C and resuspended in 2 ml P-buffer. Protoplasts were stored at -80°C in aliquots of 100 µl or used directly for transformation.

2.13 Transformation of streptomycete protoplasts with plasmid DNA

Protoplasts were transformed according to Babcock and Kendrick (1988). Frozen protoplasts were quickly thawed and centrifuged at 3,000 rpm for 7 min then resuspended in 50 - 100 µl freshly made P-buffer. In general 100 - 200 ng of plasmid DNA and 500 µl of 25% PEG1000 in T-buffer were added to the protoplasts and gently mixed by pipeting up and down three times. After 1,000 µl P-buffer had been added, the cells were plated on predried SPMR plates. Plates

were incubated at 30°C overnight and then overlaid with the appropriate antibiotic. Plates were checked for transformants after 3 - 7 days.

2.14 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out essentially as described by Sambrook and Russell (2001). DNA was separated on 0.7 - 1.5% agarose gels (according to the size of the DNA molecules) containing 0.1 µg/ml ethidium bromide. DNA solution had to be mixed with 1/5 of its volume with DNA loading buffer before loading into the wells. Electrophoresis was performed using TAE buffer (1x) at 5 - 10 V/cm. DNA was then visualized by illumination with long wavelength UV-light (302 nm). DNA fragment size was determined by comparison to conventionally used or commercially available DNA size markers.

2.15 Polymerase chain reaction (PCR)

Amplification of different probes by PCR was performed using 200 - 400 ng of the genomic DNA of each strain as a template and the selected primers for each probe (Tabs. 2.4 & 2.5). Amplifications were performed in a Personal Cycler (Biometra, Göttingen, Germany). Each assay (50 µl) contained 200 ng chromosomal DNA, 100 pmole of each appropriate primer, 0.2 mM dNTPs (Invitrogen, Karlsruhe, Germany), 3 mM $MgCl_2$, 10% DMSO to improve the denaturation of the template DNA and 2 U *Taq* DNA polymerase (Invitrogen, Karlsruhe, Germany). Sterilized mineral oil were layered over the reaction mixture to prevent evaporation.
The following general conditions for the assay were used: 98°C for 5 min, then 30 cycles [95°C for 1 min, annealing temperatures and time according to Tab. 2.4, 72°C for 1 min (normally 1 min for 1 kb)], and 72°C for 5 min (ramping rate 1°C/sec). The PCR products were cloned either by blunt end after their treatment with Klenow and kinase or restricted with the appropriate restriction endonucleases and then cloned into the appropriate vector hydrolyzed with the same restriction enzymes.

2.16 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used for the electrophoresis of proteins in vertical slab gels using the method of Laemmli (1970). Electrophoresis was based on the separation of proteins according to their molecular weight where the proteins were collected first in the focus gel and then the separation was performed in the following separation gel. The final concentration of acrylamide in the focus

gel was 5.5% where polymerization was started with 0.1% APS and 0.01% TEMED. On the other hand, the concentration of acrylamide in the separation gel was 12% and the polymerization was started with 0.1% APS and 0.005% TEMED. Before electrophoresis, an equivalent volume of sample buffer was added to the protein solution and the mixture was heated for 3 - 5 min at 100°C for denaturing the proteins. Electrophoresis started at 80 V for 1 hr after which the voltage was adjusted to a constant 120 V for the rest of the run. Gels were stained in 10 - 15 ml gel staining solution for 2 - 3 hrs or overnight with gentle shaking at 55°C. The staining solution was discarded and the gels were then destained using gel destaining solution for 1 - 2 hrs at 55°C. The protein standard (Biolabs), β-galactosidase from *E. coli* (Mr = 116.3 kDa), phosphorylase-b from rabbit muscle (Mr = 97.18 kDa), bovine serum albumin (Mr = 66.4 kDa), glutamic dehydrogenase from bovine liver (Mr = 55.5 kDa), MBP2 from *E. coli* (Mr = 42.7 kDa), lactate dehydrogenase from porcine muscle (Mr = 36.6 kDa), triosephosphate isomerase from rabbit muscle (Mr = 26.6 kDa) and trypsin inhibitor from soybean (Mr = 20.1 kDa) was applied to the gels in order to determine the molecular weight (Mr) of the proteins.

2.17 Heterologous gene expression

2.17.1 Protein production in *E. coli* using the *T7* RNA polymerase system

Expression of KanC (2-deoxy-*scyllo*-inosose synthase), KanS1 (L-glutamine: 2-deoxy-*scyllo*-inosose aminotransferase), KanE (3-amino-2,3-dideoxy-*scyllo*-inositol 1-dehydrogenase) in *E. coli* were carried out as described by Studier *et al*. (1990). Expression was performed under the control of the *T7* promoter using either of the *E. coli* Rosetta BL21 (DE3), *E. coli* JM109 (DE3) or *E. coli* BL21 (DE3) strains. Single colonies harboring the plasmid pKC16b1 (KanC), pKS16b1 (KanS1) or pKE16b1 (KanE) and the host strain harboring pET16 (empty vector) as a control were grown overnight in 3 ml LB medium containing 100 µg/ml ampicillin [plus 25 µg/ml chloramphenicol in case of *E. coli* Rosetta BL21 (DE3)] at 37°C on a shaker incubator 300 rpm. 200 µl of these precultures were used to inoculate 20 ml fresh LB medium or LB medium supplemented with 1 M sorbitol and 2.5 mM betaine in 250 ml normal flasks and let them grow to an OD_{600} of 0.5 - 0.6. *T7*-RNA polymerase production was induced by the addition of 0.5 mM IPTG. Samples (1 ml) were taken prior to induction and at the following time intervals after the addition of IPTG: 1, 2, 4, 6 hrs and overnight. The cells were centrifuged at 13,000 rpm for 2 - 3 min, washed two times with ice cold 25 mM Tris-HCl, pH 7.5 and kept at - 20°C until use or

resuspended in 50 - 100 µl of cracking buffer for sonification and appropriate amounts were subjected to SDS-PAGE gels as described in Sect. 2.16.

2.17.2 Gene expression in *S. lividans* TK23 under the control of *ermE*-up promoter

The recombinant plasmids pS1W41 (expressing KanS1) and pKEW1-2 (expressing KanE), derivatives of pUWL201PW, were transformed in *S. lividans* TK 23. A single transformant was inoculated into 10 ml TSB medium supplemented with 25 µg/ml thiostrepton. After growth for 72 - 96 hrs at 28°C as preculture, a 0.5 ml sample was inoculated into 50 ml TSB containing 25 µg/ml thiostrepton and incubated for another 2 - 3 days. Cells were harvested by centrifugation and washed twice with ice- cold 25 mM Tris-HCl buffer pH 7.5. Cells were sonified and proteins were analyzed by SDS-PAGE (Sect. 2.16).

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

The *E. coli* or *S. lividans* TK23 cells harboring the expression plasmids were harvested by centrifugation at 13,000 rpm for 2 - 3 min and washed twice with ice-cold 25 mM Tris-HCl buffer pH 7.5. The pellet was resuspended in cracking buffer (Sect. 2.3) in a ratio of 6.0 ml buffer/g for *E. coli* cells and 2 ml/g for *S. lividans* TK23 cells. The cells were disrupted by sonification (3 times, each time for 45 sec following a 30 sec interval at 60 watt). Finally, the cell debris was sedimented by centrifugation (13,500 rpm, 30 min and 4°C) and the cell free extract was then transferred to a fresh cold microfuge tube prior to protein determination.

2.19 Determination of protein concentration

Protein determination in cell-free extract was carried out according to Bradford (1976) using the Protein Assay Kit (Bio-Rad) and BSA as a standard at a wavelength (λ) of 595 nm.

2.20 Western blotting and immuno-detection of proteins

His-tagged proteins produced in *E. coli* were separated by SDS-PAGE using a prestained protein marker. The proteins were transferred from the PAGE to a Hybond™-C Extra membrane (Amersham) using a Semi-Dry Transfer cell (Bio-Rad) at 20 V for 10 min. Western blotting was applied and the method of immuno-detection of proteins was adapted here as recommended by

the provider of BM Chromagenic Western Blotting Kit (Roche-Mannheim) using an Anti-His-tag antibody.

2.21 Thin-layer chromatography (TLC)

Samples of the enzyme reactions were applied on silica thin-layer sheets (Merck) using the mobile phase (methanol/chloroform/ammonia/water = 6/4/2/1). Detection of the spots was carried out with different solutions: (i) Cer reagent for detection (blue spots after heating) of sugars or other polysaccharides (Drepper *et al*., 1996b); (ii) ninhydrin solution for the detection of aminated compounds (i.e. 2-deoxy-*scyllo*-inosamine); (iii) Ehrlich reagent for the detection of α-keto-glutaramate, where it produces a red spot on the TLC (Cooper, 1978).

2.22 High performance liquid chromatography (HPLC)

Samples of enzyme reactions were applied to HPLC (DIONEX Corporation, Chromeleon and PeakNet ® 6 Release 640, 2001) using the column Phenomenex Hypersil 3U NH2 APS (150 x 4.60 mm, 3 micron) and the mobile phase acetonitrile/water with the ratio of 80: 20 v/v. 50 µl of each enzyme reaction was injected into the HPLC using an automated pipetting system. The running time for each sample was 60 min with a flow rate of 1 ml/min. HPLC runs were carried out at room temperature and the substrates were detected by a UV-detector at λ 210 nm. Activities of KanC and KanS1 were determined using HPLC with the previously mentioned conditions and the dilutions made for each assay was described in Sects. 2.23 & 2.24.

2.23 Determination of the activity of KanC

The activity of the 2-deoxy-*scyllo*-inososе synthase (KanC) was determined according to the protocol developed for the AcbC enzyme assay described by Stratmann *et al*. (1999) with minor modifications. The enzyme assay (100 µl) was performed at 30°C for 6 - 12 hrs using 20 µl G-6-P (60 mM), 2 µl NAD (125 mM), 2 µl NaF (200 mM), 2 µl $CoCl_2$ (4 mM), 20 µl phosphate buffer (100 mM, pH 7.5) and 30 µl of the soluble fraction of the cell free extract of KanC (1.5 - 2.5 mg protein/ml). After incubation the reaction was stopped by heating at 95 - 100°C and centrifuged at 13,000 rpm for 2 - 3 min. 4 - 6 µl of the reaction supernatant was analysed by TLC and detected with Cer reagent (Sect. 2.21). 70 µl of the assays were diluted with 50 µl water and applied to the HPLC (Sect. 2.22). Standard 2-deoxy-*scyllo*-inososе was kindly provided via

chemical synthesis from the organic chemistry department, BU Wuppertal (Prof. Dr. H.J. Altenbach and Dr. M. Podeschwa).

2.24 Determination of the activity of KanS1

The activity of the putative bifunctional (L-glutamine: 2-deoxy-*scyllo*-inosose and L-glutamine: 1-keto-2,3-dideoxy-3-amino-*scyllo*-inositol) aminotransferase KanS1 was analyzed using a coupled enzyme assay with the KanC protein:

KanC assay reaction (2.23)	70 µl
100 mM L-glutamine	15 µl
150 mM Tris-HC buffer pH 7.5 with 6 mM PLP	15 µl
KanS1 extract (2.0 mg protein/ml)	30 µl

The reaction was incubated at 37°C for 4 - 6 hrs then the reaction was stopped by heating at 95-100°C and centrifuged at 13,000 rpm for 2 - 3 min. Samples of 4 - 6 µl of the reaction supernatant were analysed by TLC (Sect. 2.21), where ninhydrin solution and Ehrlich reagent were used to detect formation of both 2-deoxy-*scyllo*-inosamine and α-keto-glutaramate, respectively. Samples of 120 µl of the assays were applied to the HPLC (Sect. 2.22).

Furthermore, a spectrophotometric assay was performed to detect the activity of KanS1 using L-alanine as aminodonor. The newly formed pyruvate can be reduced to L-lactate in the presence of NADH and L-lactate dehydrogenase (550 U/mg; Roche-Mannheim). The activity of KanS1 was measured via the decrease in absorption of NADH by time at $\lambda_{340\ nm}$ using the Perkin-ELMER UV/VIS lambda spectrometer. The spectrophotometric assay was performed as follows:

20 mM 2-deoxy-*scyllo*-inosose	50 µl
100 mM L-alanine	50 µl
150 mM Tris-HC buffer pH 7.5 with 6 mM PLP	100 µl
125 mM NADH di-sodium salt	10 µl
L-lactate dehydrogenase (25 mg/2.5 ml, 550 U/mg)	10 µl
KanS1 extract (2.0 mg protein/ml)	50 µl
Water ad.	1000 µl

Similar reactions but in absence of L-alanine as aminodonor or 2-deoxy-*scyllo*-inosose as amino acceptor or using the same protein concentration of cell-free extract of the host strain harboring pET16 plasmid (empty vector) were also performed as control reactions.

Reactions were incubated at RT and the specific enzyme activity (1 U = formation of 1 µmol NAD/min/mg protein) was calculated as follows:

$$C = \frac{\Delta E \times V_1}{\Sigma \times d \times V_2 \times cp \times n} \quad \text{mol/min/mg protein}$$

C = specific enzyme activity (mol/min/mg protein)
ΔE = rate of change in absorbance at $\lambda_{340\,nm}$ (ΔA/Δt in min)
Σ = molar extinction co-efficient of NADH (6.22 x 10^6 mol/cm x 1)
n = molar ratio (mole NADH oxidized per mole pyruvate reduced)
d = width of the cuvette (cm)
V_1 = total volume of the reaction (ml)
V_2 = volume of the probe (ml)
cp = protein concentration of the probe (mg/ml)

2.25 Determination of the activity of KanE

The enzymatic activity of the putative NAD(H): Cyclitol 1-dehydrogenase (KanE) was determined using the coupled assay with KanC and KanS1 protein (Sect. 2.24). The reactions were incubated at 30°C for 12 - 20 hrs in absence and in presence of 3 - 4 mM $ZnCl_2$. Samples of 4 - 6 µl were monitored by TLC using ninhydrin solution for detection. In addition, 30 µl fresh extract of KanS1 (2.0 mg/ml protein) was added and the mixture was further incubated at 37°C for 4 - 6 hrs then samples of 4 - 6 µl were applied to TLC (Sect. 2.21). Furthermore, an attempt was made to measure the KanE activity spectrophotometrically via measurement the increase of absorbance at λ_{340} due to formation of NADH.

A- For TLC analysis the tests were performed as follows:

KanC/KanS1 assay reaction (2.24)	100 µl
NAD (125 mM)	1.0 µl
$ZnCl_2$ (100 mM)	5.0 µl
KanE extract (1.5 - 2.5 mg protein/ml)	15 µl
KanS1 extract (2.0 mg protein/ml)	30 µl

B- The spectrophotometric assays were performed as follows:

KanC/KanS1 assay reaction (2.24)	100 µl
NAD (125 mM)	8.0 µl
$ZnCl_2$ (100 mM)	10.0 µl
KanE extract (1.5 - 2.5 mg protein/ml)	30 µl
Sterilized water ad.	1000.0 µl

2.26 Computer programs

The algorithms and program packages used for the computer-assisted analysis of DNA and protein sequences were: Staden package (Staden 1996; http://staden.sourceforge.net/), FramePlot (Ishikawa and Hotta, 1999; http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl), Clustal V (Higgins *et al*., 1991), Clustal W (Thompson *et al*., 1994; http://www2.ebi.ac.uk/clustalw/), FASTA 3 (Pearson and Lipman, 1988; http://www2.ebi.ac.uk/fasta3/), BLAST (Altschul *et al*., 1990; http://www.ncbi.nlm.nih.gov/BLAST/), Nucleic acids research (SMART**;** Letunic *et al*., 2004; http://nar.oupjournals.org/cgi/content/full/32/suppl_1/D142), Sequence Manipulation Suite (http://www.cbio.psu.edu/sms/index.html), Mac-plasmap (CGC Scientific. Inc.), DNA-STRIDER^{TM} 1.2 (Marck, 1988). Restriction enzyme analysis was performed using: Restriction Enzyme Site Mapper version 3 (http://www.restrictionmapper.org/) or Webcutter 2.0 (http://www.firstmarket.com/cutter/cut2.html). PCR amplification temperatures were computed using PRIMER-FIND 3.0 (Fröbel Labor-Geräte, Lindau), pDRAW32 (http://www.acaclone.com) or Primer X (http://bioinformatics.org/primerx/).

3 Results

A Gene clusters for the production of 2-deoxystreptamine (2DOS) and related ACAGAs

3.1 Isolation and sequence analysis of the biosynthetic gene clusters

For the elucidation of the genetics and the different pathways for the production of 2DOS and related aminocyclitol-aminoglycoside antibiotics (ACAGAs), an attempt was made to identify, clone, sequence and analyse the respective gene clusters and their immediate genomic environment. For this purpose, genomic DNAs were prepared from a series of strains producing ACAGAs (see Sect. 2.9.3). Genomic cosmid libraries were constructed at Combinature Biopharm in the *Streptomyces-E. coli* shuttle-cosmid vector pOJ436 (Bierman *et al*., 1992) and screened by various homologous and heterologous probes (cf. Sect. 2.9.11). This strategy was based on the existing knowledge that in *Actinobacteria* as a rule all the genes involved in production of an antibiotic or other secondary metabolite (i.e. genes for biosynthesis, resistance, transport, and often regulation) are located in one single gene cluster. Also, resistance and very few biosynthetic genes for many ACAGAs had already been described from our and other laboratories in the past, such that the partial sequence information could be taken from the data base for a majority of cases to be studied. All probes used were isolated by PCR using homologous or heterologous primers, cloned and sequenced for verification (cf. Tabs. 2.4 & 2.5). The positively screened cosmids were identified, analysed and some of them were selected for sequencing as described in Sect. 2.9.12 (cf. Tabs. 2.6 & 2.7). The insert sequences of a single or of overlapping cosmid clones were sequenced, analysed, and submitted to the EMBL gene bank and received the accession codes as given in Tab. 3.1.

From the ACAGAs chemical point of view as well as from the analysed gene clusters of the ACAGAs regarded, the ACAGAs can be categorized into five subgroups as follows: (i) ACAGAs containing 4,5-glycosylated 2DOS which includes both pseudotetrasaccharidic antibiotics such as NMs, PMs, LM and pseudotrisaccharidic antibiotics such as RM and BUs; (ii) ACAGAs containing 4,6-glycosylated 2DOS which includes KMs, GMs and TM; (iii) ACAGAs containing 4-glycosylated 2DOS with one representative, Apr; (iv) ACAGAs containing 5-glycosylated 2DOS with one representative, HM-B; (v) Other ACAGAs related to the2DOS containing antibiotics, i.e. FTMs (or astromicin) and IM.

Tab. 3.1 Sequence data of fully sequenced ACAGA gene clusters which have been submitted to the EMBL data base and their accession codes

Strain	Number	AGA-cluster	Size of submitted sequences (kb)	Accession codes
S. kanamyceticus	DSM 40500	*kan*	41.576	AJ628422
S. fradiae	DSM 40063	*neo*	50.466	AJ629247
S. rimosus subsp. *paromomycinus*	NRRL 2455	*par*	48.169	AJ628955
S. tenebrarius	DSM 40477	*tob* *apr*	43.220 41.623	AJ810851 AJ629123
S. hygroscopicus subsp. *hygroscopicus*	DSM 40578	*hyg*	34.921	AJ628642
M. echinospora	DSM 43036	*gen*	80.880	AJ628149
M. olivasterospora	DSM 43868	*for*	47.238	AJ628421
B. circulans	ATCC 21558	*btr*	19.248	AJ781030
S. ribosidificus	NRRL B-11466	*rib*	43.190	AJ744850
S. lividus	ATCC 21178	*liv*	40.579	AJ748832
S. tenjimariensis	ATCC 31603	*ist*	69.904	AJ845083
St. hindustanus	DSM 44523	*apr*	39.979	AJ875019

apr = apramycin; *btr* = butirosin; *for* = fortimicin; *gen* = gentamicin; *hyg* = hygromycin B; *ist* = istamycin; *; kan* = kanamycin; *liv* = lividomycin; *neo* = neomycin; *par* = paromomycin; *rib* = ribostamycin; *S. = Streptomyces; St. = Streptoalloteichus*; *tob* = tobramycin.

Furthermore, all the above mentioned ACAGAs possess 2DOS or a similar diamino cyclitol as the basic aglycone subunit. In the FTMs, this unit is basically derived from the *myo*-insitol pathway (see Sect. 1.1.5.2). On the other hand, in the IM, the aminocyclitol unit is derived from 2-deoxy-*scyllo* inosamine. All these mentioned antibiotics are produced by actinomycetes except BUs which are produced by *B. circulans*. Upon screening of both cosmid banks of both *S. tenebrarius* DSM 40477 and *St. hindustanus* DSM 44523, it was found that both strains contain two separate gene clusters for the two ACAGAs, Apr and TM, these two strains produce in

parallel. It was also observed that both gene clusters were not located in vicinity to each other on the genome since no overlapping cosmids were found by hybridizing with both probes specific for *apr*- and *tob*-clusters. Furthermore, all the submitted DNA sequences exhibit typical actinomycete G+C content (70%) and codon usage except the one from *B. circulans* (Bibb *et al*., 1984). The characterization and nomenclature of the ORFs for the respective gene clusters are indicated in the appendix section (see Tabs. A.1-A.14).

3.2 Gene clusters for the individual classes of 2DOS and related ACAGAs

3.2.1 4,5-glycosylated 2DOS-ACAGAs (NM group)

The NMs, PMs, LMs, RM and BUs are members of this group (cf. Fig. 1.2). The insert sequences containing their encoding *neo*-, *par*-, *liv*-, *rib*- and *btr*- gene clusters on single or overlapping cosmids were determined and annotated. All the respective gene clusters seem to be complete except for the *btr*-cluster (Fig. 3.2). For isolation of the *neo*-cluster, the cosmid bank from the NM producer *S. fradiae* DSM 40063 was screened by use of the *neoB* probe, where a total of 10 cosmids were positively hybridized. The two overlapping cosmids SfrA10 and SfrF04 were selected for sequencing and used to determine a total of 50466 bp representing a contiguous DNA segment on which the *neo*-cluster was located (Fig. 3.1). A total of 30445 bp from both cosmids comprising 22 ORFs was annotated for the *neo*-cluster, i.e. the gene cluster was flanked by ORFs SfrA10.7c (*aphA*; encoding aminoglycoside 3'-phosphotransferase) and SfrF04.17 (*neoY*; encoding putative AGA biosynthetic protein; Fig. 3.2).

For isolation of the *par*-cluster, the cosmid bank from the PM producer *S. rimosus* subsp. *paromomycinus* NRRL 2455 was first screened by use of the *aacC7* probe (Lopez-Cabrera *et al*., 1989) where a total of 6 cosmids were positively hybridized (cf. Tab. 2.7). The cosmid SriA13 was selected for sequencing (cf. Tab. 2.7). Upon partial sequencing and analysis of the regarded cosmid, it was found that the *par*-cluster was not located in the area of the *aacC7* resistance gene. Nevertheless, a partial insert sequence (4812 bp) around the *aacC7* gene was determined from this cosmid and submitted to the EMBL data base under accession code AJ749845. Upon analysis of this DNA segment, a putative second resistance gene namely *aac(6')-IIc* encoding an aminoglycoside 6'-N-acetyltransferase type IIc enzyme was found downstream of the *aacC7* gene (Fig. 3.3; cf. Tab. A.4).

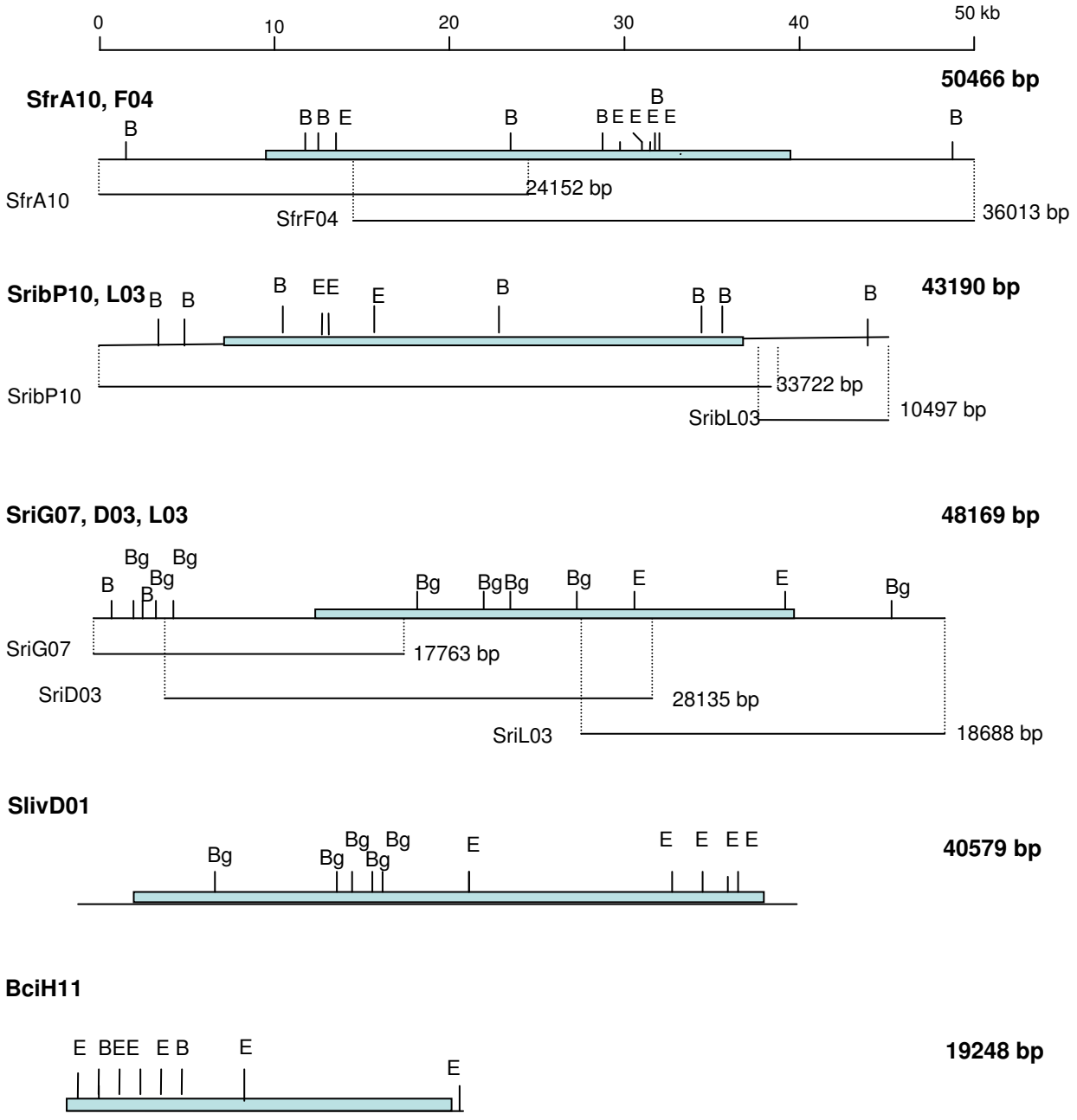

Fig. 3.1 Restriction maps and arrangement of the selected cosmids containing gene clusters of the NM group. The names in bold on the left side of the figure indicate the names of those cosmids from which the sequences were derived. The regions covering the individual gene clusters for each strain are indicated by the bar with the blue color. The numbers in bold on the right side represent the total length of the DNA segment determined in a single contig and submitted to the EMBL data base. Other numbers in the figure represent the lengths of the subsegments obtained from inserts of the individual cosmids in bp. E, B, Bg = restriction sites for *Eco*RI, *Bam*HI, *Bgl*II, respectively.

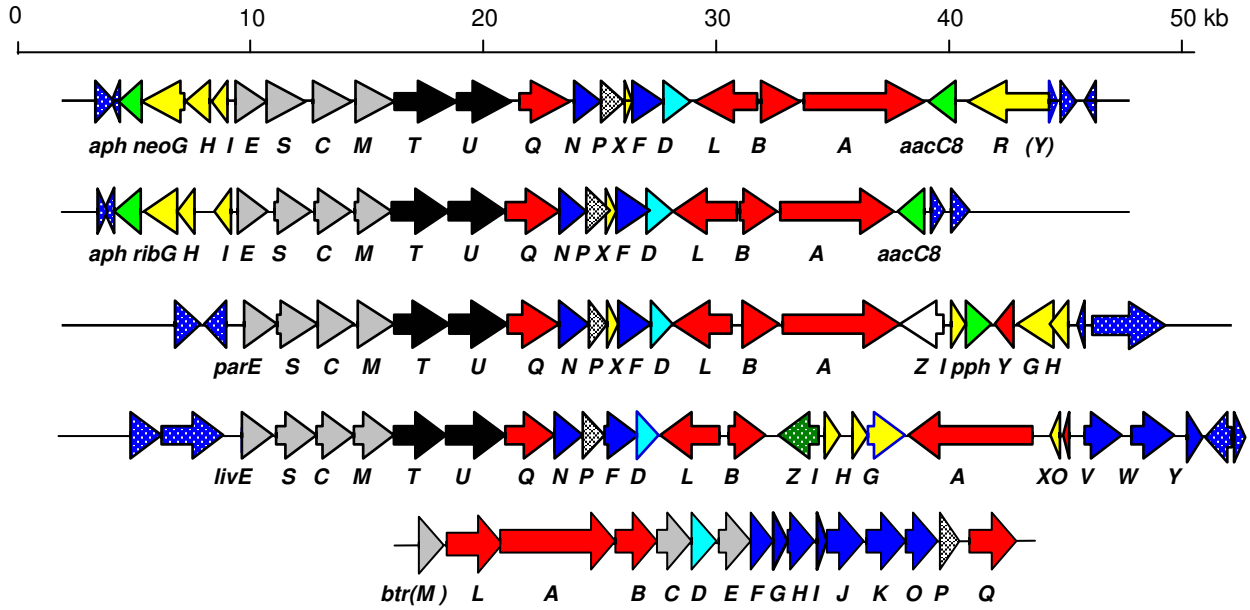

Fig. 3.2 Organization of ORFs for the gene clusters of the 4,5-glycosylated 2DOS-ACAGAs (NM group). The color codes for differentiating the genes encoding proteins putatively involved in:

- paromamine biosynthesis
- attachment and modification of sugar units
- ACAGAs biosynthesis
- UDP-D-glucosamine synthase or amidase
- extracellular aminoglycoside-phosphate phosphatase
- represent genes outside the ACAGA gene clusters
- sugar phosphate phosphatase
- transport
- regulation
- resistance

Therefore, for isolation of the *par*-cluster, two new probes (*parC* and *parA*) were prepared, using especially designed heterologous primers, for screening the same cosmid bank of *S. rimosus* subsp. *paromomycinus* (cf. Tab. 2.7). A total of 6 and 8 cosmids were positively hybridized with the *parC* and *parA* probes, respectively. Both SriG07 and SriD03 cosmids were hybridizing with *parC* as well as the cosmid SriL03 hybridizing with the *parA* probe were selected for sequencing of a total of 48169 bp of contiguous DNA (cf. Fig. 3.1). A total of 27061 bp within the submitted sequence was attributed to the *par*-cluster comprising 21 ORFs, flanked by ORFs SriD03.20 (*parE*) and SriL03.15c (*parH*; cf. Fig. 3.2).

For isolation of the *rib*-cluster, a total of 5 and 8 cosmids from the cosmid bank of the RM producer *S. ribosidificus* NRRL B-11466 were positively hybridized with *ribC* (isolated via

heterologous primers; cf. Tab. 2.6) and *rph* (Hoshiko *et al*., 1988) probes, respectively. Both SribP10 and SribL03 cosmids were determined, analysed and selected for DNA sequencing. A total of 43190 bp of a contiguous DNA segment was completely sequenced of which the *rib*-cluster covered a total of 26036 bp. The *rib*-cluster was composed of at least 23 ORFs and flanked by two resistance genes (*aacC8* and *rph*; cf. Fig. 3.2).

For isolation of the *liv*-cluster, both *livB* and *livS* probes were prepared again using the respective pairs of heterologous primers (cf. Tab. 2.5). The screening of the cosmid bank of the LM producer *S. lividus* ATCC 21178 delivered a total of 7 and 11 positively hybridizing cosmids, respectively (cf. Tab. 2.7). A total of 40579 bp of a contiguous DNA segment was sequenced from the cosmid SliD01; of these a subsegment of 31269 bp, comprising 23 ORFs, was annotated representing the *liv*-cluster (cf. Fig. 3.2).

For isolation of the *btr*-cluster, the cosmid bank of the BUs producer *B. circulans* ATCC 21558 was constructed and screened by the *btrC* probe (Ota *et al*., 2000) where 19 cosmids were positively hybridized (cf. Tab. 2.7). Analysis of the regarded cosmids resulted in selection of only one cosmid (BcirH11) for sequencing. A total of 19248 bp of a contiguous DNA segment was sequenced, where 19197 bp were dedicated to represent the major part of the already known *btr*-cluster with minor extension (cf. Fig. 3.2). For identification of the assumingly still lacking part of the *btr*-cluster upstream the *btrS* gene, the cosmid bank was rescreened by using the *btrM* and *btrS* probes (Ota *et al*., 2000) as well as a new cosmid bank was constructed and screened by the *btrS* probe. A total of 17 cosmids were hybridized positively (cf. Tab. 2.7). Unfortunately, none of the respective cosmids was turned out to contain DNA contiguous to the *btrS* gene. Possibly this genomic region near the *btrS* gene does not deliver stable cosmid clones. Obviously, the Japanese group (Ota *et al*., 2002) which had published a larger portion of the *btr*-cluster already before from another strain of *B. circulans*, had very similar difficulties in extending the *btrS* end. Therefore, the real scope and extend of the *btr*-cluster is left in unknown state at present.

Regarding the resistance genes in the respective clusters, both the *neo*- and *rib*-clusters were found to have one resistance gene at each side of the cluster; each gene encoding AGA-3'-phosphotransferase (AphA) and aminocyclitol 3-N-acetyltransferase VIII (AAC). In addition, only one of these two resistance genes (*aphA*) was found at one side of the *par*-cluster and none so far was found in the respective DNA segments covering the *liv*- and *btr*-clusters. Two genes (*neoT*, *U* and the related genes) putatively encode ABC transporters that would be involved in the export of the respective ACAGAs products were conserved among the streptomycete clusters. In

addition, three conserved genes (*neoG*, *H*, *I* and the related genes) with similarity to *cinorf*12, 13, 14 in another streptomycete antibiotic cluster (Widdick *et al*., 2003) were also found to be highly conserved in all the streptomycete clusters of this group. These three genes encode proteins (NeoI/H/G and related proteins) which by some evidences could play a role in the regulation of the respective ACAGAs biosynthesis (see Sect. 4.3).

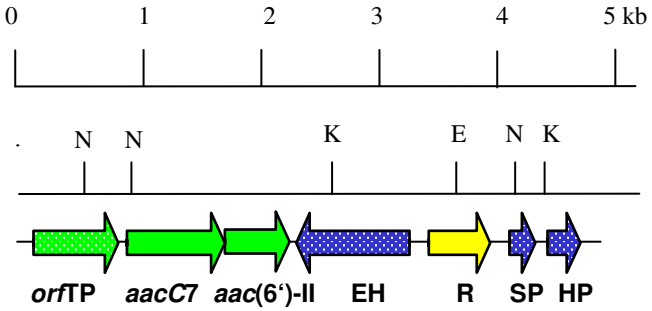

Fig. 3.3 Genetic map around the gene for the aminocyclitol acetytransferase VII (*aacC7*) from *S. rimosus* subsp. *paromomycinus* NRRL 2455: The ORFs from left to right encode an integral membrane protein (TP; incomplete), aminocyclitol 3-N-acetyltransferase type VII (*aacC7*), a putative aminoglycoside 6'-N-acetyltransferase type II (*aac6'-IIc*), putative epoxide hydrolase (EH), a putative TetR-family transcriptional regulator (R), a putative cold shock protein (SP) and a hypothetical protein (HP; incomplete). N, K, E = restriction sites for *Nco*I, *Kpn*I and *Eco*RI, respectively.

3.2.2 4,6-glycosylated 2DOS-ACAGAs (KM group)

KMs, GMs and TM are members of this group (cf. Fig. 1.3). Screening of the cosmid banks of the KM producer *S. kanamyceticus* DSM 40500, GM producer *M. echinospora* DSM 43036 and TM producer *S. tenebrarius* DSM 40477 with probes derived from *kmr* (Demydchuk *et al*., 1998), *grm* (Kelemen *et al*., 1991) and *tobS2* (via heterologous primers) resulted in identification of 4, 13 and 11 positively hybridizing cosmids, respectively (cf. Tab. 2.7). Analysis of the obtained cosmids resulted in selection of the cosmids: SkaJ19 and SkaJ15 from *S. kanamyceticus*; MecP21, MecE04 and MecG05 from *M. echinospora*; SteM07 and SteK17 from *S. tenebrarius* for sequencing. Contiguous DNA segments of a total size of 41576, 51597, 43220 bp from the inserts on these cosmids were firstly determined, respectively. The presumed *kan*-, *gen*- and *tob*-clusters were located on these segments (Fig. 3.4). From this analysis it was unclear whether the *gen*-cluster could extend downstream beyond the last ORF (*genN*). For this reason, another probe (*genI*) was prepared and used for screening of the same cosmid bank where another 9 cosmids being positively hybridized (cf. Tab. 2.7). An overlapping cosmid (MecO02) was selected for sequencing and finally the contiguous DNA segment was extended to become

80880 bp where the *gen*-cluster was fully present (cf. Fig. 3.4). As shown in Fig. 3.5, the arrangement of the three gene clusters was conserved at least in part, especially between those for the mostly related ACAGAs, KM and TM. A characteristic feature of the three gene clusters is that some of the centrally important genes, also conserved in the other ACAGAs gene clusters had been duplicated or even multiplied in the past, most probably in order to acquire new biosynthetic functions (see Sect. 4.2.4).

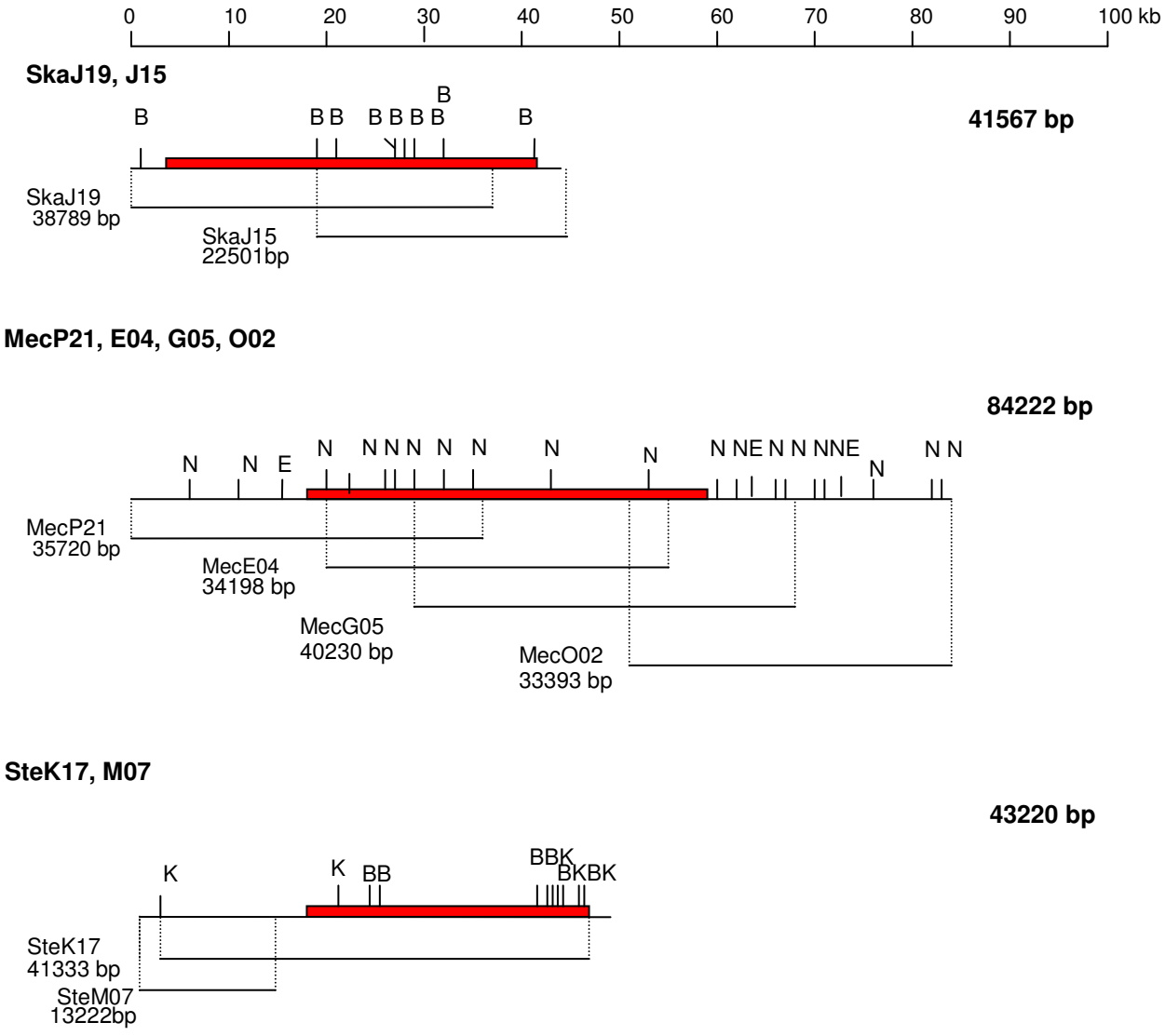

Fig. 3.4 Restriction maps and arrangement of the selected cosmids containing gene clusters of the KM group. The names in bold on the left side of the figure indicate the names of the cosmids from which the sequences were derived. The regions covering the individual gene clusters for each strain are indicated by the bar with the red color. The numbers in bold on the right side represent the total length of the DNA segment determined in a single contig and submitted to the EMBL data base. Other numbers in the figure represent the length of the subsegments obtained from inserts of the individual cosmids in bp. K, B, N, E = restriction sites for *Kpn*I, *Bam*HI, *Nco*I, *Eco*RI, respectively.

Examples are the pairs of *kanS1* and *kanS2* (and the related *genS1*, *genS2, tobS1* and *tobS2*) and *kanM1* and *kanM2* (and the related *genM1, genM2, tobM1* and *tobM2*). Another example is found in the *gen*-cluster alone, where four copies of the aminotransferase type III genes (*genB1*, *B2*, *B3* and *B4*) were found although only a single related gene is present in both *kan*- and *tob*-clusters. A more general analysis of the respective gene clusters revealed the presence of five genes (*kanC*, *S1*, *E*, *D1*, *M1* and the related genes) that are also conserved in the *neo*-like gene clusters. These five genes encode proteins that would be responsible for the biosynthesis of paromamine (see Sect. 4.2.2 and Fig. 4.1). The analysis of the *kan*- and *gen*-clusters suggested that they could be complete for the biosynthesis of the respective antibiotics (see Sect. 4.2.4 and Fig. 4.3). In addition, the *gen*-cluster was found to have a complete rRNA operon (is partially comprised in the published sequence) which seems to flank the *gen*-cluster at one side.

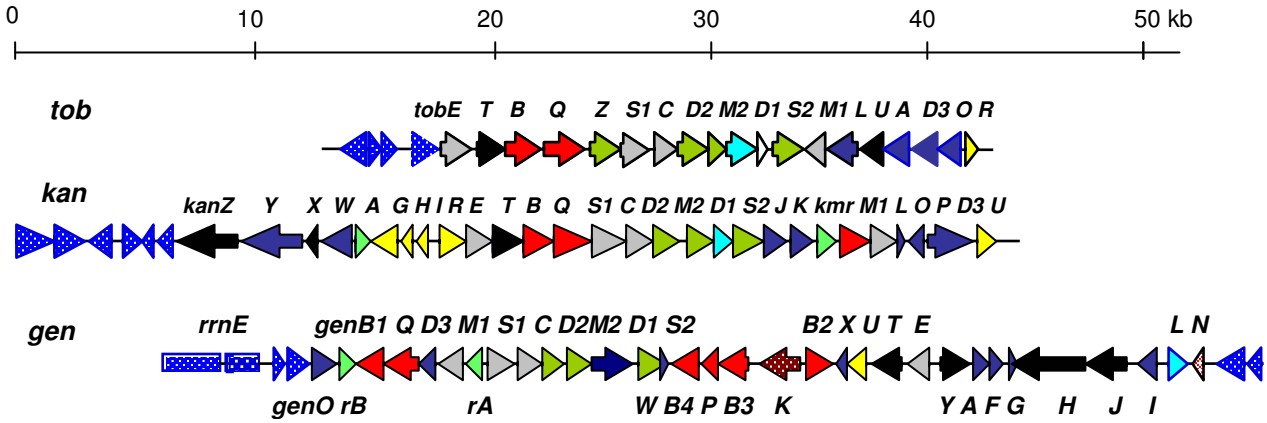

Fig. 3.5 Organization of ORFs for the gene clusters of the 4,6-glycosylated 2DOS-ACAGAs (KM group). The color codes for differentiating the genes encoding proteins putatively involved in:

- paromamine biosynthesis
- attachment of the first sugar unit at position 4 of the cyclitol ring and its modification
- attachment of the second sugar unit at position 6 of the cyclitol ring and its modification
- UDP-D-glucosamine synthase or amidase
- resistance
- ACAGAs biosynthesis
- represent genes outside the ACAGA gene clusters
- C-methyltransferase
- transport
- regulation
- N-methyltransferase

This *rrn* operon, called *rrnE* because of its closest relationship to the *rrnE* operon of *S. coelicolor* A3(2) and is completely enclosed in our sequence and located between ORFs MecP21.23c and MecE04.1 (Bentley *et al*., 2002; cf. Tab. A.9). On the other hand, analysis of the *tob*-cluster obviously lacked the gene(s) encoding proteins that could be involved in 3'-dehydroxylation. Furthermore, two unique genes (*kanJ* and *kanK*) were found only in the *kan*- and not in the *tob*-cluster which therefore, encoding KanJ and KanK which might be involved in the 2'-deamination process and formation of KM-A from KM-B. A common feature of this group seems to be the use of modification of the target site (rRNA) as a resistance mechanism instead of ACAGAs modifications. Two resistance genes were found in both *kan* (*kanA* and *kmr*) and *gen*-clusters (*gmrA* and *gmrB*); however, no equivalent resistance genes could be detected in the contiguous DNA segment covering the *tob*-cluster (cf. Fig. 3.5). The possible components of a sensor/response regulatory system (*kanG, H* and *I*) were only present in the *kan*-cluster (also conserved in the actinomycete gene clusters of the NM group; cf. Fig. 3.2 above) while other possible regulatory genes were found in all three clusters (*kanU*, *kanR*, *genU* and *tobR*). Several genes that could encode proteins involved in the transport of these ACAGAs were found in the regarded gene clusters (*kanZ*, *kanT*, *kanX*, *tobT*, *tobU*, *genT*, *genY*, *genH* and *genJ*) however, only one transporter gene (*kanT* and the related genes) was found to be highly conserved among the three clusters.

3.2.3 4-glycosylated 2DOS-ACAGAs (Apr group)

Apr is an example of this group which is produced by both *S. tenebrarius* DSM 40477 and *St. hindustanus* DSM 44523 (cf. Fig. 1.4). From each strain, a cosmid bank was constructed and screened by *kamB* (Holmes *et al*., 1991) and *aprA* (AY129957) probes (cf. Tab. 2.7). A total of 22 and 15 positively hybridizing cosmids were isolated from the cosmid banks of *S. tenebrarius* and *St. hindustanus*, respectively. Their analysis resulted in selection of the cosmids SteO08 from *S. tenebrarius* and ShinN01 from *St. hindustanus* for sequencing. From the two cosmids SteO08 and ShinN01, a total of 41623 and 39979 bp of contiguous DNA segments covering the *apr*-cluster were sequenced and annotated (cf. Tab. 3.1). Subsegments of about 38770 and 34462 bp from SteO08 and ShinN01 cosmids both comprising 26 ORFs were attributed to the *apr*-clusters, respectively. Three incomplete transposase genes (*tatA*, *B*, *C*) were found only inside the *apr*-cluster present on cosmid SteO08. The arrangement and the relatedness of the 26 ORFs in the two *apr*-clusters were highly conserved and the sequence identity was higher than 90% (Fig.

3.7). Genes that encode proteins which are likely to be involved in the biosynthesis of the 2DOS moiety (*aprC*, *S*, *E*) were also conserved in both clusters.

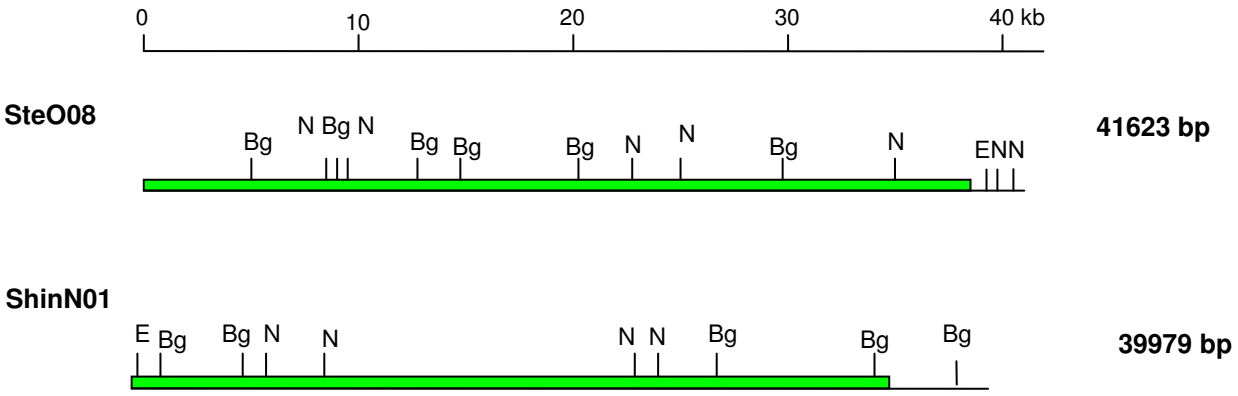

Fig. 3.6 Restriction maps and arrangement of the selected cosmids containing the *apr*-gene clusters. The names in bold on the left side of the figure represent the cosmid names from which the sequences were derived. The regions covering the *apr*-gene cluster are indicated by the bar with the green color. The numbers in bold on the right side represent the total length of the DNA segment submitted to the EMBL data base. Bg, E, N = restriction sites for *Bgl*II, *Eco*RI, *Nco*I, respectively.

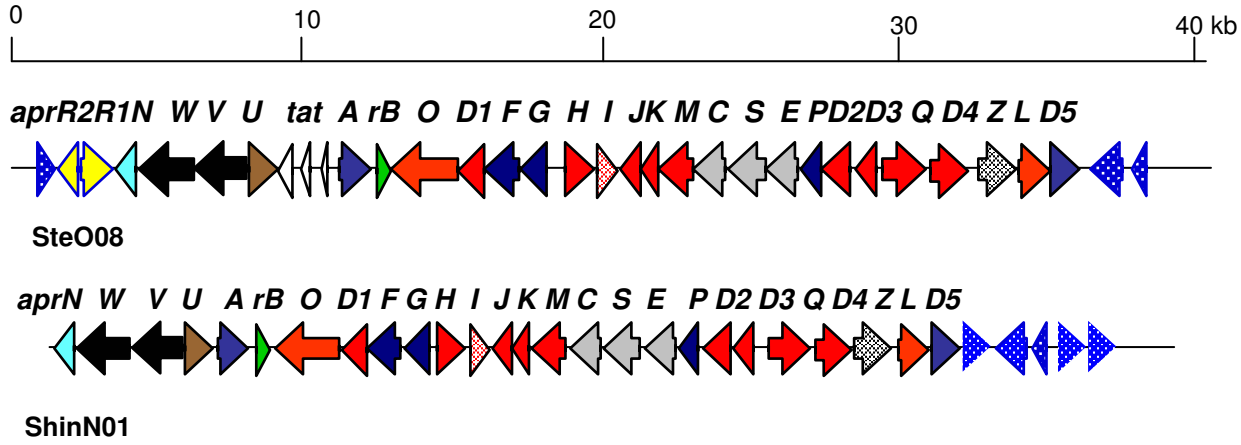

Fig. 3.7 Organization of ORFs for the two *apr*-gene clusters derived from SteO08 and ShinN01 cosmids. Both the DNA segments covering the two *apr*-gene clusters on the cosmids SteO08 and ShinN01 were derived from *S. tenebrarius* DSM 40477 and *St. hindustanus* DSM 44523, respectively. The color codes for differentiating the genes encoding proteins putatively involved in:

- 2DOS biosynthesis
- attachment of sugar units to the 2DOS moiety and their modifications
- UDP-D-glucosamine synthase or amidase
- extracellular aminoglycoside- phosphate phosphatase
- transposases
- ACAGAs biosynthesis
- represent genes outside the *apr*-clusters
- transport
- resistance
- kinase reaction
- N-methyltransferase
- regulation

The *apr*-cluster on the cosmid ShinN01 was incomplete for a small stretch, since it lacked part of the last gene, *aprN*. It was however, assumed that in both cases the complete set of the *apr*-genes was identified and therefore, no further attempt to elongate the sequence from *St. hindustanus* DSM 44523 genome was made.

