Transgenesis

APPLICATIONS OF GENE TRANSFER

Edited by
JAMES A. H. MURRAY

Transgenesis

APPLICATIONS OF GENE TRANSFER

Edited by
JAMES A. H. MURRAY
Institute of Biotechnology, University of Cambridge, UK

JOHN WILEY & SONS
Chichester · New York · Brisbane · Toronto · Singapore

Published 1992 by John Wiley & Sons Ltd, Baffins Lane, Chichester, West Sussex PO19 1UD, England

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, USA

Jacaranda Wiley Ltd, G.P.O. Box 859, Brisbane, Queensland 4001, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road, Rexdale, Ontario M9W 1L1, Canada

John Wiley & Sons (SEA) Pte Ltd, 37 Jalan Pemimpin #05-04, Block B, Union Industrial Building, Singapore 2057

Library of Congress Cataloging-in-Publication Data

Transgenesis: applications of gene transfer / edited by James A.H. Murray
p. cm.
Includes bibliographical references and index.
ISBN 0 471 93294 9
1. Microbial genetic engineering. 2. Animal genetic engineering.
3. Genetic transformation. I. Murray, James A.H.
TP248.6. T73 1992
660'.65—dc20 91-48077
CIP

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0471 93294 9

Typeset in 10/12 pt Palatino by Photographics, Honiton, Devon
Printed and bound in Great Britain by Biddles Ltd, Guildford

Contributors

Geert Angenon
Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium.

Philip J. Barr
Chiron Corporation, 4560 Horton Street, Emeryville, California 94608-2916, USA.

David H.L. Bishop
NERC Institute of Virology and Environmental Microbiology, Mansfield Road, Oxford OX1 3SR, UK.

Thomas T. Chen
Center of Marine Biotechnology and Department of Biological Science, University of Maryland, 600 E. Lombard Street, Baltimore, Maryland 21202, USA.

R. Wayne Davies
Robertson Institute of Biotechnology, University of Glagow, Church Street, Glasgow G11 5JS, UK.

Rex A. Dunham
Department of Fisheries and Allied Aquaculture, Auburn University, Auburn, Alabama 36849, USA.

Sarah J. Eccles
MRC Collaborative Centre, 1–3 Burtonhole Lane, Mill Hill, London NW7 1AD, UK.

Joachim W. Engels
Institut für Organische Chemie, J.W. Goethe-Universität, D-6000 Frankfurt-am-Main 50, Germany.

Martin J. Evans
Wellcome Trust/CRC Institute of Cancer and Developmental Biology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.

Godelieve Gheysen
Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium.

Christine Harvey
MRC Collaborative Centre, 1–3 Burtonhole Lane, Mill Hill, London NW7 1AD, UK.

Christopher C.G. Hentschel
MRC Collaborative Centre, 1–3 Burtonhole Lane, Mill Hill, London NW7 1AD, UK.

Klaus-Peter Koller
Hoechst AG, D-6230 Frankfurt-am-Main 80, Germany.

Michael Mackett
Cancer Research Campaign Department of Molecular Biology, Paterson Institute for Cancer Research, Christie CRC Research Centre, Christie Hospital and Holt Radium Institute, Manchester M20 9BX, UK.

James A.H. Murray
Institute of Biotechnology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QT, UK.

Elena Notarianni
Wellcome Trust/CRC Institute of Cancer and Developmental Biology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.
Present address:
Department of Molecular Embryology, AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK.

Robert D. Possee
NERC Institute of Virology and Environmental Microbiology, Mansfield Road, Oxford OX1 3SR, UK.

Dennis A. Powers
Hopkins Marine Station, Stanford University, Pacific Grove, California 93950, USA.

Dinko Valerio
Department of Gene Therapy, TNO Institute of Applied Radiobiology and Immunology, Lange Kleiweg 151, PO Box 5815, 2280 HV Rijswijk, The Netherlands.

V.W. van Beusechem
Department of Gene Therapy, TNO Institute of Applied Radiobiology and Immunology, Lange Kleiweg 151, PO Box 5815, 2280 HV Rijswijk, The Netherlands.

Contents

原书缺页

原书缺页

原书缺页

原

书

缺

页

2. MOLECULAR ASPECTS OF EXPRESSION

2.1. Transcription of Foreign Genes

The first step in the initiation of transcription is the binding of RNA polymerase to a promoter region in the DNA. An analysis of the DNA sequence of several *E. coli* promoters has revealed two regions of homology located upstream of the site of transcriptional initiation: one at 35 bp distance (−35 bp region, TTGACA) and one at 10 bp (−10 region, TATAAT) (Rosenberg and Court, 1979; Table 3). These conserved regions probably represent major contact points between promoter and RNA polymerase and, as predicted, synthetic promoters constructed with consensus sequences (e.g. P_{tac}; Russell and Bennett, 1982) act as strong promoters of transcription. Expression vectors are now readily available comprising either strong consensus promoters, e.g. P_{tac}, or strong bacteriophage promoters, e.g. λP_L (Drahos and Szybalski, 1981).

