Topics in Current Genetics 23

Sylvaine Renault
Philippe Duchateau *Editors*

Site-directed Insertion of Transgenes

Springer

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

Springer

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

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

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,

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).

When transgenes are moved, following their transfer into close proximity with constitutive heterochromatin, their expression can become unstable, resulting in uneven patches of gene expression. This process is regulated by a variety of proteins that are implicated in both chromatin remodeling and RNAi-based silencing. To avoid this phenomenon, it is essential to isolate the transgene sequence from its surrounding environment. One way to do this is to adjoin insulator elements. Insulators are chromatin regulatory elements, which are divided into two classes depending on their properties. When an insulator of the first class flanks a transgene it marks the boundaries of active chromatin domains and exhibits barrier activity to physically isolate the transgene from its surrounding environment. Consequently, these insulators block the encroachment of heterochromatin onto the transgene to limit its silencing and more importantly, prevent any impact of the *cis*-regulatory sequences of the transgene on the nearby genome. The second class exhibits enhancer-blocking activity from the transgene towards the adjacent genomic regions when placed between an enhancer and a promoter. This property is of great importance in limiting the genotoxic effects associated with transgene *cis*-regulatory sequences (Valenzuela and Kamakaka 2006; Gaszner and Felsenfeld 2006).

Such chromatin regulatory elements have demonstrated their efficiency in improving the expression and safety of integrative gene transfer vectors. Their use is now widespread, as nicely reviewed by Emery (2011), and insulators are currently being developed by both academia and industry (Palazzoli et al. 2011). Evans-Galea et al. (2007) provide evidence that including the chicken insulator element cHS4 into a lentiviral vector results in the suppression of clonal dominance in cultured human lymphoid cells. This finding has also been confirmed by Zychlinski et al. (2008), but to a lesser extent. Insulator sequences also have a crucial impact on the stability of the transgene expression (Dalsgaard et al. 2009; Sjeklocha et al. 2011). Insulators reduce genotoxicity since they can suppress the clonal dominance and improve the long-term expression of the transgene by blocking the spread of repressive heterochromatin

To avoid the genotoxic risks associated with the regulatory sequences present in the vector, Zychlinski et al. propose modifying SIN definition vectors by replacing retroviral enhancer-promoters by cellular promoters derived from the human elongation factor-1a (EF1a) and phosphoglycerate kinase (PGK) genes. They demonstrated that SIN vectors using the EF1a enhancer-promoter were unable to immortalize primary hematopoietic cells even in the absence of insulators, despite drastic dose escalation. The expression levels of EF1a and PGK promoters are 3- to 10-fold lower than those of a powerful MLV-derived enhancer-promoter (per transgene copy) (Schambach et al. 2007). However, the presence of the first intron of EF1a enhances gene expression mediated by this promoter by a factor of 2–3. Although it has not been formally demonstrated, the primary cause of this effect is probably improved RNA processing. Besides these first line strategies, vectors could also be equipped with a conditional suicide gene, such as the thymidine kinase gene from the Herpes Simplex virus, which acts like a time bomb if clonal expansion occurs (Ciceri et al. 2007, 2009). Caution is therefore called for when choosing an appropriate enhancer-promoter or inducible system for transgene expression (see Table 1 of Guo et al. 2010).

1.3.3.3 Ensure Expression by Modifying Vector Architecture

To optimize gene transfer using non-viral systems, the trafficking of the transgene into and within the nucleus, where the chromosomes are located, have to be considered. Indeed, various cellular barriers hinder exogenous gene transfer. The main obstacle in non-dividing cells is getting across the nuclear envelope. In their work, Lam and Dean (2010), described ways of circumventing this difficulty, notably by adding small nucleic sequences, known as DNA nuclear targeting sequences (DTSs), to the plasmid backbone. These DTSs, which have a 72-bp sequence from the SV40 enhancer as a sequence leader, lead to increased nuclear targeting of plasmids and thus, increased gene expression in multiple cell types. As neatly reviewed by Miller and Dean (2009), the development of specific DTSs to restrict gene delivery and expression to a single cell type would be of great interest. For some purposes, driving the vector into the nucleolus by adding a nucleolar localization sequence (NoLS) to insert the transgene directly into rDNA genes has also been envisaged (Dang and Lee 1989; Liu et al. 2005; Lechertier et al. 2007).

Several lessons have been learned in the course of gene transfer studies intended to enhance transgene expression (Trinh et al. 2009; Guo et al. 2010). Additional sequences, such as polyadenylation signals or supplementary introns, could be introduced into the transgene expression cassette, to reduce the transcriptional read-through (Baum et al. 2006; Schambach et al. 2007) and the alternative splicing engendered by vector sequences, respectively. The inclusion of an intron in a retroviral vector has previously been reported to facilitate processing of the mRNA transcripts, thus resulting in enhanced transgene expression (Ismail et al. 2000). Moreover, introducing both the human polyadenylation signal and the seventh intron from the human ADA gene further improved T-cell specific transgene expression (Trinh et al. 2009). Other ways to achieve reliable co-expression of two or more cDNAs from a single vector have been investigated. The use of internal ribosomal entry sequences (IRESs) or the self-cleaving 2A proteinase signal of the foot-and-mouth disease virus may lead to coordinated protein expression under the control of a single promoter (de Felipe 2002). Recently, bidirectional promoters have been shown to enhance lentiviral transgenesis (Golding and Mann 2011). Eliminating CpG by codon optimization is not only a way of preventing an immune response against the transgene, but can also lead to high-level expression of the transgene (Ward et al. 2011) and to a reduction in mRNA secondary structures that could trigger RNA interference.

Taken together, the architecture and components of the vector can reduce vector-mediated genotoxicity and immunogenicity and improve transgene expression. The design of the vector will depend on the type of vector used, the route of administration, the experimental systems, the aims of the gene transfer, and the use of tissue-specific enhancers. The various configurations designed should be rigorously validated both in *in vitro* and animal models. However, the gene networks operating in the human body can barely be taken into account by these experimental models. Most, if not all, current therapeutic protocols involve a single transgene. To produce therapeutic effects, the next generation of trials should include the design of vectors

that are able to inter-operate in host cell activities in order to promote physiological, sustained, feedback-controlled transgene expression. Moreover, to develop such new vectors, up-scaling plasmid DNA (Bower and Prather 2009; Carnes et al. 2011) or mRNA production (McKenna et al. 2007) and engineering host cells are essential.

1.3.3.4 Number of Insertions Acquired per Cell and Hot Spots of Integration Bias

Transgene expression usually increases as a function of the gene copy number, which made it tempting to transduce cells with a high viral vector titer or to transfect cells with a huge amount of a synthetic, non-viral vector (Wahlers et al. 2001; Kustikova et al. 2003). However, transgenes inserted in multiple copies could be silenced. Where is the balance between a proper level of expression and reduced genotoxicity? Studies of the integration profile show different patterns: single integrations of one or more copies in tandem, multiple single integrations, and hot spots (the accumulation of integrations at one locus).

Several independent insertion events occurring in a single cell may cause major problems related to combinatorial insertional mutagenesis in cellular proto-oncogenes. In addition, hot spots may be related to clonal expansion. Various pre-clinical models have clearly demonstrated a dose-dependent component of vector-mediated genotoxicity (Modlich et al. 2006; Montini et al. 2006; Li et al. 2009). Therefore, to avoid these side effects the number of vector insertions in the cells should be as small as possible (Baum et al. 2003). Moreover, the more effective the integrative system, the less of the vector will be used. In addition, the frequency of transgene integration could be adjusted *via* the multiplicity of viral infections or the amount of pDNA or of mRNA used. The functionality of the integrase/transposase/recombinase can also be modified assuming that hyperactive enzymes require less vector to integrate the transgene. Furthermore, Moreno-Carranza et al. (2009) show that optimization of the transgene gp91phox cDNA by codon optimization and by increasing the GC content from 47% to 61% leads to a 10-fold increase in the SIN vector titer. Similar observations have been made for vectors containing factor VIII or factor IX cDNA (Radcliffe et al. 2008; Wu et al. 2008). On the other hand, to further boost the efficiency of transposition, transposase expression levels were increased by codon optimization strategies or by changing almost every amino acid in order to derive hyperactive mutants (Cadinanos and Bradley 2007; Mates et al. 2009). So, optimizing transposase efficacy and the amount of transposase are ways to regulate the number of transgene insertions. Grabundzija et al. (2010) demonstrated that almost all clones had a single insertion when low amounts of transposase are used, whereas SB, Tol2 and PB generated insertions in the range of 2–40 copies, 1–3 copies and 1–4 copies per cell clone, respectively, when high levels of transposase were used. Since the therapeutic effect should be observed with only one or two copies per transduced cell, there is no need to insert multiple copies of the transgene into the genome of the target cells.

Finally, a comprehensive view of the vector integration loci is essential to assess vector biosafety and to monitor the fate of gene-corrected cells. For this purpose, numerous protocols have been established for high-throughput integration site profiling coupled with mass sequencing. Insertion site patterns can be determined by various PCR-based methods, the most popular protocol used is linear amplification-mediated PCR (LAM-PCR) (Schmidt et al. 2007). Recently, Paruzynski et al. (2010) have developed a non-restrictive LAM-PCR (nrLAM-PCR) that is not biased by the use of restriction enzyme cleavage of genomic DNA. Such protocols have been used to study clonal behavior and the physiology of gene-modified hematopoietic (Schmidt et al. 2003) and T cells (Cattoglio et al. 2010a, b). Other efforts have been made in this field, and have resulted in the Transposon Insertion Site Profiling chip (TIP-chip) assay (Wheelan et al. 2006). Coupled with mass sequencing, this provides comprehensive analysis of the integrome (Paruzynski et al. 2010). Recently-introduced methods exploit the large number of sequences that can be retrieved by pyrosequencing approaches. These methods use primer bar-coding that makes it possible to analyze hundreds of samples in a single sequence run (Hamady et al. 2008). Each integration site is located in the genome, and the preferential use of one site will be evaluated. This monitors the importance of a given hit to define a common insertion site (CIS). But, as pointed out by Fehse and Roeder in 2008, attention should be paid to the size of the sample (Fehse and Roeder 2008). Indeed, there is a greater probability that a given CIS may have occurred by chance when the number of analyzed integration sites increases, and so identifying two insertions at a single locus might not be sufficient to consider this insertion site biologically significant. In addition, findings should be moderate when the insertion patterns of different vectors are compared on the basis of different numbers of analyzed insertions. Otherwise, the vector for which fewer insertion sites have been analyzed might appear safer since it appears to generate fewer CISs.

1.4 Combining Long-Term Expression and Secure Integration by Specific Locus Targeting: Old Solutions to New Perspectives

We have seen the impact of the choice and design of a vector in improving secure gene transfer. Nevertheless, no complete solution has been found to any of the problems mentioned, and further efforts must be made. An alternative approach to gene therapy that would minimize the risks of insertional mutagenesis is based on targeting specific safe loci in the genome. Thus, the crucial questions are how to target the transgene to a specific locus, and what defines a good target site.

The intuitive approach is probably to replace the deficient gene by a functional one at the exact locus in the genome in a gene-to-gene correction approach. Nevertheless, homologous gene recombination is not powerful enough to treat mammalian cells (10^{-6} to 10^{-9} insertions per genome efficacy). Moreover, non-homologous recombination is 3- to 5-fold more frequent than homologous

recombination in mitotic mammalian cells. However, homologous recombination can be stimulated by strategies involving the introduction of DNA double-strand breaks (DSBs), triplex-forming oligonucleotides or DNA vectors (Sorrell and Kolb 2005). Another solution is to use and engineer the large range of enzymes, such as recombinases, viral integrases and transposases, that naturally catalyze the insertion of foreign DNA into the host genome to target the integration of the transgene in a defined locus. The aim is to achieve 1% of truly targeted integrations into the safe locus. However, only a subset of these enzymes present site-specific integrations in their natural host. Most of them interact with multiple sites in bigger eukaryotic genomes, and prokaryotic enzymes could be sensitive to eukaryotic DNA modifications. So, one solution is to couple the integrative domain of an enzyme (ability to achieve DSB) with the site-selectivity domain of another enzyme. In the ideal scenario, domains are clearly identified and isolated and each domain can be engineered to optimize its efficacy in generating DSB and defining new target sequences. In the next paragraph, we will give some information about the commonly-used integrative enzymes (for more details see the following chapters). We will then specify what constitutes a good and safe integration site, and address some questions about the integration process.

1.4.1 Engineering Integrative Enzymes

Specific DNA sequences are naturally recognized by two different mechanisms, namely the base-pairing of nucleic acids and recognition by sequence-specific DNA binding proteins. Both of these mechanisms can be exploited to mediate site-specific integration of therapeutic genes. To be useful for the purposes of genome engineering, an endonuclease must exhibit an extraordinary combination of qualities: specific recognition of long sequences (ideally, long enough to constitute a unique occurrence in a eukaryotic genome) coupled with sufficient adaptability to redefine new target sequences. Another strategy is to use DNA-modifying enzymes that have little or no sequence specificity, such as many recombinases, integrases or transposases, and confer on them site selectivity by attaching heterologous DNA binding domains (DBDs) (Review Kolb et al. 2005).

1.4.1.1 Recombinases

Two groups of recombinases, the tyrosine recombinase family and the serine recombinase family (Wirth et al. 2007; Turan et al. 2011), catalyze sequence-specific DNA insertions by producing recombination between two identical recombinase target sites. One corresponds to the genomic address, while the other one is contained in a donor plasmid. In the serine-recombinase group, the resolvase/invertase ϕC31 isolated from the Streptomyces phage ϕ31 catalyses recombination between

the attP site, in the ϕ31 phage genome, and the attB site, located in the bacterial host genome. Several pseudo-sites have been found in mammalian cells, and there are almost 100 in the human genome (Thyagarajan et al. 2001). Although transgenes can be integrated into human embryonic stem cells (Liu et al. 2010), the system is limited by the impossibility of selecting a specific integration site. Furthermore, preferred sites differ depending on the cell type. More importantly, ϕ31-mediated recombination has been associated with a relatively high rate of chromosomal translocations that are associated with cancer. Unless this potential form of genotoxicity can be eliminated, the ϕ31 system will probably not be suitable for use in human gene therapy (Ehrhardt et al. 2006).

Several tyrosine recombinases, such as the bacteriophage P1-derived Cre and the yeast-derived Flp recombinases, have been widely used for site integration in animal cells since the original pioneering work in the 1990s (Kameyam et al. 2010; Zhou et al. 2010). Their recognition targets are loxP (locus of crossover in phage P1) and FRT (Flp recombinase target) respectively. Nonetheless, an increasing number of observations show that chromosomal aberrations are caused by these particular recombinases. These inadvertent effects can be explained by less stringent control due to the presence of pseudo-sites with little resemblance with the loxP element. Thus, the unintended endonuclease activities of Cre contribute to the system's apparent toxicity, giving rise to chromosome fragmentation and fusion (Turan et al. 2011). As other experiments have shown, these systems are not as straightforward as expected; particular attention has been paid to facts and artifacts associated with their production and applications (Birling et al. 2009). Several attempts have been made to improve their efficiency and specificity by redesigning their site-specificity (Voziyanov et al. 2003; Santoro and Schultz 2002).

1.4.1.2 Meganucleases

Natural meganucleases, also known as homing endonucleases (HEs), are sequence–specific nucleases with large (14–40bp) cleavage sites. Four well-defined homing endonuclease families have been characterized on the basis of their conserved sequence motifs. They consist of the LAGLIDADG, His-Cys box, H-N-H and GIY-YIG families. The LAGLIDADG enzymes are the largest HE family with more than 200 proteins. HEs induce DSB at specific loci in cells (Thierry and Dujon 1992) to promote gene replacement through homologous recombination (HR) with a donor template. Their use has been limited by their repertoire since the sites they can recognize do not reflect the diversity of the genomic sequences. Therefore, a combinatorial approach to designing artificial endonucleases with the desired specificity derived from the LAGLIDADG-HE family has been developed (Heath et al. 1997; Smith et al. 2006). This strategy relies on the hypothetical identification of several independent DNA binding regions within the I-CreI protein, and the development of automated high-throughput screening (Arnould et al. 2006). Previous studies have identified a highly conserved core structure that is characterized by αββαββα

folding organized in two independent subdomains. The modular structure of the protein makes it possible to combine several different subdomains in order to create redesigned proteins able to recognize new specific target site. Many modified I-CreI proteins maintain their very narrow specificity, and display high levels of cleavage on new targets in living cells, showing that engineering can preserve both efficacy and specificity.

One limitation of the meganuclease strategy is that in mammalian cells, DSBs are predominantly repaired by non-homologous end joining (NHEJ) mechanisms (Cohen-Tannoudji et al. 1998). Recently, two new mega-nickases have been generated to favor HR rather than the NHEJ repair mechanism to increase specific gene replacement (Niu et al. 2008; McConnell Smith et al. 2009). Using the I-CreI dimeric LAGLIDADG meganuclease as a scaffold, Grizot et al. (2009), described the engineering of a series of endonucleases cleaving the human RAG1 gene and inducing targeted recombination at an endogenous locus in up to 6% of transfected human cells. Despite this success, it has been shown that at high doses, meganucleases can induce lethality in mice (Gouble et al. 2006). A large part of this toxicity is attributed to frequent off-site cleavage since meganucleases can only tolerate a low level of target degeneracy. So far this mutagenic effect is difficult to evaluate in gene therapy protocols.

1.4.1.3 Zinc-Finger Nucleases

Zinc-finger nucleases (ZFNs) have emerged for site-specific genome editing. They are composed of DNA binding domains of eukaryotic transcription factors (zinc-finger proteins, ZFPs) fused to the nuclease domain of the FokI restriction enzyme. ZFNs are composed of two well-characterized domains capable of independent modularity, which makes it possible to optimize each part of the protein. Each zinc finger motif binds to a triplet within the DNA *via* a single α-helix (Pavletich and Pabo 1991). Several fingers can be linked in tandem to recognize a broad spectrum of DNA sequences. Greater specificity can be achieved by adding more ZF motifs to the ZFP to specify at least 18bp of DNA per cleavage site (Beerli and Barbas 2002). On the other hand, FokI catalytic domains with enhanced cleavage capacity have been described (Guo et al. 2010). Besides, as FokI dimerizes to cleave DNA, it is possible to foster the formation of heterodimers to improve specificity (Szczepek et al. 2007). ZFNs can be designed to generate sequence-specific DSBs, which are then repaired by the natural, homology-directed DNA repair machinery or NHEJ pathway. ZFNs have been employed to trigger the targeted editing of genomes at over 50 gene loci in 11 model organisms. They are currently being investigated in three phase-I clinical trials in primary human cells for the treatment of glioblastoma and HIV (Händel and Cathomen 2011; Urnov et al. 2010). Even though bioinformatic searches have restricted off-target ZFN occurrence by selecting the most specific ZF associations *in silico*, integration site analysis during the clinical trials will constitute an important step forward in the use of ZFN in gene therapy.

1.4.1.4 Viral Integrases

Although little evidence for oncogenic insertional mutagenesis has been attributable to lentiviral vectors, targeting LV vectors to specific genome sites would be a major advance. This could be done by constructing a fusion protein between the viral integrase and a sequence-specific, DNA-binding protein. The fusion proteins direct integration into immediately adjacent regions of their target sites (Sandmeyer 2003). Two sequence-specific DNA binding proteins that have been fused with LV integrase are the *Escherichia coli* LexA repressor and the DNA binding domain of the *Lambda* bacteriophage repressor (Goulaouic and Chow 1996). However, findings revealed that these chimeric proteins were not specific enough. In 2009, Su et al. developed a targeting system using a polydactyl zinc finger protein, E2C (Su et al. 2009). This polydactyl ZFP specifically recognizes a unique 18bp DNA sequence in the human genome. This would direct the transgene integration to the 5′ untranslated region of the erB-2 gene on chromosome 17. Aubert et al. in 2011 have developed another strategy that relies on the use of engineered homing endonuclease Y2 I-AniI to confer specificity on the integrated proviral DNA (Aubert et al. 2011). The aim is to induce mutagenesis of the latent viral sequences of the long-lived memory T cells in order to inhibit the reactivation of the provirus. These cells are the main reservoir of latent HIV infection.

1.4.1.5 Transposases

Kaminski et al. (2002) were first to propose directing integration towards defined target sites by using a transposase and DNA-binding domain. Then, Maragathavally et al. (2006) used the Gal4-UAS system to achieve site-directed integration through a chimeric transposase derived from the MosI and piggyBac systems. They used a plasmid-based transposition assays in *Aedes aegypti* embryos to analyze the potential of this strategy. A consensus AT-rich palindrome is highly favored at the site of SB transposon insertion. Since, there are approximately 200,000,000 potential TA integration sites in the human genome (Vigdal et al. 2002; Liu et al. 2005), SB target site selection appears to be determined by structural constraints rather than by the primary DNA sequence. This means that improving the site selectivity of transposases is essential. The first step towards specific site integration at the genome scale has achieved by Yant et al. in 2007 with an SB transposase fused with the E2C zinc-finger protein. Unfortunately, the most of the gene addition events produced by this system are still random (Yant et al. 2007).

Subsequently, SB and PB have been modified by the addition of zinc-finger DNA-binding domains (Wilson and Georges 2010). Recently in 2011, Gersbash et al. demonstrated the ability of a zinc-finger recombinase (ZFR) to target transgene integration (Gersbash et al. 2011). The ZFR is composed of a zinc finger protein and the catalytic domain of a serine recombinase. The piggyBac transposon system was first used to distribute the ZFR sites randomly in the genome. The polyclonal population of transposon-modified cells was screened for individual

clones that had integrated the transgene in a safe locus. Finally, the ZFR targets the transgene integration to this safe locus. In spite of these efforts, off-target insertions still occur.

1.4.1.6 Other Existing Solutions

Targeted gene delivery is also possible using more indirect approaches recently reviewed by Voigt et al. (2008). These approaches involve DNA-specific binding proteins to interact directly with the DNA component of the vector or the enzyme inducing DSB. In this way, the vector is directed close to the targeting specific site where integration could occur. Targeting retroviral integration through specific protein-protein interactions has been raised using HIV integrase and lens epithelial-derived growth factor, a ubiquitous nuclear factor tightly associated with chromatin that interacts with HIV (Turlure et al. 2004; Ciuffi et al. 2005). On the other hand, some authors used the ZFR strategy to screen polyclonal populations in order to find a favorable genomic locus (Woltjen et al. 2009; Papapetrou et al. 2011).

1.4.2 What Defines a Good Insertion Site?

Studies based on first generation vectors have helped to define a safe harbor locus. Integration at such a locus is needed to support sufficient and stable gene expression in the modified cells (a locus with an open chromatin state). In addition, the integration should not interfere with normal cell function or gene regulation (a locus with no proto-oncogene or essential genes in its vicinity). Some authors have described a safe harbor site as a locus that fulfils the following five criteria: they are located (i) at least 50 kbp from the 5′ end of a gene, (ii) at least 300 kbp from any cancer-related gene, (iii) at least 300 kbp from any miRNA, (iv) outside a transcription unit and (v) outside ultraconserved regions of the human genome (Papapetrou et al. 2011). So, where are safe harbors found?

One possibility is to target repetitive sequences. The second solution is to drive the integration to specific loci already reported to be nonmutagenic. The last solution is to screen clones that have only one copy of the integrated transgene in order to characterize putative universal genomic safe harbors.

1.4.2.1 Integration into Repetitive Sequences

Integration into Widely Transcribed Repetitive Genes

There are several repetitive genes that are widely transcribed in the genome. It is supposed that disruption of one or few copies can be tolerated without any deleterious effects. For example, the human ribosomal DNA or rDNA locus contains ~400

copies per haploid genome clustered on five chromosomes (13, 14, 15, 21 and 22) (Sakai et al. 1995). The expression of these genes is robust, since rRNA accounts for nearly 80% of the total RNA in growing mammalian cells. rDNA is well conserved among organisms, especially a 50-bp region of 28S rDNA that is a target for sequence-specific, non-LTR retroposons (LINE). This short region is highly conserved in chordates and arthropods and in all eukaryotes. This could constitute a good universal integrative site. In fact, the insertion sites of six 28S rDNA–specific LINEs (R1, R2, R4, R6, R7, and RT) reside in the most highly-conserved regions in the 28S rRNA gene (Ben Ali et al. 1999). Hitherto, LINE transposons have not been developed, because of the lack of good information about the transposition mechanism.

Recently, Liu et al. (2007) have developed a non-viral targeting vector, pHrneo, which can target a foreign gene in an rDNA locus. They have successfully integrated the human factor VIII gene into the rDNA repeats of human hepatocytes. In 2008, they obtained variable targeting efficiency, but with homogenous expression driven by an RNA polymerase I promoter. These findings suggest ways in which rDNA targeting strategy could be enhanced. For example, the plasmid could be engineered with a nucleolar localization signal to drive vectors directly into the nucleolus, where rDNA is located.

Multigenic protein families, such as histones, could be other candidates. The copy number of histone families in a species are variable, but some are present in up to 1,000 copies (Kedes 1979).

Integration into Telomeric Repeats

The telomeric region usually exists as heterochromatin. This represses internal gene expression in yeasts (Tham and Zakian 2002). Despite this heterochromatic state, native SART, a LINE retrotransposon, is expressed actively (Takahashi and Fujiwara 1999). Takahashi and Fujiwara (2002) demonstrated that swapping the endonuclease domain from TRAS1 (another retrotransposon) into SART1 converts its insertion specificity into that of TRAS1. Thus, the primary determinant of *in vivo* target selection is the endonuclease domain, suggesting that modified LINEs could be used as gene therapy vectors to deliver the gene of interest to specific locations. Such a safe genomic location may provided by the subtelomeric region. Kawashima et al. (2007) have developed a novel gene delivery system combining the baculovirus AcNPV and the silkworm LINE, SART1. They have succeeded in retrotransposing the enhanced green fluorescent protein EGFP into the TTAGG telomeric repeats. This is the first demonstration of site-specific gene delivery in living larvae of *Bombyx mori*. Gene delivery is dependent on the expressed SART1 protein produced by the vector but not by endogenous elements. They have also shown that some individuals of the F1 generation retained the introduced sequences. The target sequence (CCACTGTCCCTATCTACTATCT) is highly conserved among a wide variety of organisms, including insects and vertebrates. Therefore, the combination of site-specific LINE and AcNPV elements might become a universal tool. Another

way to target integration into telomeric chromatin is to use the integrase of the retrotransposon Ty5 found in *Saccharomyces cerevisiae* (Xie et al. 2001).

There are many target-specific LINEs other than SART1 and R1 (Kojima and Fujiwara 2003, 2004). By using these elements, we could deliver transgenes to many different chromosomal sites other than telomeres or 28S rDNA.

1.4.2.2 Integration in Specific Loci

HPRT Locus

Bronson et al. (1996) described the introduction of transgenes by gene targeting into the hypoxanthine phosphoribosyltransferase (HPRT) locus of ES cells. They used an HPRT-cell line and they selected the mutant that restored the HPRT gene by gene targeting. Lack of HPRT product causes the cells to die when grown on selective media containing hypoxanthine, aminopterin, and thymidine (HAT). However, reconstitution of this locus, using a specific targeting vector, makes it possible to directly select homologous recombination events. Indeed, only properly-targeted ES cells survive HAT selection. This strategy avoids the use of selectable markers in the transgene cassette that often have deleterious effects on transgene expression.

The HPRT gene is located on the X chromosome at Xq26.1. Consequently, the transgene expression at this locus is subject to random X-inactivation, and in XY ES cells, the gene disruption can be directly selected with the appropriate media. HPRT is among the permissive loci that have been identified for several promoters and various cell types (Jasin et al. 1996). Palais et al. (2009) have achieved tightly-controlled *in vivo* expression with the tetracycline system in mice.

Locus ROSA26

Decades of research into gene targeting in mice have made it possible to develop a number of tools for knocking genes in and out of embryonic stem (ES) cells. Not all genomic sites are amenable to gene addition since many are prone to silencing or can disrupt normal gene function. The discovery of the ROSA26 locus overcomes these problems. It was first described as a gene-trapped locus on chromosome 6, from which the β-geo gene was shown to be expressed ubiquitously throughout the whole body of a mouse (Zambrowicz et al. 1997). Nowadays, the mouse ROSA26 locus has become established as one of the preferred docking sites for the ubiquitous expression of transgenes, as it can be targeted with high efficiency and is expressed in most cell types. This suggests that the genomic region is not affected by chromatin configurations that could lead to transcriptional repression of inserted transgenes. In addition to ubiquitous transgene expression *via* the endogenous promoter, transgenic constructs harboring different exogenous promoters or different transgene cassette including reporters, site-specific recombinases and, recently,

non-coding RNAs have been positioned at the ROSA26 locus (Casola 2010; Chen et al. 2011). However, the transcriptional complexity of the ROSA26 locus (Zambrowicz et al. 1997) indicates orientation-dependent effects (Strathdee et al. 2006). The lack of systematic studies has limited the general application of this method with respect to exogenous promoters. Nevertheless, Irion et al. (2007) have identified the human homologue of the mouse ROSA26 locus on chromosome 3. They demonstrated that this locus can be targeted by homologous recombination and support stable, ubiquitous expression of transgenes in human ES cells and derivative cell lineages.

AAVsite1

The natural propensity of adeno-associated virus (AAV) to persist in human cells and its non-pathogenic features have made it very attractive for developing recombinant AAV gene vectors (Henckaerts and Linden 2010). In its latent state, AAV is found integrated into the host genome at a specific site, designated AAVS1. This site has been mapped within the PPP1R12C gene (protein phosphatase 1 regulatory subunit 12C) to the chromosome 19q13.4.locus This means that AAVS1 is located in an open chromatin structure that is transcriptionally active. It contains native insulators that enable resistance against transgene silencing (Lamartina et al. 2000; Ogata et al. 2003; Philpott et al. 2004). No adverse effect on the cell due to disruption of the PPP1R12C gene has been detected. Moreover, the transcriptional competence of a transgene cassette inserted into this site occurs in various different cell types. Therefore, AAVS1 is considered to be a safe harbor for adding a transgene into the human genome (DeKelver et al. 2010). In addition, the discovery of a mouse ortholog on chromosome 7 of the human integration site has made it possible to carry out studies to address the safety aspects of the targeted integration process (Dutheil et al. 2004). To direct the transgene into this integration site Ramachandra and colleagues used a two-step process that introduced loxP sites into AAVS1, followed by the introduction of a floxed transgene by Cre recombinase (Ramachandra et al. 2011).

Several other systems of integration are discussed in terms of specificity. Lombardo et al. (2011) used the ZFN technology to target different expression cassettes into CCR5 (HIV co-receptor chemokine receptor 5) and AAVS1 of different human cell types. CCR5 was thought to be another integration site providing a safe harbor, since homologous deletion is found in apparently healthy people. However, their data have revealed transcription perturbations of flanking genes when the transgene is integrated into CCR5, but not into AAVS1.

1.4.2.3 Random Integration

Papapetrou et al. (2011) and Papapetrou and Sadelain (2011)) have developed another strategy. They hypothesized that the screening of clones of induced pluripotent stem cells (iPS) harboring a single vector copy would facilitate the discovery of

new secure harbor sites. For this purpose, they used iPS cells with a well-characterized globin LV vector in the expectation that this would meet the five criteria described above. They have examined 5,840 integration sites, and found that 17% of them met all five criteria. Transgene and endogenous gene expression in the vicinity have been also assessed.

1.5 Conclusion

Several lessons have been learned over the past 20 years. All the studies show that not only vector origin and design, but also the context of cell selection, and the molecular architecture of a given gene locus determine whether and to what degree the transgene insertion induces genotoxicity. The multilineage and self-renewing capacity of the founder cell may be a prerequisite for inducing cell immortalization, whereas less primitive cell progenitors, may require additional cooperating mutations before they induce any problems in transplanted hosts. What has been shown is that natural recombination is not very efficient in most cells so its application in clinical uses has been eliminated.

However, homologous recombination could be strongly facilitated by inducing double-strand breaks in the vicinity of the target sequence. There are different enzymes capable of introducing such breaks that have to be engineered to target the double-strand break whatever the location of the site. Site-specificity requires a minimal recognition sequence of about 16bp. In the human genome, there are about 2,000 sites that vary by 2bp. This variation could reduce the affinity of integration. Therefore, a given recombinase, integrase or modified transposase could direct integration into a related site with a similar nucleotide sequence.

In conclusion, off-target cleavages induced by these designed nucleases should be taken into account with regard to potential future clinical purposes.

References

Ahangarani RR, Janssens W, Carlier V et al (2011) Retroviral vectors induce epigenetic chromatin modifications and IL-10 production in transduced B cells via activation of toll-like receptor 2. Mol Ther 19:711–722

Arnould S, Chames P, Perez C et al (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. J Mol Biol 355:443–458

Aubert M, Ryu BY, Banks L et al (2011) Successful targeting and disruption of an integrated reporter lentivirus using the engineered homing endonuclease Y2 I-AniI. PLoS One 6:e16825

Balciunas D, Wangensteen KJ, Wilber A et al (2006) Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. PLoS Genet 2:e169

Barber GN (2011) Innate immune DNA sensing pathways: STING, AIMII and the regulation of interferon production and inflammatory responses. Curr Opin Immunol 23:10–20

Barski A, Cuddapah S, Cui K et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 29:823–837

Bartholomew C, Ihle JN (1991) Retroviral insertions 90 kilobases proximal to the Evi-1 myeloid transforming transforming gene activate transcription from the normal promoter. Mol Cell Biol 11:1820–1828

Baum C, Düllmann J, Li Z et al (2003) Side effects of retroviral gene transfer into hematopoietic stem cells. Blood 101:2099–2114

Baum C, Kustikova O, Modlich U et al (2006) Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. Hum Gene Ther 17:252–263

Beerli RR, Barbas CF 3rd (2002) Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol 20:135–141

Ben Ali A, Wuyts J, De Wachter R et al (1999) Construction of a variability map for eukaryotic large subunit ribosomal RNA. Nucleic Acids Res 27:2825–2831

Biasco L, Ambrosi A, Pellin D et al (2010) Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell. EMBO Mol Med 3:89–101

Birling MC, Gofflot F, Warot X (2009) Site-specific recombinases for manipulation of the mouse genome. Methods Mol Biol 561:245–263

Bower DM, Prather KL (2009) Engineering of bacterial strains and vectors for the production of plasmid DNA. Appl Microbiol Biotechnol 82:805–813

Bowie AG, Fitzgerald KA (2007) RIG-I: tri-ing to discriminate between self and non-self RNA. Trends Immunol 28:147–150

Breckpot K, Escors D, Arce F et al (2010) HIV-1 lentiviral vector immunogenicity is mediated by Toll-like receptor 3 (TLR3) and TLR7. J Virol 84:5627–5636

Bronson SK, Plaehn EG, Kluckman KD et al (1996) Single-copy transgenic mice with chosen-site integration. Proc Natl Acad Sci USA 93:9067–9072

Cadinanos J, Bradley A (2007) Generation of an inducible and optimized piggyBac transposon system. Nucleic Acids Res 35:e87

Carnes AE, Luke JM, Vincent JM et al (2011) Plasmid DNA fermentation strain and process-specific effects on vector yield, quality, and transgene expression. Biotechnol Bioeng 108:354–363

Casola S (2010) Mouse models for miRNA expression: the ROSA26 locus. Methods Mol Biol 667:145–163

Cattoglio C, Facchini G, Sartori D et al (2007) Hot spots of retroviral integration in human CD34+ hematopoietic cells. Blood 110:1770–1778

Cattoglio C, Maruggi G, Bartholomae C et al (2010a) High-definition mapping of retroviral integration sites defines the fate of allogeneic T cells after donor lymphocyte infusion. PLoS One 5:e15688

Cattoglio C, Pellin D, Rizzi E et al (2010b) High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. Blood 116:5507–5517

Chen CM, Krohn J, Bhattacharya S et al (2011) A comparison of exogenous promoter activity at the ROSA26 locus using a PhiC31 integrase mediated cassette exchange approach in mouse ES cells. PLoS One 6:e23376

Ciceri F, Bonini C, Marktel S et al (2007) Antitumor effects of HSV-TK-engineered donor lymphocytes after allogeneic stem-cell transplantation. Blood 109:4698–4707

Ciceri F, Bonini C, Stanghellini MT et al (2009) Infusion of suicide-gene-engineered donor lymphocytes after family haploidentical haemopoietic stem-cell transplantation for leukaemia (the TK007 trial): a non-randomised phase I-II study. Lancet Oncol 10:489–500

Ciuffi A, Llano M, Poeschla E et al (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11:1287–1289

Cohen-Tannoudji M, Robine S, Choulika A et al (1998) I-SceI-induced gene replacement at a natural locus in embryonic stem cells. Mol Cell Biol 18:1444–1448

Cui K, Zang C, Roh TY et al (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell 4:80–93

Dalsgaard T, Moldt B, Sharma N et al (2009) Shielding of sleeping beauty DNA transposon-delivered transgene cassettes by heterologous insulators in early embryonal cells. Mol Ther 17:121–130

Dang CV, Lee WMF (1989) Nuclear and nucleolar targeting sequences of c-erb-A, c-myb, N-myc, p53, HSP70, and HIV tat proteins. J Biol Chem 264:18019–18023
de Felipe P (2002) Polycistronic viral vectors. Curr Gene Ther 2:355–378
DeKelver RC, Choi VM, Moehle EA et al (2010) Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res 20:1133–1142
Deyle DR, Russell DW (2009) Adeno-associated virus vector integration. Curr Opin Mol Ther 11:442–447
Dutheil N, Yoon-Robarts M, Ward P et al (2004) Characterization of the mouse adeno-associated virus AAVS1 ortholog. J Virol 78:8917–8921
Ehrhardt A, Engler JA, Xu H et al (2006) Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. Hum Gene Ther 17:1077–1094
Elbashir SM, Harborth J, Lendeckel W et al (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494–498
Ellis J (2005) Silencing and variegation of gammaretrovirus and lentivirus vectors. Hum Gene Ther 16:1241–1246
Emery DW (2011) The use of chromatin insulators to improve the expression and safety of integrating gene transfer vectors. Hum Gene Ther 22:761–774
Evans-Galea MV, Wielgosz MM, Hanawa H et al (2007) Suppression of clonal dominance in cultured human lymphoid cells by addition of the cHS4 insulator to a lentiviral vector. Mol Ther 15:801–809
Fehse B, Roeder I (2008) Insertional mutagenesis and clonal dominance: biological and statistical considerations. Gene Ther 15:143–153
Felice B, Cattoglio C, Cittaro D et al (2009) Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. PLoS One 4:e4571
Fire A, Xu S, Montgomery MK et al (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811
Fiszer-Kierzkowska A, Vydra N, Wysocka-Wycisk A et al (2011) Liposome-based DNA carriers may induce cellular stress response and change gene expression pattern in transfected cells. BMC Mol Biol 10:12–27
Galvan DL, Nakazawa Y, Kaja A et al (2009) Genome-wide mapping of PiggyBac transposon integrations in primary human T cells. J Immunother 32:837–844
Garrison BS, Yant SR, Mikkelsen JG, Kay MA (2007) Postintegrative gene silencing within the Sleeping Beauty transposition system. Mol Cell Biol 27:8824–8833
Gaszner M, Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet 7:703–713
Gautam A, Densmore CL, Waldrep JC (2001) Pulmonary cytokine responses associated with PEI-DNA aerosol gene therapy. Gene Ther 8:254–257
Georgopoulos LJ, Elgue G, Sanchez J et al (2009) Preclinical evaluation of innate immunity to baculovirus gene therapy vectors in whole human blood. Mol Immunol 46:2911–2917
Gersbach CA, Gaj T, Rm G et al (2011) Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. Nucleic Acids Res 39:7868–7878
Geurts AM, Collier LS, Geurts JL et al (2006) Gene mutations and genomic rearrangements in the mouse as a result of transposon mobilization from chromosomal concatemers. PLoS Genet 2:e156
Glimm H, Schmidt M, Fischer M et al (2005) Efficient marking of human cells with rapid but transient repopulating activity in autografted recipients. Blood 106:893–898
Golding MC, Mann MR (2011) A bidirectional promoter architecture enhances lentiviral transgenesis in embryonic and extraembryonic stem cells. Gene Ther 18:817–826
Gouble A, Smith J, Bruneau S et al (2006) Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. J Gene Med 8:616–622
Goulaouic H, Chow SA (1996) Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli LexA protein. J Virol 70:37–46

Grabundzija I, Irgang M, Mátés L et al (2010) Comparative analysis of transposable element vector systems in human cells. Mol Ther 18:1200–1209
Grewal SI, Moazed D (2003) Heterochromatin and epigenetic control of gene expression. Science 301:798–802
Grizot S, Smith J, Daboussi F et al (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res 37:5405–5419
Guo J, Gaj T, Barbas CF (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol 400:96–107
Hacein-Bey-Abina S, von Kalle C, Schmidt M et al (2003a) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. N Engl J Med 348:255–256
Hacein-Bey-Abina S, von Kalle C, Schmidt M et al (2003b) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 302:415–419
Hamady M, Walker JJ, Harris JK et al (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods 5:235–237
Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100:57–70
Händel EM, Cathomen T (2011) Zinc-finger nuclease based genome surgery: it's all about specificity. Curr Gene Ther 11:28–37
Hayashi T, Lamba DA, Slowik A et al (2010) A method for stabilizing RNA for transfection that allows control of expression duration. Dev Dyn 239:2034–2040
Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL (1997) The structure of I-CreI, a group I intron-encoded homing endonuclease. Nat Struct Biol 4:468–476
Hematti P, Hong BK, Ferguson C et al (2004) Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. PLoS Biol 2:e423
Henckaerts E, Linden RM (2010) Adeno-associated virus: a key to the human genome? Future Virol 5:555–574
Hodges BL, Taylor KM, Joseph MF et al (2004) Long-term transgene expression from plasmid DNA gene therapy vectors is negatively affected by CpG dinucleotides. Mol Ther 10:269–278
Huang X, Yang Y (2009) Innate immune recognition of viruses and viral vectors. Hum Gene Ther 20:293–301
Huang X, Guo H, Tammana S et al (2010) Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac transposons in human primary T cells. Mol Ther 18:1803–1813
Irion S, Luche H, Gadue P et al (2007) Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nat Biotechnol 25:1477–1482
Ishii KJ, Akira S (2005) TLR ignores methylated RNA? Immunity 23:111–113
Ishii KJ, Akira S (2008) Potential link between the immune system and metabolism of nucleic acids. Curr Opin Immunol 20:524–529
Ismail SI, Kingsman SM, Kingsman AJ, Uden M (2000) Split-intron retroviral vectors: enhanced expression with improved safety. J Virol 74:2365–2371
Ivics Z, Hackett PB, Plasterk RH, Izsvák Z (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91:501–510
Ivics Z, Katzer A, Stüwe EE, Fiedler D, Knespel S, Izsvák Z (2007) Targeted sleeping beauty transposition in human cells. Mol Ther 15:1137–1144
Jacobsen L, Calvin S, Lobenhofer E (2009) Transcriptional effects of transfection: the potential for misinterpretation of gene expression data generated from transiently transfected cells. Biotechniques 47:617–624
Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33(Suppl):245–254
Jähner D, Jaenisch R (1985) Chromosomal position and specific demethylation in enhancer sequences of germ line-transmitted retroviral genomes during mouse development. Mol Cell Biol 5:2212–2220
Jähner D, Stuhlmann H, Stewart CL et al (1982) De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature 298:623–628