3.2.4 5-glycosylated 2DOS-ACAGAs (HM-B group)

HM-B is a member of this group as it possesses the 2DOS unit substituted at position 5 with a disaccharidic moiety (cf. Fig. 1.4). The cosmid bank of *S. hygroscopicus* subsp. *hygroscopicus* DSM 40578 was screened with the probe of HM-B phosphotransferase gene (*pht*; Zalacain *et al*., 1986). A total of 18 cosmids positively hybridized with this probe. Analysis of these cosmids resulted in selection of the cosmid ShyG17 for sequencing (cf. Tab. 2.7). A total contiguous DNA segment of 34921 bp, mostly likely harboring the complete *hyg*-cluster was sequenced (cf. Tab. 3.1). The respective DNA segment contained 33 ORFs of which 22 - 23 ORFs belonged to the *hyg*-cluster, from the ORFs ShyG17.9c (*hygV*) to ShyG17.31c (*hygZ*). The orientation, restriction map, location of the *hyg*-cluster and organization of ORFs are shown in Figs. 3.8 and 3.9. It is also expected that the *hyg*-cluster would contain all the genes required for the biosynthesis of HM-B (see Sect. 4.2.6). The *hyg*-cluster represents a unique mixture of genes having very few similarities with several other ACAGAs genes clusters or other sources of genes involved in rare sugar or oligosaccharide biosynthesis.

A possible sensor/response regulatory system (*hygI*, *H* and *G*) was also conserved in the *hyg*-cluster (Fig. 3.9 and see Sect. 4.3). As expected, the HM-B resistance gene, encoding APH (7'') which had been previously submitted to the data base (Zalacain *et al*., 1986) was present inside the gene cluster (*hygA*). The submitted sequences of this gene turned out to be identical to our data. Moreover, the *hyg*-cluster was observed to be flanked on each side by a transposase gene (Fig. 3.9). Regarding the structural similarity between HM-B and spectinomycin that both have a heterocyclic ring with two oxygen atoms in condensing the first sugar unit to the cyclitol moiety, the *hyg*-cluster contains a couple of genes (*hygY* and *hygF*) with a good similarity to related ones in the *spc*-cluster (*spcY* and *spcF*; cf. Figs. 1.1 and 1.4). The *hyg*F gene was anticipated to encode an UDP-galactosyltransferase where the gene product of *hygY* plays a role in heterocyclization between the first and the second sugar moiety in HM-B. In addition, the *hyg*-cluster contains a gene encoding a putative N-methyltransferase (HygM) with high level of similarity to the corresponding one in the *spc*-cluster (SpcM). Another putative glycosyltransferase gene (*hygD*)

could encode the enzyme involved in the transfer of the second sugar, a heptose to the first sugar (galactose). Both gene products of *hygV* and *hygW* would be expected to play a role in the export system for HM-B. Similar to the *apr*-cluster, a putative kinase gene (*hygN*) as well as a putative phosphatase gene (*hygU*) were also found in the *hyg*-cluster (cf. Sect. 3.2.3).

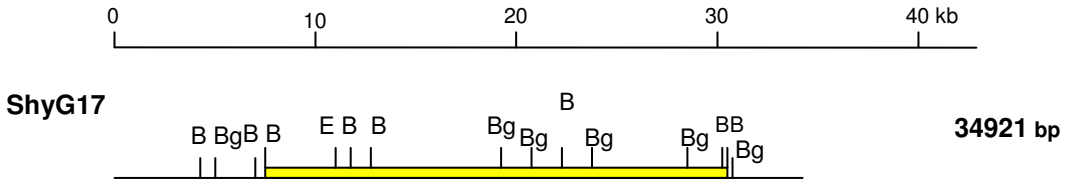

Fig. 3.8 Restriction map of the insert of cosmid ShyG17 and location of the *hyg*-cluster. The region for the *hyg*-cluster is indicated by the yellow bar. The number in bold on the right side gives the total length of the DNA segment submitted to the EMBL data base. E, Bg, B = restriction sites for *Eco*RI, *Bgl*II, *Bam*HI, respectively.

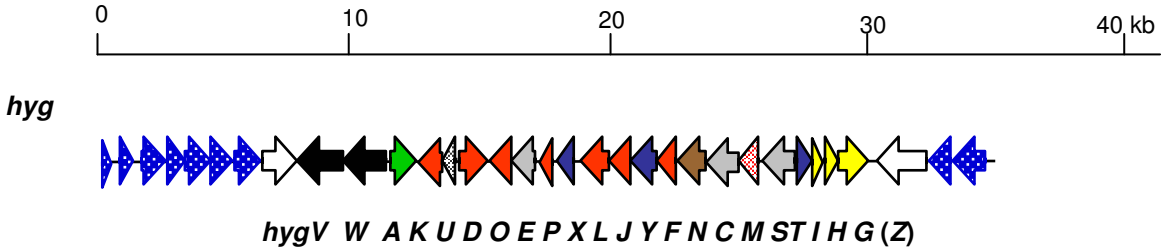

Fig. 3.9 Organization of ORFs for the *hyg*-cluster. The DNA segment was cloned into cosmid ShyG17. The color codes for differentiating the genes encoding proteins putatively involved in:

- 2DOS biosynthesis
- attachment to the 2DOS moiety and modification of sugar units
- regulation
- transposases
- ACAGA biosynthesis
- represent genes outside the *hyg*-cluster
- transport
- resistance
- sugar phosphatase
- kinase reaction
- N-methyltransferase

3.2.4 ACAGAs related to 2DOS-ACAGAs (FTM and IM group)

Both FTM-A and IM-A are pseudodisaccharidic ACAGAs produced by *M. olivasterospora* DSM 43868 and *S. tenjimariensis* ATCC 31603, respectively. They are grouped together because of their related chemical structures (cf. Fig. 1.5). However, their routes of biosynthesis are different in part. The cyclitol moiety of FTM-A is biosynthesized by a different route in comparison to the

other 2DOS-ACAGAs where *myo*-inositol is the basic starting material. On the other hand, IM-A contains 2-deoxyfortamine (a 2-deoxy-3,6-diaminocylitol derivative) as a basic cyclitol unit which originates from 2-deoxy-*scyllo*-inosose as in the other 2DOS-ACAGAs.

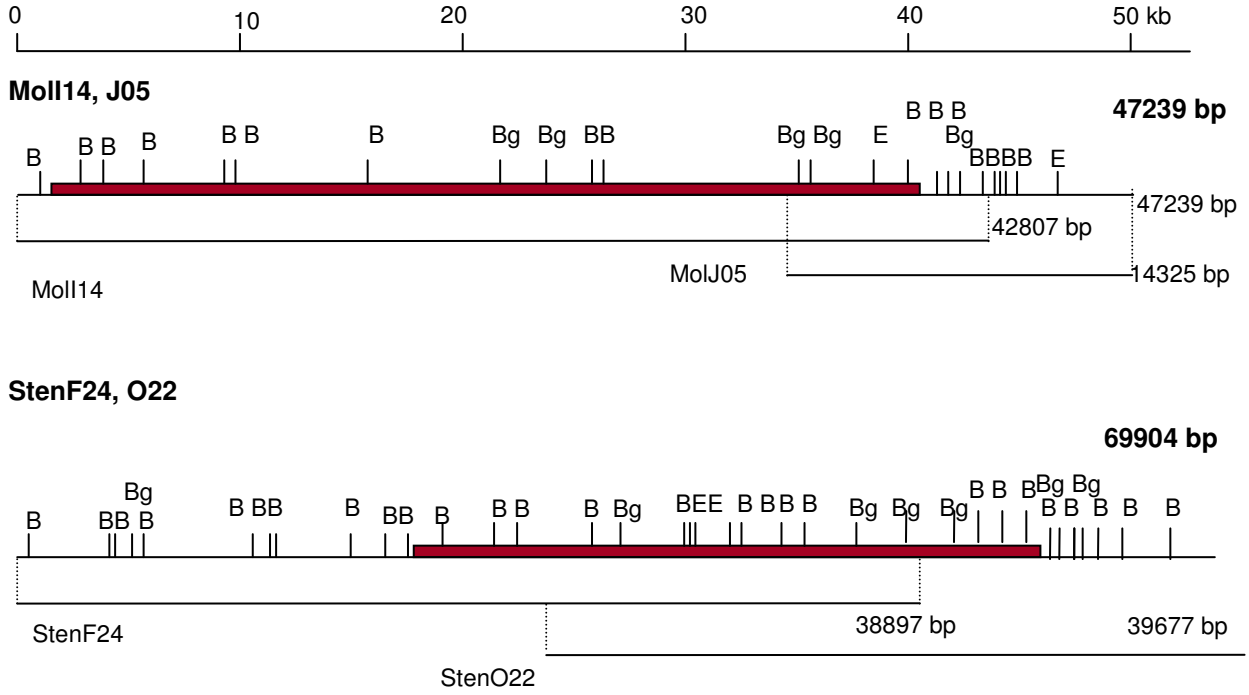

Fig. 3.10 Restriction maps and arrangement of the selected cosmids for the *for*- and *ist*-clusters. The names given in bold on the left side of the figure represent the cosmid names from which the sequences were derived. The regions covering the individual gene clusters for each strain are indicated by the brown bar. The numbers in bold on the right hand side represent the total length of the DNA segment determined in a single contig and submitted to the EMBL data base. Other numbers in the figure represents the length of the subsegments obtained from inserts of the individual cosmids in bp. Bg, B, E = restriction sites for *Bgl*II, *Bam*HI, *Eco*RI, respectively.

However, the basic routes for 3,6-diaminocyclitol biosynthesis should be quite similar. Screening of the cosmid banks of *M. olivasterospora* with *fms14* (Dairi *et al*., 1992c) and *fmrO* (Ohta and Hasegawa, 1993b) probes and of *S. tenjimariensis* with *kamA* (Ohta and Hasegawa, 1993a) and *istC* probes resulted in isolation of a total of 11 and 25 positively hybridizing cosmids, respectively (cf. Tab. 2.7). After mapping of the positive cosmids: the cosmids MolI14, MolJ05 (from *M. olivasterospora*); and StenF24, StenO22 (from *S. tenjimariensis*) were selected for sequencing. A total of 42807 and 69904 bp of contiguous DNA segments from *M. olivasterospora* and *S. tenjimariensis* covering the *for*- and *ist*-clusters were determined, respectively (cf. Tab. 3.1). The DNA segment from *M. olivasterospora* comprised 42 ORFs from which 33 ORFs were attributed to the *for*-cluster. The DNA segment from *S. tenjimariensis* covered 37 ORFs from which 28 ORFs were postulated as comprising the *ist*-cluster. Restriction maps and arrangement of genes on the selected cosmids are shown in Figs. 3.10 and 3.11. As

presented in Fig. 3.11, both *for*- and *ist*-clusters were fully sequenced and all the postulated genes products required for the biosynthesis of the respective antibiotics were putatively present (see Sect. 4.2.7; Fig. 4.6). Especially, the proteins encoded by the three genes *forA, forC* and *forS* are expected to be required for the formation of *scyllo*-inosamine from *myo*-inositol as well as the two proteins IstC and IstS required for the formation of 2-deoxy-*scyllo*-inosamine from G-6-P were found.

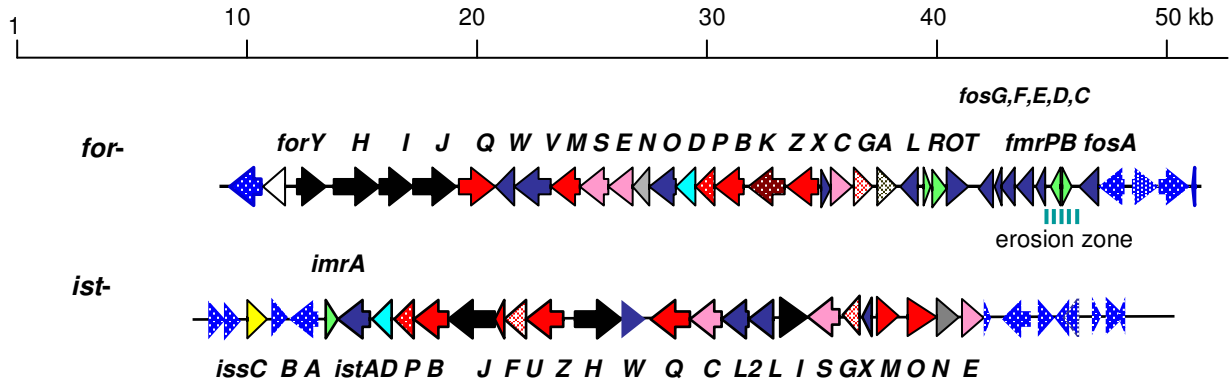

Fig. 3.11 Organization of the ORFs in the *for*- and *ist*-clusters. The symbols *for* and *ist* indicate fortimicin and istamycin gene clusters obtained from *M. olivasterospora* DSM 43868 and *S. tenjimariensis* ATCC 31603, respectively. The color codes for differentiating the genes encoding proteins putatively involved in:

- aminocyclitol biosynthesis
- attachment and modification of the amino sugar units
- UDP-D-glucosamine synthase or amidase
- ACAGAs biosynthesis
- N-methyltransferase
- C-methyltransferase
- O-methyltransferase
- represent genes outside the ACAGA gene clusters
- regulation
- transport
- resistance
- transposase
- 3'-phosphotransferase
- sugar phosphatase

Some genes were present only in one cluster, encode proteins that could be responsible for exhibiting individual functions: a putative 6'-C-methyltransferase (ForK; related to GenK in the *gen*-cluster) and a putative 6'-N-methyltransferase (IstU). Moreover, genes which would encode similar biosynthetic functions for both antibiotics were conserved in both clusters: 3'-phosphotransferases (ForP and IstP) which could be involved in the 3',4'-dehydroxylation process of the sugar moiety; putative 6-N-methyltransferases (ForG, IstG); putative 1-O-methyltransferases (ForN, IstN); putative 6'-dehydrogenases (ForE, IstE); putative UDP-D-

glucosaminyltransferases (ForM, IstM); putative 6'-dehydrogenases and 6'-aminotransferases (ForQ, ForB and IstQ, IstB). Both *for*- and *ist*-gene clusters contain putative resistance genes encoding 16S rRNA methyltransferases or some truncated proteins of this family (*fmrR*, *fmrO*, *fmrP*, *fmrB* for FTM and *imrA* for IM). Both *fmrO* and *imrA* (*kamA*) were previously submitted to the data bases and our sequence data turned out to be identical with those previously submitted (Ohta and Hasegawa, 1993a & b). Three transporter genes (*forH, I, J* and *IstH, I, J*) which encode proteins that could be involved in the export system for both antibiotics were also found to be conserved in both clusters. One possible regulatory gene (*issC*) could also be traced in the *ist*-cluster and no equivalent was found so far in the DNA segment covering the *for*-cluster.

3.3 Newly developed heterologous primers

Lack of suitable probes for screening of cosmid banks of some of the ACAGA-producing strains made the identification and design of specific heterologous primers necessary. Therefore, several attempts to design more universal primers for conserved genes in all the ACAGA gene clusters or in subgroups thereof were made.

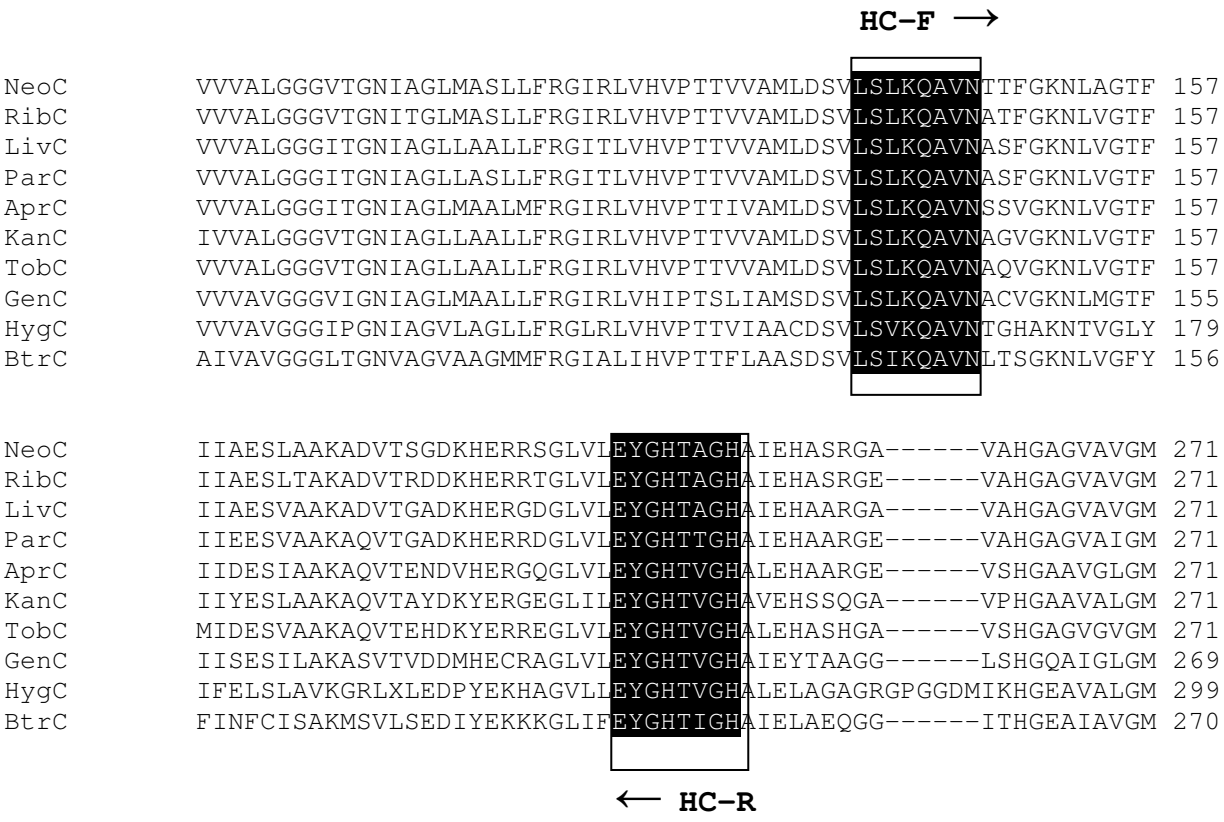

Fig. 3.12 Partial alignment of 2-deoxy-*scyllo*-inosose synthase proteins ("NeoC"-family) for the design of heterologous primers. Boxes mark the areas which were used to design the heterologous primers. The nucleotide sequences of HC-F & HC-R primers are listed in Tab. 2.5. The numbers indicate the position within the corresponding proteins.

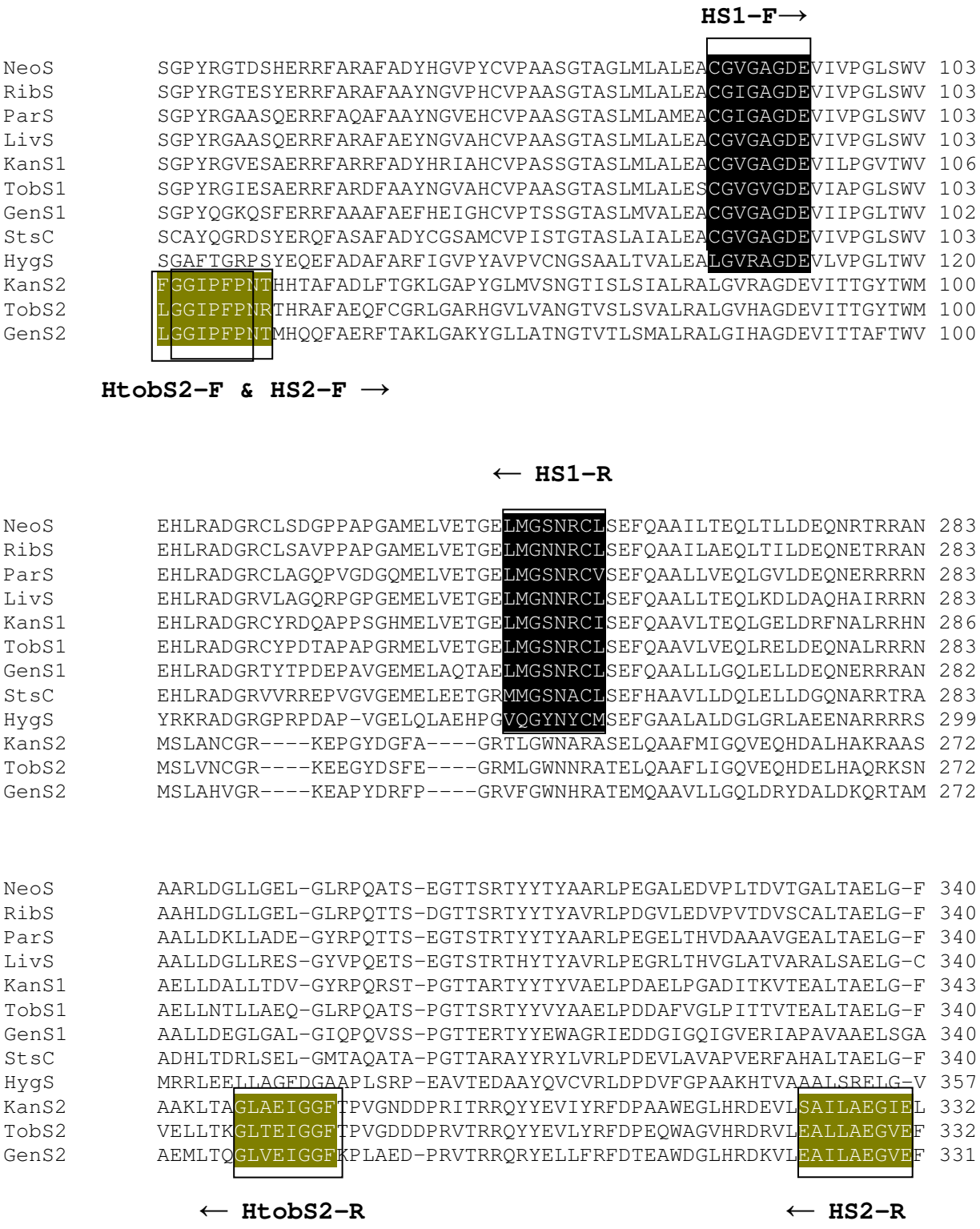

Fig. 3.13 Partial alignment of 2-deoxy-*scyllo*-inosose aminotransferase and other aminotransferase proteins ("NeoS"-family) for the design of heterologous primers. Boxes mark the areas which were used to design the heterologous primers. The nucleotide sequences of the primers selected are listed in Tab. 2.5. The numbers indicate the position within the corresponding proteins.

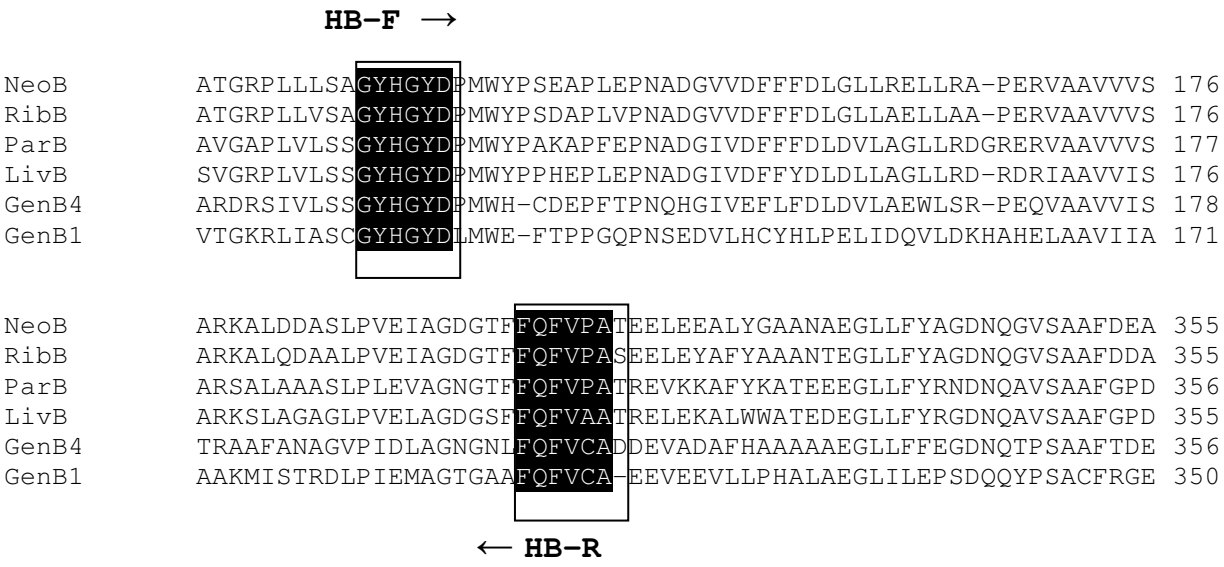

Fig. 3.14 Partial alignment of the putative 6'-aminotransferase proteins ("NeoB"-family) for the design of heterologous primers. Boxes mark the areas which were used to design the heterologous primers. The nucleotide sequences of HB-F & HB-R primers are listed in Tab. 2.5. The numbers indicate the position within the corresponding proteins.

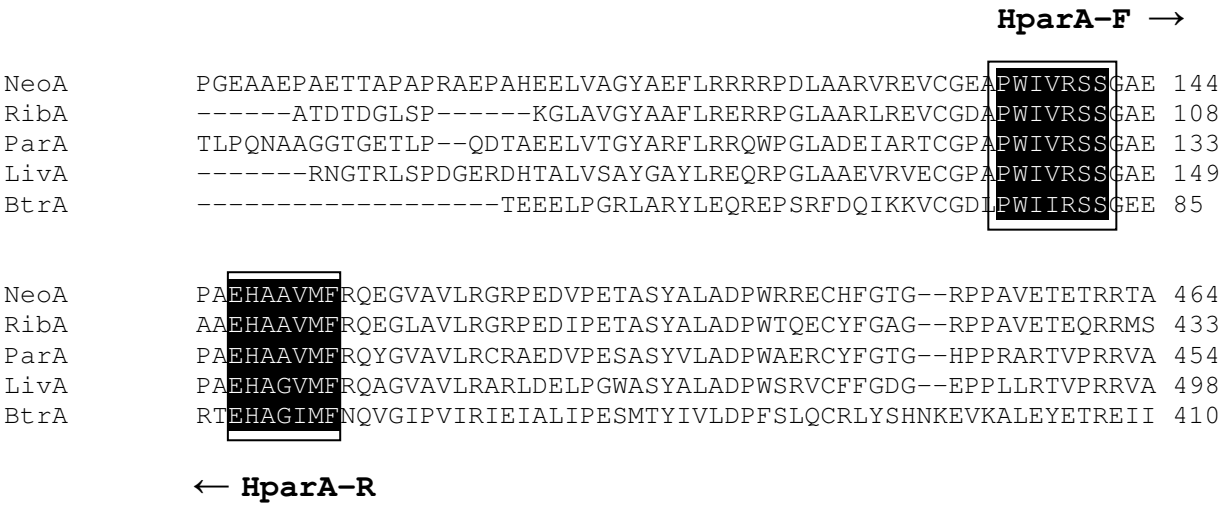

Fig. 3.15 Partial alignment of the putative ribosyltransferase proteins ("NeoA"-family) for the design of heterologous primers. Boxes mark the areas which were used to design the heterologous primers. The nucleotide sequences of HparA-F & HparA-R primers are listed in Tab. 2.5. The numbers indicate the position within the corresponding proteins.

Possible heterologous primers were generally proposed via multiple sequence alignments of the already known amino acid sequences of the respective protein families using the Clustal W program (cf. Sect. 2.26). The areas with maximum conservation of the amino acid sequences of the "NeoC", "NeoS", "NeoB", "NeoA" and "GenK" families of biosynthetic enzymes were

chosen (Figs. 3.12 - 3.16). To get the optimal nucleotide sequences for the heterologous primers, the selected amino acids were back translated into their corresponding codons taking into consideration the codon usage for actinomycetes (Bibb *et al*., 1984). The nucleotide sequences of the heterologous primers as well as the expected size for the respective PCR fragments obtained are listed in Tab. 2.5. Interestingly, it turned out to be very useful to design a pair of primers (HS2-F & HS2-R) that could distinguish between genes *kanS1* from *kanS2* (cf. Fig. 3.13; Tab. 3.2). Detection of the respective genes via PCR using the genomic DNA of the respective strains as a template and the heterologous primers is summarized in Tab. 3.2.

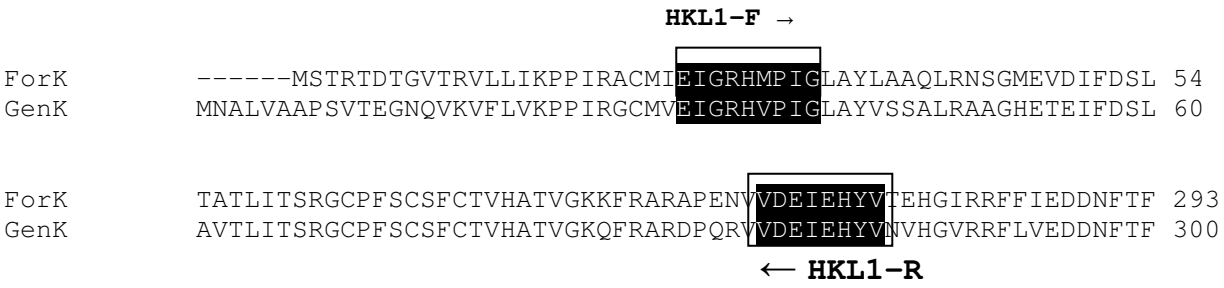

Fig. 3.16 Partial alignment of the putative 6'-C-methyltransferase proteins ("GenK"-family) for the design of heterologous primers. ForK from *M. olivasterospora* and GenK from *M. echinospora* were aligned using Clustal W program (Sect. 2.26). Boxes mark the areas which were used to design the heterologous primers. The nucleotide sequences of HKL1-F & HKL1-R primers are listed in Tab. 2.5. The numbers indicate the position within the corresponding proteins.

Tab. 3.2 Detection of biosynthetic genes in genomic DNA using heterologous primers

Primer pairs / Strain	HC-F/ HC-R	HS1-F/ HS1-R	HS2-F/ HS2-R	HB-F/ HB-R	HparAF/ HparA-R	HKL1-F/ HKL1-R
S. kanamyceticus DSM 40500	+	+	+	+	-	nt
S. fradiae DSM 40063	+	+	-	+	+	nt
S. rimosus subsp. *paromomycinus* NRRL 2455	+	+	-	+	+	nt
S. tenebrarius DSM 40477	+	+	+	+	-	nt
S. hygroscopicus subsp. *hygroscopicus* DSM 40578	+	+	-	+	-	nt
M. echinospora DSM 43036	+	+	+	+	nt	+
M. olivasterospora DSM 43868	+	+	nt	nt	nt	+
S. ribosidificus NRRL B-11466	+	+	-	+	+	nt
S. lividus ATCC 21178	+	+	nt	+	nt	nt
S. tenjimariensis ATCC 31603	+	+	nt	+	nt	-
St. hindustanus DSM 44523	+	+	nt	+	nt	nt

+ = PCR product obtained (proven by DNA-sequencing); - = no PCR product obtained (in general no homologous gene present in the gene cluster); nt = not tested.

Tab. 3.3 Biosynthetic proteins encoded by genes conserved among the ACAGA gene clusters analysed

Key enzyme (postulated function)	Members	pfam/COG	Remarks/ General function
2-deoxy-*scyllo*-inosose synthase	BtrC, NeoC, LivC, ParC, GenC, IstC, RibC, HygC, AprC, TobC, KanC.	pfam01761 COG0001	cyclase in 2DOS -pathway
L-glutamine (PLP): 2-deoxy-*scyllo*- -inosose aminotransferase (AT-I) and 1-keto-2,3-deoxy-3-amino-*scyllo*- -inositol aminotransferase (AT-II)	BtrS, NeoS, LivS, ParS, RibS, HygS, AprS, TobS1, KanS1, GenS1, ForS, IstS	pfam01041 COG0399	putative bifunctional ketocyclitol aminotransferase (AT type IV)
aminocyclitol-1-dehydrogenase	BtrE, NeoE, LivE, ParE, RibE, AprE, TobE, KanE, GenE, ForE, IstE	pfam00107 COG1063	good evidence
L-glutamine (PLP): neutral sugar Aminotransferase; putative 3"-amino- -transferase of the α-1,6- glycosyl unit of Kan, Gen and Tob.	KanS2, TobS2, GenS2	pfam01041 COG0399	AT type IV
α-1,4-glycosyltransferase (hexosaminyltransferase)	BtrM, NeoM, LivM, ParM, RibM, AprM, TobM1, IstM, GenM1, ForM, IstM, KanM1,	pfam00534 COG0438	α-1,4- glycosyl transferase I
α-1,6-glycosyltransferase (neutral sugar transferase)	KanM2, TobM2, GenM2 COG0438	pfam00534	α-1,6- glycosyl transferase II
ribosyltransferase	BtrL, NeoL, RibL, ParL, LivL	COG1926	
ribosyltransferase	BtrA, NeoA, RibA, ParA, LivA		
glycosyltransferase (hexosaminyltransferase)	NeoF, RibF, ParF, LivF	COG0438	glycosyl transferase III
6'-dehydrogenase (FAD)	BtrQ, NeoQ, RibQ, ParQ, LivQ, AprQ, TobQ, KanQ, GenQ, ForD2, IstQ	pfam05199 COG2303	
6' or (6''')-aminotransferase	BtrB, NeoB, LivB, ParB, RibB, AprB, TobB, KanB, GenB1, GenB2, GenB3, ForB, GenB4, IstB	pfam00202 COG0001	sugar aminotransferase (AT type III)
(N-acetyl)- hexosaminyl- -deacetylase or other amidase (LmbE-family)	BtrD, NeoD, LivD, ParD, RibD, AprN, TobD1, KanD1, ForD, GenL, IstD	pfam02585 COG2120	either UDP-D-glucos- -amine synthase or deacetylating amidase

3'-dehydratase	LivY, AprD3, GenD3	pfam00106 COG1028	3'-dehydroxylation
3'-oxidoreductase (Fe-S) oxidoreductase	LivW, AprD4	COG1032	3'-dehydroxylation
3''-dehydrogenase (NAD)	TobD2, KanD2, GenD2, ForC	pfam1408 COG0673	keto-sugar/cyclitol dehydrogenase
(Fe-S)-oxidoreductase (SAM)	BtrN, RibN, ParN, LivN, ForL, IstL, IstL2, LivA, ForO	pfam04055 COG0535	dehydrogenases or epimerases
6'-C-methyltransferase (SAM)	GenK, ForK	pfam04055 smart00729	only in GM and FM
6'/7'-N-methyltransferase (SAM)	ForG, IstG, AprI, GenN	pfam06325 COG0500	N-methyltransferase
3-O-methyltransferase	ForN, IstN	pfam05711	O-methyltransferase
sugar-phosphate- -phosphatase	BtrP, NeoP, ParP, LivP, RibP, AprZ, HygU, ForA	COG0406	ForA: D-*myo*-inositol- -3-phosphate- -phosphatase
NDP-heptose/hexose dehydrogenase /epimerase	HygK, AprD1	pfam01370 COG1087	unknown oxidoreductase

Tab. 3.4 Transporter proteins encoded by genes conserved among the ACAGA gene clusters analysed

Key enzyme (postulated function)	**Members**	**pfam/COG**	**Remarks/ General function**
AGA-exporter	NeoT, RibT, ParT, LivT, HygV, AprV	pfam00664 COG1132	ABC-transporter (ATP)
AGA-exporter	NeoU, RibU, ParU, LivU, HygW, AprW	pfam00005	ABC-transporter (ATP)
AGA-exporter	GenY, ForY, FosD, KanX	pfam00999 COG0475	efflux protein antiporter
AGA-exporter (permease)	TobT, KanT, GenT, IstI	COG0477	efflux protein
AGA-exporter	GenI, GenH, ForJ, ForH, ForI, IstJ, IstH	COG2271	efflux protein
AGA-exporter	KanZ	pfam00083 COG4932	efflux protein

Tab. 3.5 Regulatory proteins encoded by genes conserved among the ACAGA gene clusters analysed

Key enzyme (postulated function)	Members	pfam/COG	Remarks/ General function
response regulator system (DNA-binding protein)	NeoI, RibI, ParI, LivI, HygI, KanI	cd00093.1 CAD60534	possible component of 3 component sensor/response regulator system; similar to Cinorf12
response regulator system (sensor kinase)	NeoH, RibH, ParH, LivH, HygH, KanH	CAD60535	possible component of 3 component sensor/response regulator system; similar to Cinorf13
response regulatory system (trans-membrane protein of sensor complex)	NeoG, RibG, ParG, LivG, HygG, KanG	CAD60536	possible component of 3 component sensor/response regulator system; similar to Cinorf14
regulatory proteins	NeoX, RibX, ParX, LivX		weak evidence
regulatory proteins	NeoR, AprR1, AprR2, KanR, GenU, IssC, KanU		weak evidence

Tab. 3.6 Resistance proteins encoded in the ACAGA gene clusters analysed

Key enzyme (postulated function)	Members	pfam/COG	Remarks/ General functions
APH(3')	AphA, Rph, GenP, ForP, IstP	pfam01636 COG3281	AphA inactivate ACAGAs inside the cells; GenP, ForP, IstP are biosynthetic enzymes (3'-deydroxylation)
APH(7'') AGA resistance	HygA	pfam01636 COG3281	inactivate ACAGAs inside the cells
AAC(3) AGA resistance	AacC8 (SfrF04.15c), AacC (SribP10.14) AacC7 (SriA13.2)	pfam02522 COG2746	inactivate ACAGAs inside the cells
AAC(6') AGA resistance	KanA, AacA (SriA13.3)	pfam00583 COG2746	inactivate AGA inside the cells
16S rRNA methyltransferases	KamB, Kmr, GmrA, GmrB, FmrR, FmrO, FmrP, FmrB, ImrA	pfam07091 COG1670	modify target site

Tab. 3.7 Transposase proteins encoded in the ACAGA gene clusters analysed

Key enzyme (postulated function)	Members	pfam/COG	Remarks/ General functions
transposase neighbouring	SribP10.8c	pfam1609	All are either or inserted in AGA gene clusters; some might be non-functional (fragmentary ORFs)
transposases	SribP10.9c, ShyG17.8, SteO08.8c, (TatA), SteO08.9c (TatB), SteO08.10c (TatC)	COG3293	
transposase	ShyG17.31c (HygZ)	pfam00665 COG2826	
transposase	MolI14.2c	COG3415	
transposases	SribL03.10c, SteK17.29		

B Studies on some selected ACAGA biosynthetic enzymes

3.4 Identification of the enzymes involved in the biosynthesis of 2DOS

Following the analysis and comparison of the studied ACAGA biosynthetic gene clusters, an attempt was made for expression, characterization and biochemical study of some key enzymes which are expected to be involved in the formation of the aminocyclitol (2DOS) moiety.

Starting from G-6-P, four enzymatic steps would be necessary to form 2DOS (Fig. 3.17): i- conversion of G-6-P to 2-deoxy-*scyllo*-inosose; ii- transamination of 2-deoxy-*scyllo*-inosose to 2-deoxy-*scyllo*-inosamine; iii- oxidation of 2-deoxy-*scyllo*-inosamine at the 1- position to form 1-keto-2,3-deoxy-3-amino-*scyllo*-inositol; iv- transamination of this keto intermediate to 2DOS.

Therefore, the enzymes KanC (2-deoxy-*scyllo*-inosose synthase), KanS1 (putative bifunctional enzyme; L-glutamine (PLP): 2-deoxy-*scyllo*-inosose aminotransferase [AT-I] and 1-keto-2,3-deoxy-3-amino-*scyllo* inositol aminotransferase [AT-II]) and KanE (putative aminocyclitol 1-dehydrogenase) encoded by the *kan*-cluster of *S. kanamyceticus* were analyzed. Of these enzymes, the enzymes related to KanC (Kudo *et al.*, 1999a; Ota *et al*., 2000; Kharel *et al*., 2004) and KanS1 (Lucher *et al*., 1989; Walker, 1995; Ahlert *et al*., 1997; Tamegai *et al*., 2002b) had been previously studied enzymatically in part. It was presumed that KanS1 was a bifunctional enzyme on the evidence presented by Lucher *et al*. (1989) and Walker (1995) that the aminotransferase steps in 2DOS biosynthesis are catalysed by a bifunctional, L-glutamine dependent aminotransferase and only one KanS1 homolog was conserved in all the 2DOS-ACAGA gene clusters analyzed. Only in the KM group, two homologues (KanS1, KanS2 and the related proteins) were found (see Sect. 3.2.2).

Furthermore, the analysis of ACAGAs gene clusters revealed the presence of only one conserved gene for a putative dehydrogenase (*kanE* and the related genes; Tab. 3.3). Therefore, it was predicted that the KanE and related proteins were involved in the 1-dehydrogenation of 2-deoxy-*scyllo*-inosamine. In order to reconstitute the 2DOS pathway, the three enzymes KanC, KanS1 and KanE were expressed individually and their enzymatic roles in aminocyclitol biosynthesis were studied.

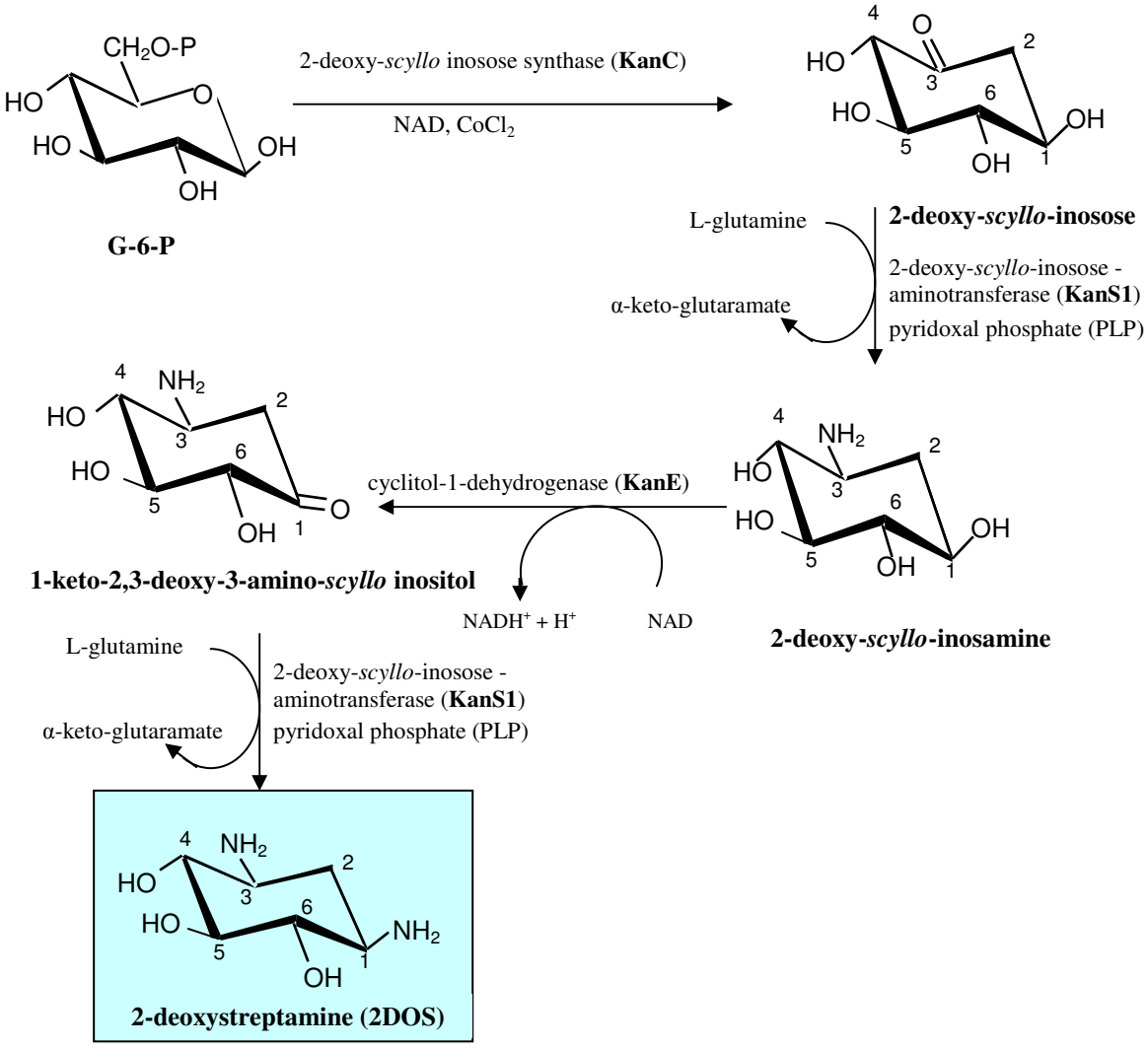

Fig. 3.17 Biosynthetic pathway for the formation of 2DOS in *S. kanamyceticus* DSM 40500.

3.5 Heterologous expression of the KanC, KanS1 and KanE proteins

3.5.1 The general strategy for the overexpression of the selected Kan-proteins

For overexpression, the individual genes were firstly amplified by PCR from the chromosomal DNA of *S. kanamyceticus* DSM 40500 using the appropriate primers and PCR conditions (cf. Tab. 2.4; Sect. 2.15). The forward primers were generally designed for the 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 suitable expression vectors. On the other hand, the reverse primers were designed for the introduction of an alternate

restriction site located immediately downstream of the natural stop codon in order to allow orientated cloning into the expression vectors. The amplified DNA fragments were cloned into i- pET16b (N-terminal His-tagged fusion protein) or pET11a (native protein) expression vectors of *E. coli* under the control of the *T7* promoter; ii- pUWL201PW *E. coli-Streptomyces* shuttle vector (native protein) under the control of the *ermE* up promoter.

3.5.2 Heterologous expression in *E. coli* under *T7* promoter

A set of recombinant plasmids was constructed (cf. Tabs. 2.2 & 2.3) to express *kanC*, *kanS1* and *kanE* genes in either of the *E. coli* BL21 (DE3), JM109 (DE3) or BL21 (DE3) Rosetta strain under the control of the *T7* promoter as described in Sect. 2.17.1 (cf. Tab. 2.1). The cloning of the corresponding genes and the construction of the expression plasmids are described in Sect. 2.4. The results of the expression of the three proteins under different conditions are summarized in Tab. 3.8.

Tab. 3.8 Summary of expression of KanC, KanS1 and KanE proteins

Protein	Vector (Promoter)	Host strain	Culture condition, time of induction	cMW (kDa)	Solubility [a]
KanC	pKC16b1 (*T7*)	*E. coli* BL21 (DE3)	LB, 37°C, 6 hrs	43,401	±
	pKC16b1 (*T7*)	*E. coli* BL21 (DE3)	LB, 37°C, 6 hrs	43,401	±
	pKC16b1 (*T7*)	*E. coli* BL21 (DE3)	LBS, 30°C, 12 hrs	43,401	+ (WB)
	pKC16b1 (*T7*)	*E. coli* JM109 (DE3)	LBS, 30°C, 12 hrs	43,401	+ (WB)
KanS1	pKS16b1 (*T7*)	*E. coli* JM109 (DE3)	LB, 37°C, 4 hrs	47,709	- (WB)
	pKS16b1 (*T7*)	*E. coli* JM109 (DE3)	LB, 30°C, 4 hrs	47,709	+ (WB)
KanE	pKE16b1 (*T7*)	*E. coli* BL21 (DE3)	LB, LBS, 30°C, 2 hrs	37,654	-
	pKE16b1 (*T7*)	*E. coli* JM109 (DE3)	LB, LBS, 30°C, 2 hrs	37,654	-
	pKE16b1 (*T7*)	*E. coli* Rosetta	LB, 37°C, 6 hrs	37,654	-
	pKE16b1 (*T7*)	*E. coli* Rosetta	LBS, 30°C, 12 hrs	37,654	+ (WB)
	pKE11a-4 (*T7*)	*E. coli* Rosetta	LBS, 30°C, 2, 4 and 12 hrs	35,881	no
	pKEW1-2 (*ermE*)	*S. lividans*TK23	YEME, 30°C, 48 hrs	35,881	+

[a]: + = soluble, - = insoluble, ± = partially soluble; no = not observed; (WB) = detection in Western blot assays; cMW = calculated molecular weight; LBS = LB medium supplemented with 1 M sorbitol and 2.5 mM betaine. Extracts from the strains underlined were used for enzyme assays.

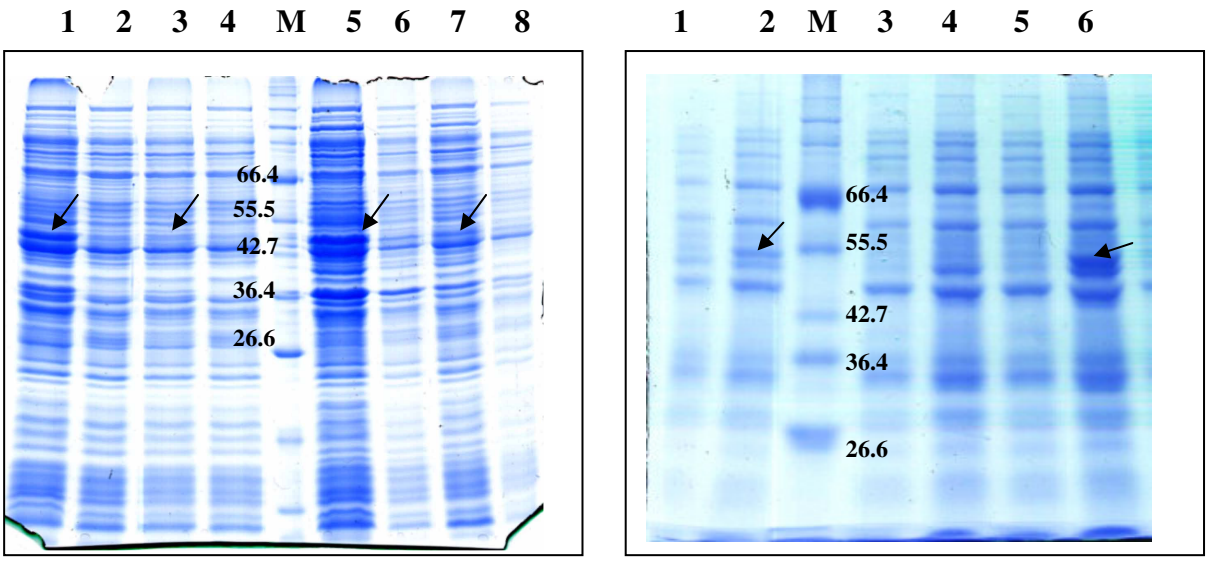

Fig. 3.18 SDS-PAGE analysis of the production of the KanC, KanS1 and KanE proteins in different *E. coli* strains. Only the soluble proteins were analyzed on 12% PAA gels. A: The production of KanC in *E. coli* JM109 (DE3; lanes 1-4) and in *E. coli* BL21 (DE3; lanes 5-8)**.** The cell-free extracts were obtained from strains harboring the following plasmids: pET16b (control; lanes 2, 4, 6 & 8); pKC16b1 (KanC; lanes 1, 3, 5, & 7); lanes 3, 4, 7 & 8 were 6 hrs after induction; lanes 1, 2, 5 & 6 were 12 hrs after induction. B: The production of KanS1 in *E. coli* JM109 (DE3). The cell-free extracts were obtained from strains harboring the following plasmids: pET16b (control; lanes 3 & 4); pKS16b1 (KanS1; lanes 1, 2, 5 and 6); lanes 1, 3 and 5 were before induction; lanes 2, 4 and 6 were 4 hrs after IPTG induction. The arrows mark additional bands corresponding to the expected size of proteins.

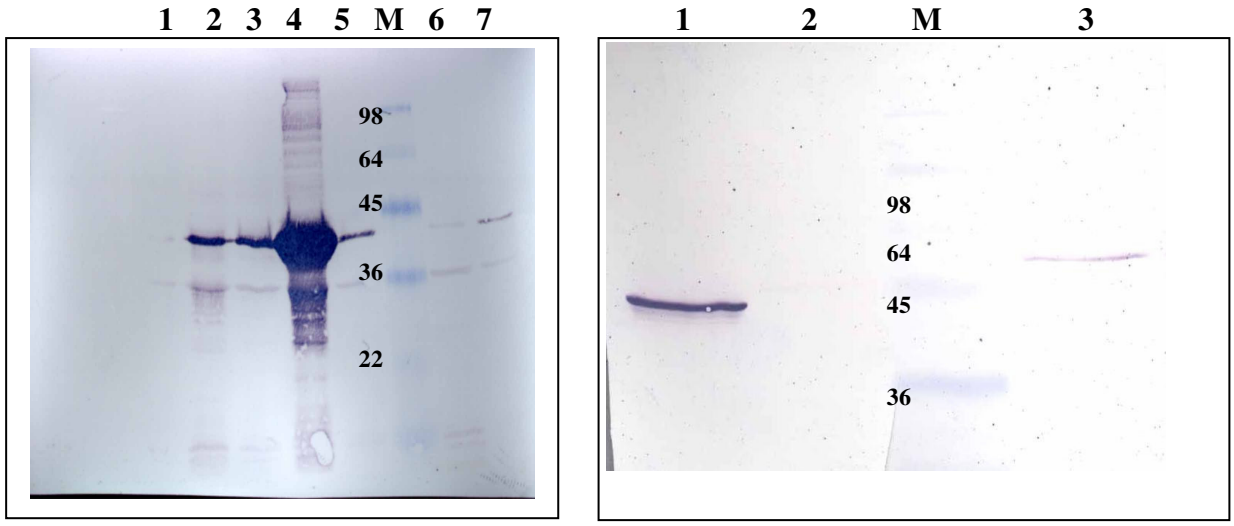

Fig. 3.19 Detection of the His-tagged proteins KanC, KanS1 and KanE by Western Blotting. The Western blotting was carried out as described in Sect. 2.20. A: Detection of N-terminal His-tagged KanE protein: Soluble cell-free extracts were from *E. coli* BL21 (DE3) Rosetta harboring the following plasmids: pET16b (control; lanes 1 & 6); pKE16b1 (KanE; lanes 2, 3, 5 & 7); and pKE16b1 (KanE; crude extract; lane 4). B: Detection of N-terminal His-tagged KanS1 protein: soluble cell-free extracts were from *E. coli* JM109 (DE3) harboring the following plasmids: pET16b (control; lane 2); pKS16b1 (KanS1; lane 3); and pKC16b1 (KanC; lane 1). The molecular masses of the marker proteins (lane M) are indicated in kDa.

