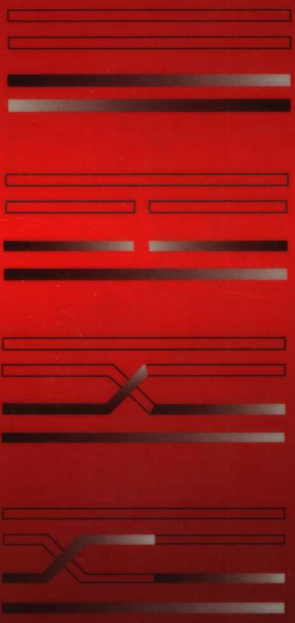
GENE TARGETING



Edited by

MANUEL A. VEGA

GENE TARGETING

Edited by

MANUEL A. VEGA

CRC Press
Boca Raton Ann Arbor London Tokyo

Library of Congress Cataloging-in-Publication Data

Gene targeting / edited by Manuel A. Vega.
p. cm.
Includes bibliographical references and index.
ISBN 0-8493-8950-X
1. Genetic engineering. I. Vega, Manuel A.
QH442.G4384 1994
575.1'0724—dc20 94-14299
CIP

International Standard Book Number 0-8493-8950-X
Library of Congress Card Number 94-14299
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Dedication

To

Lila, Manuel, Ana, Francisco,

Teresa, and Juan

Acknowledgments

The Editor specially thanks Drs. Claude Besmond and Michel Goossens (U.91-INSERM, Paris), the AFLM (Association Française de Lutte contre la Mucoviscidose), Professor Stuart G. Siddell (Wurzburg University, Germany), and SECECOM (Computer Service Center—CONICET), Bahia Blanca, Argentina.

The Editor

Manuel A. Vega, Ph.D., is chief of the Laboratory of Gene Therapy, Hospital Interzonal Dr. J. Penna, Bahia Blanca; Adjunct Professor of Medical Molecular Genetics, Universidad Nacional del Sur (UNS), Bahia Blanca; and Researcher of the National Research Council, Argentina.

Dr. Vega graduated in 1982 from the Department of Biology, UNS, and obtained his Ph.D. degree in 1987 from UNS. He did postdoctoral research at the Institut fur Virologie and Immunobiologie, Wurzburg University, Germany; at the Laboratory de Genetique Moleculaire, U.91-INSERM; Hospital H. Mondor, Creteil (Paris), France; and at the Department of Gene Therapy, TNO, Rijswijk, The Netherlands; as Wissenschaftlicher Mitarbeiter, fellow of AFLM (French Association against Cystic Fibrosis), and fellow of EMBO (European Molecular Biology Organization), respectively.

Among other awards, he has received the Dr. E. DeRobertis award from the State Secretariat for Science and Technology, Argentina, and the Dr. J.A. Balseiro award from the Forum for Science and Technology for Production from the State Ministry of Culture and Education, Argentina. He has been a recipient of research grants from the National Research Council, the Smith-Klein Beecham Foundation, the A.J. Roemmers Foundation, UNS, and private industry.

Dr. Vega has presented over 20 invited lectures at international meetings, institutes, and universities and he has published 16 research papers. His current major research interests include gene therapy for cystic fibrosis and cancer and prevention of cystic fibrosis by gamete selection.

The Contributors

Wolf M. Bertling
Paul-Ehrlich Institute
Paul-Ehrlich-Strasse
Langen, Germany

Anegela K. Cruz
Faculdade de Odontologia de Ribeira Preto
Faculdade de Medicina de Ribeira Preto
Universidade de São Paulo
São Paulo, Brazil

Jean-Louis Guenet
Unite de Genetique des Mammiferes
Institut Pasteur de Paris
Paris, France

Paul Hooykaas
Institute of Molecular Plant Sciences
Leiden University
Clusius Laboratory
Leiden, The Netherlands

Stephen C. Kowalczykowski
Division of Biological Sciences
Sections of Microbiology and of
Molecular and Cellular Biology
University of California, Davis
Davis, California

Thomas Lufkin
Brookdale Center for Molecular Biology
The Mount Sinai Medical Center
New York, New York

