

# Plasmids for Therapy and Vaccination

*Edited by M. Schleef*

 **WILEY-VCH**

Weinheim – New York – Chichester – Brisbane – Singapore – Toronto

This Page Intentionally Left Blank

**Plasmids for Therapy  
and Vaccination**

*Edited by M. Schleef*

# Plasmids for Therapy and Vaccination

*Edited by M. Schleef*

 **WILEY-VCH**

Weinheim – New York – Chichester – Brisbane – Singapore – Toronto

**Editor:**

**Dr. Martin Schleef**

Plasmid Factory GmbH & Co. KG  
Meisenstrasse 96  
D-33607 Bielefeld  
Germany

■ This book was carefully produced. Nevertheless, authors, editor, and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items inadvertently be inaccurate.

**Library of Congress Card No:** applied for  
British Library Cataloguing-in-Publication  
Data: A catalogue record for this book is  
available from the British Library

**Die Deutsche Bibliothek –  
CIP-Cataloguing-in-Publication Data**  
A catalogue record for this publication is  
available from Die Deutsche Bibliothek

© WILEY-VCH Verlag GmbH,  
D-69469 Weinheim  
(Federal Republic of Germany), 2001  
All rights reserved (including those of  
translation into other languages). No part of  
this book may be reproduced in any form –  
nor transmitted or translated into machine  
language without written permission from  
the publishers. Registered names,  
trademarks, etc. used in this book, even  
when not specifically marked as such, are  
not to be considered unprotected by law.

Printed in the Federal Republic of Germany  
Printed on acid-free paper.

**Composition** Hagedorn Kommunikation,  
Viernheim

**Printing** betz-druck, Darmstadt

**Bookbinding** Wilh. Osswald + Co.,  
Neustadt

**ISBN** 3-527-30269-7

This Page Intentionally Left Blank

## Preface

Gene therapy and vaccination with nucleic acids is one of the most striking innovations in medical and veterinary sciences. The treatment of a disease on the level of genes rather than on the phenotype level is a promising option to obtain enhanced therapeutics.

The pharmaceutical application of genetic material — modified so as to become the so-called active pharmaceutical ingredient (API) — requires developments in biotechnology and pharmacy to obtain systems for the transfer and expression of the API at the right place and time, resulting in the appropriate effects on the organism or cell. The vector for such gene transfer may be as simple as a short piece of DNA. For years now geneticists have been working with plasmids, and the recent developments in vector design and manufacturing point to potent therapeutics and vaccines.

In chapters 1 and 2 the background of plasmids is summarized and their structures are presented, since these well-known molecules still have had an unknown potency decades after their discovery. Their different size and structure turned out to be of importance for their function, and the characterization of an API made from DNA required a complete set of quality assurance in manufacturing and quality control. These aspects are explained in chapters 11 and 12.

Detailed examples of clinical applications are presented in chapters 4-6, providing an overview of the wide range of preventive and medical applications using plasmid DNA. In chapters 3 and 5, recent overviews on DNA vaccination are presented which should help to oversee the rapid development and published literature in the application of nucleic acid vaccines. Regulatory and quality assurance aspects of such new drugs are considered in chapter 13.

Chapters 7-9 describe modified vector systems based on plasmids, as well as the potency of genomic research and vector design by informatics. The link between genomics and the function of genomic information necessarily requires nucleic acids.

One example of veterinary health care is presented in chapter 10. The development of veterinary vaccines still requires some effort and the recent worldwide discussion on BSE in (at least) cattle makes obvious that health care in animals is also health care for humans. However, the treatment of animals grown for food production raise further questions on the aspect of genetically modified food.

Plasmid production on an industrial scale is necessarily linked to the question of the expected market size. Chapter 14 guides through the history of gene drug development and the expectations in today's and future pharmaceutical markets.

The development and preclinical testing of therapeutic or preventive plasmid pharmaceuticals is right at its beginning. The vision of individualized medical treatment might become reality with this type of drug system. The option to have major improvement in comparison to conventional drugs attracts research, industry and politics and is a challenge for all disciplines involved – from genomics to clinical application.

Finally, I wish to thank Karin Dembowsky from Wiley-VCH for her continuous collaboration with this project and all authors who contributed to this book, to make it what I hope it will be: A recent overview on the field and a guide through an area of useful innovation.

Bielefeld, January 2001

Martin Schleef



## Contents

**Preface** V

**List of Contributors** XIV

<b>1</b>	<b>The Biology of Plasmids</b>	<b>1</b>
1	Introduction: What are plasmids?	1
2	General properties of plasmids	2
2.1	Plasmid replication and its control	3
2.2	The molecular basis of incompatibility	6
2.3	Plasmid inheritance	7
2.4	Mechanisms of plasmid spread	8
2.4.1	Conjugation in gram-negative bacteria	9
2.4.2	Conjugation in gram-positive bacteria	11
3	Plasmid-encoded phenotypes	11
3.1	Bacteriocin production and resistance	12
3.2	The Ti plasmids	12
3.3	Heavy metal resistance	14
3.4	Other phenotypical traits	16
4	The clinical importance of plasmids	18
4.1	The spread of antibiotic resistance and the evolution of multiple antibiotic resistance	18
4.2	Transfer of antibiotic resistance genes	19
4.3	Mechanisms of antibiotic resistance	21
4.4	Bacterial virulence genes	23
5	Plasmid cloning vectors	24
6	Perspectives	27
	References	28
<b>2</b>	<b>Structures of Plasmid DNA</b>	<b>29</b>
1	Introduction	29
2	Topological structures of plasmids	30
3	Supercoiling of DNA	32
4	DNA intercalating dyes	32