For foreign gene expression to be successful there are two aspects of transcription that are likely to be important: (1) regulation; and (2) termination. Constitutive (uncontrolled) expression of foreign genes on multi-copy vectors almost invariably leads to problems of segregational instability (generation of plasmid-free cells at cell division), overgrowth

Table 3. Examples of *E. coli* promoter sequences. The 'optimum' −35 and −10 region sequences are shown on the top line. Small circles indicate the bases at which RNA polymerase initiates transcription. The promoter recognized by the bacteriophage T7 RNA polymerase does not have the classical −35 and −10 homologies, and the entire necessary promoter sequence is shown. The *tac* promoter consists of the −35 region of the *trp* promoter and −10 region of lacUV5; in the *trc* promoter the −35 to −10 spacing has been adjusted to the optimal 17 bp

Promoter	−35 to −10 distance	−35	−10
Consensus	17	TTGACA	TATAAT
LacUV5	18	AGGC TTTACA	CTTTATGCTTCCGGCTCG TATAAT GTGTGGAA
trp	17	CCTG TTGACA	ATTAATCAT CGAACTAG TTAACT AGTACGCA
trp:lac (tac)	16	CCTG TTGACA	ATTAATCAT CGGCTCG TATAAT GTGTGGAA
trp:lac (trc)	17	CCTG TTGACA	ATTAATCAT CCGGCTCG TATAAT GTGTGGAA
λP_L	17	GGTG TTGACA	TAAATACCA CTGGCGGT GATACT GAGCACA
T7	-		TAATACGACTCACTATAGGGAGA

of plasmid-free cells and/or genetic instability (plasmid DNA rearrangement). This was demonstrated by Caulcott et al. (1985) in a chemostat study of the segregational instability of plasmid pAT153 derivatives expressing calf prochymosin.

pAT153 is a multi-copy plasmid with a copy number of approximately 50 copies per cell, and under the conditions tested in chemostat culture the parent plasmid pAT153 is stable for over 100 generations. However, in strains transformed with a derivative of pAT153 capable of high-level expression of a calf prochymosin gene, plasmid loss and overgrowth of plasmid-free cells was observed before steady-state growth could be achieved. This instability correlated with a reduced growth rate for cells expressing prochymosin. These results highlight the need to separate growth and production phases in fermentation processes designed to maximize product yield.

Regulated gene expression is most easily attained through the use of promoters carrying operator sequences to which protein repressors will bind (Table 4). Binding of the repressor protein to the operator interferes with the binding of RNA polymerase to its promoter and hence effects transcriptional repression. The induction of expression therefore requires inactivation of the repressor protein and this is commonly achieved by either a temperature shift (for λP_L, controlled by the cI^{857} temperature-sensitive repressor) or the addition of a chemical that binds to the repressor (for P_{tac}, isopropyl-β-D-thiogalactoside (IPTG) is the chemical inducer).

The bacteriophage λP_L promoter controlled by the temperature-sensitive repressor protein cI^{857} is the most tightly controlled expression system (Drahos and Szybalski, 1981). At 30°C there is very little expression of the foreign gene, whereas upon a temperature shift to 42°C there is rapid induction due to inactivation of the repressor protein. Although this system has been effectively employed for a variety of heterologous genes, a significant disadvantage is the need for elevated temperature to induce expression. Recent findings suggest that growth temperatures below 30°C may favour correct folding of foreign proteins produced in *E. coli* (section

Table 4. Examples of transcriptional regulation used in foreign gene expression

Promoter	Repressor	Inducer
trp	trp	IAA (3-β-indoleacrylic acid)
lac/tac	lacI	IPTG
λP_L	cI^{857}	Temperature
RecA	lexA	Nalidixic acid
T7	lacI	IPTG[a]

[a] *lacI* regulates expression of the T7 RNA polymerase and therefore the activity of the T7 promoter.

4). An alternative approach for induced gene expression is the use of the P_{tac}/lac repressor promoter combination. However, one drawback of this system is the leakiness of the transcriptional regulation, particularly where there are multiple copies of the plasmid-borne promoter/operator and only a single copy of the repressor gene in the chromosome. This titration of the lac repressor was first documented by Backman et al. (1976) and is a consequence of the inability of a single copy of the *lacI* gene (encoding the repressor) to produce enough repressor protein to repress multiple copies of the promoter. This can be partly overcome by using a 'promoter up' mutation ($lacI^q$) which causes the over-production of repressor protein. More recently, vectors have been described that carry a copy of the *lacI* gene as well as P_{tac}, hence maintaining the gene ratio between repressor and target promoter. These vectors give improved regulation of the tac promoter (Stark, 1987; Amann et al., 1988).