Jasin M, Moynahan ME, Richardson C (1996) Targeted transgenesis. Proc Natl Acad Sci USA 93:8804–8808
Kameyam Y, Kawabe Y, Ito A, Kamihira M (2010) An accumulative site-specific gene integration system using cre recombinase-mediated cassette exchange. Biotechnol Bioeng 105: 1106–1114
Kaminski JM, Huber MR, Summers JB, Ward MB (2002) Design of a nonviral vector for site-selective, efficient integration into the human genome. FASEB J 6:1242–1247
Karikó K, Bhuyan P, Capodici J et al (2004) Exogenous siRNA mediates sequence-independent gene suppression by signaling through toll-like receptor 3. Cells Tissues Organs 177:132–138
Karikó K, Buckstein M, Ni H, Weissman D (2005) Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. Immunity 23:165–175
Kawashima T, Osanai M, Futahashi R (2007) A novel target-specific gene delivery system combining baculovirus and sequence-specific long interspersed nuclear elements. Virus Res 127:49–60
Kay MA, Glorioso JC, Naldini L (2001) Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. Nat Med 7:33–40
Kedes LH (1979) Histone genes and histone messengers. Annu Rev Biochem 48:837–870
Kim YK, Kim VN (2007) Processing of intronic microRNAs. EMBO J 26:775–783
Kojima KK, Fujiwara H (2003) Evolution of target specificity in R1 clade non-LTR retrotransposon. Mol Biol Evol 20:351–361
Kojima KK, Fujiwara H (2004) Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol Biol Evol 21:207–217
Kolb AF, Coates CJ, Kaminski JM et al (2005) Site-directed genome modification: nucleic acid and protein modules for targeted integration and gene correction. Trends Biotechnol 23:399–406
Kormann MS, Hasenpusch G, Aneja MK et al (2011) Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. Nat Biotechnol 29:154–157
Kouzarides T (2007) Chromatin modifications and their function. Cell 128:693–705
Kustikova OS, Wahlers A, Kuhlcke K et al (2003) Dose finding with retroviral vectors: Correlation of retroviral vector copy numbers in single cells with gene transfer efficiency in a cell population. Blood 102:3934–3937
Kustikova OS, Fehse B, Modlich U et al (2005) Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science 308:1171–1174
Kustikova OS, Baum C, Fehse B (2008) Retroviral integration site analysis in hematopoietic stem cells. Methods Mol Biol 430:255–267
Kustikova O, Brugman M, Baum C (2010) The genomic risk of somatic gene therapy. Semin Cancer Biol 20:269–278
Lam AP, Dean DA (2010) Progress and prospects: nuclear import of nonviral vectors. Gene Ther 17:439–447
Lamartina S, Sporeno E, Fattori E, Toniatti C (2000) Characteristics of the adeno-associated virus preintegration site in human chromosome 19: open chromatin conformation and transcription-competent environment. J Virol 74:7671–7677
Lechertier T, Sirri V, Hernandez-Verdun D, Roussel P (2007) A B23-interacting sequence as a tool to visualize protein interactions in a cellular context. J Cell Sci 120:265–375
Lesbats P, Botbol Y, Chevereau G et al (2011) Functional coupling between HIV-1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes. PLoS Pathog 7:e1001280
Lewinski MK, Yamashita M, Emerman M et al (2006) Retroviral DNA integration: viral and cellular determinants of target-site selection. PLoS Pathog 2:e60
Li CL, Xiong D, Stamatoyannopoulos G, Emery DW (2009) Genomic and functional assays demonstrate reduced gammaretroviral vector genotoxicity associated with use of the cHS4 chromatin insulator. Mol Ther 17:716–724
Liu G, Geurts AM, Yae K et al (2005) Target-site preferences of Sleeping Beauty transposons. J Mol Biol 346:161–173

Liu X, Liu M, Xue Z et al (2007) Non-viral ex vivo transduction of human hepatocyte cells to express factor VIII using a human ribosomal DNA-targeting vector. J Thromb Haemost 5: 347–351
Liu Y, Lakshmipathy U, Ozgenc A et al (2010) hESC engineering by integrase-mediated chromosomal targeting. Methods Mol Biol 584:229–268
Lombardo A, Cesana D, Genovese P et al (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat Methods 8:861–869
Lu Z, Ghazizadeh S (2007) Loss of transgene following ex vivo gene transfer is associated with a dominant Th2 response: implications for cutaneous gene therapy. Mol Ther 15:954–961
Maison C, Almouzni G (2004) HP1 and the dynamics of heterochromatin maintenance. Nat Rev Mol Cell Biol 5:296–304
Manno CS, Pierce GF, Arruda VR et al (2006) Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat Med 12:342–347
Maragathavally KJ, Kaminski JM, Coates CJ (2006) Chimeric Mos1 and piggyBac transposases result in site-directed integration. FASEB J 20:1880–1882
Markusic DM, de Waart DR, Seppen J (2010) Separating lentiviral vector injection and induction of gene expression in time, does not prevent an immune response to rtTA in rats. PLoS One 5:e9974
Marshall HM, Ronen K, Berry C et al (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS One 2:e1340
Mates L, Izsvak Z, Ivics Z (2007) Technology transfer from worms and flies to vertebrates: transposition-based genome manipulations and their future perspectives. Genome Biol 8(Suppl 1):S1
Mates L, Chuah MK, Belay E et al (2009) Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat Genet 41:753–761
McConnell Smith A, Takeuchi R, Pellenz S et al (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. Proc Natl Acad Sci USA 106:5099–5104
McKenna SA, Kim I, Puglisi EV et al (2007) Purification and characterization of transcribed RNAs using gel filtration chromatography. Nat Protoc 2:3270–3277
Meir YJ, Weirauch MT, Yang HS et al (2011) Genome-wide target profiling of piggyBac and Tol2 in HEK 293: pros and cons for gene discovery and gene therapy. BMC Biotechnol 11:28
Meister G, Tuschl T (2004) Mechanisms of gene silencing by double stranded RNA. Nature 431:343–349
Métais JY, Dunbar CE (2008) The MDS1-EVI1 gene complex as a retrovirus integration site: impact on behavior of hematopoietic cells and implications for gene therapy. Mol Ther 6:439–449
Mikkelsen JG, Yant SR, Meuse L et al (2003) Helper-independent Sleeping Beauty transposon-transposase vectors for efficient nonviral gene delivery and persistent gene expression *in vivo*. Mol Ther 8:654–665
Miller A, Dean D (2009) Tissue-specific and transcription factor-mediated nuclear entry of DNA. Adv Drug Deliv Rev 61:603–613
Mingozzi F, High KA (2011) Immune Responses to AAV in Clinical Trials. Curr Gene Ther 11:321–330
Miranda TB, Jones PA (2007) DNA methylation: the nuts and bolts of repression. J Cell Physiol 213:384–390
Miskey C, Izsvak Z, Kawakami K, Ivics Z (2005) DNA transposons in vertebrate functional genomics. Cell Mol Life Sci 62:629–641
Modlich U, Bohne J, Schmidt M et al (2006) Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. Blood 108:2545–2553
Modlich U, Navarro S, Zychlinski D et al (2009) Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. Mol Ther 17:1919–1928
Moldt B, Yant SR, Andersen PR et al (2007) Cis-acting gene regulatory activities in the terminal regions of sleeping beauty DNA transposon-based vectors. Hum Gene Ther 18:1193–1204
Montini E, Cesana D, Schmidt M et al (2006) Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. Nat Biotechnol 24:687–696

Montini E, Cesana D, Schmidt M et al (2009) The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J Clin Invest 119:964–975

Morales ME, Mann VH, Kines KJ et al (2007) PiggyBac transposon mediated transgenesis of the human blood fluke, Schistosoma mansoni. FASEB J 21:3479–3489

Moreno-Carranza B, Gentsch M, Stein S et al (2009) Transgene optimization significantly improves SIN vector titers, gp91phox expression and reconstitution of superoxide production in X-CGD cells. Gene Ther 16:111–118

Naldini L (2006) Inserting optimism into gene therapy. Nat Med 12:386–388

Nayak S, Herzog RW (2010) Progress and prospects: immune responses to viral vectors. Gene Ther 17:295–304

Niidome T, Huang L (2002) Gene therapy progress and prospects: nonviral vectors. Gene Ther 9:1647–1652

Niu Y, Tenney K, Li H, Gimble FS (2008) Engineering variants of the ISceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. J Mol Biol 382:188–202

Ogata T, Kozuka T, Kanda T (2003) Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. J Virol 77:9000–9007

Ott MG, Schmidt M, Schwarzwaelder K et al (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med 12:401–409

Palais G, Nguyen Dinh Cat A, Friedman H et al (2009) Targeted transgenesis at the HPRT locus: an efficient strategy to achieve tightly controlled in vivo conditional expression with the tet system. Physiol Genomics 37:140–146

Palazzoli F, Bire S, Bigot Y, Bonnin-Rouleux F (2011) Landscape of chromatin control element patents: positioning effects in pharmaceutical bioproduction. Nat Biotechnol 29:593–597

Papapetrou EP, Sadelain M (2011) Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector. Nat Protoc 6:1251–1273

Papapetrou EP, Lee G, Malani N et al (2011) Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. Nat Biotechnol 29:73–78

Park CW, Park J, Kren BT, Steer CJ (2006) Sleeping Beauty transposition in the mouse genome is associated with changes in DNA methylation at the site of insertion. Genomics 88:204–213

Paruzynski A, Arens A, Gabriel R et al (2010) Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. Nat Protoc 5:1379–1395

Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2 1 A. Science 252:809–817

Philpott NJ, Gomos J, Falck-Pedersen E (2004) Transgene expression after rep-mediated site-specific integration into chromosome 19. Hum Gene Ther 15:47–61

Pichon C, Billiet L, Midoux P (2010) Chemical vectors for gene delivery: uptake and intracellular trafficking. Curr Opin Biotechnol 21:640–645

Posfai G, Plunkett G III, Feher T et al (2006) Emergent properties of reduced genome *Escherichia coli*. Science 312:1044–1046

Radcliffe PA, Sion CJ, Wilkes FJ et al (2008) Analysis of factor VIII mediated suppression of lentiviral vector titres. Gene Ther 15:289–297

Ramachandra CJ, Shahbazi M, Kwang TW et al (2011) Efficient recombinase-mediated cassette exchange at the AAVS1 locus in human embryonic stem cells using baculoviral vectors. Nucleic Acids Res 39:e107

Recchia A, Bonini C, Magnani Z et al (2006) Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells. Proc Natl Acad Sci USA 103:1457–1462

Reyes-Sandoval A, Ertl HC (2004) CpG methylation of a plasmid vector results in extended transgene product expression by circumventing induction of immune responses. Mol Ther 9:249–261

Rossetti M, Gregori S, Hauben E et al (2011) HIV-1-derived lentiviral vectors directly activate plasmacytoid dendritic cells, which in turn induce the maturation of myeloid dendritic cells. Hum Gene Ther 22:177–188

Sakai K, Ohta T, Minoshima S et al (1995) Human ribosomal RNA gene cluster: identification of the proximal end containing a novel tandem repeat sequence. Genomics 26:521–526
Sakurai H, Kawabata K, Sakurai F et al (2008) Innate immune response induced by gene delivery vectors. Int J Pharm 354:9–15
Sandmeyer S (2003) Integration by design. Proc Natl Acad Sci USA 100:5586–5588
Santoro SW, Schultz PG (2002) Directed evolution of the site specificity of Cre recombinase. Proc Natl Acad Sci USA 99:4185–4190
Schambach A, Galla M, Maetzig T et al (2007) Improving transcriptional termination of self-inactivating gamma-retroviral and lentiviral vectors. Mol Ther 15:1167–73
Schmidt M, Carbonaro DA, Speckmann C et al (2003) Clonality analysis after retroviral-mediated gene transfer to CD34+ cells from the cord blood of ADA-deficient SCID neonates. Nat Med 9:463–468
Schmidt M, Schwarzwaelder K, Bartholomae C et al (2007) High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat Methods 4:1051–1057
Schroder AR, Shinn P, Chen H et al (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110:521–529
Shi X, Harrison RL, Hollister JR et al (2007) Construction and characterization of new piggyBac vectors for constitutive or inducible expression of heterologous gene pairs and the identification of a previously unrecognized activator sequence in *piggyBac*. BMC Biotechnol 7:5
Sioud M (2006) Innate sensing of self and non-self RNAs by Toll-like receptors. Trends Mol Med 12:167–176
Sjeklocha L, Chen Y, Daly MC et al (2011) β-globin matrix attachment region improves stable genomic expression of the Sleeping Beauty transposon. J Cell Biochem 112:2361–2375
Smith J, Grizot S, Arnould S et al (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 34:e149
Sorrell DA, Kolb AF (2005) Targeted modification of mammalian genomes. Biotechnol Adv 23:431–469
Strathdee D, Ibbotson H, Grant SG (2006) Expression of transgenes targeted to the Gt(ROSA)26Sor locus is orientation dependent. PLoS One 1:e4
Su K, Wang D, Ye J et al (2009) Site-specific integration of retroviral DNA in human cells using fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc-finger protein E2C. Methods 47:269–276
Sumiyoshi T, Holt NG, Hollis RP et al (2009) DB Stable transgene expression in primitive human CD34+ hematopoietic stem/progenitor cells using the Sleeping Beauty transposon system. Hum Gene Ther 20:1607–1626
Suzuki R, Oda Y, Utoguchi N, Maruyama K (2011) Progress in the development of ultrasound-mediated gene delivery systems utilizing nanoand microbubbles. J Control Release 149: 36–41
Szczepek M, Brondani V, Büchel J et al (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nature Biotech 25:786–793
Takahashi H, Fujiwara H (1999) Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm Bombyx mori. Nucleic Acids Res 27:2015–2021
Takahashi H, Fujiwara H (2002) Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. EMBO J 21:408–417
Takiguchi N, Takahashi Y, Nishikawa M et al (2011) Positive correlation between the generation of reactive oxygen species and activation/reactivation of transgene expression after hydrodynamic injections into mice. Pharm Res 28:702–711
Tham WH, Zakian VA (2002) Transcriptional silencing at Saccharomyces telomeres: implications for other organisms. Oncogene 21:512–521
Thierry A, Dujon B (1992) Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes. Nucleic Acids Res 20: 5625–5631
Thyagarajan B, Olivares EC, Hollis RP et al (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol Cell Biol 21:3926–3934

Trinh AT, Ball BG, Weber E et al (2009) Retroviral vectors encoding ADA regulatory locus control region provide enhanced T-cell-specific transgene expression. Genet Vaccines Ther 30:7–13
Turan S, Galla M, Ernst E et al (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. J Mol Biol 407:193–221
Turlure F, Devroe E, Silver PA et al (2004) Human cell proteins and human immunodeficiency virus DNA integration. Front Biosci 9:3187–3208
Un K, Kawakami S, Suzuki R et al (2011) Suppression of melanoma growth and metastasis by DNA vaccination using an ultrasound-responsive and mannose-modified gene carrier. Mol Pharm 8:543–554
Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646
Valenzuela L, Kamakaka RT (2006) Chromatin insulators. Annu Rev Genet 40:107–138
Vigdal TJ, Kaufman CD, Izsvak Z et al (2002) Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. J Mol Biol 323:441–452
Violot S, Hong SS, Rakotobe D et al (2003) The human polycomb group EED protein interacts with the integrase of human immunodeficiency virus type 1. J Virol 77:12507–12522
Voigt K, Izsvák Z, Ivics Z (2008) Targeted gene insertion for molecular medicine. J Mol Med 86:1205–1219
Voziyanov Y, Konieczka JH, Stewart AF, Jayaram M (2003) Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site. J Mol Biol 326:65–76
Wade-Martins R (2011) Developing extrachromosomal gene expression vector technologies: an overview. Methods Mol Biol 738:1–17
Wahlers A, Schwieger M, Li Z et al (2001) Influence of multiplicity of infection and protein stability on retroviral vector mediated gene expression in hematopoietic cells. Gene Ther 8:477–486
Walisko O, Izsvák Z, Szabó K et al (2006) Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1. Proc Natl Acad Sci USA 103:4062–4067
Ward NJ, Buckley SM, Waddington SN (2011) Codon optimization of human factor VIII cDNAs leads to high-level expression. Blood 117:798–807
Wegel E, Shaw P (2005) Gene activation and deactivation related changes in the three-dimensional structure of chromatin. Chromosoma 114:331–337
Wei G, Wei L, Zhu J (2009) Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity 30:155–167
Wheelan S, Scheifele LZ, Martínez-Murillo F et al (2006) Transposon insertion site profiling chip (TIP-chip). Proc Natl Acad Sci USA 103:17632–17637
Wilber A, Frandsen JL, Geurts JL et al (2006) RNA as a source of transposase for Sleeping Beauty-mediated gene insertion and expression in somatic cells and tissues. Mol Ther 13:625–630
Wilson MH, George AL Jr (2010) Designing and testing chimeric zinc finger transposases. Methods Mol Biol 649:353–363
Wirth D, Gama-Norton L, Riemer P et al (2007) Road to precision: recombinase-based targeting technologies for genome engineering. Curr Opin Biotechnol 18:411–419
Woltjen K, Mickael IP, Mohseni P et al (2009) piggyBac transposition reprograms fIbroblasts to induced pluripotent stem cells. Nature 458:766–770
Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 300:1749–1751
Wu SC, Meir YJ, Coates CJ et al (2006) PiggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. Proc Natl Acad Sci USA 103:15008–15013
Wu Z, Sun J, Zhang T et al (2008) Optimization of self-complementary AAV vectors for liverdirected expression results in sustained correction of hemophilia B at low vector dose. Mol Ther 16:280–289
Xie W, Gai X, Zhu Y (2001) Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol Cell Biol 21:6606–6614

Xue X, Huang X, Nodland SE et al (2009) Stable gene transfer and expression in cord blood-derived CD34+ hematopoietic stem and progenitor cells by a hyperactive Sleeping Beauty transposon system. Blood 114:1319–1330

Yamamoto A, Kormann M, Rosenecker J et al (2009) Current prospects for mRNA gene delivery. Eur J Pharm Biopharm 71:484–489

Yant SR, Kay MA (2003) Nonhomologous-end-joining factors regulate DNA repair fidelity during Sleeping Beauty element transposition in mammalian cells. Mol Cell Biol 23:8505–8518

Yant SR, Wu X, Huang Y et al (2005) High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol 25:2085–2094

Yant SR, Huang Y, Akache B, Kay MA (2007) Site-directed transposon integration in human cells. Nucleic Acids Res 35:e50

Yao S, Sukonnik T, Kean T, Bharadwaj RR, Pasceri P, Ellis J (2004) Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications. Mol Ther 10:27–36

Yew NS (2005) Controlling the kinetics of transgene expression by plasmid design. Adv Drug Deliv Rev 57:769–780

Yew NS, Zhao H, Przybylska M et al (2002) CpG-depleted plasmid DNA vectors with enhanced safety and long-term gene expression *in vivo*. Mol Ther 5:731–738

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–340

Yoder KE, Espeseth A, Wang XH et al (2011) The base excision repair pathway is required for efficient lentivirus integration. PLoS One 6:e17862

Yusa K, Takeda J, Horie K (2004) Enhancement of Sleeping Beauty transposition by CpG methylation: possible role of heterochromatin formation. Mol Cell Biol 24:4004–4018

Yusa K, Rad R, Takeda J, Bradley A (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. Nat Methods 6:363–369

Zaldumbide A, Hoeben RC (2008) How not to be seen: immune-evasion strategies in gene therapy. Gene Ther 15:239–246

Zambrowicz BP, Imamoto A, Fiering S et al (1997) Disruption of overlapping transcripts in the ROSA βgeo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells. Proc Natl Acad Sci USA 94:3789–3794

Zayed H, Izsvaak S, Khare D et al (2003) The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. Nucleic Acids Res 31:2313–2322

Zeh DW, Zeh JA, Ishida Y (2009) Transposable elements and an epigenetic basis for punctuated equilibria. Bioessays 31:715–726

Zhang C (2009) Novel functions for small RNA molecules 1. Curr Opin Mol Ther 11:641–651

Zhou R, Norton JE, Zhang N, Dean DA (2007) Electroporation-mediated transfer of plasmids to the lung results in reduced TLR9 signaling and inflammation. Gene Ther 14:775–780

Zhou H, Liu ZG, Sun ZW et al (2010) Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. J Biotechnol 147:122–129

Zhu J, Park CW, Sjeklocha L et al (2010) High-level genomic integration, epigenetic changes, and expression of sleeping beauty transgene. Biochemistry 49:1507–1521

Zychlinski D, Schambach A, Modlich U et al (2008) Physiological promoters reduce the genotoxic risk of integrating gene vectors. Mol Ther 16:718–725

Chapter 2
Designing Non-viral Targeted Integrating Vectors for Genome Engineering in Vertebrates

Ludivine Sinzelle and Nicolas Pollet

Abstract Genome engineering for biomedicine and biotechnology requires the integration of transgenes at specific sites in safe genomic environments. These specific integrations are needed to avoid position effects, insertional mutagenesis and chromosome abnormalities. Several efforts during the last decade led to the development of nonviral approaches based on DNA-modifying enzymes by exploiting the cellular mechanisms of cut-and-paste transposition and homologous recombination (HR). These methods include the use of zinc finger nucleases, meganucleases, site-specific recombinases such as ΦC31 integrase, Cre and Flp recombinases and transposase-based systems to achieve the integration of foreign DNA at a desired genomic position. Moreover, many teams are now investigating strategies to alter the site-specificity of these enzymes and precisely target safe insertion sites.

This review attempts to provide a general overview on recent advances in designing nonviral site-directed integrating vectors, required for both gene therapy and animal transgenesis. We discuss the advantages of such engineered vectors, the strategies employed to improve their targeting efficiency and consider their limitations in terms of safety and activity.

Keywords DNA transposons • Zinc-finger nuclease • TALE • ΦC31 integrase • Meganucleases • Site-specific recombinases

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

N. Pollet (✉)
Institute of Systems and Synthetic Biology, Genopole, CNRS, Université d'Evry Val d'Essonne, Genavenir 3, Genopole campus 1, 1, rue Pierre Fontaine, 91058 Evry, France
e-mail: Nicolas.Pollet@issb.genopole.fr

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*, Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_2,

Abbreviations

DSB	Double strand break
ZF	Zinc finger
gDNA	Genomic DNA
HR	Homologous recombination
ITR	Inverted terminal repeat

2.1 Introduction: Ideal Integrating Vectors and Safe Insertion Sites

The successful treatment of diseases using viral-mediated gene therapy has been hampered by subsequent adverse events. These events were due to retroviral insertional activation of host genes (proto-oncogenes) with uncontrolled cell proliferation or to innate immune response in the case of adenoviral vectors (for reviews, Yi et al. 2011; Thaci et al. 2011). Such negative consequences caused by random integration were encountered for example in X-Linked severe combined immunodeficiency (X-SCID) patients with the development of leukemia after the treatments using retroviral vectors (Hacein-Bey-abina et al. 2002). These reasons explain the development of efficient targeted insertion system to obtain safer gene transfer for biomedical applications. The reasoning is quite similar concerning gene transfer for synthetic biology, bioproduction or for fundamental scientific purposes.

The first question that comes to mind when considering targeted insertions is what can be considered a safe insertion site, also called a safe harbor (Silva et al. 2011). An ideal and safe integration site must fulfil two principal characteristics: i/ the possibility for the transgene to be transcribed (to ensure its long-term and stable expression) and eventually its regulation at the required therapeutic dose and ii/ a non-pathogenic impact on the function of the genome (DNA replication, repair, transcription, recombination). Whether these two characteristics constitute an achievable goal remains a central question. The bad news is that the vast majority (93%) of the human genome is transcribed (Carninci et al. 2008), and that we do not yet have the knowledge of all functional elements in the human genome. Gene deserts that cover 25% of the human genome could be of interest for gene targeting (Ovcharenko et al. 2005). They consist of large intergenic regions with no protein-coding sequences and some of these gene deserts are apparently nonessential, without any biological significance.

Nevertheless, the spatial organisation of the chromosomes is also an important consideration since it is known to impact chromatin structure and therefore influence transcriptional activities. The good news is that this knowledge increases rapidly thanks to the recent progresses of the Encyclopedia of DNA Elements (ENCODE)

type of strategies (ENCODE Project Consortium 2004), and that, after all, some transcripts might be affected in certain cells without perturbing their physiology (Dutheil et al. 2000). This aspect is maybe the most difficult problem since it turned out that insertion into transcription units may be quite safe. This is indeed the observation coming from studies on the ROSA26 locus in mice and AAVS1 in humans (Zambrowicz et al. 1997). Another candidate used in humans is the exon 3 of the chemokine receptor 5 (*CCR5*) «good for nothing» gene, since individuals homozygous for a CCR5Δ32 mutations seem healthy (Lim et al. 2006). The fact that these loci were identified without the purpose to identify safe integration sites is remarkable and can lead to three remarks. First, a rationale design of such safe integration sites would have put these aside since they are within or nearby transcriptional units. Second, we can now think of strategies to screen at random for such sites, with specific readouts pertaining to the cell types being targeted for gene therapy. Third, we can search similar sites taking into account the wealth of available genomic signatures (CpG, chromatin status, transcriptome, sequence polymorphism, evolutionary conservation). A genetic screen for such safe harbors was performed recently using such an approach (Papapetrou et al. 2011). The authors used five criterias to define a locus as being a safe harbor: 1/ reasonable distance (more than 50 kbp) from any promoter, 2/ far distance (more than 300 kbp) from any cancer-related gene, 3/ far distance (more than 300 kbp) from any microRNA gene, 4/ mapping outside of a known transcription unit and 5/ mapping outside ultraconserved regions (UCRs). After analysing 5840 lentiviral integration sites in induced pluripotent human cells (iPS), the authors found that 17.3% of integrations events met their criteria of a safe harbor. In iPS cell clones, the percentage was only 8.3% (N = 36). In comparison with a theoretical value of 38% for the existence of such sites in the human genome, these results show that safe harbors are screenable targets.

Several features of well known DNA loci or stuctures were found attractive to design safe integration sites. This is the case of rDNA loci in which successful integration and expression of a transgene without disturbing cell functions were reported (Liu et al. 2007). Based on previous studies, Kaminski et al. 2002 proposed to target some repetitive elements such as short and long interspersed nuclear elements (SINEs and LINEs). Another case of potential safe target loci to minimize position effects involve scaffold/matrix attachment regions (S/MARs) sequences that mediate structural organization of the chromatine (Stief et al. 1989).

The most popular vectors currently used in gene therapy are based on retroviruses, lentiviruses, adenovirus, adeno-associated viruses-bare (for reviews, Yi et al. 2011; Thaci et al. 2011). However, as described above, the major limitation of such viral vectors concerns their safety. Therefore, several nonviral approaches using homologous recombination (HR) were developed for gene targeting. In this review, we focus on progress made in the design of nonviral vectors. These emerging technologies are based on the use of zinc-finger nucleases, meganucleases, site-specific recombinases, native transposases as well as engineered

enzymes with modified integration site-specificity. These techniques hold great promises to target a transgene to a safe chosen genomic site for gene therapy applications.

2.2 Zinc Finger Nucleases

HR stands as a useful tool enabling specific insertion of foreign DNA, but it suffers from its low frequency in vertebrate cells. Experiments using the meganuclease I-*Sce*I showed that the generation of a chromosomal DSB strongly stimulates HR frequencies in mammalian cells compared to spontaneous HR, leading to the recruitment of the host cellular homologous repair machinery (Cohen-Tannoudji et al. 1998; Pipiras et al. 1998). Nowadays, classical HR strategies are based on the introduction of DSB into a targeted locus *via* the use of ZF nucleases (ZFN), meganucleases and site-specific recombinases (Sorrell and Kolb 2005).

Recently, the ZFN technology was successfully applied to build a wide range of modified organisms or cell lines with gene-targeted deletions or mutations (For reviews, Remy et al. 2010; Handel and Cathomen 2011). Futhermore, clinical trials based on ZF nucleases technology have been initiated.

ZFN are hybrid molecules, composed of ZF modules and the non-specific cleavage domain of *Fok*I type IIS restriction endonuclease. These nucleases catalyze a sequence-specific cleavage that induces a DNA damage response at the cleavage site to repair the double strand break (DSB). The cell can then use two mechanisms to repair such DSBs: nonhomologous end joining (NHEJ) and homologous recombination (HR). The NHEJ machinery will typically generate small insertions and deletions that often produce null or hypomorphic alleles, and this is widely used for genome editing applications to produce genetic inactivation. HR will result in sequence replacement if a suitable DNA template is provided, and this is used to perform gene knock-ins or targeted insertions.

ZFN proteins contain ZF modules, each of them being able to bind a sequence of three nucleotides. By combining three to six ZF modules, it is possible to target a unique sequence at a desired chromosomal locus in the context of a large genome. Zinc-finger nuclease must be provided in pairs to provide two DNA binding domains with adjacent specificity in the opposite orientation and with the correct spacing to cut both strands because *Fok*I must dimerize to cleave DNA. With their modular structure and function, the recognition specificities of zinc finger proteins can be easily manipulated experimentally leading to custom-designed ZF DBDs (For review, Wu et al. 2007). As an example, fusion proteins of the HIV-1 integrase and the ZF E2C have been reported to restrict integration of retroviral DNA in vitro to a 10-bp region flanking the E2C recognition sequence, located in the 5′ untranslated region of the erbB-2 gene (Tan et al. 2004). Even though ZFN are extremely useful, they suffer from certain limitations. First of all, the combination of multiple ZF modules to target unique sequences in a given genome is an empirical process. Secondly, ZFN can trigger off-target cleavage, and therefore they can be toxic to the cells.

2.3 TALE Nucleases

Studies on the interactions between plants and their pathogens have lead to the identification of a new family of DNA binding proteins named transcription activator-like effectors (TALEs) (Bonas et al. 1989). The modification of TALEs to make them bind a DNA sequence of interest was shown recently by several groups. The DNA binding capacity of TALEs is provided by simple protein motifs that were found easy to engineer (Boch 2011). The TALE protein domain that mediates DNA binding is made of an array of 33–35 amino acids repeat units, each of them being able to bind a single target base. The code used by TALEs was cracked and it was shown that a simple relationship exists between a target DNA base and two variable amino acids from the repeat monomers (Boch et al. 2009; Moscou and Bogdanove 2009). This code can be used to design proteins able to bind a DNA sequence of choice. The field of TALE nuclease is still in its infancy, but recent progress showed that TALEs can be used for genome editing using NHEJ and HR (Miller et al. 2011). Like in ZFN constructs, such genome targeting applications use an engineered TALE DNA-binding domain fused to the *Fok*I nuclease. The potential for TALE nucleases is of special interest because their DNA-binding code is flexible: it involves the specification of a single DNA base instead of a triplet for ZFN. However, the design and use of TAL effectors is a novel field and whether it competes or complements ZFNs remains to be seen.

2.4 Meganuclease-Based Vectors

Meganucleases (MNs, also-called homing endonucleases) are highly specific endonucleases that mediate site-specific DSBs and stimulate HR up to 10,000-fold. They are encoded by mobile introns or inteins in the *Saccharomyces cerevisiae* genome and behave like rare-cutting endonucleases with long (12–45 bp) target sites. MNs are classified into five families according to their sequence motif and DNA cleavage mechanism. The LAGLIDADG family and its well-studied members I-*Sce*I and I-*Cre*I are the best characterized (Grishin et al. 2010). I-*Sce*I enzyme functions as monomere containing two subdomains and targets a 18-bp non palindromic sequence whereas I-*Cre*I can be homodimeric and targets a 22-bp palindromic or pseudo-palindomic sequences.

As for Cre/*LoxP* and Flp/*FRT* recombination systems, MNs-based gene targeting has one limitation since the targeted site must contain a recognition site in its vicinity. However, vertebrate genomes analysed so far do not contain any I-*Sce*I and I-*Cre*I recognition sites and this system can not be used for targeting of endogenous genes. In light of this limitation, protein-engineering strategies methods have led to the emergence of MNs variants such as derivatives of I-*Sce*I, I-*Cre*I and I-*Dmo*I MNs with altered site-specificity (reviewed in Arnould et al. 2011). These novel enzymes can cleave chosen targets while keeping high cleavage and recombination activities with narrow specificity (Table 2.1). Several strategies were employed to create such

Table 2.1 Use of native I-SceI for gene transfer in vertebrates

Derivatives enzymes	Organism or cell line	Targeting efficiency	References
engineered single chain I-CreI (scI-ICreI)	COS cells and yeast cells	Induction of HR in yeast and mammalian cells	Epinat et al. (2003)
Chimeric MNs by fusing to I-CreI monomer the N-terminal domain of I-DmoI (DmoCre)	COS cells and yeast cells	Unable to mediate HR in yeast	Epinat et al. (2003)
Heterodimeric enzyme V2/V3 and single chain enzymes derived from I-CreI (RAG1 MNs)	Human cells (293 cell line)	Frequency of recombination up to 6% in the Human *RAG1* gene for the single-chain variant	Grizot et al. (2009)
DmoCreV5	Yeast and chinese hamster ovary cells	High target specificity	Grizot et al. (2010)
Variants of I-CreI	Chinese hamster ovary cells	Target sequences within XPC gene Gene targeting efficiency similar to I-CreI and I-SceI	Arnould et al. (2007)
Variants of I-CreI	Chinese hamster ovary cells	50 Cre variants with altered specificity and cleaved Exactly one target	Arnould et al. (2006)
Variants of V2/V3 I-CreI and variants of RAG1 MN	293 H and MRC5 human cell lines	Frequency of recombination 5.6% and 4.4% of the Transfected cells in *RAG1* gene	Munoz et al. (2011)
Variants of MsoI	In vitro assays	Cleavage of the redesigned recognition site 10 000 Times more effectively than the wild-type enzyme	Ashworth et al. (2006)
Engineered fusion of the N-terminal domain of I-DmoI to an I-CreI monomer resulting in E-DreI	In vitro assays	Highly specific	Chevalier et al. (2002)
Variants of I-SceI	In vivo assays	Greater specificity than native Isce-I for a given target	Joshi et al. (2011)

variants including the local modification of the specificity (Arnould et al. 2006), the monomerization of the enzyme (Epinat et al. 2003; Grizot et al. 2009), combinatorial strategy (Grizot et al. 2010) and domains swapping or fusion of DNA binding domains (Chevalier et al. 2002; Epinat et al. 2003). For example, the chimeric enzymes DmoCre and DmoCreV5, generated by fusing the N-terminal domain of I-*Dmo*I to a I-*Cre*I monomer, have been shown to acquire new target site specificity by cleaving the hybrid sequence corresponding to the fusion of the two half parent target sequences (Epinat et al. 2003 ; Grizot et al. 2010).

With the recent advances on engineered enzymes with altered site-specificity, MNs-based targeting are considered as one of the emerging technologies for gene therapy application (For a review, Silva et al. 2011). Several studies reported the effective targeting of desired human genes involved in monogenetic diseases. In these cases, the targeted sequences diverged from the MN native cleavage site over their entire length. An artificial engineered I-*Cre*I enzyme was created to specifically modify a human chromosomal locus within the human *XPC* gene involved in *Xeroderma pigmentosum* genetic disease (Arnould et al. 2007). Another derivative of I-*Cre*I was designed to cleave a natural sequence located in the human *RAG1* gene in which mutations cause SCID (Grizot et al. 2009; Munoz et al. 2011). I-*Cre*I variants induced high levels of targeted recombination at the endogenous locus of the *RAG1* gene in up to 6% of the transfected human cells. Recently, MNs have been used to inactivate viruses in a proof of concept study (Grosse et al. 2011). In this report, MNs were generated to recognize sequences of the Herpes simplex virus type 1(HSV1) and were shown to inhibit viral infection.

MNs-based technologies are also currently available for somatic and germinal transgenesis (Table 2.2). Transgenic animals, including various species of fish and amphibian were obtained by co-injection of the I-*Sce*I enzyme together with a reporter plasmid that contain the gene-of-interest flanked by I-*Sce*I recognition sites (Fig. 2.1a). I-*Sce*I-based transgenesis is characterized by a small copy number of integrated transgene and a high transgenesis frequency ranging from 20% to 33% (Table 2.2). However, as vertebrate genomes do not contain any I-*Sce*I recognition sites, the roles played by the MN in transgenesis remain unknown and require further investigations.

2.5 Site-Specific Recombinase-Based Integrating Vectors

In recent years, site-specific recombinases emerged as powerful new tools for gene targeting in both eukaryotic and prokaryotic genomes.

Site-specific recombinases are classified into two families according to their sequence similarity and their reaction mechanism, integrases (tyrosine recombinases) and resolvases/invertases (serine recombinases). The three most popular and well-studied site-specific gene recombination systems includes Cre recombinase, Flp recombinase and ΦC31 integrase. These systems are currently investigated for targeted recombination of transgenes into the genome of mammalian cells.

Table 2.2 Engineered meganucleases and their respective targeting efficiency

Applications	Transgene	Organism or cell line	Integration	Transgene expression and system efficiency	References
Germinal transgenesis	Various tissue-specific and ubiquitous promoters driving eGFP	Fishes (medaka, stickleback, zebrafish)	Single copy integrations	Threefold increase frequency of transgenic F1 offspring Compared to that of DNA injection.Germline transmission 25–50%	Grabher and Wittbrodt (2008)
	Various tissue-specific and ubiquitous promoters driving eGFP	Amphibians (*Xenopus*)	Small copy number of integrated transgene	Transgenesis frequency 20–30% Germline transmission rate from 17% to 69%	Ogino et al. (2006)
	Various tissue-specific and ubiquitous promoters driving eGFP	Amphibians (axolotl)	Undetermined	Transgenesis frequency 32.7% Germline transmission rate from 1% to 52%	Sobkow et al. (2006)
	Various tissue-specific and ubiquitous promoters driving eGFP	Amphibian (newts)	–	Transgenesis frequency 20%	Casco-Robles et al. (2010)
Somatic transgenesis	LagoZ gene	Mice liver	Targeted genomic recombination in toto	–	Gouble et al. (2006)
	5 different genes-of-interest	Chinese hamster ovary cell line	Efficient	Stable transgene expression over time	Cabaniols et al. (2010)
	Hygromycin gene	Mouse embryonic stem cells	Introduction of the I-Sce-I site Into the endogenous villin locus	Homologous recombination frequency 6×10-6	Cohen-Tannoudji et al. (1998)
	Neomycin	NIH 3 T3 and PCC7-S mouse cell lines	I-SceI recognition sites placed on a chromosome	Site-directed HR 10 times more frequent that random integration	Choulika et al. (1995)
	LacZ	RCCD1 rat epithelial cell line	–	A targeting event in 10–6 transfected cells	Puttini et al. (2005)

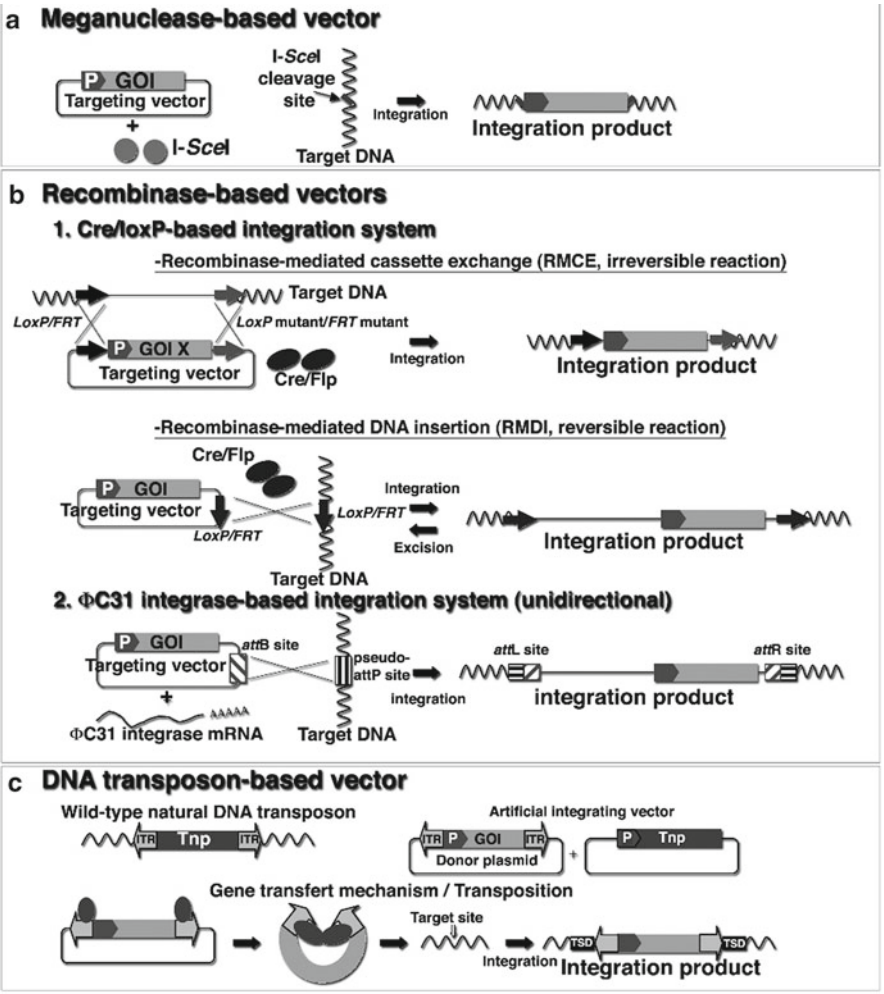


Fig. 2.1 Integration mechanisms for current non-viral integrating vector systems. (**a**) **Meganuclease I-*Sce*I-based vector**. Meganuclease-based integration systems are composed of a donor plasmid containing the gene-of-interest (GOI) flanked by one or two I-*Sce*I cleavage sites. The acceptor genomic locus can contain a I-*Sce*I recognition site. (**b**) **Recombinase-based vector. 1. Cre/*LoxP* and Flp/*FRT*-based systems.** In the recombinase-mediated cassette exchange (RMCE) technology, the targeting vector contains a *LoxP* or a *FRT* site (*dark arrows*) recognized by the recombinase Cre and Flp, respectively and a heterospecific mutant *LoxP/FRT* site (*grey arrows*), flanking the gene-of-interest. Recombination occurs between the *LoxP/FRT* site containing vector and a *LoxP/FRT* site containing gDNA, and between the *LoxP/FRT* mutant containing vector and a *LoxP/FRT* site containing DNA. The use of heterospecific target sites prevents the occurrence of reverse reactions. In recombinase-mediated DNA insertion, the recombination is mediated by the Cre/Flp between a *LoxP/FRT*-bearing vector and a *LoxP/FRT* site containing gDNA resulting in the integration of the complete targeting vector DNA. The reaction is reversible and excision can occur. **2. ΦC31 Integrase-based system.** The targeting vector contains the GOI, its promoter and an attB site. The ΦC31 integrase is supplied *in trans* and performs recombination

2.5.1 Site-Specific Tyrosine Recombinases-Based Vectors

Tyrosine recombinases are at the heart of two well-known recombination systems : Cre (causes recombination)/*loxP* (locus of crossover in phage P1) from bacteriophage P1 (*E. coli*) and Flp (Flippase) /*FRT* (Flippase Recognition Target) from *Saccharomyces cerevisiae*. Cre and Flp recombinase systems mediate recombination of a DNA fragment located between two 34-bp recognition sites, *LoxP* for Cre and *FRT* for FLP, and this without the need for cofactors (for a review, Turan et al. 2011). *LoxP* and *FRT* sites consist of two 13bp inverted repeats flanking a 8-bp unidirectional spacer core region. In contrast to ΦC31 integrase that only supports integration (unidirectional reaction), Cre and Flp recombinases catalyse reversible reaction since they create two identical cis-positioned sites which are immediate substrates for deletion or inversion when the loxP sites are oriented in the same or opposite direction, respectively.

Both systems have demonstrated their utility in conditional manipulation of gene expression by using inducible and tissue-specific recombinase expression systems in mice and in other organisms or cells such as *Xenopus* and zebrafish for the study of development or diseases (for a review, Birling et al. 2009 ; Rankin et al. 2009; Wong et al. 2010).