KanC was overproduced as a soluble N-terminal His-tagged protein in both *E. coli* BL21 (DE3) and JM109 (DE3): An additional band of about 43 kDa present in the soluble fraction of the cell-free extract corresponded in size to the expected molecular mass (43.4 kDa) of the His-tagged KanC protein was detected. The maximal overexpression of the soluble KanC protein was obtained using LB medium with 1 M sorbitol and 2.5 mM betaine at 30°C and 12 hrs after IPTG induction. Also KanS1 was overproduced as a soluble N-terminal His-tagged protein in *E. coli* JM109 (DE3) where an additional band of about 47 kDa present in the soluble fraction corresponded in size to the expected molecular mass (47.7 kDa) of the KanS1 protein (cf. Fig. 3.18). The maximal overproduction of the soluble KanS1 protein was obtained 4 hrs after IPTG induction using LB medium at 30°C. The presence of multiple His residues in the overexpressed N-terminal His-tagged KanC and KanS1 proteins were also demonstrated by Western blot assays (Fig. 3.19). On the other hand, KanE was constantly produced in insoluble form (inclusion bodies) in the crude extract of the three selected *E. coli* strains using LB medium and various incubation temperatures. An additional band that corresponded to the deduced molecular mass of KanE protein (37 kDa) was only observed in the crude extract and not in the cell-free extract; also no positive signal could be detected in a Western blot assay. Nevertheless, KanE was successfully overexpressed as soluble N-terminally His-tagged protein in *E. coli* Rosetta cells using LB medium supplemented with 1 M sorbitol and 2.5 mM betaine at 30°C and 12 hrs after IPTG induction (cf. Tab. 3.8). The soluble KanE could also be detected by a Western blot (cf. Fig. 3.19).

3.5.3 Heterologous expression of KanE in *S. lividans* TK23

The recombinant plasmid pKEW1-2 was constructed for the expression of KanE in *S. lividans* TK23 as a native protein under the control of the *ermE* up promoter. KanE was successfully overproduced as a soluble native protein. An additional band of about 35 kDa, corresponding in size to the expected molecular mass (35.8 kDa) of the KanE protein was observed in the soluble fraction of the cell-free extract *S. lividans* TK23/pKEW1-2. Maximum expression of KanE in *S. lividans* TK23 was noted after 2 days of incubation at 30°C (Fig. 3.20).

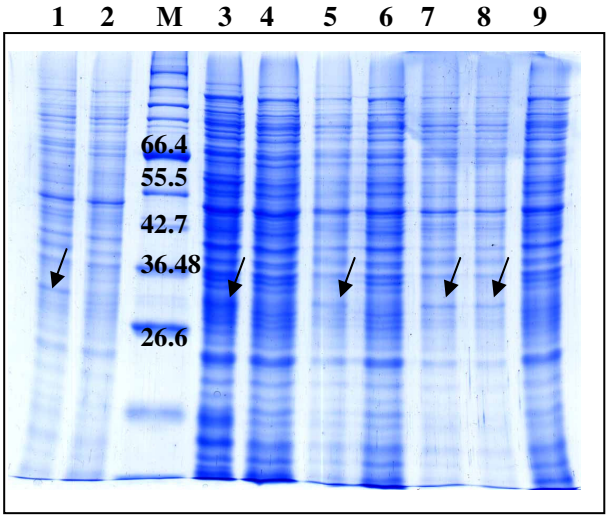

Fig. 3.20 SDS-PAGE analysis of the production of KanE in *S. lividans* TK23. Protein separation was done on 12% PAA gel. Cell extracts were obtained from *S. lividans* TK23 harboring the following plasmids: pUWL201 (control; lanes 2, 4, 6 and 9), pKEW1-2 (KanE; lanes 1, 3, 5, 7, and 8). Lanes 1, 2, 7, 8, and 9 contained cell-free extracts (soluble proteins); lanes 3, 4, 5, and 6 contained crude cell extracts. The molecular masses of the marker proteins (lane M) are indicated. The arrows mark a protein band corresponding to the molecular mass of the native KanE protein.

3.6 Biochemical characterization of the KanC, KanS1 and KanE proteins

3.6.1 Biochemical analysis of the KanC protein

Because of the stability, solubility and purification problems of the first intermediate/enzyme substrate, 2-deoxy-*scyllo*-insose as provided by chemical synthesis, as well as the absence of commercial supply, an enzymatic preparation process for synthesizing this substance seemed best fitting the purpose. Such performing a KanC-catalysed reaction was not only carried out to prove again its enzymatic function but also to provide the natural substrate for the next enzymatic steps preparatively.

The soluble fractions containing His-tagged KanC protein from either *E. coli* BL21 (DE3) or *E. coli* JM109 (DE3) were used for this purpose (cf. Tab. 3.8). 2-deoxy-*scyllo*-inosose synthase (KanC) was tested as described in Sect. 2.23. Conversion of G-6-P into 2-deoxy-*scyllo*-inosose was detected on both TLC and HPLC (Sects. 2.21 and 2.22).

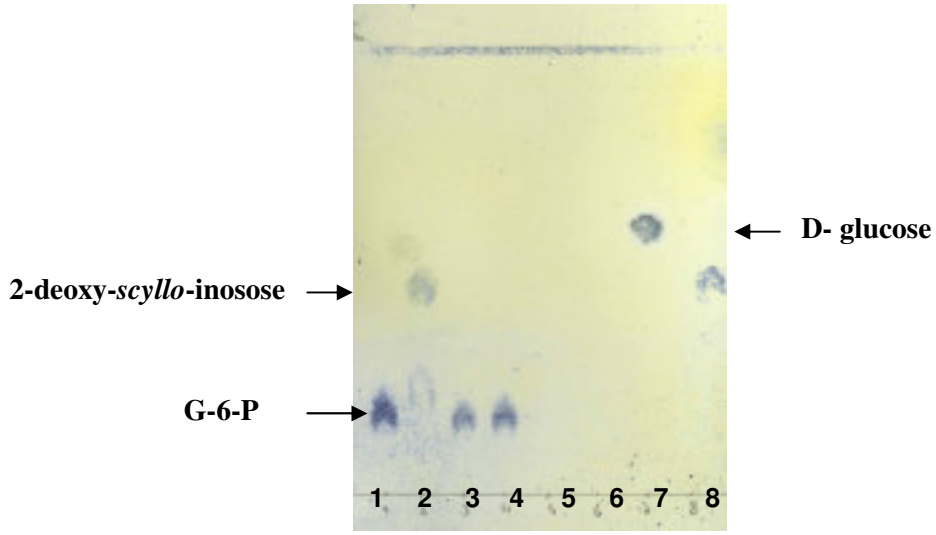

Fig. 3.21 TLC analysis of KanC catalysed formation of 2-deoxy-*scyllo*-inososse. Relevant substances were visualized using the Cer-reagent (Sect. 2.3.10). The following samples were applied to the TLC: lane 1: G-6-P; lane 2: *E. coli* BL21 (DE3)/pKC16b1 (KanC) with G-6-P; lane 3: *E. coli* BL21 (DE3)/pET16b (control) with G-6-P; lane 4: assay with cooked cell-free extract (100°C for 3 min) of *E. coli* BL21 (DE3)/pKC16b1 (KanC) with G-6-P; lane 5: *E. coli* BL21 (DE3)/pKC16b1 (KanC) in absence of G-6-P; lane 6; *E. coli* BL21 (DE3)/pET16b (control) in absence of G-6-P; lane 7: D-glucose; lane 8: standard 2-deoxy-*scyllo*-inososse.

A new spot on the TLC with an Rf-value (0.46) corresponding to the migration of the standard 2-deoxy-*scyllo*-inososse was observed in those enzyme reactions which contained soluble His-tagged KanC protein. In all the control reactions used, no conversion of G-6-P to 2-deoxy-*scyllo*-inososse was ever observed (Fig. 3.21).

In a HPLC separation, also a new peak with a retention time (6.4 -6.7 min) to that obtained with standard 2-deoxy-*scyllo*-inososse was seen in the same enzyme reaction mixture (Fig. 3.22). Again, this new peak did not occur in the control assays. Maximal conversion (about 80%) of G-6-P to 2-deoxy-*scyllo*-inososse was observed when the KanC reaction was incubated for 12 hrs at 30°C. It was also noticed that presence of the cofactors Co^{++} and NAD was essential in the KanC enzyme reaction. This proved that KanC expressed from the plasmid pKC16b1 in *E. coli* in fact had a 2-deoxy-*scyllo*-inososse synthase activity.

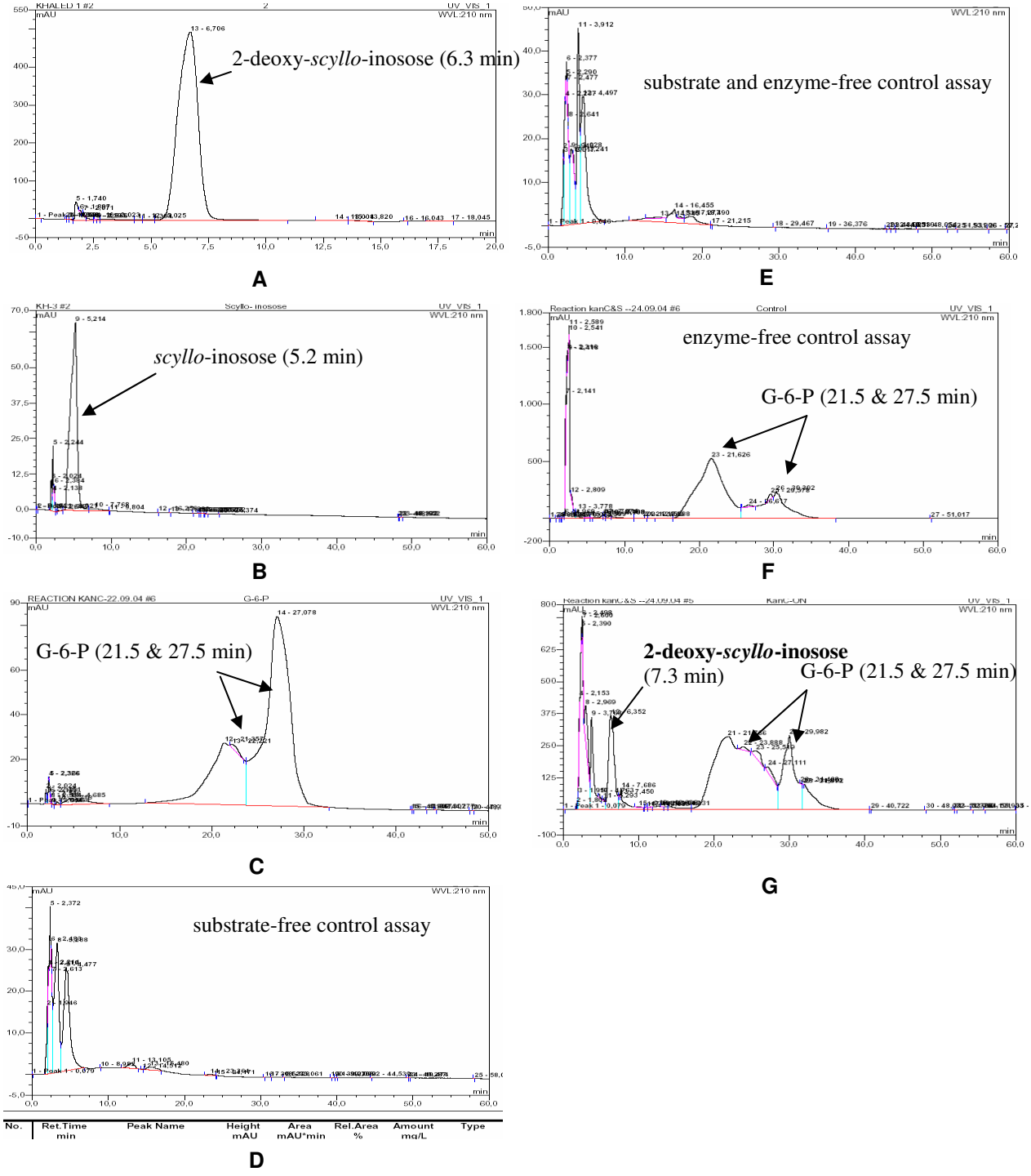

Fig. 3.22 HPLC analysis of 2-deoxy-*scyllo*-inosose formation by KanC. HPLC run was carried out as described in Sect. 2.22. A: standard 2-deoxy-*scyllo*-inosose (10 mM); B: standard: *scyllo*-inosose (8 mM); C: standard G-6-P (15 mM); D: *E. coli* BL21 (DE3)/pKC16b1 (KanC) without G-6-P; E: *E. coli* BL21 (DE3)/pET16b (control) without G-6-P; F: *E. coli* BL21 (DE3)/pET16b (control) in presence of G-6-P; G: *E. coli* BL21 (DE3)/pKC16b1 (KanC) in presence of G-6-P. Arrows indicate the peak for each substance and retension time are given between brackets.

3.6.2 Biochemical analysis of the KanS1 protein

The L-glutamine-dependent transamination of (2-deoxy-)*scyllo*-inosose to (2-deoxy-)*scyllo*-inosamine is the second enzymatic step in 2DOS biosynthesis which was previously proven in extracts from various AGA producers (Lucher *et al*., 1989; Walker, 1995; Ahlert *et al*., 1997; Tamegai *et al*., 2002b). Two genes (*kanS1* & *kanS2*) encoding related protein species belonging to this ketocyclitol aminotransferase family namely KanS1 (40% identity to BtrS from *B. circulans*; 54% identity to StsC from *S. griseus*) and KanS2 (32% identity to BtrS from *B. circulans*; 36% identity to StsC) proteins were present in the *kan*-cluster of *S. kanamyceticus* DSM 40500 (see Sect. 3.2.2). Since StsC and BtrS proteins were formerly characterized to be *scyllo*-inosose and 2-deoxy-*scyllo*-inosose aminotransferases, respectively (Ahlert et., 1997; Tamegai *et al*., 2002b), it was expected that the KanS1 protein represented the bifunctional ketocyclitol aminotransferase involved in the biosynthesis of 2DOS (Figs. 3.17 & 3.23). According to this hypothesis, KanS2 was expected (and had to be proven) not to be involved in 2DOS biosynthesis. This transamination process would be carried out in the presence of PLP as a coenzyme as well as L-glutamine as an aminodonor substrate. The cell-free extract (soluble fractions) of *E. coli* JM109 (DE3) harboring the plasmid pKS16b1 (KanS1; cf. Tab. 3.8) was tested for 2-deoxy-*scyllo*-inosose aminotransferase activity. Because of stability problems with the chemically synthesized 2-deoxy-*scyllo*-inosose, a coupled assay of KanC and KanS1 was necessary. By this way, the 2-deoxy-*scyllo*-inosose enzymatically formed in the KanC reaction in the coupled assay acted as a substrate for the KanS1 enzyme (see Sect. 3.6.1).

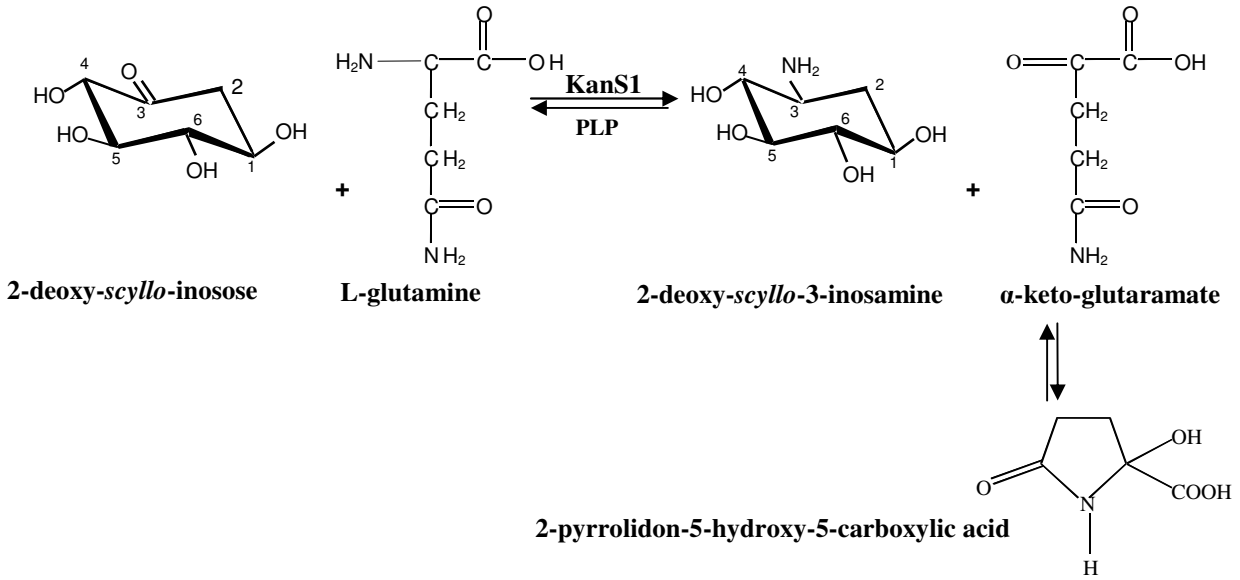

Fig. 3.23 First transamination reaction catalysed by the KanS1 enzyme using L-glutamine as aminodonor and 2-deoxy-*scyllo*-inosose as aminoacceptor.

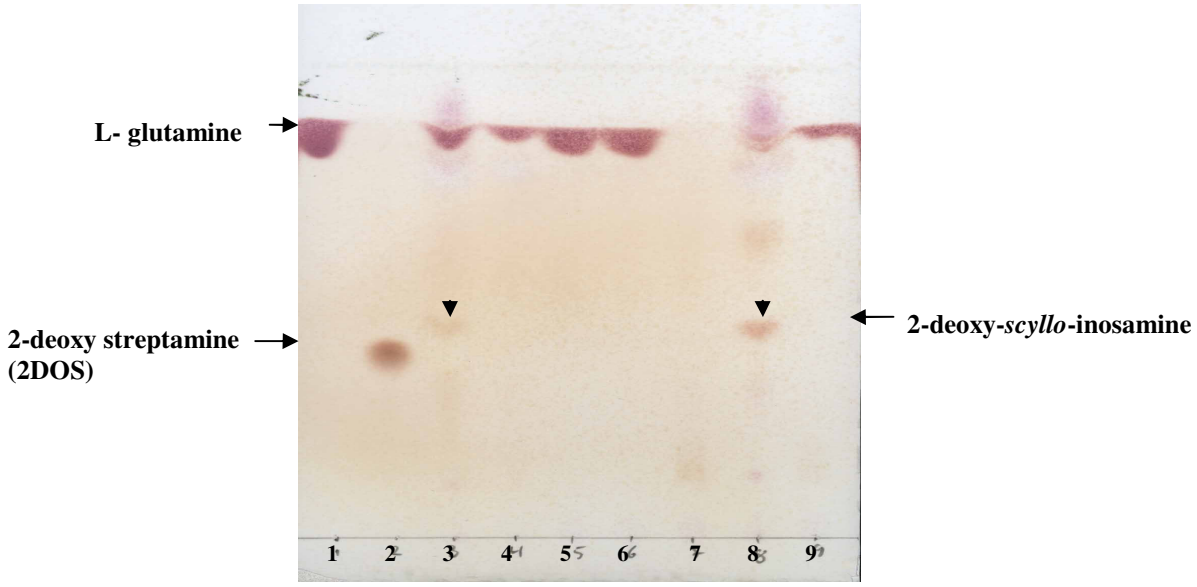

Fig. 3.24 TLC detection of 2-deoxy-*scyllo*-inosamine formed from G-6-P and L-glutamine in the coupled reaction catalysed by KanC and KanS1. The enzyme reactions were performed with cell-free extracts (Tab. 3.8) as described in Sect. 2.23. TLC was developed as described in Sect. 2.21. Spots were visualized using ninhydrin solution (Sect. 2.3.10). Lane 1: L-glutamine; lane 2: 2DOS; lane 3: 2-deoxy-*scyllo*-inosose and *E. coli* JM109 (DE3)/pKS16b1 (KanS1); lane 4: *E. coli* BL21 (DE3)/pKC16b (KanC) and *E. coli* JM109 (DE3)/pET16b (KanS1-free control); lane 5: *E. coli* BL21 (DE3)/pKC16b1 (KanC) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1) without G-6-P; lane 6: *E. coli* BL21 (DE3)/pET16b (KanC-free control) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1) without G-6-P; lane 7: *E. coli* BL21 (DE3)/pKC16b1 (KanC) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1) without L-glutamine (substrate-free control); lane 8: *E. coli* BL21 (DE3)/pKC16b1 (KanC) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1); lane 9: BL21 (DE3)/pKC16b1 (KanC) plus L-glutamine but in the absence of *E. coli* JM109 (DE3)/pKS16b1 (KanS1-free control).

As shown in Figs. 3.24 and 3.25, both 2-deoxy-*scyllo*-inosamine and α-keto-glutaramate were detectable on the TLC plates only from the coupled assays including active KanC and KanS1 enzymes in the presences of G-6-P, L-glutamine and all the co-factors necessary. In these separations, the Rf-value for 2-deoxy-*scyllo*-inosamine was 0.54 and that for α-keto-glutaramate was 0.83. Maximum conversion of 2-deoxy-*scyllo*-inosose to 2-deoxy-*scyllo*-inosamine by KanS1 was observed when the reaction was incubated at 37°C for 12 hrs. In addition, α-keto-glutaramate is unstable intermediate which easily cyclizes to 2-pyrrolidone-5-hydroxy-5-carboxylic acid (Fig. 2.23; Alert *et al*., 1997). A new peak with a retention time of 13.9 min was observed by HPLC analysis of the KanC/KanS1 reaction assays (Fig. 3.26).

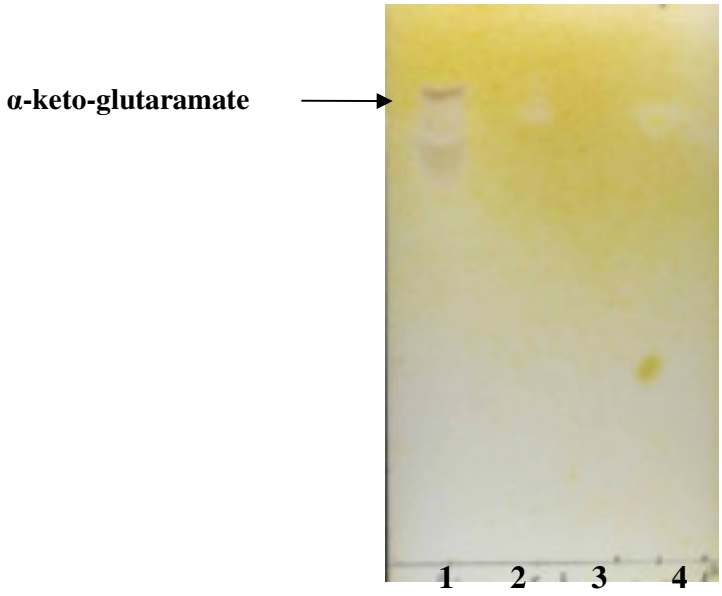

Fig. 3.25 TLC detection of α-keto-glutaramate formed in the coupled reaction catalysed by KanC and KanS1. The enzyme reactions were performed with cell-free extracts (Tab. 3.8) as described in Sect. 2.23. The TLC plate was developed as described in Sect. 2.21. Spots were visualized using the Ehrlich reagent (Sect. 2.3.10). lane 1: *E. coli* BL21 (DE3)/pKC16b1 (KanC) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1); lane 2: *E. coli* BL21 (DE3)/pKC16b1 (KanC) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1) without L-glutamine; lane-3: *E. coli* BL21 (DE3)/pET16b (control) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1); lane 4: L-glutamine. The position of α-keto-glutaramate is indicated by the arrow.

In control assays lacking either one of the enzymes or one of the two substrates, this peak was not produced (Fig. 3.26). It was assumed that this peak represented α-keto-glutaramate. Because of its instability, α-keto-glutaramate is not available commercially and could not be used as a standard for the HPLC analysis.

Furthermore, the KanS1 activity was also measured spectrophotometrically as described in Sect. 2.24 with L-alanine as an aminodonor and measuring the decrease of NADH catalysed by L-lactate dehydrogenase (Fig. 3.27). A significant decrease in the absorbance in the reaction containing KanS1 protein and all the other co-substrates in comparison to the other controls was observed (Fig. 3.27). The specific activity of KanS1 was calculated to be 0.27 U/mg protein (cf. Sect. 2.24).

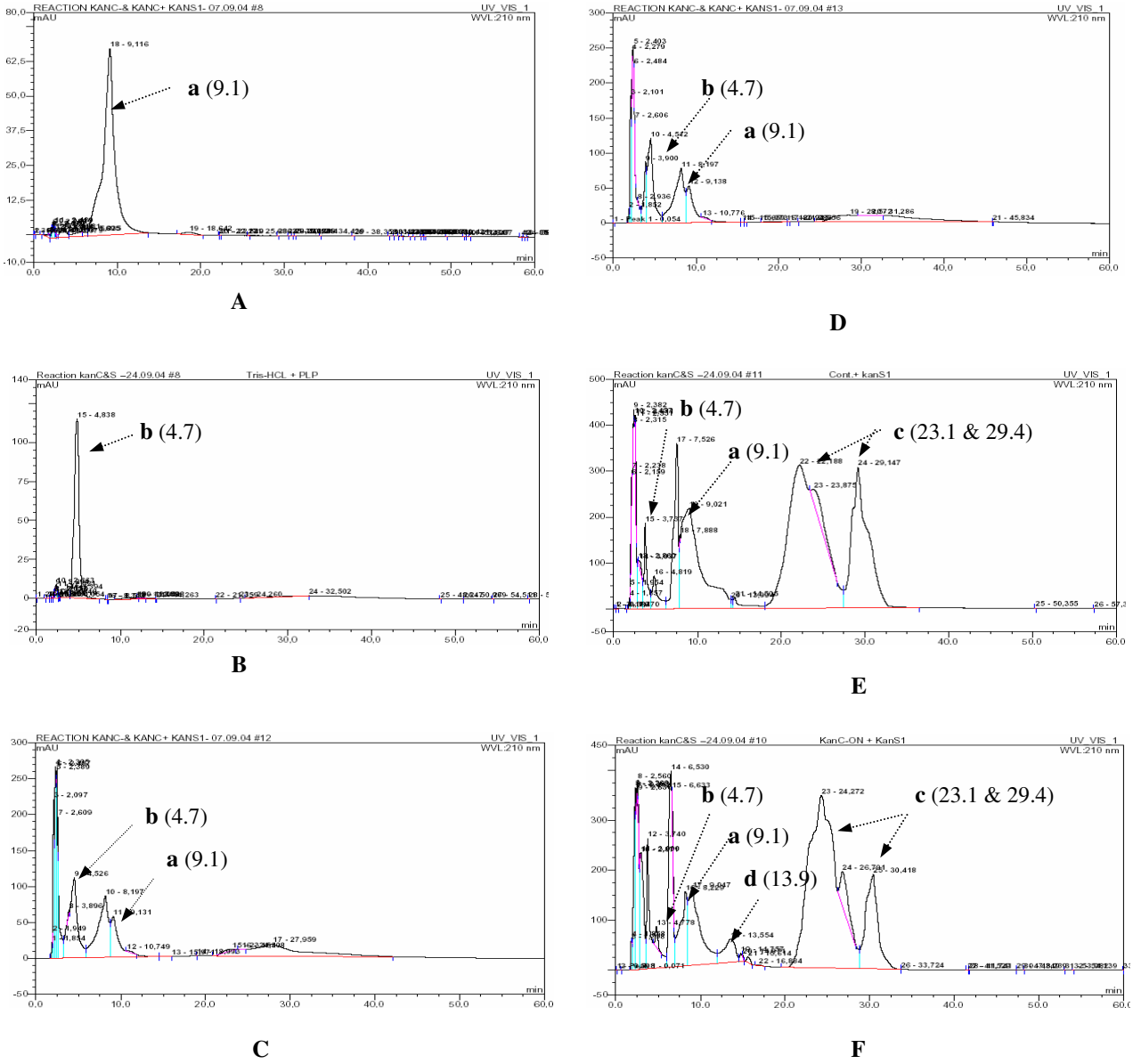

Fig. 3.26 HPLC analysis of the KanC/KanS1 assays for the production of 2-deoxy-*scyllo*-inosamine. The enzyme reaction was performed using cell-free extracts (Tab. 3.8) as previously described in Sect. 2.24. The HPLC run was carried out as described in Sect. 2.22. A: standard L-glutamine (30 mM); B: Tris-HCL, pH 7.5 (30 mM) + 1 mM PLP; **C**: *E. coli* BL21 (DE3)/pKC16b1 (KanC) without G-6-P and *E. coli* JM109 (DE3)/pKS16b1 (KanS1); D: *E. coli* BL21 (DE3)/pET16b (control) without G-6-P and *E. coli* JM109 (DE3)/pKS16b1 (KanS1); E: *E. coli* BL21 (DE3)/pET16b (control) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1); F: *E. coli* BL21 (DE3)/pKC16b1 (KanC) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1). Arrows indicate the peaks corresponding to each substance (indicated by small letters) with its retention time (min; inbrackets). Small letters indicate; a- L-glutamine (30 mM), b- Tris-HCL (30 mM, pH 7.5 + 1 mM PLP), c- G-6-P (20 mM), d- α-keto-glutaramate.

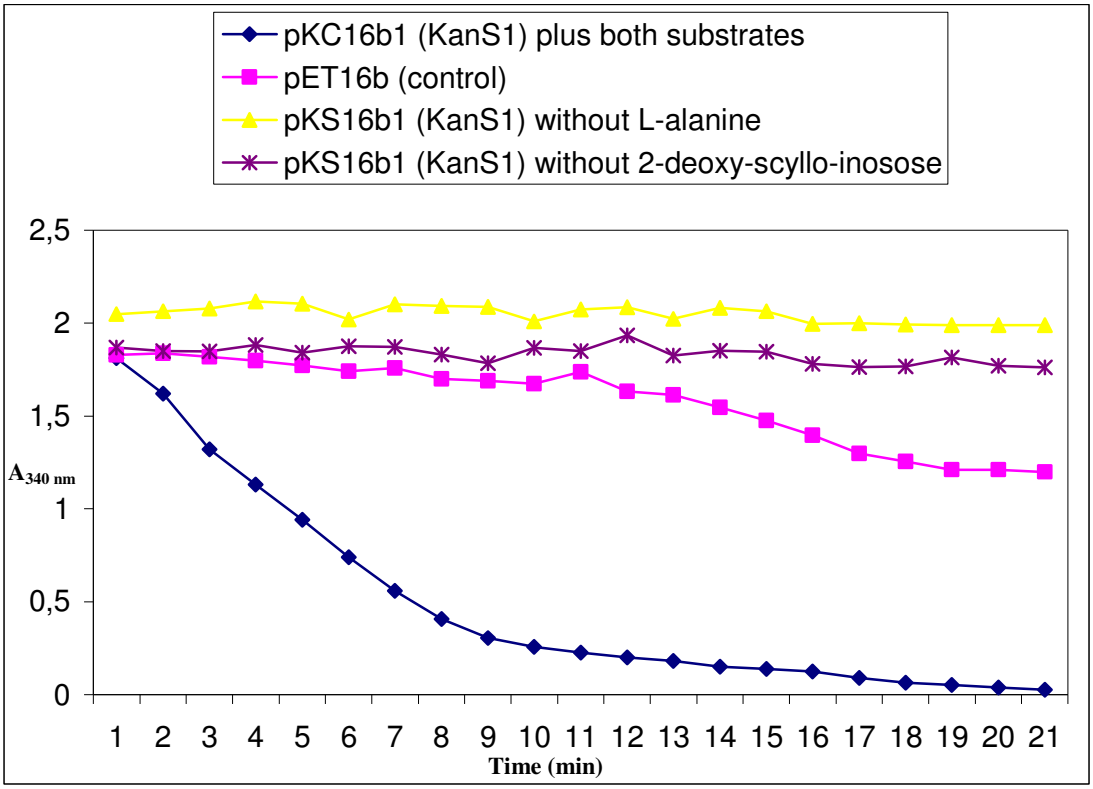

Fig. 3.27 Spectrophotometric assay for (NADH consumption) measurement of the KanS1 activity. The enzyme assay was performed as described in Sect. 2.24 using cell-free extracts (soluble fractions) of *E. coli* JM109 (DE3)/pKS16b1 (KanS1) and or of *E. coli* JM109 (DE3)/pET16 (enzyme-free control).

3.6.3 Biochemical analysis of the KanE protein

The determination of the KanE activity (putative 3-aminocyclitol 1-dehydrogenase) was carried out as described in Sect. 2.25. However, the lack of 2-deoxy-*scyllo*-inosamine as a pure substrate made the quantification of this assay difficult. Therefore, an assay containing KanE besides KanC and KanS1 was carried out starting from G-6-P as a substrate.

A weak spot on TLC with similar Rf value (0.42) to the standard 2DOS was only observed in the complete three enzyme assay, i.e. when extract of *S. lividans* TK23 harboring the plasmid pKEW1-2 (native KanE) was included (Fig. 3.28; lane 7). This spot was absent when the same reaction was done in absence of $ZnCl_2$ (cf. Fig. 3.28) or when the N-terminally His-tagged KanE protein was added instead of the native one. Additionally, spectrophotometeric measurement of NAD(P)H generation was tried for detecting the KanE activity (cf. Sect. 2.26). However, the activity could not be detected in the extracts because of a high background of NAD consumption which made the measurement of KanE activity impossible.

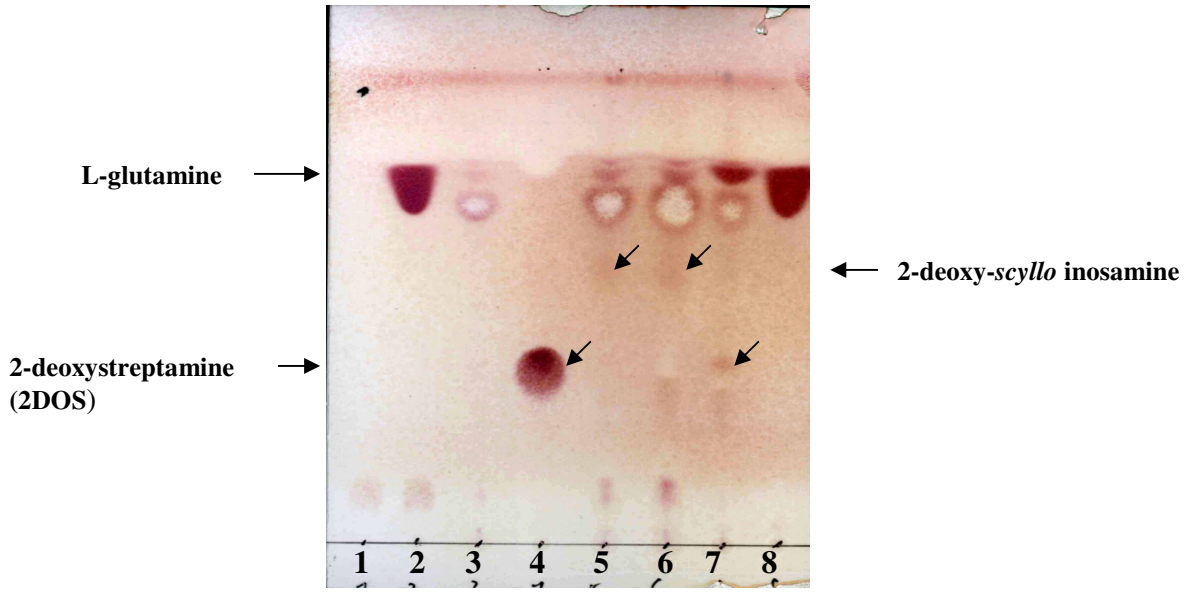

Fig. 3.28 TLC detection of 2-deoxystreptamine (2DOS) formed by the 3-enzymes catalysed reaction. The coupled enzyme reaction for KanC, KanS1 and KanE was performed using cell-free extracts (Tab. 3.8) as previously described in Sect. 2.25. TLC was developed as described in Sect. 2.21. Spots were visualized using the ninhydrin solution (Sect. 2.3.10). Lane 1: KanC/KanS1 (Sect. 2.24) without L-glutamine; lane 2: *E. coli* BL21 (DE3)/pET16b (vector control) and *E. coli* JM109 (DE3)/pKS16b1 (KanS1); lane 3: KanC/KanS1 without G-6-P; lane 4: standard 2DOS; lane 5: KanC/KanS1 and *S. lividans* TK23/pUWL201 (KanE-free control); lane 6: KanC/KanS1 and *S. lividans* TK23/pKEW1-2 (native KanE; complete assay) without $ZnCl_2$; lane 7: KanC/KanS1 and *S. lividans* TK23/pKEW1-2 (native KanE; complete assay) in presence of $ZnCl_2$; lane 8: standard L-glutamine. The positions for L-glutamine, 2DOS and 2-deoxy-*scyllo*-inosamine are indicated by arrows.

4 Discussion

4.1 General structure and genomic location of the ACAGA biosynthetic gene clusters

The major objective of this work was to elucidate the genetics and use their comparison to set up proposals for the hypothetical pathways for the production of the major groups of 2DOS and related ACAGAs in their producing strains (cf. Tab. 3.1). The immediate interpretation of the structures and locations of the respective gene clusters analysed can be summarized as follows: (i) for the NM group, the four streptomycetes clusters (i.e. the *neo*-, *rib*-, *par*- and *liv*-clusters) are homogenous in both DNA composition (G+C content between 72.1% [*rib*] and 76.8% [*neo*]) and order of the genes (cf. Fig. 3.2). However, the *btr*-cluster from *B. circulans* largely deviates from this pattern (G+C content 50.3%; cf. Fig. 3.2); (ii) for the KM group, the three *kan*-, *tob*- and *gen*-clusters show also homogenous DNA composition (G+C content between 69.7% [*gen*] and 72.8% [*tob*]). However, the order of genes inside the clusters is only conserved in part; e.g. the subclusters *kan/tobS1CD2M2D1S2* are related to *genS1CD2M2D1S2* and the subcluster *tobEBQ* is related to *kanEBQ* (cf. Fig. 3.5); (iii) the genes for the two *apr*-clusters are highly conserved in their order, number of ORFs and their sequence identity is higher than 90% (G+C content 73.3%; cf. Fig. 3.7); (iv) the *hyg*-cluster contains a mixture of genes with a few similarities with several other ACAGA gene clusters or other sources of genes involved in rare sugar or oligosaccharide biosynthesis. The *hyg*-cluster is flanked by a transposase gene and contains three genes with good similarity to corresponding ones in the *spc*-cluster, i.e. *hygFMF* related to *spcFMY* (cf. Fig. 3.9); (v) the *for*- and *ist*-clusters share gene similarities to each other; however, the order of genes is not retained (cf. Fig. 3.11). As well, both clusters share the gene sets *for*/*istHJ* which correspond to *genHJ*; (vi) the surrounding genomic regions of all the clusters studied obviously are not conserved at all however, these regions are enriched with ORFs which are highly conserved in both the completely analysed streptomycete genomes (Bentley *et al*., 2002; Ikeda *et al*., 2003); (vii) the gene clusters for nearly all of the postulated biosynthetic pathways for the individual ACAGAs treated putatively contain all the necessary biosynthetic genes, except for BU and TM (cf. Figs. 3.2 & 3.5); (viii) the gene clusters analysed obviously are not located in a commonly conserved region on the respective linear chromosomes of their producer strains. There is evidence that all the gene clusters reside in the arms (i.e. the *neo*-, *liv*-, *par*- and *rib*-clusters seem to be located on a common area of one of the variable arms)

or close to their boundary with the core of the respective genome (e.g. *kan*-cluster). (ix) There is no clear evidence that any of the cluster is located in one of the conserved core regions of the respective linear chromosomes; the presence of an rRNA operon (*rrnE*) in close vicinity to the *gen*-cluster in *M. echinospora* does not contradict this suggestion, since also in the two completely analysed streptomycete genomes, two to three of the six rRNA operons lie outside the core regions (Bentley *et al*., 2002; Ikeda *et al*., 2003).

4.2 Proposed biosynthetic pathways for the major groups of 2DOS and related ACAGAs

4.2.1 Biosynthesis of 2DOS and related diaminocyclitols

Biosynthesis of 2DOS and related 3,6-diaminocyclitol intermediates is considered the first target during the biosynthesis of the 2DOS-ACAGAs (i.e. paromamine and non-paromamine containing ACAGAs) and IM (i.e. 2-deoxyfortamine-containing ACAGAs), respectively. Only, the first two biosynthetic steps in the biosynthesis of 2DOS had been biochemically analysed by identifying the BtrC-related enzymes (2-deoxy-*scyllo*-inosose synthases or cyclases; Kudo *et al.*, 1999a; Ota *et al*., 2000; Kharel *et al*., 2004) and the BtrS-related cyclitol aminotransferases (L-glutamine: ketocyclitol aminotransferases I and II; Lucher *et al*., 1989; Walker, 1995; Tamegai *et al*., 2002b). Analysis of the gene clusters of the 2DOS-ACAGAs revealed the conserved presence of a gene for a likely candidate aminocyclitol dehydrogenase (KanE and the related proteins); this lead to the prediction for this protein to be involved in the third step of 2DOS biosynthesis. As well, evidence was presented in the present study that KanE is involved in the dehydrogenation of the 2-deoxy-*scyllo*-inosamine to form the keto intermediate (1-keto-2,3-deoxy-3-amino-*scyllo*-inositol) which is then transaminated a second time by the same ketocyclitol aminotransferase (KanS1 and related proteins; bifunctional enzymes; cf. Fig. 3.17) to form 2DOS. The proposed biosynthetic pathway for 2DOS is illustrated in Fig. 4.1.

Both FTM and IM contain a 3,6-diaminocyclitol as basic cyclitol unit (cf. Fig. 1.5); however this cyclitol unit in each antibiotic originates from different routes of biosynthesis (cf. Figs. 1.8 and 4.6 below). Analysis of the respective clusters revealed the presence of genes that encode putative enzymes needed for the biosynthesis of this aminocyclitol: 1) In case of FTM: (i) two further enzymes are found, ForA (putative D-*myo*-inositol-3-P-phosphatase) and ForC (putative *myo*-inositol-3-dehydrogenase), are required to link the biosynthetic pathway of D-*myo*-inositol-3-P to the production of *scyllo*-inosose, the ketocyclitol substrate for the first aminotransferase

(ForS) to form *scyllo*-inosamine (the *myo*-inositol-derived *scyllo*-inosamine pathway); (ii) D-*myo*-inositol-3-P is a common primary metabolite in actinomycetes, where it is essential for the synthesis of the major thiol, mycothiol, and cell wall phospholipids (Newton *et al*., 2000; Piepersberg *et al*., 2002; Movahedzadeh *et al*., 2004); (iii) Another putative *scyllo*-inosamine-6-dehydrogenase (ForE; related to KanE) is also present; however, the difference is that the ForE enzyme does not act on the monomeric substrate, like the related enzymes in the 2DOS pathways (i.e. KanE). Rather it acts on a pseudodisaccharide (FTM-FU-10) to form the substrate for the next aminotransferase step, probably also being a bifunctional enzyme (ForS; see below Fig. 4.6). Whether the ForS aminotransferase is a bifunctional enzyme as yet has to be proven in future, e.g. by mutant and biochemical studies; 2) In case of IM: (i) this cyclitol unit is derived from a 2-deoxy-*scyllo*-inosamine intermediate, like in the 2DOS pathways, where IstC and IstS are involved in its formation; (ii) the following biosynthetic steps would be similar to those in the FTM pathway, where IstE (putative 2-deoxy-*scyllo*-inosamine-3-dehydrogenase) and IstS (bifunctional aminotransferase) act on the a pseudodisaccharidic intermediate (IST-FU-10; see below Fig. 4.6).

4.2.2 Biosynthesis of paromamine

The biosynthesis of the pseudodisaccharidic intermediate, paromamine starts with a glycosyltransfer reaction coupling a D-glucosamine unit glycosidically to the 2DOS unit. A gene for a conserved putative glucosaminyltranferase enzyme (NeoM and related proteins; cf. Tab. 3.3) is found in all the paromamine-containing ACAGAs producers, with high likeliness to be involved in this glycosylation process. In addition, the likely activated sugar as a cosubstrate for this glycosyltransfer reaction could be either UDP-D-glucosamine or UDP-N-acetyl-D-glucosamine. As recently suggested by Kudo *et al*. (2005), UDP-D-glucosamine could be this activated sugar because its formation is claimed to be catalysed by BtrD protein, encoded by a conserved gene in most of the ACAGA producers. This suggestion made by Kudo *et al*. (2005) still questionable since: i) the BtrD and related proteins show significant similarity to the MitE/LmbE family of possible N-acetylhexosamine deacetylases (pfam02585; COG2120) and other amidases (e.g. mycothiol conjugate amidase; NP215598). Especially, this deacetylating function would fit to the biosynthetic pathway of the respective ACAGAs; ii) the nature of the enzymatic function of the BtrD/LmbE/MitE family of proteins (COG2120) is different from that of the nucleotidyltransferses (or UDP-D-glucosamine synthases). All other NDP-sugar synthases

(pyrophosphorylases) are strongly similar to each other but in a totally different protein family unrelated to the BtrD-like protein. However, both functions could be required during the biosynthesis of the respective ACAGAs; iii) they tried to clarify this nucleotidyl transferase function for the BtrD protein via using a protein model structure of MshB protein (1-D-*myo*-inosityl-N-acetyl-D-glucosamine deacetylase) which is involved in a completely different catalytic function. The three-dimensional structures of MshB enzyme was completely analysed to be a zinc dependant hydrolyase as previously pointed out by Maynes *et al.* (2003). A protein model structure of a BtrD protein has to be used instead in order to establish a precise structural model; iv) Kudo *et al*. (2005) also made some attempts to test for the BtrD deacetylase activity by using UDP-GlcNAc a substrate. However, it has been reported that these type of LmbE-related enzymes act only on GlcNAc-containing conjugates and not on monomeric substrate (Newton *et al*., 2000). In our case, 2'-N-acetylparomamine could be the high likely substrate for the "D" enzymes and should be used for testing the deacetylating activity of these enzymes.

For these reasons, the predicted catalytic functions attributed to this conserved protein (BtrD-like protein) could be either a deacetylase or deaminase: a) a deacetylase for the N-acetyl-D-glucosamine (GlcNAc)-containing conjugates (e.g. 2'-N-acetylparomamine), if NDP-N-acetyl-D-glucosamine would be the donor substrate for the "M" enzymes. This predicted function is supported by the occurrence of a good similarity of "D" enzymes to the 1-D-*myo*-inosityl-N-acetyl-D-glucosamine (GlcNAc-Ins) deacetylase (MshB) from *Mycobacterium tuberculosis*, which is involved in the biosynthesis of mycothiol (MSH; Newton *et al*., 2000; Maynes *et al*., 2003). This deacetylating function of the BtrD protein would also fit to the biosynthetic pathways of the respective antibiotics, particularly in the formation of paromamine or paromamine-like intermediates. The involvement of BtrD like proteins in the formation of paromamine (in this case by its deacetylating activity) would also correlate to the data obtained with blocked mutants and biochemical feeding studies by Kudo *et al.* (2005). They disrupted the *btrD* gene in the wild type strain and the resulted *btrD* mutant showed no antibiotic production which confirmed that the *btrD* gene product is involved in the biosynthesis of BUs antibiotics. In order to determine which biosynthetic step was blocked, they carried out biochemical feeding studies for this mutant strain with various biosynthetic intermediates, such as 2-deoxy-*scyllo*-inosamine, 2DOS and paromamine. They found that only with paromamine, the *btrD* mutant strain was able to recover antibiotic production which proves the involvement of the BtrD enzyme in the production of paromamine; b) a deaminase for α-ketoglutaramate, where it converts the α-ketoglutaramate (unavoidable and unstable by product of the two L-glutamine

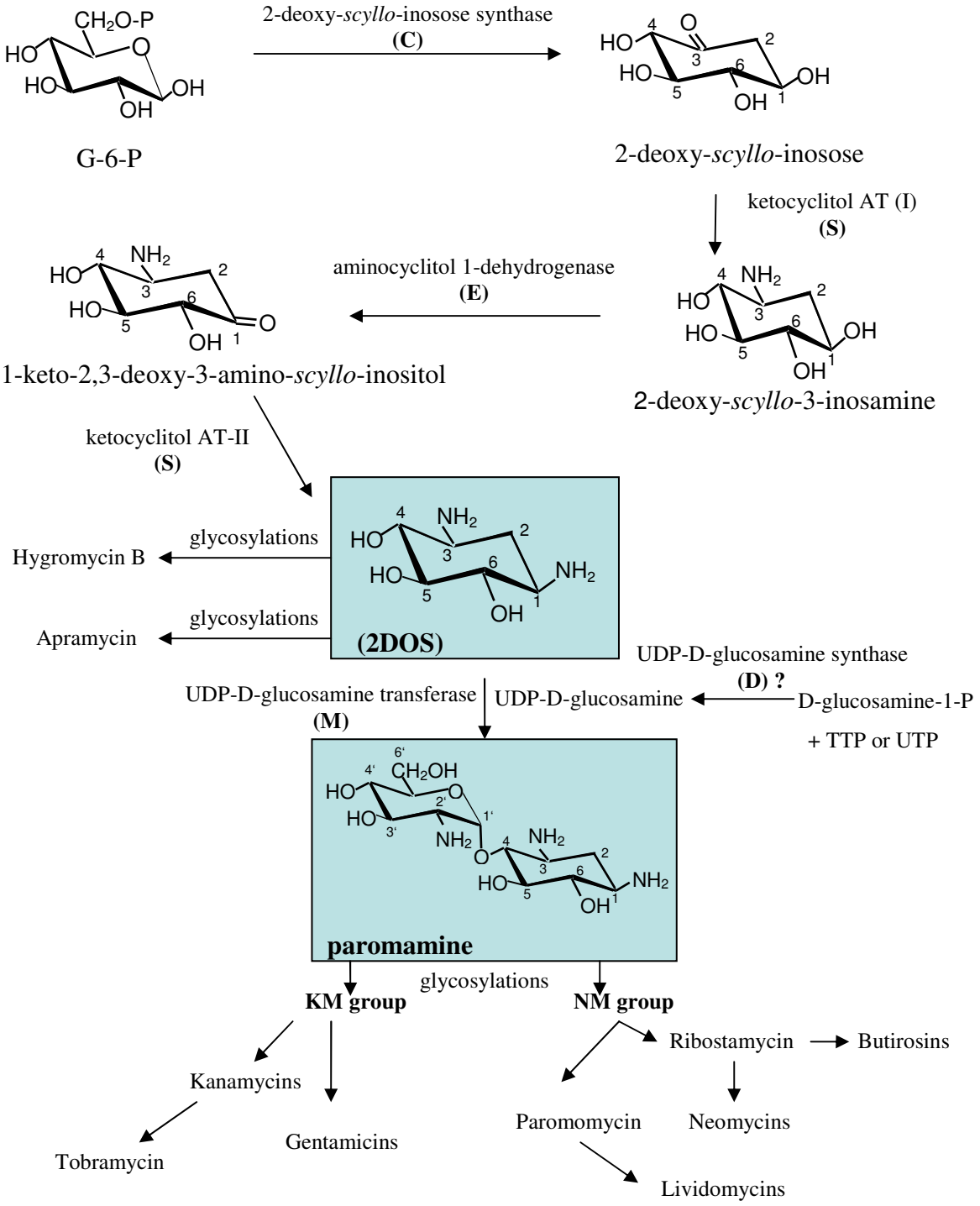

Fig. 4.1 A proposal for the biosynthetic route to paromamine in the producers of the major classes of 2DOS-ACAGAs. The bracketed letters in bold in the figure indicate the possible protein families encoded by the respective ACAGA gene clusters that could catalyse the respective reactions: C = BtrC, NeoC, LivC, ParC, RibC, HygC, AprC, TobC, KanC, GenC, IstC; S = BtrS, NeoS, LivS, ParS, RibS, HygS, AprS, TobS1, KanS1, GenS1, IstS; E = BtrE, NeoE, LivE, ParE, RibE, AprE, TobE, KanE, GenE, IstE; D = BtrD, NeoD, LivD, ParD, RibD, AprN, TobD1, KanD1, GenL, ForD, IstD; M = BtrM, NeoM, LivM, ParM, RibM, AprM, TobM1, KanM1, GenM1, IstM. NM = neomycin; KM = kanamycin.

dependant transamination steps of the 2DOS pathways; cf. Fig. 3.17) to α-ketoglutarate, before being spontaneously cyclised to 2-hydroxy-5-keto proline, a likely toxic metabolic compound. However this putative function is contradicted by the absence of the equivalent gene in the *hyg*-cluster. In addition, the BtrD protein could also act as a bifunctional enzyme (nucleotidyltransferase and 2'-N-acetylparomamine deacetylase) or maybe involved in another still unknown reaction.