Recent developments in the area of gene specific transcription have made use of bacteriophage RNA polymerase, e.g. T7 (Table 3). The gene 1 of bacteriophage T7 encodes an RNA polymerase that is responsible for the expression of most of the T7 genome. In contrast to the *E. coli* RNA polymerase, it is a single polypeptide that recognizes a unique 23-nucleotide DNA sequence within T7 promoters. The conditional expression of the RNA polymerase gene therefore provides a system in which exclusive expression of a given gene can be achieved (Tabor and Richardson, 1985). Expression vectors comprising the foreign cDNA under the transcriptional control of a T7 promoter and with the T7 RNA polymerase gene under conditional regulation have been made (Tabor and Richardson, 1985; Studier and Moffatt, 1985, 1986). More recently, the system has been developed to give dual control of expression using the lacI repressor, by: (1) using P_{lacUV5} to control T7 RNA polymerase expression; and (2) the insertion of a lac operator sequence into a T7 promoter to control foreign gene expression (Giordano et al., 1989). Using this system an increase of more than 10^5-fold in gene expression was observed upon IPTG induction.

The termination of transcription can be achieved by using defined transcriptional terminators identified for *E. coli* operons. Although this results in the generation of well-defined mRNAs, there is little evidence to suggest that it contributes to the level of protein produced. However, one reason for employing transcription terminators was identified by Stueber and Bujard (1982). They demonstrated that uncontrolled transcription from the expression cassette into the adjacent origin of DNA replication led to plasmid instability, presumably due to interference with replication origin function. This effect was reduced by the insertion of multiple transcription terminators.

In conclusion, there is a wide choice of strong promoters for use in the expression of foreign genes, but to obtain good yields of product it

is necessary to use a system which allows a tight regulation of gene expression.

2.2. Translation of Foreign mRNA

The translation of mRNA in *E. coli* requires the recognition of the mRNA by the ribosome. This recognition is achieved by a region of homology between the 16S ribosomal RNA and a short sequence termed the ribosome binding site or Shine–Dalgarno (SD) sequence 6–10 bp upstream of the translational start on the mRNA. The consensus SD sequence is AGGAGG, although shorter versions are also functional (see Stormo, 1986, for discussion). Base pairing between the SD and 16S ribosomal RNA is the first step in translational initiation. The general rules for translational initiation are simple: (1) homology between the SD and the 3′ end of the 16S ribosomal RNA; (2) a 6–10 bp spacing between the SD and the initiation codon (AUG) at which translation starts; and (3) an A–T-rich base composition between the SD and AUG (Shepard et al., 1982). However, on many occasions the adherence to these simple rules is insufficient for the achievement of efficient translation. Translational efficiency for heterologous mRNAs has been shown to vary with distance and base pair composition between the SD and AUG (Emtage et al., 1983). In general, the spacing and nucleotide composition should be determined empirically if maximum expression levels are required. In some cases, translational problems can be correlated with mRNA secondary structure formed by regions of internal homology. Most commonly, the base pairing involves either the SD or AUG regions such that these sequences are blocked from interacting with the ribosome. However, in some cases there is no obvious mRNA secondary structure and other mechanisms must be operating to block efficient translation. The influence of mRNA secondary structure on heterologous gene expression was demonstrated by a study carried out by Wood et al. (1984) on the expression of immunoglobulin μ heavy chain. The mRNA transcripts from expression constructs pNP9, 11, 12 and 14 are shown in Figure 1 and the calculated ΔG value for the proposed secondary structures. There is a strong correlation between μ chain expression and the presence of the AUG initiation codon in a single-strand form, with the best expression level being obtained when all predicted secondary structure has been removed. This result strongly suggests the need to check for secondary structure in the SD–AUG region prior to expression analysis.