A technique named recombinase-mediated cassette exchange (RMCE) is used for gene transfer, and allows a stable and targeted integration of transgenes in mammals (Feng et al. 1999 ; Lauth et al. 2000). Flp-RMCE and Cre-RMCE technologies are based on the exchange of transgenes cassettes through the use of a set of two non-interacting *FRT* or *LoxP* recombination target sites, respectively. The first step of these processes requires the insertion of two heterospecific and incompatible *loxP* or *FRT* sites at random or by HR in a chromosome (Fig. 2.1b, upper panel). Then, the recombinase can mediate the replacement of the transgene cassette between a vector carrying the same arrangements of matching and heterospecific *loxP* or *FRT* recombination sites and the chromosomal locus (Fig. 2.1b, upper panel). The use of two heterospecific target sites prevents the excision reaction and genomic insertion of donor plasmid sequence, in contrast to the recombinase-mediated DNA insertion (RMDI) process in which recombination occurs between one *loxP/FRT* site inserted

Fig. 2.1 (continued) between the attB site from the plasmid and an attP or pseudo-attP sites. The recombination product leads to the integration of the targeting plasmid with the creation of hybrid attL and attR sites that are no longer substrates for the integrase without cofactors. (**c**) **DNA transposon-derived vector.** A wild-type natural DNA transposon contains a single gene encoding the transposase (Tnp), flanked by inverted terminal repeats (ITRs, represented by *grey arrows*). In the two-component delivery system, the transposase gene is replaced by a gene-of-interest and its regulatory sequences (*P* promoter), resulting in a non-autonomous transposon carried by a donor plasmid. The transposase is maintained on a separate plasmid (helper plasmid) and is supplied *in trans*. During transposition, the transposase binds to specific binding sites within the ITRs to form a synaptic complex. This complex is excised from the donor plasmid and integrates into a new genomic DNA (*gDNA*) locus. DNA transposons integrate into a 2–4 pb sequence duplicated upon insertion that flanks the integrated transposon (target site duplication, *TSD*)

in the genome and a plasmid-containing one *loxP/FRT* site that permits reverse reaction. This RMCE process has been successfully combined with viral delivery system. As a recent example, Sorrell et al. 2010 have used an adenovirus vector to deliver the transgene cassette and Cre recombinase in human cells and showed efficient integration of the transgene into a defined genomic site.

Several variants of *loxP* or *FRT* recombination target sites that fall into two classes (spacer variants and inverted-repeat variants) have been also developed to mediate efficient and highly site-selective RMCE and tested in various cell lines (For reviews, Wirth et al. 2007; Turan et al. 2011).

Another way to achieve Cre and Flp-based targeting is to identify native genomic sequences that could be used as recombination sites. Since the 34bp *LoxP* and *FRT* recognition sites are not naturally present within eukaryotic genomes, considerable efforts have been made to identify endogenous pseudo-sites that can act as efficient substrates for these prokaryotic recombinases. Functional *LoxP* pseudo-sites identified within the mouse and human genomes, were able to allow Cre-mediated recombination in bacterial assays (Thyagarajan et al. 2000). More recently, Cre-mediated integration of a transgene into a genomic endogenous *LoxP* pseudo-site was documented in early mouse embryos (Ito et al. 2011). Transgenic mice were generated with a high transgenesis efficiency by microinjection of a Cre-expressing plasmid and a transgene containing pseudo-LoxP sites (pseudo-loxm5) into fertilized oocytes, and the transgene was found to be targeted to the endogenous pseudo-loxm5 sequence.

An alternative strategy to target efficient integration using native genomic recombination sites consists to build variants of recombinases with altered target site specificity and that are able to recognize pre-existing genomic sequences that can be used as recombination sites. Following this strategy, a sequence (termed loxH) chosen from chromosome 22 was used as a new target DNA sequence for specifically designed Cre recombinase variants (Buchholz and Stewart 2001). Similarly, derivatives of Cre recombinase generated by targeted mutagenesis were found to specifically recognize unatural recombination sites (Santoro and Schultz 2003). A new recombinase termed Tre, derived from Cre-recombinase, was evolved *in vitro* to recognize the lox-like long terminal repeat of a specific HIV-1 subtype (Sarkar et al. 2007). Recently, a web server (SeLOX) has been designed to identify lox-like sites for genomic recombination studies (Surendranath et al. 2010).

In the case of Flp recombinase, recent investigations on human genome sequences have allowed to create a database of functional *FRT*-like sequences which could serve as selection sites to generate Flp variants specific for a predetermined *FRT*-like sequence (Shultz et al. 2011). In another study, Flp variants were found to utilize two natural sites from the human interleukin 10 gene as recombination substrates (Bolusani et al. 2006).

In light of all these reports, evolving recombinase variants, able to recombine with a pre-existing genomic site, will provide great advances in gene transfer technology based on site-specific recombinases.

As for ΦC31 integrase, several improvements for enhancing the recombination efficiency and/or the site-selectivity of the recombinases through molecular and codon optimization have been described. As an illustration, to overcome the

thermo-instability of the FLP protein, Buchholz et al. 1998 have generated a thermo-stable derivative of FLP recombinase (FLPe), with a fourfold increase in recombination efficiency. A mouse codon-optimized FLP (FLPo) version has been synthetized resulting in recombination efficiency similar to that of Cre (Raymond and Soriano 2007).

In conclusion, Cre/LoxP and Flp/FRT recombination systems provide remarkably powerful tools for developping gene targeting technology. Some studies have highlighted the toxicity caused by Cre recombinase (for review, Turan et al. 2011). The molecular mechanism responsible for this problem is still unknown and so need to be addressed.

2.5.2 Site-Specific Serine Recombinase, ΦC31-Based Vectors

One of the most recently developed targeting strategy involves the ΦC31 integrase, a site-specific serine recombinase isolated from the *Streptomyces* phage ΦC31 and used in genome engineering for gene therapy (reviewed in Calos 2006; Brown et al. 2011). This enzyme catalyses the recombination of DNA fragment between two nonidentical attachment (*att*) sites in the bacterial (attB) and phage (attP) genome resulting in two new sites, attL and attR. These two attL and attR sites are not substrates for the integrase in the absence of other proteins, and thus ΦC31 catalyzes a unidirectional integration reaction (Kudriavtseva et al. 1994).

In gene transfer applications, the ΦC31 integrase mediates recombination between a plasmid containing an *attB* site and a locus in the genome containing an *attP* site. This *attP* can be naturally present or artificially preintegrated within the genome and it can be degenerated, and in this case it is called a *pseudo-attP* site (Fig. 2.1b, lower panel). The recombination in this application results in genomic integration of the attB-bearing plasmid with a higher site-specific recombination efficiency compared with that of other serine recombinases (Yamaguchi et al. 2011).

Several studies showed the potential of ΦC31 integrase-mediated transgene integration for gene-therapy (Table 2.3, and for review, Calos 2006). Moreover, several teams reported the *ex-vivo* correction of genetic human diseases in animal models of liver, skin diseases and Duchenne muscular dystrophy (Olivares et al. 2002; Ortiz-Urda et al. 2003; Quenneville et al. 2007). In mice, ΦC31-mediated integration of the gene encoding the human factor IX resulted in sustained expression of the therapeutic protein (Olivares et al. 2002). ΦC31 has been also used to generate transgenic frogs by co-microinjection of attB-containing vector and in vitro-transcribed ΦC31 integrase mRNA (Allen and weeks 2005).

The mouse, human, bovine, xenopus, zebrafish and rat genomes contain sequences with partial identity to *attP*, and these pseudo *attP* sequences have been used with significant efficiency as recombination targets (Ou et al. 2009; Lister 2010; Thyagarajan et al. 2001; Ma et al. 2006; Olivares et al. 2002; Allen and Weeks 2005). We know that a typical mammalian genome may contain 100–1,000 pseudo-attP sites. However, we also know that the enzyme preferentially mediates integration at a relatively small

Table 2.3 Features of main eucaryotic transposons used in gene transfer in vertebrates

Type	Host genome origin	Target site	New species	Applications	References
Synthetic	Salmonid fish (Ivics et al. 1997)	TA	Human, mouse, fish, frog, monkey, hamster, rabbit, dog, cow, sheep, bird, pig	Somatic and germinal transgenesis insertional mutagenesis, gene trap, gene therapy	Reviewed in Ivics and Izsvak (2010)
Synthetic	Frog (Miskey et al. 2003)	TA	Human, fish	Somatic transgenesis, gene trap	Miskey et al. (2003), Sano et al. (2009)
Natural	Fish (Clark et al. 2009)	TA	Human, monkey, pig, hamster, mouse, chicken, turkey	Somatic transgenesis	Clark et al. (2007, 2009)
Natural	Insect (Jacobson et al. 1986)	TA	Human, chicken, zebrafish	Somatic and germinal transgenesis	Reviewed in Plasterk et al. (1999)
Synthetic	Human (Miskey et al. 2007)	TA	Human, zebrafish	Somatic transgenesis	Miskey et al. (2007)
Natural	Insect (Cary et al. 1989)	TTAA	Human, mouse, pig, chicken, rabbit	Somatic and germinal transgenesis insertional mutagenesis, gene trap	Ding et al. (2005), Clark et al. (2007), Rad et al. (2010)
Natural	Fish (Koga et al. 2007)	8-bp target site	Human, mouse, zebrafish, frog, fish	Somatic and germinal transgenesis	Koga et al. (2007, 2008)
Natural	Fish (Kawakami et al. 1998)	8-bp target site	Human, chicken, mouse, frog, fish	Somatic and germinal transgenesis insertional mutagenesis, gene trap	Reviewed in Kawakami (2007)
Synthetic	Fish (Kapitonov and Jurka 2004)	15-bp consensus sequence	Human	somatic transgenesis	Sinzelle et al. (2008)

number of genomic preferred sites (hotspots), depending on the cell types and the chromatine accessibility of expressed regions (Thyagarajan et al. 2001; Chalberg et al. 2006; Maucksch et al. 2008). For example, hot-spots (called mouse pseudo-site 1 and 2, mpsL1 and mpsL2) have been identified in the mouse genome of fibroblast cells and of different tissues including liver and muscle (Thyagarajan et al. 2001; Olivares et al. 2002; Bertoni et al. 2006 ; Portlock et al. 2006; Aneja et al. 2007). Similarly, pseudo attP integration sites have been thoroughly characterized in human cells, with the establishment of consensus sequences generated from ΦC31 integration sequences. These sequences are somehow similar to the native ΦC31 attP site and share a common motif containing an inverted repeat (Chalberg et al. 2006; Thyagarajan et al. 2008; Nishiumi et al. 2009). In human embryonic kidney cells and in primary human keratinocytes, the hpsA site located on chromosome 8p22 has been characterized as the most prevalent genomic site for ΦC31-mediated integration of a attB-containing plasmid (Thyagarajan et al. 2001; Ortiz-Urda et al. 2002). In another study performed in three other somatic human cell lines, 7.5% of ΦC31 integrations isolated from individual clones were found at an unique site at chromosome 19q13.31 and 20% occurred in pseudo-sites of repetitive elements, HERV-L1 and L1 repeats (Chalberg et al. 2006). In human embryonic stem cells, some hotspots have also been identified which were different from those previously reported in other cell types (Thyagarajan et al. 2008).

To limit the accessibility of the enzyme to a large set of target sites and so restrict its integration profile, different strategies have attempted to modify the target site specificity of ΦC31. First, ΦC31 proteins were successfully mutated to obtain variants that catalyze a preferential integration at a chosen locus that deviate from the wild-type recognition sequences (Keravala et al. 2009). ΦC31 mutants were designed to display a higher preference for the previously identified hotspot 8qP22 site on human chromosome 8 (Sclimenti et al. 2001).

Several improvements of the ΦC31 recombination system have been reported. For example, a transgene placed in a reverse orientation against the attB element in combination with the use of insulators produces higher and more continuous expression of the transgene in cultured cells (Watanabe et al. 2011 ; Nishiumi et al. 2009). Integration frequency was enhanced in some ΦC31 mutants and in mouse codon optimized version (Liesner et al. 2010; Raymond and Soriano 2007; Sclimenti et al. 2001). Using mutagenesis, Keravala et al. 2009 constructed a ΦC31 mutant catalysing integration at twice the frequency of wild-type integrase and without alteration of the site-specificity.

However, the safety of ΦC31-mediated integration is doubtful and may compromise its use in gene therapy application. Recently, reports of a relatively high rate of chromosomal abnormalities (translocations, deletion, insertion and rearrangements) in the host sequence surrounding the integrated sites after ΦC31-mediated recombination were published. Several reports proposed that native ΦC31 pseudo att sites could recombine to each others (Ehrhardt et al. 2006; Liu et al. 2009). In human cells, 15% of the insertions sites contains an integrated transgene surrounded by DNA sequences originating from two different chromosomes (Ehrhardt et al. 2006).

ΦC31 integrase produces chromosomal rearrangements in primary human adult cells and in human embryonic fibroblasts leading thereafter to a DNA damage cellular response (Liu et al. 2006; Liu et al. 2009). Small deletions or insertions at the insertion sites were also observed (Thyagarajan et al. 2001; Chalberg et al. 2006; Ehrhardt et al. 2006; Nishiumi et al. 2009). However, the literature data is controversial since some reports minimize the potential negative effect of ΦC31 in producing chromosomal rearrangements (for a review, Brown et al. 2011). Indeed, the injection of ΦC31 mRNA into *Xenopus*, mice, or *Drosophila* embryos and the expression of the protein in chicken did not reveal a significant level of toxicity, suggesting that aberrant chromosomal events occur at low frequency i.e. less than 7.5% (Brown et al. 2011; Nishiumi et al. 2009).

In this context, we need to improve our understanding of the mechanisms responsible for chromosomal rearrangements and to study the biochemical properties of the enzyme before we can consider the ΦC31 recombination system as a safe and efficient integrating system for site-directed integration of transgenes.

2.6 Design of Random Integrating Vectors Based on DNA Transposons and Associated-Targeting Strategy

2.6.1 Diversity of DNA-Transposon Based Vectors

DNA transposons are mobile and repetitive genetic components of all eukaryotic genomes investigated so far. The majority of wild-type eukaryotic DNA transposons consist naturally of a gene encoding an enzyme, the transposase, which is surrounded by two inverted terminal repeats (ITRs) (Fig. 2.1c). In a transposition reaction, the transposase molecules bind to specific recognition sites within the ITRs, they excise the transposon from the donor locus and they integrate the transposon into a new chromosomal locus (acceptor locus). In a typical engineered plasmid-based transposon system, the transposase and the transposon are supplied independantly in a two-component system (Ivics and Izsvak 2010). The transposase source is provided *in trans* carried by a helper plasmid. The donor plasmid contains the gene-of-interest inserted between the two ITRs, thereby forming an expression cassette called a non-autonomous transposon. When delivered into the cells, the transposase will catalyse the precise transposition of the cassette containing the gene-of-interest from the donor plasmid to a genomic DNA locus (Fig. 2.1c).

Historically, *Sleeping Beauty* (*SB*) was the first DNA transposon shown to be active in vertebrates. SB was resurrected from an inactive fish *Tc1* transposons and it helded great promises in non-viral gene delivery technologies (Ivics et al. 1997). Following this pioneering work, a transposon toolkit was developed. This toolkit is currently expanding and used in a variety of applications, ranging from insertional

mutagenesis and transgenesis to gene therapy. Numerous studies reported the efficient transposition of these elements in various somatic tissues and germline belonging to a wide range of invertebrates or vertebrates species.

Specific features from this transposon toolkit were developed such as tropism, integration site preference, cargo capacity, overproduction inhibition etc… that might be more relevant in some applications (Ivics et al. 2009).

SB and *piggyBac* (*PB*), a naturally active DNA transposon isolated from insect genome by Cary et al. 1989, are currently the most thoroughly studied transposons because of their high transpositional activity (Ivics and Izsvak 2010). Importantly, *SB* has already been explored as a gene therapy vector in various preclinic animal models and its use in a human trial is now underway (Williams 2008; Aronovich et al. 2011).

While these elements are promising integrating vectors in gene transfer technology, a number of issues remain to be addressed. A first major problem is the relative weak integration efficiency of natural or awakened DNA transposons compared to virus-based gene transfer vectors. Considerable efforts are being made to enhance the transpositional activity by molecular engineering of the transposase. As an example, Mates et al. (2009) generated SB100, a 100-fold hyperactive version of SB transposase with integration efficiencies comparable to that of retroviral based-vectors. Similarly, a hyperactive version of PB transposase with more than tenfold higher rates of transposition than the wild-type codon-optimized PB transposase was recently engineered (Yusa et al. 2011).

The second major problem is that transposon integration is nonspecific and therefore leads to undesired insertional mutagenesis. Indeed, DNA transposons insert into a 2–8 bp specific target site duplicated upon integration. The *Harbinger3_DR* transposon system is an exception with a 15bp consensus palidromic target sequence (Kapitonov and Jurka 2004; Sinzelle et al. 2008). The insertion site is now well documented for *SB*, *piggyBac* and *Tol2* elements. A majority of *SB* integrations occurs in intergenic regions with a preference for palindromic AT repeats and the choice of the target site is influenced by the physical structure of the chromatine (reviewed in Ivics and Izsvak 2010). *Tol2* predominantly targets intergenic regions and more precisely repeat sequences whereas *piggyBac* inserts preferentially in genes and their upstream regulatory regions. Both transposons exhibit a significant preference for integrations close to CpG islands (Meir et al. 2011). Moreover, several hotspots for *piggyBac* and *Tol2* insertion were identified recently in human cells (Meir et al. 2011). Nevertheless, even if DNA transposons exhibit some preferences for given insertion sites, their integration can be considered fairly random at the genomic level.

2.6.2 Targeting Strategies of DNA Transposon-Based Systems

As mentioned above, DNA transposons hold a great potential in gene transfer technologies but lack specificity for genomic integration. To circumvent this limitation,

three experimental strategies based on the modification of the transposase/transposon DNA binding properties are currently tested (Voigt et al. 2008). The first strategy was suggested by Kaminski et al. 2002 and involves the construction of a chimeric transposase by fusing the transposase to a DNA-binding domain (DBD) that specifically recognizes a given genomic site. The second strategy involves the construction of a bifunctional targeting protein interacting on one hand with the transposome and on the other hand with the desired genomic site. Alternatively, in this second strategy, a targeting protein is constructed by fusing a protein interacting with the transposase and a DBD that specifically recognizes a chosen genomic site. These experimental strategies are detailed below.

2.6.2.1 Construction of Chimeric Transposases by Fusion to Zinc Finger DBDs

Taking advantage of these recent advances on engineered integrases for retroviral DNA targeting, several studies using fusion of PB, SB, Tol2 or Mos1 transposases and ZF DBDs such as E2C, Jazz, Gal4 have been reported (Table 2.4). Overall, these reports have demonstrated that the use of chimeric transposases can dramatically decrease transposition efficiencies (Feng et al. 2010; Wu et al. 2006; Yant et al. 2007; Ivics et al. 2007). However, some chimeric transposases based on the Gal4-UAS system or ZF E2C have proven to be effective tools for mediating site-directed transposition in plasmid to plasmid transposition assays (Table 2.4). Indeed, frequent targeted events were recovered at (or near) the expected locus for chimeric SB, Mos1 and piggyBac transposases in human and/or insect cells (Table 2.4). However, successfull *in vivo* genomic targeting experiment using such ZF chimeric transposases remains a challenge. Futhermore, some experimental target sites such as the UAS sequence do not exist in human DNA.

2.6.2.2 Construction of Bifunctional Targeting Proteins

The first proof-of-principle that manipulation of engineered transposon/transposase can be investigated for site-directed attempts was provided by Ivics et al. 2007. Two promising experimental strategies based on bifunctional targeting proteins were developed in their study (Table 2.4). The first strategy involves a fusion protein linking the SAF-box domain of the scaffold attachment factor A to the *Escherichia coli* LexA protein. The SAF-box binds the particular loop domains S/MARs in the genome. A single resulting LexA/NLS/SAF fusion protein was expected to interact with two components : the first one being a complex containing the transposase bound to its *SB* transposon DNA including a LexA operator and the second one being a MAR sequence on the genomic DNA. Therefore, this interaction should guide the transposome to the vicinity of MARs. The authors showed a bias in the insertion profile of *SB* transposon nearby MARs sequences, demonstrating that the transposon itself can be successfully targeted into physiological sites. Interestingly,

Table 2.4 Targeting strategies based on modified transposon systems

Chimeric transposases						
Transposase	DNA-binding domain	Cell -type	Transposition efficiency	Targeting efficiency	Experiment	References
Sleeping Beauty	Sp1	Human	C-terminal fusion, 85% of the unfused SB	Undetermined	Plasmid to chromosome	Wilson et al. (2005)
	ZNF202	Human	N-terminal fusion, similar to the unfused SB	Undetermined	Plasmid to chromosome	Wilson et al. (2005)
	Gal4	Human	N-terminal fusion, negligible	Undetermined	Plasmid to chromosome	Wu et al. (2006)
		Human	N-terminal fusion, 26% of the unfused SB	25% of total integration in a 443 bp targeting Window versus 2.3% for unfused SB	Plasmid to plasmid	Yant et al. (2007)
	Jazz	Human	N-terminal fusion, 15% of the unfused SB	No targeted transposition	Plasmid to chromosome	Ivics et al. (2007)
	E2C	Human	N-terminal fusion, negligible	Undetermined	Plasmid to chromosome	Ivics et al. (2007)
	E2C	Human	N-terminal fusion, 20% of the unfused SB	33.3% of total integration in a 443 bp targeting Window versus 2.3% for unfused SB	Plasmid to plasmid	Yant et al. (2007)
	TetR	Human	N-terminal fusion, negligible	Undetermined	Plasmid to chromosome	Ivics et al. (2007)
Mos1	Gal4	Insect	12.7-fold increase compared to the control Without target site	96% frequency at the same TA located 954 bp From the UAS sequence	Plasmid to plasmid	Maragathavally et al. (2006)

piggyBAc	Gal4	Human	N-terminal fusion, similar to the unfused PB	Undetermined	Plasmid to chromosome	Wu et al. (2006)
		Insect	4 and 7.5-fold increase depending On the cell type and compared to the control Without target site	71% frequence at two TTAA located 1 kb from The UAS sequence for one cell type and 36.5% Frequency at a single TTAA 1 kb from the UAS Sequence for the second cell type	Plasmid to plasmid	Wang et al. (2010)
		Insect	11.6-fold increase compared to the control Without target site	67% frequency at the same TTAA located 912 bp From the UAS sequence	Plasmid to plasmid	Maragathavally et al. (2006)
Tol2	Gal4	Human	N-terminal fusion, negligible	Undetermined	Plasmid to chromosome	Wu et al. (2006)
ISY100	Zif268 (mouse TF)	Bacteria	C-terminal fusion, 10–30% of the unfused ISY100	50% of insertions in a 20 bp region adjacent to the Zif268 binding site, up to 70% of these in a single TA	Plasmid to plasmid	Feng et al. (2010)

(continued)

Table 2.4 (continued)

Chimeric transposases						
Transposase	DNA-binding domain	Cell -type	Transposition efficiency	Targeting efficiency	Experiment	References
Engineered transposon						
Transposon	Targeting protein	Cell-type	Transposition efficiency	Targeting efficiency	Experiment	References
Sleeping Beauty	LexA/NLS/SAF	Human	Similar to the unmodified transposon system	9 insertions located within 1 kb distance from A predicted MAR with the targeting protein versus 2 insertions in the control with LexA/NLS protein	Plasmid to chromosome	Ivics et al. (2007)
	TetR/NLS/LexA	Human	Similar to the unmodified transposon system	Transposition into two TA sites in the promoter of The EGFP gene, 44 and 48 bp downstream of TRE	Plasmid to chromosome in TRE-EGFP transgenic cells	Ivics et al. (2007)
Engineered transposase						
Transposon	Targeting protein	Cell-type	Transposition efficiency	Targeting efficiency	Experiment	References
Sleeping Beauty	TetR/NLS/N-57	Human	Similar to the unmodified transposon system	10% of insertions in the vicinity of TRE locus	plasmid to chromosome in TRE-EGFP transgenic cells	Ivics et al. (2007)

previous experiments showed that expression of a reporter gene artificially integrated into these regions was independent of chromosomal position (Stief et al. 1989). Following a similar strategy, Ivics et al. 2007 used a targeting fusion protein linking the tetracycline repressor (TetR) to LexA protein in human transgenic cells containing tetracycline response element (TRE) and they obtained targeted transposition events.

The second strategy is based on protein-protein interaction and uses a targeting protein containing the N-terminal protein interaction domain of SB transposase (N-57) and TetR. The TetR/NLS/N-57 fusion protein was expected to interact wild-type SB transposase and tether the complex to a TRE chromosomal target which was previously incorporated in human cells. This construct was indeed able to target efficiently SB transposon since 10% of the insertions were detected within a 2.6-kb window in the vicinity of the TRE locus.

To conclude, these two strategies led to successful integrations into a target chromosomal locus in a genomic context without altering the transposase activity. Therefore, they represent alternative strategies to chimeric transposases and clearly warrant further studies.

2.7 Conclusion

Gene therapy and animal transgenesis applications have been driving forces for the development of nonviral site-directed integrating vectors. The field is now rich of several modular technologies that enable elaborated genetic engineering. The recent surge of synthetic biology is likely to contribute to the improvement of these technologies. It is clear, however, that the most prominent limitation is related to the safety of using, distributing, delivering and implementing these approaches in living beings.

References

Allen BG, Weeks DL (2005) Transgenic *Xenopus laevis* embryos can be generated using phiC31 integrase. Nat Methods 2:975–979

Aneja MK, Imker R, Rudolph C (2007) Phage phiC31 integrase-mediated genomic integration and long-term gene expression in the lung after nonviral gene delivery. J Gene Med 9:967–975

Arnould S, Chames P, Perez C, Lacroix E, Duclert A, Epinat JC, Stricher F, Petit AS, Patin A, Guillier S, Rolland S, Prieto J, Blanco FJ, Bravo J, Montoya G, Serrano L, Duchateau P, Paques F (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. J Mol Biol 355:443–458

Arnould S, Perez C, Cabaniols JP, Smith J, Gouble A, Grizot S, Epinat JC, Duclert A, Duchateau P, Paques F (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. J Mol Biol 371:49–65

Arnould S, Delenda C, Grizot S, Desseaux C, Paques F, Silva GH, Smith J (2011) The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. Protein Eng Des Sel 24:27–31

Aronovich EL, McIvor RS, Hackett PB (2011) The Sleeping Beauty transposon system: a non-viral vector for gene therapy. Hum Mol Genet 20:R14–R20

Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat RJ Jr, Stoddard BL, Baker D (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. Nature 441:656–659

Bertoni C, Jarrahian S, Wheeler TM, Li Y, Olivares EC, Calos MP, Rando TA (2006) Enhancement of plasmid-mediated gene therapy for muscular dystrophy by directed plasmid integration. Proc Natl Acad Sci USA 103:419–424

Birling MC, Gofflot F, Warot X (2009) Site-specific recombinases for manipulation of the mouse genome. Methods Mol Biol 561:245–263

Boch J (2011) TALEs of genome targeting. Nat Biotechnol 29:135–136

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512

Bolusani S, Ma CH, Paek A, Konieczka JH, Jayaram M, Voziyanov Y (2006) Evolution of variants of yeast site-specific recombinase Flp that utilize native genomic sequences as recombination target sites. Nucleic Acids Res 34:5259–5269

Bonas U, Stall R, Staskawic B (1989) Genetic and structural characterization of the avirulence gene avrBs3 from *Xanthomonas campestris* pv. vesicatoria. Mol Gen Genet 218:127–136

Brown WR, Lee NC, Xu Z, Smith MC (2011) Serine recombinases as tools for genome engineering. Methods 53:372–379

Buchholz F, Stewart AF (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. Nat Biotechnol 19:1047–1052

Buchholz F, Angrand PO, Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. Nat Biotechnol 16:657–662

Cabaniols JP, Ouvry C, Lamamy V, Fery I, Craplet ML, Moulharat N, Guenin SP, Bedut S, Nosjean O, Ferry G, Devavry S, Jacqmarcq C, Lebuhotel C, Mathis L, Delenda C, Boutin JA, Duchateau P, Coge F, Paques F (2010) Meganuclease-driven targeted integration in CHO-K1 cells for the fast generation of HTS-compatible cell-based assays. J Biomol Screen 15:956–967

Calos MP (2006) The phiC31 integrase system for gene therapy. Curr Gene Ther 6:633–645

Carninci P, Yasuda J, Hayashizaki Y (2008) Multifaceted mammalian transcriptome. Curr Opin Cell Biol 20:274–280

Cary LC, Goebel M, Corsaro BG, Wang HG, Rosen E, Fraser MJ (1989) Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. Virology 172:156–169

Casco-Robles MM, Yamada S, Miura T, Chiba C (2010) Simple and efficient transgenesis with I-SceI meganuclease in the newt, Cynops pyrrhogaster. Dev Dyn 239:3275–3284

Chalberg TW, Genise HL, Vollrath D, Calos MP (2005) phiC31 integrase confers genomic integration and long-term transgene expression in rat retina. Invest Ophthalmol Vis Sci 46:2140–2146

Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT, Hoelters J, Calos MP (2006) Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol 357:28–48

Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, Stoddard BL (2002) Design, activity, and structure of a highly specific artificial endonuclease. Mol Cell 10:895–905

Choulika A, Perrin A, Dujon B, Nicolas JF (1995) Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. Mol Cell Biol 15:1968–1973

Clark KJ, Carlson DF, Foster LK, Kong BW, Foster DN, Fahrenkrug SC (2007) Enzymatic engineering of the porcine genome with transposons and recombinases. BMC Biotechnol 7:42

Cohen-Tannoudji M, Robine S, Choulika A, Pinto D, EL Marjou F, Babinet C, Louvard D, Jaisser F (1998) I-SceI-induced gene replacement at a natural locus in embryonic stem cells. Mol Cell Biol 18:1444–1448

Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 122:473–483

Dutheil N, Shi F, Dupressoir T, Linden RM (2000) Adeno-associated virus site-specifically integrates into a muscle-specific DNA region. Proc Natl Acad Sci USA 97:4862–4866

Ehrhardt A, Engler JA, Xu H, Cherry AM, Kay MA (2006) Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. Hum Gene Ther 17:1077–1094

ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. Science 306:636–640

Epinat JC, Arnould S, Chames P, Rochaix P, Desfontaines D, Puzin C, Patin A, Zanghellini A, Paques F, Lacroix E (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. Nucleic Acids Res 31:2952–2962

Feng YQ, Seibler J, Alami R, Eisen A, Westerman KA, Leboulch P, Fiering S, Bouhassira EE (1999) Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. J Mol Biol 292:779–785

Feng X, Bednarz AL, Colloms SD (2010) Precise targeted integration by a chimaeric transposase zinc-finger fusion protein. Nucleic Acids Res 38:1204–1216

Fu Q, Jia S, Sun Z, Tian F, Du J, Zhou Y, Wang Y, Wang X, Zhan L (2009) PhiC31 integrase and liver-specific regulatory elements confer high-level, long-term expression of firefly luciferase in mouse liver. Biotechnol Lett 31:1151–1157

Gouble A, Smith J, Bruneau S, Perez C, Guyot V, Cabaniols JP, Leduc S, Fiette L, Ave P, Micheau B, Duchateau P, Paques F (2006) Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. J Gene Med 8:616–622

Grabher C, Wittbrodt J (2008) Recent advances in meganuclease-and transposon-mediated transgenesis of medaka and zebrafish. Methods Mol Biol 461:521–539

Grishin A, Fonfara I, Alexeevski A, Spirin S, Zanegina O, Karyagina A, Alexeyevsky D, Wende W (2010) Identification of conserved features of LAGLIDADG homing endonucleases. J Bioinform Comput Biol 8:453–469

Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, Villate M, Thomas S, Lemaire L, Montoya G, Blanco FJ, Paques F, Duchateau P (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res, 37:5405–5419

Grizot S, Epinat JC, Thomas S, Duclert A, Rolland S, Paques F, Duchateau P (2010) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. Nucleic Acids Res 38:2006–2018

Grosse S, Huot N, Mahiet C, Arnould S, Barradeau S, Clerre DL, Chion-Sotinel I, Jacqmarcq C, Chapellier B, Ergani A, Desseaux C, Cedrone F, Conseiller E, Paques F, Labetoulle M, Smith J (2011) Meganuclease-mediated Inhibition of HSV1 Infection in Cultured Cells. Mol Ther 19:694–702

Hacein-Bey-abina S, le Deist F, Carlier F, Bouneaud C, Hue C, de Villartay JP, Thrasher AJ, Wulffraat N, Sorensen R, Dupuis-Girod S, Fischer A, Davies EG, Kuis W, Leiva L, Cavazzana-Calvo M (2002) Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. N Engl J Med 346:1185–1193

Handel EM, Cathomen T (2011) Zinc-finger nuclease based genome surgery: it"s all about specificity. Curr Gene Ther 11:28–37

Held PK, Olivares EC, Aguilar CP, Finegold M, Calos MP, Grompe M (2005) In vivo correction of murine hereditary tyrosinemia type I by phiC31 integrase-mediated gene delivery. Mol Ther 11:399–408

Ito M, Yamanouchi K, Naito K, Calos MP, Tojo H (2011) Site-specific integration of transgene targeting an endogenous lox-like site in early mouse embryos. J Appl Genet 52:89–94

Ivics Z, Izsvak Z (2010) The expanding universe of transposon technologies for gene and cell engineering. Mob DNA 1:25

Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) Molecular reconstruction of sleeping beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91:501–510

Ivics Z, Katzer A, Stuwe EE, Fiedler D, Knespel S, Izsvak Z (2007) Targeted sleeping beauty transposition in human cells. Mol Ther 15:1137–1144

Ivics Z, Li MA, Mates L, Boeke JD, Nagy A, Bradley A, Izsvak Z (2009) Transposon-mediated genome manipulation in vertebrates. Nat Methods 6:415–422

Jacobson JW, Medhora MM, Hartl DL (1986) Molecular structure of a somatically unstable transposable element in Drosophila. Proc Natl Acad Sci U S A 83:8684–8688

Joshi R, Ho KK, Tenney K, Chen JH, Golden BL, Gimble FS (2011) Evolution of I-SceI homing endonucleases with increased DNA recognition site specificity. J Mol Biol 405:185–200

Kaminski JM, Huber MR, Summers JB, Ward MB (2002) Design of a nonviral vector for site-selective, efficient integration into the human genome. FASEB J 16:1242–1247

Kapitonov VV, Jurka J (2004) Harbinger transposons and an ancient HARBI1 gene derived from a transposase. DNA Cell Biol 23:311–324

Kawakami K, Koga A, Hori H, Shima A (1998) Excision of the tol2 transposable element of the medaka fish, Oryzias latipes, in zebrafish, Danio rerio. Gene 225:17–22

Kawakami K (2007) Tol2: a versatile gene transfer vector in vertebrates. Genome Biol 8(Suppl 1)**:**S7

Keravala A, Portlock JL, Nash JA, Vitrant DG, Robbins PD, Calos MP (2006) PhiC31 integrase mediates integration in cultured synovial cells and enhances gene expression in rabbit joints. J Gene Med 8:1008–1017

Keravala A, Lee S, Thyagarajan B, Olivares EC, Gabrovsky VE, Woodard LE, Calos MP (2009) Mutational derivatives of PhiC31 integrase with increased efficiency and specificity. Mol Ther 17:112–120

Koga A, Shimada A, Kuroki T, Hori H, Kusumi J, Kyono-Hamaguchi Y, Hamaguchi S (2007) The Tol1 transposable element of the medaka fish moves in human and mouse cells. J Hum Genet 52:628–635

Koga A, Cheah FS, Hamaguchi S, Yeo GH, Chong SS (2008) Germline transgenesis of zebrafish using the medaka Tol1 transposon system. Dev Dyn 237:2466–2474

Koga A, Shimada A, Kuroki T, Hori H, Kusumi J, Kyono-Hamaguchi Y, Hamaguchi S (2007) The Tol1 transposable element of the medaka fish moves in human and mouse cells. J Hum Genet 52:628–635

Koga A, Cheah FS, Hamaguchi S, Yeo GH, Chong SS (2008) Germline transgenesis of zebrafish using the medaka Tol1 transposon system. Dev Dyn 237:2466–2474

Kudriavtseva EA, Yun Yang L, Voeikova TA, Lomovskaia ND (1994) Use of a plasmid with integrative function of phage phiC31 for transfer of cloned genes into Streptomyces strains. Genetika 30:886–897

Lauth M, Moerl K, Barski JJ, Meyer M (2000) Characterization of Cre-mediated cassette exchange after plasmid microinjection in fertilized mouse oocytes. Genesis 27:153–158

Liesner R, Zhang W, Noske N, Ehrhardt A (2010) Critical amino acid residues within the phiC31 integrase DNA-binding domain affect recombination activities in mammalian cells. Hum Gene Ther 21:1104–1118

Lim JK, Glass WG, McDermott DH, Murphy PM (2006) CCR5: no longer a ““good for nothing”“ gene–chemokine control of West Nile virus infection. Trends Immunol 27:308–312

Lister JA (2010) Transgene excision in zebrafish using the phiC31 integrase. Genesis 48:137–143

Liu X, Liu M, Xue Z, Pan Q, Wu L, Long Z, Xia K, Liang D, Xia J (2007) Non-viral ex vivo transduction of human hepatocyte cells to express factor VIII using a human ribosomal DNA-targeting vector. J Thromb Haemost 5:347–351

Liu J, Skjorringe T, Gjetting T, Jensen TG (2009) PhiC31 integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. BMC Biotechnol 9:31

Ma QW, Sheng HQ, Yan JB, Cheng S, Huang Y, Chen-Tsai Y, Ren ZR, Huang SZ, Zeng YT (2006) Identification of pseudo attP sites for phage phiC31 integrase in bovine genome. Biochem Biophys Res Commun 345:984–988

Maragathavally KJ, Kaminski JM, Coates CJ (2006) Chimeric Mos1 and piggyBac transposases result in site-directed integration. FASEB J 20:1880–1882

Maucksch C, Aneja MK, Hennen E, Bohla A, Hoffmann F, Elfinger M, Rosenecker J, Rudolph C (2008) Cell type differences in activity of the Streptomyces bacteriophage phiC31 integrase. Nucleic Acids Res 36:5462–5471

Meir YJ, Weirauch MT, Yang HS, Chung PC, Yu RK, Wu SC (2011) Genome-wide target profiling of piggyBac and Tol2 in HEK 293: pros and cons for gene discovery and gene therapy. BMC Biotechnol 11:28

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29:143–148

Miskey C, Izsvak Z, Plasterk RH, Ivics Z (2003) The frog prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. Nucleic Acids Res 31:6873–6881

Miskey C, Papp B, Mates L, Sinzelle L, Keller H, Izsvak Z, Ivics Z (2007) The ancient mariner sails again: transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. Mol Cell Biol 27:4589–4600

Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501

Munoz IG, Prieto J, Subramanian S, Coloma J, Redondo P, Villate M, Merino N, Marenchino M, D'Abramo M, Gervasio FL, Grizot S, Daboussi F, Smith J, Chion-Sotinel I, Paques F, Duchateau P, Alibes A, Stricher F, Serrano L, Blanco FJ, Montoya G (2011) Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. Nucleic Acids Res 39:729–743

Nishiumi F, Sone T, Kishine H, Thyagarajan B, Kogure T, Miyawaki A, Chesnut JD, Imamoto F (2009) Simultaneous single cell stable expression of 2–4 cDNAs in HeLaS3 using psiC31 integrase system. Cell Struct Funct 34:47–59

Ogino H, McConnell WB, Grainger RM (2006) High-throughput transgenesis in Xenopus using I-SceI meganuclease. Nat Protoc 1:1703–1710

Olivares EC, Hollis RP, Chalberg TW, Meuse L, Kay MA, Calos MP (2002) Site-specific genomic integration produces therapeutic Factor IX levels in mice. Nat Biotechnol 20:1124–1128

Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q, Fang M, Calos MP, Khavari PA (2002) Stable nonviral genetic correction of inherited human skin disease. Nat Med 8:1166–1170

Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q, Calos MP, Khavari PA (2003) PhiC31 integrase-mediated nonviral genetic correction of junctional epidermolysis bullosa. Hum Gene Ther 14:923–928

Ou HL, Huang Y, Qu LJ, Xu M, Yan JB, Ren ZR, Huang SZ, Zeng YT (2009) A phiC31 integrase-mediated integration hotspot in favor of transgene expression exists in the bovine genome. FEBS J 276:155–163

Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L (2005) Evolution and functional classification of vertebrate gene deserts. Genome Res 15:137–145

Papapetrou EP, Lee G, Malani N, Setty M, Riviere I, Tirunagari LM, Kadota K, Roth SL, Giardina P, Viale A, Leslie C, Bushman FD, Studer L, Sadelain M (2011) Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. Nat Biotechnol 29:73–78

Pipiras E, Coquelle A, Bieth A, Debatisse M (1998) Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. EMBO J 17:325–333

Plasterk RH, Izsvak Z, Ivics Z (1999) Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet 15:326–332

Portlock JL, Keravala A, Bertoni C, Lee S, Rando TA, Calos MP (2006) Long-term increase in mVEGF164 in mouse hindlimb muscle mediated by phage phiC31 integrase after nonviral DNA delivery. Hum Gene Ther 17:871–876

Puttini S, Ouvrard-Pascaud A, Palais G, Beggah AT, Gascard P, Cohen-Tannoudji M, Babinet C, Blot-Chabaud M, Jaisser F (2005) Development of a targeted transgenesis strategy in highly differentiated cells: a powerful tool for functional genomic analysis. J Biotechnol 116:145–151

Quenneville SP, Chapdelaine P, Rousseau J, Tremblay JP (2007) Dystrophin expression in host muscle following transplantation of muscle precursor cells modified with the phiC31 integrase. Gene Ther 14:514–522

Rankin SA, Hasebe T, Zorn AM, Buchholz DR (2009) Improved cre reporter transgenic Xenopus. Dev Dyn 238:2401–2408

Raymond CS, Soriano P (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. PLoS One 2:e162

Remy S, Tesson L, Menoret S, Usal C, Scharenberg AM, Anegon I (2010) Zinc-finger nucleases: a powerful tool for genetic engineering of animals. Transgenic Res 19:363–371

Santoro SW, Schultz PG (2003) Directed evolution of the substrate specificities of a site-specific recombinase and an aminoacyl-tRNA synthetase using fluorescence-activated cell sorting (FACS). Methods Mol Biol 230:291–312

Sarkar I, Hauber I, Hauber J, Buchholz F (2007) HIV-1 proviral DNA excision using an evolved recombinase. Science 316:1912–1915

Sclimenti CR, Thyagarajan B, Calos MP (2001) Directed evolution of a recombinase for improved genomic integration at a native human sequence. Nucleic Acids Res 29:5044–5051

Shultz JL, Voziyanova E, Konieczka JH, Voziyanov Y (2011) A genome-wide analysis of FRT-like sequences in the human genome. PLoS One 6:e18077

Silva G, Poirot L, Galetto R, Smith J, Montoya G, Duchateau P, Paques F (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. Curr Gene Ther 11:11–27

Sinzelle L, Kapitonov VV, Grzela DP, Jursch T, Jurka J, Izsvak Z, Ivics Z (2008) Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes. Proc Natl Acad Sci USA 105:4715–4720

Sobkow L, Epperlein HH, Herklotz S, Straube WL, Tanaka EM (2006) A germline GFP transgenic axolotl and its use to track cell fate: dual origin of the fin mesenchyme during development and the fate of blood cells during regeneration. Dev Biol 290:386–397

Sorrell DA, Kolb AF (2005) Targeted modification of mammalian genomes. Biotechnol Adv 23:431–469

Sorrell DA, Robinson CJ, Smith JA, Kolb AF (2010) Recombinase mediated cassette exchange into genomic targets using an adenovirus vector. Nucleic Acids Res 38:e123

Stief A, Winter DM, Stratling WH, Sippel AE (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature 341:343–345

Surendranath V, Chusainow J, Hauber J, Buchholz F, Habermann BH (2010) SeLOX–a locus of recombination site search tool for the detection and directed evolution of site-specific recombination systems. Nucleic Acids Res 38:W293–W298

Tan W, Zhu K, Segal DJ, Barbas CF III, Chow SA (2004) Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. J Virol 78:1301–1313

Thaci B, Ulasov IV, Wainwright DA, Lesniak MS (2011) The challenge for gene therapy: innate immune response to adenoviruses. Oncotarget 2:113–121

Thyagarajan B, Guimaraes MJ, Groth AC, Calos MP (2000) Mammalian genomes contain active recombinase recognition sites. Gene 244:47–54

Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol Cell Biol 21:3926–3934

Thyagarajan B, Liu Y, Shin S, Lakshmipathy U, Scheyhing K, Xue H, Ellerstrom C, Strehl R, Hyllner J, Rao MS, Chesnut JD (2008) Creation of engineered human embryonic stem cell lines using phiC31 integrase. Stem Cells 26:119–126

Turan S, Galla M, Ernst E, Qiao J, Voelkel C, Schiedlmeier B, Zehe C, Bode J (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. J Mol Biol 407:193–221

Voigt K, Izsvak Z, Ivics Z (2008) Targeted gene insertion for molecular medicine. J Mol Med (Berl) 86:1205–1219

Wang N, Jiang CY, Jiang MX, Zhang CX, Cheng JA (2010) Using chimeric piggyBac transposase to achieve directed interplasmid transposition in silkworm Bombyx mori and fruit fly Drosophila cells. J Zhejiang Univ Sci B 11:728–734

Watanabe S, Nakamura S, Sakurai T, Akasaka K, Sato M (2011) Improvement of a phiC31 integrase-based gene delivery system that confers high and continuous transgene expression. N Biotechnol 28:312–319

Williams DA (2008) Sleeping beauty vector system moves toward human trials in the United States. Mol Ther 16:1515–1516

Wilson MH, Kaminski JM, George AL Jr (2005) Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. FEBS Lett 579:6205–6209

Wirth D, Gama-Norton L, Riemer P, Sandhu U, Schucht R, Hauser H (2007) Road to precision: recombinase-based targeting technologies for genome engineering. Curr Opin Biotechnol 18:411–419

Wong AC, Draper BW, van Eenennaam AL (2010) FLPe functions in zebrafish embryos. Transgenic Res 20:409–415

Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. Proc Natl Acad Sci USA 103:15008–15013

Wu J, Kandavelou K, Chandrasegaran S (2007) Custom-designed zinc finger nucleases: what is next? Cell Mol Life Sci 64:2933–2944

Yamaguchi S, Kazuki Y, Nakayama Y, Nanba E, Oshimura M, Ohbayashi T (2011) A method for producing transgenic cells using a multi-integrase system on a human artificial chromosome vector. PLoS One 6:e17267

Yant SR, Huang Y, Akache B, Kay MA (2007) Site-directed transposon integration in human cells. Nucleic Acids Res 35:e50

Yi Y, Noh MJ, Lee KH (2011) Current advances in retroviral gene therapy. Curr Gene Ther 11:218–228

Yusa K, Zhou L, Li MA, Bradley A, Craig NL (2011) A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci USA 108:1531–1536

Part II
Integration Based on Homologous Recombination

Chapter 3
Gene Targeting and Homologous Recombination in *Saccharomyces cerevisiae*

Judith Miné-Hattab and Rodney Rothstein

Abstract Yeast cells are particularly efficient at homologous recombination and therefore many initial successes with gene targeting occurred early on in this model system. Here we describe some of the history leading up to those developments along with several recent advances in gene replacement and targeting technology. In the last half of this chapter, we describe many of the steps that occur during homologous recombination in mitotic yeast cells. It is noteworthy that yeast still serves as an excellent system to test integration of exogenous DNA molecules.

