Topics in Current Genetics 23

Sylvaine Renault
Philippe Duchateau *Editors*

Site-directed Insertion of Transgenes

Topics in Current Genetics **23**

Series Editor: *Stefan Hohmann*

For further volumes:
http://www.springer.com/series/4735

Sylvaine Renault • Philippe Duchateau

Site-directed Insertion of Transgenes

Editors
Sylvaine Renault
UFR des Sciences et Techniques et des Sciences Pharmaceutiques
Université François Rabelais
Tours, France

Philippe Duchateau
Cellectis SA
Paris, France

ISSN 1610-2096 ISSN 1610-6970 (electronic)
ISBN 978-94-007-4530-8 ISBN 978-94-007-4531-5 (eBook)
DOI 10.1007/978-94-007-4531-5
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012943827

© Springer Science+Business Media Dordrecht 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Prologue

Advances in Chemistry have revolutionized the past century, and with the advances of the post-genomic era we can already envision the twenty-first century as the era of the Biology. Sequenced genomes, annotated genes as well as specific mutations are now available for many organisms. Thus reverse genetics is entering its golden age. These tremendous sources of data have brought new challenges and opportunities in all fields of the biology, and new tools to introduce genetic modifications in complex organisms have become a necessity. Simple analysis of gene sequences may allow for the prediction of specific protein domain bearing particular activity. However, most of the time, this approach is limited by the sequence homology search and do not permit precise characterization of the protein function. It is always necessary to study a newly identified gene sequence (or gene modification) within its biological environment i.e. a living cell or organism. In this context, transgenesis appears to have a major role to play going from the understanding of gene function to more applied aspects such as therapeutic purposes.

Transgenic complementation (i.e. introduction of exogenous coding sequence in the genome) has been the historical approach and is still an initial step that helps to elucidate the function of a gene. However, it is limited by several factors as copy number, site(s) of integration, gene expression or extinction due to neighboring sequences. In the early 1980s, pioneer works performed by Smithies and Capecchi laboratories, have paved the way for targeted gene modification in higher organisms. Since then, this technique has proven to be robust, and today homologous recombination based strategies have become a routine technique to modify mouse genome. Depending on the vector design, genomic sequences can be either replaced or deleted or exogenous sequences can be inserted. However, the frequency of targeted events recovered is quite low, and most of the time additional strategies using positive/negative selection have to be used.

Over the years, gene targeting methods have been refined and new techniques allowing precise site-directed genome modifications have been developed, offering a large palette of tools to scientists desirous to express and study its favorite gene. This book will try to summarize this evolution and will present the main strategies that one can follow to perform site-directed insertion of transgenes.

In Chaps. 1 and 2, the readers will find an overview of the main technologies used today to deliver a transgene into a genome such as free DNA, viruses or transposon. The consequences of the random integration inherent in non-targeted approaches i.e. insertional mutagenesis and inhibition of transgene expression by epigenetic regulation are also addressed. Moreover Chap. 1 also proposes solutions to allow long-term expression of transgenes by the use of insulators surrounding the transgene or by a better choice of the targeted site of integration. Over the years many tools have been developed to promote targeted gene insertion. Chapter 2 will summarize the different strategies available today going from homologous recombination (HR) induced or not through the use of nucleases (zinc-finger nuclease, meganucleases or TAL nucleases) to natural site-specific recombinases like Flp, Cre or φC31 integrase or modified recombinases like transposases and resolvases. All subsequent chapters will then allow the reader to go deeper into each strategy at his will.

Chapters 3, 4, 5 and 6 are dedicated to "Integration based on homologous recombination". Historically, the yeast *Saccharomyces cerevisiae* has been one of the earliest model used to decipher HR mechanisms and consequently to perform gene targeting. Chapter 3 presents the history and the principle of the use of homologous recombination to modify "à façon" a gene of interest within the yeast genome. The knowledge gained using this model organism provided the basis for gene targeting in mammalian cells. For the sake of clarity, Chap. 4 describes in details the different models of homologous recombination occurring in mammalian cells followed by the description of the different crucial steps of HR as well as their regulation along the cell cycle. Sequence-specific endonucleases induce DNA breaks at a precise locus within a genome and thus initiate homologous recombination at that exact position. Today three groups of endonucleases have emerged as tools for genome engineering. Zinc-finger nucleases (ZFN), meganucleases and TALE nucleases (TALEN) are presented in Chaps. 5 and 6. In these chapters, the history of development of each technology is described. In Chap. 5, the strategies developed to engineer ZFN proteins with better affinity and specificity are presented as well as the most recent successes of this technology. The first double-strand break-induced gene targeting has been achieved with the meganuclease I-SceI. Chapter 6 depicts historic milestones leading to the success of this technology as well as the advances in meganucleases engineering and their uses for site-directed genome modifications. The reader will also find within this chapter the recent developments of transcription activator like effector proteins which hold great promises for targeting transgenes.

The third part of the book, comprising Chaps. 7, 8, 9 and 10, is dedicated to the "Integration based on site-specific recombination". Recombinases are widely used technology for transgenesis and represent an attractive alternative tool for genome engineering purposes. Recombinase systems such as Cre/loxP, Flp/FRT and φC31/attL-attP sites are presented in Chaps. 7 and 8 while data on engineered recombinases such as transposases and resolvases used to target transgenes in specific sites are described in Chaps. 9 and 10. More precisely, readers will find detailed description of the mechanism of recombination via the well known Cre and Flp recombinases in Chap. 7 as well as different strategies (including RMCE) to modify mouse and

human pluripotent cells. The history, the mechanism and the potential therapeutic applications of the φC31 integrase are presented in Chap. 8, while Chap. 9, dedicated to the widely used transposases technology, exposes the strategies developed today to carry out site-specific insertions of transgenes as well as their use in different organisms from bacteria to mammalian cells. Targeting recombinase activity to a specific genomic sequence is a promising approach. Chapter 10 describes with a special emphasis on zinc-finger recombinase the recent advances in this field that will help the reader to understand the strengths and weakness of this technology.

Finally, the fourth part of this book entitled "Applications of Site-Directed Integration of Transgenes" summarizes data available today obtained with these different technologies in plants (Chap. 11), *Drosophila* (Chap. 12) and mammalian cells (Chap. 13). Chapter 11 presents a broad review of different approaches to deliver molecules such as T-DNA and oligonucleotides, the methods of selection of targeted events and the use of meganucleases, ZFN and TALEN to perform site-directed integration of exogenous sequences. Chapter 12 is dedicated to the site-specific modification of the genome of *Drosophila* using φC31 integrase, in situ generation of linear donor DNA through ZFNs. Examples of modifications for different applications are described. Chapter 13 focuses particularly on the use of ZFN and TALEN for the genome modification of human cells such as gene disruption, gene correction and gene addition. Obstacles and safety concerns for genome engineering are discussed in this chapter. Finally, Chap. 14 reviews all applications of genome modifications, in cells and animal models. It covers topics such as functional genomics, drug discovery, bioproduction, cell transformation, molecular and microbiology tools, via the use of insertional or site-directed mutagenesis and knock-down, conditional or constitutive knock-out and knock-in.