Achieving efficient translation of heterologous mRNAs has proved to be the most difficult aspect of foreign gene expression in *E. coli*, and a variety of approaches have been used to overcome instances of poor mRNA translation in which RNA secondary structure may not be the

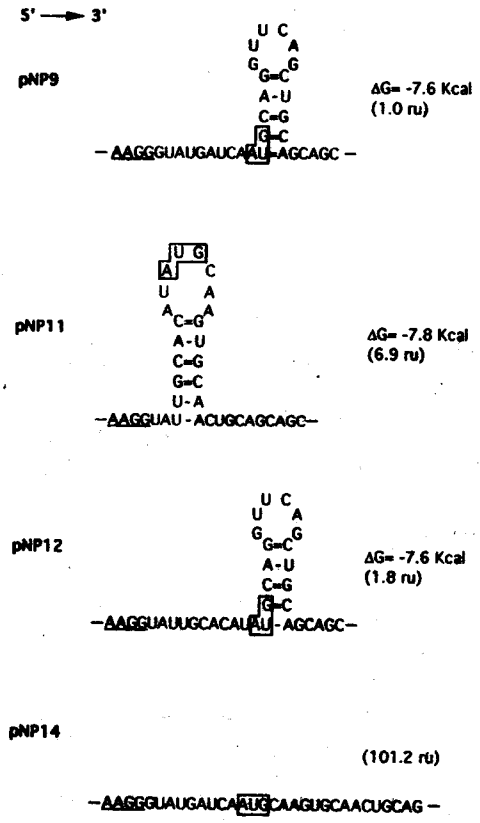

Figure 1. Possible secondary structure of μ mRNAs (see text). Sequences around the SD (underlined) and AUG (boxed) are shown. Most stable structures are identified as having the lowest ΔG value. The figures in brackets are the mean amounts of antibody produced in relative units

cause. One approach, which was successfully employed to solve the problem of bovine growth hormone (bGH) expression in *E. coli*, involves the coupling of bGH translation to that of a highly translated, short cistron which precedes the bGH sequence on an engineered polycistronic mRNA (Schoner et al., 1986; Figure 2).

For the expression of bGH, the highly translated short cistron was a 31 bp synthetic sequence. Providing that the stop codon of the first cistron is positioned correctly (i.e. in phase with the bGH AUG initiation codon and downstream of the SD of the bGH gene) then translation is increased by 20–50-fold. The mechanism by which this method of translational enhancement works is that the efficient initiation of

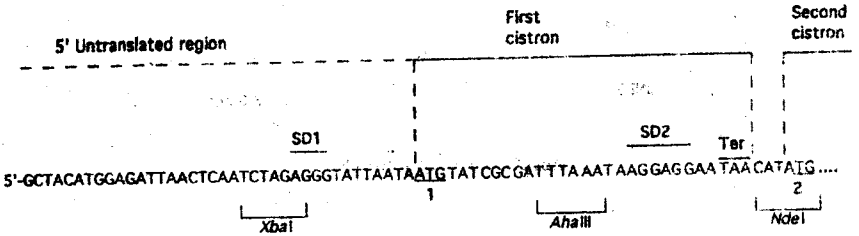

Figure 2. Two-cistron system for efficient translation using the plasmid pCZ144. The 5′ untranslated region sequence is identical to the *E. coli lpp* mRNA, and includes SD1, the ribosome binding site for the first cistron. 3′ to ATG_1 is the synthetic first cistron sequence, and SD2 the ribosome binding site for the second cistron. Termination of the translation of cistron 1 occurs shortly after the SD2 of cistron 2, and is labelled Ter (Schoner et al., 1986). A unique *NdeI* site overlapping the ATG of cistron 2 allows insertion of genes to be expressed

translation for the first cistron provides a large number of ribosomes bound to the polycistronic mRNA, and translational termination close to the SD of the second cistron allows the bound ribosomes to recognize the second SD and be positioned for initiation at the second cistron. This explains why the positioning of the first cistron stop codon relative to the SD of the second cistron is of crucial importance.

A second approach for boosting translational efficiency is to employ the translational initiation region of a highly translated mRNA, such as that for *atpE* (McCarthey et al., 1985, 1986). The *atpE* gene codes for a subunit of the *E. coli* ATP synthase (H^+-ATPase) and it has been shown that a sequence upstream of the SD of the *atpE* gene is important for enhancing efficiency of translational initiation (McCarthey et al., 1985). The relevant sequence from the mRNA of the *atpE* gene is UUUUAACU-GAAACAAA located 2–7 bp upstream of the SD, and this has been shown to have dramatic effects on the translation of interferon β (IFN-β) and IL-2 mRNAs. Expression levels were increased 6–8-fold for these two proteins when the *atpE* upstream sequence was inserted into otherwise identical expression vectors (McCarthey et al., 1986). Inducible expression vectors incorporating the *atpE* translational initiation region have been constructed and are now available (Shauder et al., 1987). The mechanism by which this enhancement is achieved is unknown, but probably involves the enhanced binding of ribosome subunits.

A third approach for boosting translational efficiency is to use the redundancy of the genetic code to vary the nucleotide sequence of the 5′ end of the gene within the coding sequence, without altering the amino acid sequence of the protein. The initial contact between mRNA