Keywords Gene targeting • "Ends-in" integration • One-step gene disruption • Allele replacement • Yeast • Homologous recombination

3.1 Introduction

3.1.1 History

Many of the fundamental studies defining the mechanism of homologous recombination were performed in the yeast *Saccharomyces cerevisiae*. These studies were in turn based on the early studies of genetic transformation in bacteria. Transformation is defined as the introduction of exogenous genetic material into a cell. The principle was first demonstrated by Griffith, who, in 1928, showed that non-virulent Pneumococcus could be transformed to virulence by co-inoculation with a heat-killed extract of a virulent strain (Griffith 1928). Nearly 20 years later,

J. Miné-Hattab • R. Rothstein (✉)
Department of Genetics and Development, Columbia University, New York 10032, USA
e-mail: rothstein@columbia.edu

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_3,

Avery, MacLeod, and McCarty demonstrated that the "active principle" in the pneumococcal transformation to virulence is DNA (Avery et al. 1944). In the 1970s, the development of bacterial plasmid technology enabled researchers to clone and isolate purified yeast genes on bacterial plasmids. Introduction of such cloned DNA into yeast proceeded along two parallel routes. First, natural yeast circular 2 μm plasmids were isolated and developed as cloning vectors so that cloned DNA could be maintained extrachromosomally (Beggs 1978). At the same time, Hinnen, Hicks and Fink reported that a yeast *LEU2* gene cloned on a bacterial plasmid integrated via homology into the genome upon transformation into yeast (Hinnen et al. 1978). The majority of the integration events occur at the chromosomal *leu2* locus and the plasmid sequence is maintained stably through many generations. Occasional integration events were detected via homologous repetitive sequences present near the *LEU2* locus.

Until the early 1980s, integration of DNA into the yeast genome was achieved by transforming circular plasmids. However, the efficiency of transformation was poor and the integration of a plasmid sequence with more than one region of genome homology occurred at multiple sites. In 1981, two fundamental ideas changed the field of gene targeting. First, it was shown that a double-strand break (DSB) within a yeast gene sequence on a plasmid dramatically increases the efficiency of transformation (Orr-Weaver et al. 1981). At the same time, if the plasmid contained two yeast sequences, integration was almost always at the genomic site homologous to the region containing the DSB and not to the genomic site of the uncut sequence. This gene targeting technique, sometimes called "ends-in" integration (Thaler and Stahl 1988), is still used for introducing DNA into the genome. Since the integration results in a direct repeat, recombination between the repeats results in loss of the plasmid and one copy of the repeat. Scherer and Davis took advantage of these observations to develop a pop-in/pop-out technique to swap one DNA sequence in the genome for another without the plasmid sequences (Scherer and Davis 1979). Shortle, Haber and Botstein showed that insertion of a gene fragment truncated at its 5′ and 3′ends resulted in a gene disruption (Shortle et al. 1982). However, due to the reverse reaction of the integration reaction, these disruptions were unstable. In 1983, following these advances, Rothstein showed that stable gene disruptions could be made using linearized fragments that contain regions of the gene of interest flanking a selectable marker (Rothstein 1983). This integration procedure, in which the homology results in "ends-out" (Thaler and Stahl 1988) gives rise to a "one-step gene replacement" and has been used extensively to disrupt gene sequences in the yeast genome.

3.1.2 Outline

In section 3.2, we describe the main techniques developed to integrate a DNA sequence into the yeast genome: "ends-in", "pop-in/pop-out", "ends-out", "hybrid methods" and plasmids construction by gap repair. In section 3.3, we describe a general model of homologous recombination in yeast and provide a step-by-step description of homologous recombination mechanisms at the molecular level.

3.2 Transformation with Linearized Plasmids: Targeted Integration

3.2.1 Ends-in Integration

3.2.1.1 General Mechanism

"Ends-in" integration refers to the integration of a cut plasmid into a yeast chromosome as illustrated in Fig. 3.1. The plasmid must contain a sequence homologous to a yeast gene and is then linearized by a restriction enzyme that cuts inside this sequence. Because DNA ends are highly recombinogenic and strongly interact with the homologous sequences in the genome, the efficiency of the integration is increased 10–1,000-fold compared to an uncut plasmid (Orr-Weaver et al. 1981). During an "ends-in" integration, a single plasmid molecule can be integrated as illustrated in Fig. 3.1b. Occasionally multiple tandem integration events occur. When two restriction enzyme cuts are made within the yeast sequence of the plasmid, an internal segment of DNA is removed generating a gapped-linear DNA molecule. Interestingly, it was shown that the absence of an internal DNA segment does not affect the ability of gapped-linear molecules to transform at high efficiency (Orr-Weaver et al. 1981). Moreover, the gap is always filled during the process of

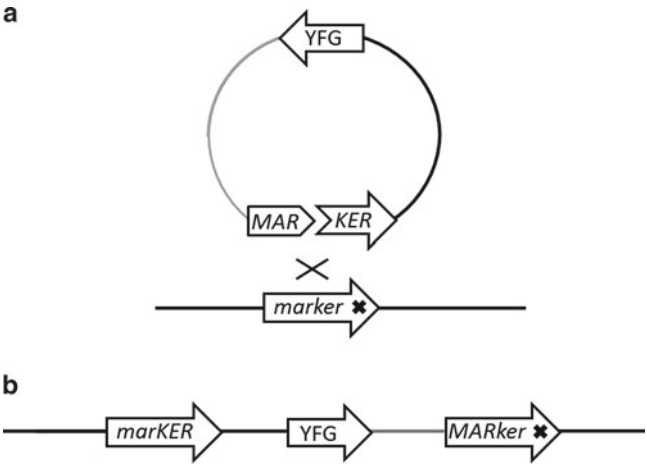


Fig. 3.1 "Ends-in" integration. (**a**): A plasmid harboring your favorite gene (*YFG)* to integrate and a wild type sequence of a yeast gene used as a marker is engineered. The plasmid is cut in the middle of the marker and transformed into a yeast strain auxotrophic for this marker (indicated as *marker* x). The "x" in the mutated marker represents the mutation. The broken ends of the plasmid are highly recombinogenic and homologous recombination occurs between the wild type copy of the marker (*MARKER*) on the plasmid and the mutated copy of the marker in the genome. (**b**): After integration, the chromosome contains a wild type copy of the marker (*marKER*), your favorite gene (*YFG*) and a mutated copy of the marker (*MARker* x). The *small* and *large letters* in "marker" indicate the position of the crossover

the integration using the chromosomal homologous sequence as a template. Integration of linear and gapped-linear molecules, but not of circular molecules, is blocked by the *rad52-1* mutation (Orr-Weaver et al. 1981) indicating that "ends-in" integrations occur through the homologous recombination pathway.

"Ends-in" integration can be performed in any recombination proficient *Saccharomyces* strain using suitable genetic markers. The integrating plasmid must contain a dominant selectable marker. In many cases the plasmid-borne marker is a wild-type allele that complements an auxotrophic mutation in the targeted strain (see below for examples). In such cases, the marker may also provide the homologous chromosomal site for plasmid integration. The genetic markers for selection of transformation events in the host strain should exhibit low reversion frequencies (less than 10^{-8}). In yeast, the most commonly used markers are *ura3, leu2, his3, trp1* and *ade2*. However, any marker can be used if a genetic selection exists. Recently, antibiotic resistance genes from bacteria and fungi have been shown to be a very powerful tool and are commonly used as markers (van den Berg and Steensma 1997). For example, the kanamycin selective marker (kanMX) confers resistance to G418. In addition, the hph gene confers dominant drug resistance to hygromycin B (Goldstein and McCusker 1999). In these cases, the dominant selectable marker is not homologous to the yeast genome, and therefore does not provide an integration site.

3.2.1.2 Optimizing "Ends-in" Integration

Different parameters can influence the efficiency of integration into the genome. First, the natural variability of different genetic regions for recombination affects integration efficiency. For example, integrations at the *HIS3* locus are 100-fold more frequent than integrations at the *LEU2* locus (Rothstein 1991). Unfortunately, there is no rule to predict which marker is optimized and each case has to be tested independently. Second, the length of homology influences integration frequency. Again, no systematic rule has been established and thus far only qualitative rules have been made from scattered experiments. In general, longer regions of homology result in increased integration frequency. For example, cut sequences with 125 nucleotides of homology (37 nucleotides on one side and 88 on the other) did not integrate efficiently, whereas targeted integration was obtained with a sequence containing 250 nucleotides of homology (40 and 210) (Rothstein 1991).

3.2.1.3 Tagging a Chromosomal Locus in Vivo by "Ends-in" Integration

"Ends-in" integration has been used for many research applications in yeast. As an example, here we describe how to fluorescently tag a specific locus on a chromosome by inserting a repetitive *Lac* or *Tet* operator array (Belmont 2001; Meister et al. 2010; Zimmer and Fabre 2011). Over the past few years, several laboratories have investigated the organization of genes inside the nucleus and

have found that nuclear position profoundly influences gene expression. To perform these studies in living cells, a method for attaching a fluorescent marker to chromosomal sites was developed using bacterial operator sequences from the *Tet* and *Lac* operons integrated at specific yeast chromosomal loci (Straight et al. 1996). To increase the signal coming from the chromosomal loci, tandem arrays of 100–300 copies of operator sequences were integrated. To visualize these loci, the *Tet* and *Lac* repressor proteins were fused to fluorescent protein tags and expressed in the cells. Using this method, it is possible to follow the motion of chromosomal loci inside both living yeast and human cells (Belmont 2001; Marshall and Sedat 1999; Meister et al. 2010).

Tagging chromosomal loci is a two-step process. The first step consists of transforming fluorescently-tagged proteins LacI or TetR into the strain, using either a circular plasmid or by integrating the plasmid into the genome (Belmont 2001; Lisby et al. 2001). Integrative plasmids are preferred because they give more reproducible levels of fluorescently tagged proteins. The second step consists of integrating the *Lac* or *Tet* operator binding site array, which is recognized by the fluorescently-tagged protein (the *Lac* or *Tet* repressor, respectively). Since the *Lac* and *Tet* operator are repetitive (100–300 copies), they are maintained in a recombination-deficient bacterial strain (e.g., SURE strain). Once extracted, the plasmid is cut in the middle of the yeast marker sequence and integrated by ends-in integration into the yeast strain (Fig. 3.2a, b). Alternatively, any genomic sequence can be introduced into this plasmid to target the integration of the repetitive array to its homologous genomic site (Meister et al. 2010). Figure 3.2c shows a yeast cell harboring a TetO array (3x112 copies) at the *URA3* locus with the TetR-RFP protein inserted elsewhere in the genome. Using wide-field microscopy, the tagged *URA3* locus is visualized as a bright spot. In this case, the Rad52 protein is tagged with CFP, which is visualized as a gray circle delineating the nucleus.

Although the "ends-in" recombination technique is used extensively in yeast genome manipulation, there are some drawbacks to such integrations. First, at some low, but measurable frequency $\approx 10^{-5}$, there is the risk of losing the integrated sequence by a "pop-out" due the direct repeat generated by the integration event. The second drawback to this technique is that plasmid backbone sequences also integrate into the chromosome, which may conceivably disrupt local chromatin structure or *cis* regulatory sequences near the integration site. The final drawback is that once a marker locus is used for the selection, it cannot be re-used for another genome manipulation.

3.2.2 Pop-In/Pop-Out Integration: Two Step Gene Replacement

To replace the chromosomal copy of a gene with another allele, Scherer and Davis developed a pop-in/pop-out method (Scherer and Davis 1979). This method, illustrated in Fig. 3.3, involves two steps: first, plasmid integration is selected using the *URA3* gene for selection and second, a plasmid excision event is counter-selected

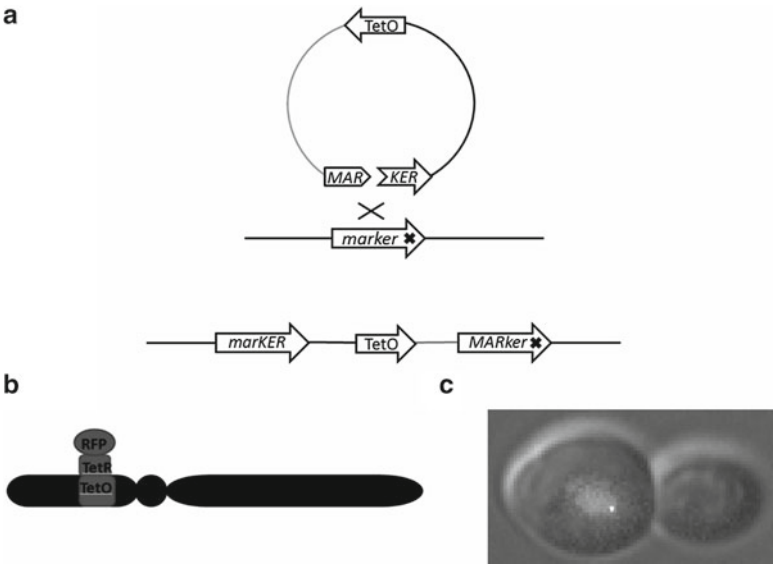


Fig. 3.2 Tagging a chromosomal locus *in vivo* by "ends-in" integration. (**a**): Schematic of an "ends-in" integration of a TetO sequence. The TetO array is integrated at the locus of the marker locus as described in Fig 3.1. (**b**): Schematic of a chromosome after integration of the TetO array. The TetR-RFP sequence is also integrated elsewhere in the genome and TetR-RFP proteins bind to the TetO repeats. (**c**): Wide-field microscopy image of a yeast cell after integration of a TetO array at the *URA3* locus. The strain contains a TetR-RFP sequence as well and Rad52-CFP proteins to mark the nucleus. In this *grayscale image*, the "red" TetO site is shown as a *bright spot* and the "cyan" Rad52 protein is shown as a *gray circle* within the mother cell on the *left*

using the same *URA3* gene. For this method, a specific alteration is introduced into the gene of interest and it is cloned to a *URA3*-based integrating plasmid. The alteration can be a deletion, an insertion or base-pair mutations. Next, the plasmid is integrated into its chromosomal location by homologous recombination, as described for the "ends-in" integration. This creates a duplication containing a wild type copy of the gene of interest and a mutant copy flanking the plasmid sequences. Since the two copies share homology, they can recombine and the excision event can be selected on 5-fluoro-orotic acid (5-FOA) medium (Boeke et al. 1984). The successful pop-out event can be screened by its resulting phenotype. Pop-in/pop-out is also useful when multiple disruptions are done in the same strain since the *URA3* selectable/counter-selectable marker can be used repeatedly.

3.2.3 "Ends-Out" Integration: One Step Gene Replacement

In 1983, a third method of gene targeting was developed that allows stable integration of the transforming DNA sequence without integration of the plasmid backbone and without duplication of the integration site (Rothstein 1983).

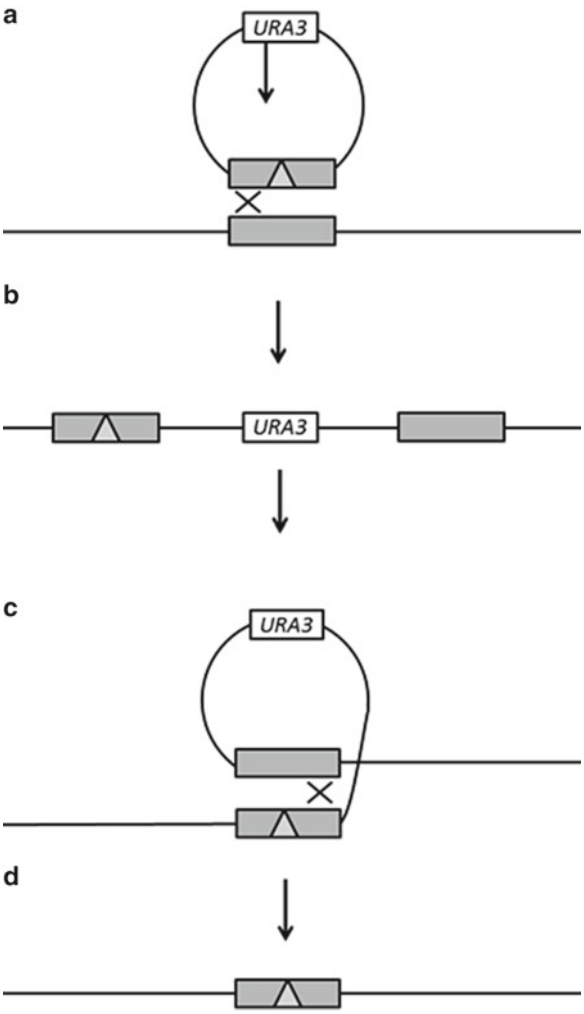


Fig. 3.3 Pop-in/pop-out integration for allele replacement. (**a**): A plasmid containing the *URA3* sequence and the mutated sequence of a gene is engineered. After cutting the plasmid in the sequence of the mutated gene (not shown), the linearized plasmid is transformed into a yeast strain auxotrophic for *ura3* (see Fig. 3.1 for details). (**b**): After selection of the transformants, the chromosome contains the sequence of the mutated gene, followed by the sequence of *URA3* and the wild type copy of the gene. (**c**): Since the sequences of the mutated and the wild type gene share homology, they can recombine and the excision event can be selected on 5-FOA medium (**d**): The product of the pop-in/pop-out event is replacement of the wild-type copy of the gene with the engineered mutation

This technique was the first example of "ends-out" integration where the two different fragments of DNA homologous to the target genomic locus are cloned flanking a dominant selectable marker as illustrated in Fig. 3.4. "Ends-out" integration requires the construction of a DNA fragment that includes the selectable marker with adjacent sequences from the upstream and downstream regions of

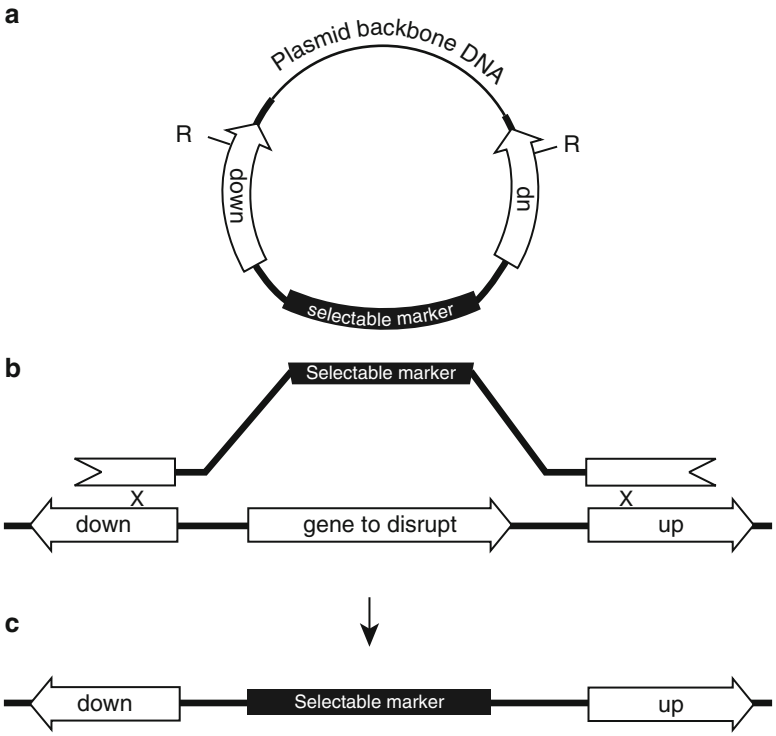


Fig. 3.4 "Ends-out" integration: one step gene disruption. (**a**): A DNA fragment is first engineered containing a selectable marker with adjacent sequences from the upstream and downstream regions of the genomic target sequences. The letters "R" on the side of the plasmid indicate restriction sites where the plasmid is cut before transformation. (**b**): The ends of this DNA fragment then interact directly with the homologous sequences in the genome. (**c**): After selection of the marker, the chromosomal copy of the gene is replaced by the marker sequence

the genomic target sequences. The ends of this DNA fragment then interact directly with the homologous sequences in the genome (Langston and Symington 2005). This method has been used to delete genes as well as modify them directly by fusing desired sequences in the plasmid construct and selecting for the appropriate crossover to recover the fusion. The advent of PCR has greatly simplified gene disruption technology. Michaelis and colleagues showed that PCR primers can be designed with the 5′ sequences homologous to the desired crossover sites in the gene and the 3′ ends homologous to the selectable marker (Oldenburg et al. 1997). The PCR product creates a substrate that can be introduced via homologous recombination to precisely replace the genomic locus. The ease with which deletions and targeted mutations can be accomplished led to the creation of the yeast gene disruption library in which every budding yeast open reading frame is disrupted (Winzeler et al. 1999).

3.2.4 Hybrid Methods

Variations on "ends-out" and "pop-in/pop-out" integrations have been developed to take advantage of these techniques. In this section, we describe two of these, gene fusion and allele replacement.

Gene fusion is a variant of the "pop-in/pop-out" technique and was first described to fuse a fluorescent protein to a gene of interest (Lisby et al. 2001). In the first step, two PCRs are performed to amplify sequences upstream and downstream of the GFP insertion site. The resulting fragments each contain an additional 19 bases pairs complementary to GFP sequence. As illustrated in Fig. 3.5, the upstream PCR product is fused by fusion PCR to the 5′-two-thirds of the *K. lactis URA3* gene and downstream product to the 3′-two-thirds of the *K. lactis URA3* gene (Fig 3.5a). The two fusion products are next transformed into a yeast strain auxotrophic for *URA3* to integrate a GFP-*K.lactis URA3*-GFP shown in Fig. 3.5b. Finally, the cells are plated on 5-FOA media that select cells in which a pop-out event occurred. The extra GFP copy is thus removed as shown on Fig. 3.5c, d, and the resulting product is a precise gene-GFP fusion with its normal promoter and terminator intact.

Using a similar strategy, allele replacement can be performed (Fig. 3.6). Here, the idea is to introduce a mutation in the gene of interest after recombination of two PCR products each containing a mutated allele followed by a pop-out event. Since the direct repeats both contain the same allele, there is no constraint on where the pop-out crossover occurs. This is unlike the pop-in/pop-out crossover shown in Fig. 3.3, which has to occur on one side of the mutation during integration and on the other side during excision. To create the integration substrates for allele replacement, the allele (*) to be transferred is either made by PCR from an existing template (transferring an allele from one strain to another) or is created *de novo* using synthetic oligonucleotides. The 3′ end of the fragment is fused to the 5′ two-thirds of the *K. lactis URA3* gene. Similarly, the 5′ end of the fragment is fused to the 3′ two-thirds of *K. lactis URA3* (Fig. 3.6a). The two PCR fragments are then co-transformed into a strain auxotrophic for the *URA3* gene. The product of this reaction is two direct repeats each containing the mutated sequence flanking the *K. lactis URA3* (Fig. 3.6b). Direct repeat recombinants are selected on 5-FOA plates (Fig. 3.6c). Since both copies of the repeat contain the desired allele, the crossover resulting in loss of the *K. lactis URA3* can occur anywhere between the two repeats giving rise to the allele replacement as shown in Fig. 3.6d.

3.2.5 Plasmid Gap Repair

The ability to construct plasmids *in vitro* allows researchers to precisely replace and clone any gene in the genome. *In vivo* gap repair is an important technique commonly used to engineer plasmids. This method is used to retrieve genetic information from a chromosomal sequence (or from a co-transformed fragment) by genetic recombination into the gapped plasmid, which also contains a yeast origin

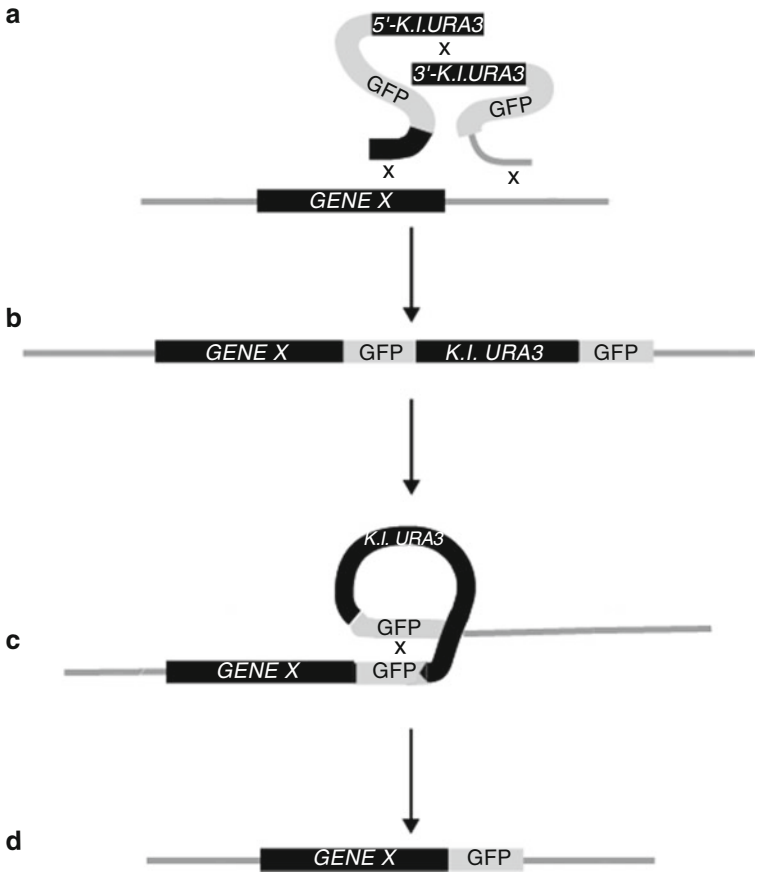


Fig. 3.5 Gene fusion. PCRs are performed to amplify sequences upstream and downstream of the GFP insertion site at the end of *gene X*. At the same time, GFP sequences and the 5′ two-thirds and 3′ two-thirds of the *K. lactis URA3* gene are also amplified. The *K. lactis URA3* gene, which is ≈ 20% diverged from *S. cerevisiae* is used to avoid recombination events at the chromosomal *ura3* site. The PCR products are fused by fusion PCR to create the fragments shown in (**a**). (**b**): Integration of the PCR fragments results in the integration of a fusion between gene X and GFP followed by *K.l URA3* and a second copy of the GFP sequence (**c**): Pop-out events: the two GFP fragments can recombine to undergo a pop-out event removing the *K.l. URA3* sequence as described in Fig. 3.3. Those events are selected by plating the cells on 5-FOA medium. (**d**): The result of this integration and pop-out is a fusion between gene X and GFP

of DNA replication and a centromere sequence. The centromere DNA insures that the linear plasmid does not integrate into the yeast genome after gap repair, but instead replicates autonomously. To perform a gap repair reaction, the desired plasmid is digested by a restriction enzyme to create a linear DNA fragment that contains upstream and downstream "ends-in" sequences that flank the region to be repaired. After transformation and selection, only cells containing the repaired plasmids can grow and the newly formed plasmid is then recovered from the parent yeast strain by DNA extraction and transformation into bacteria.

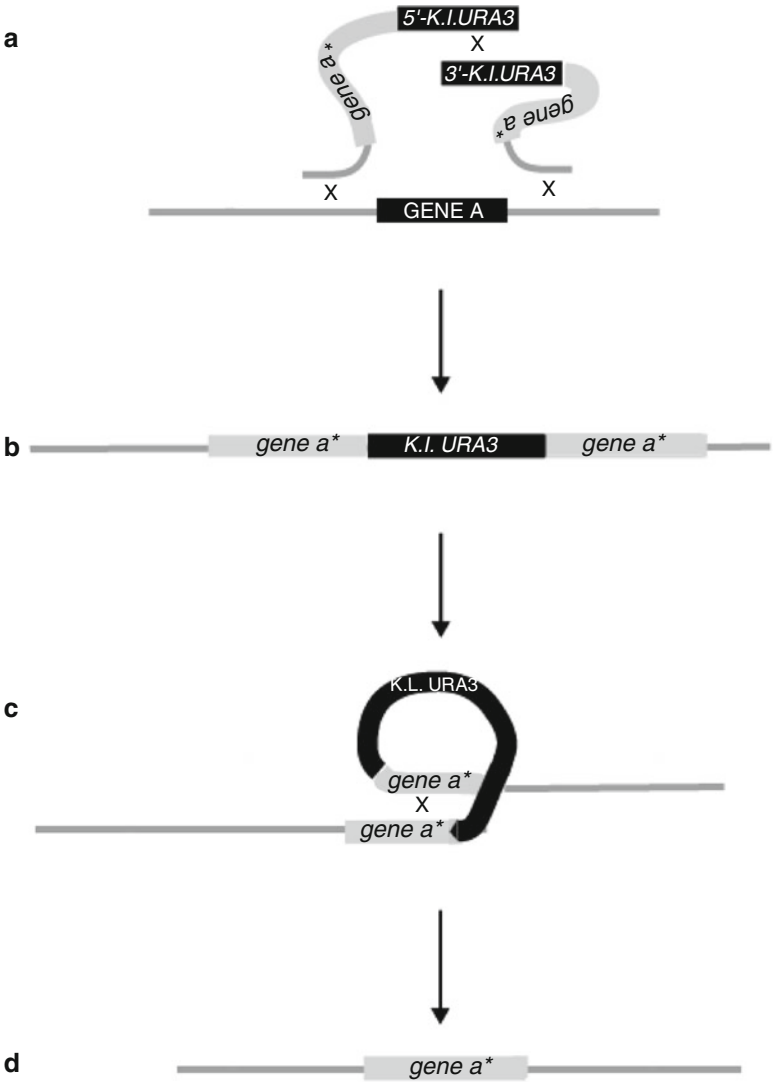


Fig. 3.6 Allele replacement. (**a**): To replace *GENE A* with a mutant allele, *gene a**, the methodology is similar to that shown in Fig. 3.5. Two PCR fragments are first engineered where the first PCR fragment contains a mutated copy of *gene a* fused at its 3′ end to the 5′ two-thirds of the *K.l. URA3* sequence. The second PCR fragment contains a mutated copy of *gene a* fused at its 5′ end to the 3′ two-thirds of the *K.l. URA3* sequence. (**b**): After recombination and selection of Ura+transformants, the chromosome contains the *K.l URA3* sequence flanked by two copies of the mutant *gene a** sequence. (**c**): Pop-out events are selected by plating the cells on 5-FOA medium. (**d**): The result of this integration is the replacement of the *GENE A* by the mutant allele *gene a**

3.3 Mechanism of Targeted Integration: A State of the Art of Homologous Recombination

3.3.1 An Overview of Homologous Recombination

Targeted integration occurs via homologous recombination. In this section, we give an update of current views concerning the mechanism of mitotic recombination and provide details in the next section. Recently, much progress has been made in understanding the mechanism of homologous recombination using budding yeast as a model system reviewed in (Prado et al. 2003). A current model of homologous recombination initiated by a DSB is illustrated in Fig. 3.7. After formation of a DSB, the 5′ ends of the break are resected by nucleases to yield 3′ single-stranded DNA (ssDNA) tails on which the repair proteins polymerize.

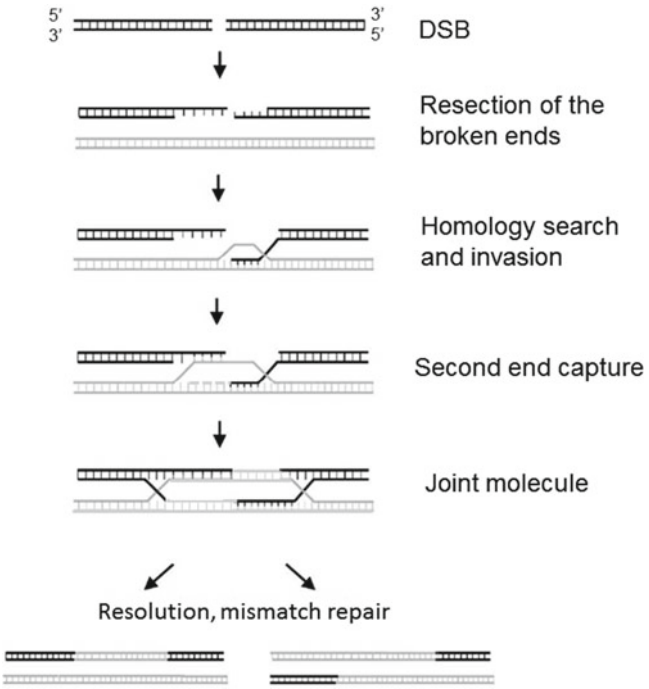


Fig. 3.7 Model of homologous recombination. After the formation of a DSB, the broken ends are resected to form ssDNA tails (*resection of the broken ends*). Those ssDNA ends are then covered by homologous recombination proteins and invade the homologous DNA sequence. A two-step invasion is described here where one 3′ end invades and is elongated displacing a D-loop. The second resected end anneals with the D-loop (*invasion and second end capture*). The newly synthesized ends are ligated forming a joined molecule containing two Holliday junctions (*joint molecule*)

This repair protein-ssDNA complex, called a nucleoprotein filament, then searches for homologous sequences among neighboring double-stranded DNA (dsDNA) molecules. A common source for an intact duplex DNA donor is the undamaged sister chromatid, especially in mitotic cells (Kadyk and Hartwell 1992; Bzymek et al. 2010). However homologous sequences on either the homologue or on a different chromosome can be captured by the presynaptic nucleofilament to perform inter-homologue recombination or ectopic recombination respectively (Aylon et al. 2003). Once homology is found, the broken ssDNA ends invade the intact donor template, displacing ssDNA loops and forming two regions of heteroduplex DNA (invasion and the second end capture step). The invading ssDNA ends then serve as primers for DNA synthesis. As new DNA synthesis proceeds, branch migration may displace the two newly synthesized strands. Mismatches present in the heteroduplex DNA can be repaired by the mismatch repair (MMR) system, generating a non-reciprocal transfer of information from a donor to the recipient chromosome (gene conversion). Ligation of newly synthesized DNA to the resected broken chromosome generates an intermediate known as a joint molecule containing two Holliday junctions. Joint molecules can be resolved to produce chromosomes that have, or have not, undergone a reciprocal exchange of genetic material flanking the repair site (crossover or non-crossover, respectively). The genetic dependencies of many of these steps have been elucidated *in vivo* by visualizing tagged recombination proteins moving in and out of repair centers in time and in various mutant backgrounds (Lisby et al. 2004).

3.3.2 Step by Step Mechanism of Homologous Recombination

3.3.2.1 Resection and Formation of the Nucleoprotein Filament

After the formation of a DSB, the Mre11 complex (Mre11-Rad50-Xrs2) is first loaded at the broken ends to keep them in close proximity. When the decision is made to repair a DSB via homologous recombination, the ends undergo 5′−>3′ degradation to generate 3′-ended single-strand tails. Several proteins are implicated in this DNA resection: the Mre11 complex, Sae2, Exo1 and Sgs1. The requirement for MRX and Sae2 in end resection depends on the nature of the DNA ends. When a DSB is induced by a restriction endonuclease, the broken ends are considered clean and resection can occur in the absence of MRX and Sae2. On the other hand, DSBs generated by irradiation or with proteins bound to the broken ends are considered complex and dirty (Henner et al. 1983; Barker et al. 2005). In that case, the MRX complex and Sae2 are required to process the resection of the broken ends. First, the Mre11 complex and Sae2 remove a small oligonucleotide from the DNA ends to form an early intermediate. Next, Exo1 and/or Sgs1/Dna2 rapidly process this intermediate to generate extensive tracts of single-stranded DNA (Mimitou and Symington 2008; Zhu et al. 2008).

The rate of resection ranges from approximately 40 nucleotides per minute (Aylon et al. 2003) to about 60 nucleotides per minute (Sugawara and Haber 1992; Vaze et al. 2002) depending on experimental systems. In diploid cells, in which homologous sequences are present, resection extends 3.5 kb from the DSB, sometimes exposing ssDNA beyond the area of shared homology (Aylon and Kupiec 2004). Although the 5′ end is extensively resected, there is little or no processing of the 3′ end at this stage (Aylon et al. 2003; Frank-Vaillant and Marcand 2002). It is not clear if both broken ends or only one broken end is resected (Aylon and Kupiec 2004).

The single-strand tail formed after resection is then covered by RPA proteins, which protect the DNA against degradation and inhibit the formation of secondary structures (Alani et al. 1992). The next step of homologous recombination is the formation of a nucleoprotein filament on the single strand tail. This step requires the activity of several members of the Rad52 group. First, the Rad52 protein is recruited to the RPA-coated ssDNA tail; it replaces RPA and recruits Rad51 proteins, the central protein of homologous recombination. Rad51 proteins polymerize on DNA to form a right-handed helix around the DNA. The Rad54 protein, a member of the Swi2/Snf2 family of chromatin-remodeling proteins is also loaded at the broken ends to stabilize Rad51 proteins on ssDNA (Mazin et al. 2003). Inside this helix, the DNA molecule is extended by a factor of 1.5 compared to B-form DNA (Stasiak and Egelman 1986; Hilario et al. 2009), and dsDNA is unwound by 18 per base pair (Arata et al. 2009). The mechanical properties of a nucleofilement change dramatically due to the presence of the repair proteins on the single stand (Stasiak and Egelman 1986; Mine et al. 2007; Sheridan et al. 2008). Indeed, the Rad51 nucleoprotein filament is extremely rigid with a persistence length of 500–800 nm whereas the persistence lengths of naked ssDNA and dsDNA are 2 nm and 60 nm respectively.

3.3.2.2 Homology Search

Once the nucleoprotein filament is formed, a genome-wide search for homologous sequences takes place. Homology search is the most enigmatic step of homologous recombination and the precise mechanism allowing a broken locus to find its homologue so efficiently is still poorly understood. However, it is known that homology search is dependent on the presence of Rad52, Rad51 and Rad55 at the broken ends (Aylon et al. 2003; Sugawara et al. 2003). Theoretical and *in vitro* studies showed that the stiffness of the Rad51 nucleoprotein filament enhances homology recognition by preventing the formation of secondary structures at the broken ends, as well as insuring multiple attachment points between the nucleoprotein filament and its target (Klapstein et al. 2004; Mine et al. 2007). The Rad54 protein is also thought to have an essential role during homology search. Using single molecule approaches, it was shown that Rad54 proteins possess a unique ability to cross-bridge or bind double-stranded DNA molecules positioned in close proximity (Bianco et al. 2007). In addition, Rad54 rapidly translocates on dsDNA while

simultaneously inducing topological loops in the DNA. The combination of the cross-bridging and double-stranded DNA translocation activities of Rad54 stimulates the formation of DNA networks, leading to a more efficient homology search (Bianco et al. 2007). The presence of Rad54 proteins on the Rad51 nucleoprotein filament was also shown to increase the movement of nucleosomes in an ATP dependent manner and thus facilitate the test for homology on dsDNA molecules (Alexeev et al. 2003; Wolner and Peterson 2005; van Attikum and Gasser 2005). Moreover, it was shown that remodeling activity of Rad54 is greatly stimulated by the Rad51-ssDNA nucleofilament indicating that Rad51 and Rad54 proteins act in concert during homology search (Alexeev et al. 2003; Kwon et al. 2007).

3.3.2.3 Strand Invasion and DNA Synthesis

In vitro studies have shown that Rad51 is able to promote strand exchange between homologous sequences on its own. This reaction can even occur in the absence of ATP (Rice et al. 2001). However, it was shown that optimization of extensive strand exchange without ATP hydrolysis requires conditions that decrease interactions between DNA molecules. The presence of multiple attachment points between homologous sequences creates entanglements that cannot be resolved in the absence of ATP hydrolysis (Rice et al. 2001). Rad55 and Rad57 proteins appear to play an important role during DNA strand exchange since invasion and DNA synthesis are not completed in *rad55Δ* and *rad57Δ* cells (Aylon et al. 2003; Sung 1997). The precise activity of Rad55 and Rad57 proteins during this step of homologous recombination is still unknown. After strand exchange, it has been shown *in vitro* that Rad54 proteins are required to dissociate Rad51 proteins from heteroduplex DNA (Solinger et al. 2002).

3.3.2.4 Resolution of the Repair

After strand exchange, the Holliday junction(s) need to be resolved to ensure proper chromosome segregation. The Mus81/Mms4 complex has been recently identified to resolve joint molecules in yeast (Ho et al. 2010; Matos et al. 2011). The Yen1 protein serves a backup function responsible for resolving intermediates in *mus81*Δ mutants. To coordinate the resolution with chromosome segregation, Cdc5 regulates the activity of Mus81-Mms4 resolvases.