Genome engineering is a fast growing field with numerous branching going from basic research to therapeutic purposes. Altogether this book tries to present and discuss the most relevant information available today regarding the main technologies used in this field. We hope that it will trigger reader's interest and help scientists to better understand each technology.

Sylvaine Renault
Philippe Duchateau

Contents

Part I Site-Directed Integration of Transgenes

1 **Transgene Site-Specific Integration: Problems and Solutions** 3
Solenne Bire and Florence Rouleux-Bonnin

2 **Designing Non-viral Targeted Integrating Vectors for Genome Engineering in Vertebrates** 41
Ludivine Sinzelle and Nicolas Pollet

Part II Integration Based on Homologous Recombination

3 **Gene Targeting and Homologous Recombination in *Saccharomyces cerevisiae*** 71
Judith Miné-Hattab and Rodney Rothstein

4 **Homologous Recombination in Mammals** 91
Aurélia Barascu, Anastazja Grabarz, and Bernard S. Lopez

5 **Engineered Zinc Finger Nucleases for Targeted Genome Editing** 121
Cherie L. Ramirez and J. Keith Joung

6 **Engineered Meganucleases for Genome Engineering Purposes** 147
Jean Charles Epinat, George H. Silva, Frédéric Pâques, Julianne Smith, and Philippe Duchateau

Part III Integration Based on Site-Specific Recombination

7 **Cre/loxP, Flp/FRT Systems and Pluripotent Stem Cell Lines** 189
Candice G.T. Tahimic, Kenji Sakurai, Kazuhiro Aiba, and Norio Nakatsuji

8 Site-Specific Recombination Using PhiC31 Integrase 211
Jonathan M. Geisinger and Michele P. Calos

9 Modified Transposases for Site-Directed Insertion of Transgenes 241
Sean Colloms and Sylvaine Renault

10 Targeted Plasmid Integration into the Human Genome by Engineered Recombinases 267
Charles A. Gersbach and Carlos F. Barbas III

Part IV Applications

11 Gene Site-Specific Insertion in Plants 287
Olivier Da Ines and Charles I. White

12 Site-Directed Insertion of Transgenes in *Drosophila* 317
Dana Carroll

13 Nuclease Mediated Targeted Genome Modification in Mammalian Cells 327
Jenny C. Barker, Richard A. Voit, and Matthew H. Porteus

14 Bio-applications Derived from Site-Directed Genome Modification Technologies 353
Christophe Delenda, Sébastien Paris, Sylvain Arnould, Edward Balbirnie, and Jean-Pierre Cabaniols

Index 385

Contributors

Kazuhiro Aiba Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Sakyo-ku, Kyoto, Japan

Sylvain Arnould Cellectis Therapeutics, Paris, France

Edward Balbirnie Ectycell, Evry, France

Aurélia Barascu Laboratoire Stabilité Génétique et Oncogenèse CNRS, UMR 8200 GR – PR2, Université Paris Sud, Villejuif, France

Institut de radiobiologie cellulaire et moléculaire, Institut de Cancérologie Gustave-Roussy CNRS, UMR217, Fontenay aux Roses, France

Carlos F. Barbas III The Skaggs Institute for Chemical Biology and the Departments of Molecular Biology and Chemistry, The Scripps Research Institute, La Jolla, CA, USA

Jenny C. Barker Medical Scientist Training Program, UT Southwestern Medical Center, Dallas, TX, USA

Solenne Bire UFR des Sciences et Techniques, Université François Rabelais GICC, UMR CNRS 6239, Tours, France

Jean-Pierre Cabaniols Cellectis Bioresearch, Paris, France

Michele P. Calos Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

Dana Carroll Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT, USA

Sean Colloms Institute of Molecular Cell and Systems Biology, University of Glasgow, Glasgow, Scotland, UK

Christophe Delenda Cellectis Bioresearch, Paris, France

Philippe Duchateau CELLECTIS SA, 8 rue de la Croix Jarry, 75013 Paris, France

Jean Charles Epinat CELLECTIS SA, 8 rue de la Croix Jarry, 75013 Paris, France

Jonathan M. Geisinger Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

Charles A. Gersbach Department of Biomedical Engineering, Duke University, Durham, NC, USA

Anastazja Grabarz Laboratoire "Biologie cellulaire de l'intégrité du génome", Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

Olivier Da Ines Génétique, Reproduction et Développement, UMR CNRS 6247 – Clermont Université – INSERM U931, Centre National de Recherche Scientifique/ Université Blaise Pascal, 24 avenue des Landais, BP 80026, 63171 Aubière CEDEX, France

J. Keith Joung Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA

Bernard S. Lopez Laboratoire Stabilité Génétique et Oncogenèse CNRS, UMR 8200 GR – PR2, Université Paris Sud, Villejuif, France

Institut de radiobiologie cellulaire et moléculaire, Institut de Cancérologie Gustave-Roussy CNRS, UMR217, Fontenay aux Roses, France

Judith Miné-Hattab Department of Genetics and Development, Columbia University, New York, NY, USA

Norio Nakatsuji Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan

Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Sakyo-ku, Kyoto, Japan

Frédéric Pâques CELLECTIS SA, 8 rue de la Croix Jarry, 75013 Paris, France

Sébastien Paris Cellectis Bioresearch, Paris, France

Nicolas Pollet Institute of Systems and Synthetic Biology, Genopole, CNRS, Université d'Evry Val d'Essonne, Evry, France

Matthew H. Porteus Department of Pediatrics, Stanford University, Stanford, CA, USA

Cherie L. Ramirez Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA

Sylvaine Renault UFR des Sciences Pharmaceutiques et des Sciences et Techniques, EA Innovation moléculaire et thérapeutique, Tours, France

Rodney Rothstein Department of Genetics and Development, Columbia University, New York, NY, USA

Florence Rouleux-Bonnin UMR CNRS 7292, GICC, Equipe LNOx "Niche leucémique & métabolisme oxydatif", Université François Rabelais, UFR de Médecine, Tours, France