5	Analysis of plasmid structures	33
5.1	Electron microscopy (EM)	33
5.2	Agarose gel electrophoresis (AGE)	35
5.3	Capillary gel electrophoresis (CGE)	38
5.4	Analytical chromatography	41
6	Conclusion	41
	References	42
<b>3</b>	<b>Genetic Vaccination with Plasmid Vectors</b>	<b>45</b>
1	Introduction	45
2	Vector design	45
2.1	Plasmid DNA	45
2.2	Construction of simple transcription units	47
2.3	Construction of complex transcription units	47
3	Strategies for DNA delivery	49
4	Priming humoral and cellular immune responses by DNA vaccines	50
5	Experimental strategies facilitated by DNA vaccination	53
6	Unique advantages of DNA vaccination	54
7	DNA vaccines in preclinical animal models	56
7.1	DNA vaccines to control infectious diseases	56
7.2	Therapeutic tumor vaccines	57
7.3	Autoimmune disease	57
7.4	Treatment of allergy by therapeutic DNA vaccination	57
8	Proposed clinical applications of DNA vaccines	58
9	Risks of nucleic acid vaccination	59
10	Future perspectives	59
	References	61
<b>4</b>	<b>A Liposomal iNOS-Gene Therapy Approach to Prevent Neointimal Lesion Formation in Porcine Femoral Arteries</b>	<b>75</b>
1	Introduction	75
2	Results and discussion	77
2.1	Therapeutic plasmid	77
2.2	The gene therapy product has a clinically acceptable format	78
2.3	Efficient gene transfer was established in a minipig femoral artery injury model	79
2.4	Transfection efficiency is dose dependent	81
2.5	Non-viral iNOS gene transfer efficiently inhibits neointimal lesion formation	82
3	Summary and perspectives	84
	References	85

<b>5</b>	<b>Immunotherapy of Chronic Hepatitis B by pCMV-S2.S DNA Vaccine</b>	<b>87</b>
1	Introduction	87
1.1	Hepatitis B: the disease	87
1.2	Hepatitis B: treatments	87
1.3	Hepatitis B: immune response to infection	88
1.4	What are DNA vaccines?	88
1.5	Which DNA vaccines for hepatitis B?	89
2	DNA vaccines for the prevention of hepatitis B	90
2.1	The mouse model	90
2.1.1	Humoral response	90
2.1.2	Cell-mediated response	92
2.1.3	Mechanisms of DNA-induced immune response to HBsAg	93
2.1.4	The primate model	94
2.1.5	DNA-based vaccination of chimpanzees against HBV	95
2.1.6	Neonatal immunization	96
3	DNA-based vaccination for chronic HBV infections	96
3.1	HBsAg transgenic mice as a model for HBV chronic carriers	96
4	Clinical trials of DNA vaccines	98
	References	99
<b>6</b>	<b>pSG.MEPfTRAP – A First Generation Malaria DNA Vaccine Vector</b>	<b>103</b>
1	Parasite life cycle and impact of malaria	103
2	Concept of vaccination against malaria	104
3	First-generation plasmid: pSG.MEPfTRAP	106
3.1	Vector backbone	106
3.2	Insert	106
3.3	Production and formulation	111
3.4	Preclinical testing of pSG.MEPfTRAP	112
3.4.1	Toxicity studies	113
3.4.2	Biodistribution	113
3.4.3	Stability testing	113
3.4.4	Potency testing	114
4	Regulatory aspects	114
5	Future perspectives	114
	References	115
<b>7</b>	<b>Polyvalent Vectors for Coexpression of Multiple Genes</b>	<b>119</b>
1	Introduction	119
2	Polycistronic expression vectors	121
2.1	Mechanisms of translation initiation	121
2.2	Characteristics of IRES elements	124
2.3	Application of IRES elements in cells and animals	126
2.4	Polycistronic vector systems	128
2.5	Expression properties of IRES vectors	130

3	Bidirectional promoters	130
3.1	Natural bidirectional promoters	130
3.2	Artificial bidirectional promoters	131
3.3	Combining polycistronic and bidirectional expression	132
4	Perspectives	133
	References	133
<b>8</b>	<b>Form Follows Function: The Design of Minimalistic Immunogenically Defined Gene Expression (MIDGE®) Constructs</b>	<b>139</b>
1	The problem	139
2	The solution	141
2.1	MIDGE – the concept	142
2.2	Simple MIDGE	143
2.3	Smart MIDGE	143
2.4	Applications	145
2.5	Practical aspects of vector sequence design	145
	References	146
<b>9</b>	<b>Synthetic Genes for Prevention and Therapy: Implications on Safety and Efficacy of DNA Vaccines and Lentiviral Vectors</b>	<b>147</b>
1	Introduction	147
2	Paradoxon: HIV-derived vaccines and gene delivery systems	151
3	Synthetic genes: Novel tools contributing to the understanding of HIV replication	152
3.1	Construction of a synthetic, HIV-1 derived <i>gag</i> gene	152
3.2	Codon usage modification in the <i>gag</i> gene abolishes Rev dependency and increases expression yields	153
3.3	Codon usage modification in the <i>gag</i> gene increases nuclear RNA stability and promotes constitutive nuclear translocation	154
3.4	Codon usage modification in the <i>gag</i> gene alters the nuclear export pathway of otherwise CRM1 dependent RNAs	156
3.5	Codon usage modification increases RNA stability, modulates nuclear RNA export and increases translational efficiency	157
4	Synthetic genes: Implications on the development of safe and effective DNA vaccines	158
4.1	Safety issues to be considered for DNA vaccine development	158
4.2	Codon optimization of a <i>gag</i> -specific candidate vaccines results in increased antibody responses	159
4.3	Enhanced <i>in vitro</i> cytokine release of splenocytes from mice immunized with synthetic <i>gag</i> plasmid DNA	160
4.4	Induction of CTL responses in mice immunized with the modified Gag expression plasmids	161