Remko Offringa
Institute of Molecular Plant Sciences
Leiden University
Clusius Laboratory
Leiden, The Netherlands

Manuel A. Vega
Laboratory of Gene Therapy
Department of Biology
Universidad Nacional del Sur
Bahia Blanca, Argentina

Alan S. Waldman
Department of Biological Sciences
University of South Carolina
Columbia, South Carolina

David A. Zarling
Cell and Molecular Biology Laboratory
SRI International
Menlo Park, California
Department of Laboratory Medicine
University of California, San Francisco
San Francisco, California

Table of Contents

Chapter 1

Gene Targeting

Wolf M. Bertling

Paul-Ehrlich Institute

Paul-Ehrlich-Strasse.

D-6070 Langen, Germany

Contents

0-8493-8950-X/95/$0.00+$.50

I. Introduction

For some time now, genome manipulation of pro- and eukaryotes and the technologies involved have been the focus of study in many labs. The expression and analysis of foreign DNA in prokaryotes is the most widely known application of these technologies. However, since these techniques have been improved and adjusted, they now allow for the introduction of such minor changes as point mutations in genomes as complex as the human genome. Especially in the field of research of higher eukaryotes, there have been many attempts to use our newly acquired knowledge to produce drugs in manipulated cells, to generate new models for diseases, and to, ultimately, perform somatic gene therapy.

Here we will give an introduction to recombination in general and to homologous recombination in particular. For this purpose, we will define the basic terminology and point out the importance of gene targeting in the mammalian genome. A brief inspection of the historical developments of this genetic tool, particularly of the problems that are associated with it, will lead to the description of current recombination models. This will be followed by new thoughts on the occurrence of homologous recombination events *in vivo* and their mechanisms, such as the involvement of repetitive elements and RNA in homologous recombination. We will discuss in detail the parameters of frequency and fidelity and will explain possible approaches to work around the inherent inefficiency of the process causing gene targeting. Finally, we will demonstrate the use of gene targeting for a wide variety of genes and purposes. Only seemingly of secondary importance are the different mechanisms for transporting DNA from the outside to the inside of a cell and into the nucleus. Therefore, a brief listing of current technologies to introduce DNA into the nuclei of cells in gene targeting experiments will also be presented. Since gene targeting in bacteria, lower eukaryotes, and plants will be discussed extensively in other chapters of this book, I will refer mainly to the development of gene targeting in mammalian cells.

II. Terminology

The term gene targeting refers to a targeted alteration of a specific DNA sequence in its genomic locus and occurs as a result of the homologous recombination of chromosomal and extrachromosomal sequences. Recombination, the exchange of genetic information, is the basic element of gene targeting, previously also defined as reassortment of a series of nucleotides along nucleic acid molecules.[1] Whereas in a non-homologous or illegitimate recombination, genetic elements of no significant homology or similarity recombine, homologous recombination is a process in which two DNA entities with a high sequence homology interact and recombine, i.e., exchange genetic information by adding or replacing their sequence elements. Generally, gene targeting is considered the recombinational interaction of an exogenous extrachromosomal and an endogenous chromosomal sequence by means of human manipulation. This is in contrast to the mainly naturally occurring processes of sister chromatide exchange or intrachromosomal recombination and is also different from the extrachromosomal recombination between two mostly exogenous, extrachromosomal elements. Every one of these processes can lead to either insertion or replacement of homologous sequences. Several reviews on gene targeting and homologous recombination, emphasizing different aspects, have been published in recent years.[2-13]

According to a definition of Koller and Smithies,[12] there are three possible mechanisms of a gene targeting reaction. For the shape of these molecules, specifically the mechanism, Smithies names these processes O- and Ω-type events, and two-step or in-and-out mechanism (Figure 1). Genetic information can be targeted with a single cross-over event, as first described by Hinnen et al.[14] and more extensively studied by Orr-Weaver et al.,[15] leading to an insertion of the new piece of DNA into the existing sequence. A replacement due to a double cross-over was first described by Rothstein.[16] If such a replacement construct is shorter, its integration creates a deletion, or if it is longer, it creates an insertion. If two excision cross-over events occur successively, the recombination product may have only a minute difference to the original.