3.3.3 Timing of the Repair by Homologous Recombination

Repair by homologous recombination is surprisingly rapid and efficient. A cell requires about 1 h to repair a single DSB (Aylon et al. 2003). Measurements in living yeast cells show that there is approximately 30–40 min between the formation

of a single DSB and the formation of a ssDNA tail at the broken ends (Aylon et al. 2003). Gene conversion events are then detected around 30 min after the resection of the broken ends (Aylon et al. 2003). Similar timing were observed for mating type switching despite the differences between the processes and strains used (White and Haber 1990). *In vitro* studies showed that the time to complete homology search and strand exchange between a 14 kb RecA nucleofilament and a 14 kb dsDNA is 30 min (Fulconis et al. 2006). Recently, we have found *in vivo* that a chromosome with an induced DSB and its homologue remain paired for approximately 20 min before the repair center disappears and the homologues separate (Miné-Hattab and Rothstein, unpublished observations).

3.4 Conclusions and Future Directions

The techniques developed in yeast for integrating a sequence into the genome by homologous recombination have proved valuable for studying biological processes in this model system. In many cases, these same techniques have been applied in a wide variety of systems to both integrate sequences in the genome and perform gene disruption experiments. Indeed, the gene knock-out technology in mammalian cells derived from the work in bacteria and yeast was awarded the 2007 Nobel Prize in Physiology or Medicine. In the future, it will be important to understand the relationship between chromatin structure and DNA dynamics to learn how to enhance targeted events. In addition, site-specific meganucleases such as those derived from I-*Sce*I or I-*Cre*I (Arnould et al. 2011) as well as zinc-finger nucleases (Carroll 2008) or TALENs (Christian et al. 2010) promise to increase gene targeting by cutting specific chromosomal sites to stimulate homologous integration from a linear DNA repair matrix. Here again, yeast will be an excellent model system to develop some of these technologies.

References

Alani E, Thresher R, Griffith JD, Kolodner RD (1992) Characterization of DNA-binding and strand-exchange stimulation properties of y-RPA, a yeast single-strand-DNA-binding protein. J Mol Biol 227:54–71

Alexeev A, Mazin A, Kowalczykowski SC (2003) Rad54 protein possesses chromatin-remodeling activity stimulated by the Rad51-ssDNA nucleoprotein filament. Nat Struct Biol 10:182–186

Arata H, Dupont A, Mine-Hattab J, Disseau L, Renodon-Corniere A, Takahashi M, Viovy JL, Cappello G (2009) Direct observation of twisting steps during Rad51 polymerization on DNA. Proc Natl Acad Sci USA 106:19239–19244

Arnould S, Delenda C, Grizot S, Desseaux C, Paques F, Silva GH, Smith J (2011) The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. Protein Eng Des Sel 24:27–31

Avery OT, Macleod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III. J Exp Med 79:137–158

Aylon Y, Kupiec M (2004) New insights into the mechanism of homologous recombination in yeast. Mutat Res 566:231–248

Aylon Y, Liefshitz B, Bitan-Banin G, Kupiec M (2003) Molecular dissection of mitotic recombination in the yeast *Saccharomyces cerevisiae*. Mol Cell Biol 23:1403–1417

Barker S, Weinfeld M, Zheng J, Li L, Murray D (2005) Identification of mammalian proteins cross-linked to DNA by ionizing radiation. J Biol Chem 280:33826–33838

Beggs JD (1978) Transformation of yeast by a replicating hybrid plasmid. Nature 275:104–109

Belmont AS (2001) Visualizing chromosome dynamics with GFP. Trends Cell Biol 11:250–257

Bianco PR, Bradfield JJ, Castanza LR, Donnelly AN (2007) Rad54 oligomers translocate and cross-bridge double-stranded DNA to stimulate synapsis. J Mol Biol 374:618–640

Boeke JD, Lacroute F, Fink GR (1984) A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet 197:345–346

Bzymek M, Thayer NH, Oh SD, Kleckner N, Hunter N (2010) Double Holliday junctions are intermediates of DNA break repair. Nature 464:937–941

Carroll D (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. Gene Ther 15:1463–1468

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186:757–761

Frank-Vaillant M, Marcand S (2002) Transient stability of DNA ends allows nonhomologous end joining to precede homologous recombination. Mol Cell 10:1189–1199

Fulconis R, Mine J, Bancaud A, Dutreix M, Viovy JL (2006) Mechanism of RecA-mediated homologous recombination revisited by single molecule nanomanipulation. EMBO J 25:4293–4304

Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. Yeast 15:1541–1553

Griffith F (1928) The significance of Pneumococcal types. J Hyg (Lond) 27:113–159

Henner WD, Rodriguez LO, Hecht SM, Haseltine WA (1983) Gamma ray induced deoxyribonucleic acid strand breaks. 3′ glycolate termini. J Biol Chem 258:711–713

Hilario J, Amitani I, Baskin RJ, Kowalczykowski SC (2009) Direct imaging of human Rad51 nucleoprotein dynamics on individual DNA molecules. Proc Natl Acad Sci USA 106:361–368

Hinnen A, Hicks JB, Fink GR (1978) Transformation of yeast. Proc Natl Acad Sci USA 75:1929–1933

Ho CK, Mazon G, Lam AF, Symington LS (2010) Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast. Mol Cell 40:988–1000

Kadyk LC, Hartwell LH (1992) Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. Genetics 132:387–402

Klapstein K, Chou T, Bruinsma R (2004) Physics of RecA-mediated homologous recognition. Biophys J 87:1466–1477

Kwon Y, Chi P, Roh DH, Klein H, Sung P (2007) Synergistic action of the *Saccharomyces cerevisiae* homologous recombination factors Rad54 and Rad51 in chromatin remodeling. DNA Repair (Amst) 6:1496–1506

Langston LD, Symington LS (2005) Opposing roles for DNA structure-specific proteins Rad1, Msh2, Msh3, and Sgs1 in yeast gene targeting. EMBO J 24:2214–2223

Lisby M, Rothstein R, Mortensen UH (2001) Rad52 forms DNA repair and recombination centers during S phase. Proc Natl Acad Sci USA 98:8276–8282

Lisby M, Barlow JH, Burgess RC, Rothstein R (2004) Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. Cell 118:699–713

Marshall WF, Sedat JW (1999) Nuclear architecture. Results Probl Cell Differ 25:283–301

Matos J, Blanco MG, Maslen S, Skehel JM, West SC (2011) Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis. Cell 147:158–172

Mazin AV, Alexeev AA, Kowalczykowski SC (2003) A novel function of Rad54 protein. stabilization of the Rad51 nucleoprotein filament. J Biol Chem 278:14029–14036

Meister P, Gehlen LR, Varela E, Kalck V, Gasser SM (2010) Visualizing yeast chromosomes and nuclear architecture. Methods Enzymol 470:535–567

Mimitou EP, Symington LS (2008) Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. Nature 455:770–774

Mine J, Disseau L, Takahashi M, Cappello G, Dutreix M, Viovy JL (2007) Real-time measurements of the nucleation, growth and dissociation of single Rad51-DNA nucleoprotein filaments. Nucleic Acids Res 35:7171–7187

Oldenburg KR, Vo KT, Michaelis S, Paddon C (1997) Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. Nucleic Acids Res 25:451–452

Orr-Weaver TL, Szostak JW, Rothstein RJ (1981) Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci USA 78:6354–6358

Prado F, Cortes-Ledesma F, Huertas P, Aguilera A (2003) Mitotic recombination in *Saccharomyces cerevisiae*. Curr Genet 42:185–198

Rice KP, Eggler AL, Sung P, Cox MM (2001) DNA pairing and strand exchange by the *Escherichia coli* RecA and yeast Rad51 proteins without ATP hydrolysis: on the importance of not getting stuck. J Biol Chem 276:38570–38581

Rothstein RJ (1983) One-step gene disruption in yeast. Methods Enzymol 101:202–211

Rothstein R (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol 194:281–301

Scherer S, Davis RW (1979) Replacement of chromosome segments with altered DNA sequences constructed in vitro. Proc Natl Acad Sci USA 76:4951–4955

Sheridan SD, Yu X, Roth R, Heuser JE, Sehorn MG, Sung P, Egelman EH, Bishop DK (2008) A comparative analysis of Dmc1 and Rad51 nucleoprotein filaments. Nucleic Acids Res 36:4057–4066

Shortle D, Haber JE, Botstein D (1982) Lethal disruption of the yeast actin gene by integrative DNA transformation. Science 217:371–373

Solinger JA, Kiianitsa K, Heyer WD (2002) Rad54, a Swi2/Snf2-like recombinational repair protein, disassembles Rad51:dsDNA filaments. Mol Cell 10:1175–1188

Stasiak A, Egelman EH (1986) Structure and dynamics of recA protein-DNA complexes as determined by image analysis of electron micrographs. Biophys J 49:5–7

Straight AF, Belmont AS, Robinett CC, Murray AW (1996) GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. Curr Biol 6:1599–1608

Sugawara N, Haber JE (1992) Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Mol Cell Biol 12:563–575

Sugawara N, Wang X, Haber JE (2003) *In vivo* roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. Mol Cell 12:209–219

Sung P (1997) Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. Genes Dev 11:1111–1121

Thaler DS, Stahl FW (1988) DNA double-chain breaks in recombination of phage lambda and of yeast. Annu Rev Genet 22:169–197

van Attikum H, Gasser SM (2005) ATP-dependent chromatin remodeling and DNA double-strand break repair. Cell Cycle 4:1011–1014

van den Berg MA, Steensma HY (1997) Expression cassettes for formaldehyde and fluoroacetate resistance, two dominant markers in *Saccharomyces cerevisiae*. Yeast 13:551–559

Vaze MB, Pellicioli A, Lee SE, Ira G, Liberi G, Arbel-Eden A, Foiani M, Haber JE (2002) Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. Mol Cell 10:373–385

White CI, Haber JE (1990) Intermediates of recombination during mating type switching in *Saccharomyces cerevisiae*. EMBO J 9:663–673

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-Macdonald P, Scherens B, Snyder

M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 285:901–906

Wolner B, Peterson CL (2005) ATP-dependent and ATP-independent roles for the Rad54 chromatin remodeling enzyme during recombinational repair of a DNA double strand break. J Biol Chem 280:10855–10860

Zhu Z, Chung WH, Shim EY, Lee SE, Ira G (2008) Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. Cell 134:981–994

Zimmer C, Fabre E (2011) Principles of chromosomal organization: lessons from yeast. J Cell Biol 192:723–733

Chapter 4
Homologous Recombination in Mammals

Aurélia Barascu*, Anastazja Grabarz*, and Bernard S. Lopez

Abstract Homologous recombination (HR) is a fundamental process that is conserved in all species and is essential for maintaining genome stability while facilitating genetic diversity but avoiding genetic instability. As such, HR is involved in numerous fundamental biological processes, controlling a balance of genetic stability/diversity instability. This chapter summarizes the different models of recombination in mammals and then presents the following molecular steps of HR: recognition of damage, chromatin remodeling, DNA single-strand resection, strand invasion and resolution of Holliday junctions. HR is a process involved in DNA repair (notably double-strand breaks), the tolerance of damages and the reactivation of stalled replication forks; it is essential for the maintenance of genome stability. Conversely, an excess of HR (for example, between dispersed repeated sequences in the genome) can lead to genetic instability. In this context, the regulation of HR restricting it to S-phase is a factor in genetic stability. Deregulation of HR (causing a defect or an excess) is often found in tumor cells and tumor predisposition situations, illustrating the need for precise control of balance in HR. Knowledge of the molecular mechanisms of HR is also of interest in applications because it may provide strategies for the optimization of targeted in gene replacement.

---

*equal contribution

A. Barascu • B.S. Lopez (✉)
Université Paris Sud, Laboratoire «Stabilité Génétique et Oncogenèse» CNRS,
UMR 8200 GR – PR2, 114 Rue Edouard Vaillant, 94805 Villejuif, France

Institut de radiobiologie cellulaire et moléculaire, Institut de Cancérologie
Gustave-Roussy CNRS, UMR217
e-mail: bernard.lopez@igr.fr; aurelia.barascu@igr.fr

A. Grabarz
Laboratoire "Biologie cellulaire de l'intégrité du génome", Institut de Génétique et de Biologie
Moléculaire et Cellulaire, Illkirch, France
e-mail: grabarz@igbmc.fr

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_4,

Keywords Homologous recombination • DNA repair • Genetic instability • Cancer • Mammalian cells

The complementary structure of the two helices of DNA enabled Watson and Crick in 1953 (Watson and Crick 1953) to provide molecular support to the hypotheses of duplication and transmission of genetic heritage by DNA. They also suggested the possibility of the exchange of one strand with an identical strand of a molecule of homologous sequence, reforming the hydrogen bonds with its new complementary partner. This mechanism of single-stranded DNA (ssDNA) exchange is the basic principle of homologous recombination (HR); it provided the first molecular basis to the "crossover" hypothesis proposed by T.H. Morgan in 1916 and demonstrated by Creighton and McClintock in 1931 (Creighton and McClintock 1931), the first models of recombination and calculations of distances between markers begun by Haldane in 1919 (Haldane 1919; Haldane and Lea 1947) and to the observation of recombination during bacterial conjugation reported by Lederberg and Tatum in 1946 (Lederberg and Tatum 1946). Finally, molecular evidences of intrachromosomal gene conversion in mammalian cells was established in 1983 (Liskay and Stachelek 1983).

HR is the exchange of DNA molecules that exhibit sequence homologies. It is a fundamental process that is made possible by the structure of the two complementary strands and that has been described in all living organisms. HR can occur anywhere in the genome but not in an equiprobable fashion because there are hot and cold regions related to the nature of the chromosomal environment, the possibility of causing damage that induces HR and/or the accessibility of DNA to the recombination machinery.

Though HR requires homology between the sequences involved, perfect identity is not required. The products of HR are gene conversion and crossovers (Fig. 4.1a), which allows HR to generate new combinations of genetic material. This ability, coupled with the fact that it allows repair of DNA breakage, places HR at the heart of the equilibrium controlling the balance between genetic stability and variability. HR is therefore implicated in numerous essential biological processes, from molecular evolution to DNA repair to meiotic differentiation, and it is also relevant to the applications of targeted gene replacement (Fig. 4.2). An understanding of the role of HR is possible only in light of the molecular models that explain how it functions.

4.1 Models and Products of HR

The need for sequence homology between the two partner molecules led to the concept of MEPS (minimal efficient processing segment). The efficiency of recombination decreases gradually with the reduction in the size of DNA segments with lengths of perfect homology (Fig. 4.1b). Below a certain length of perfect homology, which can vary among organisms, the efficiency of recombination drops sharply; the length of homology at the breaking point of the curve is defined as the MEPS. In mammals, the MEPS has been estimated to be between 200 and 250 nucleotides

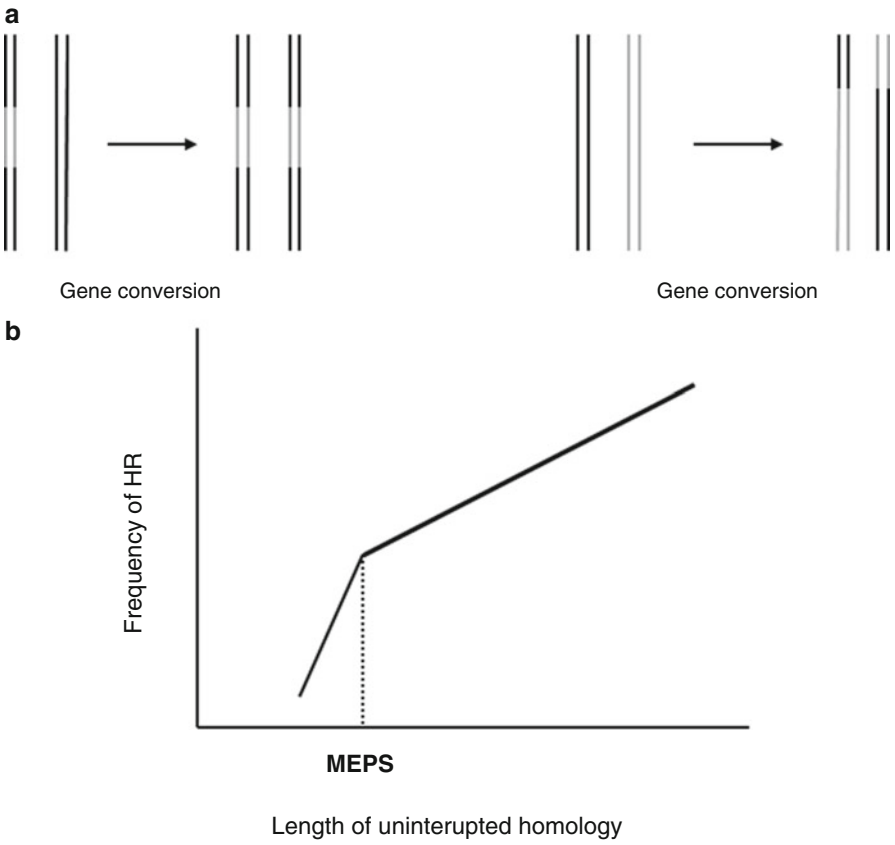


Fig. 4.1 (**a**) Gene conversion (*left panel*): non-reciprocal transfer of genetic material. Crossover (right panel): reciprocal exchange. (**b**) MEPS (minimal efficient processing segment). The efficiency of recombination increases regularly with uninterrupted length of homology between the two DNA molecules. Below a specific length of homology, (the MEPS 200–250 nt in mammals), the efficiency of HR exhibits a break in the slope, and the efficiency of recombination drops severely (without being completely inhibited)

(Liskay et al. 1987; Lopez et al. 1992; Rubnitz and Subramani 1984). The MEPS can reflect the need for extensive resection of ssDNA at the start of recombination (see below). The MEPS also has a significant impact on the efficiency of gene targeting and, therefore, on the strategies and choice of substrate to use.

Different molecular models explain the modes of function of HR (for review see Paques and Haber 1999). The DNA double-strand break repair model proposed in yeast (Szostak et al. 1983) integrates the role of HR, leading to gene conversion associated or not with crossovers (Fig. 4.3a). Moreover, this model is the basis for experimental strategies that led to the development of targeted gene replacement. HR is initiated by the resection of ssDNA, which generates a 3′ tail that then invades a double-stranded DNA of homologous sequence. This process tolerates some differences between the molecules involved, thereby creating a hybrid double-stranded

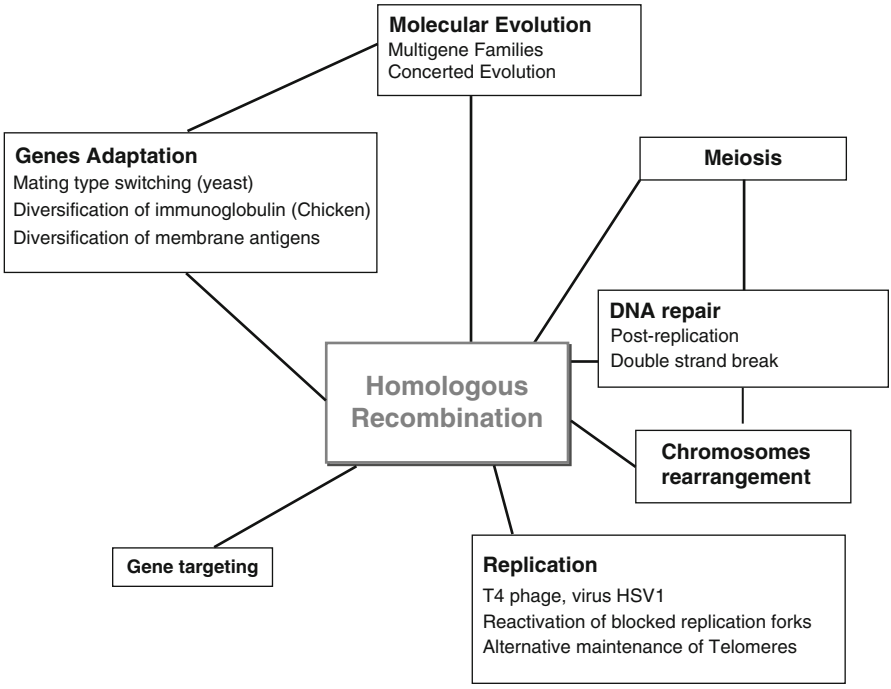


Fig. 4.2 Homologous recombination is implicated in the equilibrium between stability and variability/instability of the genome and therefore participates in numerous fundamental processes

DNA molecule, called a heteroduplex, which carries mismatches. Under the direction of mismatch repair, genetic information may or may not be converted to match the original strand, leading to gene conversion. In addition, the 3′ end of the invading strand allows for DNA synthesis via recopying the recipient molecule; a sequence absent from the invading strand can be recopied, leading to the non-reciprocal

Fig. 4.3 (continued) Initiation is similar to that of the previous model, with a single-strand resection, invasion of the homologous double-stranded DNA and DNA synthesis (*a*), but the invading strand dehybridizes and reanneals at the other end of the injured molecule (*b*). No Holliday junction forms; this model only accounts for gene conversion without crossovers (*c*). (**d**) Break-induced replication (*BIR*). Initiation is similar to that of the previous models, with a single-strand resection, invasion of the homologous double-stranded DNA and DNA synthesis (*a*), but the synthesis continues over longer distances on the chromosome arms (*b*) and even can reach the end of the chromosome (*c*). Here again, there is neither resolution of Holliday junctions nor crossover. (**e**) Single-strand annealing (SSA). A double-strand break occurs between two homologous sequences in tandem in the same orientation (*dotted arrows*). (*a*) A single-strand resection then reveals two complementary DNA strands that can hybridize (*b*). If the tandem sequences are in opposite orientations, the revealed single strands of DNA are not complementary but are identical; they therefore cannot hybridize. In this process, the break does not need to be in or near a region of homology. (*c*) Resolution of this intermediate by the ERCC1/XPF endonuclease from the system of repair by nucleotide excision and filling the gap of the single strand completes the repair of the double-strand break and leads to the deletion of the intergenic sequences between the initial repetitions

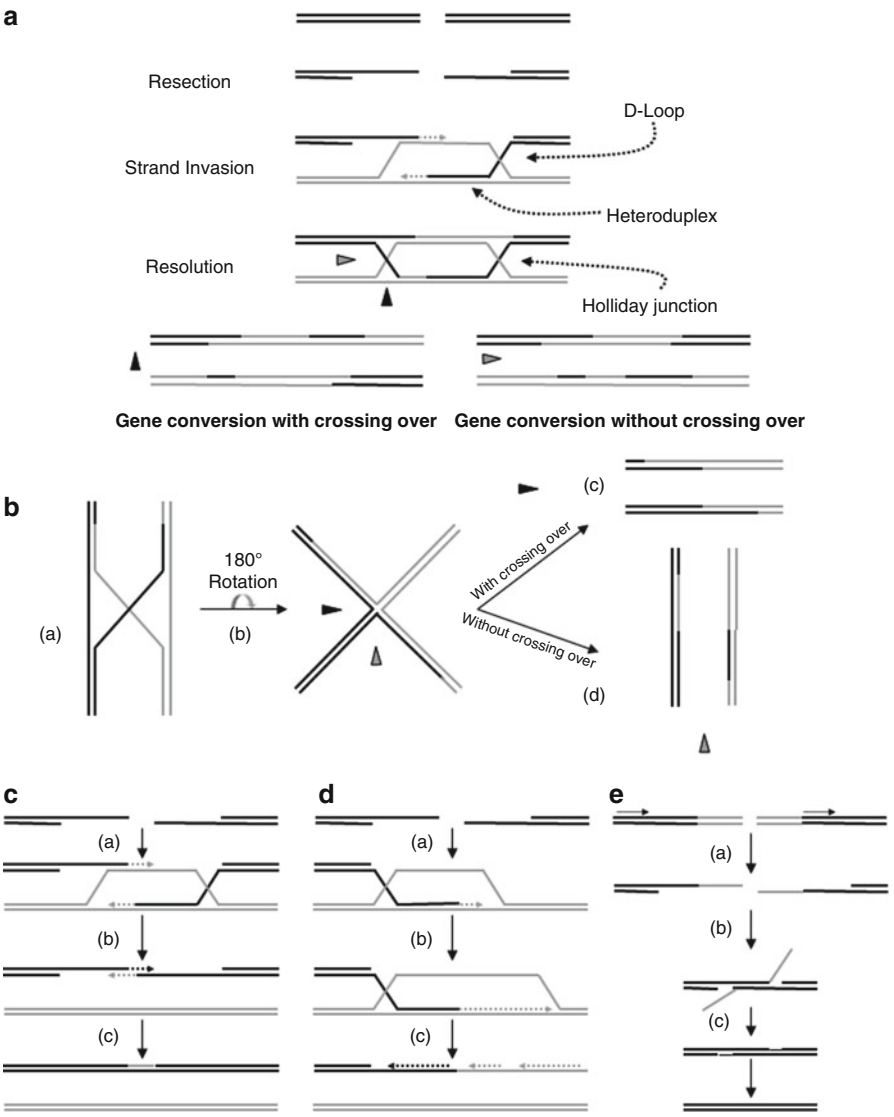


Fig. 4.3 Model for the repair of double-strand breaks. (**a**) The double-strand break repair model. Resection of ssDNA generates exiting single-stranded 3′ tails that can invade an homologous double-stranded DNA, forming a D-loop (displacement loop). Synthesis of DNA can be primed from the 3′ end of the invasive strand, causing displacement of the D-loop. This D-loop can hybridize with the second single-stranded 3′ end that is exiting, and DNA synthesis can then be initiated. DNA synthesis repairs the break and can therefore fill gaps. Cruciform junctions called Holliday junctions are formed and can migrate. Resolution of these junctions can occur in two different orientations (*black or grey triangles*) that do or do not lead to a crossover. (**b**) Resolution of Holliday junctions. After strand exchange (*a*), the cruciform junction can isomerize to 180° (*b*). The resolution (triangles) results, depending on the direction, either in the exchange of adjacent sequences (a crossover) (*c*) or no crossover (*d*). (**c**) Synthesis-dependent strand annealing (*SDSA*).

transfer of genetic information from the invaded molecule to the invading molecule and therefore to gene conversion. Strand exchange also creates cruciform structures called Holliday junctions. Holliday junctions were proposed in 1964 (Holliday 1964) to explain the formation of a crossover (Fig. 4.3b). Indeed, according the orientation of the resolution of the Holliday junction, the process will result or not in an exchange of adjacent sequences (crossover). The final result of HR is therefore gene conversion with or without a crossover (Fig. 4.3a).

Other models of HR exist (for a review, see Paques and Haber 1999), including synthesis-dependent strand annealing (SDSA) and break-induced replication (BIR). These two models do not involve resolution of the Holliday junction (Figs. 4.3c, d). However, all models are initiated by common steps, beginning with resection of ssDNA, followed by invasion and exchange of a homologous DNA strand, which is therefore a pivotal step in HR. These models are now considered more and more prevalent for mitotic and meiotic recombination without crossover. Furthermore, BIR seems to be the mechanism underlying the ALT system that allows telomere maintenance in the absence of telomerase (Dunham et al. 2000).

Finally, the last model, single-strand annealing (SSA), was initially described in mouse cells (Lin et al. 1984). This process can occur between two sequences in tandem and is initialized by a single-strand resection (Fig. 4.3e) but, contrary to the models described previously, this step is not followed by invasion of the DNA duplex and strand exchange. In fact, when the two sequences are in direct orientation, the ssDNA sequences revealed are complementary and can hybridize (which is different from the need for an invasion of the duplex DNA), forming a branched structure. The following concerns should be noted in relation to the SSA model: it is a non-conservative process leading inevitably to a deletion at the repair junction; it cannot occur between inverted repeat sequences because the strands revealed by the resection are not complementary but identical; and finally, if two breaks occur simultaneously in ectopic homologous sequences, SSA can generate translocations (Richardson and Jasin 2000).

4.2 Molecular Mechanisms of the Different Steps of HR

The cellular response to double-strand DNA breaks requires the coordination of a complex network that ensures maintenance of genome integrity, the regulation of which is carried out by PI3K kinases such as ATM, ATR and DNA-PK. These kinases lead to the activation of signaling pathways to enable effective management and repair the damage. HR occurs in successive steps that must be coordinated.

4.2.1 Recognition of Damage

The initial recognition of a double-strand break is accomplished by the MRN complex (MRE11, Rad50 and NBS1) acting in an activation loop with ATM kinase to allow for the rapid phosphorylation of the H2AX histones around the break.

This process leads to the recruitment of MDC1, which then stabilizes MRN at the break and thereby promotes signaling by ATM. The phosphorylation of MDC1 by ATM allows for the recruitment of ubiquitin ligases RNF8 and RNF168, which are necessary for chromatin remodeling. The relaxation of the chromatin then exposes epigenetic markers, such as methyl-histone H4, which leads to the subsequent recruitment of 53BP1 and RAP80-BRCA1 to the site of damage (for review, see Stewart 2009).

4.2.2 Chromatin Remodeling

The nucleosome is a barrier that blocks the access of enzymes involved in transcription, replication or repair of the DNA. Chromatin remodeling, by way of post-translational modifications of histones, occurs during the detection and repair of DNA damage and is dependent on two categories of enzymes. The first class of enzymes alters histone-histone and histone-DNA interactions by phosphorylation of serine and threonine residues; by acetylation, ubiquitination and sumoylation of lysine residues; and by methylation of lysine and arginine residues in the N- and C-terminal domains of relevant proteins (Kouzarides 2007; Marmorstein and Berger 2001). The second class of enzymes uses energy released by ATP hydrolysis to modify histone organization, thereby impeding the movement and repositioning of nucleosomes (Kouzarides 2007; Lusser and Kadonaga 2003; Strahl and Allis 2000).

Phosphorylation of serine-139 at the C-terminus of H2AX is a crucial chromatin modification in response to double-strand DNA break (Fillingham et al. 2006; Kinner et al. 2008; Rogakou et al. 1999; Sedelnikova et al. 2003). This phosphorylation, which begins at the break and extends beyond 2 Mb in higher eukaryotes (Rogakou et al. 1999), can be carried out by the PI3 kinases ATM, ATR or DNA-PK (Burma et al. 2001; Stiff et al. 2004; Ward and Chen 2001). It has been suggested that the phosphorylation of H2AX stabilizes the interactions among repair proteins such as 53BP1 (Ward et al. 2003), BRCA1 and NBS1 (Celeste et al. 2003; Paull and Gellert 2000) at the break site and facilitates the downstream accumulation of other proteins implicated in damage responses. Phosphorylation of γH2AX is then recognized by the BRCA1 carboxy-terminal (BRCT) domain of mediator of DNA damage checkpoint (MDC1) (Lee et al. 2005; Stucki et al. 2005). This complex regulates the level of phosphorylation of H2AX and is an early step that allows the downstream accumulation of key protein complexes such as MRE11-RAD50-NBS1 (Chapman and Jackson 2008; Melander et al. 2008) and ubiquitin ligase UBC13/RNF8 (Mailand et al. 2007; Wang and Elledge 2007). Phosphorylation of H2AX is also implicated in the recruitment of chromatin remodeling factors such as the NuA4 and Ino80 complexes (Downs et al. 2004; Morrison et al. 2004; van Attikum et al. 2004). The NuA4 complex consists of two types of chromatin remodelers, p400 and Tip60, a histone acetyltransferase (HAT). The Ino80 complex involved in DNA transcription and repair has various activities, including a 3′–5′ helicase,

nucleosome sliding and DNA-dependent ATPase (Osley et al. 2007). Ino80 has also been implicated in the generation of ssDNA by participating in the 3′–5′ resection of DNA. Indeed, Ino80 mutants demonstrate a defect in the recruitment of Mre11 nuclease and in the formation of ssDNA (van Attikum et al. 2004). It is important to note that these chromatin remodeling factors are necessary for the phosphorylation of H2AX (Park et al. 2006).

During repair, MDC1, via its interaction with H2AX, opposes the action of the phosphatases (Stucki et al. 2005). However, when repair is completed, histone γ-H2AX must be dephosphorylated for the repaired chromatin to regain the original structure. In mammals, the phosphatases PP2A (Chowdhury et al. 2005), PP4 (Chowdhury et al. 2008) and PP2C gamma (Kimura et al. 2006) participate in the dephosphorylation of histone H2AX.

In addition to phosphorylation, histones undergo other modifications during DNA repair. The methylation of histone H3 at lysine 79 (H3K79me) by Dot1 methyltransferase is accessible and recognized, after damage, by the Tudor domain of 53BP1 protein (Botuyan et al. 2006; Huyen et al. 2004). 53BP1 binds directly to histone H4 methylated at lysine 20 (H4K20me) (Botuyan et al. 2006).

In response to ionizing radiation, H2A and H2AX are both polyubiquitinated by the UBC13/RNF8 ubiquitin ligase complex, which occurs at the beginning of the ubiquitination process (Mailand et al. 2007), and the RNF168 ubiquitin ligase, which amplifies the phenomenon (Doil et al. 2009; Stewart et al. 2009). Rap80 acts as a mediator in recognizing polyubiquitin chains generated by RNF8 and in sending polyubiquitinated proteins to the site of the break. Rap80 is also able to interact with the Abraxas complex composed of BRCA1, Abraxas and Brcc36 (Kim et al. 2007; Sobhian et al. 2007; Wang and Elledge 2007). A recent study showed that the BRCA1-RAP80 complex, in S/G2, reduces HR by limiting the access of nucleases to the breakage site, thereby reducing resection (Coleman and Greenberg 2011). Thus, after irradiation, RNF8 plays a major role in the recruitment of RAP80, BRCA1, Abraxas, BRCC6 and 53BP1. RNF8 is also necessary for the accumulation of H2AX and MDC1 proteins at the repair sites at the earliest stages. The proposed hypothesis is that by ubiquitinating the histones, RNF8 restructures chromatin at the breakage site to facilitate the accumulation of factors that act later in repair, such as 53BP1 and BRCA1, near the lesion and therefore favor maintenance of the integrity of the genome.

Numerous studies have shown the importance of acetylation of histone lysine residues in DNA repair, facilitating the access of repair proteins to the site of damage. Changes in the dynamics of histone acetylation have been observed during HR (Tambini et al. 1997) and play a key role in post-HR viability. Histone acetylation is controlled by the balance between the activities of acetyltransferases (HATs) and deacetylases (HDACs). Transient acetylation of histones H3 and H4 has been found at the sites of double-strand breaks (van Attikum and Gasser 2005). In mammals, the Tip60 complex (comprising, among others, the Tip60 HAT, p400 and TRRAP) is responsible for acetylation. Tip60 is recruited to the break, where it binds and acetylates H4 with its cofactor TRRAP (Ikura et al. 2007). Recent data showed that Tip60 promotes the ubiquitination

of γ-H2AX (Ikura et al. 2007). In addition, acetylation by the Tip60-UBC13 complex causes detachment of H2AX from chromatin. It therefore seems that the sequential acetylation and ubiquitination of H2AX affect histone dynamics at the break site. It was recently shown that the changes in nucleosome and chromatin structure induced by p400 favor the ubiquitination of the chromatin and the accumulation of BRCA1 and 53BP1 at the breakage site (Xu et al. 2010). An interaction between Tip60 and MRN was also described recently (Chailleux et al. 2010).

4.2.3 Resection of ssDNA

An essential step in HR is the resection of ssDNA, leading to the production of a single-stranded 3′ tail (see Fig. 4.3). This step determines the choice of mechanism for double-strand break repair, and 53BP1, in binding to the ends, protects them from resection. Recent data show the essential role of BRCA1 in HR, in cooperation with CtIP, in removing 53BP1 from the ends and allowing the initiation of resection, thus promoting the process of HR (Bouwman et al. 2010; Bunting et al. 2010; Cao et al. 2009).

Recent studies in *S. cerevisiae* showed that the resection necessary for initiation of HR occurs in two steps. The first step of limited resection involves the Mre11, Rad50 and Xrs2 (MRX) protein complex and the Sae2 protein. These proteins remove several tens of nucleotides from the ends of the double-strand break. The single-stranded end formed is a preferred substrate for exonuclease Exo1 or for the Sgs1/endonuclease DNA2 complex. These enzymes perform a more processive resection, allowing the production of an outgoing single-stranded 3′ DNA to invade the intact homologous molecule necessary for repair by HR (Mimitou and Symington 2009a; Mimitou and Symington 2009b). In mammals, studies have suggested a similar two-step model; however, the mechanism has yet to be determined, particularly for the second step of the resection (for a review, see Huertas 2010). The role of the MRN complex (for Mre11, Rad50, Nbs1) in the initiation of resection during HR in mammalian cells is well established. The partner of the MRN complex, CtIP protein (a homologue of yeast Sae2), interacts with the MRN complex and stimulates the endonuclease activity of MRE11 *in vitro*. It was recently shown that deacetylation by SIRT6 has an important role in favoring resection mediated by CtIP (Kaidi et al. 2010). In agreement with the two-step model, the first initiation of resection is followed by the concerted action of nucleases and helicases. The involvement of exonuclease Exo1 in HR was shown by Bolderson et al. (Bolderson et al. 2010). A recent study also showed that recruitment of Exo1 depends on MRE11 and CtIP (Eid et al. 2010) and that the CtIP protein can inhibit the exonuclease activity of Exo1 *in vitro*. The involvement of BLM (a RecQ helicase mutated in Bloom Syndrome) in the process of resection was also suggested, but these data are

controversial. One study showed that BLM and Exo1 act in parallel but independent pathways (Gravel et al. 2008). Other studies showed a specific interaction between these two proteins. BLM could stimulate the exonuclease activity of Exo1 *in vitro* (Nimonkar et al. 2008), suggesting their cooperation during resection. This result is contradicted by a more recent *in vitro* biochemical study by the same group, which showed the involvement of the two independent machineries in the resection of ssDNA. Thus, on the one hand, BLM interacts directly with Exo1, independently of its helicase activity, to increase its affinity for DNA (DNA binding) but not its nuclease activity. On the other hand, a parallel mechanism involving BLM and DNA2 could be implemented. BLM unwinds the double helix of DNA to allow DNA2 to carry out the endonucleolytic cleavage of DNA. In this case, the RPA protein would be necessary to ensure proper directionality of the resection in the 5′–3′ direction because DNA2 possesses both 5′–3′ and 3′–5′ resection activity. However, it should be noted that an active role of BLM in the initiation of HR is inconsistent with the phenotype of cells from patients with Bloom Syndrome (deficient in BLM), which show traits of hyperrecombination rather than deficiency in HR.

4.2.4 Strand Invasion and Exchange

Single-stranded DNA from the resection stage is protected by the heterotrimer RPA and may be an element mediating cell signaling. In yeast, RAD52 displaces RPA to charge RAD51, the protein backbone of HR (an ortholog of RECA in bacteria), on the ssDNA to form the active filament. This function does not seem to be performed by RAD52 in mammals, given that RAD52 mutants have almost no phenotype (Rijkers et al. 1998). In mammals, it is the BRCA complex, consisting of the BRCA1 and BRCA2 proteins linked by PALB2, which is responsible for the localization of RAD51 at ssDNA (for review see Holthausen et al. 2010). BRCA1 is recruited to the damage site by the MRN complex or by RAP80. PALB2 (FANC-N), the key protein physically linking BRCA1 and BRCA2 in the complex, is essential for the formation of BRCA2 and RAD51 foci, but it is BRCA2 that interacts directly with RAD51 and ensures its localization to ssDNA coated with RPA.

Once the RAD51 filament has formed, the 3′ end of the strand invades an intact, homologous DNA duplex, effectively creating a 3′ DNA primer and allowing for polymerization and the formation of a displacement loop (see Fig. 4.3). Polymerization displaces the D-loop that can capture the liberated ssDNA, which is complementary (see Fig. 4.3), thereby creating two Holliday junctions. Polymerization fills the breach, and the Holliday junction can migrate, thus extending the heteroduplex region (see Fig. 4.3). Five paralogs of RAD51 exist (RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3), and each facilitates different steps; however, their exact molecular roles remain debated.

4.2.5 Resolution of Holliday Junctions

Holliday junction resolution is essential to ensure the proper segregation of chromosomes during cell division. Holliday junctions can be resolved by the action of structure-specific endonucleases (SSEs) or dissolved by the helicase activity of the BLM protein in cooperation with Top3α and RMI1 (Wu and Hickson 2003). Holliday junction dissolution leads to the occurrence of events without crossover, while their resolution can lead to events with or without crossover, depending on the direction of the cleavage performed by resolvases (see Fig. 4.3). The first resolvase activity in mammals, called ResA, was detected in 1990 and purified in 2001 (Elborough and West 1990), but it was not until 2008 that the gene was identified as GEN1. In eukaryotes, the following three large classes of SSEs are found: ERCC4 (with XPF and MUS81), UvrC (Slx1) and XPG (Gen1). MUS81 acts in a heterodimeric complex with its non-catalytic partner, the EME1 protein. This complex allows asymmetric cleavage of a mobile Holliday junction and is therefore not considered a "classic" resolvase. Mice lacking MUS81 are viable and fertile and show no significant defect in meiotic recombination (Dendouga et al. 2005), unlike *S. pombe*, in which MUS81 and EME1 are essential for proper meiosis (Boddy et al. 2001). In 2008, the first "classic" resolvases possessing ResA-type activity were identified in *S. cerevisiae* (Yen1) and in human cell extracts (GEN1) (Ip et al. 2008). GEN1/Yen1 performs the symmetrical cleavage of a static Holliday junction. Recent work identified another resolvase, SLX4, conserved in multicellular eukaryotes. This protein is a scaffold for the assembly of multiple SSEs and other proteins responsible for DNA repair and telomere maintenance. SLX4 interacts with the following three endonucleases: SLX1, ERCC1/XPF(ERCC4) and MUS81-EME1. The SLX1-SLX4 complex exhibits classic resolvase activity, cleaving junctions in a symmetrical fashion (Fekairi et al. 2009; Svendsen et al. 2009). Its absence leads to the defective repair of collapsed replication forks with increased levels of H2AX and of 53BP1 phosphorylation (Fekairi et al. 2009; Munoz et al. 2009). A mutation affecting both alleles of the SLX4 gene was recently found in two patients with clinical features characteristic of Fanconi anemia. It was found that these phenotypes can be rescued at the cellular level by wild-type SLX4 and that this mutation is responsible for a new sub-type of Fanconi anemia, called FANC-P (Kim et al. 2011).

4.2.6 Rad54

Rad54 is a key protein in HR that plays a major role in the maintenance of genome stability. This protein is a member of the SWI2/SNF2 complex in the SF2 family of helicases. Its ATPase activity is stimulated by DNA (Pollard and Peterson 1998). In interacting with its partners, Rad54 links early and late stages of HR. Indeed, in the early stages, Rad54 facilitates, by its interaction with Rad51, the formation of

the Rad51 filament along the ssDNA (Sigurdsson et al. 2002). By stimulating the activity of Rad51 with regard to DNA and by participating in chromatin remodeling, Rad54 promotes HR (Wolner and Peterson 2005; Zhang et al. 2007). In later stages, Rad54 promotes the migration of Holliday junctions by binding to the junction as a multimeric complex in an ATP-independent fashion. In turn, Rad51 stimulates the loading of Rad54 on Holliday junctions (Rossi and Mazin 2008). By direct interaction, Rad54 also stimulates the activity of the Mus81-Eme1(Mms4) complex that plays an important role in the resolution of HR intermediates by cleaving DNA (Mazina and Mazin 2008). Moreover, Rad54 participates in the recruitment of endonuclease Mus-81 to Holliday junctions (Matulova et al. 2009; Mazina and Mazin 2008). A paralog of Rad54, Rad54B, exhibits great similarities in sequence, genetic function and biochemical properties to Rad54 (Hiramoto et al. 1999). Rad54B augments the function of Rad54 in DNA repair in cells undergoing mitosis, but also during recombination in meiosis, through its interaction with the RecA homologue Dmc1 (Dresser et al. 1997; Klein 1997).

4.2.7 Some Aspects of Post-translational Regulation

A complex post-translational regulation of HR exists and it is for the moment not possible to present an unifying view. Thus we only will present few examples. Post-translational regulation acts through the phosphorylation, ubiquitination and sumoylation of central actors of HR. Moreover all the steps of HR (recognition, resection, strand exchange, resolution of the intermediates) can be affected by post-translational modification.