5	Synthetic genes: Implications on the development of safe lentiviral vectors for gene delivery into quiescent cells	162
5.1	Safety issues to be considered for lentiviral vector development	163
5.2	Construction and characterization of synthetic <i>gagpol</i> expression plasmids	163
5.3	Production of lentiviral vectors using synthetic <i>gagpol</i> genes	164
5.4	Transduction of non-dividing cells	164
5.5	Absence of replication-competent recombinants (RCRs)	165
6	Future perspectives	165
	References	166
<b>10</b>	<b>Plasmids in Fish Vaccination</b>	<b>169</b>
1	Introduction	169
2	Fish	170
3	Fish immunology	171
3.1	Innate defence mechanisms	171
3.2	Adaptive defence mechanisms	171
4	Vaccination of fish	173
5	Nucleic acid vaccination of fish	174
6	Plasmid constructs used in fish studies	175
7	Routes of plasmid administration	176
7.1	Intramuscular injection of plasmid DNA	176
7.2	Other routes of plasmid administration	178
8	Fate of injected plasmid DNA	178
9	Magnitude, distribution and longevity of expressed antigen	179
10	Responses of fish to injection with plasmid DNA	181
10.1	Inflammatory responses	181
10.2	Avirulent antigens	181
10.3	Virulent antigens	182
11	Regulatory issues and future directions	187
	References	188
<b>11</b>	<b>Plasmid Manufacturing – An Overview</b>	<b>193</b>
1	Introduction	193
2	Structure of nucleic acids	194
2.1	Brief structural description of DNA and RNA structures	194
2.2	DNA supercoiling	196
3	Plasmid DNA Manufacturing	199
3.1	Major impurities and main product specifications	201
3.1.1	Host nucleic acids	202
3.1.2	Proteins	203
3.1.3	Endotoxins	203

3.2	Factors influencing the production of plasmid DNA: Some considerations on the upstream processing and fermentation stages	204
3.2.1	The plasmid vector	204
3.2.2	The bacterial host strain	205
3.2.3	Plasmid fermentation	206
3.3	Downstream processing of plasmid DNA	208
3.3.1	Cell lysis	210
3.3.2	Pre-chromatography processing: clarification and concentration	213
3.3.3	Chromatographic processing: purification of supercoiled plasmid DNA	214
3.4	Purification Strategies	228
4	Concluding remarks	229
	References	231
<b>12</b>	<b>Quality Control of pDNA</b>	<b>237</b>
1	Introduction	237
2	Characterization and quality control of pDNA	237
3	Validation of test procedures	239
4	GLP	240
5	Detailed description of the characterization of pDNA (final product)	241
5.1	Sterility	241
5.2	Purity	243
5.2.1	Content	243
5.2.2	Homogeneity	244
5.2.3	Host DNA	246
5.2.4	Host cell protein impurities	248
5.2.5	Endotoxins	248
5.3	Identity	249
5.3.1	Restriction analysis	249
5.3.2	Determination of the DNA sequence	251
6	Conclusion	251
	References	252
<b>13</b>	<b>From Research Data to Clinical Trials</b>	<b>255</b>
1	Introduction	255
2	Approaching regulators	256
3	Vaccine manufacture	256
4	Preclinical safety testing	257
5	Clinical trials	258
6	Approval for clinical trials	259
7	Clinical trial applications in Germany	259
	References	260

<b>14</b>	<b>Market Potential for DNA Therapeutics</b>	<b>261</b>
1	Definition of biotechnology	261
2	History	262
3	Process of pharmaceutical development	263
4	Human society and technical revolution	264
5	From sequence to product: Applications of biotechnology	265
5.1	Milestones in biotechnology: The Human Genome Project	267
5.2	The future is now: Examples for existing therapeutic approaches using gene products	268
5.2.1	Gene therapy in cardiovascular diseases	268
6	Legal aspects of gene technology and pDNA derived products	270
6.1	The extension of patent law to living creatures and their components	270
6.2	Impacts of biomedical patents	271
6.3	Resisting corporate ownership of life forms	272
7	Health care in the light of biotech is different in Europe and US	274
7.1	Biotech in the US from an economical point of view	274
7.2	Emergence of new companies	274
7.3	Biotech in Germany	275
8	Who is the health care industry and their clients, a paradigm?	276
8.1	Health care industry share prices	277
8.2	HMO enrolment rose	278
8.3	Limitations to the access of health care services	278
8.4	US uninsured population rose	279
9	Economical evaluation of the biotech marked in the future	279
10	Conclusion	280
	<b>Index</b>	<b>281</b>

This Page Intentionally Left Blank



## List of Contributors

Jens Alfken  
MOLOGEN Berlin  
Fabeckstrasse 30  
D-14195 Berlin  
Germany  
Alfken@mologen.com

Ruth Baier  
QIAGEN GmbH  
Max-Volmer-Strasse 4  
D-40724 Hilden  
Germany  
ruth.baier@web.de

Kurt Bieler  
Institut für Medizinische  
Mikrobiologie und Hygiene  
Universität Regensburg  
Franz-Josef Strauß Allee 11  
D-93053 Regensburg  
Germany  
kurt.bieler@klinik.uni-regensburg.de

Joaquim M. S. Cabral  
Centro de Engenharia Biológica  
e Química  
Instituto Superior Técnico  
Av Rovisco Pais 1049-001  
Lisboa  
Portugal  
gmmferreira@ist.utl.pt

Klaus Cichutek  
Abteilung für Medizinische  
Biotechnologie  
Paul-Ehrlich-Institut  
Paul-Ehrlich-Strasse 51-59  
D-63225 Langen  
Germany  
cickl@pei.de

Juergen Dobmeyer  
IMS HEALTH GmbH & Co. OHG  
Hahnstrasse 30-32  
D-60528 Frankfurt/Main  
Germany  
dobmeyer@gmx.net

Rita Dobmeyer  
Johann-Klotz-Strasse 3  
D-60528 Frankfurt/Main  
Germany  
dobmeyer@t-online.de

Julia Dörge  
Cardion AG  
Max-Planck-Strasse 15a  
D-40699 Erkrath  
Germany  
doerge@cardion-ag.de

Andreas Fels  
NewLab Bioquality AG  
Max-Planck-Strasse 15A  
D-40699 Erkrath  
Germany  
fels@newlab.de

Guilherme N. M. Ferreira  
Centro de Engenharia Biológica  
e Química  
Instituto Superior Técnico  
Av Rovisco Pais 1049-001  
Lisboa  
Portugal  
gnmferreira@ist.utl.pt

Erwin Flaschel  
Universität Bielefeld  
Technische Fakultät  
AG Fermentationstechnik  
D-33501 Bielefeld  
Germany  
efl@fermtech.techfak.uni-bielefeld.de