Similar to the replacement process is the so-called gene conversion, which leads to the adaptation of the sequence of one strand to the sequence of another strand. Gene conversion is generally defined as locally restricted copying of genetic information from one strand to another. Therefore, it may be considered as a different form of homologous recombination, not necessarily being associated with cross-overs and, therefore, being non-reciprocal (Figure 2). The main difference is that homologous recombination products resulting from one or more cross-over events are reciprocal, because both parental double strands carry elements from another after the recombination. These two concepts of recombination are frequently associated, since both involve DNA breakage and repair and the enzymes necessary for these events.[17] Enzymatic activities involved in DNA breakage and repair were known before to occur in

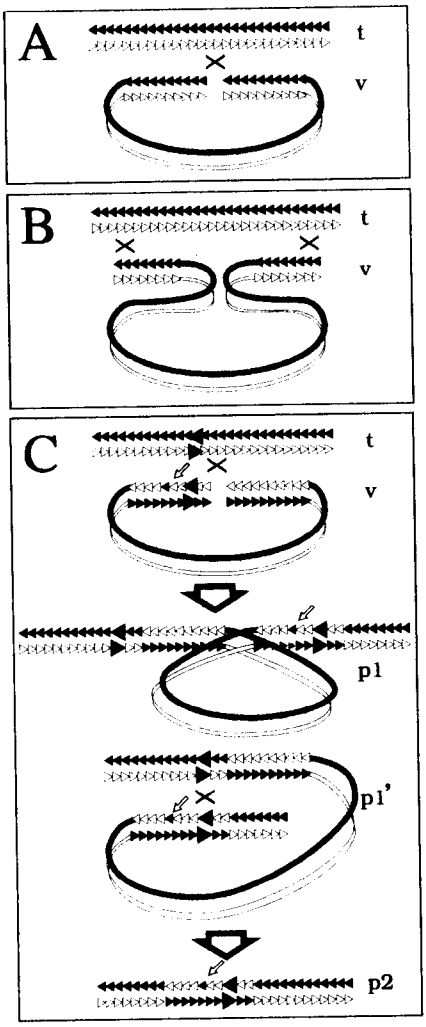


Figure 1. *Schematic drawing of three forms of gene targeting.* ***A*** *shows a single-step O-type insertion event. This type of insertion leads to a duplication of the target sequences. The chromosomal DNA ((t), upper two lanes of triangles) inserts the construct (v), which is linearized in the homologous part. The non-homologous parts are shown as black and white lines, respectively. The Ω-type replacement in* ***B*** *involves two sites of recombination. Note that no duplication of the target site results and that the orientation of the marker (in the non-homologous loop) is opposite to the product one would receive in* ***A****. In the so-called in-out targeting procedure, two successive steps are necessary (****C****). After an initial O-type recombination, a primary insertion product forms (****p1****). The large dotted triangles symbolize specific marker areas; the small dotted triangles symbolize minute sequence differences in the homologous region. The genetic marker is duplicated after the first cross-over. Note that the duplicated sequences are inverted compared to the origional. After reshaping of this product (****p1****), a new recombination leads to the desired product (****p2****) by excision cross-over. It resembles the original sequence, except for the mutation, e.g., a point mutation (small dotted triangles).*

human cells.[18,19] The phenotypical and mechanistic similarities of gene conversion and homologous recombination will be discussed in greater detail when the current models are introduced.

Different from these processes is another class of events: site specific recombination (for review, see Reference 20). Although a targeted DNA segment also recombines at a specific site, which in this case shows an extensive nucleotide sequence homology, it especially involves mechanisms that target sequences to either short specific signal sequences or even to sites where specific proteins are present. Sequence pattern recognition is used for the integration of λ in *E. coli*, as well as for the insertion of other viruses into pro- and eukaryotic genomes, for the rearrangement of vertebrate immuno-

globulines, and to a certain extent, for the movement of transposable elements. One group of transposable elements, the so-called retrotransposons, however, seems to target either sites similar to the ones that they originated from or sites where specific proteins such as RNA polymerases or DNA polymerases and related factors bind to DNA.