Kenji Sakurai Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan

Biologics Research Lab, Kyowa Hakko Kirin Co. Ltd, Machida, Tokyo, Japan

George H. Silva CELLECTIS SA, 8 rue de la Croix Jarry, 75013 Paris, France

Ludivine Sinzelle Institute of Systems and Synthetic Biology, Genopole, CNRS, Université d'Evry Val d'Essonne, Evry, France

Julianne Smith Cellectis Therapeutics, Paris, France

Candice G.T. Tahimic Endocrine Research Unit, San Francisco Veterans Affairs Medical Center, Building 6 Room 312, 4150 Clement St., San Francisco, California 94121, USA

NCIRE/San Francisco Veterans Affairs Medical Center, San Francisco, CA, USA

Richard A. Voit Medical Scientist Training Program, UT Southwestern Medical Center, Dallas, TX, USA

Charles I. White Génétique, Reproduction et Développement, UMR CNRS 6247 – Clermont Université – INSERM U931, Centre National de Recherche Scientifique/ Université Blaise Pascal, 24 avenue des Landais, BP 80026, 63171 Aubière CEDEX, France

Part I
Site-Directed Integration of Transgenes

Chapter 1
Transgene Site-Specific Integration: Problems and Solutions

Solenne Bire and Florence Rouleux-Bonnin

Abstract Integrative gene transfer performed by viral and non-viral vectors have demonstrated their effectiveness, but have been linked to some adverse events, such as clonal expansion and tumorigenesis. These observations have raised serious concerns about the safety of gene transfer methods, and have led to many attempts to find new solutions. In this chapter, we summarize the major problems encountered with viral and non-viral-vectors and various ways of avoiding insertional mutagenesis, the induction of innate immunity and transgene silencing are described. We also list the main strategies for optimizing vector architecture so as to ensure safe and long-term expression of the transgene. Several new approaches have succeeded in targeting transgene integration to a specific locus using recombinases, homing endonucleases, zinc finger nucleases, integrases and transposases or a combination of them. Here, we report various criteria that can be used to define what is a good insertion site in the human genome.

Keywords Transgene targeting • Site specificity • Insertional mutagenesis • Transgene silencing • Safe harbor locus • Genotoxicity

Abbreviations

ADA Adenosine deaminase deficiency
AAV Adeno-associated virus
CIS Common insertion site

S. Bire
UFR des Sciences et Techniques, Université François Rabelais GICC,
UMR CNRS 6239, Parc de Grandmont, 37200 Tours, France

F. Rouleux-Bonnin (✉)
UMR CNRS 7292, GICC, Equipe LNOx “Niche leucémique & métabolisme oxydatif”,
Université François Rabelais, UFR de Médicine, Tours, France

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_1,
© Springer Science+Business Media Dordrecht 2013

DBD	DNA binding domain
DSB	Double-strand break
dsRNA	double-stranded RNA
DTS	DNA nuclear targeting sequence
EF1a	Elongation factor-1a
ES	Embryonic stem cells
iPS	induced pluripotent stem
IRES	Internal ribosome entry sequence
HAT	Hypoxanthine aminopterin, and thymidine
LAM-PCR	Linear amplification-mediated PCR
LINEs	Long interspersed nuclear elements
LTR	Long terminal repeat
LV	Lentiviral vector
HE	Homing endonuclease
HSS	DNAseI-hypersensitive site
HR	Homologous recombination
miRNAs	microRNAs
MMLV	Moloney murine leukemia virus
NHEJ	Non-homologous end joining
NoLS	Nucleolar localization sequence
nrLAM-PCR	Non-restrictive LAM-PCR
PB	PiggyBac
PGK	Phosphoglycerate kinase
PRRs	Pattern-recognition receptors
RISC	RNA Induced Silencing Complex
γRV	Gammaretroviruses
SB	Sleeping Beauty
SIN	Self-inactivating retrovirus
siRNAs	small interfering RNAs
TFBS	Transcription factor binding site
TLR	Toll-like receptor
TSS	Transcription start site
TIP-chip	Transposon insertion site profiling chip
ZFN	Zinc-finger nuclease
ZFP	Zinc-finger protein
ZFR	Zinc-finger recombinase

1.1 Introduction

The basic principle of gene therapy is based on the genetic modification of somatic cells. This is achieved by inserting a copy of a therapeutic gene in order to restore the proper expression and function of a damaged gene, or by correcting nucleotide

mutations by homologous recombination. Gene therapy protocols should meet several specifications: they must (i) deliver the therapeutic genes with high efficiency, specifically into the relevant cells, (ii) be adaptable to changing needs in terms of vector design, (iii) minimize the risk of genotoxicity, and (iv) be cost-effective. Gene-based technologies have undergone massive development during the past decade. Gene transfer vector engineering aims to procure a secure and lasting tissue-specific expression of the therapeutic gene. These objectives will be dependent on and determined by the type of vectors, the number of gene-transfer particles to which the cells are exposed, the copy number of transgenes per cell, the transcription rate, the efficiency of RNA processing, features of the protein such as activity or stability, the population size of the targeted cells and the life span of the modified cells.

Many protocols use gene vectors that maintain their episomal status as extrachromosomal self-replicating systems in post-mitotic cells (Deyle and Russell 2009; Wade-Martins 2011). However, use of this kind of vectors is limited, and more work is required; in particular with regard to how the transgene is transmitted and maintained during cell division. Transgene integration into the chromosomes therefore appears to be a more useful approach. So far, two well-characterized integrative systems have been described that rely on viral and non-viral vectors respectively. First, natural elements, such as viruses and more recently non-viral transposable elements, have been turned into gene vehicles. However, use of the first generation of viral vectors for gene therapy has been associated with uncontrolled cell proliferation. Along the way, we have gained further insights into the effects of transgene insertion using these two kinds of vector. These issues involve the efficiency of delivery, the host's immune responses towards the transgene and/or its product, insertional mutagenesis and also transgene remobilization and postintegrative promoter silencing.

To avoid these serious adverse effects, the "holy grail" of gene therapy is to improve the vector design in such a way as to ensure both site-directed integration in a safe locus and long-term expression of the transgene. The questions that remain include: what is an ideal vector and what defines a good integration site.

1.2 Random Integration and Clonal Dominance: Reality or Myth?

The first strategies developed to integrate a gene of interest involved turning infectious agents into therapy vehicles (Kay et al. 2001). For instance, viral particles can encapsulate a modified genome carrying a therapeutic gene in place of their own genome. The modified transducing particles still retain their ability to penetrate effectively into the targeted cells to introduce functional genetic information, but are no longer infectious.