Karl Friehs  
Universität Bielefeld  
Technische Fakultät  
AG Fermentationstechnik  
D-33501 Bielefeld  
Germany  
kfr@fermtech.techfak.uni-bielefeld.de

Sarah C. Gilbert  
Wellcome Trust Centre for  
Human Genetics  
Roosevelt Drive  
Headington  
Oxford OX3 7BN  
UK  
sarah.gilbert@ndm.ox.ac.uk

Hansjörg Hauser  
Abteilung für Genregulation  
and Differenzierung  
GBF – Gesellschaft für  
Biotechnologische Forschung  
Mascheroder Weg 1  
D-38124 Braunschweig  
Germany  
hha@gbf.de

Angela Heischmann  
Cardion AG  
Max-Planck-Strasse 15a  
D-40699 Erkrath  
Germany  
heischmann@cardion-ag.de

Andreas Herrmann  
Cardion AG  
Max-Planck-Strasse 15a  
D-40699 Erkrath  
Germany  
herrmann@cardion-ag.de

Adrian V. S. Hill  
Level 7  
NDM  
John Radcliffe Hospital  
Headington  
Oxford OX3 9DU  
UK  
adrian.hill@Molecular-Medicine.  
oxford.ac.uk

Simon R. M. Jones  
Department of Fisheries and Oceans  
Pacific Biological Station  
3190 Hammond Bay Road  
Nanaimo, British Columbia V9R 5K6  
Canada  
joness@pac.dfo-mpo.gc.ca

Claas Junghans  
MOLOGEN Berlin  
Fabeckstrasse 30  
D-14195 Berlin  
Germany  
Junghans@mologen.com

Sven A. Koenig-Merediz  
MOLOGEN Molecular Medicines S. L.  
CNBF  
Instituto de la Salud Carlos III  
28220 Majadahonda  
Madrid  
Spain  
Koenig@mologen.com

Andrea Kröger  
Abteilung für Genregulation  
und Differenzierung  
GBF – Gesellschaft für  
Biotechnologische Forschung  
Mascheroder Weg 1  
D-38124 Braunschweig  
Germany  
hha@gbf.de

Marcin Kwissa  
Universität Ulm  
Institut für Medizinische  
Mikrobiologie  
Helmholtzstrasse 8/1  
D-89081 Ulm  
marcin.kwissa@medizin.uni-ulm.de

Marie-Louise Michel  
UREG, INSERM U163  
Dpartement des Retrovirus  
Institut Pasteur  
75015 Paris  
France  
maloum@pasteur.fr

Andreas Muhs  
Cardion AG  
Max-Planck-Strasse 15a  
D-40699 Erkrath  
Germany  
muhs@cardion-ag.de

Peter P. Müller  
Abteilung für Genregulation  
and Differenzierung  
GBF – Gesellschaft für  
Biotechnologische Forschung  
Mascheroder Weg 1  
D-38124 Braunschweig  
Germany  
hha@gbf.de

André Oumard  
Abteilung für Genregulation  
and Differenzierung  
GBF – Gesellschaft für  
Biotechnologische Forschung  
Mascheroder Weg 1  
D-38124 Braunschweig  
Germany  
hha@gbf.de

Duarte M. F. Prazeres  
Centro de Engenharia Biológica  
e Química  
Instituto Superior Técnico  
Av Rovisco Pais 1049-001  
Lisboa  
Portugal  
gnmferreira@ist.utl.pt

Joerg Reimann  
Universität Ulm  
Institut für Medizinische Mikrobiologie  
Helmholtzstrasse 8/1  
D-89081 Ulm  
Germany  
joerg.reimann@medizin.uni-ulm.de

xviii | *List of Contributors*

Andreas Richter  
Facility Manager  
NewLab Bioquality AG  
Max-Planck-Strasse 15A  
40699 Erkrath  
richter@newlab.de

James S. Robertson  
National Institute for Biological  
Standards and Control  
Blanche Lane  
South Mimms  
Potters Bar  
Herts EN6 3QG  
UK  
jrobertson@nibsc.ac.uk

Manfred Rüdiger  
Cardion AG  
Max-Planck-Strasse 15a  
D-40699 Erkrath  
Germany  
ruediger@cardion-ag.de

Florian Sack  
MOLOGEN Berlin  
Fabeckstrasse 30  
D-14195 Berlin  
Germany  
Sack@mologen.com

Reinhold Schirmbeck  
Universität Ulm  
Institut für Medizinische Mikrobiologie  
Helmholtzstrasse 8/1  
D-89081 Ulm  
Germany  
reinhold.schirmbeck@medizin.  
uni-ulm.de

Martin Schleef  
PlasmidFactory GmbH & Co. KG  
Meisenstrasse 96  
D-33607 Bielefeld  
Germany  
Martin.Schleef@PlasmidFactory.com

Jens Schletter  
Cardion AG  
Max-Planck-Strasse 15a  
D-40699 Erkrath  
Germany  
schletter@cardion-ag.de

Torsten Schmidt  
PlasmidFactory GmbH & Co. KG  
Meisenstrasse 96  
D-33607 Bielefeld  
Germany  
Torsten.Schmidt@PlasmidFactory.com

Joerg Schneider  
Oxxon Pharmaccines Ltd  
The Oxford BioBusiness Centre  
Littlemore Park  
Littlemore  
Oxford OX4 4SS  
UK  
joerg@oxxonpharmaccines.com

Jürgen Schrader  
Abteilung für Physiologie  
Heinrich-Heine-Universität  
D- Düsseldorf  
Germany  
schrader@uni-duesseldorf.de

Matthias Schroff  
MOLOGEN Berlin  
Fabeckstrasse 30  
D-14195 Berlin  
Germany  
Schroff@mologen.com

Wolfgang Schumann  
Institut für Genetik  
Universität Bayreuth  
D-95440 Bayreuth  
Germany  
wolfgang.schumann@uni-bayreuth.de

Colin Smith  
MOLOGEN Berlin  
Fabeckstrasse 30  
D-14195 Berlin  
Germany  
Smith@mologen.com