Due to the low complexity of bacterial genomes, many basic mechanisms and factors involved in recombination have been identified and studied in bacteria, especially *E. coli*,[2] where the majority of these proteins has been characterized. Much of the information currently available about the genetics of eukaryotic recombination has been generated in lower eukaryotes such as yeast and other low fungi. This information is now used to explore new ways to allow for the genetic manipulation of mammals, an admittedly anthropocentric goal.

III. Historical developments and current recombination models

Relatively early in the 1900s, recombination was seen as a means to generate species diversity, although the elements to comprise the genetic information were unknown.

In the 1960s and 1970s, other terms for homologous recombination had been introduced, such as general, generalized, chromosomal or equatorial.[21] Some of these terms, such as normal or non-specific,[22-25] would be considered misleading nowadays, considering that, in gene targeting attempts, the majority of constructs insert at random sites[7] rather than at the predetermined, homologous site. So it is not normal to detect a homologous recombination event in which an exogenous construct integrated at a specific site, the homolo-

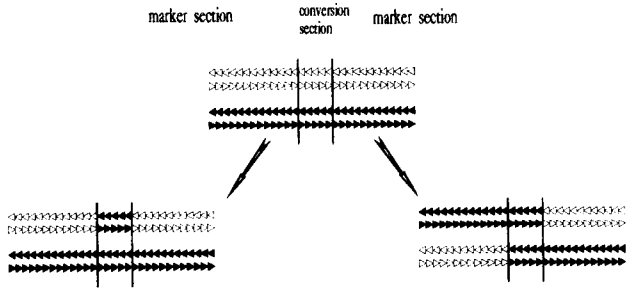


Figure 2. *The principle of gene conversion events with or without accompanying cross-over. Marker sections of two homologous "white" (target) and "black" double-stranded DNAs are separated by vertical lines from the section of conversion. The non-reciprocal product on the lower left was not subject to a cross-over, whereas a reciprocal exchange accompanied the conversion event of the product on the lower right.*

gous region. However, the situation is different *in vivo*, and interchromosomal recombination, for instance, is primarily homologous.

In 1928, Griffith[26] observed genetic recombination in bacteria and termed it transformation. In the early 1960s, the selection process of a homologous recombination reaction was first attributed to basepairing of complementary single strands to generate a heteroduplex region near the point of exchange.[27,28] This led to the development of molecular models for the mechanism underlying recombination, which is discussed in detail in Chapter 2.

A. Conventional models

The oldest and most significant model to describe homologous recombination is the Holliday model (Figure 3). A so-called Holliday structure, formed as an intermediate of the recombination process, resolves either with or without cross-over events. Fluorescence energy transfer studies indicated that the four-way DNA junction in such a Holliday structure is a right-handed cross of antiparallel molecules.[29]

This model uses the same basic structure to explain gene conversion and cross-over-driven homologous recombination. The heteroduplex in the case of gene conversion has to be subjected to repair enzymes to clear mismatched bases. So both concepts ask for some cutting and religation activities acting on the DNA strands. If such a heteroduplex is transient and moves along the DNA, which may resolve with or without a preceding cross-over, we are following Sobell's model.[30] A very important contribution to models of recombination comes from Radding[31] (for review, see References 32 and 33) (Figure 1) and is introduced in detail in Chapter 2. Nicks on two homologous double-stranded molecules initiate the strand exchange of homologous duplex DNAs. The subsequent resolution of intermediate Holliday structures leads to newly arranged heteroduplexes. This model and its variations base on the discovery of recA in *E. coli*.[34,35] This protein promotes the pairing of homologous DNA molecules.[36] Meanwhile, equivalent proteins and activities have also been discovered and described in lower and higher eukaryotes.[37-40] According to the Meselson/Radding model, which is also called the Aviemore model,[31] recA promotes a pairing of an at least partially single-stranded DNA with duplex DNA in three steps. In a presynaptic phase, the protein binds to single-stranded DNA and polymerizes to a nucleoprotein filament. In the following synaptic phase, this filament binds non-specifically to duplex DNA and migrates along the DNA until a homologous region is found. *In vitro* this homologous pairing is initiated by DNA synthesis.[41] In the last step, the actual strand exchange occurs.