Among viral vectors, retroviruses have been extensively used for gene therapy applications. Although they were first thought to insert randomly into the host genome, preferential integration loci have now been well established by several studies (Kustikova et al. 2010). Initially insertional mutagenesis was thought to be negligible, but since 2002 the thinking has changed as a result of the successful pioneering work carried out to treat SCID (X-linked severe combined immunodeficiency) with the murine leukemia virus-derived vector (MLV). This work highlighted a potential link between integration of the transgene, insertional activation of the LMO2 proto-oncogene, and clonal dominance inducing leukemogenesis (Hacein-Bey-Abina et al. 2003a, b).

Clonal dominance occurs when a population of modified cells becomes prevalent due to dysregulation of one or more genes by insertional mutagenesis. Generally, dysregulation of these genes affects their self-renewal or competitive fitness, conferring on them the advantage of stronger growth than other clones. However, the situation is not quite so simple, since clonal dominance is not always synonymous with malignancy. In a study of clinical gene transfer to treat chronic granulomatous disease a clone harboring an insertion in the SETBP1 gene that is implicated in the proliferation of leukemia cells was detected (Glimm et al. 2005). These gene-corrected neutrophils cells were viable and functional, but they progressively increased in number after transplantation, and then stabilized. Further, the clone has not expanded after 7 years of follow up (Ott et al. 2006). Even though the trial was a success it is possible that, in the absence of such clonal expansion, there would not be enough gene-corrected cells to achieve a sustained therapeutic effect (Naldini 2006). Similarly, Recchia et al. in 2006, showed that even though one fifth of the integrations achieved using a retroviral vector in T cells affect the expression of nearby genes, no clonal selection has been observed up to 9 years after T cell reimplantation (Recchia et al. 2006). Kustikova et al. (2005) have shown that a single integration in the Mds1-Evi1 locus encoding zinc finger transcription factors, led to long-term *in vivo* clonal dominance. Although this single clone dominance has not turned leukemic, it has been associated with poor transgene expression (Métais and Dunbar 2008). Moreover, clones that dominated hematopoiesis as a whole for a period of time, eventually disappeared (Ott et al. 2006; Kustikova et al. 2008).

It is noteworthy that the distribution of insertions in long-term reconstituting cells significantly differs from that in freshly-transduced cells, suggesting that *in-vivo* selection has occurred. Current data as a whole do not support the idea that induced clonal dominance is the first step in inadvertent malignant transformation. The induction of cancer as a result of insertional mutagenesis in gene therapy is thought to be a multistep process that requires specific forms of oncogene collaboration (Hanahan and Weinberg 2000). Nevertheless, concerns still persist about the influence of the vector used upon the host genome and the potential genotoxic risks.

1.2.1 Viral Vectors and Their Integration Profiles

Virus-derived vectors display differing integration preferences. Understanding the molecular mechanism of natural, site-directed integration is crucial if we are to design safer vectors, especially if intragenic or regulatory sequences are targeted. For example, gamma retroviruses (γRVs) preferentially target regulatory sequences with a propensity for growth-regulating genes. They are also dependent on the cell cycle to gain access to chromosomes (Cattoglio et al. 2007). In contrast, MLV integration events occur only after mitosis and preferentially around transcription start sites (TSS) and CpG islands. In these regions, the transcriptional enhancers contained in the viral long terminal repeats (LTRs) are likely to interfere with gene regulation (Wu et al. 2003). The human immunodeficiency virus (HIV) and other lentiviral vectors (LVs) perform insertions at any time during the cell cycle, preferentially within active transcription units (Schroder et al. 2002; Hematti et al. 2004). Much attention has been paid to γRVs, since they show a 10-fold greater propensity to insert near proto-oncogenes than lentiviral vectors (Montini et al. 2009). In fact, the choice of the integration site is partly due to the enzyme that catalyzes the integration process. In 2009, Felice et al. proposed that LV integrase plays a crucial role in directing retroviral integration towards regions of the genome containing high levels of transcription factor binding sites (TFBSs) (Felice et al. 2009). The hypothesis they propose is that transcription factors bound to the viral U3 enhancer cooperate with the integrase in directing pre-integration complexes towards regulatory regions actively engaged in the transcriptional machinery. Indeed, proteins interacting with HIV integrase include components of the SWI/SNF chromatin-remodeling (Lesbats et al. 2011), DNA-repair complexes (Yoder et al. 2011) and Polycomb-group proteins (Violot et al. 2003). Moreover, the cellular lens epithelium-derived growth factor (LEDGF/p75) binds HIV integrase, and is partially responsible for promoting integration into genes (Ciuffi et al. 2005; Marshall et al. 2007).

Evidence supporting the idea that the integrase determines the selection of the integration site comes from the creation of chimeric viruses. In 2006, Lewinski et al. found that transferring the MLV integrase coding region into HIV makes the hybrid integrate with specificity similar to that of MLV (Lewinski et al. 2006). On the other hand, LV with chimeric γRV LTR results in strong, dose-dependent acceleration of tumor onset, as observed for its γRV counterpart (Montini et al. 2009). So, integrase and LTR mainly determine the bias of viral vector integration profiles. Furthermore, virus-based vectors are known to induce immune activation against viral particles, the modified cells, and the transgene product. Their production for use in clinical trials is still subject to technical and regulatory hurdles. Consequently, non-viral integrative vectors have been developed as an alternative.

1.2.2 Non-viral Vectors and Their Integration Profiles

Non-viral vectors are considered to be less immunogenic because they do not have a biological capsule. In addition, clinical-grade stable DNA plasmids (pDNA) can be easily prepared at lower cost. However, non-viral gene transfer systems necessitate

synthetic delivery systems to achieve cellular uptake. Chemical transfection reagents such as cationic lipids or cationic polymers are widely used to enable them to enter cells (reviewed in Pichon et al. 2010), but some physical methods such as electroporation or micro-injection have also been used (reviewed in Suzuki et al. 2011; Niidome and Huang 2002). Progress is continually making these methods more effective for delivering nucleic acids.

Non-viral vectors also raise the issue of stable integration, because plasmids carrying the gene of interest are rapidly diluted and/or degraded in dividing cells, unless the enzymes that catalyze integration events, such as recombinases, integrases, or transposase, are provided. DNA transposable elements are natural genetic elements that move from one chromosomal location to another *via* a conservative cut-and-paste mechanism. They are composed of an ORF coding for the transposase, and flanked by inverted terminal repeats. These latter are recognized and fixed by the transposase, which then catalyzes the excision and integration of the transposon. This process is the key feature that makes natural DNA transposons particularly attractive as potential gene delivery tools. Indeed, the molecules that support the integrative process are relatively easy to engineer since only two separate plasmids are co-delivered to the cells. One plasmid carries the transposon (donor plasmid), in which the original transposase gene has been replaced by a transgene of interest driven by an appropriate promoter, and the other plasmid carries the transposase expression cassette (helper plasmid). Alternatively, the transgene and the transposase expression cassettes can be placed on the same plasmid to simplify the process (Mikkelsen et al. 2003).