For instance, RAD51 can be phosphorylated by CHK1, which is activated in response to replication stress. ATM can also affect HR through the phosphorylation of BRCA1, and indirectly through the regulation of RAD51 by the kinase cABL. The cABL tyrosine kinase is activated by different genotoxic treatments among them ionizing radiation. Treatment with ionizing radiation induces cABL-dependent phosphorylation of RAD51 (Chen et al. 1999). The Tyr-phosphorylation of RAD51 leads to an apparent paradox. Indeed, on one hand, *in vitro* cABL-dependent phosphorylation on Tyr54 of Hs RAD51 or ScRAD51 inhibits their binding to single stranded DNA. It also inhibits the *in vitro* strand exchange activity of ScRAD51 (Yuan et al. 1998). These data suggest that phosphorylation of RAD51 inhibits its recombination activity and since RAD51 is a limiting factor for HR, it should thus inhibit the whole recombination process. On another hand, Tyr315-phosphorylation of RAD51 depends on both ATM and cABL and enhances the complex formation between RAD51 and RAD52, suggesting that it facilitates the HR complex assembly and potentially the HR process (Chen et al. 1999). Finally, high level of Tyr-phosphorylated RAD51 is observed in MEF cABL$^{-/-}$, but not in Tyr315, suggesting that other kinases can phosphorylate RAD51 in other site than Tyr315 (Chen et al. 1999).

Using a two-hybrid screen, RAD51 has been found to interact with UBC9/UBE21, a protein belonging to the family of ubiquitin-conjugating enzymes

(Kovalenko et al. 1996; Shen et al. 1996a). RAD51 has also been found to interact with UBL1 (Ubiquitin-Like protein 1) (Shen et al. 1996b) also called PIC1 (PML-interacting clone 1), GMP1 (GAP modifying protein 1), SUMO-1 (Small Ubiquitin-related Modifier 1), sentrin. In addition UBL1 and UBC9 interact in two-hybrid systems suggesting that UBC9 is able to conjugate RAD51 to UBL1 (Shen et al. 1996a). Overexpression of UBL1 down-regulates DSB-induced HR in CHO cells and reduces resistance to ionizing radiation in HT 1080 cells (Li et al. 2000). However, a mutant UBL1 that is incapable of being conjugated, retains the RAD51 binding and the HR repression activities. The author points out that the precise role of endogenous level of UBL1 in regulating HR remains unclear (Li et al. 2000).

Recent data established the implication of SUMO in post-translation modification of recombination proteins. In parallel to ubiquitination, these modifications play important roles in chromatin remodeling. With SUMO1 and SUMO2/3, SUMO E3 ligases PIAS1 and PIAS4 are required for the recruitment of BRCA1 and 53BP1 to the foci induced by ionizing radiations. (Galanty et al. 2009; Morris et al. 2009). Cells defective in E3 ligase PIAS4, which show a strong decrease in the recruitment of RNF168 and RNF8, thus of the ubiquitination at the damage site, show DNA repair defects associated with increased sensitivity to genotoxic stresses. These cells also show a decrease in RPA loaded on ssDNA, suggesting a defect in HR through reduced resection. Recently, it has been shown that the SUMOylation of proteins involved in Holliday junctions resolution, such as BLM, might affect its cell distribution (Galanty et al. 2009). It has also been shown that cells expressing a non-phosphorylable form of BLM, treated with hydroxyurea (which depletes the nucleotide pools generating thus a replication stress) show a defect in sister chromatids exchange due to a defect of localization of RAD51 to the repair foci (Eladad et al. 2005; Ouyang et al. 2009). They also showed that RAD51 directly interacts with SUMO and that its SUMOylation increases its binding to BLM.

4.3 HR Is at the Heart of the Genetic Stability-Diversity-Instability Equilibrium

Through the combination of its different products and because of its ability to repair DNA, HR contributes to the maintenance of genome stability or, conversely, to generation of genetic diversity or even genetic instability; it is therefore involved in many fundamental biological processes (see Fig. 4.2).

4.3.1 Repair of Double-Strand DNA Breaks; Consequences for Radioresistance and Meiosis

As suggested by models of HR, this process can participate in the repair of DNA double-strand DNA breaks. Accordingly, mutants defective in HR are sensitive to ionizing radiation (which generates double-strand breaks in DNA).

During meiotic division, HR can meet a double challenge. The first is assuring the balance of segregation of homologous chromosomes, and the second is ensuring the mixture of alleles to create genetic diversity. Indeed, during reductional division, the cell must segregate the two homologous chromosomes into the two different daughter cells. However, these chromosomes do not have a physical link that would distinguish them (in mitosis, the chromosomes are linked by centromeres). This physical link is assured in meiosis by HR, which generates Holliday junctions; in addition, the generation of crossovers allows rearrangement of alleles, ensuring genetic diversity. Meiotic recombination is initiated by a double-stranded break generated by the enzyme SPO11; the repair of this break essentially uses the same systems used for repair of breaks induced by ionizing radiation (for review see Buard and de Massy 2007). It should be noted that defects in HR lead to increased sensitivity to radiation and to sterility linked to problems encountered during meiotic division. It has been proposed that cells specifically adapted damage repair to meiosis (Kleckner 1996). However, it is noteworthy that some genes are specific to meiosis, such as DMC1, which is a paralog of RAD51.

4.3.2 HR and DNA Replication

HR plays a central role in the maintenance of genome stability, assisting DNA replication when the DNA has sustained damage that may be mutagenic or could block the progress of replication forks. Indeed, HR provides a mechanism to allow tolerance of some damage and ensure the restart of stalled replication forks. It should be noted that cells deficient in HR spontaneously show slower speed of replication fork elongation by a mechanism that remains to be characterized (Daboussi et al. 2008).

When replication copies a damaged DNA template, the cell has two options (Fig. 4.4); one is replication across the lesion, called trans-lesional synthesis, but this process is mutagenic, and the other is template switching by recombination with the undamaged sister chromatid, allowing replication to continue without risk of mutation. It should be noted that mutants deficient in HR exhibit a hypermutated phenotype, after genotoxic stress such as ultraviolet radiation (Lambert and Lopez 2002).

Replication forks regularly encounter obstacles that impede their progress (Hyrien 2000). These obstacles may be damage to DNA, DNA-protein interactions or regions of the genome that are difficult to replicate RNA/DNA hybrids regions. In this context, the following several points should be highlighted:

- Prolonged stalling of replication forks generates double-strand breaks that can be repaired by HR (Petermann et al. 2010; Saintigny et al. 2001).
- Chk1, which participates in the signaling of replication stress, phosphorylates and activates RAD51 (Sorensen et al. 2005).
- Generally, HR actively participates in restarting stalled forks and in reactivation of various collapsed forks (Fig. 4.4b).
- Many genes mutated in Fanconi anemia are directly implicated in HR. This metabolic pathway allows the cell to avoid succumbing to the damage and to

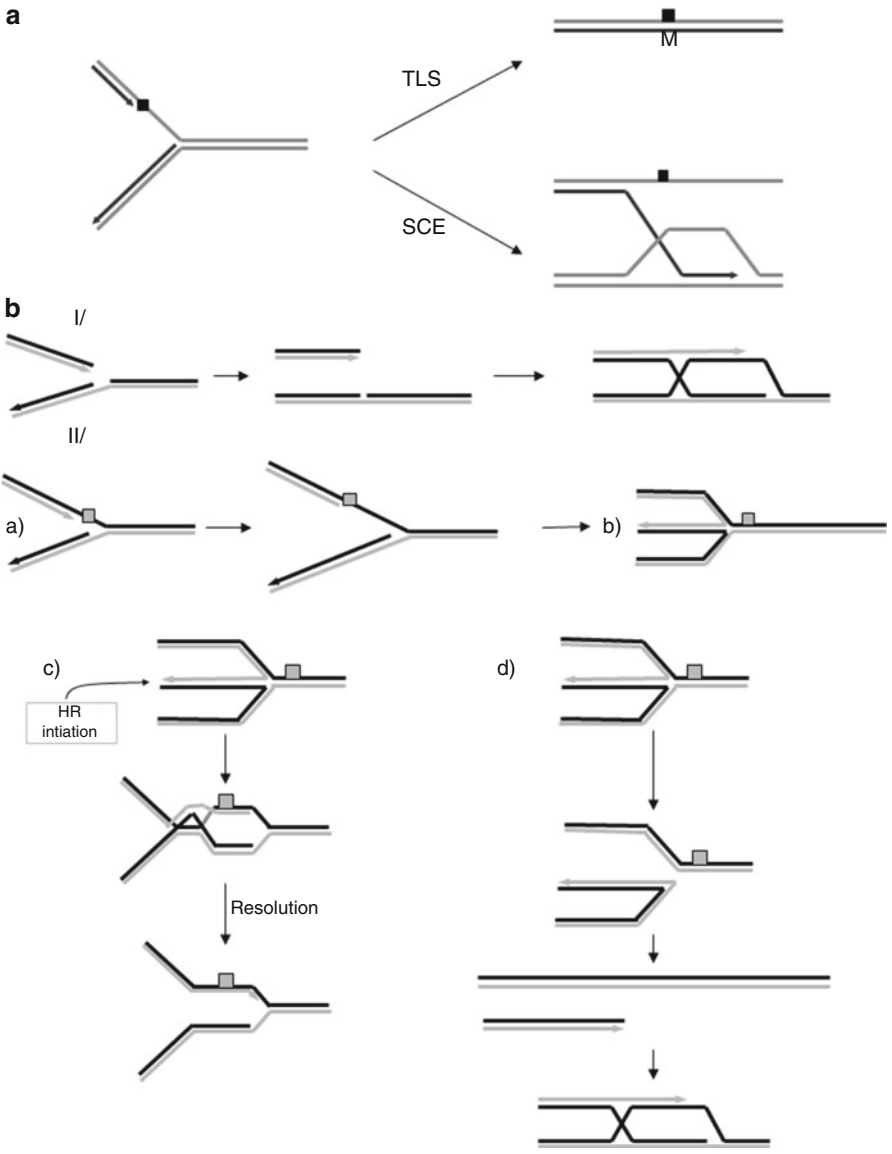


Fig. 4.4 Homologous recombination and replicative stress. (**a**) Tolerance of damage. When the replication fork reaches a lesion (*black square*) on the target strand, the cell has two options. Either replication passes through the lesion (translesion synthesis: *TLS*), although this process is mutagenic (*left panel*), or exchange with a sister chromatid (*SCE*: sister chromatid exchange) by recombination allows replication to continue. In these cases, the damage is not repaired but is bypassed or tolerated. (**b**) Restarting stalled replication forks. (*I*) When a replication fork reaches a single-stranded gap in the template, a collapsed fork results. Recombination with the sister chromatid allows the restarting of the fork via a process similar to BIR (see Fig. 4.3d). (*II*) When the fork reaches a blocking lesion (*a*), a reversion of this lesion can occur, forming a structure called "chicken foot" (*b*); (*c*) recombination can be initiated at the double end because the sequences are homologous to both sides of the cruciform structure (due to the reversion); (*d*) alternatively, the cruciform structure can be resolved (in the manner of a Holliday junction), resulting in a double end that can then initiate recombination with the sister chromatid and allowing replication to restart by a BIR-like mechanism

replicate its DNA after treatment with agents that create interstrand crosslinks in DNA, which block the progress of the replication fork. The Fanconi pathway allows activation of HR in these conditions of interstrand crosslinks (Nakanishi et al. 2011). The proteins of the Fanconi pathway fall into the following three groups that correspond to the three steps of their mode of action: core proteins (first group) that allow monoubiquitination of FANCD2 and FANCI (second group), and the group of proteins (third group) that activate HR. Among the proteins of the third group, FANCD1 corresponds to BRCA2, FANCJ corresponds to BACH1/BRIP1 and FANCN corresponds to PALB2.

This replication/recombination interface has clinical implications. Indeed, the inhibitor of PARP (poly-ADP-ribose polymerase) generates a replicative stress that leads to the formation of double-strand breaks in the S phase. Tumors deficient in HR, such as those characterized by a defect in BRCA1 or BRCA2, are extremely sensitive to these inhibitors (Bryant et al. 2005; Farmer et al. 2005), which are in clinical trials.

Finally, HR can be an alternative mechanism for telomere maintenance, called ALT, when telomerase is defective. In this case, the exchange of telomeric sequences by a BIR-like process (see Fig. 4.3d) allows for the maintenance of telomeres (Dunham et al. 2000).

4.3.3 Consequences of a Lack or Excess of HR on Genome Stability

Maintenance of genome stability requires a very precise balance of the regulation of HR. Indeed, recombination is a double-edged sword because either a defect in or an excess of HR can lead to genetic instability.

A lack of HR leads to genetic instability because cells deficient in HR accumulate DNA breaks that can lead to the loss of fragments of chromosomes or that can leave the field free to less faithful means of repair. In addition, HR protects against instability in mitosis, by mechanisms that have yet to be identified. Indeed, cells defective in HR show spontaneous mitotic defects such as supernumerary centromeres or interchromosomal bridges in metaphase (Bertrand et al. 2003; Daboussi et al. 2005; Deng 2002; Griffin et al. 2000; Kraakman-van der Zwet et al. 2002; Lahkim Bennani-Belhaj et al. 2010; Laulier et al. 2011b); these processes can lead to aneuploidy.

Excessive HR can also lead to genetic instability or to cellular toxicity. Indeed, intermediates of abortive or unresolved HR are toxic and/or responsible for genetic rearrangements (Gangloff et al. 2000). In addition, mammalian genomes contain numerous repeated sequences dispersed throughout the genome. In humans, more than a third of the genome consists of repeated sequences, and sequences arranged in tandem can occupy up to 10% of the human genome. The majority of repeated sequences are found outside of genes, while the rest code for multigene families. Crossovers between repeated sequences can lead to profound chromosomal rearrangements (Fig. 4.5a). Gene conversion between two heteroalleles can lead to the loss of heterozygosity (Fig. 4.5b). Finally, gene conversion between a

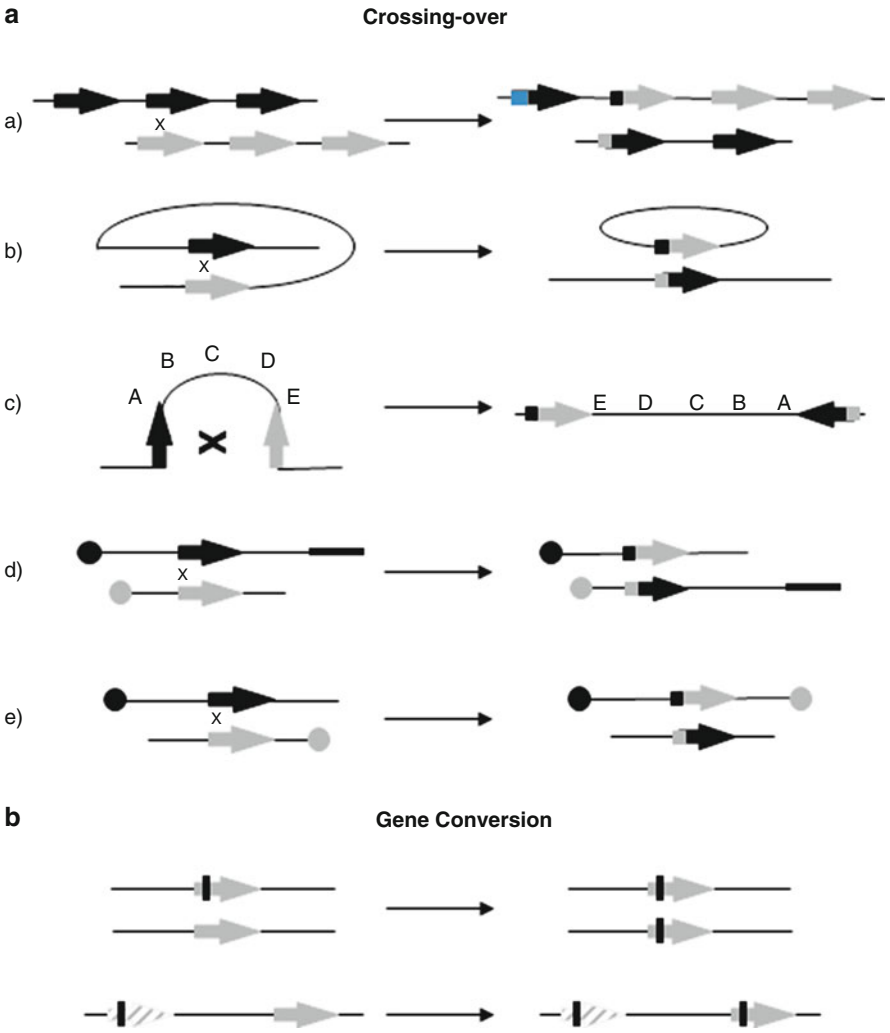


Fig. 4.5 Genetic instability associated with excess homologous recombination. (**a**) Chromosomal rearrangements resulting from crossovers (*CO*). (*a*) CO between repetitions carried on two chromosomes or following unequal sister chromatids exchange; the result is amplification of one molecule and deletion of the other. (*b*) Intra-chromosomal CO between two direct repeat sequences, resulting in excision of the intervening sequence. (*c*) CO between two inverted repeat sequences, resulting in inversion of the internal fragment. (*d*) and (*e*) Inter-chromosomal CO. According to the orientation of the sequences with respect to the centromers (blue or red circles), the process will generate a translocation (*d*) or one dicentric and one acentric chromosome (*e*). (**b**) Genetic modifications resulting from gene conversion between two heteroalleles, leading to a loss of heterozygosity. (**c**) Gene conversion between a pseudogene (hatched), which often contains nonsense mutations, and a gene, leading to the inactivation of the latter. Mutations are shown in black

pseudogene (bearing a stop codon) and a gene can lead to the inactivation of the latter (Fig. 4.5c). These cases have been described in various human diseases (Amor et al. 1988; Chen et al. 2007; Purandare and Patel 1997).

Precise control of HR is therefore essential for maintaining a balance and for avoiding a defect or excess of HR.

4.4 The Importance of Sister Chromatids in the Cell Cycle for the Maintenance of Genome Stability

Prolonged blockage of replication leads to double-strand breaks that can be addressed by HR or by non-homologous end joining (NHEJ) (Couedel et al. 2004; Mills et al. 2004; Saintigny et al. 2001). However, unlike double-strand breaks produced by enzymes or ionizing radiation, breaks produced by replication stops have only one double-strand end (see Fig. 4.4b). Ligation of the two ends of the double strand (which are therefore far apart) by NHEJ inevitably leads to chromosomal rearrangements. In contrast, during S phase, HR can use the intact sister chromatid. Because the two chromatids have identical sequences, the genetic impact is minimal. A potential risk is exchange between shifted sister chromatids, or crossovers with sequences carried by different chromosomes. However, the close proximity of the sister chromatids (particularly due to cohesins) favors the use of sister chromatids as templates for the repair of double-strand breaks (Johnson and Jasin 2000). In addition, in somatic cells, gene conversion without crossover is favored, limiting the risks associated with crossovers (Liskay et al. 1984).

In G1, HR, should it occur, does not have the sister chromatid available and would therefore use sequences carried by other chromosomes, jeopardizing genome stability (see Fig. 4.5). Maintenance of genome stability therefore requires the restriction of HR to the S and G2 phases of the cell cycle.

4.4.1 Regulation of HR During the Cell Cycle

The conclusion that HR is active from mid-S phase to G2 and that it is quelled at the end of G2 phase, particularly by AKT1 signaling (Xu et al. 2010), is supported by the following data:

1. Measurements of HR with episomic plasmids microinjected into nucleus (Wong and Capecchi 1987) or with intrachromosomal reporter substrates show that recombination occurs preferentially in S phase (Saintigny et al. 2007; Saleh-Gohari and Helleday 2004).
2. Vertebrate cells deficient in HR are more sensitive to radiation at the end of S phase and in G2 (Cheong et al. 1994; Hinz et al. 2005; Rothkamm et al. 2003; Saintigny et al. 2007; Takata et al. 1998).

3. RAD51 foci, which are induced after genotoxic stress and are considered as recombination centers, are not assembled in G1 (Aten et al. 2004; Saintigny et al. 2007; Yuan et al. 1998), although one study contradicts these conclusions (Kim et al. 2005).
4. Phosphorylation of BRCA2 by cyclin-dependent kinase causes disassembly of the RAD51-BRCA2 interaction only in G1 (Esashi et al. 2005).
5. MRE11 can promote single-strand resection in S phase and requires CDK activity.
6. The CtIP protein is phosphorylated by CDK kinases in S phase, promoting its interaction with BRCA1 and the MRN complex. The association with BRCA1 and MRE11 allows the recruitment of CtIP to damage sites and efficient resection of DNA (Chen et al. 2008; Limbo et al. 2007; Sartori et al. 2007).

4.5 Deregulation of HR and Tumor Predisposition

The presence of double-strand breaks and activation of the DNA damage response has been observed in pretumorous tissue, which has been interpreted as a result of endogenous replicative stress (Bartkova et al. 2005; Gorgoulis et al. 2005). Indeed, replication forks are regularly blocked by a variety of endogenous stresses such as DNA structures that are difficult to replicate, RNA/DNA hybrid regions or lesions resulting from cell metabolism (Hyrien 2000). The replication/recombination interface is therefore at the forefront for protecting against genetic instability and spontaneous tumor initiation.

In this context, it should be noted that among the 12 genes in which a germline mutation is associated with a familial predisposition to breast cancer, 11 are directly implicated in DRR (Erkko et al. 2008; Walsh and King 2007). The two genes most often mutated, BRCA1 and BRCA2, are two major players in HR (Moynahan et al. 1999; Moynahan et al. 2001). The eleventh gene, PTEN, indirectly affects HR, specifically via the expression of RAD51 (Shen et al. 2007). PTEN is also mutated in numerous other types of cancer. This overrepresentation of genes involved in DNA damage response and the communication between replication and recombination suggest the importance of these specific metabolic pathways in the etiology of breast cancer and raise the question of characteristics common to the causation of sporadic and hereditary breast cancer. This question is even more important because, of all breast cancers, hereditary cancers represent only 5–10% of cases and mutations in BRCA1 only 1–2% of cases. Several studies have reported hyperactivation of the oncogenic kinase AKT1 in 40–60% of sporadic breast cancers and in 40% of sporadic ovarian cancers (Plo et al. 2008; Sun et al. 2001; Yang et al. 2006). It must be noted that PTEN is an antagonist of AKT1. Recently, several studies have shown connections between AKT1 and DNA damage response (for a review, see Guirouilh-Barbat et al. 2010). In particular, overexpression of AKT1 induces the sequestration of BRCA1 and RAD51 in the cytoplasm, leading to inhibition

of HR (Plo et al. 2008; Plo and Lopez 2009). Taken together, these data underline the importance of HR in protection against breast cancer and reveal that the AKT1 signaling pathway is the missing link between hereditary and sporadic breast cancers.

Other examples of inhibition of HR exist in situations of predisposition to cancer. For example, Bc1-2 is an inhibitor of the intrinsic pathway of apoptosis induction, and its activation confers a predisposition for tumor formation. Bc1-2 was initially found to be overexpressed in B cell lymphoma with recurrent translocation t(14:18), but it is also overexpressed in numerous tumors. Overexpression of Bc1-2 leads to the relocalization of BRCA1 in endomembranes (endoplasmic reticulum, mitochondria), resulting in inhibition of HR (Laulier et al. 2011a).

Conversely, there are also situations that lead to the predisposition for tumors that show hyper-recombinogenic phenotypes. For example, Bloom syndrome shows a greatly elevated predisposition to spontaneous tumors in all tissues, even those not directly exposed; Bloom syndrome results from the inactivation of BLM protein, a member of the RecQ helicase family of proteins that plays an important role in the resolution of intermediates of HR and in the management of blocked replication forks. Cells from patients afflicted with Bloom syndrome show increased levels of exchange between sister chromatids and of hyper-recombination phenotypes (Chu and Hickson 2009).

The tumor-suppressing p53 gene is the most frequently mutated gene in all types of cancers. It has been shown that the p53 protein represses HR; tumors deficient in p53 show a hyper-recombination phenotype (for a review, see Bertrand et al. 2004).

The fusion oncogene BCR/ABL derives from the translocation of the cABL gene from chromosome 9 to the BCR gene locus on chromosome 22: Philadelphia chromosome t(9:22). This translocation is present in chronic myelogenous leukemia (CML) patient and in many acute lymphocytic leukemia patients. The BCR/ABL fusion produced proteins (p230, p210 or p185) exhibit constitutive tyrosine kinase activity. The resistance to DNA damages induced by therapeutic drugs of BCR/ABL tumors depends on the kinase activity of the fusion protein. The expression of BCR/ABL increases the intracellular level of RAD51 protein by different mechanisms (Slupianek et al. 2001). First, signaling from the BCR/ABL src homogy-3 (SH3) and SH2 domains stimulates RAD51 transcription *via* the activation of the signal transducer and activation transcription 5 (STAT5). The transcription of the paralogs RAD51B, RAD51D and XRCC2 is also stimulated whereas transcription of RAD51C and XRCC3 is decreased. Second, BCR/ABL inhibits the caspase 3 activation and thus RAD51 protein degradation. Indeed, BCR/ABL stimulates HR monitored between tandem repeat sequences. In addition, BCR/ABL interacts with RAD51 and results in high level of constitutive Tyr315-phosphorylation. This Tyr315-phosphorylation and RAD51-dependant HR seem to control resistance to Cisplatin and Mitomycin C (Slupianek et al. 2001). BCR/ABL expression inhibits DNA-PK activity, which is involved in non-homologous end-joining, a competitor pathway to HR for DNA double strand break repair (Deutsch et al. 2001). This

suggests that the regulation of the balance between HR and NHEJ can be modified by BCR/ABL.

These examples illustrate the necessity for the precise regulation of the balance in HR to avoid either a deficiency or an excess, both of which are harmful.

4.6 HR in Molecular Evolution: Concerted Evolution

HR is a force in the evolution of multi-genes families; crossovers leading to unequal exchanges between sister chromatids are particularly responsible for variation in the repetition of duplicated sequences (see Fig. 4.5).

During evolution, most of the duplicated sequences diverge; the genes of one species derived from a common ancestor are paralogs. Due to selective pressure, there are generally fewer divergences between homologous genes of two different species (orthologs), than between their respective paralogs.

However, in some families of repeated genes, the divergence between the duplicated units is less significant within one species than when compared to a different species, even one that is evolutionarily close. For example, primates possess two α globin genes. Despite an ancient duplication (approximately 300 million years ago), the α1 and α2 genes show high homology within each species; in humans, the coding sequences are identical and only a few differing nucleotides are found in intron II. In this case, the duplicated genes did not evolve independently but instead co-evolved; this phenomenon is called "concerted evolution" (Arnheim 1983; Liao 1999). The signature of concerted evolution is higher sequence conservation between paralogs than with the orthologs of another species.

Gene conversion is the driving force behind homogenization of duplicated sequences, and therefore of concerted evolution. Gene conversion allows the transfer (or deletion) of a mutation appearing in one duplicated unit to the second duplicated unit by a non-reciprocal mechanism of transfer of genetic information (see Figs. 4.1 and 4.2).

Concerted evolution is a universal biological phenomenon that occurs in bacteria, yeast, plants and animals, including mammals. Repeated ribosomal, histone, ubiquitin, α globin and even mitochondrial (though gene conversion has not been demonstrated in this organelle) genes are subjected to concerted evolution, as are non-coding sequences ranging from α satellite sequences to dispersed repeats such as LINE sequences.

Sequence heterologies thwart gene conversion and should therefore be barriers to concerted evolution. It has been suggested that introns, which can accumulate differences in sequence without affecting the function of the coded protein, can be protective barriers against HR between repeated sequences, thereby favoring the maintenance of the structural organization of the genome (Kourilsky 1983; Kricker et al. 1992). In this context, it is tempting to speculate that introns are an evolutionary force antagonistic to concerted evolution, directing evolution towards the divergence of repeated sequences.

4.7 Comments/Considerations Regarding Optimization of Targeted Gene Replacement

HR has applications in genome surgery (see other chapters of this work). However, in mammals, the effectiveness of gene targeting is modest. Different strategies have been developed to improve this process. Knowledge of the following molecular mechanisms of HR underlines the importance of taking different cellular parameters for the regulation of HR into account to optimize strategies for gene targeting:

- The structure of chromatin can be a barrier to the machinery of HR when accessing the target sequence and the DNA marker.
- The phase of the cell cycle is essential because HR occurs only in the S and G2 phases. Targeting is therefore only possible in proliferating cells; forcing HR in G0 or G1 remains yet to be demonstrated, in mammalian cells.
- The length of homology is a parameter that must be taken into account in the construction of the corrective substrate.
- A strategy using a crossover between the target DNA and the corrective DNA will be at a disadvantage because crossovers are suppressed in vegetative cells for the benefit of gene conversion without crossover (as opposed to meiotic cells). Other models such as SSA may be used.
- Strategies using a cut in the target can find themselves confronted by antagonistic repair pathways such as NHEJ (non-homologous end-joining), which is very effective in mammals and is active throughout the cell cycle (Guirouilh-Barbat et al. 2008; Rothkamm et al. 2003).
- Some metabolic pathways can more or less directly repress HR. This is the case in p53 (for a review, see Bertrand et al. 2004), in the activation of AKT1 (Plo et al. 2008; Plo and Lopez 2009) or in expression of members of the Bc1-2 family (Laulier et al. 2011a). Other cellular metabolic pathways affecting HR are likely to be discovered.

References

Amor M, Parker KL, Globerman H, New MI, White PC (1988) Mutation in the CYP21B gene (Ile-172—Asn) causes steroid 21 – hydroxylase deficiency. Proc Natl Acad Sci USA 85:1600–1604

Arnheim N (1983) Concerted evolution of multigene families. In: Koehn RK, Nei M (eds) Evolution of genes and proteins. Sinauer, Sunderland, pp 38–61

Aten JA, Stap J, Krawczyk PM, van Oven CH, Hoebe RA, Essers J, Kanaar R (2004) Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. Science 303:92–95

Bartkova J, Horejsi Z, Koed K, Kramer A, Tort F, Zieger K, Guldberg P, Sehested M, Nesland JM, Lukas C, Orntoft T, Lukas J, Bartek J (2005) DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature 434:864–870

Bertrand P, Lambert S, Joubert C, Lopez BS (2003) Overexpression of mammalian Rad51 does not stimulate tumorigenesis while a dominant-negative Rad51 affects centrosome fragmentation, ploidy and stimulates tumorigenesis, in p53-defective CHO cells. Oncogene 22:7587–7592

Bertrand P, Saintigny Y, Lopez BS (2004) p53's double life: transactivation-independent repression of homologous recombination. Trends Genet 20:235–243

Boddy MN, Gaillard PH, McDonald WH, Shanahan P, Yates JR 3rd, Russell P (2001) Mus81-Eme1 are essential components of a Holliday junction resolvase. Cell 107:537–548

Bolderson E, Tomimatsu N, Richard DJ, Boucher D, Kumar R, Pandita TK, Burma S, Khanna KK (2010) Phosphorylation of Exo1 modulates homologous recombination repair of DNA double-strand breaks. Nucleic Acids Res 38:1821–1831

Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J, Mer G (2006) Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell 127:1361–1373

Bouwman P, Aly A, Escandell JM, Pieterse M, Bartkova J, van der Gulden H, Hiddingh S, Thanasoula M, Kulkarni A, Yang Q, Haffty BG, Tommiska J, Blomqvist C, Drapkin R, Adams DJ, Nevanlinna H, Bartek J, Tarsounas M, Ganesan S, Jonkers J (2010) 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. Nat Struct Mol Biol 17:688–695

Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ, Helleday T (2005) Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature 434:913–917

Buard J, de Massy B (2007) Playing hide and seek with mammalian meiotic crossover hotspots. Trends Genet 23:301–309

Bunting SF, Callen E, Wong N, Chen HT, Polato F, Gunn A, Bothmer A, Feldhahn N, Fernandez-Capetillo O, Cao L, Xu X, Deng CX, Finkel T, Nussenzweig M, Stark JM, Nussenzweig A (2010) 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell 141:243–254

Burma S, Chen BP, Murphy M, Kurimasa A, Chen DJ (2001) ATM phosphorylates histone H2AX in response to DNA double-strand breaks. J Biol Chem 276:42462–42467

Cao L, Xu X, Bunting SF, Liu J, Wang RH, Cao LL, Wu JJ, Peng TN, Chen J, Nussenzweig A, Deng CX, Finkel T (2009) A selective requirement for 53BP1 in the biological response to genomic instability induced by Brca1 deficiency. Mol Cell 35:534–541

Celeste A, Fernandez-Capetillo O, Kruhlak MJ, Pilch DR, Staudt DW, Lee A, Bonner RF, Bonner WM, Nussenzweig A (2003) Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. Nat Cell Biol 5:675–679

Chailleux C, Tyteca S, Papin C, Boudsocq F, Puget N, Courilleau C, Grigoriev M, Canitrot Y, Trouche D (2010) Physical interaction between the histone acetyl transferase Tip60 and the DNA double-strand breaks sensor MRN complex. Biochem J 426:365–371

Chapman JR, Jackson SP (2008) Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage. EMBO Rep 9:795–801

Chen G, Yuan SS, Liu W, Xu Y, Trujillo K, Song B, Cong F, Goff SP, Wu Y, Arlinghaus R, Baltimore D, Gasser PJ, Park MS, Sung P, Lee EY (1999) Radiation-induced assembly of Rad51 and Rad52 recombination complex requires ATM and c-Abl. J Biol Chem 274:12748–12752

Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP (2007) Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8:762–775

Chen L, Nievera CJ, Lee AY, Wu X (2008) Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. J Biol Chem 283:7713–7720

Cheong N, Wang X, Wang Y, Iliakis G (1994) Loss of S-phase-dependent radioresistance in irs-1 cells exposed to X-rays. Mutat Res 314:77–85

Chowdhury D, Keogh MC, Ishii H, Peterson CL, Buratowski S, Lieberman J (2005) gamma-H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double-strand break repair. Mol Cell 20:801–809

Chowdhury D, Xu X, Zhong X, Ahmed F, Zhong J, Liao J, Dykxhoorn DM, Weinstock DM, Pfeifer GP, Lieberman J (2008) A PP4-phosphatase complex dephosphorylates gamma-H2AX generated during DNA replication. Mol Cell 31:33–46

Chu WK, Hickson ID (2009) RecQ helicases: multifunctional genome caretakers. Nat Rev Cancer 9:644–654

Coleman KA, Greenberg RA (2011) The BRCA1-RAP80 complex regulates DNA repair mechanism utilization by restricting end resection. J Biol Chem 286:13669–13680

Couedel C, Mills KD, Barchi M, Shen L, Olshen A, Johnson RD, Nussenzweig A, Essers J, Kanaar R, Li GC, Alt FW, Jasin M (2004) Collaboration of homologous recombination and nonhomologous end-joining factors for the survival and integrity of mice and cells. Genes Dev 18:1293–1304

Creighton HB, McClintock B (1931) A correlation of cytological and genetical crossing-over in *Zea mays*. Proc Natl Acad Sci USA 17:492–497

Daboussi F, Thacker J, Lopez BS (2005) Genetic interactions between RAD51 and its paralogues for centrosome fragmentation and ploidy control, independently of the sensitivity to genotoxic stresses. Oncogene 24:3691–3696

Daboussi F, Courbet S, Benhamou S, Kannouche P, Zdzienicka MZ, Debatisse M, Lopez BS (2008) A homologous recombination defect affects replication-fork progression in mammalian cells. J Cell Sci 121:162–166

Dendouga N, Gao H, Moechars D, Janicot M, Vialard J, McGowan CH (2005) Disruption of murine Mus81 increases genomic instability and DNA damage sensitivity but does not promote tumorigenesis. Mol Cell Biol 25:7569–7579

Deng CX (2002) Roles of BRCA1 in centrosome duplication. Oncogene 21:6222–6227

Deutsch E, Dugray A, AbdulKarim B, Marangoni E, Maggiorella L, Vaganay S, M'Kacher R, Rasy SD, Eschwege F, Vainchenker W, Turhan AG, Bourhis J (2001) BCR-ABL down-regulates the DNA repair protein DNA-PKcs. Blood 97:2084–2090

Doil C, Mailand N, Bekker-Jensen S, Menard P, Larsen DH, Pepperkok R, Ellenberg J, Panier S, Durocher D, Bartek J, Lukas J, Lukas C (2009) RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. Cell 136:435–446

Downs JA, Allard S, Jobin-Robitaille O, Javaheri A, Auger A, Bouchard N, Kron SJ, Jackson SP, Cote J (2004) Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. Mol Cell 16:979–990

Dresser ME, Ewing DJ, Conrad MN, Dominguez AM, Barstead R, Jiang H, Kodadek T (1997) DMC1 functions in a Saccharomyces cerevisiae meiotic pathway that is largely independent of the RAD51 pathway. Genetics 147:533–544

Dunham MA, Neumann AA, Fasching CL, Reddel RR (2000) Telomere maintenance by recombination in human cells. Nat Genet 26:447–450

Eid W, Steger M, El-Shemerly M, Ferretti LP, Pena-Diaz J, Konig C, Valtorta E, Sartori AA, Ferrari S (2010) DNA end resection by CtIP and exonuclease 1 prevents genomic instability. EMBO Rep 11:962–968

Eladad S, Ye TZ, Hu P, Leversha M, Beresten S, Matunis MJ, Ellis NA (2005) Intra-nuclear trafficking of the BLM helicase to DNA damage-induced foci is regulated by SUMO modification. Hum Mol Genet 14:1351–1365

Elborough KM, West SC (1990) Resolution of synthetic Holliday junctions in DNA by an endonuclease activity from calf thymus. EMBO J 9:2931–2936

Erkko H, Pylkas K, Karppinen SM, Winqvist R (2008) Germline alterations in the CLSPN gene in breast cancer families. Cancer Lett 261:93–97

Esashi F, Christ N, Gannon J, Liu Y, Hunt T, Jasin M, West SC (2005) CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. Nature 434:598–604

Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC, Ashworth A (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 434:917–921

Fekairi S, Scaglione S, Chahwan C, Taylor ER, Tissier A, Coulon S, Dong MQ, Ruse C, Yates JR, Russell P, Fuchs RP, McGowan CH, Gaillard PH (2009) Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. Cell 138:78–89

Fillingham J, Keogh MC, Krogan NJ (2006) GammaH2AX and its role in DNA double-strand break repair. Biochem Cell Biol 84:568–577

Galanty Y, Belotserkovskaya R, Coates J, Polo S, Miller KM, Jackson SP (2009) Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. Nature 462:935–939

Gangloff S, Soustelle C, Fabre F (2000) Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases. Nat Genet 25:192–194

Gorgoulis VG, Vassiliou LV, Karakaidos P, Zacharatos P, Kotsinas A, Liloglou T, Venere M, Ditullio RAJ, Kastrinakis NG, Levy B, Kletsas D, Yoneta A, Herlyn M, Kittas C, Halazonetis TD (2005) Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature 434:907–913

Gravel S, Chapman JR, Magill C, Jackson SP (2008) DNA helicases Sgs1 and BLM promote DNA double-strand break resection. Genes Dev 22:2767–2772

Griffin CS, Simpson PJ, Wilson CR, Thacker J (2000) Mammalian recombination-repair genes XRCC2 and XRCC3 promote correct chromosome segregation. Nat Cell Biol 2:757–761

Guirouilh-Barbat J, Huck S, Lopez BS (2008) S-phase progression stimulates both the mutagenic KU-independent pathway and mutagenic processing of KU-dependent intermediates, for nonhomologous end joining. Oncogene 27:1726–1736

Guirouilh-Barbat J, Wilhelm T, Lopez BS (2010) AKT1/BRCA1 in the control of homologous recombination and genetic stability: the missing link between hereditary and sporadic breast cancers. Oncotarget 1:691–699

Haldane JB (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Haldane JB, Lea DE (1947) A mathematical theory of chromosomal rearrangements. J Genet 48:1–10

Hinz JM, Yamada NA, Salazar EP, Tebbs RS, Thompson LH (2005) Influence of double-strand-break repair pathways on radiosensitivity throughout the cell cycle in CHO cells. DNA Repair (Amst) 4:782–792

Hiramoto T, Nakanishi T, Sumiyoshi T, Fukuda T, Matsuura S, Tauchi H, Komatsu K, Shibasaki Y, Inui H, Watatani M, Yasutomi M, Sumii K, Kajiyama G, Kamada N, Miyagawa K, Kamiya K (1999) Mutations of a novel human RAD54 homologue, RAD54B, in primary cancer. Oncogene 18:3422–3426

Holliday R (1964) A mechanism for gene conversion in fungi. Genet Res 5:282–306

Holthausen JT, Wyman C, Kanaar R (2010) Regulation of DNA strand exchange in homologous recombination. DNA Repair (Amst) 9:1264–1272

Huertas P (2010) DNA resection in eukaryotes: deciding how to fix the break. Nat Struct Mol Biol 17:11–16

Huyen Y, Zgheib O, Ditullio RA Jr, Gorgoulis VG, Zacharatos P, Petty TJ, Sheston EA, Mellert HS, Stavridi ES, Halazonetis TD (2004) Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature 432:406–411

Hyrien O (2000) Mechanisms and consequences of replication fork arrest. Biochimie 82:5–17

Ikura T, Tashiro S, Kakino A, Shima H, Jacob N, Amunugama R, Yoder K, Izumi S, Kuraoka I, Tanaka K, Kimura H, Ikura M, Nishikubo S, Ito T, Muto A, Miyagawa K, Takeda S, Fishel R, Igarashi K, Kamiya K (2007) DNA damage-dependent acetylation and ubiquitination of H2AX enhances chromatin dynamics. Mol Cell Biol 27:7028–7040

Ip SC, Rass U, Blanco MG, Flynn HR, Skehel JM, West SC (2008) Identification of Holliday junction resolvases from humans and yeast. Nature 456:357–361

Johnson RD, Jasin M (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. EMBO J 19:3398–3407

Kaidi A, Weinert BT, Choudhary C, Jackson SP (2010) Human SIRT6 promotes DNA end resection through CtIP deacetylation. Science 329:1348–1353

Kim JS, Krasieva TB, Kurumizaka H, Chen DJ, Taylor AM, Yokomori K (2005) Independent and sequential recruitment of NHEJ and HR factors to DNA damage sites in mammalian cells. J Cell Biol 170:341–347

Kim H, Chen J, Yu X (2007) Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. Science 316:1202–1205

Kim Y, Lach FP, Desetty R, Hanenberg H, Auerbach AD, Smogorzewska A (2011) Mutations of the SLX4 gene in Fanconi anemia. Nat Genet 43:142–146

Kimura H, Takizawa N, Allemand E, Hori T, Iborra FJ, Nozaki N, Muraki M, Hagiwara M, Krainer AR, Fukagawa T, Okawa K (2006) A novel histone exchange factor, protein phosphatase 2Cgamma, mediates the exchange and dephosphorylation of H2A-H2B. J Cell Biol 175:389–400

Kinner A, Wu W, Staudt C, Iliakis G (2008) Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. Nucleic Acids Res 36:5678–5694

Kleckner N (1996) Meiosis: how could it work? Proc Natl Acad Sci USA 93:8167–8174

Klein HL (1997) RDH54, a RAD54 homologue in Saccharomyces cerevisiae, is required for mitotic diploid-specific recombination and repair and for meiosis. Genetics 147:1533–1543

Kourilsky P (1983) Genetic exchanges between partially homologous nucleotide sequences: possible implications for multigene families. Biochimie 65:85–93

Kouzarides T (2007) Chromatin modifications and their function. Cell 128:693–705

Kovalenko OV, Plug AW, Haaf T, Gonda DK, Ashley T, Ward DC, Radding CM, Golub EI (1996) Mammalian ubiquitin-conjugating enzyme Ubc9 interacts with Rad51 recombination protein and localizes in synaptonemal complexes. Proc Natl Acad Sci USA 93:2958–2963