Another, variant model to explain homologous recombination is the double strand break model, initially mentioned by Resnick[42] and later modified by others[43-46] (Figure 4). A model that leaves one recombinant duplex, and as an intermediate, two DNA ends (which have to be closed in a second round of half cross-over), is called sequential half reciprocal recombination[5] (Figure 5).

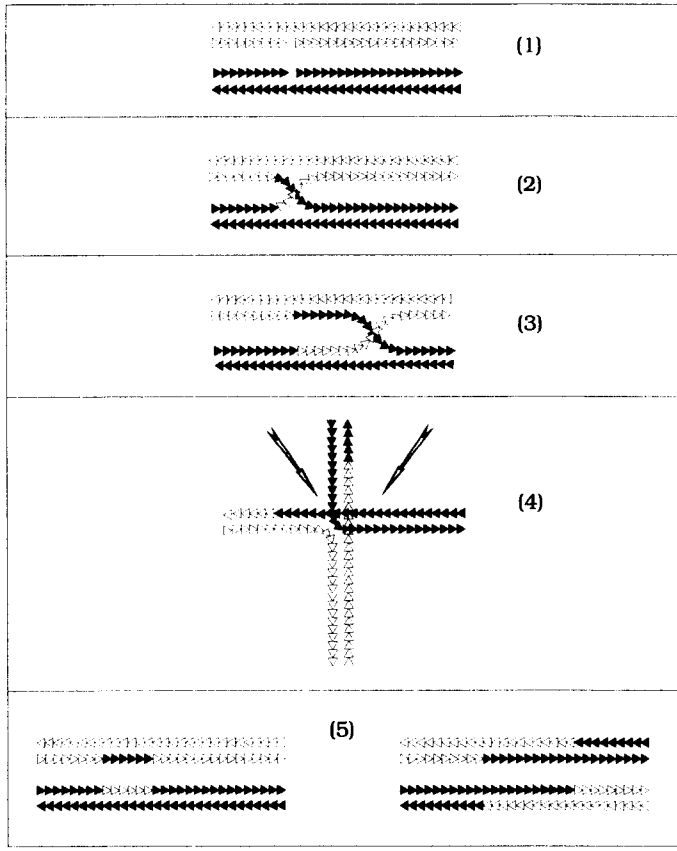


Figure 3. *The Holliday model. An internal nick* ***(1)*** *in the two participating double-stranded molecules (white and black triangles) leads to a local strand exchange* ***(2)****. This point of exchange can migrate (branch migration)* ***(3)****. A different view of this migration product* ***(4)*** *shows how this Holliday structure can be resolved. The arrows point to one set of potential cutting and rejoining sites. The two reciprocal products that result are given in [(5) left, right]. Using alternate cleaving and rejoining sites, by twisting the molecule* ***(4)****, yields a different set of products. Further processes will repair differences in complementary strands.*

The double strand break model in fungi[43,47] and prokaryotes[2] has been reviewed extensively and is in its early development based on work of Stahl[48,49] in prokaryotes, in yeast,[15,50] and in higher eukaryotes.[42] An exonuclease cleaves DNA; another enzyme has DNA annealing activity. The initial breakpoint is expanded to a gap. These ends invade a homologous part of the DNA and are used either as primers for synthesis or trigger an exchange, leaving the other ends reactive and causing a cross-over.

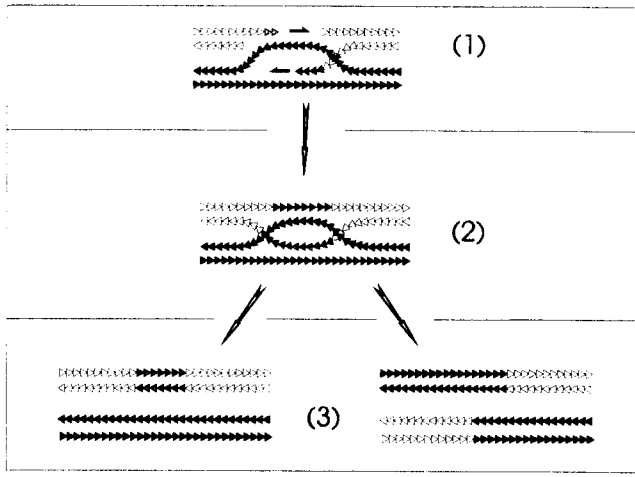


Figure 4. *The double strand break model for homologous recombination. According to this model the exchange of genetic information begins with a double strand break of one partner (white triangles in (1)), which is expanded to a gap by exonucleases. The following strand displacement (black triangles in (1)), enables the reconstitution of that part of the sequence lost in the gap by DNA polymerase I extension. After ligation to their original strands a double Holliday structure (2) results. The resolution of this Holliday structure can [(3) right] or cannot [(3) left] involve cross-over.*