Until Sleeping Beauty (SB) was reactivated in 1997, no transposon-based tool was available for genome manipulations in vertebrates and mammals. SB is a Tc1/mariner transposable element isolated from the salmonid fish genome (Ivics et al. 1997). The SB tool has been successfully used for genetic modifications in a wide variety of vertebrates, including human cell lines (Miskey et al. 2005; Mates et al. 2007; Ivics et al. 2007). Two other transposon systems could also be developed for use as vectors for gene therapy: Tol2 and piggyBac (PB). Indeed, Tol2 can transfer large transgenes of up to 11 kbp with minimal loss of transposition activity (Balciunas et al. 2006), and is less subject to overproduction inhibition of transposition (Wu et al. 2006). However, molecular engineering of Tol2 transposase that affects its N-terminus abolishes its activity, as has been described for SB (Wu et al. 2006). The piggyBac system shows the same characteristics as Tol2, but the transposase is less susceptible to the effects of engineering in its C or N-termini (Wu et al. 2006). Recently, the PB vector has been used to reprogram fibroblasts to produce induced pluripotent stem cells (iPS) after ectopic expression of a defined combination of the 4 transcription factors, c-Myc, Kfl4, Oct4 and Sox2 (Yusa et al. 2009). SB and PB have been used in pre-clinical studies to modify CD34+ hematopoietic stem cells for which retroviruses and lentiviruses have been the preferred vectors (Sumiyoshi et al. 2009; Xue et al. 2009). By mapping transposon insertion sites, several studies have underlined the potential differences in the pattern of genomic integrations (Yant et al. 2005; Galvan et al. 2009; Grabundzija et al. 2010; Meir et al. 2011). They have set the molecular mechanism of integration site targeting. Interestingly, PB and Tol2 have

revealed similar bias towards inserting into genes (nearly 49% insert into RefSeq genes for all cell types) and especially into introns or close to transcription start sites (Grabundzija et al. 2010). It seems that an open chromatin state surrounding transcriptionally-active chromosomal regions favors the piggyBac and Tol2 integration process. For SB, the frequency of gene insertion is 31–39%, depending on the cell type (Wu et al. 2006). Host factors probably play an important role in the SB transposition process through interactions with components of the transposable elements. Indeed, Zayed et al. 2003 showed that the high-mobility group protein HMGB1 acts as a co-factor of SB transposition in mammalian cells by favoring DNA bending (Zayed et al. 2003). SB transposase also interacts with the Ku protein, a component of the non-homologous end-joining (NHEJ) pathway of double-strand DNA break repair (Yant and Kay 2003), and with Miz-1, a transcriptional regulator of genes involved in regulating the cell-cycle (Walisko et al. 2006). No host factor has been described for PB and Tol2. The control of integration at the chromatin level is poorly understood, and remains to be elucidated.

Nevertheless, random integrations and subsequent risks of insertional oncogenesis cannot be excluded although transposons do not seem to display any pronounced bias in their integration pattern (Mates et al. 2009). Another concern relates the potential for inadvertent genomic integration of the transposase-encoding construct. If the transposase is continuously expressed, this could result in uncontrolled transposition or hopping of the integrated transgene that could contribute to increasing the genotoxic risk.

1.3 Principal Drawbacks of Gene Transfer Integrative Systems and Solutions

Two major limitations of gene transfer could impair the transgenesis efficiency and long-term expression of the transgene whatever type of vector is used. One consists of vector-on-host effects, in other words, how and to what extent the vector disrupts genome organization. The other involves the host-on-vector effects determined by whether the modified cell reacts towards the alien by inducing an innate immune response or silencing transgene expression.

1.3.1 Vector-on-Host Effects: Insertional Mutagenesis

Unless specifically targeted to a safe locus, transgene integrations may introduce high genomic diversity depending on the chromosomal integration pattern of the vector used and the copy number of transgene insertions. These phenomena are known collectively as ‘insertional mutagenesis’. On the one hand, random transgene integrations can disrupt an essential gene, which could dramatically compromise cell viability and lead to cell lethality. On the other hand, integrative vectors

used for gene therapy usually carry strong enhancers in order to ensure high and persistent transgene expression. These regulatory elements (promoter, enhancer, silencer) can influence the expression of adjacent cellular genes over distances as great as 90 kbp (Bartholomew and Ihle 1991). In extreme cases, regulatory elements present in the vector sequences could lead to the oncogenic transformation of an individual cell clone.

Most of these genotoxic effects are well documented for retrovirus vectors. They involve the presence of enhancers in the LTR sequence that activate proto-oncogenes implicated in cell growth or differentiation (Baum et al. 2006; Kustikova et al. 2010). The second-generation of recombinant retroviruses may address some of the inadvertent side effects, such as insertional oncogenesis. For example, self-inactivating (SIN) MLV and HIV-1-derived vectors with deletions in the U3 enhancer region of the LTR have been developed, and display higher biosafety, as a result of the abolition of the enhancer activity, and they have stronger transgene expression than the unmodified parental vectors (Schambach et al. 2007; Modlich et al. 2009). For transposon-based vectors, Moldt et al. demonstrated in 2007 that the inverted terminal repeats of Sleeping Beauty present *cis*-acting regulatory activities and act as promoters (Moldt et al. 2007). However, the most potent of these promoters was about 60-times less active than the SV40 promoter. Similarly, promoter analysis experiments performed in mammalian cells, revealed that the 5′ and 3′ terminal-repeats of PB do act as promoters. The 5′ promoter is 5-fold stronger than its 3′ counterpart (Cadinanos and Bradley 2007), which exhibits remarkably enhanced activity (Shi et al. 2007). No comparison with other strong promoters has been done to account for the propensity of these regulatory elements to induce clonal expansion by transcriptional dysregulation of adjacent genes. Nevertheless, in 2009 Galvan et al. showed that the frequency of PB integrants into or near known proto-oncogenes was not different from simulated random integrations, i.e. that there is no integration hot-spot near proto-oncogenes (Galvan et al. 2009). Secondly, in 2011 Meir et al. developed a more secure system based on a highly effective micro-PB transposon system (Meir et al. 2011). The micro-PB vector relies on the smallest terminal repeats in mammalian cells. Indeed, most activator sequences are excluded. This suggests that the internal regulatory *cis* elements of integrated transposons-derived vectors are less likely to influence the expression of flanking cellular genes than LTR promoters of viral vectors. However, further *in vivo* experimentation will be required to fully evaluate the genotoxicity of PB or SB transposons.