4.2.3 4,5-glycosylated 2DOS-ACAGAs (NM group)

4.2.3.1 Biosynthesis of the NM-related ACAGAs

The similarities in the chemical structures of the members of this group are mirrored by the greater similarity in the respective actinomycete gene clusters, gene content and sequence similarity in the individual genes/enzymes (cf. Fig. 3.2). The set of postulated enzymes encoded by the respective clusters are putatively complete for the biosynthesis of the individual ACAGAs except for BU. All the respective members of this group share the common intermediate paromamine and all the enzymes needed for this purpose are conserved in the producers of this subgroup (cf. Fig. 4.1). All the respective NM-related ACAGAs contain a ribosyl unit attached via a glycosidic link at position 5 of the cyclitol residue which distinguishes this group from the other subgroups of 2DOS-ACAGAs (cf. Fig. 1.2). Therefore, glycosylation of paromamine with an activated ribosyl unit would be expected to occur in order to form the common pseudotrisaccharidic intermediate (5-ribosylparomamine; Fig. 4.2). A pair of genes (*neoA, neoL* and related genes) is conserved only in this subgroup with a relatively good likeliness to encode proteins involved in the activation and transfer of this ribosyl unit. However, the enzymatic mechanism of this process still unknown. We expected that NeoA and NeoL (and related proteins) would be involved in the ribosyl activation and transfer, respectively, since: (i) NeoL shares some aa similarity with other phosphoribosyltransferase proteins and other glycosyltransferases (NP752321, 32% identity in 152 aa overlap; ZP00166262, 43% identity in 57 aa overlap; NP248665, 47% identity in 57 aa overlap); (ii) the type of activation of the ribose residue for this unique biosynthetic step should be completely new since the normal activated ribose in the cell is 5-P-ribose-1-pyrophosphate (PRPP), which is needed or the biosynthesis of nucleotides at several stages and hence no conserved domains with other known proteins for the NeoA protein was found in the data base; (iii) both proteins are only present in the producers of this subgroup of ACAGAs and absent in others; (iv) both proteins are also conserved in both RM

and BU producers (pseudotrisaccharidic ACAGAs) and contain a ribosyl unit which means that these two proteins would have nothing to do with the third glycosylation process as in the pseudotetrasaccharidic ACAGAs (i.e. NM, PM and LM).

After formation of the 5-ribosylparomamine, a need for reactions catalysed by a 6'-dehydrogenase and a transaminase would be necessary to produce the pseudotrisaccharide RM which seems to be the end product in *S. ribosidificus* (Fig. 4.2). A pair of genes (*neoQ and neoB*) is found to be conserved in all gene clusters for the 2DOS and related ACAGAs except in the *hyg*-cluster. These two genes are postulated to encode the Q- and B-proteins (an oxidoreductase and a type III aminotransferase; see Tab. 3.3) for performing these two biosynthetic steps. The reasons attributed for this hypothesis are: (i) the two amino groups in the cyclitol moiety are produced by the NeoS (and related proteins) as previously analysed (Sect. 4.2.1; cf. Fig. 4.1); (ii) the 2'-amino group originates from the (N-acetyl-)D-glucosamine precursor which is a regular metabolite inside each cell; (iii) both proteins are also encoded by the *rib*-cluster where RM (end product) contains only one possible site for these two candidate enzymes (the 6'-OH group); (iv) the NeoQ (and related proteins) share putative conserved domains to other alcohol dehydrogenases from the pfam05199 (COG2303); (v) a protein homologous to NeoQ is not encoded by the *hyg*-cluster and HM-B lacks a 6'-amino residue or an equivalent one; (vi) NeoB (and related proteins) shares a conserved domain structure to other aminotransferases type III (pfam00202; COG0001) with good identity to a glutamate-1-semialdehyde aminotransferase-like protein (CAF34036). Both the Q- and B-related characterised enzymes are generally involved in the oxidoreduction/transamination of primary alcoholic/aldehyde groups (rather than ring positioned hydroxyl or keto groups).

Interestingly, only one pair of homologous proteins (ParQ/LivQ and ParB/LivB) is encoded by the *par*- and *liv*-clusters. However their end products (PM-I and LM) lack a 6'-amino group. Furthermore, the non-functionality of the respective enzymes necessary for fulfilling a similar job in PM/LM is not evident from their primary structures. The presumed reasons for that are: (i) both enzyme pairs encoded by the *par*- and *liv*-clusters show different protein folding and hence their 3-dimensional structures are different from the other homologous proteins encoded by the *neo*-, *rib*- or *btr*-clusters; (ii) the presence of this gene pair in the *par*- and *liv*-clusters could mean that they are also responsible for introduction of the 6'''-amino group (Fig. 4.2); (iv) a phylogram of BtrB homologous shows the closer relatedness of ParB/LivB and their divergence from the rest of the homologous proteins (NeoB, RibB, BtrB, etc.; cf. Fig A.7) which again could underline their separate substrate specificity. Therefore, it is postulated that the ParQ/B,

LivQ/B protein pairs act only on the pseudotetrasaccharidic substrate for the introduction of the 6’’’-amino group and on the other hand, the NeoQ/B pair alone acts as a pair of bifunctional enzymes on both the pseudotrisaccharidic and pseudotetrasaccharidic intermediates for the introduction of both the 6’- and 6’’’-amino groups.

Another surprising phenomenon of unexpected occurrence in the *rib*-cluster is the *ribF* gene, the equivalents of which putatively encode the second hexosaminyltransferase in the NM, PM and LM pathways. Furthermore, both the *rib*- and *neo*-clusters are highly conserved and the only major difference found was that in the *rib*-cluster, a natural frame shift mutation created by two compensating frame shifts (+1 bp in position 544 and -1 in position 771) was seen in the *ribN* ORF (encoding a putative 5’’’-epimerase). Since frame-shifts of this kind can also occur as artefacts via either reading errors during automatic analysis of DNA sequences or via mutational alteration in heterologous genetic backgrounds after cloning of the DNA, we analysed the particular DNA sequence twice again from independently isolated fragments obtained directly by PCR-amplification from the genomic DNA of *S. ribosidificus*, which confirmed the frame shifts. These results can be interpreted by either one or a combination of the following possibilities: (i) the RibF protein has lost its function; however, its production can not be excluded because there are no features to be seen in the DNA sequence which would indicate its exclusion from transcription or translation; (ii) the natural frame-shifts occurring in the *ribN* ORF lead to an inactive RibN protein which could negatively affect this glycosylation process (RibN could be involved in other as yet unknown accompanying function(s), so that causing fidelity of the third glycosylation step); (iii) the trisaccharidic nature of RM could be a favourite substrate for the exporter system (RibT/U) and therefore, compete with the third glycosylation (RibF) which leads to pumping of RM outside the cell before being further glycosylated.

Another pair of ORFs (*livY*/*W*) are only found in the *liv*-cluster with a good reclamation to encode proteins (LivY and LivW) having putative 3’,4’-dehydratase and 3’,4’-oxidoreductase activities, respectively. Both enzymes are anticipated to be necessary for the 3’-dehydroxylation, the unique biosynthetic step during biosynthesis of LM. The reasons attributed for that are: (i) both genes are found only in the *liv*-cluster and are absent in the other clusters of this subgroup; (ii) both proteins show similarities to other protein families encoding dehydratase and/or oxidoreductase functions summarized under the data base entries pfam00106 (COG1028) for LivY and COG1032 for LivW; (iii) both proteins exhibit good identities to related ones in the *apr*-cluster, putatively involved in the same type of catalytic function (3’-dehydroxylation) in the

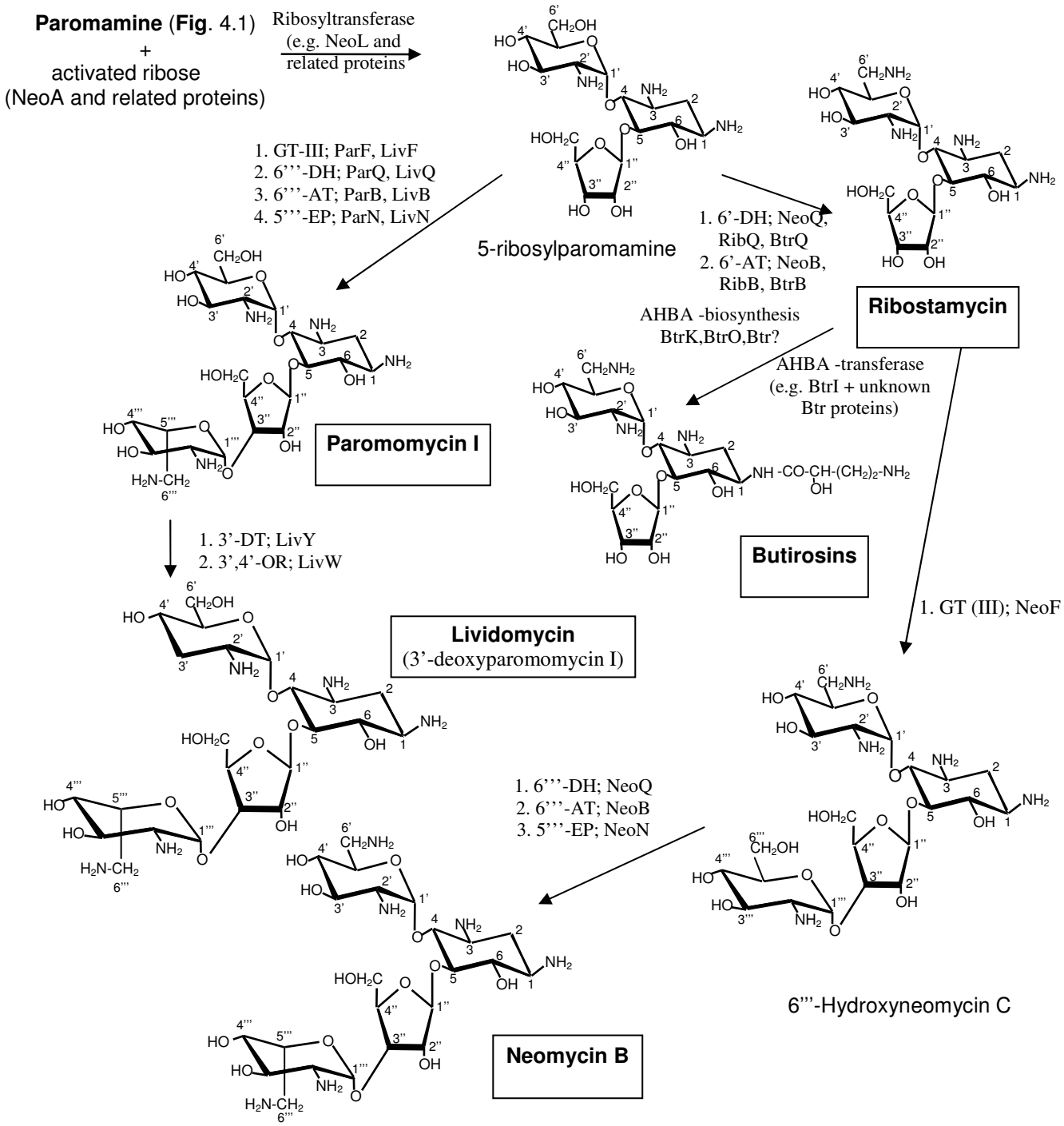

Fig. 4.2 Proposal of a general pathway design for the biosynthesis of the 4,5-glycosylated 2DOS-ACAGAs starting from paromamine. The ACAGAs given in bold are the main end products in their corresponding ACAGAs producers (cf. Tab. 2.1). AT = aminotransferase; AHB = α-hydroxy-γ-aminobutyryl residue; DH = dehydrogenase; DT = dehydratase; EP = epimerase; GT = glycosyltransferase; OR = oxidoreductase.

apramycin pathway (LivY, 58.4% identity in 229 aa overlap with AprD3; and LivW, 85.3% identity in 458 aa overlap with AprD4; cf. Figs. A.9 and A.10; cf. Tab. 3.3).

4.2.3.2 Resistance mechanisms for the NM-related ACAGAs

The producers of NM-related ACAGAs rely on aminoglycoside modification, *O*-phosphorylation or *N*-acetylation, as resistance mechanisms for their self-protection (Cundliffe, 1989; Shaw *et al*., 1993; Piepersberg, 1997). The gene clusters reported contain at least one resistance gene (generally *aphA*-derivatives) in the cases of the *neo*-, *rib*-, and *par*-clusters; however, none so far was detected in the *liv*-cluster and the DNA segments covering at least the larger part of the *btr*-cluster. However, a BU phosphotransferase gene (*aphA*-family) is located probably outside the *btr*-cluster (Herbert *et al*., 1986). The *neo*- and *rib*-clusters were found to be flanked by two resistance genes (*aphA* and *aacC*); the *par*-cluster contained only one resistance gene (*aphA*) at one side of the cluster however, the *aacC* gene and a putative *aac6'-IIc* gene are located outside the *par*-cluster (see Sect. 3.2.1; cf. Figs. 3.5 &. 3.3). Some of these genes had already been described in earlier studies (Thompson and Gray, 1983; Hoshiko *et al*., 1988; Lopez-Cabrera *et al*., 1989; Salauze *et al*., 1991). Because LM is the only antibiotic which lacks the 3'-OH group (target site of AphA enzyme), it can therefore, not be 3'-phosphorylated as a resistance mechanism and hence the *aphA* gene became deleted from the *liv*-cluster during evolution. Since, in the *liv*-cluster, the sole case among the respective clusters, *livZ* is present which putatively encodes an extracellular alkaline phosphatase similar to the StrK from the streptomycin producer *S. griseus* (Mansouri and Piepersberg, 1991; cf. Fig. 3.5 and Tab. A.5). This strongly implies that LM, like streptomycin and the other NMs is phosphorylated inside the cells and that, after secretion, it becomes dephosphorylated in the medium. This additionally implies that another (non-*aphA*-like) LM phosphotransferase gene occurs in *S. lividus*, but outside the *liv*-cluster. However, AGA phosphotransferases modifying NM-like 2DOS-ACAGAs in other positions are not known so far. Furthermore, another additionally conserved gene (*neoP* and related genes) which encodes a putative sugar phosphate phosphatase could also be involved in the 3'-dephosphorylation of the respective phosphorylated-ACAGAs outside the cells (cf. Fig. 3.2).

Also, a pair of genes was found to be highly conserved among the respective gene clusters (*neoT*, *U* and related genes) which encodes two transporter proteins (NeoT/U; cf. Tab. 3.4). These two proteins show a significant similarity to classical, two-components ABC transporters (e.g. NP823966 and NP823967), which are mostly encoded by two adjacent genes in an operon

(cf. Fig. 3.2). Presumably, both enzymes may be involved in the export for of the respective ACAGAs outside the cells (cf. Fig. 3.2 and Tab. 3.4) and therefore, could contribute to the resistance phenotype of the producer cells as drug exporters. Future investigations will have to clarify these open questions.

4.2.3.3 Possible sensor/response regulatory system

The presence of the three conserved genes *neoG, H, I* (and the related genes) among the actinomycete genes clusters of the NM-family (*neo-, par-, liv-*, and *rib-*) gave an expectation that their gene products could play a role in the regulation of the produced metabolites (cf. Fig. 3.2). From our lab discussion, there was a conclusion that, the respective three conserved gene products (G/H/I) would possibly act as a new type of sensor/response regulatory system for transmitting signals via the cytoplasmic membrane and hence control the pathway-specific gene expression (Galperin, 2004). Some reasons which lead to that conclusion are: i) the equivalent gene products are also encoded by *hyg-* and *kan*-clusters as well as by the *cin*-cluster (cinnamycin gene cluster of *S. cinnamoneus* subsp. *cinnamoneus* DSM 40005; Cinorf12, 13 and 14; Widdick *et al*., 2003). In all cases, the three conserved genes are present in a common operon except in the *par*-cluster (cf. Figs. 3.2, 3.5 and 3.9). Equivalent genes are absent in the other related gene clusters (i.e. *gen-*, *for-*, *tob-*, *ist-* and *apr*-clusters; cf. Figs. 3.5, 3.7 and 3.11). This means that the presence of these genes does not correlate with the nature and structure of metabolic products formed by the respective clusters. In addition, they are not required for any other metabolic functions such as, biosynthesis, transport or resistance; ii) regarding the structural analysis of the respective protein sets, the NeoI-related proteins are in the range of 160-175 aa long having a possible helix-turn-helix (HTH) DNA-binding motif at positions 51-90 (see Fig. A.11). The NeoH-related proteins are in the range of 168-195 aa long containing a strongly conserved hydrophilic and histidine-rich residues at positions 89-100 and possibly contain a single trans-membrane helix at positions 42-62 (see Fig. A.12). The NeoG-related proteins are in the range of 382-431 aa long having 6 or 7 trans-membrane helices (in the first 230-250 aa; see Fig. A. 13). Presumably, the NeoG/H-related proteins would interact on the cytoplasmic side of the membrane forming a trans-membrane/sensor kinase complex. When signals come via the membrane, the NeoG/H complex would become active which in turn activate the postulated transcription regulator (NeoI-related proteins) leading to a control pathway-specific gene expression. Future physiological analysis in wild type and knock-out mutant strains will help to clarify this working hypothesis.

4.2.4 4,6-glycosylated 2DOS-ACAGAs (KM group)

Biosynthetic feeding studies with isotope labeled precursors, induction of series of blocked mutants and their crossfeeding with accumulated intermediates was carried out extensively for the GMs and FTMs (Kase *et al*., 1982; Hasegawa, 1992). In their biosynthetic pathways proposed on the basis of such results, the two families, KMs and GMs share the intermediate formation of the pseudodisaccharide paromamine (D-glucosamine-alpha-1,4-2DOS) with the NM group (see Sect. 3.2.2). KMs (including TM) are structurally differentiated from GMs by some additional modifications, the latter share with the FTMs, not belonging to the 2DOS class (Rinehart, 1979; Kase *et al*., 1982; Umezawa et al., 1986; Hasewaga, 1992; Piepersberg, 1997). Thus KMs and GMs are both 3''-aminated during the later stages of modification in the neutral sugar unit attached to the 6-hydroxyl group of the 2DOS aminocyclitol. In turn, the GMs and FTMs share typical 6'-C-methylation and 3',4'-didehydroxylations (cf. Figs. 1.3 & 1.5). Recently, two other groups have sequenced the *kan*-cluster of *S. kanamyceticus* (AJ582817, a segment of 47050 bp; Kharel *et al*., 2004b) or part of it (AB164642; a segment of 25467 bp; Yanai and Murakami, 2004). The submitted sequence of the group of Kharel *et al*. (2004b) was found to be different from position 44915 (equivalent to position 41573 bp of our sequence) up to the end of their sequence compared to our sequence. The difference is created by a recombination in a *Sau*3AI recognition site (GATC) in our submitted sequence (SkaJ15.18). This in turn means that, the DNA segment (21314 bp) downstream to this recombinant site is not located adjacent to upstream region harboring the *kan*-cluster. However, both DNA segments were ligated during construction of the cosmid bank. In addition, the same group of Kharel has recently submitted part of the *tob*-cluster (AJ579650; a segment of 13802 bp; Kharel *et al*., 2004a) where the overlapping sequences to our submission was found to be identical. Also Kharel group (AJ575934; 32668 bp; Kharel *et al*., 2004a) and another group (AY524043; 38146 bp; Unwin *et al*., 2004) have partially sequenced the *gen*-cluster from *M. echinospora* ATCC 15835. The overlapping sequences of these clusters were identical to our submission.

Analyses of the *kan*-, *tob*- and *gen*-clusters revealed two characteristic features, (1) some of the centrally located genes are duplicated or even multiplied most probably to encode proteins responsible for unique biosynthetic purposes. This phenomenon could be explained as follows: a) presence of an amino group at position 3''of the respective ACAGAs could explain the duplication of the *S*-like genes (*kanS1/S2* and the related gene pairs). The hypothesis on which this conclusion is based are : i) only one related gene is required [like in the *neo*- and other gene

clusters; also on the evidence formerly provided by Lucher *et al*. (1989) and Walker (1995)] to encode a bifunctional aminotransferase needed for the introduction of the two amino groups into the 2DOS cyclitol ring; ii) the KanS2 would act on 3''-keto-intermediate originated from the secondary sugar hydroxyl group which resemble to certain degree the structures of the ketocyclitol intermediates (substrate of KanS1); iii) the presence of *kanD2* (and related genes) only in these gene clusters which putatively encode 3''-dehydrogenase complements this hypothesis; b) duplication of *M*-like genes (*kanM1/M2* and the related genes), suggests that the second glycosylation step (probably catalysed by KanM2 and the related proteins; 6-glycosylation of 2DOS) needs a very similar enzyme and cosubstrates during the biosynthesis of KMs and GMs (UDP-D-glucosamine or UDP-xylose, respectively); c) the four fold occurrence of a B-type aminotransferase gene in the *gen*-cluster only (*genB1/B2/B3/B4*) is still enigmatic and needs particular investigation. At present the most likely explanation for this phenomenon is that the encoded pyridoxal phosphate-dependant enzymes are not only involved in the 6'-transamination steps. Rather, at least in part in the stereoisomerization of the 6'-C-atom to yield the various variants and branching alternatives of GM family intermediates/end products. Future biochemical studies of the encoded proteins as well as knock-out mutants of the respective genes would give more clear knowledge about their precise catalytic functions.

(2) Analyses of the *tob*-cluster in comparison to the related *kan*-cluster revealed the following: a) absence of some genes that would encode proteins required for the 3'-dehydroxylation processes during the biosynthesis of TM. However, the only hypothesis related to that observation is that, this unique biosynthetic step could be catalysed by the two proteins encoded by the co-existing *apr*-cluster in the same producer (AprD3 and AprD4; Sect. 4.1.2); b) no resistance genes were detected in the *tob*-cluster however, it is also assumed that the *kamB* resistance gene (Holmes *et al*., 1991) found in the *apr*-cluster would also counter resistance to TM; c) presence of the two genes, *kanK* (putatively encodes 2'-deaminase) and *kanJ* (putatively encodes 2',3'-oxidoreductase/reductase) only in the *kan*-cluster would explain their necessity for the conversion of KM-B to KM-A (the two main end products) in *S. kanamyceticus*. This prediction is made on the following arguments: i) both *kanJ* and *kanK* genes are unique to the *kan*-cluster; ii) absence of these two genes from the *tob*-cluster, since these catalytic functions are not needed for TM (only one end product; TM = 3'-deoxykanamycin B); iii) UDP-(N-acetyl)-D-glucosamine is the precursor required for the first glycosylation step (KanM1); therefore, the hypothesis that UDP-D-glucose could act as an alternative precursor to form KM-A seems to be high unlikely;

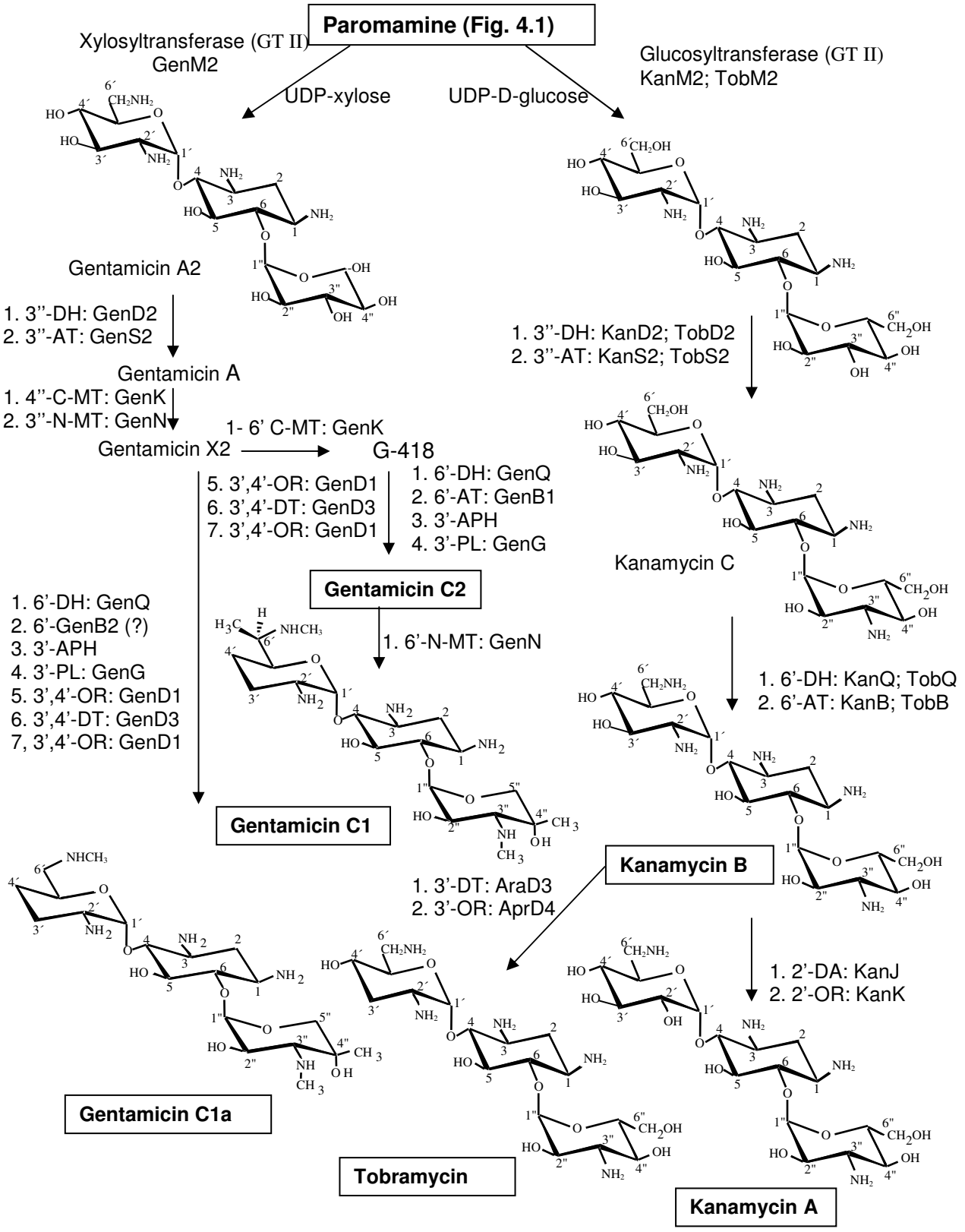

Fig. 4.3 Proposal of a general pathway design for the biosynthesis of the 4,6-glycosylated 2DOS-ACAGAs starting from paromamine. ACAGAs given in bold are the main products in their corresponding producers (Tab. 2.1). AT = aminotransferase; DA = deaminase; DH = dehydrogenase; DT = dehydratase; MT = methyltransferase; OR = oxidoreductase ; PL = phospholyase.

iv) KanJ has significant similarity with the MmcH protein (mitomycin biosynthetic protein; 31% identity in 204 aa overlap; COG5285) from *S. lavendulae* which could have a similar biosynthetic function (deaminase); v) KanK has a significant similarity (43% identity in 61 aa overlap; NP626465) with a putative reductase. Certain genes encoding proteins needed for unique biosynthetic functions in GMs were present only in the *gen*-cluster: putative 6'/(4'')-C-methyltransferase (GenK), which is closely related to a characteristic enzyme encoded by the *for*-cluster (ForK; Kuzuyama *et al*., 1995); and the putative 6'-N-methyltransferase (GenN) which was also found to be conserved in the *ist*-cluster (cf. Figs. 3.5 & 3.11). A proposal for the biosynthetic pathways for the three ACAGAs starting from the paromamine intermediate is outlined in Fig. 4.3. It was known before that the producers of KM-related ACAGAs rely on modification of the target site (16S rRNA methyltransferase) as a resistance mechanism (Kelemen *et al*., 1991; Shaw *et al*., 1993; Piepersberg, 1997; Demydchuk *et al*., 1998). Two resistance genes were found in both the *kan*- and *gen*-clusters (cf. Fig. 3.5). In case of the producer strain *S. kanamyceticus*, the genes *kmr* (CAA75800.1; encoding 16S rRNA methyltransferase) and *kanA* (AB164230.1; encoding aminoglycoside 6'-acetyltransferase) were known from earlier studies (Murakami *et al*., 1983; Demydchuk *et al*., 1998). Both genes were found in our data set and turned out to be identical with those submitted to the data base earlier. In the *gen*-cluster, *gmrA* and *gmrB* (both encoding 16S rRNA methyltransferases) were detected; however, only the *gmrA* gene was already known from earlier analyses and data base submissions (Kelemen *et al*., 1991; M55520). The *kan*-, *tob*- and *gen*-clusters contained genes which encoded putative efflux proteins that could play a role in the export system of the respective antibiotics (cf. Sect. 3.2.2). Out of these, a *kanT*-related gene was conserved in all the clusters (*kanT*, *tobT* and *genT*), however, a *kanX*-related gene was conserved only in the *gen*-cluster (cf. Fig. 3.5).

Analysis of the *gen*-, *kan*- and *for*-clusters revealed evidences that the *gen*-cluster could have basically originated from a fusion of a *kan*- and a *for*-cluster. Comparison of the DNA and protein sequences showed a significant similarity between the *gen*- and *kan*-clusters as well as between the *gen*/*for*-clusters. However, no significant similarity could be detected between the *kan*- and *for*-clusters. The flanking segments of the *gen*-cluster showed a strong similarity (55 to 95% sequence identity) to the *for*-cluster. Especially in most of the genes which are not present in any other of the analysed ACAGA gene clusters, i.e. *forHIJ/genHJ*, *forQ/genQ*, *forPBK/genB4PB3K*, *forT/genI*, *fosC/genW* as well as for several truncated (obviously non-functional) ORFs. The opposite of that was found between the *kan*- and *gen*-clusters where the

central part of the *gen*-cluster showed a strong similarity to the *kan*-clusters. Especially in most of the genes conserved among the majority of the 2DOS-ACAGA gene clusters, which share the pseudodisaccharidic intermediate, paromamine in their biosynthesis (Okuda and Ito, 1982; Piepersberg, 1997). Examples of these gene sets are: *kanBQ/genBQ*, *kanS1CD2M2/genS1CD2M2*, *kanE/genE*, *kanS2/genS2*, *kanM1/genM1*. Structural comparison of GM-C1, KM-B and FTM-A with regard to their biosynthetic origin and evolution of a hybrid pathway is presented in Fig. A. 14.

4.2.5 4-glycosylated 2DOS-ACAGAs (Apr group)

Apr is the only important product in this group and is produced by various different actinomycetes, e.g. *S. tenebrarius* DSM 40477 and *St. hindustanus* DSM 44523. Organization and order of ORFs for the two *apr*-clusters derived from these two strains are highly conserved (cf. Fig. 3.7). This could mean that their occurrence in such distant organisms is the result of a rather recent event of horizontal gene transfer of the full length gene cluster. The postulated biosynthetic pathway for Apr starting from 2DOS is illustrated in Fig. 4.4. The genes encoding proteins (AprC, AprS and AprE) involved in the biosynthesis of the 2DOS moiety were conserved in both clusters (cf. Fig. 3.7). The first sugar unit in Apr could be formed via an (NDP-) octosamine sugar precursor probably derived from a C7- or C8- pentose phosphate pathway intermediate (heptoses and octoses are typical bacterial cell wall sugars). This octosamine unit would later undergo some intramolecular arrangement to form the stable double ring sugar unit attached to position 4 of the 2DOS unit. The genetic evidences in favour for this hypothesis are: (i) presence of the *aprK* gene that encodes a protein related to ADP-heptose synthases with good identity to other proteins from this family (e.g. ZP00291440 with 38% identity in 119 aa overlap). So, it seems that the AprK protein would putatively encode an octose- (or heptose-) or octosamine-activating enzyme which in presence of NTP forms an NDP-activated precursor of an octosamine; (ii) presence of *aprJ* and *aprD1* genes that encode proteins with good similarity to known phosphosugar mutases and NDP-sugar epimerases, respectively. These enzymes could also be involved in the formation and modification of the postulated NDP-octosamine precursor; (iii) presence of a gene, *aprH*, for a putative glycosyltransferase (AprH; related to ZP00054179.1; 32% identity in 321 aa overlap; COG3980) which could be involved in 4-octos(amin)yl transfer of the C8-sugar moiety.

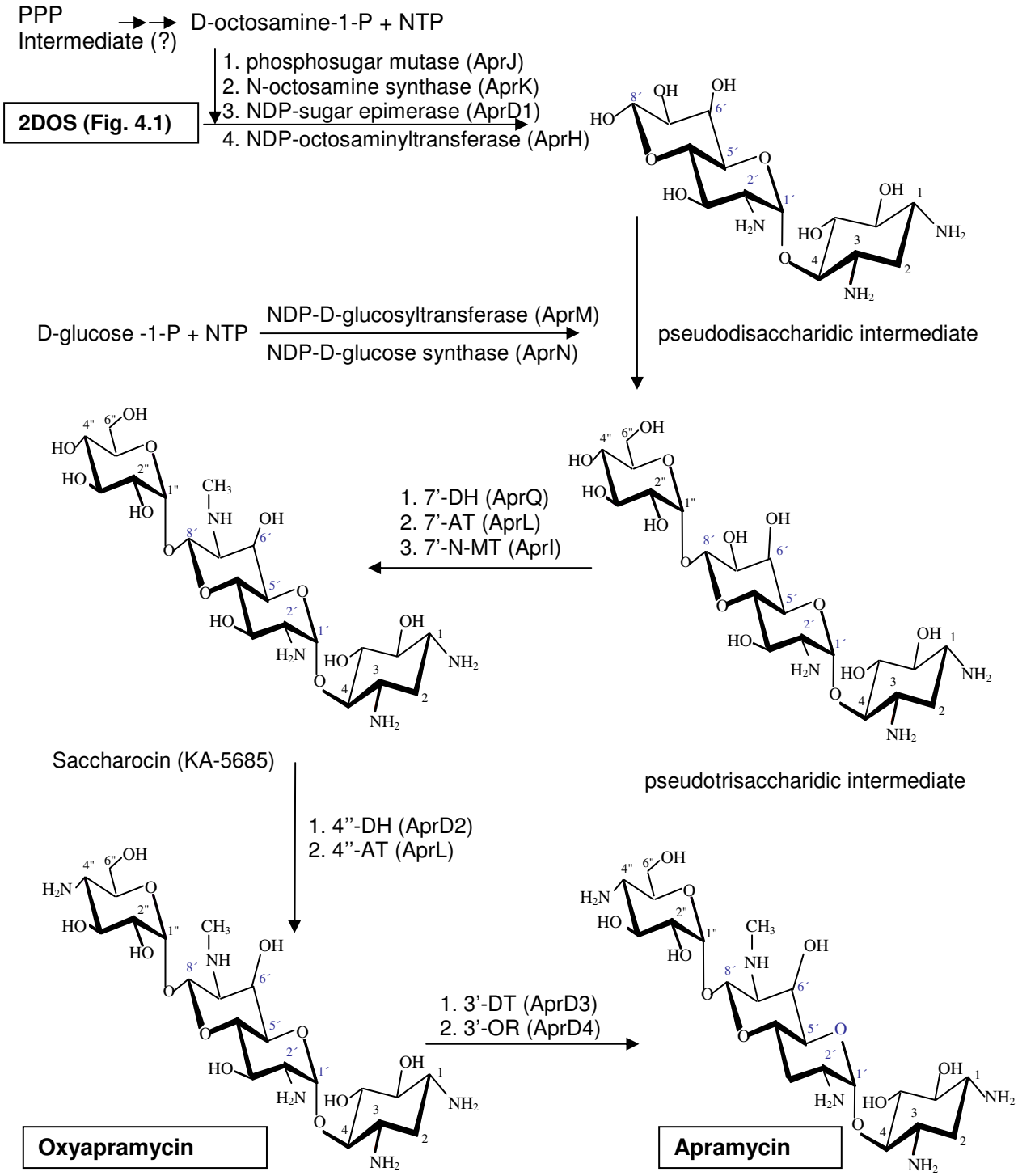

Fig. 4.4 A proposal for a biosynthetic pathway for the 4-glycosylated 2DOS-ACAGAs (apramycins) starting from 2DOS. ACAGAs given in bold are the main products in the producing strains (cf. Tab. 2.1). AT = aminotransferase; DH = dehydrogenase; DT = dehydratase; MT = methyltransferase; OR = oxidoreductase; PPP = pentose phosphate pathway.

However, the exact function and their placement in the pathway of both the AprJ- and AprD1-catalysed steps are highly speculative at present, i.e. before or after the glycosylation to form the pseudodisaccharidic intermediate (see Fig. 4.4). Another conserved gene in the *apr*-cluster (*aprM*), putatively encodes a second glycosyltransferase (AprM), relatives of which are conserved in other ACAGA gene-clusters, which is likely to be involved in the second glycosylation step. Related activities were attributed above to the KanM2 and TobM2 proteins which are probable UDP-D-glucosyltransferases (cf. Tab. 3.3; see Fig. 4.3). Formation of a UDP-D-glucose precursor could be catalysed by the AprN (BtrD homolog) as discussed above (see Sect. 4.2.2). After formation of the pseudotrisaccharidic intermediate, introduction of an amino group at position 7' followed by its methylation would be the next step in order to form the first byproduct, saccharocin or KA-5685 (see Fig. 4.4; Piepersberg, 1997). Therefore, the three genes (*aprQ, aprL* and *aprI*) encoding proteins with putative 7'-dehydrogenase, 7'-aminotransferase and 7'-N-methyltransferase activities are expected to be involved in these biosynthetic steps.

For the production of oxyapramycin (second byproduct) from saccharocin, a need for introduction of an amino group at position 4''of the second neutral sugar is necessary. Therefore, both AprD2 (putative 4''-dehydrogenase) and AprL (putative 4''-aminotransferase) are expected to be involved in this conversion (cf. Fig. 4.4). Production of Apr (end product) does occur via 3'-dehydroxylation of oxyapramycin in which we postulate that the enzymes AprD3 (putative 3',4'-dehydratase) and AprD4 (putative 3',4'-oxidoreductase) are involved (see also Sect. 4.2.3).

4.2.6 5-glycosylated 2DOS-ACAGAs (HM-B group)

HM-B is at present the sole member of this group (cf. Fig. 1.4). The postulate on the biosynthetic pathway for HM-B is based on an interpretation of the gene products encoded by the *hyg*-cluster as illustrated in Fig. 4.5. It was also observed that the *hyg*-cluster could be located at the end of the respective chromosome because the last 2 ORFs (ShyG17.32c and ShyG17.33c) on the cloned DNA segment harboring the *hyg*-cluster are homologous to those (NP631871.1 and NP624364.1) at the end of the linear chromosome of the fully sequenced genome of *S. coelicolor* A3(2) (Bentley *et al*., 2002). The following proteins were assigned to biosynthetic functions in the HM-B: a) The biosynthesis of the aglycone and its modification: the conserved proteins HygC, HygS, HygE clearly are involved in 2DOS biosynthesis (cf. Fig. 4.1); HygM putatively represents a 3-N-methyltransferase due to its homology to the SpcM protein (AF145038) which

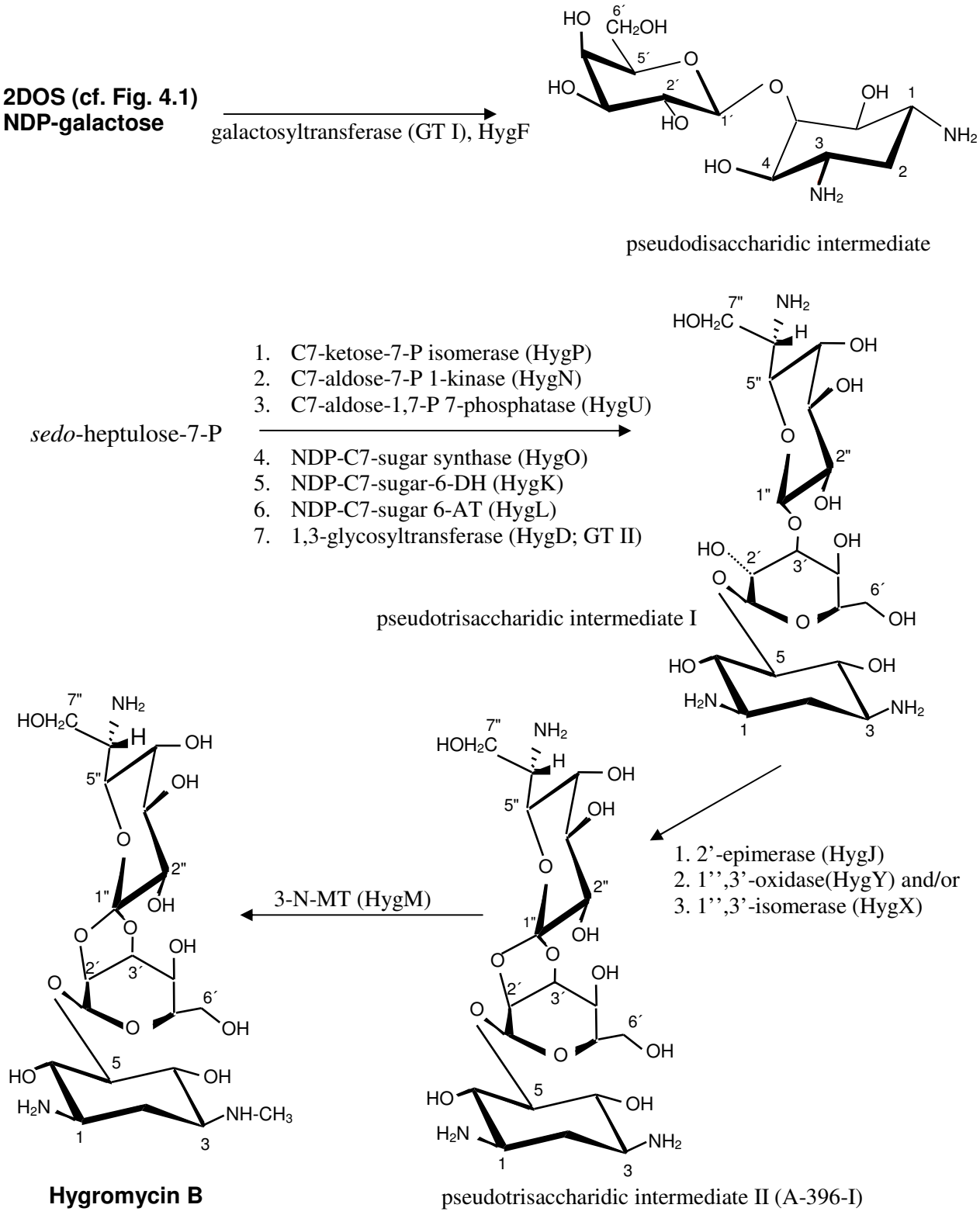

Fig. 4.5 Proposal of a biosynthetic pathway for the 5-glycosylated 2DOS-ACAGAs (HM-B) starting from 2DOS. AT = aminotransferase; DH = dehydrogenase; GT = glycosyltransferase; MT = methyltransferase; NDP = nucleotidyldiphosphate; P = phosphate.

was postulated to be responsible for a similar catalytic function in the biosynthesis of spectinomycin (see Sect. 3.2.4); b) The biosynthesis of the glycosidic units and their modifications: From the chemical structure of HM-B (cf. Fig. 1.4) as well as from the analysis of the *hyg*-cluster, it seems that both sugar units would likely be derived from a hexose and a heptose, respectively (Piepersberg, 1997). The hexose precursor seems to be a galactose rather than glucose. The heterocyclic ring formed between the first and second sugar, gives rigidity to the molecule and hinders the normal rotation and flexibility between the two sugar units in the disaccharidic molecule. Therefore, for molecular stability, the hydroxyl group at position 4 of the first hexose should be in an axial conformation and above the molecular plane, which is the case in galactose, and not in glucose. Presence of the *hygJ* gene, putatively encoding an NDP-hexose 3'-epimerase (HygJ), which could be required to change the stereochemistry of the hydroxyl group at the position 3 of the hexose in order to facilitate formation of this heterocyclic ring. It is also expected that formation of this heterocyclic ring between the two sugars units occurs at one of the later steps during the biosynthesis of HM-B. HygF is a putative NDP-galactosyltransferase with good identity to a similar protein encoded in the *spc*-cluster (SpcF; Tab. A.13).

It was predicted that *sedo*-heptulose-7-phosphate is the precursor of the second heptose sugar. *Sedo*-heptulose-7-phosphate is supposed to undergo several modifications, before being glycosylated by the putative 1,3-glycosyltransferase (HygD, 35% identity in 170 aa overlap to a glycosyltransferase from *Methylococcus capsulatus*; AAU92274) to the preformed disaccharidic intermediate. Modifications that are postulated to occur to the *sedo*-heptulose-7-phosphate precursor by use of the putative proteins HygP, HygN, HygU, HygO, and HygL are outlined in Fig. 4.5. Heterocycle formation is expected to be catalysed by one or more proteins, such as a 1",2'- or 1'',3'-oxidase (HygY; a SpcY homolog is present in the spectinomycin producer where a similar heterocyclization process is involved in AGA biosynthesis) or maybe others, such as HygX. The gene product HygA, the APH(7'') enzyme would be responsible for self resistance in the producing strain. HygA was already known from the earlier studies (Zalacain *et al*., 1986). The two genes *hygV* and *hygW* encode proteins putatively involved in the export of HM-B-7''-phosphate outside the cells and therefore could be relevant for the resistance phenotype.

4.2.7 ACAGAs related to 2DOS-containing compounds (FTM/IM group)

A preliminary overview about the partial structure of the *for*-(*fms*-) cluster was already published by Hasegawa and collaborators in the past (Odakura *et al*., 1984; Dairi *et al*., 1992a, b, c; Ohta *et*

al., 1992a, b; Kuzuyama *et al*., 1995). These studies were based on both the isolation of DNA-fragments encoding purified biosynthetic enzymes or complementing several mutations blocked in FTM-A/dactimicin/sannamycin production in *M. olivasterospora* ATCC 21819 and in *S. sannaensis* IFO 14239 (producer of the FTM-related sannamycin). All gene loci had been mapped in a cluster located on a cosmid covering a larger genomic segment and carrying at least ten genes (*fms1*/*forA*, fms3/*forM*, fms4/*forS*, fms5/*forE*, fms7/*forK*, fms8/*forP*, fms10/*forN*, fms11/*forL*, fms12/*forO*, and fms13/*forV*; cf. Tab. A.10). Some of these genes had also been sequenced and/or functionally identified by these authors: *fms7* (*forK*; encoding FTM-KL1 methyltransferase), *fms8* (*forP*; encoding FTM-KK1 3'-phosphotransferase), *fms11* (*forL*; encoding 1-O-methyl-epimerase), *fms13* (*forV*; encoding FTM-B glycyltransferase), *fms14* (*forZ*; encoding FAD-dependent FTM-A oxidase or N-formimidoyl FTM-A [dactimicin] synthase).

Previous studies on the biosynthesis of FTM (and the related pseudodisaccharidic antibiotics) gave indications that the second transamination step in the synthesis of the diamino-cyclitol unit took place on the disaccharidic intermediate, which is otherwise not the case in all the 2DOS-ACAGAs biosynthetic pathways (cf. Figs. 4.1 & 4.6). Thus, it seemed likely that all producers of FTM-like aminoglycosides (including IM, etc.) contain highly related gene clusters (cf. Fig. 3.11), originating from a common evolutionary source, with some minor modifications, such as the use of different gene sets for the formation of the alternative cyclitol precursors, (2-deoxy)-*scyllo*-inosose (see proposed pathway in Fig. 4.6).

Analyses of the blocked mutants and/or conversion studies with IM intermediates corresponding to the last five intermediates of the FTM pathway (cf. Fig. 4.6) in *S. sannanensis* and *S. tenjmariensis* supported this conclusion (Hotta *et al*., 1989a and b; Ohta *et al*., 1992a; Hotta *et al*., 1994). Analyses of genes present in both the *for*- and *ist*-clusters, were found to encode the proteins putatively required for the biosynthesis of FTM and IM, respectively (cf. Fig. 4.6). Moreover, the coupled 3',4'-dehydroxylation is a tailoring biosynthetic process that does not only occur in the formation of FTM and IM, but also in that of GM (cf. Fig. 4.3). The biochemical tools and mechanisms used for this process, would include five subsequent steps namely, phosphorylation-phospholysis-reduction-dehydratation-reduction (cf. Fig. 4.6). The putatively proteins that are responsible for this dehydroxylation process are indicated in Fig. 4.6.

The first step in 3'-phosphorylation would be catalysed by an AphA-related APH(3') enzyme (ForP, IstP) encoded by the respective gene clusters, which means its involvement in a biosynthetic step rather than in the resistance phenotype.

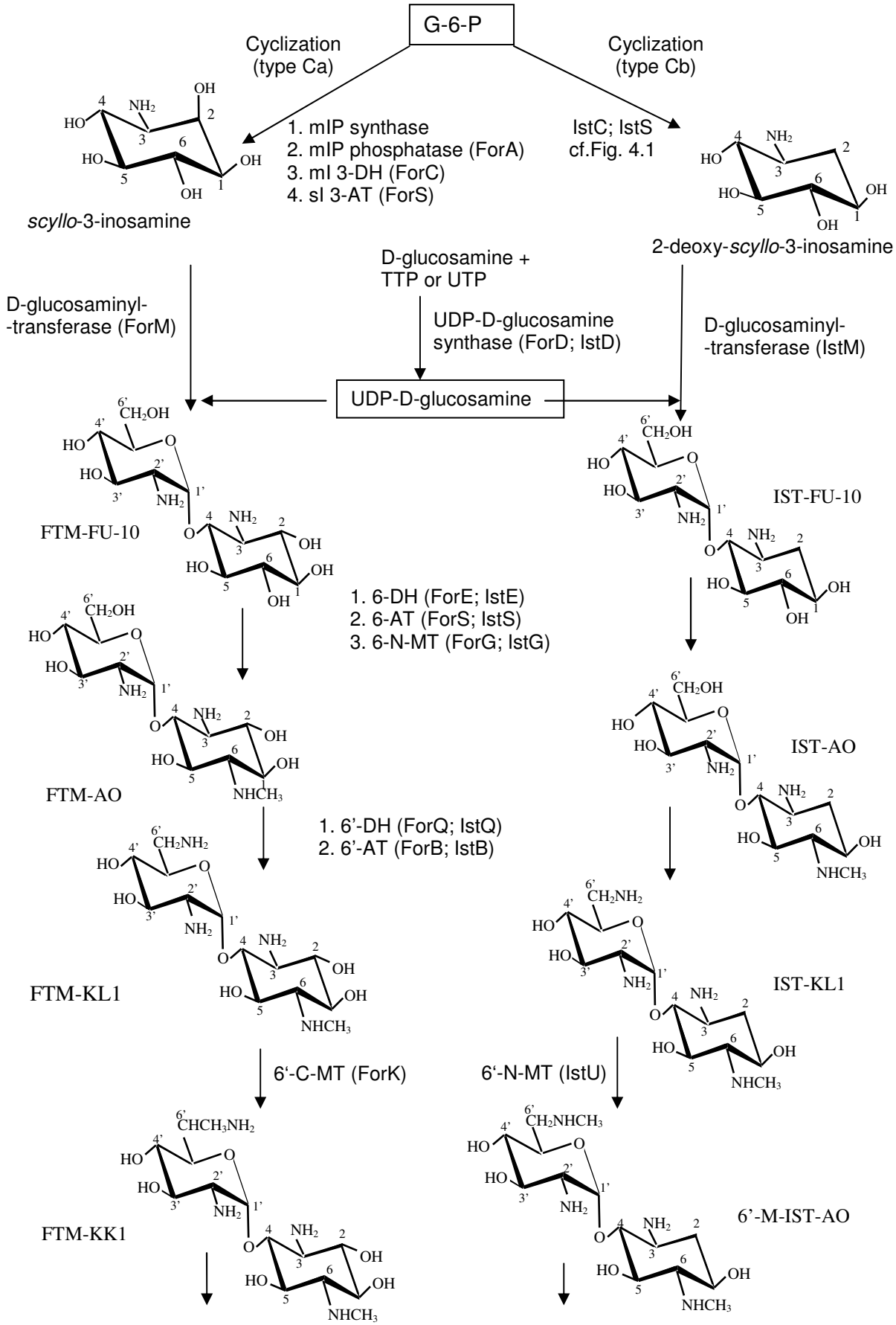

Fig. 4.6; continued on next page

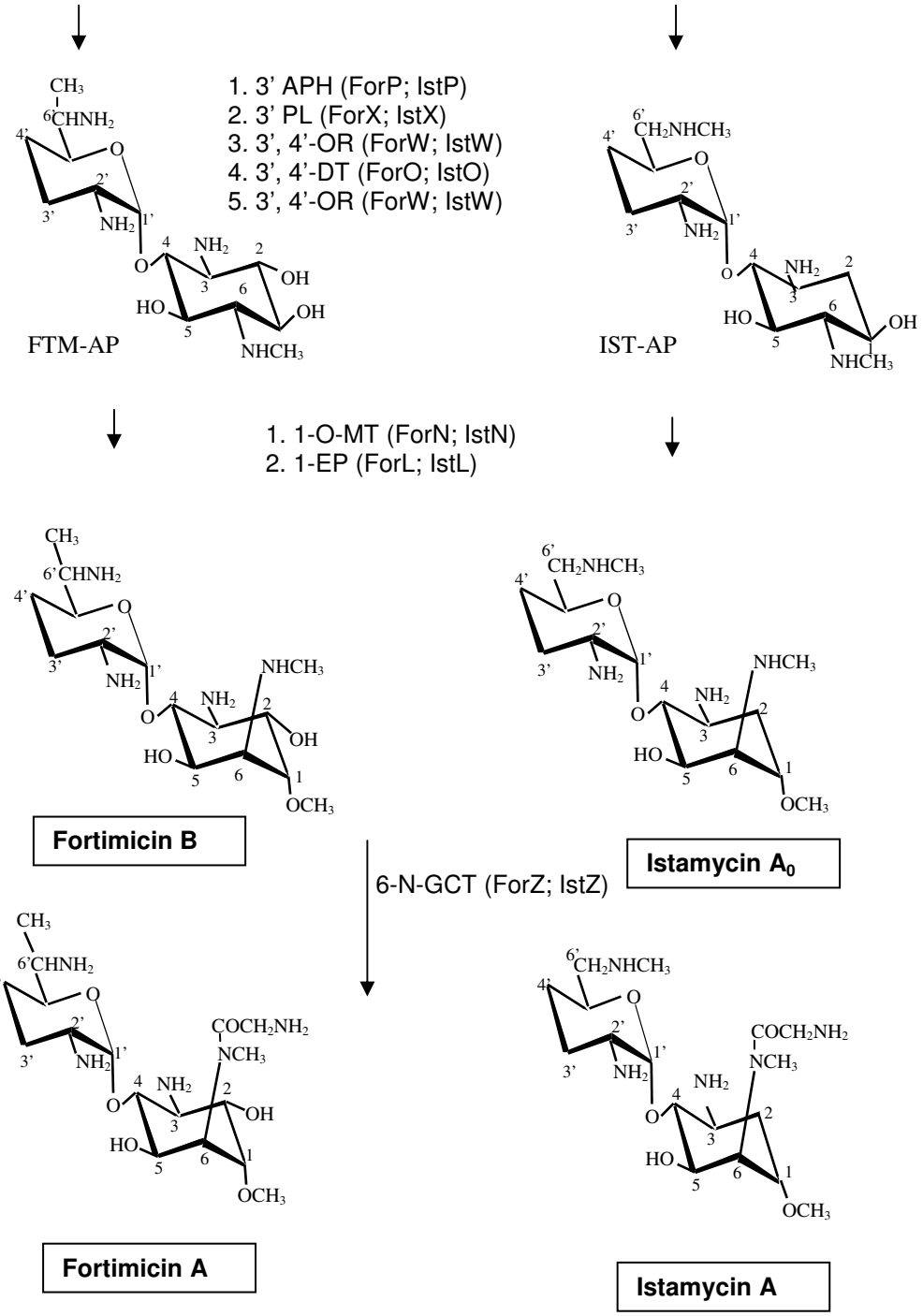

Fig. 4.6 Proposal of a pathway for the biosynthesis of the FTM and IM group ACAGAs starting from G-6-P. ACAGAs given in bold are the main products in the producing strains (cf. Tab. 2.1). APH = aminoglycoside phosphotransferase; AT = aminotransferase; DH = dehydrogenase; DT = dehydratase; EP = epimerase; GCT = glycyltransferase; mI = *myo*-inositol; mIP = *myo*-inositolphosphate; MT = methyltransferase; OR = oxidoreductase;PL = phospholyase; sI = *scyllo*-inosose.