Currently, there are two strand break models. In one model, a single-strand cut is introduced in each DNA duplex. The strands are exchanged to form a Holliday structure, and this structure is then resolved by cleavage and exchange of the other pair of strands.[51,52] Depending on the direction and extent of strand rotation preceding the resolution of the Holliday structure, these exchanges of single strands can lead to different products. This model bears resemblance to the half crossing over model,[5] which also leaves one or two ends at the end of the first round after two parental duplexes generate one recombinant duplex and reactive ends, which initiate a second round of half crossing over. If one of these ends engages in a second site cut during the cross-over at a short distance from the first exchange, the product resembles a gene conversion, and the half cross-over results then in a half reciprocal recombination. There are actually situations when one round is sufficient, such as in the case of telomere conversion in Trypanosoma.[5,53]

In the second model, double strand cuts are made at both parental DNAs, and double strands are exchanged by single rotation before resolution of the intermediate Holliday structure and ligation.[51]

B. Novel mechanisms

Recently a three-stranded intermediate structure[54] (Figure 6) was proposed by several groups[55] to facilitate homologous recombination. In this structure,

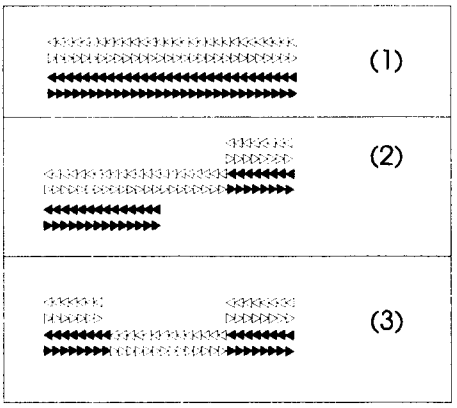


Figure 5. *The principle of the half crossing-over model. Two intact double-stranded DNA molecules* ***(1)*** *lead to one intact recombinant molecule and to two ends* ***(2)****. The ends engage in a second half cross-over at a different site, probably due to exonucleolytic digestion of free DNA ends in* ***(3)****. The resulting double-strand gap in the "white" DNA molecule can be repaired in different ways (see also Figure 4).*

the invading single-stranded DNA is interacting with the major groove of the duplex DNA. Homology-specific strands in the triple helix are in parallel orientation. Four-stranded DNA could also be involved in the pairing of recombination.[55]

There are also other mechanistic possibilities leading to homologous recombination and/or gene conversion that should be considered. A recombination of RNA with RNA during viral replication, or by using the splicing apparatus or the recombination of RNA with DNA during reverse transcription of, e.g., endogenous retroviruses and retroposons, could lead to homologous recombination or gene conversion. RNA could also be used as template to close DNA double strand breaks in analogy to DNA as depicted in Figure 6.

Generally, it is assumed that recombination, particularly homologous recombination, is taking place among DNA molecules. However, a number of observations indicate that RNA is also a suitable template for recombination. Obviously, a recombination of RNA molecules occurs during splicing, although this does not require homologies of the newly joint products either for the normal *cis*-splicing or for the seldom observed *trans*-splicing.[56,57] Mechanisms of homologous recombination among RNAs have recently been described during reverse transcription of retroviral RNA.[58] It is assumed that the reverse transcriptase switches templates during the synthesis of the minus strand, when the newly generated minus-strand DNA has a sufficient homology to the plus-strand template. The argument is that the recombination occurs only after packaging of these different RNAs into viral particles. However, a direct recombination between double-stranded RNAs by a mechanism similar to those models introduced for DNAs (see above) would need complementary