Other side effects are restricted to transposon-derived vectors. For instance, transposase activity needs to be regulated because excessive and uncontrolled transposition can result in genomic instability, leading to inversions, deletions and translocations that could mask singular transposition events (Geurts et al. 2006). Therefore, to avoid multiple insertions and remobilizations of the transgene, the transposase should be present in the cell only during a very narrow expression window. This restriction would allow a one-shot transposition process to occur with only one integrated copy. For this purpose, the transposase could be provided as an mRNA or directly as the protein (Wilber et al. 2006; Morales et al. 2007). While purified transposase production requires sophisticated and onerous procedures,

in-vitro transcription is now a commonly-used and easy way to obtain sufficient amounts of functional mRNA using commercially available kits. As reviewed by Yamamoto et al. in 2009, the mRNA strategy has several advantages over pDNA (Yamamoto et al. 2009). Since mRNA exerts its function in the cytoplasm, there is no need to cross the nuclear envelope. Consequently mRNA is also effective in non-dividing cells. mRNA is not able to integrate into the host genome and does not contain any promoter sequences. This eliminates the risk of perturbing the general network of gene regulation. Moreover, repeated applications are possible since vector-induced immunogenicity may be avoidable.

1.3.2 Host-on-Vector Effects

1.3.2.1 Innate Immunity

The ability to distinguish foreign nucleic acids from the abundant "self" nucleic acids is essential to protect the host from natural invaders. However, an excessive immune response against the vector carrier (virus or transfection reagent) and against the nucleic acid it carries would lead to the elimination of the transgene and so no therapeutic effect. Several innate immune surveillance systems have evolved to detect exogenous nucleic acids and trigger cellular responses depending on their recognition by Toll-like receptors (TLRs), the nature of their sequence (CpG content), or the structure of the foreign molecule.

TLRs are the best-studied family of pattern-recognition receptors (PRRs) that recognize conserved microbial components. Although most of these receptors sense pathogen components on the surface of various immune cells, such as macrophages and dendritic cells, a subset of them, TLR3, TLR7, TLR8 and TLR9 recognizes viral and bacterial nucleic acids in endosomal compartments. However, there is also accumulating evidence that supports the existence of TLR-independent mechanisms of virus-sensing by cytosolic PRR such as the antiviral helicases RIG-I and Mda5, or the nucleotide-binding, oligomerization domain NOD-like receptors (Sioud 2006; Bowie and Fitzgerald 2007; Sakurai et al. 2008; Huang and Yang 2009; Barber 2011). In 2010, Breckpot et al. showed that the activation of dendritic cells cultures *via* TLR3 and TLR7 was independent of their lentiviral pseudotype, but dependent on cell entry and reverse transcription (Breckpot et al. 2010). In their experiments performed in 2011, Rossetti et al. demonstrated that the induction of transgene-specific immunity is mediated by TLR7 and TLR9, a response that closely resembles that induced by functional viruses (Rossetti et al. 2011). Markusic et al. in 2010, suggest that the expression of foreign proteins in muscle could lead to an immune response, even if the viral vector capsid proteins have been removed (Markusic et al. 2010).

AAV-mediated gene transfer has not been associated with significant inflammatory responses or toxicity in animal or human models irrespective of the presence of neutralizing antibodies due to preexisting immunity to the common AAV serotype

(Mingozzi and High 2011). However, neutralizing antibodies induced against surface antigens of viral vectors could strongly interfere with the further readministration of such vectors. So, gene transfers based on AAV are limited even at their first use by preexisting neutralizing antibodies induced by natural infections. On the other hand, adenoviral vector particles tend to elicit strong innate immune responses, and 90% of the vector DNA is cleared from the tissue within 24 h following intravenous administration. Gene replacement therapy for hemophilia B was found to be limited by the induction of an immune response against components of the AAV transfer vector, ultimately resulting in elimination of the genetically modified cells (Manno et al. 2006; Nayak and Herzog 2010).

It has been clearly established that non-viral gene delivery methods induce a less severe immune response than virus-mediated delivery systems. Nevertheless, immunity problems due to the way pDNA and mRNA are produced and internalized in the cells still persist. Indeed, if synthetic vectors enter the cell through endocytosis, DNA and RNA can encounter endosomal TLRs. One way to circumvent the Toll-like-mediated response is to deliver vectors directly into the cytosolic compartment using physical delivery methods or lipidic formulations (fusion with the cytoplasmic membrane), but even this alternative is not absolutely reliable since DNA and RNA can be recognized by RIG-I or NOD elements. Recent studies have demonstrated that non-viral vector carriers, such as lipoplexes and polyplexes, can induce an innate immune response as a result of high levels of cytokine synthesis when delivered intravenously (Sakurai et al. 2008). Alternatively when the DNAs are complexed with polyethylenimine and delivered by aerosol, high level of transgene expression is achieved without inducing high levels of cytokine responses (Gautam et al. 2001). Consequently, the route of administration and the nature of the molecules transfected (pDNA or mRNA) determine the type and level of innate immune response induced.

Unlike viral vectors, plasmid expression vectors do not carry or encode antigens other than the transgene product, but bacterial sequences present in the plasmid backbone can still trigger a strong host immune reaction. Concerns have been raised regarding the immunostimulatory prokaryotic CpG motifs in plasmids. In fact, bacterial CpG motifs are either not methylated at all or less methylated than CpG islands from the human genome and can, therefore, interact with TLR9 in the cells of the innate immunity system. Activation of innate immunity results in the loss of cells harboring the vector DNA (Hodges et al. 2004).

Transfection of mRNA molecules is now widely used in gene transfer protocols However, the transient nature of mRNA has limited its use except when a brief pulse of transgene expression is desired or when working with stabilized mRNA (Hayashi et al. 2010). It has been shown that single strand RNA and *in vitro* transcribed mRNA have dynamic secondary structures that form double-stranded sequences and that are recognizable by TLR3 (Karikó et al. 2004; Ishii and Akira 2005).