As previously indicated, methylation of the 16S rRNA is considered to be the main resistance mechanism in the producers of FTM and IM (Ohta and Hasegawa, 1993a & b). The three genes (*forI, H, J* and the related genes) that putatively encode proteins involved in the export of the respective antibiotics were also conserved in both clusters (cf. Tab. 3.4). The proposed biosynthetic pathways for both FTM and IM are mostly similar except for two particular biosynthetic phases: (i) the different cyclitol biosynthesis as described above (see Sect. 4.2.1); (ii) the 6'-C-methyltransfer in the FTM pathway (ForK; which also occurs in the GM pathway by GenK) and instead the 6'-N-methyltransfer in the formation of IM (ForU; cf. 4.6).

4.3 Reconstitution of the 2DOS pathway in vitro

The second main goal of this work was to test and elucidate the biosynthetic steps for the biosynthesis of 2DOS, the aglycone unit in all the 2DOS-ACAGAs. The first and second enzymatic steps had already been studied previously and characterized in part from other pathways (Lucher *et al*., 1989; Walker, 1995; Ahlert *et al*., 1997; Kudo *et al*., 1999a; Ota *et al*., 2000; Tamegai *et al*., 2002b; Kharel *et al*., 2004; Kharel *et al*., 2005). After identification and analysis of most of the biosynthetic gene clusters for 2DOS-ACAGAs, three conserved genes were likely candidates to be involved in the biosynthesis of the 2DOS moiety (see Sect. 4.2.1 and Fig. 4.1). For achieving this, the respective three genes from the *kan*-cluster (*kanC*, *kanS1*, *kanE*) were selected for expressing them heterologously in several *E. coli* strains as well as in *S. lividans* TK23 and characterizing their catalytic activities.

4.3.1 Heterologous expression of KanC, KanS1 and KanE proteins

In order to overproduce the soluble proteins of KanC, KanS1 and KanE for enzyme assays, different expression systems in *E. coli* and in *S. lividans* as well as different expression conditions were tested. *E. coli* is the most frequently used prokaryotic expression system for the high level production of heterologous proteins (Hannig and Makrides, 1998; Binnie *et al*., 1997). Due to the high G+C content of the heterologous genes from *Streptomyces* to be expressed in *E. coli*, sometimes codon bias arises as an obstacle during expression (Novy *et al*., 2001). Codon bias leads to insufficient tRNA pools which could in turn lead to stalling, premature termination, frame shifting and amino acids misincorporation during translation. Overcoming the codon bias in *E. coli* for enhanced protein expression was achieved by using the *E. coli* Rosetta strain, because it harbors the pRARE plasmid, on which most of rare tRNAs are encoded (Novy *et al*.,

2001). In this work, the soluble N-terminally His-tagged KanE protein was only achieved via heterologous expression in the *E. coli* Rosetta strain (cf. Tab. 3.8).

4.3.2 Biochemical characterization of KanC, KanS1 and KanE proteins

In order to reach the final goal, to reconstitute the 2DOS formation in a "one-pot-assay", the heterologous expressed and soluble proteins, KanC (2-deoxy-*scyllo*-inosose synthase), KanS1 (putative bifunctional enzyme, L-glutamine (PLP):2-deoxy-*scyllo*-inosose aminotransferase [AT-I]; and 1-keto-2,3-deoxy-3-amino-*scyllo* inositol aminotransferase [AT-II]) and KanE (putative aminocyclitol 1-dehydrogenase) were at first tested individually for their enzymatic activities. Conversion of G-6-P to 2-deoxy-*scyllo*-inosose by KanC was confirmed (see Sect. 3.6.1). In all the control reactions, formation of 2-deoxy-*scyllo*-inosose was never detected which in turn delivered an additional detail to the substrate specificity of the KanC-catalysed reaction. Similar results were obtained by others with KanC homologous from other pathways (BtrC, Kudo *et al*., 1999a; TbmA, Kharel *et al*., 2004). It was also confirmed that both Co^{++} and NAD are necessary for exhibiting KanC activity, since both cofactors were necessary not only for the enzyme reaction but also for stabilizing the 2-deoxy-*scyllo*-inosose synthase as pointed out by Kudo *et al*., (1999a). The mechanism as well as the stereochemistry of this reaction catalysed by the BtrC enzyme was previously illustrated by the group of Kakinuma and others in more detail (Kudo *et al*., 1999a; Nango *et al*., 2003; Nango *et al*., 2004). The BtrC enzyme catalyses intramolecular carbocyclization of G-6-P to 2-deoxy-*scyllo*-inosose via a multi-step process which includes the first oxidation at the C-4 position of G-6-P followed by elimination of the phosphate group from the activated ulose to form an enol or enolate intermediate. Subsequent reduction at C-4, followed by intramolecular aldol condensation between C-1 and C-6 gives rise to 2-deoxy-*scyllo*-inosose.

The next step in the 2DOS biosynthesis, the transamination of the keto group of 2-deoxy-*scyllo*-inosose, was also confirmed to be catalysed by the KanS1 enzyme. This enzymatic step was previously analyzed by different KanS1 homologous from other producers of ACAGAs (Lucher *et al*., 1989; Walker, 1995; Ahlert *et al*., 1997; Tamegai *et al*., 2002b; Kharel *et al*., 2005). Lack of 2-deoxy-*scyllo*-inosose as a pure substrate forced the use of a coupled assay of KanC and KanS1 for testing this activity and delivered the first indication that a combined "one-pot-system" could be achievable (see Sect. 3.6.3). In an L-alanine/NADH-coupled assay, the specific activity of KanS1 enzyme was calculated to be 0.27 U/mg protein in the cell free extract which is in a reasonable range for this aminotransferase activity (cf. Sect. 3.6.2).

The third enzymatic process during the biosynthesis of 2DOS, the oxidation (dehydrogenation) of the 2-deoxy-*scyllo*-inosamine at the C-3 position was postulated by us first to be catalysed by KanE-related oxidoreductases (see Fig. 3.17; Sect. 3.4; Tab. 3.3). KanE is a member of Zn^{++}-dependent alcohol dehydrogenases (pfam00107; COG1063). This reflected by the presence of two conserved histidine and three conserved cysteine amino acid residues in the same positions as in other members of this class that could play role in the attachment to Zn^{++} ions (cf. Fig. A.19). Accordingly, an attempt was made to prove these suggestions and also the dependence on Zn^{++} ions. However, absence of 2-deoxy-*scyllo*-inosamine as a pure substrate had brought an additional obstacle about performing this step in a single reaction. Despite of lacking of pure 2-deoxy-*scyllo*-inosamine, an attempt was made to perform the KanE assay in coupling it with the KanC/KanS1-promoted reactions (see Sect. 2.25). However, it was previously indicated that BtrC (KanC homolog) activity was completely inhibited by the presence of Zn^{++} ions (Kudo *et al*., 1999a). Therefore, the complete "one-pot-system" was carried out in two separate stages, i.e. the KanC/KanS1-catalysed reaction was at first proceeded until 2-deoxy-*scyllo*-inosamine was formed and then the KanE/KanS1-catalysed reactions were started thereafter (cf. Sect. 2.25). In this case, inhibition of KanC by Zn^{++} ions could be avoided. Formation of 2DOS was demonstrated preliminary this way and gave an indication that the activity of KanE was necessary. It also proves that the second transamination in the 2DOS biosynthesis was again catalysed by the bifunctional KanS1 aminotransferase (see Sect. 3.4; Fig. 3.17).

The reasons that 2DOS was detected only in the presence of KanE as a native protein, but not when using the soluble N-terminal His-tagged KanE protein (cf. Fig. 3.28) could be that the tagged His residues interact with and chelate Zn^{++} ions from the medium. In this case, either the protein changes to inactive conformation or Zn^{++} ions would be scavenged and no longer be free to play their role in the KanE-catalysed reaction. In order to better characterize KanE activity further in the future, the following prerequisites should be fulfilled: a) purification of the native KanE protein; b) preparation of 2-deoxy-*scyllo*-inosamine as a pure substrate; c) as well, *kanE*-mutants should be carried out in the wild type strain.

4.4 Prospective of this work

This study gives good basis for the following aspects to be further investigated in future:

1. Biosynthetic studies for these important ACAGAs; This item could be achieved by cloning and expression of all the genes that putatively involved in the biosynthesis and regulation of ACAGAs. This would be done by testing the biochemical activities of the resulted proteins and

hence get a detailed overview about their metabolic roles in the formation of these important metabolites.

2. Evolutionary studies about the origin and relatedness of these important metabolites; This could also be achieved via checking both DNA and protein sequences encoded by the respective gene clusters. Phylogram of all the related proteins encoded by these gene clusters are also a useful tool for testing evolution and origin of these metabolites.

3. Biocombinatorial formation of new ACAGAs; This approach now is considered one of the most important tool for getting new members of these valuable metabolites especially those conferring resistance to the clinically relevant pathogens.

4. Finding new producers of natural, still unknown relatives of the known ACAGAs, e.g. isolation and screening various antibiotic-producing isolates from soil and testing them using the newly developed heterologous primers designed in this work (see Sect. 3.3).

5 References

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

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

Ando O, Satake H, Itoi K, Sato A, Nakajima M, Takahashi S, Haruyama H, Ohkuma Y, Kinoshita T, Enokita R (1991) Trehazolin, a new trehalase inhibitor. *J. Antibiot*., **44**: 1165-1168.

Asano N, Takeuchi M, Kameda Y, Matsui K, Kono Y (1990a) Trehalase inhibitors, validoxylamine A and related compounds as insecticides. *J. Antibiot*., **43**: 722-726.

Asano N, Kameda Y, Matsui K, Horii S, Fukase H (1990b) Validamycin H, a new pseudo-tetrasaccharide antibiotic. *J. Antibiot.*, **43**: 1039-1041.

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

Ban N, Freeborn B, Nissen P, Penczek P, Grassucci RA, Sweet R, Frank J, Moore PB, Steitz TA (1998) A 9 A resolution X-ray crystallographic map of the large ribosomal subunit**.** *Cell*, **93**: 1105-1115.

Ban N, Nissen P, Hansen J, Capel M, Moore PB, Steitz TA (1999) Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit. *Nature*, **400**: 841-847.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution**.** *Science*, **289**: 905-920.

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417**: 141-147.

Berkeley RCW, Logan NA, Shute LA, Capey AG (1984) Identification of *Bacillus* species. In: *Methods in Microbiology*. Vol. **16**, pp. 291-328, Ed.: Bergan T. Academic Press, London.

Beye M, Poch A, Burgtorf C, Moritz RFA, Lehrach H (1998) A gridded genomic library of the honeybee (*Apis mellifera*): A reference library system for basic and comparative genetic studies of a hymenopteran genome. *Genomics*, **49**: 317-320.

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.

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

Binnie C, Cossar JD, Stewart DI (1997) Heterologous biopharmaceutical protein expression in *Streptomyces*. *Trends Biotechnol*., **15**: 315-320.

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

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

Brodersen DE, Clemons WM, Carter AP, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V (2000) The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell*, **103**: 1143-1154.

Brodersen DE, Clemons WM JR, Carter AP, Wimberly BT, Ramakrishnan V (2002) Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA**.** *J. Mol. Biol*., **316**: 725-768.

Brodersen DE, Clemons WM Jr, Carter AP, Wimberly BT, Ramakrishnan V (2003) Phasing the 30S ribosomal subunit structure. *Acta Crystallogr. D. Biol. Crystallogr*., **59**: 2044-2050.

Burgtorf C, Welzel K, Hasenbank R, Zehetner G, Weis S, Lehrach H (1998) Gridded genomic libraries of different chordate species: a reference library system for basic and comparative genetic studies of chordate genomes. *Genomics*, **52**: 230-232.

Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, **407**: 340-348.

Cate JH, Yusupov MM, Yusupov GZ, Earnest TN, Noller HF (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science*, **285**: 2095-2104.

Cech TR (2000) Structural biology. The ribosome is a ribozyme**.** *Science,* **289**: 878-879.

Chadwick D, Whelan J (1992) Secondary Metabolites: Their Function and Evolution, Ciba Foundation Symp. 171. John Wiley and Sons, Chichester.

Challis GL, Hopwood DA (2003) Synergy and contingency as deriving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. In conference: *Chemical Communication in a Post-Genomic World*. Vol. **100**, pp. 14555-14561, Ed.: Sackler AM. National Academy of Sciences and Engineering, Irvine, Canada.

Chater KF, Hopwood DA, Kieser T, Thomposon CJ (1982) Gene cloning in *Streptomyces*. *Curr. Top. Microbio. Immunol*., **96**: 69-95.

Clemons WM, May JL, Wimberly BT, McCutcheon JP, Capel MS, Ramakrishnan V (1999) Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution. *Nature*, **400**: 833-840.

Cooper AJL (1978) Spot test for the determination of α-ketoglutaramic acid in human cerebrospinal fluid. *Analytical Biochem*., **90**: 444-446.

Cundliffe E (1989) How antibiotic-producing organisms avoid suicide. *Annu. Rev. Microbiol*., **43**: 207-233.

Dairi T, Hasegawa M (1989) Common biosynthetic feature of fortimicin-group antibiotics. *J. Antibiot.*, **42**: 934-943.

Dairi T, Ohta T, Hashimoto E, Hasegawa M (1992a) Self cloning in *Micromonospora olivasterospora* of fms genes for fortimicin A (astromicin) biosynthesis. *Mol. Gen. Genet*., **232**: 262-270.

Dairi T, Ohta T, Hashimoto E, Hasegawa M (1992b) Organization and nature of fortimicin A (astromicin) biosynthetic genes studied using a cosmid library of *Micromonospora olivasterospora* DNA. *Mol. Gen. Genet*., **236**: 39-48.

Dairi T, Yamaguchi K, Hasegawa M (1992c) N-formimidoyl fortimicin A synthase, a unique oxidase involved in fortimicin A biosynthesis: purification, characterization and gene cloning. *Mol. Gen. Genet*., **236**(1): 49-59.

DeLey J (1978) Modern molecular methods in bacterial taxonomy: Evaluation, applications, prospects. Proc. 4th Int. Conf. *Plant Path. Bacteriol*. pp. 347-357.

Demain AL, Inamine E (1970) Biochemistry and regulation of streptomycin and mannosidstreptomycinase (α-D-mannosidase) formation. *Bacteriol. Rev*., **34**: 1-19.

Demydchuk J, Oliynyk Z, Fedorenko V (1998) Analysis of a kanamycin resistance gene (kmr) from *Streptomyces kanamyceticus* and a mutant with increased aminoglycoside resistance. *J. Basic Microbiol*., **38**: 231-239.

Distler J, Klier K, Piendl W, Böck A, Kresze G, Piepersberg W (1985) Streptomycin biosynthesis in *Streptomyces griseus*. I. Characterizaton of streptomycin-idiotrophic mutants. *FEMS Microbiol. Lett*., **30**: 145-150.

Distler J, Ebert A, Mansouri K, Pissowotzki K, Stockmann M, Piepersberg W (1987) Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: nucleotide sequence of three genes and analysis of transcriptional activity. *Nucleic Acids Res*., **15**: 8041-8056.

Doumith M, Weingarten P, Wehmeier UF, Salah-Bey K, Benhamou B, Capdevila C, Michel JM, Piepersberg W, Raynal MC (2001) Analysis of genes involved in 6-deoxyhexose biosynthesis and transfer in *Saccharopolyspora erythraea. Mol. Gen. Genet*., **264**: 477-485.

Drepper A, Peitzmann R, Pape H (1996) Maltokinase (ATP: maltose 1-phosphotransferase) from *Actinoplanes* sp.: demonstration of enzyme activity and characterization of the reaction product. *FEBS Lett*., **388**: 177-179.

Dring GJ, Gould GW, Eller DJ (1984) Fundamental and applied aspects of bacterial spores, Academic Press, London.

Enquist LW, Bradley SG (1971) Characterization of deoxyribonucleic acid from *Streptomyces venezuelae* species. *Dev. Ind. Microbiol.*, **12:** 225-236.

Ensign JC (1978) Formation, properties, and germination of actinomycete spores. *Annu. Rev. Microbiol.*, **32**: 185-219.

Famy F, Flossdorf J, Claus D (1985) The DNA base composition of the type strains of the genus *Bacillus*. *Syst. Appl. Microbiol*., **6**: 60-65.

Fedorenko VA, Golets LM, Demidchuk IuA, Kriugel G (1998) Analysis of genome rearrangements in *Streptomyces kanamyceticus* mutants. *Antibiot. Khimioter.*, **43**: 14-19.

Fiedler HP, Welzel K, Vente A, Bechthold A, Trefzer A, Pelzer S, Schimana J, Stockert S, Bihlmaier C (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob. Agents Chemother.*, **46**: 1174-1182.

Fong DH, Berghuis AM (2002) Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry. *EMBO* (*European Molecular Biology Organization*), **21**: 2323-2331.

Fourmy D, Recht MI, Blanchard SC, Puglisi JD (1996) Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science*, **274**: 1367-1371.

Frommer W, Junge B, Muller L, Schmidt D, Truscheit E (1981) New enzyme inhibitors from microorganisms *Planta Med.,* **35**: 195-217.

Galperin MY (2004) Bacterial signal transduction network in a genomic perspective. *Environ. Microbiol.*, **6**: 552-567.

Golovacheva RS, Loginova Lg, Salikhov TA, Kolesnikov AA, Zaitseva GN (1975). A new thermophilic species, *Bacillus thermocatenulatus*. *Mikrobiologiya*, **44**: 265-268; Microbiology, **44**: 230-233.

Grisebach H (1978) Biosynthesis of sugar components of antibiotic substances. *Adv. Carbohydr. Chem. Biochem.*, **35**: 81-126.

Haddad J, Vakulenko S, Mobashery S (1999) An antibiotic cloaked by its own resistance enzyme, *J. Am. Chem. Soc.*, **121**: 11922-1192.

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

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

Harayama S, Yamamoto S, (1996) Phylogenetc identification of Pseudomonas strains based on a comparison of *gyrB* and *rpoD* sequences. In: *Molecular Biology of Pseudomonads*. pp. 250-258, Ed.: Nakazawa T, Furukawa K, Haas D, Silver S. American Society for Microbiology, Washington, USA.

Hardick DJ, Hutchinson DW, Trew SJ, Wellington, EMH (1992). Glucose is a precursor of 1-deoxynojirimycin and 1-deoxymannonojirimycin in *Streptomyces subrutilus. Tetrahedron,* **48**: 6285-6296.

Hasegawa M (1992) A novel, highly efficient gene-cloning system in *Micromonospora* applied to the genetic analysis of fortimicin biosynthesis. *Gene*, **115**: 85-91.

Herbert CJ, Sarwar M, Ner SS, Giles IG, Akhtar M (1986) Sequence and interspecies transfer of an aminoglycoside phosphotransferase gene (APH) of *Bacillus circulans*. Self-defence mechanism in antibiotic-producing organisms. *J. Biochem.*, **233**: 383-393.

Higgins DG, Bleasby AJ, Fuchs R (1991) CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.*, **8**: 198-191.

Holm SE, Hill B, Lowestad A, Maller R, Vikefors A (1983) A prospective, randomized study of amikacin and gentamicin in serious infections with focus on efficacy, toxicity and duration of serum levels above the MIC. *J. Antimicrob. Chemother*., **12:** 393-402.

Holmes DJ, Drocourt D, Tiraby G, Cundliffe E (1991) Cloning of an aminoglycoside-resistance-encoding gene, *kamC*, from *Saccharopolyspora hirsuta*: comparison with *kamB* from *Streptomyces tenebrarius*. *Gene*, **102**: 19-26.

Hopwood DA, Wright HM (1978) Bacterial protoplast fusion: recombination in fused protoplasts of *Streptomyces coelicolor*. *Mol. Gen. Genet*., **162**: 307-317.

Hopwood DA (2003) The *Streptomyces* genome-be prepared*! Nat. Biotechnol*., **21**: 505-506.

Hoshiko S, Nojiri C, Matsunaga K, Katsumata K, Satoh E, Nagaoka K (1988) Nucleotide sequence of the ribostamycin phosphotransferase gene and of its control region in *Streptomyces ribosidificus*. *Gene*, **68**: 285-296.

Hotta K, Morioka M, Tohyama H, Okami Y (1989a) Biosynthesis of istamycins by *Streptomyces tenjimariensis*. In*: Trends* in *Actinomycetology In Japan*. pp. 61-64, Ed.: Koyama Y. Society for Actinomycetes, Tokyo, Japan.

Hotta K, Morioka M, Okami Y (1989b) Biosynthetic similarity between *Streptomyces tenjimariensis* and *Micromonospora olivasterospora* which produce fortimicin-group antibiotics. *J. Antibiot.*, **42**: 745-751.

Hotta K, Davies J, Yagisawa M (1994) Aminogylcosides and aminocyclitols (other than streptomycin). (Vining L, Stuttard C, Eds.), In: *Biochemistry and Genetics of Antibiotics Biosynthesis*. pp. 571-595, Stoneham, Butterworth-Heinemann.

Hurst A, Gould GW (1984) The Bacterial Spore. Vol. 2. Academic press, London.

Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnol*., **21**: 526-531.

Inoue M, Nonoyama M, Okamoto R, Ida T (1994) Antimicrobial activity of arbekacin, a new aminoglycoside antibiotic, against methicillin-resistant *Staphylococcus aureus*. *Drugs Exp. Clin. Res.*, **20**: 233-239.

Ishikawa J, Hotta K (1999) Frameplot: a new implementation of the Frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiol. Lett*., **174**: 251-253.

Itoh S, Odakura Y, Kase H, Satoh S, Takahashi K, Iida T, Shirahata K, Nakayama K (1984) Biosynthesis of astromicin and related antibiotics. I. Biosynthetic studies by bioconversion experiments. *J. Antibiot*., **37**: 1664-1669.

Jendrossek D, Tomasi G, Kroppenstedt RM (1997) Bacterial degradation of natural rubber: a privilege of actinomycetes. *FEMS Microbiol. Lett*., **150**: 179-188.

Joe YA, Goo YM (1998) Kanamycin acetyltransferase gene from kanamycin-producing *Streptomyces kanamyceticus* IFO 13414. *Arch. Pharm. Res*., **21**: 470-474.

Joung KB, Côté JC (2002) Evaluation of ribosomal RNA gene restriction patterns for the classification of *Bacillus* species and related genera. *J. Applied Microbiology*, **92**: 97-108.

Kameda Y, Asano N, Yoshikawa M, Takeuchi M, Yamaguchi T, Matsui K, Horii S, Fukase H (1984) Valiolamine, a new α-glucosidase inhibiting aminocyclitol produced by *Streptomyces hygroscopicus*. *J. Antibiot.*, **37**: 1301-1307.

Karagouni AD, Vionis AP, Baker PW, Wellington EMH (1993) The effect of soil moisture content on spore germination, mycelium development and survival of a seeded streptomycetes in soil. *Microbiol. Release*, 2: 47-51.

Kasai H, Watanabe K, Gasteiger E, Bairoch A, Isono K, Yamamoto S, Harayama S (1998) Construction of the *gyrB* database for the identification and classification of bacteria. Universal Academy Press, *Genome Inform.*, **9**: 13-21.

Kase H, Odakura Y, Takazawa Y, Kitamura S, Nakayama K (1982) Biosynthesis of sagamicicn and related aminoglycosides. in: *Trends in Antibiotic Research. Genetics, Biosyntheses, Actions, and New Substances.* pp. 195-212, Ed.: Umezawa H, Demain AL, Hata T, Hutchinson CR. Antibiotics Research Association, Tokyo, Japan.

Kawaguchi H, Naito T, Nakagawa S, Fujisawa KI (1972) BB-K 8, a new semisynthetic aminoglycoside antibiotic. *J. Antibiot.*, **25**: 695-708.

Kelemen GH, Cundliffe E, Financsek I (1991) Cloning and characterization of gentamicin-resistance genes from *Micromonospora purpurea* and *Micromonospora rosea*. *Gene*, **98**: 53-60.

Kharel MK, Subba B, Lee HC, Liou K, Woo JS, Sohng JK (2003) An approach for cloning biosynthetic genes of 2-deoxystreptamine-containing aminocyclitol antibiotics: isolation of a biosynthetic gene cluster of tobramycin from *Streptomyces tenebrarius*. *Biotechnol. Lett*., **25**: 2041-2047.

Kharel MK, Basnet DB, Lee HC, Liou K, Woo JS, Kim BG, Sohng JK (2004a) Isolation and characterization of the tobramycin biosynthetic gene cluster from *Streptomyces tenebrarius*. *FEMS Microbiol. Lett*., **230**: 185-190.

Kharel MK, Subba B, Basnet DB, Woo JS, Lee HC, Liou K, Sohng JK (2004b) A gene cluster for biosynthesis of kanamycin from *Streptomyces kanamyceticus*: comparison with gentamicin biosynthetic gene cluster. *Arch. Biochem. Biophys*., **429**: 204-114.

Kharel MK, Subba B, Lee HC, Liou K, Sohng JK (2005) Characterization of L-glutamine:2-deoxy-*scyllo*-inosose aminotransferase (*tbmB*) from *Streptomyces tenebrarius*. *Bioorg. Med. Chem. Lett.,* **15**: 89-92.

Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA (2000) Practical *Streptomyces* genetics. The John Innes Foundation, John Innes centre, Norwich, UK.

Kobayashi Y, Shiozaki M (1994) Synthesis of trehazolin β-anomer. *J. Antibiot*., **47**: 243-246.

Koch C, Kroppenstedt RM, Stackebrandt E (1996a) Intrageneric relationships of the actinomycete genus *Micromonospora*. *Int. J. Syst. Bacteriol*., **46**: 383-387.

Koch C, Kroppenstedt RM, Rainey FA, Stackebrandt E (1996b) 16S ribosomal DNA analysis of the genera *Micromonospora*, *Actinoplanes, Catellatospora*, *Catenuloplanes*, *Couchioplanes*, *Dactylosporangium*, and *Pilimelia* and emendation of the family *Micromonosporaceae*. *Int. J. Syst. Bacteriol*., **46**: 765-768.

Kojic M, Topisirovic L, and Vasiljevic B (1992) Cloning and characterization of an aminoglycoside resistance determinant from *Micromonospora zionensis*. *J. Bacteriol*., **174**: 7868-7872.

Kondo S, Iinuma K, Yamamoto H, Maeda K, Umezawa H (1973) Letter: Syntheses of 1-n-(S)-4-amino-2-hydroxybutyryl)-kanamycin B and 3',4'-dideoxykanamycin B active against kanamycin-resistant bacteria. *J. Antibiot*., **26**: 412-415.

Kondo S, Horiuchi Y, Ikeda D, Gomi S, Hotta K, Okami Y, Umezawa H (1982) 2"-N-formimidoylistamycin A and B produced by *Streptomyces tenjimariensis*. *J. Antibiot.*, **35**: 1104-1106.

Krassilnikov NA (1950) Actinomycetes antagonists and antibiotic substances. J. Zd. A. Kad. Nauk. SSR. Moscow, USSR. pp. 214-236.

Kudo F, Hosomi Y, Tamegai H, Kakinuma K (1999a) Purification and characterization of 2-deoxy-*scyllo*-inosose synthase derived from *Bacillus circulans*. A crucial carbocyclization enzyme in the biosynthesis of 2-deoxystreptamine-containing aminoglycoside antibiotics. *J. Antibiot*., **52**: 81-88.

Kudo F, Tamegai H, Fujiwara T, Tagami U, Hirayama K, Kakinuma K (1999b) Molecular cloning of the gene for the key carbocycle-forming enzyme in the biosynthesis of 2-deoxystreptamine-containing aminocyclitol antibiotics and its comparison with dehydroquinate synthase. *J. Antibiot.*, **52**: 559-571.

Kudo F, Kawabe K, Kuriki H, Eguchi T, Kakinuma K (2005) A new family of glucose-1-phosphate/glucosamine-1-phosphate nucleotidylyltransferase in the biosynthetic pathways for antibiotics. *J. Am. Chem. Soc*., **127**: 1711-1718.

Kuzuyama T, Seki T, Dairi T, Hidaka T, Seto H (1995) Nucleotide sequence of fortimicin KL1 methyltransferase gene isolated from *Micromonospora olivasterospora*, and comparison of its deduced amino acid sequence with those of methyltransferases involved in the biosynthesis of bialaphos and fosfomycin. *J. Antibiot.*, **48**: 1191-1193.

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

Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Research*, **32**: D142-D144.

Lopez-Cabrera M, Perez-Gonzalez JA, Heinzel P, Piepersberg W, Jimenez A (1989) Isolation and nucleotide sequencing of an aminocyclitol acetyltransferase gene from *Streptomyces rimosus* forma *paromomycinus*. *J. Bacteriol*., **171**: 321-328.

Lucher LA, Chen YM, Walker JB (1989) L-glutamine: keto-*scyllo*-inositol aminotransferase, an enzyme required for biosynthesis of aminocyclitol antibiotics. *Antimicrob. Agents Chemother*., **33**: 452-459.

Luedemann GM, Brodsky B (1964) *Micromonospora carbonacea* sp., an everninomicin-producing organism. *Antimicrobial. Agents Chemother*., **10**: 47-52.

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

Maniatis T, Fritch EF, Sambrook J (1989) Molecular cloning. A laboratory manual. Second edition, Cold Spring Harbour Laboratory Press, New York.

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

Marck C (1988) DNA Strider: a ´C` program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res*., **16**: 1829-1836.

Mayfield CI, Williams ST, Ruddick SM, Hatfield HL (1972) Studies on the ecology of actinomycetes in soil. IV: Observation on the form and growth of strepomycetes in soil. *Soil Biol. Biochem*., **4**: 1213-1220.

Maynes JT, Garen C, Cherney MM, Newton G, Arad D, Av-Gay Y, Fahey RC, James MN (2003) The crystal structure of 1-D-*myo*-inosityl 2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) from *Mycobacterium tuberculosis* reveals a zinc hydrolase with a lactate dehydrogenase fold. *J. Biol. Chem*., **278**: 47166-47170.

Mehling A, Wehmeier UF, Piepersberg W (1995a) Nucleotide sequences of streptomycete 16S ribosomal DNA: towards a specific identification system for streptomycetes using PCR. *Microbiology*, **141**: 2139-2147.

Mehling A, Wehmeier UF, Piepersberg W (1995b) Application of random amplified polymorphic DNA (RAPD) assays in identifying conserved regions of actinomycete genomes. *FEMS Microbiol. Lett*., **128**: 119-126.

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold spring Harbor, pp. 433.

Mingeot-Leclercq MP, Glupczynski Y, Tulkens PM (1999) Aminoglycosides: activity and resistance. *Antimicrob. Agents Chemother*., **43**: 727-737.

Mingeot-Leclercq MP, Tulkens PM (1999) Aminoglycosides: Nephrotoxicity. *Antimicrob. Agents Chemother*., **43**: 1003-1012.

Miyadoh S (1990) A history of systematics and a concept of species in streptomycetes. *Actinomycerol*., **4**: 41-48.

Moazed D, Noller HF (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature*, **327**: 389-394.

Müller L (1989) Chemistry, biochemistry and therapeutic potential of microbial α-glucosidase inhibitors., *In: Novel Microbial products for medicine and agriculture*. pp. 109-116, Ed.: Demain AL, Somkuti GA, Hunter-Creva JC, Rossmoore HW. Elsevier Science Publishers BV, Amsterdam.

Murakami T, Nojiri C, Toyama H, Hayashi E, Katumata K, Anzai H, Matsuhashi Y, Yamada Y, Nagaoka K (1983) Cloning of antibiotic-resistance genes in *Streptomyces*. *J. Antibiot*. **36**: 1305-1311.

Nagabhushan TL, Cooper AB, Tsai H, Daniels PJ, Miller GH (1978) The syntheses and biological properties of 1-N-(S-4-amino-2-hydroxybutyryl)-gentamicin B and 1-N-(S-3-amino-2-hydroxypropionyl)-gentamicin B. *J. Antibiot.*, **31**: 681-687.

Nakamura LK, Swezey J, (1983) Taxonomy of *Bacillus circulans*, Jordan 1890: base composition and reassociation of deoxyribonucleic acid. *Int. J. Syst. Bacteriol*., **33**: 46-52.

Nakano MM, Butsuya I, Ogawara H (1989) Expression of the kanamycin resistance gene in a kanamycin-producing strain of *Streptomyces kanamyceticus*. *J. Antibiot*., **42**: 423-430.

Nam DH, Ryu DY (1985) Relationship between butirosin biosynthesis and sporulation in *Bacillus circulans*. *Antimicrob. Agents and Chemo*., **27**: 798-801.

Nango E, Kudo F, Eguchi T, Kakinuma K (2003) Reaction stereochemistry of 2-deoxy-*scyllo*-inosose synthase, the key enzyme in the biosynthesis of 2-deoxystreptamine. *Chemistry Letters*, **32**: 438-439.

Nango E, Eguchi T, Kakinuma K (2004) Active site mapping of 2-deoxy-*scyllo*-inosose synthase, the key starter enzyme for the biosynthesis of 2-deoxystreptamine. Mechanism-based inhibition and identification of lysine-141 as the entrapped nucleophile. *J. Org. Chem*., **69**: 593-600.

Neonakis I, Gikas A, Scoulica E, Manios A, Georgiladakis A, Tselentis Y (2003) Evolution of aminoglycoside resistance phenotypes of four Gram-negative bacteria: an 8-year survey in a University Hospital in Greece. *Int. J. Antimicrob. Agents*, **22**: 526-531.

Newton GL, Bewley CA, Dwyer TJ, Horn R, Aharonowitz Y, Cohen G, Davies J, Faulkner DJ, Fahey RC (1995) The structure of U17 isolated from *Streptomyces clavuligerus* and its properties as an antioxidant thiol. *Eur. J. Biochem*., **230**: 821-825.

Newton GL, Av-Gay Y, Fahey RC (2000) A novel mycothiol-dependent detoxification pathway in mycobacteria involving mycothiol S-conjugate amidase. *Biochemistry*, **39**: 10739-10746.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis**.** *Science*, **289**: 920-930.

Nizetic D, Zehetner G, Monaco AP, Gellen L, Young BD, Lehrach H (1991) Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proc. Natl. Acad. Sci*., **88**: 3233-3237.

Norris JR, Berkeley RCW, Logan NA (1981) The genera *Bacillus* and *Sporolactobacillus*. In: *The Prokaryotes*; A handbook on habitats, isolation and identification of bacteria. Vol. **2**, pp. 1711-1741, Ed.: Starr MP, Stolp H, Truper HG. Springer-Verlag, New York.

Norin A, Van Ophem PW, Piersma SR, Persson B, Duine JA, Jornvall H (1997) Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium-chain dehydrogenase/reductase, phylogenetically links different eukaroytic alcohol dehydrogenases-primary structure, conformational modelling and functional correlations. *Eur. J. Biochem*., **248**: 282-289.

Novy R, Drott D, Yaeger K, Mierendorf R (2001) Overcoming the codon bias of *E. coli* for enhanced protein expression. *InNovations*, newsletter of Novagen inc. Advanced products and protocols for molecular biology research, *InNovations*, **12**: 1-3.

Numata K, Satoh F, Hatori M, Miyaki T, Kawaguchi H (1986) Isolation of 3,3'-neotrehalosadiamine (BMY-28251) from a butirosin-producing organism. *J. Antibiot.*, **39**: 1346-1348.

Odakura Y, Kase H, Itoh S, Satoh S, Takasawa S, Takahashi K, Shirahata K, Nakayama K (1984) Biosynthesis of astromicin and related antibiotics. II. Biosynthetic studies with blocked mutants of *Micromonospora olivasterospora*. *J. Antibiot*., **37**: 1670-1680.

Odakura Y, Kase H, Nakayama K (1983) Sagamicin and the related aminoglycosides: fermentation and biosynthesis. III. Isolation and characterization of *Micromonospora sagamiensis* mutants blocked in gentamicin C1 pathway. *J. Antibiot*., **36**: 125-130.

Oh SH, Chater KF (1997) Denaturation of circular or linear DNA facilitates targeted integrative transformation of *Streptomyces coelicolor* A3(2): Possible relevance to other organisms. *J. Bacteriol*., **179**: 122-127.

Ohta T, Dairi T, Hasegawa M (1993) Characterization of two different types of resistance genes among producers of fortimicin-group antibiotics. *J. Gen. Microbiol*., **139**: 591-599.

Ohta T, Hasegawa M (1993a) Analysis of the nucleotide sequence of *fmrT* encoding the self-defence gene of the istamycin producer, *Streptomyces tenjimariensis* ATCC 31602; comparison with the sequences of *kamB* of *Streptomyces tenebrarius* NCIB 11028 and *kamC* of *Saccharopolyspora hirsuta* CL102. *J. Antibiot*., **46**: 511-517.

Ohta T, Hasegawa M (1993b) Analysis of the self-defence gene (*fmrO*) of a fortimicin A (astromicin) producer, *Micromonospora olivasterospora*: comparison with other aminoglycoside-resistance-encoding genes. *Gene*, **127**: 63-69.

Ohta T, Hashimoto E, Hasegawa M (1992a) Characterization of sannamycin A-nonproducing mutants of *Streptomyces sannanensis*. *J. Antibiot*., **45**: 289-291.

Ohta T, Hashimoto E, Hasegawa M (1992b) Cloning and analysis of a gene (*sms13*) encoding sannamycin B-glycyltransferase from *Streptomyces sannanensis* and its distribution among actinomycetes. *J. Antibiot*., **45**: 1167-1175.

Okuda T, Ito Y (1982) Biosynthesis and mutasynthesis of aminoglycoside antibiotics, In: *Aminoglycoside Antibiotics.* pp. 111-203, Ed.: Umezawa H, Hooper IR. Springer-Verlag, Berlin.

Omura S (1992) Thom Award Lecture. Trends in the search for bioactive microbial metabolites. *J. Ind. Microbiol*., **10**: 135-156.

Ørskov J (1923) Investigation into the morphology of the ray fungi. Levin and Munksgaard. Copenhagen, Denmark.

Ota Y, Tamegai H, Kudo F, Kuriki H, Koike-Takeshita A, Eguchi T, Kakinuma K (2000) Butirosin-biosynthetic gene cluster from *Bacillus circulans*. *J. Antibiot*., **53**: 1158-1167.

Pape T, Wintermeyer W, Rodnina MV (2000) Conformational switch in the decoding region of 16S rRNA during aminoacyl-tRNA selection on the ribosome. *Nature Struct. Biol*., **7:** 104-107.

Pearce CJ, Rinehart KL (1981) Biosynthesis of aminocyclitol antibiotics. In: *Antibiotics Biosynthesis.* Vol. **4**, pp. 74-100, Ed.: Corcoran JW. Springer-Verlage, Berlin.

Person WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.*, **85**: 2444-2448.

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

Piendl W, Bock A, Cundliffe E (1984) Involvement of 16S ribosomal RNA in resistance of the aminoglycoside-producers *Streptomyces tenjimariensis*, *Streptomyces tenebrarius* and *Micromonospora purpurea*. *Mol. Gen. Genet.*, **197**: 24-29.

Piepersberg W (1992) In: *Secondary Metabolites*: Their function and Evolution. pp. 294-304, Ed.: Chadwick D and Whelan J. Ciba Foundation Symp. 171, John Wiley and Sons, Chichester.

Piepersberg W (1993) *Streptomyces* and *Corynebacteria*. In: *Biotechnology*, pp. 434-468, Ed.: Rehm H, Reed G, Pühler A, Stadler P. VCH, Weinheim.

Piepersberg W (1994) Pathway engineering in secondary metabolite-producing actinomycetes. *Crit. Rev. Biotechnol.*, **14**: 251-285.

Piepersberg W (1995) Streptomycin and related aminoglycosides. *Biotechnology*, **28**: 531-570.

Piepersberg W (1997) Molecular biology, biochemistry, and fermentation of aminoglycoside antibiotics (Ch. 4). In: *Biotechnology of Industrial Antibiotics* (2nd eds.). Ed.: Strohl WR. pp. 81-163, Marcel-Dekker Inc., New York.

Piepersberg W, Distler J (1997) Aminoglycosides and sugar components in other secondary metabolites. In: *Biotechnology* (2nd eds.); Products of Secondary Metabolism, Vol. **7**, pp. 397-488, Ed.: Rehm HJ, Reed G, Pühler A, Stadler P; Vol. Ed.: Kleinkauf H, Döhren HV. VCH-Verlagsgesellschaft, Weinheim.

Piepersberg W (2001) Glycosylation of antibiotics and other agents from actinomycetes. In: *Novel frontiers in the production of compounds for biomedical use*. Vol. **1**, pp. 161-168, Ed.: Van Broekhoven A, Shapiro F, Anne J. Kluwer academic publishers, Dordrecht, Boston, London.

Piepersberg W (2002) In: *Molecular Medical Microbiology*. Vol. **1**, pp. 562-594. Ed.: Sussman M. Academic Press, San Diego.

Piepersberg W, Diaz-Guardamino Uribe PM, Stratmann A, Thomas H, Wehmeier U, Zhang CS (2002) Recent developments in the biosynthesis and regulation of aminoglycosides. In: *Microbial Secondary Metabolites: Biosynthesis, genetics and regulation.* pp. 27-41, Ed.: Fierro F, Martin JF. Research Signpost, Kerala, India.

Piepersberg W (2003) *Streptomyces* and *Corynebacteria.* In: *Biotechnology*. pp. 434-468, Ed.: Rehm H, Reed G, Pühler A, Stadler P. VCH, Weinheim.

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

Priest FG (1993) Systemics and ecology of *Bacillus subtilis* and other Gram positive bacteria. In: *Biochemistry, Physiology and Molecular Genetics*. pp. 3-16, Ed.: Sonenshein AL, Hoch JA, Losick R. American Society for Microbiology, Washington.

Prins JM, Weverling GJ, de Blok K, van Ketel RJ, Speelman P (1998) Validation and nephrotoxicity of a simplified once-daily aminoglycoside dosing schedule and guidelines for monitoring therapy. *Antimicrob. Agents Chemother*., **40**: 2494-2499.

Retzlaff L, Distler J (1995) The regulator of streptomycin gene expression, StrR, of *Streptomyces griseus* is a DNA binding activator protein with multiple recognition sites. *Mol. Microbiol*., **18**: 151-162.
Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet*., **16**: 276-277

Richardson MA, Kuhstoss S, Solenberg P, Schaus NA, Rao RN (1987) A new shuttle cosmid vector, pKC505, for streptomycetes: its use in the cloning of three different spiramycin-resistance genes from a *Streptomyces ambofaciens* library. *Gene*, **61**: 231-241.

Rinehart KL, Stroshane RM (1976) Biosynthesis of aminocyclitol antibiotics. *J. Antibiot*., **19**: 319-353.

Rinehart KL (1979) Biosynthesis and mutasynthesis of aminocyclitol antibiotics. *J. Antibiot*., **32**: 32-46.

Rinehart KL, Snyde WC, Staley AL, Lau RCM (1992). In: *Secondary Metabolite Biosynthesis and Metabolism*. pp. 41-60, Ed.: Petrosski RJ, McCormick SP. Plenum Press, New York.

Rheims H, Schumann P, Rohde M, Stackebrandt E (1998) *Verrucosispora gifhornensis* gen. nov., sp. nov., a new member of the actinobacterial family *Micromonosporaceae*. *Int. J. Syst. Bacteriol*., **48**: 1119-1127.

Robson GD, Kuhn PJ, Trinci AP (1988) Effects of validamycin A on the morphology, growth and sporulation of *Rhizoctonia cerealis*, *Fusarium culmorum* and other fungi. *J. Gen. Microbiol*., **134**: 3187-3194.

Sakuda S, Isogai A, Matsumoto S, Suzuki A (1987) Search for microbial insect growth regulators. II. Allosamidin, a novel insect chitinase inhibitor. *J. Antibiot*., **40**: 296-300.

Salauze D, Perez-Gonzalez JA, Piepersberg W, Davies J (1991) Characterisation of aminoglycoside acetyltransferase-encoding genes of neomycin-producing *Micromonospora chalcea* and *Streptomyces fradiae*. *Gene*, **101**: 143-148.

Sanger F, Nicklen S, Coulsen AR (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci.*, **74**: 5463-5467.

Sambrook J, Russell DW (2001) *Molecular cloning*: *a laboratory manual*, 3rd Eds. Cold Spring harbor laboratory Press, Cold Spring Harbor, New York.

Schleifer KH, Kandler O (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.*, **36**: 407-477.

Schlünzen F, Zarivach R, Harms J, Bashan A, Tocilj A, Albrecht R, Yonath A, Franceschi F (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, **413**: 814-821.

Shaw KJ, Rather PN, Hare RS, Miller GH (1993) Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminogylcoside-modifying enzymes. Microbiological reviews, American Society for Microbiology, **57**: 138-163.

Skerman VBD, McGowan V, Sneath PHA (1980) Approved lists of bacterial names. *Int. J. Syst. Bacteriol*., **30**: 225–420.

Sneath PH, Mair NS, Sharpe ME, Holt JG (1986) Bergey's Manual of Systemic Bacteriology, Vol.2, pp. 1104-1138. Williams and Wilkins, Baltimore, London, Loss Angeles, Sydney.

Stackebrandt E, Rainey FA, Ward-Rainey NT (1997) Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int. J. Syst. Bacteriol*., **47**: 479-491.

Stratmann A, Mahmud T, Lee S, Distler J, Floss HG, Piepersberg W (1999) The AcbC protein from *Actinoplanes* species is a C7-cyclitol synthase related to 3-dehydroquinate synthases and is involved in the biosynthesis of α-glucosidase inhibitor acarbose. *J. Biol. Chem*., **274**: 10889-10896.

Studier W, Rosenberg A, Dunn J, Dubendroff J (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol*., **185**: 61-89.

Suzukake K, Tokunaga K, Hayashi H, Hori M, Uehara Y, Ikeda D, Umezawa H (1985) Biosynthesis of 2-deoxystreptamine. *J. Antibiot*., **38**: 1211-1218.

Takeuchi M, Takai N, Asano N, Kameda Y, Matsui K (1990) Inhibitory effect of validamine, valienamine and valiolamine on activities of carbohydrases in rat small intestinal brush border membranes. *Chem. Pharm. Bull*., **38**: 1970-1972.

Tamegai H, Eguchi T, Kakinuma K (2002a) First identification of *Streptomyces* genes involved in the biosynthesis of 2-deoxystreptamine-containing aminoglycoside antibiotics-genetic and evolutionary analysis of L-glutamine:2-deoxy-*scyllo*-inosose aminotransferase genes. *J. Antibiot*., **55**: 1016-1018.

Tamegai H, Nango E, Kuwahara M, Yamamoto H, Ota Y, Kuriki H, Eguchi T, Kakinuma K (2002b) Identification of L-glutamine: 2-deoxy-*scyllo*-inosose aminotransferase required for the biosynthesis of butirosin in *Bacillus circulans*. *J. Antibiot*., **55**: 707-714.

Tamegai H, Nango E, Koike-Takeshita A, Kudo F, Kakinuma K (2002c) Significance of the 20-kDa subunit of heterodimeric 2-deoxy-*scyllo*-inosose synthase for the biosynthesis of butirosin antibiotics in *Bacillus circulans*. *Biosci. Biotechnol. Biochem*., **66**: 1538-1545.

Tamura T, Hayakawa M, Hatano K (1997) A new genus of the order Actinomycetales, *Spirilliplanes* gen. nov., with description of *Spirilliplanes yamanashiensis* sp. nov. *Int. J. Syst. Bacteriol*., **47**: 97-102.

Tanaka N, Matsunaga K, Hirata A, Matsuhisa Y, Nishimura T (1983) Mechanism of action of habekacin, a novel amino acid-containing aminoglycoside antibiotic. *Antimicrob. Agents Chemother*., **24**: 797-802.

Thompson CJ, Gray GS (1983) Nucleotide sequence of a streptomycete aminoglycoside phosphotransferase gene and its relationship to phosphotransferases encoded by resistance plasmids. *Proc. Natl. Acad. Sci*., **80**: 5190-5194.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.,* **22**: 4673-4680.

Topley T, Wilson S (1998) Microbiological and microbial infections, 9th (Eds.), Chapter 31, *Bacillus*, Rosovitz CMJ, Voskuil MI, Chambliss GH, **2**: 709-730.

Trefzer A, Pelzer S, Schimana J, Stockert S, Bihlmaier C, Fiedler HP, Welzel K, Vente A, Bechthold A (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob. Agents Chemother*., **46**: 1174-1182.

Truscheit L, Frommer W, Junge B, Müller L, Schmidt DD, Wingeder W (1981) Chemistry and biochemistry of bacterial α-glucosidase inhibitors, Angew Chem. Int., (Eds.), **20**: 744-761.

Tsuno T, Ikeda C, Numata K, Tomita K, Konishi M, Kawaguchi H (1986) 3,3'-Neotrehalosadiamine (BMY-28251), a new aminosugar antibiotic. *J. Antibiot*., **39**: 1001-1003.

Umezawa H, Hooper IR (1982) *Aminoglycoside Antibiotics* (eds), Springer-Verlag, Berlin.

Umezawa S, Kondo S, Ito Y (1986) In: *Biotechnology*. Vol. 4, pp. 309-357, Ed.: Rehm HJ, Reed G. VCH verlagsgesellschaft, Weinheim.

Umezawa H, Gomi S, Yamagishi Y, Obata T, Ikeda T, Hamada M, Kondo S (1987) 2"-N-formimidoylsporaricin A produced by *Saccharopolyspora hirsuta* subsp. *kobensis*. *J. Antibiot*., **40**: 91-93.

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

Waksman SA (1959) The Actinomycetes, nature, occurrence and activities. The Williams and Wilkins Co. Baltimore, **1**: 573.

Waksman SA, Lechevalier HA (1962) The Actinomycetes. Vol. I-III. Williams and Wilkins Co. Baltimore, pp. 339.

Walker JB (1971) Enzymatic reactions involved in streptomycin biosynthesis and metabolism. *Lioydia*, **34**: 363-371.

Walker JB (1990) Possible evolutionary relationships between streptomycin and bluensomycin biosynthetic pathways: detection of novel inositol kinase and O-carbamoyltransferase activities. *J. Bacteriol*., **172**: 5844-5851.