Kraakman-van der Zwet M, Overkamp WJ, van Lange RE, Essers J, van Duijn-Goedhart A, Wiggers I, Swaminathan S, van Buul PP, Errami A, Tan RT, Jaspers NG, Sharan SK, Kanaar R, Zdzienicka MZ (2002) Brca2 (XRCC11) deficiency results in radioresistant DNA synthesis and a higher frequency of spontaneous deletions. Mol Cell Biol 22:669–679

Kricker MC, Drake JW, Radman M (1992) Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. Proc Natl Acad Sci USA 89:1075–1079

Lahkim Bennani-Belhaj K, Rouzeau S, Buhagiar-Labarchede G, Chabosseau P, Onclercq-Delic R, Bayart E, Cordelieres F, Couturier J, Amor-Gueret M (2010) The Bloom syndrome protein limits the lethality associated with RAD51 deficiency. Mol Cancer Res 8:385–394

Lambert S, Lopez BS (2002) Inactivation of the RAD51 recombination pathway stimulates UV-induced mutagenesis in mammalian cells. Oncogene 21:4065–4069

Laulier C, Barascu A, Guirouilh-Barbat J, Pennarun G, Le Chalony C, Chevalier F, Palierne G, Bertrand P, Verbavatz JM, Lopez BS (2011a) Bcl-2 inhibits nuclear homologous recombination by localizing BRCA1 to the endomembranes. Cancer Res 71:3590–3602

Laulier C, Cheng A, Stark JM (2011b) The relative efficiency of homology-directed repair has distinct effects on proper anaphase chromosome separation. Nucleic Acids Res 39:5935–5944

Lederberg J, Tatum EL (1946) Gene recombination in *Escherichia coli*. Nature 158:558

Lee MS, Edwards RA, Thede GL, Glover JN (2005) Structure of the BRCT repeat domain of MDC1 and its specificity for the free COOH-terminal end of the gamma-H2AX histone tail. J Biol Chem 280:32053–32056

Li W, Hesabi B, Babbo A, Pacione C, Liu J, Chen DJ, Nickoloff JA, Shen Z (2000) Regulation of double-strand break-induced mammalian homologous recombination by UBL1, a RAD51-interacting protein. Nucleic Acids Res 28:1145–1153

Liao D (1999) Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64:24–30

Limbo O, Chahwan C, Yamada Y, de Bruin RA, Wittenberg C, Russell P (2007) Ctp1 is a cell-cycle-regulated protein that functions with Mre11 complex to control double-strand break repair by homologous recombination. Mol Cell 28:134–146

Lin FL, Sperle K, Sternberg N (1984) Homologous recombination in mouse L cells. Cold Spring Harb Symp Quant Biol 49:139–149

Liskay RM, Stachelek JL (1983) Evidence for intrachromosomal gene conversion in cultured mouse cells. Cell 35:157–165

Liskay RM, Stachelek JL, Letsou A (1984) Homologous recombination between repeated chromosomal sequences in mouse cells. Cold Spring Harb Symp Quant Biol 49:183–189

Liskay RM, Letsou A, Stachelek JL (1987) Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics 115:161–167

Lopez BS, Corteggiani E, Bertrand-Mercat P, Coppey J (1992) Directional recombination is initiated at a double strand break in human nuclear extracts. Nucleic Acids Res 20:501–506
Lusser A, Kadonaga JT (2003) Chromatin remodeling by ATP-dependent molecular machines. Bioessays 25:1192–1200
Mailand N, Bekker-Jensen S, Faustrup H, Melander F, Bartek J, Lukas C, Lukas J (2007) RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell 131:887–900
Marmorstein R, Berger SL (2001) Structure and function of bromodomains in chromatin-regulating complexes. Gene 272:1–9
Matulova P, Marini V, Burgess RC, Sisakova A, Kwon Y, Rothstein R, Sung P, Krejci L (2009) Cooperativity of Mus81.Mms4 with Rad54 in the resolution of recombination and replication intermediates. J Biol Chem 284:7733–7745
Mazina OM, Mazin AV (2008) Human Rad54 protein stimulates human Mus81-Eme1 endonuclease. Proc Natl Acad Sci USA 105:18249–18254
Melander F, Bekker-Jensen S, Falck J, Bartek J, Mailand N, Lukas J (2008) Phosphorylation of SDT repeats in the MDC1 N terminus triggers retention of NBS1 at the DNA damage-modified chromatin. J Cell Biol 181:213–226
Mills KD, Ferguson DO, Essers J, Eckersdorff M, Kanaar R, Alt FW (2004) Rad54 and DNA Ligase IV cooperate to maintain mammalian chromatid stability. Genes Dev 18:1283–1292
Mimitou EP, Symington LS (2009a) Nucleases and helicases take center stage in homologous recombination. Trends Biochem Sci 34:264–272
Mimitou EP, Symington LS (2009b) DNA end resection: many nucleases make light work. DNA Repair (Amst) 8:983–995
Morris JR, Boutell C, Keppler M, Densham R, Weekes D, Alamshah A, Butler L, Galanty Y, Pangon L, Kiuchi T, Ng T, Solomon E (2009) The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. Nature 462:886–890
Morrison AJ, Highland J, Krogan NJ, Arbel-Eden A, Greenblatt JF, Haber JE, Shen X (2004) INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. Cell 119:767–775
Moynahan ME, Chiu JW, Koller BH, Jasin M (1999) Brca1 controls homology-directed DNA repair. Mol Cell 4:511–518
Moynahan ME, Pierce AJ, Jasin M (2001) BRCA2 is required for homology-directed repair of chromosomal breaks. Mol Cell 7:263–272
Munoz IM, Hain K, Declais AC, Gardiner M, Toh GW, Sanchez-Pulido L, Heuckmann JM, Toth R, Macartney T, Eppink B, Kanaar R, Ponting CP, Lilley DM, Rouse J (2009) Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. Mol Cell 35:116–127
Nakanishi K, Cavallo F, Perrouault L, Giovannangeli C, Moynahan ME, Barchi M, Brunet E, Jasin M (2011) Homology-directed Fanconi anemia pathway cross-link repair is dependent on DNA replication. Nat Struct Mol Biol 18:500–503
Nimonkar AV, Ozsoy AZ, Genschel J, Modrich P, Kowalczykowski SC (2008) Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. Proc Natl Acad Sci USA 105:16906–16911
Osley MA, Tsukuda T, Nickoloff JA (2007) ATP-dependent chromatin remodeling factors and DNA damage repair. Mutat Res 618:65–80
Ouyang KJ, Woo LL, Zhu J, Huo D, Matunis MJ, Ellis NA (2009) SUMO modification regulates BLM and RAD51 interaction at damaged replication forks. PLoS Biol 7:e1000252
Paques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 63:349–404
Park JH, Park EJ, Lee HS, Kim SJ, Hur SK, Imbalzano AN, Kwon J (2006) Mammalian SWI/SNF complexes facilitate DNA double-strand break repair by promoting gamma-H2AX induction. EMBO J 25:3986–3997
Paull TT, Gellert M (2000) A mechanistic basis for Mre11-directed DNA joining at microhomologies. Proc Natl Acad Sci USA 97:6409–6414

Petermann E, Orta ML, Issaeva N, Schultz N, Helleday T (2010) Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. Mol Cell 37:492–502

Plo I, Lopez B (2009) AKT1 represses gene conversion induced by different genotoxic stresses and induces supernumerary centrosomes and aneuploidy in hamster ovary cells. Oncogene 28:2231–2237

Plo I, Laulier C, Gauthier L, Lebrun F, Calvo F, Lopez BS (2008) AKT1 inhibits homologous recombination by inducing cytoplasmic retention of BRCA1 and RAD51. Cancer Res 68:9404–9412

Pollard KJ, Peterson CL (1998) Chromatin remodeling: a marriage between two families? Bioessays 20:771–780

Purandare SM, Patel PI (1997) Recombination hot spots and human disease. Genome Res 7:773–786

Richardson C, Jasin M (2000) Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells. Mol Cell Biol 20:9068–9075

Rijkers T, van den Ouweland J, Morolli B, Rolink AG, Baarends WM, van Sloun PP, Lohman PH, Pastink A (1998) Targeted inactivation of mouse RAD52 reduces homologous recombination but not resistance to ionizing radiation. Mol Cell Biol 18:6423–6429

Rogakou EP, Boon C, Redon C, Bonner WM (1999) Megabase chromatin domains involved in DNA double-strand breaks in vivo. J Cell Biol 146:905–916

Rossi MJ, Mazin AV (2008) Rad51 protein stimulates the branch migration activity of Rad54 protein. J Biol Chem 283:24698–24706

Rothkamm K, Kruger I, Thompson LH, Lobrich M (2003) Pathways of DNA double-strand break repair during the mammalian cell cycle. Mol Cell Biol 23:5706–5715

Rubnitz J, Subramani S (1984) The minimum amount of homology required for homologous recombination in mammalian cells. Mol Cell Biol 4:2253–2258

Saintigny Y, Delacote F, Vares G, Petitot F, Lambert S, Averbeck D, Lopez BS (2001) Characterization of homologous recombination induced by replication inhibition in mammalian cells. EMBO J 20:3861–3870

Saintigny Y, Delacote F, Boucher D, Averbeck D, Lopez BS (2007) XRCC4 in G1 suppresses homologous recombination in S/G2, in G1 checkpoint-defective cells. Oncogene 26:2769–2780

Saleh-Gohari N, Helleday T (2004) Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. Nucleic Acids Res 32:3683–3688

Sartori AA, Lukas C, Coates J, Mistrik M, Fu S, Bartek J, Baer R, Lukas J, Jackson SP (2007) Human CtIP promotes DNA end resection. Nature 450:509–514

Sedelnikova OA, Pilch DR, Redon C, Bonner WM (2003) Histone H2AX in DNA damage and repair. Cancer Biol Ther 2:233–235

Shen Z, Pardington-Purtymun PE, Comeaux JC, Moyzis RK, Chen DJ (1996a) Associations of UBE2I with RAD52, UBL1, p53, and RAD51 proteins in a yeast two-hybrid system. Genomics 37:183–186

Shen Z, Pardington-Purtymun PE, Comeaux JC, Moyzis RK, Chen DJ (1996b) UBL1, a human ubiquitin-like protein associating with human RAD51/RAD52 proteins. Genomics 36:271–279

Shen WH, Balajee AS, Wang J, Wu H, Eng C, Pandolfi PP, Yin Y (2007) Essential role for nuclear PTEN in maintaining chromosomal integrity. Cell 128:157–170

Sigurdsson S, van Komen S, Petukhova G, Sung P (2002) Homologous DNA pairing by human recombination factors Rad51 and Rad54. J Biol Chem 277:42790–42794

Slupianek A, Schmutte C, Tombline G, Nieborowska-Skorska M, Hoser G, Nowicki MO, Pierce AJ, Fishel R, Skorski T (2001) BCR/ABL regulates mammalian RecA homologs, resulting in drug resistance. Mol Cell 8:795–806

Sobhian B, Shao G, Lilli DR, Culhane AC, Moreau LA, Xia B, Livingston DM, Greenberg RA (2007) RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. Science 316:1198–1202

Sorensen CS, Hansen LT, Dziegielewski J, Syljuasen RG, Lundin C, Bartek J, Helleday T (2005) The cell-cycle checkpoint kinase Chk1 is required for mammalian homologous recombination repair. Nat Cell Biol 7:195–201

Stewart GS (2009) Solving the RIDDLE of 53BP1 recruitment to sites of damage. Cell Cycle 8:1532–1538

Stewart GS, Panier S, Townsend K, Al-Hakim AK, Kolas NK, Miller ES, Nakada S, Ylanko J, Olivarius S, Mendez M, Oldreive C, Wildenhain J, Tagliaferro A, Pelletier L, Taubenheim N, Durandy A, Byrd PJ, Stankovic T, Taylor AM, Durocher D (2009) The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. Cell 136:420–434

Stiff T, O'Driscoll M, Rief N, Iwabuchi K, Lobrich M, Jeggo PA (2004) ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation. Cancer Res 64:2390–2396

Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403:41–45

Stucki M, Clapperton JA, Mohammad D, Yaffe MB, Smerdon SJ, Jackson SP (2005) MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell 123:1213–1226

Sun M, Wang G, Paciga JE, Feldman RI, Yuan ZQ, Ma XL, Shelley SA, Jove R, Tsichlis PN, Nicosia SV, Cheng JQ (2001) AKT1/PKBalpha kinase is frequently elevated in human cancers and its constitutive activation is required for oncogenic transformation in NIH3T3 cells. Am J Pathol 159:431–437

Svendsen JM, Smogorzewska A, Sowa ME, O'Connell BC, Gygi SP, Elledge SJ, Harper JW (2009) Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. Cell 138:63–77

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25–35

Takata M, Sasaki MS, Sonoda E, Morrison C, Hashimoto M, Utsumi H, Yamaguchi-Iwai Y, Shinohara A, Takeda S (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J 17:5497–5508

Tambini CE, George AM, Rommens JM, Tsui LC, Scherer SW, Thacker J (1997) The XRCC2 DNA repair gene: identification of a positional candidate. Genomics 41:84–92

van Attikum H, Gasser SM (2005) ATP-dependent chromatin remodeling and DNA double-strand break repair. Cell Cycle 4:1011–1014

van Attikum H, Fritsch O, Hohn B, Gasser SM (2004) Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. Cell 119:777–788

Walsh T, King MC (2007) Ten genes for inherited breast cancer. Cancer Cell 11:103–105

Wang B, Elledge SJ (2007) Ubc13/Rnf8 ubiquitin ligases control foci formation of the Rap80/Abraxas/Brca1/Brcc36 complex in response to DNA damage. Proc Natl Acad Sci USA 104:20759–20763

Ward IM, Chen J (2001) Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. J Biol Chem 276:47759–47762

Ward IM, Minn K, Jorda KG, Chen J (2003) Accumulation of checkpoint protein 53BP1 at DNA breaks involves its binding to phosphorylated histone H2AX. J Biol Chem 278:19579–19582

Watson JD, Crick FH (1953) Genetical implications of the structure of deoxyribonucleic acid. Nature 171:964–967

Wolner B, Peterson CL (2005) ATP-dependent and ATP-independent roles for the Rad54 chromatin remodeling enzyme during recombinational repair of a DNA double strand break. J Biol Chem 280:10855–10860

Wong EA, Capecchi MR (1987) Homologous recombination between coinjected DNA sequences peaks in early to mid-S phase. Mol Cell Biol 7:2294–2295

Wu L, Hickson ID (2003) The Bloom's syndrome helicase suppresses crossing over during homologous recombination. Nature 426:870–874

Xu N, Hegarat N, Black EJ, Scott MT, Hochegger H, Gillespie DA (2010) Akt/PKB suppresses DNA damage processing and checkpoint activation in late G2. J Cell Biol 190:297–305

Yang H, Wen YY, Zhao R, Lin YL, Fournier K, Yang HY, Qiu Y, Diaz J, Laronga C, Lee MH (2006) DNA damage-induced protein 14-3-3 sigma inhibits protein kinase B/Akt activation and suppresses Akt-activated cancer. Cancer Res 66:3096–3105

Yuan ZM, Huang Y, Ishiko T, Nakada S, Utsugisawa T, Kharbanda S, Wang R, Sung P, Shinohara A, Weichselbaum R, Kufe D (1998) Regulation of Rad51 function by c-Abl in response to DNA damage. J Biol Chem 273:3799–3802

Zhang Z, Fan HY, Goldman JA, Kingston RE (2007) Homology-driven chromatin remodeling by human RAD54. Nat Struct Mol Biol 14:397–405

Chapter 5
Engineered Zinc Finger Nucleases for Targeted Genome Editing

Cherie L. Ramirez and J. Keith Joung

Abstract Zinc finger nucleases (ZFNs) are artificial proteins consisting of engineered zinc finger proteins fused to the FokI endonuclease domain. These nucleases bind to specific DNA recognition sites and introduce double-strand breaks (DSBs). Repair of these DSBs by normal cellular processes can be exploited to either disrupt genes or significantly increase the frequency of homologous recombination with a user-defined repair template. Several platforms have been developed that enable engineering of zinc finger proteins that bind with high affinity and specificity to 9–18 bp target DNA sequences. ZFNs have already been used in basic research, most prominently with the development of methods to efficiently modify various model organisms and cell lines.

Keywords DNA repair • FokI • Gene targeting • Protein engineering • Zinc finger nucleases • Engineered zinc fingers

5.1 Introduction

Since its conception in the 1970s, the idea of manipulating genomic DNA to cure genetic diseases has remained a challenging goal (Wigler et al. 1979). Despite the multitude of advances recently made to improve delivery methods and minimize unwanted side-effects, many gene therapy technologies are still moving cautiously toward clinical applications (Jensen et al. 2011). The finding by Jasin and colleagues that double-strand breaks (DSBs) introduced at specific genomic locations can significantly stimulate gene repair by homologous recombination with a corrective

C.L. Ramirez • J.K. Joung (✉)
Molecular Pathology Unit, Massachusetts General Hospital,
149 13th St. – 6th floor, Charlestown, MA 02129, USA
e-mail: cramirez@fas.harvard.edu; joung@partners.org

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_5,

DNA template (Rouet et al. 1994) has motivated efforts to utilize this particular strategy for gene correction. Zinc finger nuclease (ZFN) technology enables the creation of targeted DSBs in a wide variety of different cell types and organisms at essentially any desired gene of interest. Gene modification mediated by ZFNs may have important therapeutic applications and has already provided an important tool for many key areas of basic research, including the generation of methods to create genetic alterations in novel model organisms. Here we review methods for engineering ZFNs and the various ways in which these nucleases have been used for genome manipulation.

5.2 Zinc Fingers: Structure and Function

Proteins with Cys_2-His_2 zinc finger domains (hereafter simply zinc fingers) comprise the most abundant class of transcription factors in the human genome, with as many as 900 of ~20,000 predicted proteins bearing the motif (Urnov 2002; Tupler et al. 2001). The consensus sequence of a zinc finger ((F/Y)-X-C-X_{2-4}-C-X-X-X-(F/Y)-X-X-X-X-X-γ-X-X-H-X_{3-5}-H, where X is any amino acid and γ is a hydrophobic residue) was first observed in the sequence of the *Xenopus* transcription factor TFIIIA (Miller et al. 1985). Subsequent studies demonstrated that sequences matching this consensus adopt a ββα fold that coordinates a zinc atom using the conserved cysteine and histidine pairs (Lee et al. 1989; Parraga et al. 1988).

The concept of engineering zinc finger domains with new DNA-binding specificities originated from analyzing the co-crystal structure of DNA-bound zinc fingers from the Zif268/Egr1 protein, a transcriptional regulator that determines thymus size in mouse (Fig. 5.1a) (Pavletich and Pabo 1991; Bettini et al. 2002). The DNA binding domain of Zif268 consists of three tandem zinc fingers that wrap around B-DNA. Primary contacts are made between residues -1, +2, +3, and +6 of each zinc finger's α-helix (or "recognition helix") and the edges of bases in the major groove from the G-rich strand of the Zif268 binding site. Partially overlapping 3–4 bp "subsites" are bound by each finger. The order of zinc finger binding is such that the amino-terminal finger (F1) binds to the 3′-most subsite, the middle finger (F2) binds to the middle subsite, and the carboxy-terminal finger (F3) binds to the 5′-most subsite within the target site. Artificial zinc finger proteins with altered binding specificities have primarily been engineered by selection from combinatorial libraries because of the complexity of zinc finger protein-DNA interactions (see Sect. 5.4 below).

5.3 Zinc Finger Fusion Proteins

5.3.1 Zinc Finger Nucleases (ZFNs)

Zinc finger nucleases (ZFNs) are fusion proteins composed of engineered zinc finger arrays fused to a non-specific nuclease domain (Fig. 5.1b). In contrast to many restriction enzymes in which catalytic activity and cleavage are inextricably

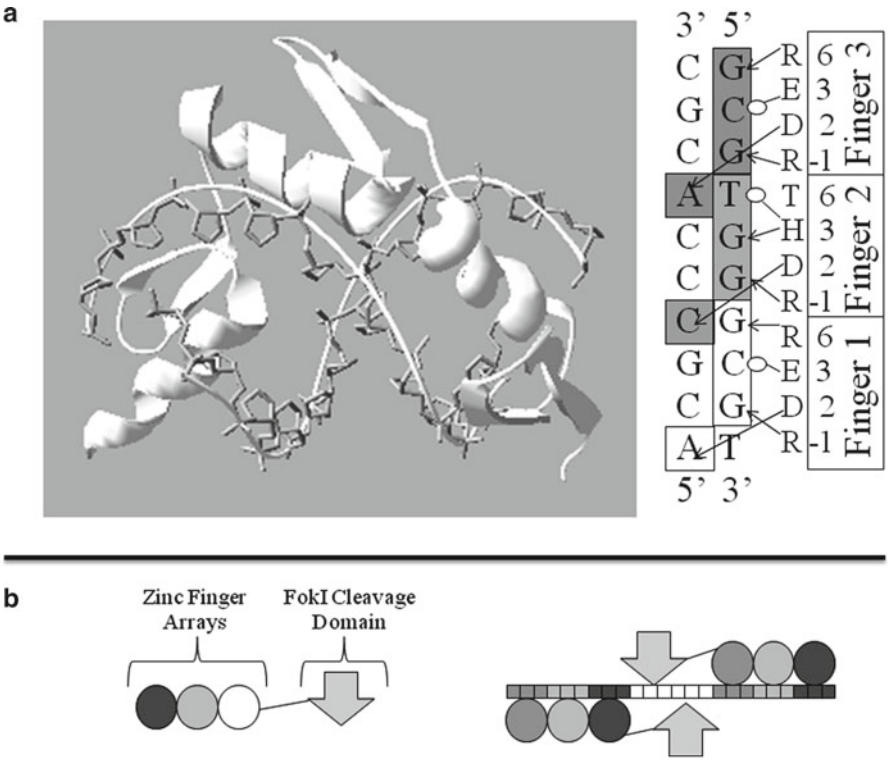


Fig. 5.1 Zinc finger proteins and zinc finger nucleases: (**a**) Structure of Zif268 Zinc Fingers Bound to DNA. In the panel to the *right*, contacts between amino acids in the recognition helix and bases in the target DNA site are summarized. *Arrows* and *open circles* indicate hydrogen bonds and hydrophobic interactions, respectively. Figure adapted from Wolfe et al. (2000). (**b**) Zinc finger nucleases. A zinc finger nuclease is a fusion of an engineered zinc finger array (*circles*) and the non-specific FokI domain (*arrow*). Zinc finger nucleases bind and cleave their target sites as dimers

linked in a single domain, members of the Type IIS subfamily cleave at least one DNA strand outside of the recognition site (Pingoud et al. 2005). Construction of the first ZFN was enabled by the discovery that the cleavage domain of the Type IIS FokI restriction enzyme is functionally independent from its natural DNA-binding domain (Li et al. 1992). The FokI cleavage domain functions enzymatically as a dimer in the native protein (Kim et al. 1996; Bitinaite et al. 1998). Interestingly, it has been suggested that, unlike the wild-type FokI enzyme (Bitinaite et al. 1998), ZFNs may have some minor potential for cleavage when only one of the monomers is bound to DNA and the other is weakly bound to DNA or in solution (Halford et al. 2011; Bitinaite et al. 1998; Catto et al. 2008). Each ZFN monomer typically recognizes a 9–18 bp "half-site" separated by 5–7 bps of intervening "spacer" sequence (within which the FokI domain cleaves). ZFN dimers therefore cleave 18–36 bp target sites, a length of sequence with the potential to be unique in complex genomes.

5.3.2 *Non-nuclease Zinc Finger Fusions*

In addition to the FokI nuclease domain, a variety of other functional domains including transcriptional activation and repression domains have been fused to engineered zinc finger arrays and targeted to specific genomic sites (Rebar et al. 2002; Ooi et al. 2006; Beerli et al. 2000; Zhang et al. 2000; Liu et al. 2001; Ren et al. 2002; Falke et al. 2003; Jamieson et al. 2006; Stolzenburg et al. 2010; Jouvenot et al. 2003). Zinc finger fusions to epigenetic modifiers including DNA and histone methylases have also been constructed; although desired methylation events or expected effects on expression can be detected in some instances, the full range of undesired methylation events have yet to be characterized and minimized (Snowden et al. 2002; Meister et al. 2009; Nomura and Barbas 2007). In addition, zinc finger-integrase and -recombinase fusions have been used to direct modifications to specific DNA sequences (Peng et al. 2002; Tan et al. 2004, 2006; Gordley et al. 2007, 2009; Gersbach et al. 2010; Gaj et al. 2011; Gersbach et al. 2011). Restoration of split luciferase, GFP or β-galactosidase activity by proper co-localization of two zinc finger domains is also being developed for applications in molecular diagnostics (Furman et al. 2011; Ooi et al. 2006; Stains et al. 2005; Ghosh et al. 2006).

5.4 Zinc Finger Engineering Platforms

5.4.1 *Engineering Single Zinc Finger Domains with Novel DNA-Binding Specificities*

Phage-display of combinatorial zinc finger libraries was one of the first effective methods for engineering single individual ZFs with novel DNA-binding specificities. In these studies, large libraries of single zinc finger variants were created by randomizing residues in the recognition helix. These single finger variants were created in the context of a three-finger array where the other two fingers were held constant to serve as "anchors" that positioned the randomized finger over a desired target 3–4 bp target site of interest. Phage display was used successfully to interrogate these libraries to identify individual fingers that bound to a wide variety of different 3–4 bp target sites (Choo and Isalan 2000; Rebar and Pabo 1994; Jamieson et al. 1994; Choo and Klug 1994; Wu et al. 1995). In later studies, a bacterial cell-based "two-hybrid" system was also used successfully to identify fingers from combinatorial libraries with novel binding specificities (Joung et al. 2000).

Although modifying the binding specificities of individual zinc fingers using combinatorial libraries and selection has been straightforward, the engineering of multi-finger arrays has proven to be much more challenging. Various methods have been described for creating engineered multi-finger arrays and we review these approaches below.

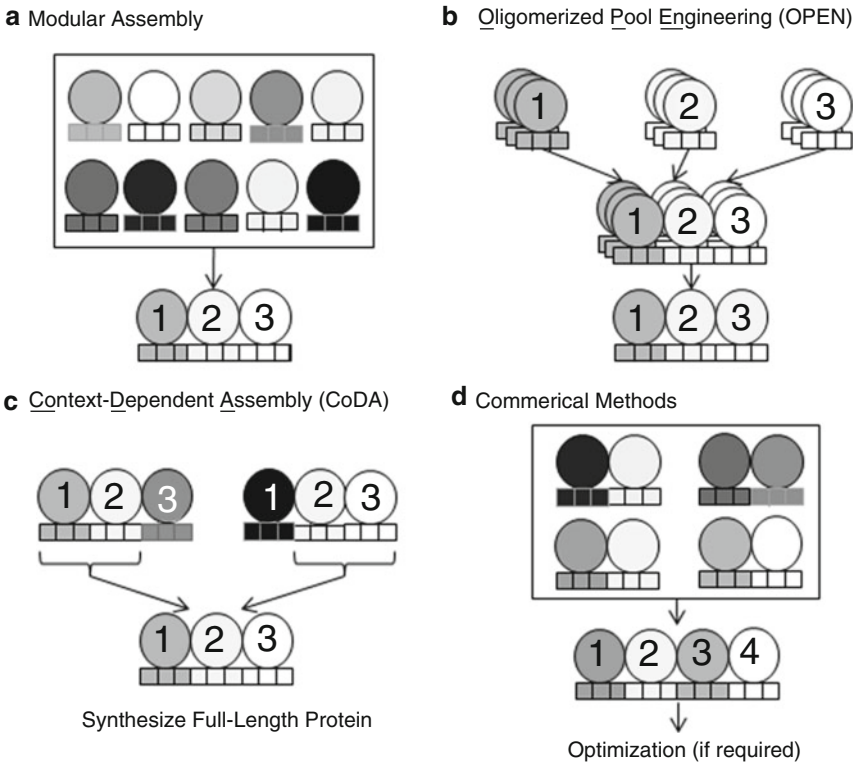


Fig. 5.2 Overview of various zinc finger engineering strategies: (**a**) Modular assembly, (**b**) Oligomerized Pool Engineering (*OPEN*), (**c**) Context-dependent Assembly (*CoDA*), and (**d**) Commercial methods. Zinc fingers represented as *circles* and individual base pairs of DNA are shown as *small squares*. Detailed descriptions of these methods are provided in the text

5.4.2 Modular Assembly

"Modular assembly" was one of the first methods used to create multi-finger arrays (Fig. 5.2a) (Bhakta and Segal 2010). Klug and colleagues first used this approach to engineer three-finger arrays for a target in the BCR-ABL translocation (Choo et al. 1994). Subsequently, other groups also created large collections of individual engineered and naturally occurring zinc finger domains with *in vitro* binding activity for particular DNA triplets (Beerli et al. 1998; Segal et al. 2003; Bae et al. 2003; Liu et al. 2002). Along with detailed protocols, plasmids encoding these modules were later made publicly available on a standardized cloning platform by the Zinc Finger Consortium: http://www.zincfingers.org (Wright et al. 2006). With these pre-made module archives, users could assemble multi-finger proteins by iterative cloning steps until the desired number of modules had been joined together to recognize a particular binding site.

Although modular assembly is straightforward to practice, the success rate of this method has been reported to be low. A large-scale study by our group found that proteins made by modular assembly and tested for DNA binding activity in a bacterial two-hybrid transcriptional activation assay have on average a 76% failure rate, which was as high as 100% for sites lacking GNN subsites (Ramirez et al. 2008; Kim et al. 2009; Joung et al. 2010). Another large-scale analysis by Kim and colleagues also revealed a low success rate for modularly assembled ZFNs in human cells (Kim et al. 2009). The most likely reason for these low success rates is that modular assembly largely ignores the well documented context-dependent behaviors of zinc fingers that can occur within an array (Elrod-Erickson et al. 1998; Wolfe et al. 1999, 2001). This limitation of the method may explain why modularly assembled proteins have in some cases been reported to have low affinities and specificities *in vitro* (Hurt et al. 2003) and low activities and high toxicities in cultured human cells (Cornu et al. 2008).

Other groups have also suggested that certain modules may perform consistently better than others, which may be useful in guiding the choice of binding site and design of proteins with a higher likelihood of success (Kim et al. 2009, 2010; Sander et al. 2009). However, restricting use to this smaller subset of modules in turn leads to a more limited targeting range for this method.

5.4.3 Context-Sensitive Methods for Engineering Multi-finger Arrays

In this section, we review various methods for engineering multi-finger arrays that, in contrast to modular assembly, explicitly attempt to account for the context-dependent DNA-binding activities of fingers within the larger array. An ideal method would randomize the recognition helices of all fingers in the array and then use selection to identify the combination of variant fingers that best work together to mediate recognition of the target DNA site. However, as noted above, the size of a single finger combinatorial library can be quite large (e.g. 2×10^8 members) and therefore such randomized multi-finger libraries would be far too large to reasonably construct or interrogate using current molecular biology and selection-based approaches. Each of the methods we describe in this section seeks to manage the size of combinatorial library selections while still accounting for context-dependent activities of zinc fingers.

5.4.3.1 Sequential Optimization

The sequential optimization method seeks to select one finger at a time in the context of its neighboring finger or fingers to identify combinations of fingers that function well together for recognition (Greisman and Pabo 1997). This strategy

begins with the wild-type fingers from Zif268. In a first selection step, the C-terminal Zif268 finger is removed and replaced with a combinatorial library of fingers. Selections are then performed to identify members of this library that can bind to the 3′ subsite of the final desired target site (positioned adjacent to subsites bound by the remaining Zif268 "anchor" fingers). Following this selection step, the N-terminal Zif268 finger is removed and another library of fingers is placed at the C-terminal end. A second selection is then performed for the middle subsite of the final desired target site (positioned adjacent to the 3′ subsite of the final site and the subsite bound by the remaining middle finger of Zif268). This process is performed a third time to select fingers for the 5′ subsite of the final target site. This approach has been shown to yield three-finger proteins with high affinities and specificities for their target binding sites (Greisman and Pabo 1997; Wolfe et al. 1999). However, this method is difficult for most laboratories to practice because it is labor-intensive, requiring the serial construction of multiple, experiment-specific combinatorial libraries as well as the performance of selections with each of those libraries.

5.4.3.2 Bi-partite Assembly

Choo and colleagues developed a bi-partite assembly method that requires the construction of only two master randomized libraries (Isalan et al. 2001). With this approach, they conceptually "divide" a three-finger array into two halves with one-and-a-half fingers each binding to half of the full target site. Using this strategy, they constructed two randomized libraries in which one-and-a-half fingers of each half of the Zif268 three-finger array were randomized. Selections were then performed from these libraries for target half-sites and full-length proteins assembled from the resulting one-and-a-half fingers. This approach has been shown to yield high quality three-finger arrays. Although bi-partite assembly requires the construction of only two randomized one-and-a-half finger libraries, the combinatorial diversity of fingers in these master libraries is more constrained due to the need to randomize a greater number of positions in one-and-a-half fingers compared with a single finger. This limitation in the initial libraries may therefore reduce the number and quality of fingers that come from selections.

5.4.3.3 Oligomerized Pool Engineering

Oligomerized Pool Engineering (OPEN) was developed by the Zinc Finger Consortium to provide a robust, publicly available method for engineering zinc finger arrays that could be practiced by any academic scientist with standard molecular biology experience (Maeder et al. 2008, 2009; Hurt et al. 2003) (Fig. 5.2b). The key to practicing this method is an archive of zinc finger "pools", each containing up to 95 unique fingers for a different target 3 bp subsite. These pools were selected from randomized zinc finger libraries under low stringency selection conditions to enable the identification of many different zinc fingers each capable of binding to a

target 3 bp subsite of interest. To engineer a three-finger array for a 9 bp target site of interest, a recombined library consisting of the three appropriate pools is generated. This relatively small library (with a combinatorial diversity of less than $95^3 = 857{,}375$ members) is then interrogated using a bacterial two-hybrid system to identify the combination of fingers that work well together to recognize the target site. OPEN ZFNs have been used to modify or mutate endogenous genes in zebrafish (Foley et al. 2009b), tobacco (Townsend et al. 2009), *Arabidopsis* (Zhang et al. 2010), human somatic cells (Maeder et al. 2008), and human pluripotent stem cells (Zou et al. 2009; Sebastiano et al. 2011).

At present, OPEN pools exist for 66 of the 192 potential 3 bp subsites (64 possible 3 bp subsites for each of the three finger positions = 192 subsites). This range of pools enables targeting of a dimeric ZFN site once in every ~200 bps of random DNA sequence (Maeder et al. 2008). Binding sites targetable by OPEN within user-defined sequences of interest can be identified with the publicly available online tool ZiFiT: http://zifit.partners.org (Sander et al. 2007, 2010a). Data derived from the outcomes of 135 successful and failed OPEN selections have also been used to create a machine learning algorithm to assign relative probabilities that a ZFN pair will show activity. These confidence scores have been integrated into the ZiFiT software program to facilitate the process of identifying the most promising ZFN sites from user-specified sequences (Sander et al. 2010b). To further aid researchers interested in modifying endogenous loci in frequently studied organisms, the related software program ZFNGenome (http://bindr.gdcb.iastate.edu/ZFNGenome/) was developed as a tool that can be used to browse genomes to visualize any of the 11.6 million potential target sites of OPEN-generated ZFNs (Reyon et al. 2011). Both ZiFiT and ZFNGenome are freely available on the web without any requirement for user registration.

A protocol similar to OPEN has also been used successfully to engineer ZFNs that function efficiently in zebrafish (Meng et al. 2008). This method begins from partially randomized single finger libraries and utilizes a bacterial "one-hybrid" selection system derived from the original bacterial two-hybrid system used in OPEN. Software and protocols for practicing the one-hybrid method are available on the web at: http://pgfe.umassmed.edu/ZFPsearch.html.

5.4.3.4 Context-Dependent Assembly (CoDA)

More recently, a simple assembly-based method known as Context-Dependent Assembly (CoDA) has been described for engineering multi-finger arrays (Fig. 5.2c). With CoDA, combinations of fingers known to work well with one another are assembled together. By analyzing the results of dozens of OPEN selections, the observation was made that certain F2 fingers appear repeatedly within various adjacent finger contexts (Sander et al. 2011). This led to the idea that adjacent fingers selected to work well next to a particular F2 finger could be joined together to create three-finger arrays with novel specificities, without the need to perform selections. The Joung lab has engineered sets of 18 fixed F2 units and associated

319 N-terminal-end fingers (F1 units) and 344 C-terminal-end fingers (F3 units). With these reagents, a ZFN pair can be rapidly assembled for a site once in every ~500 bp of random DNA sequence. We have used ZFNs made by CoDA to successfully modify 25 endogenous genes in *Arabidopsis*, soybean, and zebrafish with an overall success rate of >50% (Sander et al. 2011) and other groups have also used the method to generate active ZFNs (e.g. Osborn et al. 2011). The ZiFiT software program contains a module that enables users to search for sites targetable by CoDA in their sequence of interest.

5.4.4 Commercial Zinc Finger Engineering

In addition to the publicly available zinc finger engineering platforms described above, one can also choose to purchase ZFNs through Sigma-Aldrich (Pearson 2008; Urnov et al. 2005; Carbery et al. 2010). Although Sigma-Aldrich uses proprietary methods developed by Sangamo BioSciences, some details of the approach used have been disclosed in publications. For example, using a previously published strategy originally described by Choo and colleagues (Moore et al. 2001; Isalan and Choo 2001; Isalan et al. 2001; Jamieson et al. 2003), four- and six-finger arrays are apparently assembled from previously validated two-finger modules (Perez et al. 2008; Doyon et al. 2008), derived from a proprietary archive owned by Sangamo (Fig. 5.2d). Two finger units are sometimes joined together by canonical linkers (TGEKP) or by non-canonical "disrupted linkers" (e.g.--TGSQKP) (Perez et al. 2008; Moore et al. 2001). An open-source method was recently described for designing 4- to 6-finger zinc finger proteins using fingers selected via a commercially available yeast one-hybrid selection system (Herrmann et al. 2011).

5.5 Modified ZFN Architectures

To minimize the formation of unwanted homodimeric ZFNs, two groups independently developed obligate heterodimeric versions of the FokI nuclease domain. Structural analysis and functional characterization of variants by these groups yielded two sets of heterodimer architectures in which hydrophobic and electrostatic interactions have been modified, including the EL:KK (also termed '−'and '+', respectively) (Miller et al. 2007) and DD:RR frameworks (Szczepek et al. 2007).

Although heterodimeric FokI cleavage domains exhibit reduced toxicity, they also possess less robust nuclease activity compared to the original homodimeric FokI domain. Novel mutations (N496D and H537R, numbered relative to the wild-type FokI sequence) that favor the formation of salt bridges to strengthen interactions between the aforementioned heterodimeric FokI domains directly addressed this problem. The introduction of these mutations not only enhance ZFN activity to

levels comparable to that of homodimeric FokI ZFNs but also serve to further suppress homodimerization (Doyon et al. 2010b). Several of these heterodimer frameworks are orthologous in their interactions and therefore enable the introduction of two pairs of ZFNs simultaneously into cells with minimal cytotoxicity (Sollu et al. 2010; Doyon et al. 2010b), a capability that will be useful for applications requiring simultaneous modification of two genomic loci. Furthermore, an engineered hyperactive mutant of FokI ("Sharkey") has been shown to be compatible with these newest heterodimeric FokI domains, yielding an additive increase in activity (Doyon et al. 2010b; Guo et al. 2010).

The length and sequence composition of linkers between zinc finger arrays and the FokI cleavage domain have also been varied and analyzed for function. These studies have yielded a repertoire of linkers that enable ZFN binding to target sites with 5–18 bp spacers between the half-sites (Bibikova et al. 2001; Handel et al. 2009).

Attempts have even been made at engineering ZFNs that function as monomers. These studies have met with only limited success, primarily due to generally reduced activity levels relative to ZFN dimers and increased frequency of non-specific cleavage events (Minczuk et al. 2008; Mino et al. 2009; Mori et al. 2009). A notable achievement in these efforts is the targeting of a single-chain ZFN to mitochondria to induce selective degradation of defective organelles in heteroplasmic cells (Minczuk et al. 2008, 2010; Minczuk 2010).

5.6 Types of ZFN-Inducible Genome Modifications

5.6.1 Gene Mutagenesis by ZFN-Induced Non-homologous End-Joining (NHEJ)

In a broad range of organisms and cell types, the short (typically 4–5 bp) complementary 5′ overhangs of ZFN-induced DSBs can be re-ligated either perfectly or with the introduction of small insertions or deletions (indels) by NHEJ (Hartlerode and Scully 2009; Orlando et al. 2010). NHEJ is the primary mechanism for correction in the event that a repair template for homology-directed repair (HDR) cannot be found (van Gent et al. 2001). Even in the absence of selection, gene disruption rates exceeding 50% can be achieved under optimal conditions (Doyon et al. 2010b). Transient incubation of cultured cells at 30°C has also been shown to increase NHEJ rates 1.5–15-fold, although it is not yet known how generalizable this finding is (Doyon et al. 2010a; Handel and Cathomen 2010; Herrmann et al. 2011).

Short indels from ZFN-mediated mutagenesis can be detected by mismatch-sensitive endonuclease assays (Cel-I/Surveyor or T7 endonuclease I) with a detection limit of about 1% (Oleykowski et al. 1998; Guschin et al. 2010; Kim et al. 2009). To identify lower frequency mutations, Sanger or next-generation sequencing is required (Perez et al. 2008; Foley et al. 2009a; Foley et al. 2009b). Longer deletions

can also be introduced by ZFNs, but these can be missed by mismatch-sensitive endonuclease assays or by sequencing (Morton et al. 2006).

ZFN-induced NHEJ can also be harnessed to induce larger deletions and translocations. Lee et al. reported a diverse range of genomic deletions using ZFNs targeted in the vicinity of the human *CCR5* locus, including one which was 15 kb in length. These large deletions were shown to be caused by the simultaneous cleavage of cognate sites at *CCR5* and *CCR2* (Lee et al. 2009). Translocations can be induced when ZFN-mediated DSBs are introduced into sites present on different chromosomes (Brunet et al. 2009; Simsek et al. 2011).

NHEJ-mediated repair of ZFN-induced DSBs can also be exploited to mediate the insertion of desired exogenous linear donor DNAs with efficiencies as high as 10% in the absence of selection (Orlando et al. 2010). This process is most efficient if the 5′ overhang of the fragment to be inserted is complementary to the overhang generated by ZFN cleavage and if phosphorothioate modifications are present in the donor to confer greater stability within cells. In addition, simultaneous cutting by two ZFN pairs in *cis* can result in up to 50% donor insertion (Orlando et al. 2010).

5.6.2 *Genome Editing by ZFN-Induced Homology-Directed Repair*

ZFNs have been used to mediate highly efficient introduction of specific alterations or insertions by homology-directed repair (HDR). Site-specific DSBs can stimulate HDR by several orders of magnitude (Jasin 1996; Rouet et al. 1994) although it is generally believed that ZFN-induced alterations or insertions should be within 100 bp of the DSB site for maximal activity due to the relatively short conversion tract length in mammals and the observation that the efficiency of gene conversion is inversely related to distance (Elliott et al. 1998; Donoho et al. 1998; Hartlerode and Scully 2009).

Whereas NHEJ promotes DNA repair throughout the cell cycle, HDR is active predominantly during the S and G2 phases to counteract damage associated with genome replication (Hartlerode et al. 2011; Takata et al. 1998; Mao et al. 2008). This observation has been exploited to increase the rates of repair for ZFN-induced DSBs by treatment with cytostatic drugs like vinblastine (Urnov et al. 2005) and nocodazole (Olsen et al. 2010), which arrest cultured cells in G2/M. Increasing levels of donor in the nucleus can also bias repair toward HDR (Gellhaus et al. 2009; Lombardo et al. 2007).