Ex-vivo gene therapy is believed to provide a safer and less immunogenic approach than *in-vivo* gene transfer since it should avoid activating the immune system. In fact, no interaction occurs between the vector and human blood components. However, Lu and Ghazizadeh (2007) have observed that *ex-vivo* transduced

keratinocytes did induce an immune response and the clearance of the genetically modified cells when introduced into mice. This implies that a better understanding of the mechanisms of the immune reaction following *ex vivo* or *in vivo* gene transfer is essential. The use of animal models makes it possible to study immunity in a whole organism, but notable differences have been observed from the human species. So far, Georgopoulos et al. (2009) have developed an *ex vivo* human blood loop system to evaluate the innate immune response.

For efficient gene transfer, circumventing the immune response triggered by the vector is a major challenge regardless of the type of vector used (Zaldumbide and Hoeben 2008). Indeed, an immune response raised against a gene transfer vector may lead to elimination of the vector, the transgene and/or the transfected cells. This results in a decrease in both the intensity and the duration of the transgenic protein expression. Furthermore, as observed during infection by microorganisms, the immune response against gene therapy vectors involves the production of pro-inflammatory cytokines and chemokines that have harmful effects on transgene expression. Nonetheless, gene therapists are learning to circumvent, manipulate, or suppress unwanted immune responses. Advances in vector design (such as capsid engineering, regulated expression cassettes, etc.) delivery techniques, administration to privileged immune sites, immune suppression and modulation regimens and taking advantage of organ-specific immune responses, are all promising strategies for overcoming immunological hurdles. However in other contexts, the aim of gene transfer is in fact precisely to induce strong immunity. In this regard, DNA vaccination has attracted much attention as a way of preventing metastasis and relapse of malignant tumors (Un et al. 2011).

1.3.2.2 Transgene Regulation and Silencing

In addition to the innate immune response against the vector and transgene, other host-on-vector hurdles need to be circumvented. These include so-far unresolved problems related to how the cell controls the integration of the transgene and regulates its expression through its chromatin structure, epigenetic modulation and extrinsic environmental factors.

Chromatin Structure

Monitoring chromatin conformation is essential to correlate genome accessibility during the various cell cycle phases with the overall distribution of the transgene integration sites. This could be done through analysis of the DNA structure and distribution of DNAseI-hypersensitive sites (HSS), CpG islands or TSS analysis.

DNaseI HSS are usually related to an open chromatin state and the presence of active DNA binding sites for regulatory proteins. As reported by Huang et al. (2010), genome-wide integration analysis has demonstrated that SB, Tol2, and PB-mediated integrations occur indiscriminately in all the chromosomes. SB integrations are

randomly distributed (Grabundzija et al. 2010; Huang et al. 2010), but significant bias with regard to the integration site selection in the primary DNA sequence as well as local DNA structures has been reported (Zayed et al. 2003; Geurts et al. 2006; Walisko et al. 2006). Furthermore, Tol2 and PB integration sites have mainly been found near TSS, CpG islands and DNaseI hypersensitive sites. The integration preferences of γRV vectors are similar to those of PB and Tol2. Moreover, LV vectors show strong preferences for integrating near promoter regions and inside active genes. These loci are in favor with transgene expression, and also in genome dysregulation, driving the cell to turn off the newly integrated gene.

Epigenetic Modulation

It is well established that the expression of transgenes is often silenced once they have been integrated. This phenomenon raises the issue of whether the therapeutic protein will be efficiently and sufficiently expressed over time. In order to explore this post-integrative silencing of transgenes, investigation of the epigenetic status of the genome before and after gene transfer or cell infusions appears to be crucial. Indeed, DNA CpG methylation and histone modifications (methylation and acetylation) are important epigenetic markers of the open/closed chromatin states implied in the transcription regulatory network. The integration and expression of DNA transposons, retrotransposons, and retroviruses, are regulated in living organisms through CpG methylation and histone modifications in order to maintain genomic integrity. These modifications also influence the accessibility of transgene integration and the transcription-permissive state.

It is noteworthy that H3K27me3 is the only modification that makes chromatin inaccessible for integration events (Biasco et al. 2010), and it has been shown that H3K27me3 distribution is both gene- and cell-specific and is significantly modified when hematopoietic progenitor cells differentiate *in vitro* (Wei et al. 2009; Cui et al. 2009). When H3K27me3 marks were mapped on unrelated target cells, the distribution of the integration site with regard to the distribution of this histone modification was more similar to the random reference. As demonstrated by Grabundzija et al. (2010), Tol2 insertion sites are under-represented within chromosomal regions displaying H3K27me3 marks associated with transcriptionally-repressed chromatin. Histone 3 modifications are not only associated with lysine 27, but also with lysine 9. For instance, H3K9me3 is disfavoured by the MLV vector (Biasco et al. 2010), whereas, histones associated with TSS are heavily regulated through methylation (Barski et al. 2007). Cattoglio et al. (2010a, b) have shown that in pre-infusion T-cells, MLV integrations cluster in DNA regions containing particularly high levels of H3K4me1, H3K4me3, H3K9ac, H2A.Z and PolII, associated with an open chromatin state and gene activation. They also showed that 58% of these MLV integration clusters were shared in the post-infusion dataset, indicating that the distribution of integration sites is rather determined by MLV integrase rather by *in vivo* selection. The transposition process initiated by Sleeping Beauty in the mouse genome is associated with changes in

DNA methylation at the site of insertion (Park et al. 2006). CpG methylation of the SB transposon increases transposition efficiency (Yusa et al. 2004). The authors also showed that integration of methylated transposon was observed in histone H3K9me3, whereas the unmethylated transposon formed a relaxed euchromatin structure, as revealed by the enrichment of acetylated histone H3 in mouse cells. Zhu et al. re-evaluated these results in 2010 (Zhu et al. 2010).

It has been widely reported that the expression of the transgene declines over time, whereas the concentration of the vector DNA remains constant in the cells. Moreover, various different expression levels of the transgene have been observed, a phenomenon known as variegation or mosaicism. Silencing and variegation are both thought to be induced by epigenetic modifications such as DNA hypermethylation and histone hypoacetylation (Jaenisch and Bird 2003). Cytosine methylation of CpG islands found in promoters is involved in the formation of compacted chromatin to repress gene expression (Miranda and Jones 2007). Acetylation at specific histone H3 lysine residues, particularly in the promoter regions, can regulate gene expression. Hypoacetylation of histones leads to chromatin condensation through interactions involving the free lysine residues, resulting in suppression of gene expression (Kouzarides 2007). Once initiated, the formation of heterochromatin from the transgene sequence spreads to the upstream and downstream neighboring cellular genes (Grewal and Moazed 2003 and Maison and Almouzni 2004) through the recruitment of silencing factors, such as Swi6/HP1 or Sir3. This further promotes the recruitment of the deacetylases and methyltransferases that modify the adjacent histones, and create another binding site for the silencing factors (Wegel and Shaw 2005). In fact, since 1980, DNA methylation has been known to be associated with the inactivation of proviruses and transgenes delivered by different MLV-based retroviral vectors (Jähner et al. 1982; Jähner and Jaenisch 1985; Yao et al. 2004; Ellis 2005).