Heiko E. von der Leyen  
Cardion AG  
Max-Planck-Strasse 15a  
D-40699 Erkrath  
Germany  
vdleyen@cardion-ag.de

Ralf Wagner  
Institut für Medizinische Mikrobiologie  
und Hygiene  
Universität Regensburg  
Franz-Josef Strauß Allee 11  
D-93053 Regensburg  
Germany  
and  
geneart GmbH  
BioPark Regensburg  
Josef-Engert-Strasse 9  
D-93053 Regensburg  
Germany  
Ralf.Wagner@geneart.de

Dagmar Wirth  
Abteilung für Genregulation  
and Differenzierung  
GBF – Gesellschaft für  
Biotechnologische Forschung  
Mascheroder Weg 1  
D-38124 Braunschweig  
Germany  
hha@gbf.de

Bürghardt Wittig  
Abteilung für Molekularbiologie  
und Bioinformatik  
Fachbereich Humanmedizin  
Freie Universität Berlin  
Arnimallee 22  
D-14195 Berlin  
Germany  
and  
MOLOGEN Berlin  
Fabeckstrasse 30  
D-14195 Berlin  
Germany  
Wittig@mologen.com

In the text corresponding authors are marked with an asterisk.

# 1

## The Biology of Plasmids

Wolfgang Schumann

### 1

#### Introduction: What are plasmids?

Plasmids are autonomously replicating entities which can be found essentially in all bacterial species and which play a significant role in bacterial adaptation and evolution. Furthermore, plasmids are studied for their own sake and serve as important tools in studies of molecular biology. Plasmids are normally circular, although linear forms have also been described and vary widely in size from 1 kb to 200 kb. Larger plasmids of up to 1,000 kb termed megaplasmids have also been identified. The copy number per chromosome also varies among plasmids, and bacterial cells can carry more than one type.

Like chromosomes, plasmids code for RNA molecules and proteins, replicate as the cell grows, and equal numbers are normally distributed to the two daughter cells upon cell division. However, plasmids do not code for functions which are essential to bacterial growth in the absence of any stressful situation, e. g., in the absence of an antibiotic.

The first plasmids which have been described were named after their presence of known phenotypes. Therefore, the first plasmid discovered in the early 50s of the last century has been designated F factor which stands for fertility since this plasmid is involved in the exchange of genetic information in *Escherichia coli* and related Enterobacteriaceae. In the late 1950s, so-called R factor plasmids have been discovered in Japan, when *E. coli* and *Shigella* strains resistant to a number of antibiotics were isolated from the fecal flora of patients. The ColE1 plasmid harbors a gene the product of which named colicin E1 can kill bacteria not carrying this plasmid. The Ti plasmid of *Agrobacterium tumefaciens* contains genes which can be transferred to plant cells where they induce tumors.

This system of nomenclature has led to some confusion since several plasmids carry genes coding for phenotypes different from that for which they were originally named. To avoid further confusion, the nomenclature of bacterial plasmids has now been standardized. Plasmid names start with a small "p" for plasmid followed by capital letters that either describe the plasmid or the location where it has been

constructed or give the initials of the person or persons who isolated or constructed it. These letters are then followed by numbers to identify the particular construct. To give two examples, the plasmid pBR322 was constructed by Bolivar and Rodriguez, pUC19 stands for University of California.

Why are there plasmids? If plasmid genes, such as those for antibiotic resistance and toxin synthesis, were part of the chromosome all cells would benefit from those genes. There is one disadvantage and one advantage to explain the presence of plasmids. The disadvantage is that the bacteria would have the burden to replicate and to maintain a larger chromosome. And we know from experiments that bacterial cells with smaller genomes will outgrow those with larger ones, and bacterial cells without a plasmid will outcompete those with a plasmid, provided that the plasmid function(s) is not needed. The advantage of having plasmids is that genetic information not essential for growth under all conditions can be easily distributed among cells of a population and even to cells of other species depending on the host range of the plasmids. While some have a very narrow host range (e. g., enteric species) others can be transferred to and replicated into all gram-negative species.

The presence of plasmids in a bacterial host can change its phenotype in a variety of ways. Four basic groups of plasmid genes have been determined:

1. Determinants involved in plasmid replication and segregation to daughter cells:  
These determinants control plasmid characteristics such as copy number per chromosome, incompatibility and host range.
2. Conjugational transfer determinants:  
These control plasmid transmissibility and associated characteristics of the plasmid-bearing strain, such as sensitivity to donor-specific phages.
3. Determinants which regulate interactions with other replicons and extra-chromosomal elements:  
These include plasmid sequences involved in chromosome mobilization and genes the products of which inhibit the fertility of other conjugative plasmids in the same host, inhibit the growth of specific bacteriophages, or prevent the host from acting as a recipient in conjugal crosses (surface exclusion).
4. Determinants which affect the interaction of the plasmid-bearing cell with the environment:  
These include genes for bacteriocin production and immunity, adhesion and pathogenicity factors, resistance to antibacterial agents (antibiotics, ions and radiation), and metabolism of environmental substrates (sugars, organic acids, aromatic and aliphatic hydrocarbons, detergents and pesticides).

## 2

### General properties of plasmids

Important properties of bacterial plasmids are strict control of their replication to achieve stable coexistence with its host. Besides control of the copy number, plasmids must be equally segregated into the two daughter cells, especially when they

are present in a few copies only. Since bacterial cells can take up and inherit more than one plasmid, what will happen if two plasmids belong to the same incompatibility group? All plasmids can be transferred to other cells of the same or of other species. Conjugative plasmids code for the genetic information to catalyze their own transfer from the donor to the recipient cell. Mobilizable plasmids need the presence of a conjugative plasmid within the same cell to achieve transfer to the recipient cell.

## 2.1

### Plasmid replication and its control

When the bacterial cell grows, it replicates its chromosome(s) while plasmids present within the cell are also replicated. While autonomous replication of plasmids is a fundamental characteristic, they must synchronize their replication with the division of the host. To guarantee stable coexistence with its host, each plasmid must replicate, on the average, once every generation.