Walker JB (1995) Enzymatic synthesis of aminocyclitol moieties of aminoglycoside antibiotics from inositol by *Streptomyces* sp.: detection of glutamine-aminocyclitol aminotransferase and diaminocyclitol aminotransferase activities in a spectinomycin producer. *J. Bacteriol*., **177**: 818-822.

Walker JB (2002) Enzymatic synthesis of aminoglycoside antibiotics: novel adenosylmethionine:2-deoxystreptamine N-methyltransferase activities in hygromycin B- and spectinomycin-producing *Streptomyces* sp. and uses of the methylated products. *Appl. Environ. Microbiol*., **68**: 2404-2410.

Walker RJ, Duggin GG (1988) Drug nephrotoxicity. *Annu. Rev. Pharmcol. Toxicol*., **28**: 331-345.

Weber T, Welzel K, Pelzer S, Vente A, Wohlleben W (2003) Exploiting the genetic potential of polyketide producing streptomycetes. *J. Biotechnol*., **106**: 221-232.

Wehmeier U (1995) New multifunctional *Escherichia coli-Streptomyces* shuttle vectors allowing positive selection on X-gal plates. *Gene*, **165**: 149-150.

Wellington EMH, Cresswell N, Herron PR (1992) Gene transfer between streptomycetes in soil. *Gene,* **115**: 193-198.

Widdick DA, Dodd HM, Barraille P, White J, Stein TH, Chater KF, Gasson MJ, Bibb MJ (2003) Cloning and engineering of the cinnamycin biosynthetic gene cluster from *Streptomyces cinnamoneus* DSM 40005. *Proc. Natl. Acad. Sci*., **100**: 4316-4321.

Woese CR, Fox GE, Zablen L, Uchida T, Bonen L, Pechman K, Lewis BJ, Stahl D (1975) Conservation of primary structure in 16S ribosomal RNA. *Nature*, **254**: 83-86.

Wright F, Bibb MJ (1992) Codon usage in the G+C-rich *Streptomyces* genome**.** *Gene*, **113**: 55-65.

Yanai K, Murakami T (2004) The kanamycin biosynthetic gene cluster from *Streptomyces kanamyceticus*. *J. Antibiot*., **57**; 351-354

Yamamoto S, Harayama S (1995) PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl. Environ. Microbiol*., **61**: 1104-1109.

Yamamoto S, Harayama S (1996) Phylogenetic analysis of *Acinetobacter* strains based on the nucleotide sequences of *gyrB* genes and on the amino acid sequences of their products. *Int. J. Syst. Bacteriol.*, **46**: 506-511.

Yamamoto S, Harayama S (1998) Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int. J. Syst .Bacteriol.*, **48**: 813-819.

Yamamoto S, Bouvet PJ, Harayama S (1999) Phylogenetic structures of the genus *Acinetobacter* based on *gyrB* sequences: comparison with the grouping by DNA-DNA hybridization. *Int. J. Syst. Bacteriol.*, **49**: 87-95.

Yamauchi N, Kakinuma K (1993) Biochemical studies on 2-deoxy-*scyllo*-inosose, an early intermediate in the biosynthesis of 2-deoxystreptamine. IV. A clue to the similarity of 2-deoxy-*scyllo*-inosose synthase to dehydroquinate synthase. *J. Antibiot.*, **46**: 1916-1918.

Yukita T, Nishida H, Eguchi T, Kakinuma K (2003) Biosynthesis of (2R)-4-amino-2-hydroxybutyric acid, unique and biologically significant substituent in butirosins. *J. Antibiot*., **56**: 497-500.

Zaehner H, Fielder FP (1995) Fifty years of antimicrobials: past perspectives and future trends. In: *The Need for New Antibiotics*: *Possible Ways Forward.* pp. 67-84, Ed.: Hunter PA, Darby GK, Russell NJ. Fifty-Third Symposium of the Society for General microbiology, Cambridge, UK.

Zalacain M, Gonzalez A, Guerrero MC, Mattaliano RJ, Malpartida F, Jimenez A (1986) Nucleotide sequence of the hygromycin B phosphotransferase gene from *Streptomyces hygroscopicus*. *Nucleic Acids Res*., **14**: 1565-1581.

Zembower TR, Noskin GA, Postelnick MJ, Nguyen C, Peterson LR (1998) The utility of aminoglycosides in an era of emerging drug resistance. *Int. J. Antimicrob. Agents*, **10**: 95-105.

Zhang CS, Stratmann A, Block O, Bruckner R, Podeschwa M, Altenbach HJ, Wehmeier UF, Piepersberg W (2002) Biosynthesis of the C(7)-cyclitol moiety of acarbose in *Actinoplanes* species SE50/110. 7-O-phosphorylation of the initial cyclitol precursor leads to proposal of a new biosynthetic pathway. *J. Biol. Chem*., **277**: 22853-22862.

6.1 Tables of proteins encoded in the genomic region covering the biosynthetic gene clusters of the ACAGAs analysed

Tab. A.1 Proteins encoded in the genomic region covering the *neo*-cluster of *S. fradiae* DSM 40063 (accession code: AJ629247)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in NM pathway or else [a]
SfrA10.1		544		NP_823514.1	*S. avermitilis* MA-4680	HP
SfrA10.2		311		NP_823513.1	*S. avermitilis* MA-4680	HP
SfrA10.3		1793		NP_823512.1	*S. avermitilis* MA-4680	multidomain regulator (AfsR-family)
SfrA10.4		82		NP_827553.1	*S. avermitilis* MA-4680	pectin esterase (fragment)
SfrA10.5		180		NP_629028.1	*S. coelicolor* A3(2)	membrane protein (fragment)
SfrA10.6c		98		NP_828422.1	*S. avermitilis* MA-4680	HP (fragment)
SfrA10.7c	*aphA*	268	AphA	AAA26699.1	*S. fradiae*	NM phosphotransferase, APH(3')-V, NM-resistance
SfrA10.8c	*neoG*	431	NeoG	CAD60536.1	*S. cinnamoneus*	component of sensor/response regulator syst. (?)
SfrA10.9c	*neoH*	173	NeoH	CAD60535.1	*S. cinnamoneus*	component of sensor/response regulator syst. (?)
SfrA10.10c	*neoI*	175	NeoI	CAD60534.1	*S. cinnamoneus*	component of sensor/response regulator syst. (?)
SfrA10.11	*neoE*	340	NeoE	CAE22477.1	*S. tenebrarius*	aminocyclitol 1-dehydrogenase
SfrA10.12	*neoS*	424	NeoS	CAE22472.1	*S. tenebrarius*	L-glutamine:ketocyclitol aminotransferase, AT-I + -II
SfrF04.2	*neoC*	430	NeoC	CAE22471.1	*S. tenebrarius*	2-deoxy-*scyllo*-inosose synthase (cyclase)
SfrF04.3	*neoM*	421	NeoM	BAD20768.1	*S. kanamyceticus*	2DOS 4-glucosaminyltransferase, GT-I
SfrF04.4	*neoT*	666	NeoT	NP_823966.1	*S. avermitilis* MA-4680	ABC transporter, ATP binding; NM(-P) export
SfrF04.5	*neoU*	594	NeoU	NP_629587.1	*S. coelicolor* A3(2)	ABC transporter, ATP binding; NM(-P) export
SfrF04.6	*neoQ*	541	NeoQ	AAR98543.1	*M. echinospora*	NM 6'-(6'''-) dehydrogenase
SfrF04.7	*neoN*	299	NeoN	NP_623366.1	*Thermoanaerobacter tengcongensis*	Fe-S oxidoreductase; (5'''-epimerase or 6'''-DH)
SfrF04.8	*neoP*	233	NeoP	BAC41220.1	*B. circulans*	sugar phosphate phosphatase (?)
SfrF04.9	*neoX*	83	NeoX	AAL18481.1	*Photorhabdus luminescens*	cHP, (5'''-epimerase ?)
SfrF04.10	*neoF*	366	NeoF	NP_630571.1	*Nostoc* sp. PCC 7120	glycosyltransferase, put. GT-III
SfrF04.11	*neoD*	279	NeoD	BAD20763.1	*S. kanamyceticus*	TTP-D-glucosmamine synthase
SfrF04.12c	*neoL*	660	NeoL	BAC41207.1	*B. circulans*	unknown, poss. involved in ribosyltransfer (?)
SfrF04.13	*neoB*	416	NeoB	AAR98542.1	*M. echinospora*	NM 6'-/6'''-aminotransferase
SfrF04.14	*neoA*	1293	NeoA	AAL18480	*Photorhabdus luminescens*	unknown, poss. involved in ribosyltransfer (?)
SfrF04.15c	*aacC8*	287	AacC8	AAA26685.1	*S. fradiae*	aminoglycoside N-3-acetyltransferase, NM resistance

Tab. A.1 Proteins encoded in the genomic region covering the *neo*-cluster of *S. fradiae* DSM 40063 (accession code: AJ629247), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in NM pathway or else [a]
SfrF04.16c	*neoR*	886	NeoR	NP_822883.1	*S. avermitilis* MA-4680	put. regulator
SfrF04.17	(*neoY*)	73	(NeoY)	O53810	*Mycob. tuberculosis*	cHP
SfrF04.18		173		NP_631038.1	*S. coelicolor* A3(2)	cHP
SfrF04.19c		140		NP_631360.1	*S. coelicolor* A3(2)	cHP
SfrF04.20c		454		NP_628830.1	*S. coelicolor* A3(2)	serine protease
SfrF04.21c		316		NP_628831.1	*S. coelicolor* A3(2)	regulator (lysR family)
SfrF04.22	*icdA*	739		NP_828390.1	*S. avermitilis* MA-4680	isocitrate dehydrogenase
SfrF04.23		629		NP_733614.1	*S. coelicolor* A3(2)	heat shock protein 70
SfrF04.24		113		NP_630232.1	*S. coelicolor* A3(2)	HP
SfrF04.25		152		NP_628193.1	*S. coelicolor* A3(2)	membrane protein
SfrF04.26		254		AAN85502.1	*S. atroolivaceus*	HP
SfrF04.27c		128		AAB71209.1	*S. cinnamoneus*	membrane protein (partial ORF))

[a] Abbreviations: AacC = aminoglycoside 3-N-acetyltransferase; acc. code = accession code; AphA = aminoglycoside phosphotransferase; Apr = apramycin; AT = aminotransferase; BU = butirosin; cl. = cluster; (c)HP = conserved hypothetical protein; CPS = cell wall polysaccharide; DH = dehydrogenase; 2DOS = 2 deoxystreptamine; fam. = family; FTM = fortimicin; Glu = glutamine; GM = gentamicin; GT = glycosyltransferase; HM-B = hygromycin B; *icdA* = isocitrate dehydrogenase; IM = istamycin; KM = kanamycin; LM = lividomycin; *Magnetospir* = *Magnetospirillum*; *Mycob.* = *Mycobacterium*; MT = methyltransferase; NM = neomycin; OMe = O-methyltransferase; PM = paromomycin; pr. = protein; poss. = possible; put. = putative; RM = ribostamycin; SAM-d. = S-adenosylmethionin dependant; sim. = similar, sp. = species; syst. = system; THB = tetrahydrobiopterin; TM = tobramycin; try = tryptophane.

Tab. A.2 Proteins encoded in the genomic region covering the *rib*-cluster of *S. ribosidificus* NRRL B-11466 (accession code: AJ744850)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in RM pathway or else [a]
SribP10.1c		135		AAQ20787.1	*S. hygroscopicus*	ketosyntase/methylmalonyl-CoA transferase
SribP10.2c		105		ZP_733531.1	*Pseudomonas fluorescens*	penicillin acylase related protein
SribP10.3		288		NP_828769.1	*S. avermitilis* MA-4680	HP
SribP10.4c		340		AAP20835.1	*Oryza sativa*	HP
SribP10.5		487		NP_942801.1	*Ralstonia eutropha*	HP
SribP10.6c		147		ZP_00081914.1	*Geobacterium metallireducens*	HP
SribP10.7		375		ZP_00058465.1	*Thermobifida fusca*	Streptomycin 6-kinase
SribP10.8c		206		NP_639614.1	*S. coelicolor* A3(2)	transposase (partial ORF)
SribP10.9c		151		NP_862085.1	*S. lividans*	transposase (partial ORF)
SribP10.10c		139		AAF26370.1	*S. coelicolor* A3(2)	HP, probable transposition function
SribP10.11c		80		NP639616.1	*S. coelicolor* A3(2)	HP
SribP10.12c		253		NP862083.1	*S. lividans*	HP
SribP10.13c		149		NP_862091.1	*S. lividans*	helicase-like protein
SribP10.14c	*aacC*	287	AacC	AAA26685.1	*S. fradiae*	aminoglycoside 3-N-acetyltransferase; RM resistance
SribP10.15c	*ribA*	1242	RibA	AAL18480.1	*Photorhabdus luminescens*	unknown; poss. involved in ribosyltransfer
SribP10.16c	*ribB*	416	RibB	AAR98542.1	*M. echinospora*	RM 6'-aminotransferases (AT-III)
SribP10.17	*ribL*	652	RibL	BAC41207.1	*B. circulans*	unknown; poss. involved in ribosyltransfer
SribP10.18c	*ribD*	278	RibD	NP_634124.1	*Methanosarcina mazei* Goe1	TTP-D-glucosmamine synthase
SribP10.19c	*ribF*	352	RibF	NP_630571.1	*S. coelicolor* A3(2)	glycosyltransferase III (hexosamiyltransferase II)
SribP10.20c	*ribX*	82	RibX	AAL18481.1	*Photorhabdus luminescens*	cHP, (5'''-epimerase ?)
SribP10.21c	*ribP*	223	RibP	BAC41220.1	*B. circulans*	sugarphosphate phosphatase (?)
SribP10.22c	*ribN*	299	RibN	NP_578073.1	*Pyrococcus furiosus*	Fe-S oxidoreductase; (5'''-epimerase or 6'''-DH ?)
SribP10.23c	*ribQ*	541	RibQ	AAR98543.1	*M. echinospora*	aminoglycoside 6'- (and 6'''-)dehydrogenase (?)
SribP10.24c	*ribU*	594	RibU	NP_629587.1	*S. coelicolor* A3(2)	ABC transporter, ATP binding; RM(-P) export
SribP10.25c	*ribT*	617	RibT	NP_823966.1	*S. avermitilis* MA-4680	ABC transporter, ATP binding; RM(-P) export
SribP10.26c	*ribM*	419	RibM	AAR98545.1	*M. echinospora*	UDP-(NAc)glucosamine:2DOS 4-glycosyltransferase; GT-I or paromamine synthase
SribP10.27c	*ribC*	391	RibC	CAE22471.1	*S. tenebrarius*	2-deoxy-*scyllo*-inosose synthase (cyclase)
SribP10.28c	*ribS*	424	RibS	CAE22472.1	*S. tenebrarius*	L-glutamine:ketocyclitol aminotransferase (AT-I and -II)

Tab. A.2 Proteins encoded in the genomic region covering the *rib*-cluster of *S. ribosidificus* NRRL B-11466 (accession code: AJ744850), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in RM pathway or else [a]
SribP10.29c	*ribE*	340	RibE	CAE22477.1	*S. tenebrarius*	aminocyclitol 1-dehydrogenase
SribP10.30	*ribI*	166	RibI	CAD60534.1	*S. cinnamoneus*	component of sensor/response regulator syst.
SribP10.31	*ribH*	177	RibH	CAD60535.1	*S. cinnamoneus*	component of sensor/response regulator syst.
SribP10.32	*ribG*	390	RibG	CAD60536.1	*S. cinnamoneus*	component of sensor/response regulator syst.
SribL03.8	*aphA*	468	AphA	AAC32025.1	*S. ribosidificus*	aminoglycoside 3'-phosphotransferase, RM resistance
SribL03.9		81		AAQ82565.1	*S.* sp. FR-008	polyketide synthase (partial ORF); outside *rib*-cluster
SribL03.10c		107		NP_626990.1	*S. coelicolor* A3(2)	transposase (partial ORF)
SribL03.11		154		ZP_00047805.1	*Magnetospir. magnetotacticum*	HP
SribL03.12		103		XP_323407.1	*Neurospora crassa*	HP
SribL03.13		245		NP_733533.1	*S. coelicolor* A3(2)	HP
SribL03.14c		186		AAP92498.1	*S. vinaceus*	type II thioesterase (partial ORF)
SribL03.15c		1089		AAP92497.1	*S. vinaceus*	nonribosomal peptide synthetase
SribL03.16c		739		AAP92496.1	*S. vinaceus*	nonribosomal peptide synthetase (partial ORF)

[a] Abbreviations: see under Tab. A.1

Tab. A.3 Proteins encoded in the genomic region covering the *par*-cluster of *S. rimosus* subsp. *paromomycinus* NRRL 2455 (accession code: AJ749845)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in PM pathway or else [a]
SriG07.16c		79		NP_336806.1	*Mycobact. tuberculosis*	P450 heme-thiolate protein
SriG07.17c		413		AAN85514.1	*S. atroolivaceus*	P450 hydroxlase
SriG07.18c		1110		AAK57184.1	*Stigmatella aurantiaca*	non-ribosomal peptide synthetase
SriD03.11c		158		NP_628587.1	*S. coelicolor* A3(2)	unknown
SriD03.12		260		NP_627740.1	*S. coelicolor* A3(2)	membrane protein
SriD03.13		136		NP_531113.1	*Agrobacterium* S. C58	membrane protein
SriD03.14		211		AAP21653.1	*S. hygroscopicus* subsp. *yingchengensis*	unknown
SriD03.15c		372		ZP_00052846.1	*Magnetosp. magnetotacticum*	amidinotransferase
SriD03.16c		169		NP_625008.1	*S. coelicolor* A3(2)	transcriptional regulator
SriD03.17		80		NP_822222.1	*S. avermitilis* MA-4680	transcriptional regulator
SriD03.18		304		NP_627627.1	*S. coelicolor* A3(2)	unknown
SriD03.19		271		NP_822224.1	*S. avermitilis* MA-4680	unknown
SriD03.20	*parE*	339	ParE	NP_385683.1	*Sinorhizobium meliloti*	aminocyclitol 1-dehydrogenase
SriD03.21	*parS*	424	ParS	AAD45549.1	*S. netropsis*	L-glutamine:ketocyclitol aminotransferase (AT-I and -II)
SriD03.22	*parC*	386	ParC	BAA83344.1	*B. circulans*	2-deoxy-*scyllo*-inosose-synthase (cyclase)
SriD03.23	*parM*	417	ParM	CAC93943.1 BAC41206.1	pIJ6021 *B. circulans*	UDP-(NAc)glucosamine:2DOS 4-glycosyltransferase; GT-I or paromamine synthase
SriD03.24	*parT*	604	ParT	NP_823966.1	*S. avermitilis* MA-4680	ABC transporter, ATP binding; PM(-P) export
SriD03.25	*parU*	628	ParU	NP_823967.1	*S. avermitilis* MA-4680	ABC transporter, ATP binding; PM(-P) export
SriD03.26	*parQ*	546	ParQ	NP_823106.1	*S. avermitilis* MA-4680	PM 6'''-dehydrogenase (?)
SriD03.27	*parN*	298	ParN	AAD32720.1	*S. lavendulae*	Fe-S oxidoreductase; (5'''-epimerase or 6'''-DH ?)
SriD03.28	*parP*	231	ParP	BAC41220.1	*B. circulans*	sugarphosphate phosphatase (?)
SriD03.29	*parX*	90	ParX	AAL 18481.1	*Photorhabdus luminescens*	cHP, (5'''-epimerase ?)
SriD03.30	*parF*	367	ParF	NP_826550.1	*S. avermitilis* MA-4680	put. glycosyltransferase III (hexosaminyl GT-II)
SriD03.31	*parD*	253	ParD	BAC41211.1	*B. circulans*	TTP-D-glucosmamine synthase
SriD03.32	*parL*	637	ParL	BAC41207.1	*B. circulans*	unknown; poss. involved in ribosyltransfer
SriD03.33	*parB*	417	ParB	BAC41209.1	*B. circulans*	PM 6'''-aminotransferase
SriD03.34	*parA*	1302	ParA	BAC41208.1	*B. circulans*	unknown; poss. involved in ribosyltransfer

Tab. A.3 Proteins encoded in the genomic region covering the *par*-cluster of *S. rimosus* subsp. *paromomycinus* NRRL 2455 (accession code: AJ749845), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in PM pathway or else [a]
SriL03.20c	*parZ*	455	ParZ	AAP92510.1	*S. vinaceus*	extracellular PM-phosphate phosphatase
SriL03.19	*parI*	163	ParI	CAD60534.1	*S. cinnamoneus*	component of sensor/response regulator system
SriL03.18	*aphA*	262	AphA	AAA26699.1	*S. fradiae*	paromomycin phosphotransferase; Aph(3')-V
SriL03.17c	*parY*	229	ParY	ZP_00004610.1	*Rhodobacter sphaeroides*	cHP; unknown
SriL03.16c	*parG*	400	ParG	ZP_00110163.1	*S. cinnamoneus*	component of sensor/response regulator system
SriL03.15c	*parH*	184	ParH	NP_828190.1	*S. cinnamoneus*	component of sensor/response regulator system
SriL03.14c	(*pasA*)	76	(PasA)	AAF10386.1	*Deinococcus radiodurans*	UDP-N-acetylglucosamine synthase; outside *par*-cluster
SriL03.13	(*pasB*)	775	(PasB)	NP_630071.1	*S. coelicolor* A3(2)	chitinase; outside *par*-cluster
SriL03.12	(*pasC*)	284	(PasC)	ZP_00227284.1	*Kineococcus radiotolerans*	NDP-sugar epimerase; outside *par*-cluster
SriL03.11	(*parR1*)	209	(ParR1)	ZP_00226580.1	*Kineococcus radiotolerans*	transcriptional regulator; outside *par*-cluster
SriL03.10	(*parR2*)	292	(ParR2)	NP_824491.1	*S. avermitilis* MA-4680	transcriptional regulator; outside *par*-cluster
SriL03.9	(*pasD*)	285	(PasD)	ZP_00201781.1	*Methylobacterium flagellatus*	oxidoreductase, DH; outside *par*-cluster
SriL03.8		40		NP_631315.1	*S. coelicolor* A3(2)	conserved hypothetical protein

[a] Abbreviations: see under Tab. A.1

Tab. A.4 Proteins encoded in the genomic regions covering the *aacC7* of *S. rimosus* subsp. *paromomycinus* NRRL 2455 (accession codes: AJ628955)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in PM pathway or else [a]
SriA13.1		246		A32245	*S. rimosus* subsp. *paromomycinus*	integral membrane protein
SriA13.2	*aacC7*	287	AacC7	P30180	*S. rimosus* subsp. *paromomycinus*	3-N-acetyltransferase; PM resistance
SriA13.3	*aacA*	190	AacA	AAA25688.1	*Pseudomonas aeruginosa*	6'-N-acetyltransferase; ACAGA resistance(?)
SriA13.4c		333		NP_821315.1	*S. avermitilis* MA-4680	epoxide hydrolase
SriA13.5		180		NP_821317.1	*S. avermitilis* MA-4680	transcriptional regulator (TetR-family)
SriA13.6		76		NP_624841.1	*S. coelicolor* A3(2)	cold-shock protein
SriA13.7		94		NP_901291.1	*Chromobacterium violaceum*	sensor/response regulatory hybrid protein

[a] Abbreviations: see under Tab. A.1

Tab. A.5 Proteins encoded in the genomic region covering the *liv*-cluster of *S. lividus* ATCC 31603 (accession code: AJ748832)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in LM pathway or else [a]
SliD01.1		687		EAA46978.1	*Magnaporthe grisea* 70-15	predicted protein
SliD01.2c		131		NP_626993.1	*S. coelicolor* A3(2)	put. secreted protein
SliD01.3		329		NP_631450.1	*S. coelicolor* A3(2)	ABC transporter, ATP binding; iron (III) dicitrate transport system
SliD01.4		325		NP_631449.1	*S. coelicolor* A3(2)	transport lipoprotein; iron (III) dicitrate transport system
SliD01.5		692		NP_631448.1	*S. coelicolor* A3(2)	put. membrane transport protein
SliD01.6	*livE*	339	LivE	CAE22477.1	*S. tenebrarius*	aminocyclitol 1-dehydrogenase
SliD01.7	*livS*	424	LivS	CAE22472.1	*S. tenebrarius*	L-glutamine:ketocyclitol aminotransferase (AT-I and -II)
SliD01.8	*livC*	384	LivC	CAE22471.1	*S. tenebrarius*	2-deoxy-*scyllo*-inosose synthase (cyclase)
SliD01.9	*livM*	414	LivM	BAD20768.1	*S. kanamyceticus*	UDP-(NAc)glucosamine:2DOS 4-glycosyltransferase; GT-I or paromamine synthase
SliD01.10	*livT*	601	LivT	NP_823966.1	*S. avermitilis* MA-4680	ABC transporter, ATP binding; LM(-P) export
SliD01.11	*livU*	621	LivU	NP_823967.1	*S. avermitilis* MA-4680	ABC transporter, ATP binding; LM(-P) export
SliD01.12	*livQ*	546	LivQ	AAR98543.1	*M. echinospora*	aminoglycoside 6' (and 6"'-)dehydrogenase (?)
SliD01.13	*livN*	299	LivN	ZP_00054205.1	*Magnetospir. magnetotacticum*	Fe-S oxidoreductase; (5"'-epimerase or 6"'-DH ?)
SliD01.14	*livP*	229	LivP	BAC41220.1	*B. circulans*	cHP; sugarphosphate phosphatase (?)
SliD01.15	*livF*	357	LivF	NP_630571.1	*Nostoc* sp. PCC 7120	glycosyltransferase III (hexosaminyl GT-II)
SliD01.16	*livD*	252	LivD	BAD20763.1	*S. kanamyceticus*	TTP-D-glucosmamine synthase
SliD01.17c	*livL*	617	LivL	BAC41207.1	*B. circulans*	unknown; poss. involved in ribosyltransfer (?)
SliD01.18	*livB*	416	LivB	AAR98542.1	*M. echinospora*	LM 6"'-aminotransferase
SliD01.19c	*livZ*	456	LivZ	AAP92510.1	*S. vinaceus*	extracellular LM-phosphate phosphatase (?)
SliD01.20	*livI*	166	LivI	CAD60534.1	*S. cinnamoneus*	component of sensor/response regulator syst. (?)
SliD01.21	*livH*	170	LivH	NP_828190.1	*S. cinnamoneus*	component of sensor/response regulator syst. (?)
SliD01.22	*livG*	407	LivG	ZP_00110163.1	*S. cinnamoneus*	component of sensor/response regulator syst. (?)
SliD01.23c	*livA*	1355	LivA	AAL18480.1	*Photorhabdus luminescens*	unknown; poss. involved in ribosyltransfer (?)
SliD01.24c	*livX*	80	LivX	AAL18481.1	*Photorhabdus luminescens*	cHP, (5"'-epimerase)
SliD01.25c	(*livO*)	62	(LivO)	ZP_00087677.1	*Pseudomonas fluorescens*	fragmentary ORF, topoisomerase IA
SliD01.26	*livV*	430	LivV	NP_960853.1	*Mycob. avium*	cHP
SliD01.27	*livW*	458	LivW	NP_614275.1	*Methanopyrus kandleri*	Fe-S oxidorededucatse; 3'-dehydroxylation (?)

Tab. A.5 Proteins encoded in the genomic region covering the *liv*-cluster of *S. lividus* ATCC 31603 (accession code: AJ748832), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in LM pathway or else [a]
SliD01.28	*livY*	238	LivY	AAR98544.1	*M. echinospora*	oxidoreductase; 3'-dehydroxylation (?)
SliD01.29c	*(liwA)*	407	(LiwA)	NP_738548.1	*Corynebacterium efficiens*	coenzyme PQQ synthesis protein; outside *liv*-cluster ?
SliD01.30		152		NP_624549.1	*S. coelicolor* A3(2)	nitrate reductase alpha chain

[a] Abbreviations: see under Tab. A.1

Tab. A.6 Proteins encoded in the genomic region covering part of the *btr*-cluster of *B. circulans* ATCC 21558 (accession code: AJ781030)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in BU pathway or else [a]
BciH11.1	*btrM*	(389)	BtrM	BAC41206.1	*B. circulans*	hexosaminyltransferase (GT-I); paromamine synthase
BciH11.2	*btrL*	604	BtrL	BAC41207.1	*B. circulans*	unknown; poss. involved in ribosylation(?)
BciH11.3	*btrA*	1225	BtrA	BAC41208.1	*B. circulans*	unknown; poss. involved in ribosylation(?)
BciH11.4	*btrB*	432	BtrB	BAC41209.1	*B. circulans*	hexosaminyl 6'-AT
BciH11.5	*btrC*	368	BtrC	BAC41210.1	*B. circulans*	2-deoxy-*scyllo*-inosose synthase (cyclase)
BciH11.6	*btrD*	275	BtrD	BAC41211.1	*B. circulans*	TTP-D-glucosmamine synthase
BciH11.7	*btrE*	349	BtrE	BAC41212.1	*B. circulans*	oxidoreductase; aminocyclitol 1-DH
BciH11.8	*btrF*	232	BtrF	BAC41213.1	*B. circulans*	oxidoreductase; hexose-DH?
BciH11.9	*btrG*	156	BtrG	BAC41214.1	*B. circulans*	HP, unknown
BciH11.10	*btrH*	302	BtrH	BAC41215.1	*B. circulans*	HP, unknown
BciH11.11	*btrI*	87	BtrI	BAC41216.1	*B. circulans*	acyl carrier protein; AHBA synthesis/transfer
BciH11.12	*btrJ*	419	BtrJ	BAC41217.1	*B. circulans*	carboxylase; AHBA synthesis?
BciH11.13	*btrK*	428	BtrK	BAC41218.1	*B. circulans*	Glu-decarboxylase; AHBA synthesis
BciH11.14	*btrO*	341	BtrO	BAC41219.1	*B. circulans*	FMN monoxygenase, AHBA synthesis
BciH11.15	*btrP*	213	BtrP	BAC41220.1	*B. circulans*	shorter than BAC41220.1; phosphatase or mutase
BciH11.16	*btrQ*	504	BtrQ	AAK89143.1	*Agrobacterium tumefaciens*	oxidoreductase; hexosaminyl 6'-DH

[a] Abbreviations: see under Tab. A.1

Tab. A.7 Proteins encoded in the genomic region covering the *kan*-cluster of *S. kanamyceticus* DSM 40500 (accession code: AJ628422)

ORF design.	Gene symbol	Gene product aa	desig.	Similarity to other proteins in the data base acc. code	organism	Put. function in KM pathway or else [a]
SkaJ19.1		516		NP_826945.1	*S. avermitilis*	ABC transporter
SkaJ19.2		413		NP_626652.1	*S. colicolor* A3(2)	ABC transporter permease
SkaJ19.3c		406		BAA32132.1	*S. griseus*	methyltransferase
SkaJ19.4		278		NP_624822.1	*S. colicolor* A3(2)	NAD(+) synthase
SkaJ19.5c		146		NP_522161.1	*Ralstonia solanacearum*	unknown, cHP
SkaJ19.6c		248		NP_823077.1	*S. avermitilis*	unknown, cHP
SkaJ19.7c	(*kaoA*)	832	KaoA	NP_823078.1	*S. avermitilis*	cHP, poss. regulator, (out of cl.?)
SkaJ19.8c	(*kanZ*)	811	KanZ	NP_630566.1	*S. colicolor* A3(2)	efflux protein, (out of cl.?)
SkaJ19.9c	(*kanY*)	147	KanY	NP_630565.1	*S. colicolor* A3(2)	unknown, HP, (out of cl.?)
SkaJ19.10c	(*kanX*)	420	KanX	NP_828653.1	*S. avermitilis*	efflux protein, (out of cl.?)
SkaJ19.11c	(*kanW*)	64	KanW	AAO65787.1	*S. cinnamonensis*	unknown, HP, (out of cl.?)
SkaJ19.12	*kanA*	184	KanA	NP_105746.1	*Mesorhizobium loti*	resistance, N-6'-acetyltransferase, AAC(6')
SkaJ19.13c	*kanG*	382	KanG	CAD60536.1	*S. cinnamonensis*	unknown, poss. regulator system
SkaJ19.14c	*kanH*	195	KanH	CAD60535.1	*S. cinnamonensis*	unknown, poss. regulator system
SkaJ19.15c	*kanI*	160	KanI	CAD60534.1	*S. cinnamonensis*	unknown, poss. regulator system
SkaJ19.16	*kanR*	369	KanR	CAA55579.1	*S. glaucescens*	KM-specific pathway regulator
SkaJ19.17	*kanE*	343	KanE	NP_626648.1	*S. colicolor* A3(2)	cyclitol 1-dehydrogenase
SkaJ19.18	*kanT*	418	KanT	NP_625738.1	*S. colicolor* A3(2)	export protein.
SkaJ19.19	*kanB*	392	KanB	BAC41209.1	*B. circulans*	hexosamine-6'-AT
SkaJ19.20	*kanQ*	508	KanQ	NP_356358.1	*Agrobactacterium tumefaciens*	hexosamine-6'-DH
SkaJ19.21	*kanS1*	427	KanS1	CAA70012.1	*S. griseus*	cyclitol AT, AT I
SkaJ19.22	*kanC*	390	KanC	BAA83344.1	*B. circulans* BtrC	cyclitol cyclase,
SkaJ19.23	*kanD2*	369	KanD2	NP_698342.1	*Brucella suis*	(cyclitol-1-) or hexose-3"-DH
SkaJ19.24	*kanM2*	388	KanM2	BAC41206.1	*B. circulans*	6-glucosyl-GT (GT II)?
SkaJ19.25	*kanD1*	249	KanD	NP_634124.1	*S. colicolor* A3(2)	TTP-D-glucosmamine synthase
SkaJ19.26	*kanS2*	419	KanS2	CAA70012.1	*S. griseus*	cyclitol or hexose-3"-aminotransferase; AT II
SkaJ19.27	*kanJ*	322	KanJ	AAD32731	*S. lavendulae*	poss. to be involved the 2' deamination process
SkaJ19.28	*kanK*	329	KanK	NP_809643.1	*Bacteroides thetaiotaomicron*	poss. to be involved 2' deamination process
SkaJ19.29	*kmr*	277	Kmr	CAA75800.1	*S. kanamyceticus*	16S rRNA methylase; KM resistance

Tab. A.7 Proteins encoded in the genomic region covering the *kan*-cluster of *S. kanamyceticus* DSM 40500 (accession code: AJ628422), continued

ORF design.	Gene symbol	Gene product aa	desig.	Similarity to other proteins in the data base acc. code	organism	Put. function in KM pathway or else [a]
SkaJ19.30	*kanM1*	414	KanM1	BAC41206.1	*B. circulans*	4-hexosaminyl-GT (GT I)?
SkaJ19.31	(*kanL*)	335	KanL	NP_217745.1	*Mycob. tuberculosis*	unknown (GTPase?), (out of cl.?)
SkaJ19.32	(*kanO*)	132	KanO	NP_630561.1	*S. colicolor* A3(2)	unknown, cHP, (out of cl.?)
SkaJ19.33c	(*kanP*)	223	KanP	NP_823084.1	*S. avermitilis*	unknown, HP, (out of cl.?)
SkaJ15.16	(*kanD3*)	593	KanD3	NP_823085.1	*S. avermitilis*	FadC1, dehydrogenase, (out of cl.?)
SkaJ15.17	(*kanU*)	316	KanU	AAK19882.1	*S. cinnamonensis*	transcript. repressor (TetR fam.), (out of cl.?)

[a] Abbreviations: see under Tab. A.1

Tab. A.8 Proteins encoded in the genomic region covering the *tob*-cluster of *S. tenebrarius* DSM 40477 (accession code: AJ810851)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in TM pathway or else [a]
SteM07.32		185		ZP_00292780.1	*Thermobifida fusca*	acyl-CoA synthetase (NDP forming)
SteK17.1c		336		NP_625346.1	*S. coelicolor* A3(2)	put. integral membrane protein
SteK17.2		302		NP_960675.1	*Mycob. avium*	hypothetical protein MAP1741c
SteK17.3c		90		NP_926603.1	*Gloeobacter violaceus*	transcriptional regulatory pr. (MerR family; fragmentary)
SteK17.4		986		ZP_00294447.1	*Thermobifida fusca*	predicted helicase
SteK17.5		574		ZP_00294448.1	*Thermobifida fusca*	HP, unknown
SteK17.6		200		ZP_00294449.1	*Thermobifida fusca*	HP, unknown
SteK17.7		384		ZP_00294450.1	*Thermobifida fusca*	HP, unknown
SteK17.8		281		NP_828715.1	*S. avermitilis* MA-4680	HP, unknown
SteK17.9		269		ZP_00294452.1	*Thermobifida fusca*	transcriptional regulator
SteK17.10		366		NP_952445.1	*Geobacter sulfurreducens*	CRISPR-associated protein Cas1
SteK17.11		88		ZP_00268340.1	*Rhodospirillum rubrum*	HP, unknown
SteK17.12c		56		XP_500831.1	*Yarrowia lipolytica*	HP, unknown (fragmentary)
SteK17.13		147		XP_504139.1	*Yarrowia lipolytica*	HP, unknown
				ZP_00294864.1	*Methanosarcina barkeri*	transposase and inactivated derivatives
SteK17.14		119		NP_627825.1	*S. coelicolor* A3(2)	HP, unknown
SteK17.15		361		ZP_00226505.1	*Kineococcus radiotolerans*	cHP, unknown
				NP_285525.1	*Deinococcus radiodurans* R1	superoxide dismutase (Cu-Zn family)
SteK17.16	(*tobY*)	165	(TobY)	ZP_00296600.1	*Methanosarcina barkeri*	GTPases (G3E family)
SteK17.17	(*tobX*)	217	(TobX)	NP_436925.1	*Sinorhizobium meliloti* 1021	HP, unknown
SteK17.18	*tobE*	339	TobE	CAE22477.1	*S. tenebrarius*	dehydrogenase
SteK17.19	*tobT*	436	TobT	CAE22476.1	*S. tenebrarius*	transport protein
SteK17.20	*tobB*	395	TobB	CAE22475.1	*S. tenebrarius*	aminoglycoside 6'-aminotransferase
SteK17.21	*tobQ*	508	TobQ	CAE22474.1	*S. tenebrarius*	aminoglycoside 6'-dehydrogenase
SteK17.22	*tobZ*	579	TobZ	CAE22473.1	*S. tenebrarius*	6"-carbamoyltransferase
SteK17.23	*tobS1*	424	TobS1	CAE22472.1	*S. tenebrarius*	2-deoxy-*scyllo*-inosose aminotransferase and ketoaminocyclitol aminotransferase II
SteK17.24	*tobC*	386	TobC	CAE22471.1	*S. tenebrarius*	2-deoxy-*scyllo*-inosose synthase
SteK17.25	*tobD2*	347	TobD2	CAE22470.1	*S. tenebrarius*	hexose-3"-dehydrogenase

Tab. A.8 Proteins encoded in the genomic region covering the *tob*-cluster of *S. tenebrarius* DSM 40477 (accession code: AJ810851), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in TM pathway or else [a]
SteK17.26	*tobM2*	420	TobM2	CAE22469.1	*S. tenebrarius*	6-glucosyltransferase
SteK17.27	*tobD1*	261	TobD1	BAD20763.1|	*S. kanamyceticus*	TTP-D-glucosmamine synthase
SteK17.28	*tobS2*	416	TobS2	BAD20764.1	*S. kanamyceticus*	L-glutamine:hexose-3"-aminotransferase
SteK17.29		123		XP_470790.1	*Oryza sativa*	retrotransposon protein (fragmentary)
SteK17.30	*tobM1*	416	TobM1	BAD20768.1	*S. kanamyceticus*	aminoglycoside 4-glucosaminyltransferase
SteK17.31c	(*tobL*)	410	(TobL)	NP_782304.1	*Clostridium tetani* E88	carbamoyl-phosphate synthase large chain
SteK17.32c	(*tobU*)	325	(TobU)	ZP_00294424.1	*Thermobifida fusca*	transporter (DMT-superfamily)
SteK17.33c	(*tobA*)	372	(TobA)	NP_577850.1	*S. coelicolor* A3(2)	aminotransferase
SteK17.34c	(*tobD3*)	358	(TobD3)	ZP_00187505.2	*S. coelicolor* A3(2)	dehydrogenase (FMN-dependent)
SteK17.35c	(*tobO*)	327	(TobO)	NP_626927.1	*S. coelicolor* A3(2)	oxygenase (put. secreted protein)
SteK17.36	(*tobR*)	163	(TobR)	NP_108143.1	*Mesorhizobium loti*	transcriptional regulator (AsnC family)

[a] Abbreviations: see under Tab. A.1

Tab. A.9 Proteins encoded in the genomic region covering the *gen*-cluster of *M. echinospora* DSM 43036 (accession code: AJ628149)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in GM pathway or else [a]
MecP21.15		357		NP_828616.1	*S. avermitilis*	DNA-polymerase beta chain
MecP21.16c		324		NP_733531.1	*S. coelicolor* A3(2)	integral membrane protein
MecP21.17		157		NP_628738.1	*S. coelicolor* A3(2)	HP
MecP21.18		556		NP_938521.1	*Corynebacterium diphtheriae*	cHP
MecP21.19c		339		AAS07044.1	*Chlamydomonas reinhardtii*	sim. to agglutinin
MecP21.20c		192		ZP_00058638.1	*Thermobifida fusca*	HP
MecP21.21c		200		NP_629559.1	*S. coelicolor* A3(2)	cholesterol esterase
MecP21.22		568		NP_602043.1	*Corynebacterium glutamicum*	vancomycin resistance, secreted protein
MecP21.23c		208		NP_629602.1	*S. coelicolor* A3(2)	cHP
MecE04.1		228		ZP_00034543.1	*Burkholderia fungorum*	GntT, probably not in GM pathway; COG2232: predicted ATP-dependent carboligase;
MecE04.2		359		NP_219070.1	*Treponema pallidum*	GntU; probably not in GM pathway; Trp-tRNA ligase
MecE04.3	*genO*	385	GenO	NP_622810.1	*Thermoanaerob. tengcong.*	GntV; ForO; transferase in GM biosynthesis (?)
MecE04.4	*gmrB*	272	GmrB	A45282	*M. zionensis*	GmrO; FmrB,O,R; 16S rRNA MT; GM resistance
MecE04.5c	*genB1*	417	GenB1	BAC41209.1	*B. circulans*	GntW; GM (6-hexosaminyl-?) AT; GM AT I
MecE04.6c	*genQ*	557	GenQ	ZP_00032020.1	*Burkholderia fungorum*	GntX; oxidoreductase (flavoprotein): GM 6'-DH
MecE04.7c	*genD3*	269	GenD3	NP_822086.1	*S. avermitilis*	GntY; oxidoreductase, put. 3',4'dehydratase
MecE04.8c	*genM1*	415	GenM1	CAC93943.1	cloning vector pIJ6021	GntZ; GT I, hexosaminyl-GT, paromamine synthase
MecE04.9c	*gmrA*	274	GmrA	P24618	*M. purpurea*	GmrA; GM resistance¸ 16S rRNA MT
MecE04.10	*genS1*	420	GenS1	CAA70012.1	*S. griseus* DSM 40236	GntA; L-glutamine:ketocyclitol AT I (and II ?)
MecE04.11	*genC*	397	GenC	BAA83344.1	*B. circulans*	GntB; 2-deoxy-*scyllo*-inosose synthase (cyclase)
MecE04.12	*genD2*	341	GenD2	NP_541843.1	*Brucella melitensis*	GntC; cyclitol 1-DH or pentose-3"-DH (?)
MecE04.13	*genM2*	390	GenM2	BAC41206.1	*B. circulans*	GntD; GT II, xylosyl-GT (GM-A synthase)
MecE04.14	*genD1*	659	GenD1	NP_267324.1	*Lactoccocus* sp.	GntE; HP; GM (Fe-S-) oxidoreductase (COG1032)
MecE04.15	*genS2*	418	GenS2	CAA70012.1	*S. griseus* DSM 40236	GntF; L-glutamine:GM 3"-AT (?)
MecE04.16	*genW*	145	GenW	NP_842285.1	*Nitrosomonas europaea*	GntG; FosC; unknown; GTP cyclohydrolase fam.
MecE04.17c	*genB4*	445	GenB4	ZP_00112608.1	*Prochlorococcus marinus*	GntH; ForB; hexosamine-6'- or pentose-3"-AT; GM AT IV
MecE04.18c	*genP*	268	GenP	P13250	*S. ribosidificus*	GntI; ForP; APH(3'), hexosamine 3'-kinase
MecE04.19c	*genB3*	490	GenB3	ZP_00057034.1	*Thermobifida fusca*	GntJ; ForB; GM-6'- or -3"-AT; GM AT III

Tab. A.9 Proteins encoded in the genomic region covering the *gen*-cluster of *M. echinospora* DSM 43036 (accession code: AJ628149), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in GM pathway or else [a]
MecE04.20c	*genK*	638	GenK	BAA08420.1	*M. olivasterospora*	GntK, ForK/Fms14; put. 6'-*C*-MT (and 4"-*C*-MT ?)
MecE04.21	*genB2*	414	GenB2	BAB18041.1	*B. circulans*	GntL; hexosamine-6'- or pentose-3"-AT; GM AT II
MecE04.22c	*genX*	170	GenX	NP_828190.1	*S. avermitilis* MA-4680	GntM; HP, unknown
MecE04.23c	*genU*	311	GenU	ZP_00110163.1	*Nostoc punctiforme*	GntN; WD-repeat protein, regulator
MecG05.15c	*genT*	468	GenT	AY228175.1	*S. murayam*	GntO; put. efflux protein
MecG05.16c	*genE*	336	GenE	NP_252842.1	*Pseudomonas aeruginosa* PAO1	GntP; cyclitol 1-DH
MecG05.17	*genY*	504	GenY	M80527.1	*S. toyocaensis*	GntQ; FosD; antiporter/exporter (?)
MecG05.18	*genA*	245	GenA	ZP_00015638.1	*Rhodospirillum rubrum*	GntR; FosE; PP-loop ATPase, unknown
MecG05.19	*genF*	212	GenF	NP_881273.1	*Bordetella pertussis*	GntS; FosF; unknown, HP
MecG05.20	*genG*	117	GenG	NP_251356.1	*Pseudomonas aeruginosa* PAO1	FosG; put. phospholyase (6-pyruvoyl-THB synthesis)
MecG05.21	*genH*	1138	GenH	NP_823189.1	*S. avermitilis*	ForH+ForI; efflux protein complex (?)
MecG05.22	*genJ*	645	GenJ	(MecG05.21)	*M. echinospora*	ForJ; efflux protein complex (?)
MecG05.23	*genI*	312	GenI	Q08329	*M. olivasterospora*	ForT; GM production pr.; unknown
MecG05.24	*genL*	206	GenL	CAD91216.1	*Nonomuraea* sp.	TTP-D-glucosmamine synthase
MecG05.25	*genN*	321	GenN	NP_579402.1	*Pyrococcus furiosus*	GM (*N*-)MT(?)
MecG05.26c		167				HP; unknown
MecG05.28c		438		NP_624865.1	*S. coelicolor* A3(2)	two-components syst. histidine kinase
MecO02.15c		231		NP_826427.1	*S. avermitilis* MA-4680	two- components syst. response regulator
MecO02.16		161		NP_636672.1	*Xanthomonas campstris*	secreted serine protease; fragmentary ORF
MecO02.17c		386		ZP_00293467.1	*Thermobifida fusca*	DNA polymerase III, epsilon subunit
MecO02.18c		217		ZP_00267533.1	*Rhodospirillum rubrum*	HP; glycosyltransferase
MecO02.19c		345		EAA38054.1	*Giardia lamblia* ATCC 50803	HP; unknown
MecO02.20c		155		NP_631062.1	*S. coelicolor* A3(2)	HP
MecO02.21c		218		NP_631061.1	*S. coelicolor* A3(2)	RNA polymerase sigma factor
MecO02.22c		430		NP_285664.1	*Deinococcus radiodurans R1*	serine protease/subtilase family
MecO02.23		882		NP_631059.1	*S. coelicolor* A3(2)	HP
MecO02.24c		167		NP_823881.1	*S. avermitilis*MA-4680	HP
MecO02.25		452		YP_121867.1	*Nocardia farcinica*	HP
MecO02.26		126		ZP_00298424.1	*Geobacterium metallireducens*	glycosyltransferase/PMT family

Tab. A.9 Proteins encoded in the genomic region covering the *gen*-cluster of *M. echinospora* DSM 43036 (accession code: AJ628149), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in GM pathway or else [a]
MecO02.27c		97		NP_636569.1	*Xanthomonas campestris*	HP
MecO02.28c		347		NP_693914.1	*OceanoB iheyensis*	HP
MecO02.29		375		CAC93708.1	*Lechevalieria aerocolonigenes*	put. regulatory protein,
MecO02.30c		60		NP_241565.1	*B. halodurans*	ribosomal N-acetyltransferase
MecO02.31		761		NP_242912.1	*B. halodurans*	HP
MecO02.32		779		NP_242908.1	*B. halodurans*	HP
MecO02.33		323		NP_148662.1	*Aeropyrum pernix* K1	ATP-binding protein
MecO02.34		202		BAA95020.1	*S. kasugaensis*	membrane protein

[a] GntA-Z = the alternative nomenclature of identical gene products in the *gnt*-cluster of *M. echinospora* ATCC 15835 (accession code AY524043); all other abbreviations are presented under Tab. A.1

Tab. A.10 Proteins encoded in the genomic region covering the *for*-cluster of *M. olivasterospora* DSM 43868 (accession code: AJ628421)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in FTM pathway or else [a]
MolI14.1c		470		NP_250185.1	*Pseudomonas aeruginosa*	cHP
MolI14.2c		306		NP_769118.1	*Bradyrhizobium japonicum*	Transposase
MolI14.3	*forY*	440	ForY	AAK81829.1	*S. lavendulae*	cation antiporter, (outside cl.?); sim. to MecG05.17
MolI14.4	*forH*	668	ForH	EAA49750.1	*Magnaporthe grisea*	efflux transporter system, FTM-efflux (?)
MolI14.5	*forI*	503	ForI	NP_823189.1	*S. avermitilis*	efflux transporter system, FTM-efflux (?)
MolI14.6	*forJ*	620	ForJ	EAA49750.1	*Magnaporthe grisea*	efflux transporter system, FTM-efflux (?)
MolI14.7	*forQ*	559	ForQ	ZP_00110912.1	*Nostoc punctiforme*	FTM-AO 6'-DH, flavoprotein
MolI14.8c	*forW*	287	ForW	ZP_00086974.1	*Pseudom. fluorescens* PfO-1	cytochrome in oxidoreductase complex?
MolI14.9c	*forV*	516	ForV	P0047E11.11	*Oryza sativa*	Fms13(?); glycyltransferase (?) PQQ biosynthesis; poss. Fe-S pr.
MolI14.10c	*forM*	427	ForM	BAC41206.1	*B. circulans*	Fms3(?); UDP-N-acetylglucosamine: 3-amino-3-deoxy-*scyllo*-inositol 4-GT, FTM-FU-10 synthase
MolI14.11c	*forS*	423	ForS	AAD28492.1	*S. spectabilis*	Fms4(?); L-glutamine: ketocyclitol 3- and 6-AT
MolI14.12c	*forE*	353	ForE	NP_626648.1	*S. coelicolor* A3(2)	Fms5(?); cyclitol (3-) or 6-DH
MolI14.13c	*forN*	244	ForN	AAD41819.1	*S. fradiae*	Fms10(?); FTM-AP 1-O-MT
MolI14.14c	*forO*	379	ForO	AAK76881.1	*Clostridium acetobutylicum*	Fms12(?); unknown; cHP (AstB/NirJ/MoaA-related pr.) or 3',4'-reductase(?) (Fe-S oxidoreducatse) PQQ biosynthesis; heme biosynthesis
MolI14.15c	*forD*	302	ForD	ZP_00051545.1	*Magnetospir. magnetotacticum*	TTP-D-glucosmamine synthase
MolI14.16c	*forP*	274	ForP	AAB21326.1	*M. chalcea*	Fms8; GenF; FTM-KK1 3'-phosphotransferase
MolI14.17c	*forB*	442	ForB	NP_893936.1	*Prochlorococcus marinus*	Fms6(?); GenB3,B4; hexosamine 6'-AT
MolI14.18c	*forK*	553	ForK	BAA08420.1	*M. olivasterospora*	Fms7; GenK; FTM-KL1 6'-C-MT
MolI14.19c	*forZ*	482	ForZ	BAA00940.1	*M. olivasterospora*	Fms14; glycyl-N-formimidoyltransferase, FAD-dependent FI-FTM-A synthase
MolI14.20	*forX*	155	ForX	BAC52860.1	*Bradyrhizobium japonicum*	unknown; cHP
MolI14.21	*forC*	353	ForC	BAC69696.1	*S. avermitilis*	cyclitol 3- (or 6-)DH
MolI14.22	*forG*	280	ForG	ZP_00074622.1	*Trichodesmium* sp.	SAM-d. MT, 6-*N*-MT
MolI14.23	*forA*	281	ForA	ZP_00045395.1	*Magnetococcus* sp.	Fms1(?); D-*myo*-inositol-3-phosphate phosphatase
MolI14.24c	*forL*	262	ForL	NP_929964.1	*Photorhabdus luminescens*	Fms11(?); 1-OMe-epimerase; Fe-S oxidoreductase