As previously reported, transposon integration could be directed into regions that are inherently restrictive of gene expression. Consequently, integration efficiency is probably underestimated. Incubating cells with DNA methyltransferase or histone deacetylase inhibitors could therefore have reactivated silenced clones, as demonstrated by Garrison et al. 2007. In fact, DNA methylation is considered to be a fundamental cellular defense mechanism preventing the expression of potentially harmful viruses or the mobilization of intragenomic transposable elements in mammalian genomes (Yoder et al. 1997).

Alternatively, integration into a highly active region of a chromosome could be subject to selective repression of the invading sequence by RNA interference or through transcriptional interference. mRNA transcribed from the transgene integrated copy could generate double-stranded RNA (dsRNA) after spatial folding. It has been demonstrated by Fire et al. (1998) that in *Caenorhabditis elegans* dsRNA can trigger the destruction of complementary mRNA. The RNAi mechanism has also been reported in mammalian cells (Elbashir et al. 2001). The dsRNA is converted into small interfering RNAs (siRNAs) of 21–25 nucleotides by the action of the RNAse III Dicer enzymes. These siRNAs are incorporated into a multi-protein complex known as the RNA Induced Silencing Complex (RISC), which guides the

cleavage of cognate mRNAs (Meister and Tuschl 2004). Another class of small non-coding RNAs, microRNAs (miRNAs), is involved in gene expression inhibition. miRNAs are encoded by the host genome and transcribed as part of a long primary transcript from PolII promoters (Kim and Kim 2007). miRNAs contain a hairpin structure that is cleaved by the nuclear enzyme Drosha. After their exportation into the cytoplasm, they are matured by Dicer. Like siRNAs, the mature miRNAs are incorporated into the RISC where they generally mediate translational repression of their target mRNA sequences. So far more than 700 human miRNAs have been identified. They are predicted to regulate the expression of at least one third of all human genes (Zhang 2009).

Extrinsic Environmental Factors

Finally, *in vivo* modified cells, the engraftment of *ex vivo* modified cells and transgene expression could all be influenced by extrinsic environmental factors. Numerous studies have highlighted the risks associated with the methods used to isolate and culture cells during the transgene integration process, with how the transgene is addressed to the cell and integrated into the genome, with how modified stem cells differentiate *in vitro,* and with how the *in vivo* modified cells interact with their environment. All of these steps could generate stresses that could impact on the epigenome of adjacent sequences (Fiszer-Kierzkowska et al. 2011; Takiguchi et al. 2011; Ahangarani et al. 2011; Jacobsen et al. 2009; Zeh et al. 2009).

1.3.3 Optimizing the Vector Architecture: Solutions to Ensure Safe and Long-Term Expression

1.3.3.1 Limiting the Innate Immune Response

We have previously described that the type and level of the host immune response is largely governed by how the vector enters the cell, notably depending on whether the endosomal pathways are involved or not. For example, electroporation-mediated plasmid delivery has been shown to reduce TLR-9 induced inflammatory responses of IL-6, TNF-α, and interferon-γ in animals by directly transferring constructs into the cytoplasm of cells, thus avoiding the endosome TLR-9 response (Zhou et al. 2007).

As pDNA vectors originate from prokaryotes, they have a higher frequency of unmethylated CpGs than mammalian DNA. One method for suppressing the inflammatory response consists of eliminating or methylating CpG sequences in plasmids (Yew et al. 2002; Reyes-Sandoval and Ertl 2004). One way to eliminate CpG is codon optimization; another solution involves engineering bacterial strains (Bower and Prather 2009). Plasmids produced from the usual *E. coli* K12 or B

strains are distinct from those produced from strains that have specific epigenetic nucleotide methylation. In 2011, Carnes et al. have reported the impact of differences in epigenetic dcm methylase-directed cytosine methylation on plasmid production, transgene expression, and immunogenicity (Carnes et al. 2011). In 2009 Bower et al. reminded us that contamination of pDNA by mobile elements is a serious regulatory concern, as these elements can alter the biological properties and safety profile of the vector DNA (Bower et al. 2009). For example, the DH10B strain was found to have a mutation rate 13.5-fold higher than the wild-type *E. coli* MG1655 strain. This difference is mostly due to a significantly higher rate of insertion sequence transposition. This concern is not purely theoretical; insertion sequence mediated mutagenesis was recently reported in an industrial process for the selection of HIV pDNA vaccine candidates. Consequently, bacterial strains have now been engineered to remove all their mobile elements (Posfai et al. 2006).

It has been shown that single-strand RNA and *in vitro* transcribed mRNA have dynamic secondary structures that form double-stranded sequences recognizable by TLR3 (Karikó et al. 2005; Ishii and Akira 2008). Human TLR3 localizes in the endosomal compartments in myeloid dendritic cells, while it localizes both on the cell surface and interior in fibroblasts and epithelial cells. To avoid the destruction of mRNA involved in some gene transfer protocols, Karikó et al. (2005) suggested that innate immune recognition of RNA by TLR3, TLR7 or TLR8 could be controlled by modifying nucleotides by processes including methylation. Kormann et al. demonstrated in 2011 the therapeutic utility of using chemically-modified mRNA as an alternative to pDNA-based gene therapy (Kormann et al. 2011). They show that replacing 25% of uridine and cytidine with 2-thiouridine and 5-methylcytidine, respectively, synergistically reduced the interaction with the TLR receptors and RIG-I, resulting in lower immunogenicity and greater stability in mice.

1.3.3.2 Limiting Genotoxicity Using Regulatory Components

The combination of adverse events in clinical trials and studies of transgene patterns of integration have stimulated an intense search for new approaches to vector design. Component optimization will increase the efficacy of gene transfer while decreasing toxicity at the systemic and cellular levels.

Because viral or non-viral vectors contain different regulatory elements, such as promoters or enhancers, it is important to look at how these elements interact with the surrounding genome at the integration site. In fact, modification of the general expression network of the cell could lead to potential genotoxicity due to activation of a cellular proto-oncogene or disruption of an oncosuppressor and induce innate immunity, resulting in the destruction of the modified cells. Genotoxicity could therefore be reduced through the judicious choice of vector sequences. Moreover, optimizing transcriptional regulatory elements would improve the quality of transgene expression (Yew 2005) and the immune-escape strategies (Zaldumbide and Hoeben 2008).