Each plasmid normally has one origin of replication, or *oriV* site where replication begins. In addition, the plasmid must encode at least one protein that enables replication to initiate at the *oriV* site. All other required proteins such as DNA polymerase, ligase, primase and helicase are taken from the host cell. Each type of plasmid replicates by one of two general mechanisms: theta or rolling circle replication.

Plasmids replicating via the theta mechanism start by opening the two strands of DNA at the *oriV* region, creating a structure that looks like the Greek letter  $\theta$  – hence the name theta replication. Replication can proceed from the *oriV* in one or in both directions depending on the plasmid. In the first case, one leading primer is synthesized by the primase which is extended by DNA polymerase III, and one replication fork moves around the whole plasmid DNA (*unidirectional replication*). In the second case, two leading primers are synthesized within the *oriV* region, two replication forks start in opposite directions at the *oriV* which meet somewhere on the other side of the molecule (*bidirectional replication*). The theta mechanism is the most common form of plasmid DNA replication. It is used by most plasmids such as ColE1, RP4, F factor and P1 prophage.

In the other type of replication, a single-strand break (nick) is introduced in one strand of the double-stranded plasmid at the so-called plus origin. The free 3' OH end serves as a primer, and replication proceeds around the circle, displacing the opposite or minus strand. On the displaced strand, replication initiates at specific sites, minus-strand origins, to synthesize the second strand. With some plasmids minus-strand origins do not function properly resulting in the accumulation of single-stranded plasmids. These aberrant forms of single-stranded plasmids are found in some gram-positive plasmids such as pUB110 and pC194 from *Staphylococcus aureus*.

In most plasmids, the genes coding for replication proteins are located adjacent to the *oriV*. Thus, only a very small region surrounding the plasmid *oriV* site is required for replication. This finding prompted the construction of relatively

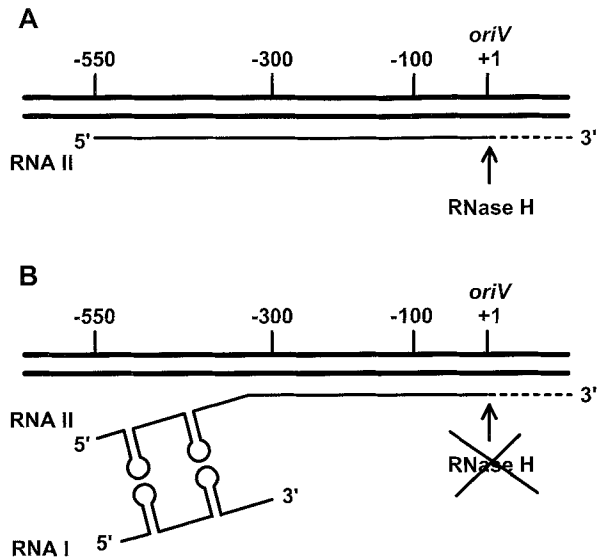


small cloning vectors by using the essential backbone of prototype plasmids (see below).

The *host range* of a plasmid includes all types of bacteria in which the plasmid can replicate, and the host range is usually determined by the *oriV* region. Some plasmids such as ColE1 and the derived cloning vectors pBR322 and pUC have very narrow host ranges. These plasmids will replicate only in *E. coli* and closely related bacteria belonging to the group of Enterobacteriae. In contrast, plasmids with a broad host range include RP4, RSF1010 and the rolling circle plasmids of gram-positive bacteria. Plasmids with the host range of RP4 will replicate in most, if not all gram-negative bacteria and RSF1010 will even replicate in some types of gram-positive bacteria. Broad-host range plasmids must encode all of their own proteins required for initiation of replication, therefore being independent from the host cell for any of these functions. Furthermore, they must be able to express these genes in many types of bacteria. This means that the promoters and ribosome initiation sites for the replication genes will be recognized in a wide variety of bacterial species.

The copy number of plasmids is determined mostly by their *oriV* region. All plasmids must regulate their replication. While some plasmids replicate enough to populate the cell with hundreds of copies, others such as the F plasmid of *E. coli* or the P1 prophage replicate only once or a few times during the cell cycle. Plasmids that have high copy numbers, such as ColE1, only need to have a mechanism that inhibits the initiation of plasmid replication when the copy number of plasmids in the cell reaches a certain number. These molecules are called *relaxed plasmids*. By contrast, low-copy number plasmids must have a tighter mechanism for regulating their replication and are therefore called *stringent plasmids*. Replication is controlled by the plasmid itself. Three different strategies will be discussed: regulation by RNA, regulation by RNA and protein and regulation by a protein.

Replication of ColE1 is regulated through the effects of a small plasmid-encoded RNA called RNA I. This 108 nt long RNA inhibits plasmid replication by interfering with the processing of another RNA called RNA II which serves as the primer for plasmid DNA replication (Figure 1). This RNA is first synthesized as an about 550 nt preprimer and forms an RNA–DNA hybrid at the replication origin. Then, RNA II is cleaved by RNase H specific for DNA–RNA hybrids, releasing a 3' hydroxyl group that serves as the primer for replication catalyzed by DNA polymerase I. RNA II will only function as a primer when it is processed properly. RNA I and RNA II are complementary to each other, since they are made from the same region of the DNA, but from opposite strands. Both RNAs can base pair to form a double-stranded RNA helix, and this helix is refractory to processing. In addition, base pairing between RNA I and RNA II is enhanced by the homodimeric Rom protein, a *trans*-acting inhibitor of replication. Rom increases the rate at which RNA I binds to the preprimer transcript by stabilizing the initial complex between the two RNAs. RNA I is rapidly synthesized and is extremely unstable ( $t_{1/2} = 2$  min) and this instability is related to polyadenylation carried out by poly(A) polymerase I. More RNA I will be made when the copy number of ColE1 increases maintaining the number of ColE1 molecules at around 20 per cell.



**Fig. 1.** Regulation of replication initiation of the ColE1 plasmid. **(A)** Replication is initiated by synthesis of a primer by the RNA polymerase where its 5' end is located 555 nucleotides upstream of the *oriV* and its 3' end is variable; this primer RNA has been designated RNA II. Next, RNase H (an RNase which specifically recognizes DNA-RNA hybrids) processes these RNA II molecules of heterogeneous

length so that their 3' ends are located within the *oriV* which are then elongated by DNA polymerase I. **(B)** Processing of RNA II can be inhibited by interaction with RNA I which is derived from the complementary DNA strand and acts as an antisense RNA. Initial contact between RNA I and RNA II occurs via loops of their secondary structures also called the kissing complex.