Tab. A.10 Proteins encoded in the genomic region covering the *for*-cluster of *M. olivasterospora* DSM 43868 (accession code: AJ628421), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in FTM pathway or else [a]
MolI14.25	*fmrR*	99	FmrR	CAF34044.1	*M. echinospora*	GM/FTM resistance, 16S rRNA MT
MolI14.26	*fmrO*	293	FmrO	Q08325.1	*M. olivasterospora*	FTM-resistance, 16S rRNA MT
MolI14.27	*forT*	313	ForT	Q08329.1	*M. olivasterospora*	unknown
MolI14.28c	*fosG*	118	FosG	ZP_00087522.1	*Pseudomonas fluorescens*	unknown; 6-pyruvoyl THB synthase family
MolI14.29c	*fosF*	212	FosF	NP_421953.1	*Caulobacterium crescentus*	unknown
MolI14.30c	*fosE*	233	FosE	ZP_00015638.1	*Rhodospirillum rubrum*	unknown; ATP/GTP-bp. of PP-loop superfamily
MolI14.31c	*fosD*	137	FosD	AAF78928	*Crypthecodinium cohnii*	unknown
MolI14.32c	*fosC*	123	FosC	NP_842285.1	*Nitrosomonas europaea*	unknown; GTP cyclohydrolase family
PORF1		40		CAE06507.1	*M. echinospora*	AT-II-fragment
MolI14.33	*fmrP*	158	FmrP	P24619	*M. echinospora*	GM/FTM resistance, 16S rRNA MT
PORF2		103		CAF34042.1	*M. echinospora*	oxidoreductase
MolI14.34c	*fmrB*	269	FmrB	CAF34044.1	*M. echinospora*	GM/FTM resistance, 16S rRNA MT
MolI14.35c	*fosA*	385	FosA	NP_622810.1	*Thermoanaerobacterium* sp.	FTM biosynthetic transferase (?)
MolI14.36		376		NP_631515.1	*S. coelicolor* A3(2)	HP, unknown
MolI14.37		471		NP_147838.1	*Aeropyrum pernix*	tryptophanase
MolJ05.1c		195		ZP_00058483.1	*Thermobifida fusca*	GTPase
MolJ05.2c		121		ZP_00056729.1	*Thermobifida fusca*	HP, unknown
MolJ05.3c		143		NP_822753.1	*S. avermitilis*	cHP, unknown
MolJ05.4c		493		ZP_00058480.1	*Thermobifida fusca*	signal transduction histidine kinase
MolJ05.5c		343		NP_628328.1	*S. coelicolor* A3(2)	acetyltransferase

[a] Fms1-14 = are the alternative nomenclature of identical gene products in the *for*-cluster of *M. olivasterospora* previously reported by Dairi *et al*. (1992a); all other abbreviations are presented under Tab. A.1

Tab. A.11 Proteins encoded in the genomic region covering the the *apr*-cluster of *S. tenebrarius* DSM 40477 (accession code: AJ629123)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in Apr pathway or else
SteO08.1	*aprX*	205	AprX	NP_733638.1	*S. coelicolor* A3(2)	lipoprotein
SteO08.2c	*aprR2*	244	AprR2	NP_923854.1	*Gloeobacter violaceus*	transcriptional regulator (AraC family)
SteO08.3	*aprR1*	370	AprR1	NP_824342.1	*S. avermitilis* MA-4680	transcriptional regulator (AraC family)
SteO08.4c	*aprN*	244	AprN	NP_634124.1	*Methanosarcina mazei*	TTP-D-glucosmamine synthase
SteO08.5c	*aprW*	657	AprW	NP_823966.1	*S. coelicolor* A3(2)	ABC transporter; exporter
SteO08.6c	*aprV*	621	AprV	NP_823966.1	*S. avermitilis* MA-4680	ABC transporter; exporter
SteO08.7	*aprU*	335	AprU	NP_294117.1	*Deinococcus radiodurans*	kinase
SteO08.8c	*tatA*	181	TatA	NP_639821.1	*S. coelicolor* A3(2)	transposase
SteO08.9c	*tatB*	93	TatB	NP_639821.1	*S. coelicolor* A3(2)	transposase (fragmentary ORF)
SteO08.10c	*tatC*	84	TatC	NP_106540.1	*Mesorhizobium loti*	transposase (fragmentary ORF)
SteO08.11	*aprA*	373	AprA	AAN05728.1	*S. tenebrarius*	unknown biosynthetic protein
SteO08.12	*kamB*	155	KamB	P25920	*S. tenebrarius*	16S rRNA methyltranserase; Apr resistance
SteO08.13c	*aprO*	785	AprO	NP_622451.1	*Thermoanaerobacter* sp.	put. disaccharide hydrolase or phosphorylase
SteO08.14c	*aprD1*	312	AprD1	NP_738423.1	*Corynebacterium efficiens*	NDP-hexose DH or epimerase
SteO08.15	*aprF*	418	AprF	AAQ99278.1	*S. tenebrarius*	unknown biosynthetic protein
SteO08.16c	*aprG*	296	AprG	AAQ99277.1	*S. tenebrarius*	unknown biosynthetic protein
SteO08.17	*aprH*	359	AprH	ZP_00054179.1	*Magnetospiri. magnetotacticum*	glycosyltransferase; Apr-GT I or II
SteO08.18	*aprI*	262	AprI	NP_293969.1	*Deinococcus radiodurans*	oxidase or N-methyltransferase
SteO08.19c	*aprJ*	238	AprJ	NP_737699.1	*Corynebacterium efficiens*	phosphosugar mutase
SteO08.20c	*aprK*	189	AprK	NP_772062.1	*Bradyrhizobium japonicum*	NDP-heptose or -octose synthase
SteO08.21c	*aprM*	438	AprM	AAR98545.1	*M. echinospora*	glycosyltransferase; Apr-GT II or I
SteO08.22c	*aprC*	384	AprC	AAR98548.1	*M. echinospora*	2-deoxy-*scyllo*-inosose synthase, cyclase
SteO08.23c	*aprS*	454	AprS	AAR98547.1	*M. echinospora*	ketocyclitol aminotransferase I and II
SteO08.24c	*aprE*	373	AprE	AAR98562.1	*M. echinospora*	aminocyclitol 1-dehydrogenase
SteO08.26c	*aprP*	249	AprP	T48846	*Pseudomonas* sp.	amidohydrolase
SteO08.27c	*aprD2*	329	AprD2	ZP_00044795.1	*Magnetococcus* sp. MC-1	oxidoreductase
SteO08.28c	*aprD3*	260	AprD3	AAR98544.1	*M. echinospora*	UDP-N-acetylglucosamine 4,6-dehydratase
SteO08.29c	*aprY*	285	AprY	EAA49425.1	*Magnaporthe grisea* 70-15	(poss. not an active reading frame)
SteO08.31	*aprQ*	503	AprQ	NP_769359.1	*Bradyrhizobium japonicum*	hexosamine 6-dehydrogenase

Tab. A.11 Proteins encoded in the genomic region covering the the *apr*-cluster of *S. tenebrarius* DSM 40477 (accession code: AJ629123), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in Apr pathway or else [a]
SteO08.32	*aprD4*	457	AprD4	NP_614275.1	*Methanopyrus kandleri*	biosynthetic Fe-S oxidoreductase
SteO08.33	*aprZ*	462	AprZ	P09401	*S. griseus*	extracellular aminoglycoside-phosphate phosphatase
SteO08.34	*aprL*	373	AprL	NP_819727.1	*Coxiella burnetii* RSA 493	aminotransferase class IV
SteO08.35	*aprD5*	348	AprD5	NP_819706.1	*Coxiella burnetii* RSA 493	NAD-dependent epimerase/dehydratase
SteO08.36c	(*aprT*)	392	(AprT)	NP_817764.1	Mycobacteriophage Rosebush	PP-loop superfamily ATPase
SteO08.37c		61			no significant similarity	HP
SteO08.38c		158			no significant similarity	HP

[a] Abbreviations: see under Tab. A.1

Tab. A.12 Proteins encoded in the genomic region covering the *apr*-cluster of *St. hindustanus* DSM 44523 (accession code AJ875019)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in Apr pathway or else [a]
ShinN01.1c	*aprN*	170	AprN	NP_634124.1	*Methanosarcina mazei*	TTP-D-glucosmamine synthase
ShinN01.2c	*aprW*	626	AprW	NP_823966.1	*S. coelicolor* A3(2)	ABC transporter; exporter
ShinN01.3c	*aprV*	600	AprV	NP_823966.1	*S. avermitilis* MA-4680	ABC transporter; exporter
ShinN01.4	*aprU*	336	AprU	NP_294117.1	*Deinococcus radiodurans*	kinase
ShinN01.5	*aprA*	373	AprA	AAN05728.1	*S. tenebrarius*	unknown biosynthetic protein
ShinN01.6	*kamB*	155	KamB	P25920	*S. tenebrarius*	16S rRNA methyltranserase; Apr resistance
ShinN01.7c	*aprO*	780	AprO	NP_622451.1	*Thermoanaerobacter* sp.	put. disaccharide hydrolase or phosphorylase
ShinN01.8c	*aprD1*	312	AprD1	NP_738423.1	*Corynebacterium efficiens*	NDP-hexose DH or epimerase
ShinN01.9c	*aprF*	425	AprF	AAQ99278.1	*S. tenebrarius*	unknown biosynthetic protein
ShinN01.10c	*aprG*	298	AprG	AAQ99277.1	*S. tenebrarius*	unknown biosynthetic protein
ShinN01.11	*aprH*	358	AprH	ZP_00054179.1	*Magnetospirr. magnetotacticum*	glycosyltransferase; Apr-GT I or II
ShinN01.12	*aprI*	262	AprI	NP_293969.1	*Deinococcus radiodurans*	oxidase or N-methyltransferase
ShinN01.13c	*aprJ*	237	AprJ	NP_737699.1	*Corynebacterium efficiens*	phosphosugar mutase
ShinN01.14c	*aprK*	163	AprK	NP_772062.1	*Bradyrhizobium japonicum*	NDP-heptose or -octose synthase
ShinN01.15c	*aprM*	469	AprM	AAR98545.1	*M. echinospora*	glycosyltransferase; Apr-GT II or I
ShinN01.16c	*aprC*	386	AprC	AAR98548.1	*M. echinospora*	2-deoxy-*scyllo*-inosose synthase, cyclase
ShinN01.17c	*aprS*	424	AprS	AAR98547.1	*M. echinospora*	ketocyclitol aminotransferase I and II
ShinN01.18c	*aprE*	338	AprE	AAR98562.1	*M. echinospora*	aminocyclitol 1-dehydrogenase
ShinN01.19c	*aprP*	272	AprP	T48846	*Pseudomonas* sp.	amidohydrolase
ShinN01.20c	*aprD2*	319	AprD2	ZP_00044795.1	*Magnetococcus* sp. MC-1	UDP-N-acetylglucosamine 4,6-dehydratase
ShinN01.21c	*aprD3*	248	AprD3	AAR98544.1	*M. echinospora*	biosynthetic short chain oxidoreductase
ShinN01.22	*aprQ*	503	AprQ	NP_769359.1	*Bradyrhizobium japonicum*	hexosamine 6-dehydrogenase
ShinN01.23	*aprD4*	456	AprD4	NP_614275.1	*Methanopyrus kandleri*	biosynthetic Fe-S oxidoreductase
ShinN01.24	*aprZ*	466	AprZ	P09401	*S. griseus*	extracellular aminoglycoside-phosphate phosphatase
ShinN01.25	*aprL*	373	AprL	NP_819727.1	*Coxiella burnetii* RSA 493	aminotransferase (SMAT-type)
ShinN01.26	*aprD5*	345	AprD5	NP_819706.1	*Coxiella burnetii* RSA 493	NAD-dependent epimerase/dehydratase
ShinN01.27		449		NP_817764.1	*Thermus thermophilus* HB27	phenylacetyl-CoA ligase
ShinN01.28c		267		NP_626295.1	*S. coelicolor* A3(2)	membrane protein
ShinN01.29c		180		YP_056267.1	*Propionibacterium acnes*	membrane protein

Tab. A.12 Proteins encoded in the genomic region covering the *apr*-cluster of *St. hindustanus* DSM 44523 (accession code AJ875019), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in Apr pathway or else [a]
ShinN01.30		239		NP_626287.1	*S. coelicolor* A3(2)	membrane protein
ShinN01.31		232		BAA02038.1	*S. fradiae*	glutamyl endopeptidase II precursor

[a] Abbreviations: see under Tab. A.1

Tab. A.13 Proteins encoded in the genomic region covering the *hyg*-cluster of *S. hygroscopicus* subsp. *hygroscopicus* DSM 40578 (accession code: AJ628642)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in HM-B pathway or else [a]
ShyG17.1		164		Q50434	*Mycob. smegmatis*	cHP
ShyG17.2		109		ZP_00092220.1	*Azotobacter vinelandii*	cHP
ShyG17.3		332		NP_625292.1	*S. coelicolor* A3(2)	poss. iron transport lipoprotein
ShyG17.4		254		NP_250611.1	*Pseudom. aerugin.*	cHP
ShyG17.5		348		PSPTO3257	*Pseudom. syringae* pv. *tomato*	iron ABC transporter
ShyG17.6		272		AGR_L_2857p	*Agrobacterium tumefaciens*	FecE protein
ShyG17.7		344		NP_737792.1	*Corynebacterium efficiens*	cHP, not in cluster
ShyG17.8		474		NP_624443.1	*S. coelicolor* A3(2)	transposase
ShyG17.9c	*hygV*	616	HygV	NP_823967.1	*S. avermitilis* MA-4680	ABC transporter; HM-B exporter
ShyG17.10c	*hygW*	604	HygW	NP_629588.1	*S. coelicolor* A3(2)	ABC transporter; HM-B exporter
ShyG17.11	*hygA*	332	HygA	P09979	*S. hygroscopicus* subsp. *hygrosc.*	APH(7"), HM-B-resistance
ShyG17.12c	*hygK*	319	HygK	NP_627354.1	*S. coelicolor* A3(2	NDP-heptose-(or hexose) dehydrogenase
ShyG17.13c	*hygU*	162	HygU	ZP_00080098.1	*Geobacterium metallireducens*	put. dehydratase or phosphatase, HisB/LmbK homolog
ShyG17.14	*hygD*	401	HygD	BAC55142.1,	*Methylobacterium* sp. 12S	glycosyltransferase. put. GT II
ShyG17.15c	*hygO*	287	HygO	NP_627355.1	*S. coelicolor* A3(2	NDP-heptose synthase
ShyG17.16c	*hygE*	339	HygE	NP_627680.1	*S. coelicolor* A3(2	aminocyclitol 1-dehydrogenase
ShyG17.17c	*hygP*	198	HygP	NP_214627.1	*Mycobacterium tuberculosis*	phosphoheptose isomerase
ShyG17.18c	*hygX*	258	HygX	NP_034856.1	*S. avermitilis* MA-4680	dioxygenase or hydroxylase (orthoester formation?)
ShyG17.19c	*hygL*	377	HygL	NP_440612.1	*Synechocystis* sp.	class IV aminotransferase; (NDP-)heptose AT
ShyG17.20c	*hygJ*	308	HygJ	NP_614008.1	*Methanopyrus kandleri*	(NDP-)heptose (or -hexose) DH or epimerase
ShyG17.21c	*hygY*	343	HygY		*S. netropsis*	FeS-cluster oxidoreductase condensation; SpcY-related
ShyG17.22c	*hygF*	270	HygF	AAD45551.1	*S. netropsis*	HM-B GT I; SpcF-related; UDP-Gal: 2DOS galactosyltransferase
ShyG17.23c	*hygN*	351	HygN	ZP_00019295.1	*Chloroflexus aurantiacus*	heptose-(7-P) 1-kinase
ShyG17.24c	*hygC*	410	HygC	BAA83344.1	*B. circulans*	2-deoxy-*scyllo*-inosose synthase, cyclase
ShyG17.25c	*hygM*	272	HygM	AAD28488.1	*S. spect*abilis ATCC27741	N-methyltransferase; SpcM-related
ShyG17.26c	*hygS*	434	HygS	AAD45549.1	*S. netropsis*	ketocyclitol aminotransferase I and II

Tab. A.13 Proteins encoded in the genomic region covering the *hyg*-cluster of *S. hygroscopicus* subsp. *hygroscopicus* DSM 40578 (accession code: AJ628642), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in HM-B pathway or else [a]
ShyG17.27	*hygT*	182	HygT		no significant similarity	HP, unknown function
ShyG17.28	*hygI*	163	HygI	CAD60534.1	*S. cinnamoneus*	sensor-dependant regulator?; conserved in other antibiotic biosynthesis clusters (Cinorf12)
ShyG17.29	*hygH*	168	HygH	CAD60535.1	*S. cinnamoneus*	sensor-dependant regulator?; conserved in other antibiotic biosynthesis clusters (Cinorf13)
ShyG17.30	*hygG*	392	HygG	CAD60536.1	*S. cinnamoneus*	sensor-dependant regulator?; conserved in other antibiotic biosynthesis clusters (Cinorf14)
ShyG17.31c	(*hygZ*)	662	(HygZ)	AAD45557.1	*S. netropsis*	transposase; end of *hyg* cluster
ShyG17.32c		286		NP_631871.1	*S. coelicolor* A3(2)	cHP, end of chromosome (?)
ShyG17.33c		447		NP_624364.1	*S. coelicolor* A3(2)	put. DNA-binding protein, end of chromosome (?)

[a] Abbreviations: see under Tab. A.1

Tab. A.14 Proteins encoded in the genomic region covering the *ist*-cluster of *S. tenjimariensis* ATCC 31603 (accession code: AJ845083)

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in IM pathway or else
StenF24.1		321		NP_821300.1	*S. avermitilis* MA-4680	transport permease
StenF24.2		314		NP_825884.1	*S. avermitilis* MA-4680	adenine glycosylase
StenF24.3		321		NP_825883.1	*S. avermitilis* MA-4680	RNA polymerase ECF-subfamily sigma factor
StenF24.4		206		NP_825882.1	*S. avermitilis* MA-4680	cHP; unknown
StenF24.5		234		NP_627566.1	*S. coelicolor* A3(2)	two-component system response regulator
StenF24.6		444		NP_627567.1	*S. coelicolor* A3(2)	two-component system sensor kinase
StenF24.7c		238		NP_627576.1	*S. coelicolor* A3(2)	peptidase (put. secreted protein)
StenF24.8c		841		NP_627581.1	*S. coelicolor* A3(2)	Clp-family ATP-binding protease
StenF24.9		237		NP_627582.1	*S. coelicolor* A3(2)	proline-rich protein
StenF24.10c		47		NP_627584.1	*S. coelicolor* A3(2)	fragmentary put. acetyltransferase
StenF24.11c		616		2124370K	*Amycolatopsis methanolica* plasmid pMEA300	plasmid transfer protein
StenF24.12c		86		NP_824907.1	*S. avermitilis* MA-4680	cHP; unknown
StenF24.13c		100				HP; unknown
StenF24.14c		111		NP_628786.1	*S. coelicolor* A3(2)	cHP; unknown
StenF24.15c		353		NP_106920.1	*Mesorhizobium loti*	conjugal transfer protein
StenF24.16c		93		NP_824897.1	*S. avermitilis* MA-4680	cHP; unknown
StenF24.17c		88		NP_628789.1	*S. coelicolor* A3(2)	cHP; unknown
StenF24.18c		71		NP_880746.1	*Bordetella pertussis* Tohama I	isocitrate lyase (fragmentary)
StenF24.19	(*issC)*	251	(IssC)	NP_824898.1	*S. avermitilis* MA-4680	GntR-family transcriptional regulator; not in cluster?
StenF24.20	(*issB*)	229	(IssB)	NP_862161.1	*S. violaceoruber*	ATP/GTP-binding protein; not in cluster?
StenF24.21c	(*issA*)	388	(IssA)	ZP_00282077.1	*Burkholderia fungorum*	hydrolase or acyltransferase; not in cluster?
StenF24.22	*imrA*	220	ImrA	D13170	*S. tenjimariensis*	kanamycin-apramycin-resistance methyltransferase
				P25919	*Saccharopolyspora hirsuta*	(16S rRNA methyltransferase)
StenF24.23c	*istA*	397	IstA	AAN05728.1	*S. tenebrarius*	cHP; unknown; ArpA-related
StenF24.24c	*istD*	238	IstD	BAD20763.1	*S. kanamyceticus*	TTP-D-glucosmamine synthase
StenF24.25c	*istP*	276	IstP	AAR98555.1	*M. echinospora*	APH(3'); biosynthetic phosphotransferase
StenF24.26c	*istB*	448	IstB	AAR98554.1	*M. echinospora*	AGA 6'-aminotransferase

Tab. A.14 Proteins encoded in the genomic region covering the *ist*-cluster of *S. tenjimariensis* ATCC 31603 (accession code: AJ845083), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in IM pathway or else
StenF24.27c	*istJ*	621	IstJ	EAA49750.1	*Magnaporthe grisea 70-15*	component in transmembrane efflux complex
StenF24.28c	*istF*	135	IstF	EAA48169.1	*Magnaporthe grisea 70-15*	HP; unknown
StenF24.29c	*istU*	277	IstU	NP_641155.1	*Xanthomonas axonopodis*	N-methyltransferase (LmbJ related)
StenF24.30c	*istZ*	484	IstZ	S30159	*M. olivasterospora*	N-formimidoyl fortimicin A synthase
StenF24.31	*istH*	617	IstH	EAA49750.1	*Magnaporthe grisea 70-15*	component in transmembrane efflux complex
StenF24.32	*istW*	320	IstW	NP_899815.1	*Chromobacterium violaceum*	glucose-6-P 1-dehydrogenase
StenF24.33c	*istQ*	545	IstQ	AAR98543.1	*M. echinospora*	AGA 6'-dehydrogenase
StenF24.34c	*istC*	396	IstC	BAA83344.1	*S. kanamyceticus*	2-deoxy-*scyllo*-inosose synthase/cyclase
StenF24.35c	*istL2*	348	IstL2	NP_558504.1	*Pyrobaculum aerophilum*	metallo cofactor biosynthesis protein
StenF24.36c	*istL*	327	IstL	Q08324	*M. olivasterospora*	Fortimicin KH epimerase (FTM KH epimerase)
StenO22.6	*istI*	483	IstI	NP_883648.1	*M. echinospora*	transmembrane efflux protein in
				NP_883648.1	*Bordetella bronchiseptica* RB50	complex
StenO22.7c	*istS*	424	IstS	CAE22472.1	*S. tenebrarius*	2-deoxy-*scyllo*-inosose aminotransferase
StenO22.8c	*istG*	257	IstG	ZP_00202638.1	*Ralstonia eutropha* JMP134	put. 6'-N-methyltransferase
StenO22.9c	*istX*	150	IstX		*M. olivasterospora*	unknown
				ZP_00275047.1	*Ralstonia metallidurans* CH34	
StenO22.10	*istM*	427	IstM	BAD20768.1	*S. kanamyceticus*	glycosyltransferase
StenO22.11	*istO*	385	IstO		*M. olivasterospora*	put. Fe-S oxidoreductase or 3',4' reductase
				NP_375972.1	*Sulfolobus tokodaii*	
StenO22.12	*istN*	247	IstN		*M. olivasterospora*	put. 1-O-methyltransferase
				AAD41819.1	*S. fradiae*	
StenO22.13	*istE*	331	IstE		*M. olivasterospora*	cyclitol (3-/or 6-) dehydrogenase
				CAF60530.1	*S. Kanamyceticus*	
StenO22.14	(*ist*?)	164	Ist?	NP_968608.1	*Bdellovibrio bacteriovorus*	put. tRNA/rRNA MT; not in cl.?
StenO22.15c	(*ist*?)	471	Ist?	NP_772500.1	*Bradyrhizobium japonicum*	transcriptional regulatory protein; not in cl.?
StenO22.16		222		YP_116837.1	*Nocardia farcinica*	translocator (amino acid efflux)
StenO22.17c		217		BAC74556.1	*S. avermitilis* MA-4680	integral membrane protein
StenO22.18c		167		NP_631504.1	*S. coelicolor* A3(2)	HP, unknown

Tab. A.14 Proteins encoded in the genomic region covering the *ist*-cluster of *S. tenjimariensis* ATCC 31603 (accession code AJ845083), continued

ORF design.	Gene symbol	Gene product aa	design.	Similarity to other proteins in the data base acc. code	organism	Put. function in IM pathway or else
StenO22.19		216		NP_826538.1	*S. avermitilis* MA-4680	tyrosinase co-factor protein
StenO22.20		288		NP_826539.1	*S. avermitilis* MA-4680	tyrosinase
StenO22.21c		135		NP_626933.1	*S. coelicolor* A3(2)	small membrane protein
StenO22.22c		91		NP_626937.1	*S. coelicolor* A3(2)	membrane protein
StenO22.23c		182		ZP_00222667.1	*Burkholderia cepacia*	HP, unknown
StenO22.24		128		NP_627098.1	*S. coelicolor* A3(2)	HP, unknown
StenO22.25		331		NP_627099.1	*S. coelicolor* A3(2)	HP, unknown
StenO22.26		399		NP_627100.1	*S. coelicolor* A3(2)	HP, unknown
StenO22.27c		389		NP_898643.1	*Rhodococcus erythropolis*	HP, unknown
StenO22.28		230		YP_077641.1	*B. licheniformis*	HP, unknown
StenO22.29c		1065		AAL25727.1	*Rhodococcus ruber*	transposase
StenO22.30		210		NP_821633.1	*S. avermitilis* MA-4680	hsp18 transcriptional regulator
StenO22.31		119		XP_479904.1	*Oryza sativa*	HP (fragmentary), unknown
StenO22.32		92				HP (fragmentary), unknown
StenO22.33		194		YP_138136.1	*Haloarcula marismortui*	HP, unknown
StenO22.34c		169		NP_907570.1	*Wolinella succinogenes*	MT, unknown
StenO22.35c		149		NP_795557.1	*Streptococcus pyogenes* phage 315.4	recombination protein
StenO22.36c		170		AAK02044.1	*Salmonella enterica*	resolvase

[a] Abbreviations: see under Tab. A.1

6.2 Figures for alignments and phylograms of various proteins encoded by genes conserved among the ACAGA clusters analysed

```
NeoC       ----------------------MQTTRIAMEDASFPYRLGTDCAEDVVARLAALEASSY 37
RibC       ----------------------MQVTPIAMEDASFAYRLGTECTEDVVARLATLGASSY 37
LivC       ----------------------MHVTAITMEDANFPYRLGTDCAEELVARLGVRAASRY 37
ParC       ----------------------MHVTAITMEDTSFPYRLGTECAEEIVARLGERAASRY 37
AprC       ----------------------MKVTTIAMGDVRFPYRLGTSCVDEIVDRLADLDASRY 37
KanC       ----------------------MQVTTITMDDVQYPYRLGTDCLDGIVTRLGELGASRY 37
TobC       ----------------------MQTTTITMGDVQYPYRLGTGCVDGIVTRLGELEASHY 37
GenC       ----------------------M-EVEIRLGSVRYPFRLGTDCLGAIVEDLVAMSASRL 36
HygC       MDLQAREPAAAADATAAPPLPSGLRERSFDIAGKRVRVVLGEHIEAEMAAALAELDADRF 60
BtrC       ----------------------MTTKQICFADRCFNFAFGEHVLESVESYIPRDEFDQY 37

NeoC       LVVADTTVAELYGAALTAHIDKEAGPSHLLTHEVGEVHKTLATVSALAEQALGRGADRRS 97
RibC       LVVADTTVAGLYGHDLTARIDKEAGPAHLLTHESGEVHKDLTTVSVLAEQALERGADRRS 97
LivC       LVVCDTTVAALYGHDLVARLEKDAGPAVLLTHQVGEVHKDVTTVGALAEQALGAGADRRS 97
ParC       LVVCDTTVAALYGRDLVARLEKDAGPAVLLTHPAGEVHKRIGTVGDLAEQALAAGADRRS 97
AprC       LVVADTTVAGLYGHDLVRRIRGTAGPADLLLHQGGEEHKGLGTLTELVESALRLGADRRS 97
KanC       LIVSDPRVAELYGQGLRERLAEQAGPAELITHASGEQNKGLPALHDLAEEALRRGADRQS 97
TobC       LVLCDATVAELYGHDLAARLRRSAGPASVLTHPAGEEHKGLGTLDTLADAALHAGVDRRG 97
GenC       LIVCDSNTGPLFGAELVERLSPRV-PANLLIHRAGEPYKDLQAVGTLADSALQLGADRAS 95
HygC       VVVTDARVAELAAGALAERLERTT-PTLLLSHPPGEEFKNLATLSRFVDLALDFGVTRRS 119
BtrC       IMISDSGVPDSIVHYAAEYFGKLA-PVHILRFQGGEEYKTLSTVTNLQERAIALGANRRT 96

NeoC       VVVALGGGVTGNIAGLMASLLFRGIRLVHVPTTVVAMLDSVLSLKQAVNTTFGKNLAGTF 157
RibC       VVVALGGGVTGNITGLMASLLFRGIRLVHVPTTVVAMLDSVLSLKQAVNATFGKNLVGTF 157
LivC       VVVALGGGITGNIAGLLAALLFRGITLVHVPTTVVAMLDSVLSLKQAVNASFGKNLVGTF 157
ParC       VVVALGGGITGNIAGLLASLLFRGITLVHVPTTVVAMLDSVLSLKQAVNASFGKNLVGTF 157
AprC       VVVALGGGITGNIAGLMAALMFRGIRLVHVPTTIVAMLDSVLSLKQAVNSSVGKNLVGTF 157
KanC       IVVALGGGVTGNIAGLLAALLFRGIRLVHVPTTVVAMLDSVLSLKQAVNAGVGKNLVGTF 157
TobC       VVVALGGGVTGNIAGLLAALLFRGIRLVHVPTTVVAMLDSVLSLKQAVNAQVGKNLVGTF 157
GenC       VVVAVGGGVIGNIAGLMAALLFRGIRLVHIPTSLIAMSDSVLSLKQAVNACVGKNLMGTF 155
HygC       VVVAVGGGIPGNIAGVLAGLLFRGLRLVHVPTTVIAACDSVLSVKQAVNTGHAKNTVGLY 179
BtrC       AIVAVGGGLTGNVAGVAAGMMFRGIALIHVPTTFLAASDSVLSIKQAVNLTSGKNLVGFY 156

NeoC       YQPVEVLADTAALRTLPPREIRSGMGEVVKNALAIRPAMLDRLAGALRPDTRYDDETMRW 217
RibC       YQPVEVLADTAFLRTLPPREIRSGLGEVVKNALAIRPAMLDRLGDALRADARYDDETLRW 217
LivC       YQPAEVLADTAMLRTLPEREVRSGMGEVVKNALAVRPAMADRLAGLLRPDARYDDDALRW 217
ParC       YQPAEVLADTAMLRTLPARELRSGMGEVVKNALAIRPSMIERLAAELRPDARYEDAAMRW 217
AprC       YQPVEVLADTEFLRTLSRREIRSGLCEVVKNALAIRPSMIDRLAGLLPLDERYDDEAMRW 217
KanC       YQPVEVLADTAMLRTLPVREVRSGMCEVVKNSLAIRPSMIDQLSAGLRPDGRYPDDTMHW 217
TobC       YPPVEVLADTAMLGTLPVREIRSGLCEVVKNALAIRPSMIDFLAAELRPDGRYADDVLRW 217
GenC       YAPESVLADTAMLRSLPFRETVSGLCEVVKNSLAIRPSMVEMLRTSLRQDAVYDDETMYE 215
HygC       HIPEAVLVDLAVVANGSPRDLRSGTCETVKNALAIEPAQIPRLRRLLRREADYSPADLGE 239
BtrC       YPPRFVFADTRILSESPPRQVKAGMCELVKNMLILENDNKEFTEDDLNSANVYSPKQLET 216

NeoC       IIAESLAAKADVTSGDKHERRSGLVLEYGHTAGHAIEHASRGA------VAHGAGVAVGM 271
RibC       IIAESLTAKADVTRDDKHERRTGLVLEYGHTAGHAIEHASRGE------VAHGAGVAVGM 271
LivC       IIAESVAAKADVTGADKHERGDGLVLEYGHTAGHAIEHAARGA------VAHGAGVAVGM 271
ParC       IIEESVAAKAQVTGADKHERRDGLVLEYGHTTGHAIEHAARGE------VAHGAGVAIGM 271
AprC       IIDESIAAKAQVTENDVHERGQGLVLEYGHTVGHALEHAARGE------VSHGAAVGLGM 271
KanC       IIYESLAAKAQVTAYDKYERGEGLILEYGHTVGHAVEHSSQGA------VPHGAAVALGM 271
TobC       MIDESVAAKAQVTEHDKYERREGLVLEYGHTVGHALEHASHGA------VSHGAGVGVGM 271
GenC       IISESILAKASVTVDDMHECRAGLVLEYGHTVGHAIEYTAAGG------LSHGQAIGLGM 269
HygC       IFELSLAVKGRLLLEDPYEKHAGVLLEYGHTVGHALELAGAGRGPGGDMIKHGEAVALGM 299
BtrC       FINFCISAKMSVLSEDIYEKKKGLIFEYGHTIGHAIELAEQGG------ITHGEAIAVGM 270
```

```
NeoC        TLAAEVSRRLGHADAGLVALHRELVAAAGVEPAVPDHVDTALVKNWLAYDNKRGYLDSPP 331
RibC        TIAAEVSRRLGHAGPDFVALHRELVAAAGVEPAVPAHVDPALVKNWLAYDNKRGYLDSPP 331
LivC        VIAAEVSHRLGHASASFVARHRELISKAGLEDTVPACVRTDDVKNWLTYDNKRGYLDCAA 331
ParC        IVAAEVSRLLGHASGDLVGLHRELVAKAGLEGSVPALVDPADVKHWLTYDNKRGYMPCPP 331
AprC        LAEAEIARRLGLAGDDMVQLHRELIGRVGIDLEFPSHVSADDVRFRLRYDNKRGYLSDAG 331
KanC        IAAAQVSHRAGWASAELVDLHRELVAKTGVARRIPSDIPLSAVRHRLSFDNKRGYLPASA 331
TobC        VAAAEVARRLGHVDADLVELHRELVGKVGVATTLPADVPTEEITYRLGFDNKRGYQPLPA 331
GenC        VVAAEVSHRLGHLDQEAVALHRELLTRAGAMVTIPEEVDLDEVMHRLRFDNKRGYLADPA 329
HygC        TAAADVAHRAGLLDAEGLAVHEELIDRIGICRCLTVGIAPETVLHHIAYDNKRGYRSCDG 359
BtrC        IYAAKIANRMNLMPEHDVSAHYWLLNKIGALQDIPLKSDPDSIFHYLIHDNKRGYIKLDE 330

NeoC        GHTPMVLLSAPGEVLHTGTMPLVPVPLALLEEVVDESAARGGAGGGAAEPAAARTGPVPD 391
RibC        GHTPMVLLSAPGEVLHTGPMPLVPVPLALLEEAVDEAARRG----RDAAPAAAYVG---- 383
LivC        DTTPMVLLAGPGRPLRTGGMPLVPVPLAVLNETVDALAAPG--RAGTDHRTAVPV----- 384
ParC        AATPMVLLSAPGEVLRSGPLPLVPVPLELLGRAVDALAAPAGQSAGAERLSPAPA----- 386
AprC        DGYVIVLLKAPGVPLQSEGLPLTVADPAVVDAVVSEFAHPA---TAAAIGRRTPVG---- 384
KanC        DTYPMVLLESPGKVLRSEGTVLTAAPRDLVDAVVDELAEPPRPAAARTDDAATVLGGAG- 390
TobC        DHYAMVLLADVGQPLYQDGLPLTPAPRALVDEVVRELADAP----SRIGASVGSAGGAS- 386
GenC        ESSAMVLLGGLGEPLWHDGRPLVSVPMALVGEVVNEIARPEIPNFELVAPVETVEEGRVP 389
HygC        SEVTMVLLRAVGAPLNAEQRYLTPVPQTLVAAAVHDLLRRG----KECTAGHGPS----- 410
BtrC        DNLGMILLSGVGKPAMYNQTLLTPVRKTLIKEVIREGL---------------------- 368
```

Fig. A.1 Alignment of BtrC homologous. Alignment was performed using Clustal W program (cf. Sect. 2.26). The numbers indicate the position within the corresponding proteins.

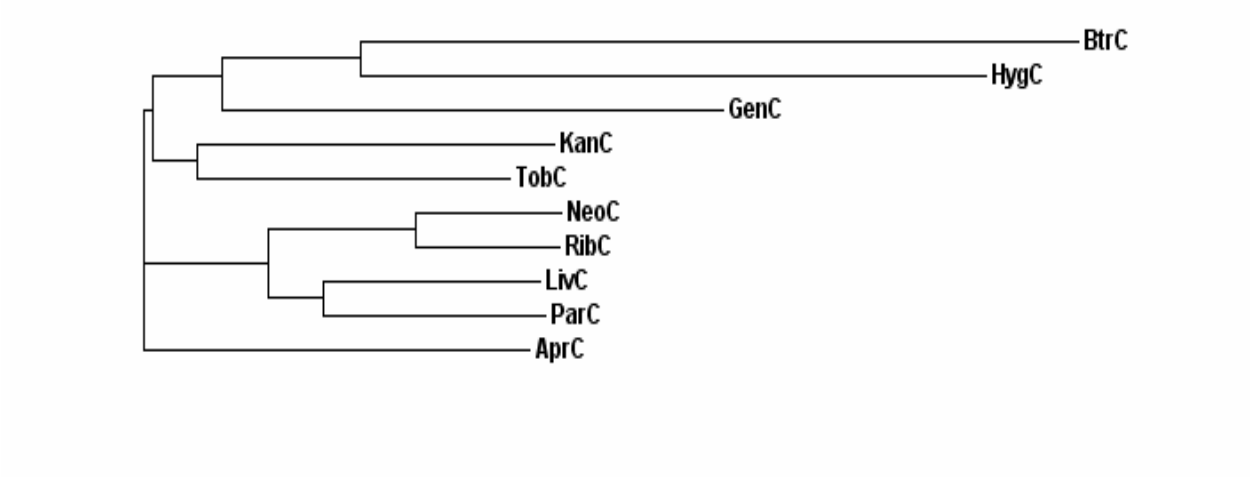

Fig. A.2 Phylogram of BtrC homologous. Phylogram was calculated using Clustal W program (cf. Sect. 2.26).

```
BtrS        --------MTIP------------------FDHWPEWPQHSDRTRRKIEEVFQSNRWAI 33
HygS       MPGPVSGVTPIPGTFPGRTAGGPHRRFLMTDAQPWPAWPVHTPDTERWLLEVVRSGRWAV 60
ParS        --------MTRS---LAVQGGSPVR------TRPWPLWPQPAPGAVRALDGVLTSGRWSI 43
LivS        --------MAQS---LAVQGGSAVR------TRPWPVWPRPAAGAAEAVQQVLSSGRWSI 43
NeoS        --------MVSP---LAVKGGEALR------TRPWPAWPQPAPGVPAAVAEVLGSGRWSI 43
RibS        --------MVSQ---LAVKGGEALR------TRPWPAWPQPAPGVPDAVADVLGSGRWSI 43
KanS1       --------MPLQSSRLAVDNGTPVR------GKPWPVWPQPTDGTLDALSRVLRSGRWAI 46
TobS1       --------MPVH---LAINNGTPVR------TRPWPVWPQPARGALDALERVLRSGRWAI 43
GenS1       --------MTLL----AVNGGSPIR------SQQWPLWPAPAPGALDALNEVLHSGRWAI 42
StsC        --------MDSS---LAISGGPRLS------NREWPRWPQPGDRALKSLEDVLTSGRWTI 43
KanS2       --------MSKK---LALFGGTPVR------NEEFYDGPHIGPHDLDRLKSVLDSG---N 40
TobS2       --------VTSE---LALFGGTPVR------TEPFPDGPRFRERDLERIREVLESG---S 40
GenS2       --------MTQK---LAILGGDPVR------TRPWPEWPHVGPEDVDRLRTVIESR---N 40

BtrS        SGYWTGEESMERKFAKAFADFNGVPYCVPTTSGSTALMLALEALGIGEGDEVIVPSLTWI 93
HygS        SGAFTGRPSYEQEFADAFARFIGVPYAVPVCNGSAALTVALEALGVRAGDEVLVPGLTWV 120
ParS        SGPYRGAASQERRFAQAFAAYNGVEHCVPAASGTASLMLAMEACGIGAGDEVIVPGLSWV 103
LivS        SGPYRGAASQERRFARAFAEYNGVAHCVPAASGTASLMLALEACGVGAGDEVIVPGLSWV 103
NeoS        SGPYRGTDSHERRFARAFADYHGVPYCVPAASGTAGLMLALEACGVGAGDEVIVPGLSWV 103
RibS        SGPYRGTESYERRFARAFAAYNGVPHCVPAASGTASLMLALEACGIGAGDEVIVPGLSWV 103
KanS1       SGPYRGVESAERRFARRFADYHRIAHCVPASSGTASLMLALEACGVGAGDEVILPGVTWV 106
TobS1       SGPYRGIESAERRFARDFAAYNGVAHCVPAASGTASLMLALESCGVGVGDEVIAPGLSWV 103
GenS1       SGPYQGKQSFERRFAAAFAEFHEIGHCVPTSSGTASLMVALEACGVGAGDEVIIPGLTWV 102
StsC        SCAYQGRDSYERQFASAFADYCGSAMCVPISTGTASLAIALEACGVGAGDEVIVPGLSWV 103
KanS2       FGGIPFPNTHHTAFADLFTGKLGAPYGLMVSNGTISLSIALRALGVRAGDEVITTGYTWM 100
TobS2       LGGIPFPNRTHRAFAEQFCGRLGARHGVLVANGTVSLSVALRALGVHAGDEVITTGYTWM 100
GenS2       LGGIPFPNTMHQQFAERFTAKLGAKYGLLATNGTVTLSMALRALGIHAGDEVITTAFTWV 100

BtrS        ATATAVLNVNALPVFVDVEADTYCIDPQLIKSAITDKTKAIIPVHLFGSMANMDEINEIA 153
HygS        ACASAVLRVGAVPVLVDISPKTLCLSVEAAAASVTPRTRAVMAVHLHGSAVDMDALTELA 180
ParS        ASGSTVLGVNAVPVFCDVDPDTLCLDPAAVESALTERTKAIVVVHLYSAVAAMDALRALA 163
LivS        ASGSTVLGVNAVPVFCDVDPRTLCLDPEAVEAAVTERTKAIVVVHLYSAVADMDALTALA 163
NeoS        ASGSTVLGVNAVPVFCDVDPDTLCVSPEAVEALITERTRAVVVVHLYSAVADMDGLTRVA 163
RibS        ASGSTILGVNAVPIFCDVDPDTLCLSPEAVEAAITEHTRAIVVVHLYSALADMDALSAIA 163
KanS1       ASASTVVGVNAVPVFADIDPDTLCLDPDAVEAAITPATKAIVVVHLYAAVADLTRLKEVA 166
TobS1       ASASTIVGVNAVPVLVDIDPRTLCLDPAAVEAAITPATKAVVVVHLYSAVADLDALRAVA 163
GenS1       ANASTVAGVNAVPVPVDVDPQTLCLDPAAVERAITPRTAAIVVVHLYSAVADLDALTAIA 162
StsC        ASASAVLGINAVPVLVDVDPATYCLDPAATEAAITERTRAITVVHAYSAVADLDALLDIA 163
KanS2       GTAAAIVHINAVPVLVDIDPTTWCIDPAAVEAAITPRTKVIVPVHLGNQIADLDALRAIA 160
TobS2       GTAASIVHINAVPVLVDIDPNTWCIDPAAVEAAITPRTRAIVPVHLANQIADLDALLEIA 160
GenS2       GTVAGIVHVNAVPVLADISDDNWCIDPVKVEEAITDRTRAIMVVHLGNQVADMDALLDIC 160

BtrS        QEHNLFVIEDCAQSHGSVWNNQRAGTIGDIGAFSCQQGKVLTAGEGGIIVTKNPRLFELI 213
HygS        RRHDLAVLEDCSHVHGAAWRDRRLGSLGDAAAFSLQQTKLLTSGEGGVATTSDPAVHELL 240
ParS        DRHGLPLLEDCAQAHGAEYRGVKVGALATAGTFSMQHSKVLTSGEGGAVITRDAEFARRV 223
LivS        ERHSLPLIEDCAQAHGAAYRGVKVGALATAGTFSMQHSKMLTSGEGGAVITRDADFARRV 223
NeoS        ERHGLPLVEDCAQAHGASYRGVKVGALATAGTFSMQHSKVLTSGEGGAVITRDADLARRV 223
RibS        ERHGLPLIEDCAQAHGATYRGVKVGALATAGTFSMQHSKVLTSGEGGAVITRDEDFARRV 223
KanS1       DRHGIVLIEDCAQAHGAEFEGHKVGTFGAVGTFSMQQSKVLTSGEGGAAITADPVLARRM 226
TobS1       DRHGLPLIEDCAQAHGAEHRGRKVGSVGDVGTFSMQHSKVLTSGEGGAAITNSAELARRM 223
GenS1       ERHEIPLIEDCAQAHGARYRDRRVGTFGAFGTFSMQHSKVLTSGEGGAVITGDAALSRRA 222
StsC        RRHGLPLIEDCAHAHGAGFRGRPVGAHGAAGVFSMQGSKLLTCGEGGALVTDDADVALRA 223
KanS2       DKHGLAILEDTAHGHFAEWRGQCVGTHGDAGSFSFESSKIMTAGEGGFLVARDEDVYQRM 220
TobS2       RKHDLVVLEDCAHAHFAEWRGRCVGTHGDAGSFSFESSKIMTSGEGGFLVSGDETTHHRA 220
GenS2       RRHNLLLIEDCAHAHFAEWRGRCVGTIGDAGSYSFETSKIMTSGEGGFLVTATEEAFHRA 220
```

```
BtrS        QQLRADSRVYCDDSSELMHGDMQLVKKGDIQGSNYCLSEFQSAILLDQLQELDDKNAIRE 273
HygS        YRKRADGRGPRPDAP---VGELQLAEHPGVQGYNYCMSEFGAALALDGLGRLAEENARRR 297
ParS        EHLRADGRCLAGQPVG--DGQMELVETGELMGSNRCVSEFQAALLVEQLGVLDEQNERRR 281
LivS        EHLRADGRVLAGQRPG--PGEMELVETGELMGNNRCLSEFQAALLTEQLKDLDAQHAIRR 281
NeoS        EHLRADGRCLSDGPPA--PGAMELVETGELMGSNRCLSEFQAAILTEQLTLLDEQNRTRR 281
RibS        EHLRADGRCLSAVPPA--PGAMELVETGELMGNNRCLSEFQAAILAEQLTILDEQNETRR 281
KanS1       EHLRADGRCYRDQAPP--SGHMELVETGELMGSNRCISEFQAAVLTEQLGELDRFNALRR 284
TobS1       EHLRADGRCYPDTAPA--PGRMELVETGELMGSNRCLSEFQAAVLVEQLRELDEQNALRR 281
GenS1       EHLRADGRTYTPDEPA--VGEMELAQTAELMGSNRCLSEFQAALLLGQLELLDEQNERRR 280
StsC        EHLRADGRVVRREPVG--VGEMELEETGRMMGSNACLSEFHAAVLLDQLELLDGQNARRT 281
KanS2       MSLANCGRKEPGYDGF----------AGRTLGWNARASELQAAFMIGQVEQHDALHAKRA 270
TobS2       MSLVNCGRKEEGYDSF----------EGRMLGWNNRATELQAAFLIGQVEQHDELHAQRK 270
GenS2       MSLAHVGRKEAPYDRF----------PGRVFGWNHRATEMQAAVLLGQLDRYDALDKQRT 270

BtrS        KNAMFLNDALSKIDGIKVMKRP-PQVSRQTYYGYVFRFDPVKFGGLNADQFCEILREKLN 332
HygS        RSMRRLEELLAGFDGAAPLSRP-EAVTEDAAYQVCVRLDPDVFGPAAKHTVAAALSRELG 356
ParS        RNAALLDKLLADE-GYRPQTTS-EGTSTRTYYTYAARLPEGELTHVDAAAVGEALTAELG 339
LivS        RNAALLDGLLRES-GYVPQETS-EGTSTRTHYTYAVRLPEGRLTHVGLATVARALSAELG 339
NeoS        ANAARLDGLLGEL-GLRPQATS-EGTTSRTYYTYAARLPEGALEDVPLTDVTGALTAELG 339
RibS        ANAAHLDGLLGEL-GLRPQTTS-DGTTSRTYYTYAVRLPDGVLEDVPVTDVSCALTAELG 339
KanS1       HNAELLDALLTDV-GYRPQRST-PGTTARTYYTYVAELPDAELPGADITKVTEALTAELG 342
TobS1       RNAELLNTLLAEQ-GLRPQATS-PGTTSRTYYVYAAELPDDAFVGLPITTVTEALTAELG 339
GenS1       ANAALLDEGLGAL-GIQPQVSS-PGTTERTYYEWAGRIEDDGIGQIGVERIAPAVAAELS 338
StsC        RAADHLTDRLSEL-GMTAQATA-PGTTARAYYRYLVRLPDEVLAVAPVERFAHALTAELG 339
KanS2       ASAAKLTAGLAEIGGFTPVGNDDPRITRRQYYEVIYRFDPAAWEGLHRDEVLSAILAEGI 330
TobS2       SNVELLTKGLTEIGGFTPVGDDDPRVTRRQYYEVLYRFDPEQWAGVHRDRVLEALLAEGV 330
GenS2       AMAEMLTQGLVEIGGFKPLAED-PRVTRRQRYELLFRFDTEAWDGLHRDKVLEAILAEGV 329

BtrS        MGTFYLHPPYLPVHKNPLFCPWTKNRYLKSVR-KTEAYWRGLHYPVSERAS-GQSIVIHH 390
HygS        V---HVEPIDSALNRNDLYQPLAS-VWATGDR-RRELDPARFRLPVAEAAG-ENAVVLPH 410
ParS        ---FPVAPCYAPITRNRLYDPRSRGRFALGVQHESRIDPKRFELPVCEEAA-RRTVTVHH 395
LivS        ---CTVAPSYAPITRNRLYDPASRRRFALGVEHQALTDPKRFELPVAEDAA-RRVLTLHH 395
NeoS        ---FPVQPCYAPIPANRLYAPQTRRRYTLGPDHEARIDPKRFALPVCEDTA-RRTVTLHH 395
RibS        ---FPVLPSYAPIPANRLYTPHTRRRYTLGLDHERRIDPKRFALPVCEDAA-RRTVTLHH 395
KanS1       ---FPVAPAYSPLNANPLYDPASRSRFALGPQHEKLIDPARFVLPVSGRLT-RRLVTFHH 398
TobS1       ---FPISPAYAPLHTNRLYAPASRRRFALGEEHEKRIDPARFHLPVCERLT-RRLITFHH 395
GenS1       G--AAIYASYPPMNHNRLYQPATRARFKG----IAGLDLTGYSLPVAEDAG-QRVVTIHH 391
StsC        ---FAVTQTHRPLNDNPLNRPSSRRRFATDARYLERVDPSRFDLPAAKRAH-ESVVSFSH 395
KanS2       E--LEGDAFYPPVHKSELFAVDAVHWPMIAERYGDRIGPDSVDLPVADRAAADESVWVHH 388
TobS2       E--FEGITFYPPLHRDSLFTVSAEDWPMIRDRYGDRMGPEDFHLPVSERAAYDESVWVHH 388
GenS2       E--FEGNTFYPPMHRDELFHITADDWPAIRERYGEKIEPDAFHLPVAERVAFDEAVWIHH 387

BtrS        AILLAEPSHLSLLVDAVAELARKFCVTH--- 418
HygS        RMFLAEPERMG---EIVAALEKARAAS---- 434
ParS        AALLGDESDMHDIATAFGKVVRHGALLTG-- 424
LivS        AALLGAEDDMRDIAAAFGKVLRHGADLTA-- 424
NeoS        AALLGDAEDMADIAAAFAKVLRHGADLAT-- 424
RibS        AALLGDADDMGDIAAAFAKVLRHGAGLMH-- 424
KanS1       AALLGDESDMRDIAEAFTKVLQHRAVLAA-- 427
TobS1       AALLGDESDMHDIAAAVAKVLRHHGELRA-- 424
GenS1       SALLGDESDMKDIVRAFEKVFANHRELRG-- 420
StsC        EVLLAPLDAIDDIARAFRKVLDNVREVSR-- 424
KanS2       ALLTGDDKDLGDILEAVAKVRDNLRELHDAS 419
TobS2       SLLTGPATDVDQILEAVAKVRRNVDALR--- 416
GenS2       SLLSVEPEDVQDMLDAVVKVRDNLGALKKSL 418
```

Fig. A.3 Alignment of BtrS homologous. Alignment was performed using Clustal W program (cf. Sect. 2.26). The numbers indicate the position within the corresponding proteins.