Another extensively studied plasmid is R1 where the copy number is regulated by RNA and protein. Initiation of replication is regulated by the RepA protein. The *repA* gene is transcribed from two promoters, one of them,  $P_{\text{copB}}$ , transcribes *copB* and *repA* and the second,  $P_{\text{repA}}$ , located within *copB* makes an mRNA that encodes only the RepA protein. When the R1 plasmid enters a cell, a short burst of synthesis of RepA from  $P_{\text{repA}}$  causes the plasmid to replicate until it attains its copy number. Then,  $P_{\text{repA}}$  is repressed by the CopB protein, and the *repA* gene can be transcribed only from the  $P_{\text{copB}}$  promoter. Once the plasmid has attained its copy number, the regulation of synthesis of RepA is regulated by an RNA called CopA. The *copA* gene is transcribed from its own promoter located within *repA* and reading into the opposite direction and the RNA product affects the stability of the mRNA made from  $P_{\text{copB}}$ . Since both RNAs are complementary, they can pair and, therefore, prevent binding of the ribosomes to the Shine-Dalgarno sequence of the *repA*. In summary, the replication of R1 will be regulated by the concentration of CopA RNA which in turn will depend on the concentration of the plasmid. The higher the concentration of the plasmid, the more CopA RNA will be made and the less RepA protein will be synthesized maintaining the plasmid copy number.

The third strategy regulates replication by a protein. This mechanism is used by the F factor, the P1 prophage and by pSC101, R6K and RP4. The *oriV* region of these plasmids contains several tandem DNA repeats termed the *iteron* sequences which are typically 17 bp to 22 bp long and exist in about three to seven tandem repeats in the *oriV* region. Additionally, there usually are copies of these repeated sequences a short distance apart. The plasmid pSC101 contains within *oriV* region the gene *repA* coding for the protein required for initiation of replication, and three repeated iteron sequences called R1, R2 and R3 through which RepA regulates the copy number. Iteron replication is regulated by two superimposed mechanisms. First, the RepA protein represses its own synthesis by binding to the inverted repeats IR1 and IR2 to autoregulate its own gene. The higher the concentration of pSC101 within the cell, the more RepA will be synthesized leading to a stronger repression. Therefore, the concentration of RepA protein is maintained within narrow limits, and the initiation of replication is strictly regulated.

In the second mechanism, RepA regulates the initiation of replication through binding to the three iterons. While binding of RepA to only one iteron favors initiation of replication, binding to all three represses replication. Furthermore, if all three iterons are occupied two plasmids can interact via the RepA molecules, and this is called the coupling or handcuffing model.

## 2.2

### The molecular basis of incompatibility

Many bacteria isolated from their natural habitat carry more than one type of plasmid which stably coexist in the bacterial cell and remain there even after many generations. However, not all types of plasmids can stably coexist in bacterial cells. Some types will interfere with each other's replication or partitioning so that if two such plasmids are introduced into the same cell, one or the other will be lost when the cells divide. This phenomenon is called plasmid incompatibility. Two plasmids A and B are *compatible* with each other when they stably coexist in the same cell in the absence of selective pressure. If plasmid A excludes plasmid B from the cell, both plasmids are *incompatible* with each other. This finding has led to the classification of plasmids into different incompatibility (Inc) groups. There may be hundreds of different Inc groups, and, e.g., plasmids RP4 and RK2 both belong to the IncP (incompatibility group P) while RSP1010 is an IncQ plasmid. This means that RP4 and RSP1010 can be stably maintained within one single cell since they belong to different incompatibility groups. The knowledge about the Inc phenotype of plasmids is important in gene technology experiments when more than one plasmid is used per cell.

What is known about the molecular basis of incompatibility? Plasmids will be incompatible when they either share the same mechanism of replication control or the same *par* (partitioning) function. If two plasmids are replicated by the same replication control system, they are randomly selected for replication. Assuming that a cell starts with two plasmids A and B belonging to the same incompatibility group which are both present in ten copies each, during the first re-

plication round the cell will select seven copies of A which are each replicated four times summing up to 28 copies in total. Furthermore, three copies of B are selected and replicated four times each totaling to twelve copies of B which are then distributed to the two daughter cells. During the next replication round this imbalance is further increased, and after a few generations, plasmid B gets lost from the population. In general, the smaller plasmid is retained while the larger plasmid will disappear. The reason for this finding is that the smaller plasmid is replicated faster than the larger one and, therefore, has a better chance to be reused for the next replication round.

The second molecular basis for incompatibility is based on the *par* function. Many plasmids, especially those with a low copy number, carry *partitioning* systems ensuring that at least one copy of the plasmid segregates into each daughter cell during cell division. The genetic determinants have been termed *par* functions. In the case of the P1 plasmid, the *par* region consists of the *cis*-acting *incB* sequence and the two genes *parA* and *parB*. How do the partitioning functions operate? The following model has been suggested: The *incB* site binds three molecules of the 39 kDa ParB protein. Two molecules interact with a perfect 13 bp palindrome and the third binds to a region separated from the palindrome by a recognition site for the IHF (integration host factor). The IHF is a host-encoded heterodimeric protein which bends the DNA by 140° upon binding, thereby bringing all three ParB molecules in close proximity. This complex then interacts with the membrane at the division site. ParA belongs to the group of ATPases and binds close to the *incB* site. Possibly, ATP hydrolysis by the ParA–ParB *incB* complex provides the energy to separate the products of plasmid replication in a process analogous to mitosis in eukaryotic cells. When two plasmids A and B share the same *par* function, one or the other will always interact with the membrane site and be distributed into the daughter cell during division while the other will randomly segregate into one of the two daughter cells. However, sometimes one daughter cell will receive plasmid A and the other cell will receive plasmid B, producing cells cured of one or the other plasmid.