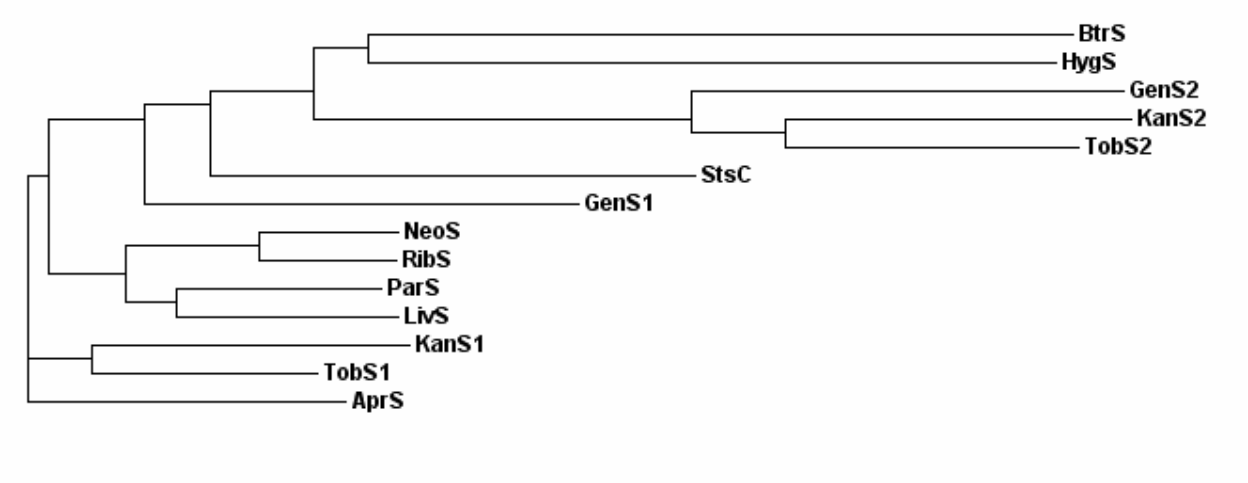

Fig. A.4 Phylogram of BtrS homologous. Phylogram was calculated using Clustal W program (cf. Sect. 2.26).

```
AprE-Ste    MTGPERPRPGRRGRRRADPRRRREPFAAATTRGERMKALVFDAPERALLLDKDVPEPARG 60
AprE-Shin   ----------------------------------MKALVFDAPERALLRDKDVPEPARG 25
KanE        ----------------------------------VKALVFHSPEKATFEQRDVPTPRPG 25
TobE_       ----------------------------------VKALSFEAPGEAVFGTREVPVPAPG 25
NeoE        ----------------------------------MKALVFEAPERAVLTHRDIPAPAPG 25
RibE        ----------------------------------MKALVFEAPERALLTHRDIPDPAPG 25
ParE        ----------------------------------MKALMFKAPLQAVLTERDVPEPAPG 25
LivE        ----------------------------------MKALMFEAPERAVLIQRDLPEPAPG 25
GenE        ----------------------------------MRALQFHGGHKAQIIDTPRPEAPPG 25
HygE        ----------------------------------MRAAVLNGPRDITVTEVPGPRLPEG 25

AprE-Ste    EVLVKIAYNSLCG-SDLSLYRGVWHGFTYPVVPGHEWSGTVVETNGVD---GDLRGARVT 116
AprE-Shin   EALVKIAYNSLCG-SDLSLYRGVWHGFTYPVVPGHEWSGTVVEVNGAD---ESLRGAKVT 81
KanE        EALVHIAYNSICG-SDLSLYRGVWHGFGYPVVPGHEWSGTVVEINGANGHDQSLVGKNVV 84
TobE_       EALIHLGYNSICG-SDLSLYRGVWHGFSYPVVPGHEWSGTVVEVNGPG---AELVGRDVV 81
NeoE        EALVRVAYNSVCG-SDLSFYKGVWHGFTYPVVPGHEWSGSVVDVNGPRG--ADLVGRNVV 82
RibE        EALVRIAYNSVCG-SDLSFYKGVWHGSTYPVVPGHEWSGTVVDVNGPLG--AELIGTNVV 82
ParE        EALVKLAYNSICG-SDLSFYKGVWHGFTYPVVPGHEWSGTVVEAGG-GA--ADLVGQNVV 81
LivE        EALVRLAYNSVCG-SDLSFYKGVWHGSTYPVVPGHEWSGTVVDVNG-AD--SGLVGVDVV 81
GenE        WAVVKVHYCCLCG-SDLWLYRGKWHGNRYPIVPGHEWAG-VVDSAPEGY--ESWVGRPVT 81
HygE        WALVRVAYNSICG-SDVSLYNNAWHGTAFPAVPGHEWSG-VVEQAPPG---QVAPGDRVV 80

AprE-Ste    GDLTVSCSRCAPCERGTPVLCENLQELGFNRDGACAEYMTIPAHNVYRLPDGLSLKAACQ 176
AprE-Shin   GDLTVSCSRCAPCERGTPVLCDNLQELGFNRDGACAEYMTIPVHNVYRLPDALDLKAACQ 141
KanE        GDLTCACGNCAACGRGTPVLCENLQELGFTKDGACAEYMTIPVDNLRPLPDALSLRSACQ 144
TobE_       GDLTCACGSCAACGRGTPVLCENLQELGFTRDGACAEYMTIPTGNLHVLPEGLSLRAACQ 141
NeoE        GDLTCSCGTCAHCAAGTPTLCEDLGELGFTRDGACAEYMTVPVANLRPLPDTLPLRTACQ 142
RibE        GDLTVSCGTCAHCTARTPTLCEDLGELGFTRDGACAEYMTIPVGNLRRLPDTLPLRAACQ 142
ParE        GDLTVACGSCRHCTVGKPTLCADLQELGFTRDGACAEYMTVPTGNLHRLPEGLSLREATQ 141
LivE        GDLTCSCGTCRHCAEGKPTLCGDLQELGFTRDGACAEYMTIPAANLHRLPPGTSLRAACQ 141
GenE        GDLIVGCQGCGPCRDGLPVMCENLIEIGFTVDGGCAGYVAVPITNLYLLPEGMDLAAASQ 141
HygE        ADLTLSCGQCRWCRRSQPVMCPGLREFGFTDPGGCADYVAVPAANLVRLPPDTDLLAATQ 140
```

```
AprE-Ste    VEPLAVALHAVATAGVEEGDRVAVLGAGGIGLLLMQAAALRGATVTTVSEPVAERRARAR 236
AprE-Shin   VEPLAVALHAVANAGIGEGDKVAVLGAGGIGLLLMQAAALRGATVTTVSEPVTERRERAR 201
KanE        VEPLAVALNAVSIAGVAPGDRVAVMGAGGIGLMLMQVARHLGGEVTVVSEPVAERRAVAG 204
TobE_       VEPVAVALHAVSTVGVEPGERVAVLGAGGIGLMLMQVARQRGGVITTVGEPVAERRAVAA 201
NeoE        VEPLAVALNAVDRLGVTPGEKVAVMGAGGIGLLLVQAVRLRGGTVTAVAEPVPERRAAAL 202
RibE        VEPLAVALNAVDRLGVTAGEKVAVMGAGGIGLLLAQAVRLRGGSVTAVAEPVPERRAAAL 202
ParE        VEPLAVALNAVDRLAVTAGEKVAITGAGGIGLLLAQAVRLRGAEVTVLAEPVTERRQAAH 201
LivE        VEPLAVALNAVDRLAVVPGEKVAITGAGGIGLLLALAVRQRGGEVTAFAEPVAERREAAA 201
GenE        TEPLAVALHAVDRINLRPAERVAVLGAGGIGQLILQSARATGATVTLATDLVAERRKIAE 201
HygE        AEPLAVSLHALSRVRLAPGETVAVLGCGGIGLTLLQAAQVAGAQVVLAVDPLPGRARTAG 200

AprE-Ste    DLGALTVTSAAPGELAALVE-RRPELTPDVVLEASGYPAAVQEAIEVVRPGGRIGLVGYR 295
AprE-Shin   ELGARTVTSAAPGELAALVE-SRPELVPDVVLEASGYPVAVQEAIEVVRPGGRIGLVGYR 260
KanE        QLGATELCSAEPGQLAELVA-RRPELTPDVVLEASGYPAALQEAIEVVRPGGRIGLIGYR 263
TobE_       QLGARTVTTGRPGELAELVA-KHPDLTPDVVLEASGYPVAVQEAIEVVRPGGRIGLVGYR 260
NeoE        ALGVPAAVGGDPGALVELTR-SDPAAVPDVVLEASGYPTAVQEAVEAVRPGGRVGLVGYR 261
RibE        ALGVPAAVSGEPGALVELTR-THPDAVPDVVLEASGYPTAVQESLEAVRPGGRVGLVGYR 261
ParE        ALGVPHTVGGDPGELVGFVE-KHPELTPDVVLEASGYPLAVQEAVEAVRSGGRIGLIGYR 260
LivE        ALGVSATVGGAPGDLVALVE-KHPELRPDVVLEASGYPSAVQEAIEAVRSGGRVGLIGYR 260
GenE        ESGAAAAVH--PSELPELTSYADK---VDVVFEASGDPESVVRALDLVRPGGRVCLVGYQ 256
HygE        LLGAGAALNT-PEEVAGWIADAGPDGLPDVVLEASGEPEAIRAATELVVPGGRVALVGYR 259

AprE-Ste    VEESGVMAP------HHATVKALTIRGSLGPGGRFAEAIEVLASG-----AINTDALLSH 344
AprE-Shin   VEETGVMAP------HHATVKALSIRGSLGPGGRFAEAIEVLASG-----AINTDALLSH 309
KanE        VEETGPMSP------QHIAVKALTLRGSLGPGGRFDDAVELLAKGD----DIAVEPLLSH 313
TobE_       VEEVGPMAT------HHVAVKALTIRGSLGPGGRFPEAIDLLARG-----EIEVEPLLSH 309
NeoE        IEEAAVMAP------HHIVLKVLTVRASMGPGTRFEEAVDVLASG-----AVDVDALLSH 310
RibE        IGETAVMAP------HHIVLKVLTIRASMGPGTRFEEAIEVLASG-----AVTVDALLSH 310
ParE        IEEAATMAP------HHIVVKVLSLQASMGPGDRFGEAIELLAAG-----AVDVDALLSH 309
LivE        VEEVGPMAP------QHIVIKMLTLQASMGPGTRFGEAIGLLSSG-----AVDVDALLSH 309
GenE        VGAEHALET------ARLPLSYASLVGVMGPGGKYREAVDLLANG-----AIDTQPILTD 305
HygE        VGRQVELES------ARWPLKLMETVGTMGPGRFMGAAAALVARG-----ALRTDLVVTD 308

AprE-Ste     EFGLDDHAAALDLALRRTDGNTRSYFHFA-- 373
AprE-Shin    EFGLDDYADALDLALRRTNGNTRSYFNLG-- 338
KanE         EFGLADYATALDLALSRTNGNVRSFFNLRD- 343
TobE_        EFALDDHARALDLALRRAEGNVRSFFNLRA- 339
NeoE         EFALDDYAKALDVALRRADGNTRSYFNLRA- 340
RibE         EFALDDYAKALDVALRRADSNTRSYFNLQA- 340
ParE         EFGLADHDRALDVALRRADGNTRSYFNLRA- 339
LivE         EFALEDHARALDVALRRADGNTRSFFNLRT- 339
GenE         IVTLDDYAPAIDRAINRTDGTVRVVFDLRNE 336
HygE         VLDLTAADEAFRLAGAPGTDTIRVAVRADGA 339
```

Fig. A.5 Alignment of BtrE homologous. Alignment was performed using Clustal W program (cf. Sect. 2.26). The numbers indicate the position within the corresponding proteins. The underlined C (cysteine) and H (histidine) residues given in bold are conserved amino acids that could play a role in the attachment to Zn^{++} ions.

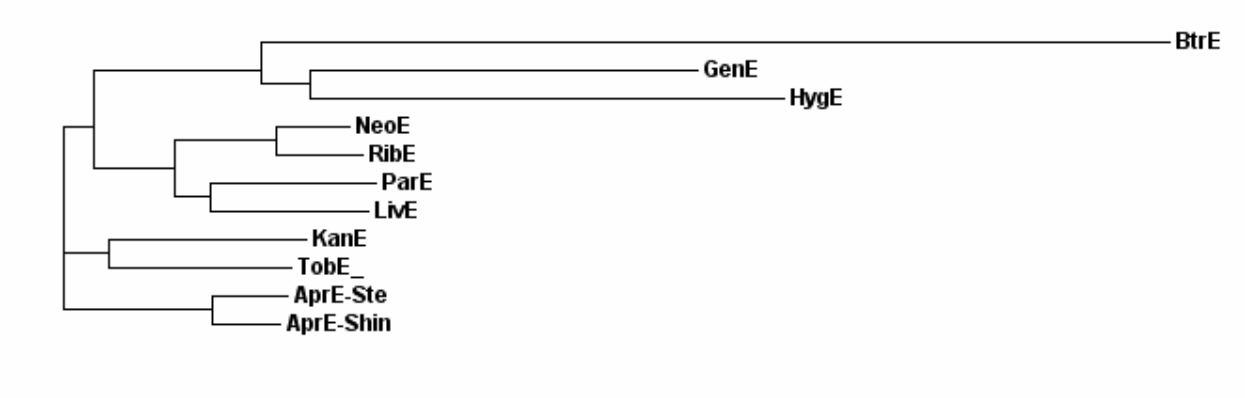

Fig. A.6 Phylogram of BtrE homologous. Phylogram was calculated using Clustal W program (cf. Sect. 2.26).

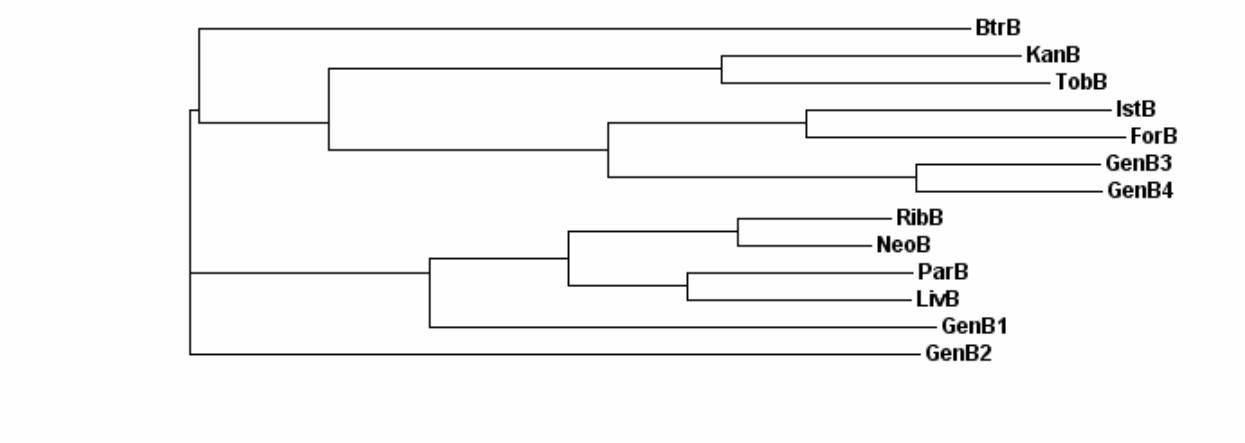

Fig. A.7 Phylogram of BtrB homologous. Phylogram was calculated using Clustal W program (cf. Sect. 2.26).

```
RibB        ------------------------------------------MTENS---SLLAEFP 12
NeoB        ------------------------------------------MTKNS---SLLAEFP 12
ParB        ------------------------------------------MTQNF---ATLAEIP 12
LivB        ------------------------------------------MTRNS---STLAEKP 12
GenB1       ------------------------------------------MTIDIGAGKLLAQEP 15
GenB2       ---------------------------------------------------MIIANAD 7
BtrB        ---------------------------------------MKQETVKSSEQLLSVLGTYI 20
KanB        ----------------------------------MSTHPVLDWSRSAEHLRRSHGVTT 24
TobB        ------------------------------MTRTNETIATITDWSRSAEHLRRSHEVTT 29
IstB        ---------------------------------------MTTTGGQSNLETIERARRVTA 21
ForB        ------------------------------------------MKNLETIDRARRVTA 15
GenB3       MAVADHRSSEPSWRAGRTARRRSRWHSWARVKSAREGSQDMDSANLTNRGLVERARRVTA 60
GenB4       ------------------------------------------MNYRELIERARRTTA 15
```

```
RibB        TCPRDVQDRPRVFTS-ASGAYVTDKSGFCWIDFDNARGSIVLGHGDPRVVEAVGRAATGA 71
NeoB        TCPRDEKDRPRVFTA-ASGAWLTDESGFRWIDFDNARGSILLGHGDPVVAEAVARAATGA 71
ParB        TCPRNAEGNPRVFVS-ASGAYVTDDAGKRWIDFDNTRGSVLLGHGDPEVAEAVGRAATGA 71
LivB        TCPKDADGNPRIFTA-ARGARLTDRAGKEWIDFDNARGSVVLGHADPEIAEAVARAASGA 71
GenB1       TCPRDADGRPRVFVE-GSGAYLTDPDGRRWIDFDNARGSVVLGHGDEEVAEAIARAARGR 74
GenB2       GCTPYEVARGVTIVR-GEGAYVYDAEGRGLIDLSNSFGSVMLGHQDPVVTEAVLKTVRSG 66
BtrB        DSPVDPFRKERVMFSRGSGAYLFDYDGGNYIDLMNGKGSIILGHNDPSVNAALRNFLEQD 80
KanB        DPRPDEDGHYPCVLTRGSGTRVYDLDGNAYLDLTGSFGSVLIGHAEPAVVRAVTDVLSE- 83
TobB        NPRPDEDGQYPCVLTRGSGARVHDMDGNAYLDLTCSFGTVLIGHAEREVTDAIRACLDE- 88
IstB        AEDYDIGTRFPSVFVSAEGSWMRDVEGNRVLDVTAASGSLLLGNRHPAVVEAVTKCVAEH 81
ForB        AEQYDIGTRFPSVFVKAQGSWMEDVEGRRVLDVTAASGALLLGNRHPAVVEAITRYIAEH 75
GenB3       AENYDIGTRFSAMIQSGEGAWLTDVEGNRYVDLTASSGTIILGHRNQAVTEAITRQIRDF 120
GenB4       AEEYDISGRYPSVIAHAEGAWMTDLSGNRYVDLTGADAAVILGYRHPAVNEAITRQIRDY 75

RibB        DGTATGWSPRVDAVLERLH--ALCGGEVVGLFRSGTAAVRAAGLAVREATGRPLLVSAGY 129
NeoB        DGTATGWSRRVDAVLERLH--ALCGGEVVGLFRSGTAAVRAAVLAVREATGRPLLLSAGY 129
ParB        LGTATGWSPLLDTVTSRLL--ELCGGEVVGLFRTGTSAVRAAVLAVREAVGAPLVLSSGY 129
LivB        AGTATGWSPLVDTVADRLL--ALCGGEVVGLFRTGTSAVRAAVLAVRESVGRPLVLSSGY 129
GenB1       SGVGTAWSPVLDSLLGQLQ--EVCGGDVVGLYRTGTAALRSVTCAVRDARDRSIVLSSGY 132
GenB2       VPAAASLDLQNHLAEQIAG--DLPGDQRVAFFKTGTAATRAAASAARQVTGKRLIASCGY 124
BtrB        REVVTGPSKPIIDLAERIKKDSALPDAKVSFYTTGTAACRAAVYAARDYSGKKIVLSSGY 140
KanB        GNLFYTGASPRRLALAERLLDWFPWSEQAIFYRTGSCAVSAAARLAQHATGRNRVLSSGY 143
TobB        GNLFYTGPSPRRLALAERLLDWFPWADQALFYRTGSCAVSSVARLSQHVTGRTAVLTSGY 148
IstB        GTVFASTISLPRLELAERLCERYPAGEKAVFCKTGSEATLAAIRLARAYTGRDLILTSGY 141
ForB        GTVFASTLSLPRIELAERLCERYPAGEKAVFSKSGSEATTAAIRMARAASGRDIIVTSGY 135
GenB3       GTAFASTLSVPRVELAERLCERYECAEKVVFHKTGSEGTAMAARLARAATGRELILSCGY 180
GenB4       GTTFASTLSVPRVELAERMCERYECAEKVVFHKTGTEGTAMAVRLARAATGRELVLSSGY 135

RibB        HGYDPMWYPSDAPLV--PNADGVVDFFFDLGLLAELLAA-PERVAAVVVSPDHMHLSAAW 186
NeoB        HGYDPMWYPSEAPLE--PNADGVVDFFFDLGLLRELLRA-PERVAAVVVSPDHMHLSPGW 186
ParB        HGYDPMWYPAKAPFE--PNADGIVDFFFDLDVLAGLLRDGRERVAAVVVSPDHMHLSPRW 187
LivB        HGYDPMWYPPHEPLE--PNADGIVDFFYDLDLLAGLLRD-RDRIAAVVISPDHMHLTPGW 186
GenB1       HGYDPMWH-CDEPFT--PNQHGIVEFLFDLDVLAEWLSR-PEQVAAVVISPDHMHLGERW 188
GenB2       HGYDLMWE-FTPPGQ--PNSEDVLHCYHLPELIDQVLDKHAHELAAVIIAPDYIHVSPEY 181
BtrB        HGWDPMWRQQGPLLE--PNEDGVIEFYFIPELLERALTAHKDQVALVIFSPDYTYLSAST 198
KanB        HGWHDWHLEAVPEAKPKTFESYATEFHNDLALYRSWLDRHGEEIAAVVVTPEPHRFDHAY 203
TobB        HGWHDWHLEAVPEAK--LFPSYATEFHDDLDVYRAYLDRHADEIAAVVVTPEPTRHPLEH 206
IstB        HGWHDWQLTYLNMGF--DPASRVVNFGYNETALARLLEQFGSEIGGVIVTPEPAWLGADY 199
ForB        HGWHDWHLSYLNIGY--NPGTRIACYGYNETALRRMLHEFAGEIAAVIVTPEPAWFDEAY 193
GenB3       HGWHEWQLAGETFGY--QQTTGVVGFGYNEKALAKMLEAFGNEVAGVLISPELLYFDVEF 238
GenB4       HGWHEWQMAGEEFGY--QQSTGVVGFGYNEKALAKMLEAFGEQVAGVIVSPEVLYFDLDH 193

RibB        YRELRQLCSAAGVVLIADEVKVGLRYSPG--LSTAELLNPDVWVVAKGMANGHAVSAVGG 244
NeoB        YRELRRLCSAAGVVLVADEVKVGLRYAPG--LSTAELLAPDVWVVAKGMANGHAVSAVGG 244
ParB        YERARALLAEAAVPLIVDEVKVGLRYGPG--LSTAGLLDADVWVTAKGVANGFPTAAVGG 245
LivB        YARARALLADAGVPLIVDEVKVGLRYGPG--LSTADLLDADVWVVAKGMANGYPTAAVGG 244
GenB1       YTEFTRLTKEADVPVIADEVKVGLRYRAG--LSTP-LLDPAVWIVAKCLANGSPVAAVGG 245
GenB2       IADLFERCERVGVVTIADEVKHGYRLRQGA-SVTEASVVADMYTYAKGISNGWPLSCVAG 240
BtrB        MERILGICRAHGVLVCCDDVKQGYRHRQGSSLELVTTEKADMYVFSKGLSNGHRISCVVS 258
KanB        YQELREVAKEHGCLFVVDEVKTGFRAGAGG-FSALAGIEPDAVTVSKGMANGHSISAVVG 262
TobB        YRTLRDLAAEAGCLFVLDEVKTGMRAGKGG-LSARAGLEPDAVTVSKGLANGHSISAVVG 265
IstB        YRRLSELCARHDVPFILDEVITGVRWGARG-LNGTGGVRADLITVSKGLSNGHALSAVLG 258
ForB        YRRVSQLCAEHGVYFIIDEVITAFRYGHRG-LNGTGEVPADAITMSKGLGNGHSISAVVG 252
GenB3       YQRMYALCARYDVPFMMDEVYTGFRAGPKG-VHGLG-VPADVVVVSKGLANGHSLAAVMG 296
GenB4       YRRMSALCARYDVPFMLDEVYTGFRAGPKG-VHGLG-VPADVVVLGKGLANGHSLAAVMG 251
```

```
RibB        PRALLRPLKEVSFTSFF--EPTILAAADAVLSRVATGEPQRALREAGDRFLCHARKALQD 302
NeoB        SRRLLKPLKEVSFTSFF--EPTILAAADAALARVATGEPQRAVREAGDRFLRHARKALDD 302
ParB        SRTLLKPLREVSYTSFF--EPTVLAAAERTLARVATGEPQRTVRETGDLFVEHARSALAA 303
LivB        SRTLLKPLREVSFTSFF--EPTVLAAAERTLARVATGAPQRTVREAGERFLTHARKSLAG 302
GenB1       DAHLLAALEDVSFTSYF--EPTAMAAATTTLRRMATGEPQQAIRAAGDRFIAHTRAAFAN 303
GenB2       DERFLKPLAEFVSTLTF--EAPSFAAASATLDRLAELDVQAQLAIDGARFVSEAAKMIST 298
BtrB        SDEIMAETKEHTYTAYY--QMLPILSSLETLKKMESGKGYDLIRSYGQTLTGNLKELFVQ 316
KanB        QRQLTQELSEAHVWSTYQNEQVGFAAALASLDFLERHDVAAVTRRTGEAVRQGVLQLFAE 322
TobB        SRRITEGLAEAHVWSTYQNEQVGYAAALSTVDFLLREDVAGVVERTGRTVERAFRSAFAE 325
IstB        KREIVDAYDKAGIAGTYTREVPPMAAALAVLDAIEDGSVHEHCEKMGRKLKDGMREILKK 318
ForB        RREILDAYDKAGVAGTYTREVPPMAAALAVLDVTEDGSVHEHTQRMGATLRDGMRDILAA 312
GenB3       RRDIIDAYDVSGIQGTYTREVPPMAAAMAVLDVLDTPGVYEHAEAMGRRLADGMREILTG 356
GenB4       RRDIIDAYDVSGIQGTYTREVPPMAAALAVFEVLDTPGVYEHAEAMGRRLADGMREILTG 311

RibB        AALP--VEIAGDGTFFQFVPASEELEYAFYAAANTEGLLFYAGDNQGVSAAFDDAVLTDA 360
NeoB        ASLP--VEIAGDGTFFQFVPATEELEEALYGAANAEGLLFYAGDNQGVSAAFDEAVLGEA 360
ParB        ASLP--LEVAGNGTFFQFVPATREVKKAFYKATEEEGLLFYRNDNQAVSAAFGPDVLDDA 361
LivB        AGLP--VELAGDGSFFQFVAATRELEKALWWATEDEGLLFYRGDNQAVSAAFGPDVLDDA 360
GenB1       AGVP--IDLAGNGNLFQFVCADDEVADAFHAAAAAEGLLFFEGDNQTPSAAFTDEVVEDA 361
GenB2       RDLP--IEMAGTGAAFQFVCA-EEVEEVLLPHALAEGLILEPSDQQYPSACFRGEVVDDA 355
BtrB        SSLP--IEVNGS-SIFQLVFGDEELEEAFYREAFIQGLILFEGDNQSLSLCMDKDVQVDL 373
KanB        HGLPVGAPGWGPMFELDFDAADEGLAERLEAALLRHGIFC-DTGDDFNMMFHTAEHTDEL 381
TobB        RGLPVEVHGWGPMFDLDFSAAEEDLPERLQLALLRHGVFC-DVGDDFNLMYRMADHLDEL 384
IstB        VGIPAYIT--GPDMMFDVVVESESLSWDIYRAAFDYGAYFEDSGTHMVTAAYGDEEVEHA 376
ForB        VGIPAFVN--GPPMMFDVVVPSEQLSWDIYRAAYDHGAYFEDSGTHMVTAAFGQAEVDHA 370
GenB3       EGIPNWVG--GPALMFDTVLPNDDLGWEIYKTAHDFGVYFEDSGTQLVTTAFDEAAVDHA 414
GenB4       EGIPNWVG--GPALMFDVVLPNDDLGWEIYKTAHDFGVYFEDSGTQLVTAAFDEAAVDHA 369

RibB        EQRFTRVCDRLSAY---AGGEPVDEAARYRVAWNVMDGLREAERDRAATDGMLARLLDD- 416
NeoB        ERRFARVCERLAPY---AGGEPVGDAARYRVAWNVMDGLRQAPRDREETTGLLARLLDD- 416
ParB        RARFSRVCDRLAPF---AAAGPVGEEARYQAAWSVMDGLREAERDARETREWVDRFLDD- 417
LivB        EARFTRVCDRLAPY---ASDAPVSEEARYLAAWNVIDGLRDAARDDRTTREWISRLLDD- 416
GenB1       CGRIDRVSAALTGR---FTDRELTEESWYASAWGAMDGLADRPRTREETTAIVERLWED- 417
GenB2       LERLDRALTTMAAARPDLVGREVTQLDRVNAAFCQMDGLPGRP-DGWSLDQCVEYVTAQL 414
BtrB        IRRFANVTDVLSEQFKHLRGKEVTTEQTFRTAWNMIDGASDLLPYEKQLKLLDNLIGGG- 432
KanB        LERFAAALGDL---------------------------------------------------- 392
TobB        LERVTAAIASV---------------------------------------------------- 395
IstB        LHAFEMGAQRVVA----TREPDFGELPEDRRMAFPLEAFGGNIRDDETVLKRIEDIVRQV 432
ForB        LEAFEKGAREIAR----ITSFDFDALPESRLHQFASEAFGGALHDDEAVLRRIDDTVHAI 426
GenB3       LTAFRKATRQVIADRPDIAPTSGGELTEERKLDFAEEAFGGLLRDDERTNALIDETIEKV 474
GenB4       LTAFRKATRQVVADRPDIAPTSGGELTEERKLDFAEEAFGGLLRDDERTNALDRRDHREG 429

RibB        ----------------
NeoB        ----------------
ParB        ----------------
LivB        ----------------
GenB1       ----------------
GenB2       ----------------
BtrB        ----------------
KanB        ----------------
TobB        ----------------
IstB        EQRDRTLGGTLDPACG 448
ForB        ANRDPALARLLDPSCG 442
GenB3       VNRDRSIKPVLIPAQN 490
GenB4       GQPGPQHQAVLFPAQN 445
```

Fig. A.8 Alignment of BtrB homologous. Alignment was performed using Clustal W program (cf. Sect. 2.26). The numbers indicate the position within the corresponding proteins.

```
LivY      MGAQATESVMSGEKRYVLITGANRGLGLHSARLLAGRGWSVILACRKPEAAAPALEKLR- 59
AprD3     ----------MEQRYVLVTGASRGLGRRSAEELAGRGWTVVLACRDVADVASLLATVRG 49
GenD3     MDTAPANTHVRPRRGHVLVTGGNRGLGLATATLLVDDGWSVLLGCRDERRGAAAAEALRR 60
                       ** **  ****   *  *   ** * * **     *      *

LivY      -AYDVTSVRLDVTDPESVSAAVRTVRG-TAPKLHALVNNAGIFEH-AEEHLSDAPDSVSR 116
AprD3     RGARAHAVPMDVTDPDSVAAAVDAVRE-VGGRLHALVNNAGVFRH-AEERFPGLTPGDAL 107
GenD3     RGGRAAHVPLEVTSPASIAAAVDLVADRTGGRLAGLVNNAGVFLDERDADLESVTAEAVH 120
                 **  ** * * ****  *       *  ****** *

LivY      DLLLTNAWGPLMVTRHFLPLLRAAGGASVVNVTSQDADPENLDGTFTGYCMSKAALNVMT 176
AprD3     DILLTNTYGPLVVTRAFLPLLRAAGGAAVVNVTSRDADEDTFDGEFTCYRASKAALNAMT 167
GenD3     ELLAVNAVGPLLVTRAMVPLLRAAAGAAVVNVTADDADPATADGEATGYRMSKAALNIMT 180
            *  *  *** ***   ****** ** *****  ***    **  * *  ****** **

LivY      ANLAIALRADRVVVNGVDPGWIPTDMGGPEAPDAVADAAETVAAAVALAGTARSGDILRA 236
AprD3     RNLAVALRPDRIVVNAVDPGWIPTDMGGPEAPDSLDAAVTAVVDAVELAGSDRTGVLLRA 227
GenD3     VNLAVALREQDVVVNAVDPGWIPTDMGGPEAPDDTAAAARLVAWAVTAAREHGTTGQVLS 240
           *** ***    *** *****************    *   *  **  *

LivY      VR------------------------------ 238
AprD3     GRDPAPPVAAPSEPNQRHQPDLPPPPVPRWSAS 260
GenD3     VRQPPDGLRADPVTVGAPPTVVPPHGQPQ---- 269
           *
```

Fig. A.9 Alignment of LivY homologous (putative 3',4'-dehydratase). Alignment was performed using Clustal W program (cf. Sect. 2.26). ***** = identical aa residues. The numbers indicate the position within the corresponding proteins.

```
LivW          MRHMRLGKVMLVSPKTSFGRDLQRTYAGGLGTVCKDEDFLLPPLDLMRLAGVLREAADDI 60
AprD4         MRRMRLGTVLLVSPKTSFGRDLQRTYAGGLGTVCKDEDFLLPPLDLMRLAGVLREDADDI 60
              ** **** * ********************************************** ****

LivW          SVVDEEVTGTV-SASPGSIVICQVSLPSLHEDAERIAAFRAQGARCYAYAGMRSPAQWRV 119
AprD4         AIADEEVTGVVPSVEPGTTVICHVSLPSLLEDAERLATFRAQGARCYAYTSIRSPAQWRT 120
                 ****** * *  **  *** ****** ***** * ***********   *******

LivW          LFEQGGCEGILLPESISFARAALEGDTTVPGCVTPENLLEPRHHQPQFGDLAAEPLPARD 179
AprD4         LFERGGCEGVLLPESISFARAALAGDHTVPGLVTPDSLLDPRHHQPAFGDLAAEPLPARD 180
              *** ***** ************* ** **** ***  ** ****** *************


LivW          LVDHTPYMFPPIALTGITSINGSFGCPYPCRFYCPYPLSEGRKIRTYPVERIVAEFRQCA 239
AprD4         LVDHTPYMFPPIARTGITSINGSFGCPYPCRFYCPYPLSEGRKIRTYPVERIVAEFRQCA 240
              ************* **********************************************

LivW          ELGITAAVFRDPVFSFHRGRTLELCRAIEAAGTHVPWWCETRIDRLDEEVVAALVAAGCV 299
AprD4         ELGITAAVFRDPVFSFHRDRTLELCQALKAADTGVPWWCETRIDRLDEEVVAALVDAGCV 300
              ****************** ****** *  ** * ********************** ****


LivW          GVEVGVESGDPEMQAKAVRKRLDLDAVRRFHALAREMNLELVFLFLVGLPDETRMSIRRT 359
AprD4         GVEVGVESGDPDMQATAVRKRLNLDTVRKFHAVARKAGLKLVFLFLIGLPRETRMSIRRT 360
              *********** *** ****** ** ** *** **   * ****** *** *********


LivW          FDFILELGLADTEFNLSIITPYPGTELHRIARDNGWIDGEQNAFTSHNAVMHTDRLSVAD 419
AprD4         FDFILELGLADTEFNLSVITPYPGTELHQIAVDKGWIDGSQNAFTSHNAVMHTDELSIGD 420
              ***************** ********** ** * ***** ************** **  *
```

```
LivW          LEQATSFVDELHAVCRKTATPQDRAEFQRRLHAWSEQDA 458
AprD4         LERASRFVDELHAVCK--AGPAERAEFQARVHAWSTGDA 457
              ** *  *********   * *  ***** * ****  **
```

Fig. A.10 Alignment of LivW and AprD4 proteins (putative 3',4'-dehydrogenase). Alignment was performed using Clustal W program (cf. Sect. 2.26). ***** = identical aa residues conserved in all proteins. The numbers indicate the position within the corresponding proteins.

```
NeoI          MA-GSGSAPEPSERPLLRRFDQRLNELIAALYPDE-RRRPGYARLAKEIRETTGGAISGT 58
RibI          MA-GSGSAPEPSRRPLLRKFDQRLNELIAALYPDE-RKRPGYARLAKEIRETTGGAISGT 58
ParI          MS-DRTSATDQEGRPLLRAFGQRLNELIAALHPDE-RKRPGYARLAKEIRETTGGAISGT 58
LivI          MG-DNASASDQAGRPLVRKFDQRLNELIAALHPDE-RKRPGYARLAKEISESTGGTISGT 58
Cinorf12      MS-DSRPAPDRNDRPLLRQFDQRLSELIATTAGAEGNKRPGYARLAKEIRDTTGRTISGT 59
KanI          MR-DGAAERSESGRLLLDKLDQRIKDLICAIYPDK-GTRPGYARLAQDIRETTGGTISGT 58
HygI          MSSEHDTAKPTPDRPVPGGFDERLARLITIYYAGE-RKRPGYARLAREISEKTGRAISGT 59
              *           *       **   **          ********      *** ****

NeoI          YLWELATGKKRNVTLEQLDVLAEFFGVPPEYFLNEETAARVNSQLKLATALRDSRVRNLA 118
RibI          YLWELATGKKRNVTLEQLDVLAEFFGVPPEYFLNEETADRVNAQLKLATALRDARVRNLA 118
ParI          YLWELATGKKRNVTLEQLDVLAEYFGVPPEYFLSDDVAERVNSQLRLAVALRDTRVRNLA 118
LivI          YLWELATGKKRNVTLEQLDVLAEYFGVPPEYFLSEEVAERVSSQLKLATALRDARVRNLA 118
Cinorf12      YLWELATGKKRNVTLEQLDVLAEFFGVPPEYFLDDETGRRIDDRRRLAIALRDAKVRNLA 119
KanI          YLWELATGKKRNVTLEQLGVLADYFGVPPEYFLNDEVSERVNAQLALATALRDNRIRNLA 118
HygI          YLWELATGRKHNVTLHQVGVLADFFGVPTDYFVDDEVAARVDAQLQLAAALRDTRVRNLA 119
              ****************  ***  ****  **       *     ** ****   ****

NeoI          LRAEGLSPESLDALLIMVQEARRVQNLASVDDQDHTDRTDHTAAPDGGPAGTSGSTA 175
RibI          LRAEGLSAESLDALLIMVNEARKVQNLSSVGDQ---------AAPDGGTTGTSGRTT 166
ParI          LRAEGLSPAALDALLTIVDEARKVQNLAPPE-----------AGGDGTQEPPAAAP- 163
LivI          LRADGLSPATLDALLVMVNEARKVQNLASVDDL---------ADGSGTANGSVAGSG 166
Cinorf12      LRADGLSPDCLDALIAMVNEARKTQNLSSIDD-----------DDDTATTLTSSG-- 163
KanI          LRAEGLSPSTLDALLVMVNEARKIQNLSPPAD----------EDGRGPQHTE--- 160
HygI          LRADGLSPATLDALLTMVNEARKVQNLPVDDS----------AGPPPNEGGPGA-  163
              *v* ***   *****    **** ***
```

Fig. A.11 Alignment of NeoI homologous. Underlined aa represent the conserved HTH DNA binding motif. The numbers indicate the position within the corresponding proteins. Alignment was performed using Clustal W program (cf. Sect. 2.26). ***** = identical aa residues conserved in all proteins

```
NeoH          ---------------------MSYR-IARLRRRCRPLLDQLVLPRPFSLEPLCAQLSARR 38
RibH          ---------------------MSYR-IARLRRRCRPLLGQLALPHPFSIESLCAHVSANR 38
ParH          --------MRCPSQKKWQFSHMLYG-YSTLRRRCRAKLTDITLPTPFSVEEFCRSISEQR 51
KanH          MHVRRTAYYRCPAQG-LRKEHMSYI-YVKLRQRCQAIVDGLSLPHPFTVESFCTRLAAQR 58
LivH          ---------------------MPYR-YVQLRRRCQQKLDQFDLPRPFSMDALCESVSRQR 38
Cinorf13      MPRRTYVAYRRCLRKGQREPEMPYTSDAALRRRCRALLARVSLPEPFSVEVLCRHLGEQR 60
HygH          ---------------------MSLCDTRLRRRRCEAAMAGLRLPQPFSVPELCRQVAERR 39

NeoH          GRP**LHLHPLPIQAAAAGTCGLWLA**TDTDDHIFYEQRTARPHQEHIVLHEIGHMLFDHHTV 98
RibH          RRP**LHLHPLPLQAAVAGACGLWVA**TETDDHIFYEQRTARPHQEHIVLHEIGHMLFNHHAV 98
ParH          QRP**LHLHALPTHVAKMAACGLWLA**TDTDDHIFFEARTAPLHQEHIVLHEIGHLLFDHQML 111
KanH          QRP**LHLHALPRQAAQAGACGLWLA**TDIDDHIFFEQRTARVHQDHIILHEIGHMLFGHQSL 118
LivH          GRP**LYLHPLPDQAATAGACGLWLA**TATDDHIFFEQHTVRLHQEHIVLHEIGHMLFDHHAV 98
Cinorf13      GRP**IHLHPLPEQAALAGACGLWLA**TATDDHIFHERHTVRPHQEHIVLHEIGHMLFDHHSL 120
HygH          GRP**LHLHPLPAAAAAAGTCGLWLG**TEAADHIFYEQRTGRLHQEHIVLHELAHILLDHHRI 99
```

```
NeoH          D----AGPAGLPGRLFGDLSPRLVRRVLGRTDYSTRQEQEAEMLATLIRTGDTAHRCLPP 154
RibH          TDIATTPTAGTPGGLLGDLSPRLVRRILTRTDYSSHQEQEAEMLATLIRTGETG-----P 153
ParH          A---DGN-SGGISSVLADLSPRLIQRMLARTNYSTRQEQEAEMLASLIRTHATGP----- 162
KanH          E---LGEGWGDAPQLFGDLSPQLIQRLLARTDYTTRREQEAEMLASLIRTRGEQP----- 170
LivH          D---ARTGSAILPDLFLDLDPRLVRRLLARTNYSTRQEQEAEMLASLIRAGDHGP----- 150
Cinorf13      A---PAG--GPAGALLADLDPRLIRRLLARTNYSTRQEREAEMLASLIRTSVRAG----- 170
HygH          V-----PGGAGTERLLPDLDRQLVRRLMPRADYSTPQEREAETLATMLRLGGAAP----- 149

NeoH          DPYGKLHEALGMAGDHDLRCLPPDPYGKLHEALGMAGDHDL 195
RibH          D-----------------RCPPADPYGKLHAALGVSTDHAL 177
ParH          ----------------GRGPQRCATTDRADRALGITHD--- 184
KanH          ----------------AGERHADDVLHYLGEALGVGATRGG 195
LivH          ----------------GGE--RDGVLRNLGAALGVSGR--- 170
Cinorf13      ----------------TGER-PPGALGRLQAALGVVGSHGR 194
HygH          ------------------ARRPDQVLDRLRDATGVWP---- 168
```

Fig. A.12 Alignment of NeoH homologous. Alignment was performed using Clustal W program (cf. Sect. 2.26). Underlined aa residues represent the hydrophilic and histidine-rich central domain; aa in bold represent a possible single trans-memberane domain. The numbers indicate the position within the corresponding proteins.

```
NeoG          MTSDLPGNLGRVLQTAGVLLLWAAVLLRARPAARHPGQRGLWLAVACAAAALTLHLEAVS 60
RibG          MRCDLPG-------HCSLAVLWAALLLRAGPAVRNPHQRGLWLAVASAAVAMTLHVEPVS 53
LivG          -VTDAVAVLGPWYLVAGVSCLWSAVLLRCRSAVRSPHQRGLWLAVATAAVAMTMDLPAVD 59
KanG          -----------VAQYVSAAAMGAVVLLRAPAAARSPAQRGLWLAVATAAVAMALQLPVVI 49
HygG          -MAAHVPAWLVWYQPLTTGFLVAAVLSSLPGAVRRPRQRLLWGGVALAAVAMVLDLPPFR 59
Cinorf14      ------------------------MVLRCPYAVRHPAQRGLWLAVATAALAMTLTTSIGS 36
ParG          MTDRGPHGLFVQLEWLGLALMWVVTLLRAPVAVRTPQQRPLWLAVGLASVAMTLQLPQVS 60

NeoG          AYAAALTGSPRAVGLVKNQAGIVSAGAILHFAVHTTGSRRRTGAVLATTAAVMAALLALA 120
RibG          AYAAGLTGSPRAVALAKNQAGVVSVAAVLHFAVYATRGRR-TGTVLAAMAAVMAALLALA 112
LivG          SFLQARTGLGHLHHLARDLFGLLSAGAVLRF-VAASGNCRYGRGVYVAVAALTAALITLD 118
KanG          RAALDLTGSVHEVGLARNLFGVLSAGCVLYFVTAAAQHRWWRWGLVTSVALVVLALVLLD 109
HygG          -HASRTSGAPWALPLARNAAGVLSAGAVLLFVAQAADKPRLRTAVRAG-APIALTLLSAL 117
Cinorf14      VVPEAALG------LAGNLTGMVSAGAVLGFVITIMGGRRLHTWACGTVAATALALTVLG 90
ParG          SVATAVLDP-HTVFLAKEIVGLLAATAVFEFISAIVGLRRTQLAILAAAVLTLAALVTLC 119

NeoG          GLGSPHAPGHGPYPTLA-CPRCAAYWLLLVTVHVAACALCARVCWRYGRSGPHRSVNLGL 179
RibG          ALGAAPTSGQGPPFTLA-RPPCVAYWILLITAHVVACAACVRVCWTYARRGTNRSVNLSL 171
LivG          LLAAPHG-EHTAHGARS-TTETTAYWLILIAMHLFAGAMCFRLCWRYAGRADSRSLRLGL 176
KanG          QVREQHDNPGGP-------SSLSAYWAILMGSHIVANTVCVHVCLWQGCRASAPSLRLSL 162
HygG          VLAERTRAPTPVDGVLR-VTHSALYWLVLGGIHLAANAACATLCLRHGRRTADRGLGLSL 176
Cinorf14      VTSRAHLSYGTIAEIP---PTATAYRLLLIGTHLAVNAACIAVCWRYGRGPSRSPLALGL 147
ParG          VLSPPHTRHVVPSGPGQEASPDTYYWVVLLGYHLVANTACVAACWACGRQVKNTPTRAGL 179

NeoG          TLLGWGSASAGLYWLAHYYLMAVEG-RPGG--VLRLVLSAHAVLCAAVLMVPTALQLRRA 236
RibG          ALFGWGTALAGLFWLAHYVLLATGS-RPGT--VLCLLISLHAVLVGAALLVPSVLQLRQA 228
LivG          WLFGWGTASAGLFWAGHFLLFCAEAQEPVP--WLRLLMVVQGVLWAAAILLPVAVASAGA 234
KanG          WIFGAGTALVGIYWCVALGRLIIGE-PPLR--NLSLLMSLHGFLRAAALLVPLWGSMRSQ 219
HygG          RLFGAGAVCAGIYWVGRLAMALTPASLPGEGVALRLVIGLHAVLRGAALLVPLLGAARRA 236
Cinorf14      RLFGIGTVLAELYWLRLFAGLFTTS---------------------------------- 172
ParG          LLFGTGIALSAVLMALSLVHVLSRDATIPH--LFPAISGAEAAFMAAGAALPLLRRPWRT 237
```

```
NeoG        AGHARTVWRIWPLWRDLVEAVPHVALAGTRCRPLTLLRPQLPWRQLAYRKVIEVRDALLV 296
RibG        AGHARTIWRIWPLWRDLVDAVPHVALSATRSRLLVLLHPHLSWRLIAYRKVIEIRDAILA 288
LivG        VGHIGTVWRLWPLWRHLVDAAPHVALMKPRNRLLELLRPRHSWRLLGYRKVIETRDAILG 294
KanG        PSQIRTLWVLWPLWRTLVQAVPHVALHRRRSRIVEVVWPRAPRRLAVYRRMIEIRDAILS 279
HygG        AADARALWRLWPLWHDLVRASPQVVLDGYRSRCAELLTPPSHWRFLAYRKVIETRDALLV 296
Cinorf14    ------------------------------------------------------DALLR 177
ParG        ARDMAVMYRLHPLWRTLVRVTPDVTLTDPRSRAVDLLSAVRHSDLRLYRRAIEIRDGLLA 297


NeoG        LSHYTDPAVSRAARAHVARWGIPAERADAHVTACVLRGARAARLAGADPDPAAEVLAADR 356
RibG        LSHYTDPAVSRSARAHVARWGVPADRVEAHVTACALSRAHAAKFAGAEPDPAAEVLTTGQ 348
LivG        LRGYADPTVPGLARRYTATAGLADGEADLVALACELSEARRAKLAGLPRRTAPDGWFGG- 353
KanG        LHGYVHPALPGAVKSRVEQLGLHGRSADAMTLACLLHVALRARRTGAAKDFDAS-LSHGW 338
HygG        LRDRAGPAALDCTQRRPADP---LGSEEAVEPACLFGGG-CAQQPRSAPPAAPGAGPMGD 352
Cinorf14    Y------------------------------------------------------------ 178
ParG        LEDYSPEGTLARAREHVRAARVADASVEAAVTGCWASAAVANKRSGAATSGRPSGLPQS- 356


NeoG        EACGLAAETAVLLRLTEAYLSPCARAFDP--AGTSGTGPAGDAGRTGSAGAPPPRTRMSE 414
RibG        EAGDLAAETAFLLRLTEAYLSPCVRDFDT--AAAPAHATAGGDR--------------- 390
LivG        DSRDLDDETAFLLRLSRAYHSPHVREFHLYMAVQDAAGHAAQDG----TGRPPTRSAS-- 407
KanG        DSVNISGEESFLLELARAFRTPVARSLAGQLGGRTGGLTEVGDG--------------- 382
HygG        GGTAVTAEVAFLHHAS--HHCPALRLFRSDAGPGEDGSRRTA----------------- 392
Cinorf14     ------------------------------------------------------------
ParG        GGGDMAEELDFLLQLARFFRSPLAASFLSAEEQRHRAAPTGQAS--------------- 400


NeoG        TPTGPDERRGTASPPLP 431
RibG        ----------------- 
LivG        ----------------- 
KanG        ----------------- 
HygG        ----------------- 
Cinorf14    ----------------- 
ParG        ----------------- 
```

Fig. A.13 Alignment of NeoG homologous. Alignment was performed using Clustal W program (cf. Sect. 2.26). The Cinorf14 protein was incomplete**.** The numbers indicate the position within the corresponding proteins

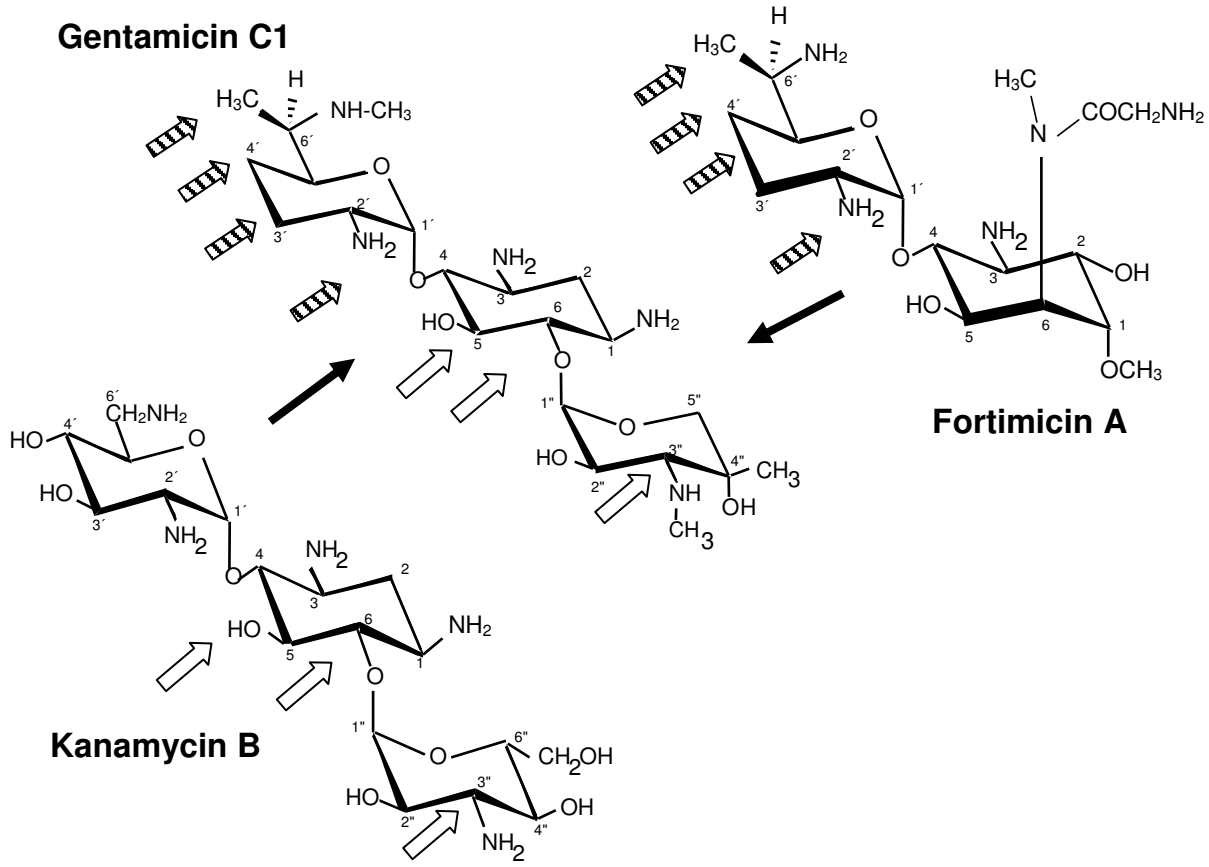

Fig. A.14 Structural comparison of gentamicin C1, kanamycin B and fortimicin A with regard to their biosynthetic origin and evolution of a hybrid pathway.

indicates similar biosynthetic origin between gentamicin C1 and fortimicin A

indicates similar biosynthetic origin between gentamicin C1 and kanamycin B

indicates that gentamicin C1 biosynthesis could be originally derived from a merger of fortimicin and kanamycin biosynthetic gene clusters.