## 2.3

### Plasmid inheritance

Cells that have lost their plasmid during cell division are said to be *cured* of the plasmid. Since cells without plasmids will grow a little faster than those carrying plasmids within a population, several mechanisms prevent curing: site-specific recombinases that destroy multimers and addiction systems.

The possibility of losing a plasmid during cell division is increased, if the plasmid forms multimers during replication or via homologous replication. Since the total number of origins will always be the same independent whether they occur on many monomers or a few multimers, multimers will increase the chance that daughter cells arise without the plasmid, since a few multimers might segregate within just one daughter cell. To avoid this problem, many plasmids have acquired site-specific recombination systems to resolve multimers. These resolution

systems are composed of a site-specific recombinase and a short DNA sequence (about 30 bp) to which the recombinase binds. The principle of monomerization will be illustrated with two examples: P1 prophage and ColE1.

Prophage P1 codes for the sequence-specific Cre recombinase which interacts with the *loxP* sequence. When two P1 monomers form a dimer, the two *loxP* sequences occur as direct repeats. The Cre recombinase will bind to both of them, align both sequences, carry out recombination and thereby resolve the dimer into two monomers. Another example is ColE1 where the resolution site is called *cer*. This *cis*-acting 240 bp DNA recombination sequence is recognized by the two site-specific recombinases XerC and XerD which are encoded by the *E. coli* chromosome. Here this recombination system also resolves ColE1 multimers into monomers.

Some plasmids code for a system which kills cured cells to prevent their disappearance from a culture due to faster growth of plasmidless cells. These systems are called *addiction* or *post-segregational killing systems* and consist of two components. While one component is always a protein toxic to the cell, the other is either an antisense RNA or a second protein. The antisense RNA prevents translation of the mRNA coding for the toxic protein, and the second protein interacts with the toxic protein thereby preventing its lethal action. As long as the cell carries the plasmid, action of the toxic protein is prevented due to the antidote. When a daughter cell without the plasmid arises, the toxic protein becomes active due to the shorter half-life of either the antisense RNA or the second protein. The F factor codes for the antidote CcdA (coupled cell division), a 8.3 kDa polypeptide and the toxic protein CcdB (11.7 kDa). The CcdB toxin alters the activity of DNA gyrase and causes double-stranded breaks in the bacterial chromosome. The addiction system of plasmid R1 consists of the two genes *hok* (host killing) and *sok* (suppression of killing). They are encoded on opposite strands with a 128 bp overlap at their 5' ends. While the Hok polypeptide (5.5 kDa; destroys the cellular membrane potential thereby causing loss of cellular energy) is responsible for the host cell killing, *sok* codes for an antisense RNA which prevents translation of the *hok* mRNA.

## 2.4

### Mechanisms of plasmid spread

The most important characteristic of plasmids is their ability to transfer themselves to other cells of the same or even different bacterial species. These mechanisms are called conjugation and mobilization and will be described in detail. Besides this mechanism of self-transmission, plasmids released from lysed cells can also be taken up by surrounding cells in a process termed transformation. And the mechanisms of plasmid spread involve phages which may package plasmid DNA into their heads and inject them into new hosts. We will consider here only the conjugative mechanisms which need genetic information encoded by the plasmid itself.

### 2.4.1 Conjugation in gram-negative bacteria

One remarkable feature of some plasmids is the ability to transfer themselves from one cell to another in a process called *conjugation*. This process was discovered in 1947, when mixing different of *E. coli* strains resulted in strains that were genetically different from their originals. Later, it was discovered that conjugative plasmids can cause the transfer of non-conjugative plasmids into other cells. This process was termed *mobilization*.

During conjugation or mobilization, a plasmid is transferred from the *donor* to the *recipient* cell, and the recipient bacterium that has received DNA is called *trans-* or *exconjugant*. Conjugative plasmids in gram-negative bacteria such as in *E. coli* need three different genetic elements for successful transfer of their DNA. These are the *cis*-active *oriT* (origin of transfer), the *tra* (transfer) and the *mob* (mobilization) genes. The process of conjugation will be illustrated with the example of the F factor, the most extensively studied conjugative plasmid with a size of 100 kb. Plasmid transfer specifically initiates at the *oriT* site which is about 250 bp long and is recognized by the TraI protein (in general called Mob protein) which introduces a single-strand break in one strand at a specific site. Then, the TraI protein is covalently linked to via a tyrosine residue to the 5' phosphoryl group and acts as a 5'-3' helicase to unwind the donor duplex. At the same time, DNA polymerase III binds to the free 3' end to start replication, and the displaced strand is transferred to the recipient strain.

Transfer needs close contact between the donor and the recipient cell. This is accomplished by the F or *sex pilus*, an appendage of the cell surface of the donor cell. At least 14 *tra* genes are involved in construction of the F pilus, a hollow cylinder of 8 nm in diameter with a 2 nm axial hole. The pilus is constructed from pilin subunits which are encoded by *traA* and processed by the product of *traQ*. After the initial contact between the tip of the pilus and the recipient cell, the pilus retracts by depolymerization and thereby brings donor and recipient into close contact. Transfer of the single-stranded DNA occurs either through the hollow pilus or through the membranes of the two attached cells. It has to be stressed that the donor cell contains several pili on its outer surface enabling one donor to make contact with several recipient cells leading to the formation of mating aggregates. When the complete single-strand DNA molecule has arrived within the recipient, the bound TraI protein acts as a DNA ligase, seals the two ends and dissociates from the circular molecule which is then converted to a double-stranded molecule. In summary, the TraI protein has three enzymatic activities: sequence-specific endonuclease, helicase and DNA ligase. It should be mentioned here that the F pilus serves as a receptor for filamentous phages such as M13 and f1, which are used in gene technology experiments. These phages package single-stranded DNA and are used as vectors whenever single-stranded DNA is needed, e. g., during sequence-specific mutagenesis.

Besides replicating as an autonomous replicon, the F factor can also become integrated into the *E. coli* chromosome leading to an *Hfr* strain (high frequency of recombination). Two different mechanisms are responsible for recombination of the F factor into the chromosome, and both depend on one of the three IS (ins-