The delivery method and configuration of the repair template donor can have dramatic effects on rates of correction and fidelity. For instance, plasmid DNA with 1.5 kb homology (~750 bp homology arms) was used to introduce a silent point mutation in up to 18% of *IL2Rγ* loci in transformed human cells (Urnov et al. 2005). It was later demonstrated that sequences ranging from 12-bp epitope tags to 8 kb

cassettes harboring multiple transgenes could be knocked-in with 16% and 6% efficiency, respectively, in the absence of selection (Moehle et al. 2007). Up to 50% endogenous gene correction or addition by HDR was achieved in human cells by delivery of ZFN and donor via integration-defective lentiviral (IDLV) vectors (Lombardo et al. 2007). Adeno-associated viral (AAV) vectors have even been delivered to whole organisms yielding partial phenotypic correction of hemophilia in a mouse model (Li et al. 2011).

Despite surprisingly little shared sequence homology compared to the previously mentioned donor types, it has also been demonstrated that PCR-derived double-stranded linear DNA donors with 50 bp homology arms can stimulate transgene addition in 5–10% of chromosomes (Orlando et al. 2010). Although initial reports suggested that single strand oligonucleotide donors could only achieve up to <1% gene conversion (Olsen et al. 2009; Radecke et al. 2006; Radecke et al. 2010), rates as high as 57% for incorporation of insertions have been published for single-stranded donors flanked by ~40 bp of homology on either side of the ZFN cut site (Chen et al. 2011). In addition, appropriately designed single-stranded donors can be used to mediate deletion of sequences of up to 100 kb (Chen et al. 2011).

5.7 Cell Line and Organismic Modifications Using ZFNs

ZFNs have dramatically simplified and accelerated the generation of cell lines and model organisms bearing desired mutations of interest. Prior to the use of ZFNs, many of these cell lines and organisms were previously genetically intractable (Table 5.1). The high efficiency rates observed with ZFNs enable the identification of founder organisms able to pass mutations through the germ line without the need for selection (Foley et al. 2009b). Cell lines with alterations at up to three different endogenous loci can be readily created using ZFNs (Cost et al. 2009; Liu et al. 2010). Despite these successes, it is important to note that ZFNs may be less efficient in certain genomic contexts (e.g. regions of heterochromatin or DNA methylation), which can vary between cell types (Liu et al. 2001; Maeder et al. 2008). As an alternative, the introduction of genes into putative "safe harbor" loci such as *PPP1R12C/AAVS1* or *CCR5* may be a viable therapeutic strategy. In proof-of-concept studies, ZFN-stimulated integration was achieved in up to 50% of alleles surviving selection in human embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) at safe harbor loci (Hockemeyer et al. 2009; Hockemeyer and Jaenisch 2011). To further validate these putative loci, the epigenetic and transcriptional effects of ZFN-integrated expression cassettes has been evaluated in detail for the *PPP1R12C/AAVS1* and *CCR5* sites (Lombardo et al. 2011); although it has previously been thought that safe harbor loci should be in gene deserts to avoid disruption of transcription, these authors argue that having the ability to characterize (and in many cases, through adequate design considerations, minimize) perturbations in genes neighboring integration events may be preferable to altering a genetic landscape with unknown consequences.

Table 5.1 Summary of ZFN-mediated organismic and cell line modifications

Organism	References
Frog	Young et al. (2011) and Bibikova et al. (2001)
Fruit fly	Bibikova et al. (2002), Beumer et al. (2008), Bozas et al. (2009), Beumer et al. (2006) and Bibikova et al. (2003)
Arabidopsis	Sander et al. (2011), Zhang et al. (2010) and Lloyd et al. (2005)
Nematode	Morton et al. (2006) and Wood et al. (2011)
Tobacco	Maeder et al. (2008) and Wright et al. (2005)
Zebrafish	Doyon et al. (2008), Foley et al. (2009b), Meng et al. (2008) and Sander et al. (2011)
Maize	Shukla et al. (2009)
Rat	Geurts et al. (2009a, b), Cui et al. (2010), Moreno et al. (2011) and Mashimo et al. (2010)
Mouse	Goldberg et al. (2010), Carbery et al. (2010), Cui et al. (2010), Meyer et al. (2010) and Li et al. (2011)
Pig	Watanabe et al. (2010), Whyte et al. (2011) and Yang et al. (2011)
Sea Urchin	Ochiai et al. (2010)
Soybean	Sander et al. (2011)
Silkworm	Takasu et al. (2010)
Rabbit	Flisikowska et al. (2011)
Cell Line	
Chinese hamster ovary (CHO) cells	Santiago et al. (2008), Cost et al. (2009) and Liu et al. (2010)
Immortalized or transformed human cell lines	Szczepek et al. (2007), Radecke et al. (2010), Porteus and Baltimore (2003), Urnov et al. (2005), Maeder et al. (2008), Moehle et al. (2007) and Alwin et al. (2005)
Primary human T cells	Urnov et al. (2005), Perez et al. (2008) and Wilen et al. (2011)
Human induced pluripotent stem cells (iPSCs)	Hockemeyer et al. (2009) and Zou et al. (2009, 2011)
Human mesenchymal stromal cells	Benabdallah et al. (2010)
Human hematopoietic stem cells (HSCs)	Holt et al. (2010)

5.8 Cytotoxicity and Off-Target Effects

5.8.1 Assessing and Attenuating ZFN-Mediated Cytotoxicity

Given their ability to introduce DSBs, it is not surprising that ZFNs lead cytotoxic effects in cells. However, the degree of cytotoxicity appears to be different for each ZFN pair and is determined in part by the specificity with which a ZFN can cleave its intended unique target and can avoid cleavage at alternative "off-target" sites. Consistent with this idea, obligate ZFN heterodimers show greatly reduced

cytotoxicity (See Sect. 5.5) (Doyon et al. 2010b; Miller et al. 2007; Szczepek et al. 2007). Despite the use of heterodimeric ZFNs, high levels of nuclease expression can still lead to significant consequences, as with the generation of developmentally abnormal zebrafish "monsters" (Meng et al. 2008; Doyon et al. 2008).

Because excess protein concentration exacerbates cytotoxic effects, ZFN expression should ideally be as low and transient as possible to achieve the desired modification without otherwise affecting genomic integrity and cellular viability (Pruett-Miller et al. 2008; Beumer et al. 2006). A variety of delivery options have been explored as means of maximizing ZFN efficacy, including plasmids (Holt et al. 2010; Porteus and Baltimore 2003), mRNAs (Zou et al. 2010; Doyon et al. 2008; Meng et al. 2008), and viral vectors—adeno-associated- (Porteus et al. 2003), adeno- (Perez et al. 2008), and integrase-deficient lenti- (Lombardo et al. 2007; Cornu and Cathomen 2007) viruses. Temporal regulation of ZFN expression has been achieved by fusion to drug-dependent domains, but this system exhibits leaky baseline expression and a limited dynamic range (Pruett-Miller et al. 2009) and therefore could benefit from additional optimization.

A variety of assays have been used to evaluate cytotoxicity of ZFNs including apoptosis (Alwin et al. 2005), cell survival (Alwin et al. 2005; Maeder et al. 2008; Pruett-Miller et al. 2008; Cornu et al. 2008; Sollu et al. 2010; Porteus and Baltimore 2003), and teratogenicity (Meng et al. 2008; Doyon et al. 2008; Foley et al. 2009b). Visualizing DNA repair factors recruited to sites of DSBs *in vivo* by staining of γ-H2AX or 53BP1 have also been used to assess genotoxicity (Miller et al. 2007; Szczepek et al. 2007; Cornu et al. 2008). However, these assays only provide a snapshot of DSB repair activity and may lack the sensitivity to determine whether DSBs are occurring at off-target sites over the background of intended heterodimeric ZFN-induced DSBs and those occurring naturally in dividing cells (see, for example, Ramalingam et al. 2010; Perez et al. 2008).

5.8.2 Profiling and Validating Off-Target Cleavage Events

DNA-binding properties of zinc finger monomers have been evaluated using a variety of techniques, including Enzyme-Linked Immunosorbent Assays (ELISA) (Segal et al. 1999, 2003), microarrays (Bulyk et al. 2001), Electrophoretic Shift Assays (EMSA) (Hurt et al. 2003), Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Liu and Stormo 2005; Perez et al. 2008; Wolfe et al. 1999), bacterial two-hybrid transcriptional activation assay (Wright et al. 2006), Cyclical Amplification and Selection of Targets (CAST) (Brayer et al. 2008; Segal et al. 2003), bacterial one-hybrid profiles (Gupta et al. 2010; Meng et al. 2007), Bind-n-Seq (Zykovich et al. 2009), fluorescent anisotropy (Sander et al. 2009), molecular modeling (Yanover and Bradley 2011), and purely computational approaches (Cradick et al. 2011; Sander et al. 2010a). A subset of these various methods can be used to obtain DNA specificity profiles for zinc finger proteins. This information can then be used to computationally search for potential off-target sites in genomes of interest. The greatest limitation of this type of approach is that the DNA-binding

profiles of the zinc finger domains of each ZFN monomer must be used to extrapolate the cleavage specificity of the ZFN dimer.

Given that dimerization of the FokI nuclease domains may lead to cooperative binding of ZFN monomers in a dimer (Bitinaite et al. 1998), it is not entirely surprising that assays using the monomeric DNA-binding specificities of ZFNs do not identify the full range of potential off-target cleavage events in a cell. A novel and recently developed selection scheme for elucidating actual cleavage preferences of an active ZFN dimer *in vitro* has enabled the identification of new potential ZFN off-target sites in human cells. With this approach, eight previously unknown off-target sites for ZFNs targeted to the *CCR5* locus were identified in human K562 cells (Pattanayak et al. 2011). Another approach in which off-target DSBs within human cells were mapped by insertion of integration-defective lentiviruses; this study led to the identification of eight additional off-target sites for the same *CCR5* ZFNs (Gabriel et al. 2011). Both of these methods identified off-target sites in cells that had not been predicted computationally from monomeric ZFN binding data. Interestingly, these methods yielded non-overlapping lists of novel off-target cleavage sites for the *CCR5* ZFNs, strongly suggesting that neither approach comprehensively identifies off-target sites at the genome-wide level.

Confirmation of off-target sites has been typically performed by mismatch-sensitive endonuclease assays and next-generation sequencing to look for indel mutations introduced by NHEJ-mediated repair (Perez et al. 2008; Gupta et al. 2010; Hockemeyer et al. 2009). Despite biases introduced during PCR and next-generation sequencing library preparation, it may be possible to estimate mutation frequencies of up to 1:1000 with reasonable accuracy with this approach (Herrmann et al. 2011). As the cost of sequencing continues to drop, the possibility of subjecting clonal cell populations or small organisms such as nematodes to whole genome sequencing analysis will become an increasingly viable option (Wood et al. 2011; Rahman et al. 2011). However, in applications for which polyclonal cell populations will be used therapeutically, the need for comprehensive off-target characterization will continue to be highly relevant (Cathomen and Schambach 2009). Given that chromosomal translocations are a potential risk with ZFN treatment, cytogenetic karyotype analysis should be performed to screen for gross abnormalities (Cathomen and Schambach 2009; Hockemeyer et al. 2009; Brunet et al. 2009; Simsek et al. 2011). Once ZFN off-target cleavage profiles can be derived with greater accuracy and confidence and for less cost, these methods will enable routine identification of ZFN off-target sites. An important question for future research will be to develop methods for further optimizing the cleavage profiles of ZFNs so as to minimize particularly deleterious off-target effects (Gupta et al. 2010).

5.9 Conclusions

The ability to use ZFNs to efficiently induce gene mutations or alterations is already opening the doors for novel explorations of fundamental biology, including the study of isogenic cell lines with precisely defined variations (DeKelver et al. 2010;

Soldner et al. 2011; Hockemeyer et al. 2009), factors determining histone localization (Goldberg et al. 2010), splicing regulation (Cristea et al. 2011), and clathrin-mediated endocytosis (Doyon et al. 2011). Creative therapeutic strategies have also emerged, such as the disruption of the *CCR5* gene to mimic a naturally-occurring mutation conferring T cells with resistance to HIV infection (NCT01044654, NCT00842634, NCT01252641) (Perez et al. 2008; Holt et al. 2010). With the multitude of options available to academic researchers for engineering ZFNs, there exists a broad range of research and therapeutic applications that can be explored by the field as it continues to move forward.

Acknowledgements We thank members of the Joung lab for helpful comments and suggestions. We thank Victor Quesada for assistance with the development of Fig. 5.1. C.L.R. is partially supported by the National Science Foundation and the Ford Foundation. J.K.J is supported by National Institutes of Health (NIH) R01 GM088040, an NIH Director's Pioneer Award (DP1 OD006862), and the Massachusetts General Hospital Jim and Ann Orr Research Scholar Award.

References

Alwin S, Gere MB, Guhl E, Effertz K, Barbas CF 3rd, Segal DJ, Weitzman MD, Cathomen T (2005) Custom zinc-finger nucleases for use in human cells. Mol Ther 12:610–617

Bae KH, Kwon YD, Shin HC, Hwang MS, Ryu EH, Park KS, Yang HY, Lee DK, Lee Y, Park J, Kwon HS, Kim HW, Yeh BI, Lee HW, Sohn SH, Yoon J, Seol W, Kim JS (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. Nat Biotechnol 21:275–280

Beerli RR, Segal DJ, Dreier B, Barbas CF 3rd (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci USA 95:14628–14633

Beerli RR, Dreier B, Barbas CF 3rd (2000) Positive and negative regulation of endogenous genes by designed transcription factors. Proc Natl Acad Sci USA 97:1495–1500

Benabdallah BF, Allard E, Yao S, Friedman G, Gregory PD, Eliopoulos N, Fradette J, Spees JL, Haddad E, Holmes MC, Beausejour CM (2010) Targeted gene addition to human mesenchymal stromal cells as a cell-based plasma-soluble protein delivery platform. Cytotherapy 12:394–399

Bettini M, Xi H, Milbrandt J, Kersh GJ (2002) Thymocyte development in early growth response gene 1-deficient mice. J Immunol 169:1713–1720

Beumer K, Bhattacharyya G, Bibikova M, Trautman JK, Carroll D (2006) Efficient gene targeting in Drosophila with zinc-finger nucleases. Genetics 172:2391–2403

Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, Gall JG, Carroll D (2008) Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci USA 105:19821–19826

Bhakta MS, Segal DJ (2010) The generation of zinc finger proteins by modular assembly. Methods Mol Biol 649:3–30

Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, Chandrasegaran S (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol 21:289–297

Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161:1169–1175

Bibikova M, Beumer K, Trautman JK, Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. Science 300:764

Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci USA 95:10570–10575

Bozas A, Beumer KJ, Trautman JK, Carroll D (2009) Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila. Genetics 182:641–651

Brayer KJ, Kulshreshtha S, Segal DJ (2008) The protein-binding potential of C2H2 zinc finger domains. Cell Biochem Biophys 51:9–19

Brunet E, Simsek D, Tomishima M, Dekelver R, Choi VM, Gregory P, Urnov F, Weinstock DM, Jasin M (2009) Chromosomal translocations induced at specified loci in human stem cells. Proc Natl Acad Sci USA 106:10620–10625

Bulyk ML, Huang X, Choo Y, Church GM (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. Proc Natl Acad Sci USA 98:7158–7163

Carbery ID, Ji D, Harrington A, Brown V, Weinstein EJ, Liaw L, Cui X (2010) Targeted genome modification in mice using zinc-finger nucleases. Genetics 186:451–459

Cathomen T, Schambach A (2009) Zinc-finger nucleases meet iPS cells: Zinc positive: tailored genome engineering meets reprogramming. Gene Ther 17:1–3

Catto LE, Bellamy SR, Retter SE, Halford SE (2008) Dynamics and consequences of DNA looping by the FokI restriction endonuclease. Nucleic Acids Res 36:2073–2081

Chen F, Pruett-Miller SM, Huang Y, Gjoka M, Duda K, Taunton J, Collingwood TN, Frodin M, Davis GD (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods 8(9):753–755

Choo Y, Isalan M (2000) Advances in zinc finger engineering. Curr Opin Struct Biol 10:411–416

Choo Y, Klug A (1994) Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. Proc Natl Acad Sci USA 91:11168–11172

Choo Y, Sanchez-Garcia I, Klug A (1994) In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature 372:642–645

Cornu TI, Cathomen T (2007) Targeted genome modifications using integrase-deficient lentiviral vectors. Mol Ther 15:2107–2113

Cornu TI, Thibodeau-Beganny S, Guhl E, Alwin S, Eichtinger M, Joung JK, Cathomen T (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. Mol Ther 16:352–358

Cost GJ, Freyvert Y, Vafiadis A, Santiago Y, Miller JC, Rebar E, Collingwood TN, Snowden A, Gregory PD (2009) BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. Biotechnol Bioeng 105:330–340

Cradick TJ, Ambrosini G, Iseli C, Bucher P, McCaffrey AP (2011) ZFN-Site searches genomes for zinc finger nuclease target sites and off-target sites. BMC Bioinformatics 12:152

Cristea S, Gregory PD, Urnov FD, Cost GJ (2011) Dissection of splicing regulation at an endogenous locus by zinc-finger nuclease-mediated gene editing. PLoS One 6:e16961

Cui X, Ji D, Fisher DA, Wu Y, Briner DM, Weinstein EJ (2010) Targeted integration in rat and mouse embryos with zinc-finger nucleases. Nat Biotechnol 29:64–67

Dekelver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijsing SH, Sancak Y, Cui X, Steine EJ, Miller JC, Tam P, Bartsevich VV, Meng X, Rupniewski I, Gopalan SM, Sun HC, Pitz KJ, Rock JM, Zhang L, Davis GD, Rebar EJ, Cheeseman IM, Yamamoto KR, Sabatini DM, Jaenisch R, Gregory PD, Urnov FD (2010) Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res 20:1133–1142

Donoho G, Jasin M, Berg P (1998) Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. Mol Cell Biol 18:4070–4078

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, Hocking TD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol 26:702–708

Doyon Y, Choi VM, Xia DF, Vo TD, Gregory PD, Holmes MC (2010a) Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. Nat Methods 7:459–460

Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, Xia DF, Miller JC, Urnov FD, Gregory PD, Holmes MC (2010b) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods 8:74–79

Doyon JB, Zeitler B, Cheng J, Cheng AT, Cherone JM, Santiago Y, Lee AH, Vo TD, Doyon Y, Miller JC, Paschon DE, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Drubin DG (2011) Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. Nat Cell Biol 13:331–337

Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M (1998) Gene conversion tracts from double-strand break repair in mammalian cells. Mol Cell Biol 18:93–101

Elrod-Erickson M, Benson TE, Pabo CO (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. Structure 6:451–464

Falke D, Fisher M, Ye D, Juliano RL (2003) Design of artificial transcription factors to selectively regulate the pro-apoptotic bax gene. Nucleic Acids Res 31:e10

Flisikowska T, Thorey IS, Offner S, Ros F, Lifke V, Zeitler B, Rottmann O, Vincent A, Zhang L, Jenkins S, Niersbach H, Kind AJ, Gregory PD, Schnieke AE, Platzer J (2011) Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. PLoS One 6:e21045

Foley JE, Maeder ML, Pearlberg J, Joung JK, Peterson RT, Yeh JR (2009a) Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. Nat Protoc 4:1855–1867

Foley JE, Yeh JR, Maeder ML, Reyon D, Sander JD, Peterson RT, Joung JK (2009b) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). PLoS One 4:e4348

Furman JL, Mok PW, Badran AH, Ghosh I (2011) Turn-On DNA Damage Sensors for the Direct Detection of 8-Oxoguanine and Photoproducts in Native DNA. J Am Chem Soc 133(32):12518–12527

Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppel C, Nowrouzi A, Bartholomae CC, Wang J, Friedman G, Holmes MC, Gregory PD, Glimm H, Schmidt M, Naldini L, von Kalle C (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol 29(9):816–823

Gaj T, Mercer AC, Gersbach CA, Gordley RM, Barbas CF 3rd (2011) Structure-guided reprogramming of serine recombinase DNA sequence specificity. Proc Natl Acad Sci USA 108:498–503

Gellhaus K, Cornu TI, Heilbronn R, Cathomen T (2009) Fate of recombinant adeno-associated viral vector genomes during DNA double-strand break-induced gene targeting in human cells. Hum Gene Ther 21:543–553

Gersbach CA, Gaj T, Gordley RM, Barbas CF 3rd (2010) Directed evolution of recombinase specificity by split gene reassembly. Nucleic Acids Res 38:4198–4206

Gersbach CA, Gaj T, Gordley RM, Mercer AC, Barbas CF 3rd (2011) Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. Nucleic Acids Res 39(17):7868–7878

Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Menoret S, Anegon I, Davis GD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jacob HJ, Buelow R (2009a) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325:433

Geurts AM, Cost GJ, Remy S, Cui X, Tesson L, Usal C, Menoret S, Jacob HJ, Anegon I, Buelow R (2009b) Generation of gene-specific mutated rats using zinc-finger nucleases. Methods Mol Biol 597:211–225

Ghosh I, Stains CI, Ooi AT, Segal DJ (2006) Direct detection of double-stranded DNA: Molecular methods and applications for DNA diagnostics. Mol Biosyst 2:551–560

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, Wen D, Chapgier A, Dekelver RC, Miller JC, Lee YL, Boydston EA, Holmes MC, Gregory PD, Greally JM, Rafii S, Yang C, Scambler PJ, Garrick D, Gibbons RJ, Higgs

DR, Cristea IM, Urnov FD, Zheng D, Allis CD (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140:678–691
Gordley RM, Smith JD, Graslund T, Barbas CF 3rd (2007) Evolution of programmable zinc finger-recombinases with activity in human cells. J Mol Biol 367:802–813
Gordley RM, Gersbach CA, Barbas CF 3rd (2009) Synthesis of programmable integrases. Proc Natl Acad Sci USA 106:5053–5058
Greisman HA, Pabo CO (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. Science 275:657–661
Guo J, Gaj T, Barbas CF 3rd (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol 400:96–107
Gupta A, Meng X, Zhu LJ, Lawson ND, Wolfe SA (2010) Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. Nucleic Acids Res 39:381–392
Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ (2010) A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol 649:247–256
Halford SE, Catto LE, Pernstich C, Rusling DA, Sanders KL (2011) The reaction mechanism of FokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. Biochem Soc Trans 39:584–588
Handel EM, Cathomen T (2010) Zinc-finger nuclease based genome surgery: it's all about specificity. Curr Gene Ther 11:28–37
Handel EM, Alwin S, Cathomen T (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Mol Ther 17:104–111
Hartlerode AJ, Scully R (2009) Mechanisms of double-strand break repair in somatic mammalian cells. Biochem J 423:157–168
Hartlerode A, Odate S, Shim I, Brown J, Scully R (2011) Cell cycle-dependent induction of homologous recombination by a tightly regulated I-SceI fusion protein. PLoS One 6:e16501
Herrmann F, Garriga-Canut M, Baumstark R, Fajardo-Sanchez E, Cotterell J, Minoche A, Himmelbauer H, Isalan M (2011) p53 gene repair with zinc finger nucleases optimised by Yeast 1-Hybrid and validated by Solexa sequencing. PLoS One 6:e20913
Hockemeyer D, Jaenisch R (2011) Gene targeting in human pluripotent cells. Cold Spring Harb Symp Quant Biol 75:201–209
Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, Dekelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng X, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol 27:851–857
Holt N, Wang J, Kim K, Friedman G, Wang X, Taupin V, Crooks GM, Kohn DB, Gregory PD, Holmes MC, Cannon PM (2010) Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nat Biotechnol 28:839–847
Hurt JA, Thibodeau SA, Hirsh AS, Pabo CO, Joung JK (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. Proc Natl Acad Sci USA 100:12271–12276
Isalan M, Choo Y (2001) Rapid, high-throughput engineering of sequence-specific zinc finger DNA-binding proteins. Methods Enzymol 340:593–609
Isalan M, Klug A, Choo Y (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. Nat Biotechnol 19:656–660
Jamieson AC, Kim SH, Wells JA (1994) In vitro selection of zinc fingers with altered DNA-binding specificity. Biochemistry 33:5689–5695
Jamieson AC, Miller JC, Pabo CO (2003) Drug discovery with engineered zinc-finger proteins. Nat Rev Drug Discov 2:361–368
Jamieson AC, Guan B, Cradick TJ, Xiao H, Holmes MC, Gregory PD, Carroll PM (2006) Controlling gene expression in Drosophila using engineered zinc finger protein transcription factors. Biochem Biophys Res Commun 348:873–879

Jasin M (1996) Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet 12:224–228

Jensen NM, Dalsgaard T, Jakobsen M, Nielsen RR, Sorensen CB, Bolund L, Jensen TG (2011) An update on targeted gene repair in mammalian cells: methods and mechanisms. J Biomed Sci 18:10

Joung JK, Ramm EI, Pabo CO (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. Proc Natl Acad Sci USA 97:7382–7387

Joung JK, Voytas DF, Cathomen T (2010) Reply to "Genome editing with modularly assembled zinc-finger nucleases". Nat Methods 7:91–92

Jouvenot Y, Ginjala V, Zhang L, Liu PQ, Oshimura M, Feinberg AP, Wolffe AP, Ohlsson R, Gregory PD (2003) Targeted regulation of imprinted genes by synthetic zinc-finger transcription factors. Gene Ther 10:513–522

Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci USA 93:1156–1160

Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res 19:1279–1288

Kim JS, Lee HJ, Carroll D (2010) Genome editing with modularly assembled zinc-finger nucleases. Nat Methods 7:91, author reply 91–2

Lee MS, Gippert GP, Soman KV, Case DA, Wright PE (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science 245:635–637

Lee HJ, Kim E, Kim JS (2009) Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res 20:81–89

Li L, Wu LP, Chandrasegaran S (1992) Functional domains in Fok I restriction endonuclease. Proc Natl Acad Sci USA 89:4275–4279

Li H, Haurigot V, Doyon Y, Li T, Wong SY, Bhagwat AS, Malani N, Anguela XM, Sharma R, Ivanciu L, Murphy SL, Finn JD, Khazi FR, Zhou S, Paschon DE, Rebar EJ, Bushman FD, Gregory PD, Holmes MC, High KA (2011) In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature 475:217–221

Liu J, Stormo GD (2005) Combining SELEX with quantitative assays to rapidly obtain accurate models of protein-DNA interactions. Nucleic Acids Res 33:e141

Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC, Zhong X, Lee YL, Eisenberg SP, Spratt SK, Case CC, Wolffe AP (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A J Biol Chem 276:11323–11334

Liu Q, Xia Z, Zhong X, Case CC (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. J Biol Chem 277:3850–3856

Liu PQ, Chan EM, Cost GJ, Zhang L, Wang J, Miller JC, Guschin DY, Reik A, Holmes MC, Mott JE, Collingwood TN, Gregory PD (2010) Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. Biotechnol Bioeng 106:97–105

Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci USA 102:2232–2237

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25:1298–1306

Lombardo A, Cesana D, Genovese P, Di Stefano B, Provasi E, Colombo DF, Neri M, Magnani Z, Cantore A, Lo Riso P, Damo M, Pello OM, Holmes MC, Gregory PD, Gritti A, Broccoli V, BONINI C, NALDINI L (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat Methods 8(10):861–869

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Muller-Lerch F, Fu F, Pearlberg J, Gobel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, McCray PB Jr, Cathomen T, Voytas DF, Joung JK (2008) Rapid "open-source" engineering

of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31:294–301
Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, Joung JK (2009) Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. Nat Protoc 4:1471–1501
Mao Z, Bozzella M, Seluanov A, Gorbunova V (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. Cell Cycle 7:2902–2906
Mashimo T, Takizawa A, Voigt B, Yoshimi K, Hiai H, Kuramoto T, Serikawa T (2010) Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. PLoS One 5:e8870
Meister GE, Chandrasegaran S, Ostermeier M (2009) Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. Nucleic Acids Res 38:1749–1759
Meng X, Thibodeau-Beganny S, Jiang T, Joung JK, Wolfe SA (2007) Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method. Nucleic Acids Res 35:e81
Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol 26:695–701
Meyer M, de Angelis MH, Wurst W, Kuhn R (2010) Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. Proc Natl Acad Sci USA 107:15022–15026
Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J 4:1609–1614
Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, Gregory PD, Pabo CO, Rebar EJ (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25:778–785
Minczuk M (2010) Engineered zinc finger proteins for manipulation of the human mitochondrial genome. Methods Mol Biol 649:257–270
Minczuk M, Papworth MA, Miller JC, Murphy MP, Klug A (2008) Development of a single-chain, quasi-dimeric zinc-finger nuclease for the selective degradation of mutated human mitochondrial DNA. Nucleic Acids Res 36:3926–3938
Minczuk M, Kolasinska-Zwierz P, Murphy MP, Papworth MA (2010) Construction and testing of engineered zinc-finger proteins for sequence-specific modification of mtDNA. Nat Protoc 5:342–356
Mino T, Aoyama Y, Sera T (2009) Efficient double-stranded DNA cleavage by artificial zinc-finger nucleases composed of one zinc-finger protein and a single-chain FokI dimer. J Biotechnol 140:156–161
Moehle EA, Rock JM, Lee YL, Jouvenot Y, Dekelver RC, Gregory PD, Urnov FD, Holmes MC (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. Proc Natl Acad Sci USA 104:3055–3060
Moore M, Klug A, Choo Y (2001) Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. Proc Natl Acad Sci USA 98:1437–1441
Moreno C, Hoffman M, Stodola TJ, Didier DN, Lazar J, Geurts AM, North PE, Jacob HJ, Greene AS (2011) Creation and characterization of a renin knockout rat. Hypertension 57:614–619
Mori T, Kagatsume I, Shinomiya K, Aoyama Y, Sera T (2009) Sandwiched zinc-finger nucleases harboring a single-chain FokI dimer as a DNA-cleavage domain. Biochem Biophys Res Commun 390:694–697
Morton J, Davis MW, Jorgensen EM, Carroll D (2006) Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. Proc Natl Acad Sci USA 103:16370–16375
Nomura W, Barbas CF 3rd (2007) In vivo site-specific DNA methylation with a designed sequence-enabled DNA methylase. J Am Chem Soc 129:8676–8677
Ochiai H, Fujita K, Suzuki K, Nishikawa M, Shibata T, Sakamoto N, Yamamoto T (2010) Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. Genes Cells 15:875–885

Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT (1998) Mutation detection using a novel plant endonuclease. Nucleic Acids Res 26:4597–4602

Olsen PA, Solhaug A, Booth JA, Gelazauskaite M, Krauss S (2009) Cellular responses to targeted genomic sequence modification using single-stranded oligonucleotides and zinc-finger nucleases. DNA Repair (Amst) 8:298–308

Olsen PA, Gelazauskaite M, Randol M, Krauss S (2010) Analysis of illegitimate genomic integration mediated by zinc-finger nucleases: implications for specificity of targeted gene correction. BMC Mol Biol 11:35

Ooi AT, Stains CI, Ghosh I, Segal DJ (2006) Sequence-enabled reassembly of beta-lactamase (SEER-LAC): a sensitive method for the detection of double-stranded DNA. Biochemistry 45:3620–3625

Orlando SJ, Santiago Y, Dekelver RC, Freyvert Y, Boydston EA, Moehle EA, Choi VM, Gopalan SM, Lou JF, Li J, Miller JC, Holmes MC, Gregory PD, Urnov FD, Cost GJ (2010) Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. Nucleic Acids Res 38:e152

Osborn MJ, Defeo AP, Blazar B, Tolar J (2011) Synthetic zinc finger nuclease design and rapid assembly. Hum Gene Ther 22(9):1155–1165

Parraga G, Horvath SJ, Eisen A, Taylor WE, Hood L, Young ET, Klevit RE (1988) Zinc-dependent structure of a single-finger domain of yeast ADR1. Science 241:1489–1492

Pattanayak V, Ramirez CL, Joung JK, Liu DR (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods 8(9):765–770

Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252:809–817

Pearson H (2008) Protein engineering: the fate of fingers. Nature 455:160–164

Peng WJ, Chang CM, Lin TH (2002) Target integration by a chimeric Sp1 zinc finger domain-Moloney murine leukemia virus integrase in vivo. J Biomed Sci 9:171–184

Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, Wang N, Lee G, Bartsevich VV, Lee YL, Guschin DY, Rupniewski I, Waite AJ, Carpenito C, Carroll RG, Orange JS, Urnov FD, Rebar EJ, Ando D, Gregory PD, Riley JL, Holmes MC, June CH (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26:808–816

Pingoud A, Fuxreiter M, Pingoud V, Wende W (2005) Type II restriction endonucleases: structure and mechanism. Cell Mol Life Sci 62:685–707

Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763

Porteus MH, Cathomen T, Weitzman MD, Baltimore D (2003) Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. Mol Cell Biol 23:3558–3565

Pruett-Miller SM, Connelly JP, Maeder ML, Joung JK, Porteus MH (2008) Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. Mol Ther 16:707–717

Pruett-Miller SM, Reading DW, Porter SN, Porteus MH (2009) Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. PLoS Genet 5:e1000376

Radecke F, Peter I, Radecke S, Gellhaus K, Schwarz K, Cathomen T (2006) Targeted chromosomal gene modification in human cells by single-stranded oligodeoxynucleotides in the presence of a DNA double-strand break. Mol Ther 14:798–808

Radecke S, Radecke F, Cathomen T, Schwarz K (2010) Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. Mol Ther 18:743–753

Rahman SH, Maeder ML, Joung JK, Cathomen T (2011) Zinc-finger nucleases for somatic gene therapy: the next frontier. Hum Gene Ther 22(8):925–933

Ramalingam S, Kandavelou K, Rajenderan R, Chandrasegaran S (2010) Creating designed zinc-finger nucleases with minimal cytotoxicity. J Mol Biol 405:630–641

Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA, Cathomen T, Voytas DF, Joung JK (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods 5:374–375

Rebar EJ, Pabo CO (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science 263:671–673

Rebar EJ, Huang Y, Hickey R, Nath AK, Meoli D, Nath S, Chen B, Xu L, Liang Y, Jamieson AC, Zhang L, Spratt SK, Case CC, Wolffe A, Giordano FJ (2002) Induction of angiogenesis in a mouse model using engineered transcription factors. Nat Med 8:1427–1432

Ren D, Collingwood TN, Rebar EJ, Wolffe AP, Camp HS (2002) PPARgamma knockdown by engineered transcription factors: exogenous PPARgamma2 but not PPARgamma1 reactivates adipogenesis. Genes Dev 16:27–32

Reyon D, Kirkpatrick JR, Sander JD, Zhang F, Voytas DF, Joung JK, Dobbs D, Coffman CR (2011) ZFNGenome: a comprehensive resource for locating zinc finger nuclease target sites in model organisms. BMC Genomics 12:83

Rouet P, Smih F, Jasin M (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci USA 91:6064–6068

Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D (2007) Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. Nucleic Acids Res 35:W599–W605

Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D (2009) An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. Nucleic Acids Res 37:506–515

Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D (2010a) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res 38:W462–W468

Sander JD, Reyon D, Maeder ML, Foley JE, Thibodeau-Beganny S, Li X, Regan MR, Dahlborg EJ, Goodwin MJ, Fu F, Voytas DF, Joung JK, Dobbs D (2010b) Predicting success of oligomerized pool engineering (OPEN) for zinc finger target site sequences. BMC Bioinformatics 11:543

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, Pierick CJ, Hoffman E, Maeder ML, Khayter C, Reyon D, Dobbs D, Langenau DM, Stupar RM, Giraldez AJ, Voytas DF, Peterson RT, Yeh JR, Joung JK (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8:67–69

Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN (2008) Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proc Natl Acad Sci USA 105:5809–5814

Sebastiano V, Maeder ML, Angstman JF, Haddad B, Khayter C, Yeo DT, Goodwin MJ, Hawkins JS, Ramirez CL, Batista LF, Artandi SE, Wernig M, Joung KJ (2011) In Situ Genetic Correction of the Sickle Cell Anemia Mutation in Human Induced Pluripotent Stem Cells Using Engineered Zinc Finger Nucleases. Stem Cells 29(11):1717–1726

Segal DJ, Dreier B, Beerli RR, Barbas CF 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. Proc Natl Acad Sci USA 96:2758–2763

Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, Huber A, Koksch B, Lund CV, Magnenat L, Valente D, Barbas CF 3rd (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry 42:2137–2148

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X. et al. (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459:437–441

Simsek D, Brunet E, Wong SY, Katyal S, Gao Y, McKinnon PJ, Lou J, Zhang L, Li J, Rebar EJ, Gregory PD, Holmes MC, Jasin M (2011) DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet 7:e1002080

Snowden AW, Gregory PD, Case CC, Pabo CO (2002) Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. Curr Biol 12:2159–2166

Soldner F, Laganiere J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, Khurana V, Golbe LI, Myers RH, Lindquist S, Zhang L, Guschin D, Fong LK, Vu BJ, Meng X, Urnov FD, Rebar EJ, Gregory PD, Zhang HS, Jaenisch R (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146:318–331

Sollu C, Pars K, Cornu TI, Thibodeau-Beganny S, Maeder ML, Joung JK, Heilbronn R, Cathomen T (2010) Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. Nucleic Acids Res 38:8269–8276

Stains CI, Porter JR, Ooi AT, Segal DJ, Ghosh I (2005) DNA sequence-enabled reassembly of the green fluorescent protein. J Am Chem Soc 127:10782–10783

Stolzenburg S, Bilsland A, Keith WN, Rots MG (2010) Modulation of gene expression using zinc finger-based artificial transcription factors. Methods Mol Biol 649:117–132

Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol 25:786–793

Takasu Y, Kobayashi I, Beumer K, Uchino K, Sezutsu H, Sajwan S, Carroll D, Tamura T, Zurovec M (2010) Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. Insect Biochem Mol Biol 40:759–765

Takata M, Sasaki MS, Sonoda E, Morrison C, Hashimoto M, Utsumi H, Yamaguchi-Iwai Y, Shinohara A, Takeda S (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J 17:5497–5508

Tan W, Zhu K, Segal DJ, Barbas CF 3rd, CHOW SA (2004) Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. J Virol 78:1301–1313

Tan W, Dong Z, Wilkinson TA, Barbas CF 3rd, CHOW SA (2006) Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. J Virol 80:1939–1948

Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK, Voytas DF (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459:442–445

Tupler R, Perini G, Green MR (2001) Expressing the human genome. Nature 409:832–833

Urnov FD (2002) A feel for the template: zinc finger protein transcription factors and chromatin. Biochem Cell Biol 80:321–333

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435:646–651

van Gent DC, Hoeijmakers JH, Kanaar R (2001) Chromosomal stability and the DNA double-stranded break connection. Nat Rev Genet 2:196–206

Watanabe M, Umeyama K, Matsunari H, Takayanagi S, Haruyama E, Nakano K, Fujiwara T, Ikezawa Y, Nakauchi H, Nagashima H (2010) Knockout of exogenous EGFP gene in porcine somatic cells using zinc-finger nucleases. Biochem Biophys Res Commun 402:14–18

Whyte JJ, Zhao J, Wells KD, Samuel MS, Whitworth KM, Walters EM, Laughlin MH, Prather RS (2011) Gene targeting with zinc finger nucleases to produce cloned eGFP knockout pigs. Mol Reprod Dev 78:2

Wigler M, Sweet R, Sim GK, Wold B, Pellicer A, Lacy E, Maniatis T, Silverstein S, Axel R (1979) Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell 16:777–785

Wilen CB, Wang J, Tilton JC, Miller JC, Kim KA, Rebar EJ, Sherrill-Mix SA, Patro SC, Secreto AJ, Jordan AP, Lee G, Kahn J, Aye PP, Bunnell BA, Lackner AA, Hoxie JA, Danet-Desnoyers GA, Bushman FD, Riley JL, Gregory PD, June CH, Holmes MC, Doms RW (2011) Engineering HIV-resistant human CD4+ T cells with CXCR4-specific zinc-finger nucleases. PLoS Pathog 7:e1002020

Wolfe SA, Greisman HA, Ramm EI, Pabo CO (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. J Mol Biol 285:1917–1934

Wolfe SA, Grant RA, Elrod-Erickson M, Pabo CO (2001) Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. Structure 9:717–723

Wolfe SA, Ramm EI, Pabo CO (2000) Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. Structure, 8:739–750

Wood AJ, Lo T, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Meyer BJ (2011) Targeted Genome Editing Across Species Using ZFNs and TALENs. Science 333(6040):307

Wright DA, Townsend JA, Winfrey RJ Jr, Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD, Voytas DF (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 44:693–705

Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF, Joung JK (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nat Protoc 1:1637–1652

Wu H, Yang WP, Barbas CF 3rd (1995) Building zinc fingers by selection: toward a therapeutic application. Proc Natl Acad Sci USA 92:344–348

Yang D, Yang H, Li W, Zhao B, Ouyang Z, Liu Z, Zhao Y, Fan N, Song J, Tian J, Li F, Zhang J, Chang L, Pei D, Chen YE, Lai L (2011) Generation of PPARgamma mono-allelic knockout pigs via zinc-finger nucleases and nuclear transfer cloning. Cell Res 21:979–982

Yanover C, Bradley P (2011) Extensive protein and DNA backbone sampling improves structure-based specificity prediction for C_2H_2 zinc fingers. Nucleic Acids Res 39(11):4564–4576

Young JJ, Cherone JM, Doyon Y, Ankoudinova I, Faraji FM, Lee AH, Ngo C, Guschin DY, Paschon DE, Miller JC. et al. (2011) Efficient targeted gene disruption in the soma and germ line of the frog Xenopus tropicalis using engineered zinc-finger nucleases. Proc Natl Acad Sci USA 108:7052–7057

Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, Raschke EE, Jamieson AC, Rebar EJ, Wolffe AP, Case CC (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. J Biol Chem 275:33850–33860

Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, Joung JK, Voytas DF (2010) High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. Proc Natl Acad Sci USA 107:12028–12033

Zou J, Maeder ML, Mali P, Pruett-Miller SM, Thibodeau-Beganny S, Chou BK, Chen G, Ye Z, Park IH, Daley GQ, Porteus MH, Joung JK, Cheng L (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell 5:97–110

Zou J, Cochran R, Cheng L (2010) Double knockouts in human embryonic stem cells. Cell Res 20:250–252

Zou J, Sweeney CL, Chou BK, Choi U, Pan J, Wang H, Dowey SN, Cheng L, Malech HL (2011) Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. Blood 117:5561–5572

Zykovich A, Korf I, Segal DJ (2009) Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. Nucleic Acids Res 37:e151

Chapter 6
Engineered Meganucleases for Genome Engineering Purposes

Jean Charles Epinat, George H. Silva, Frédéric Pâques, Julianne Smith, and Philippe Duchateau

Abstract Over the past 15 years, site-directed genome modification has proven to be an efficient and robust approach. Meganucleases, sequence-specific endonucleases with long recognition sites, represent one of several tools described in this book that can be used for this purpose. This chapter will review the early stages of the technology (with the first strand break-induced gene targeting with I-SceI), and describe the recent advances in protein engineering that have led to the making of tailored meganucleases. We will then summarize the data and strategies available today regarding their use for site-directed genome modification. In the last section, we will discuss the latest data available on Transcription activator like effectors proteins as they have recently emerged as a promising new tool for genome modifications.

Keywords I-SceI • Meganuclease • Homing endonuclease • TALEN • Gene modifications

6.1 Introduction

For thousands of years, human beings have domesticated and improved plants and animals to their own advantages. This could be achieved through iterated crossing and/or selection for a particular phenotype. Today, with the recent advances in

J.C. Epinat • G.H. Silva • F. Pâques • P. Duchateau
CELLECTIS SA, 8 rue de la Croix Jarry, 75013 Paris, France
e-mail: jean-charles.epinat@cellectis.com; george.silva@cellectis.com; Frederic.paques@cellectis.com; philippe.duchateau@cellectis.com

J. Smith (✉)
Cellectis Therapeutics, 8 rue de la Croix Jarry, 75013 Paris, France
e-mail: julianne.smith@cellectis.com

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*, Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_6,

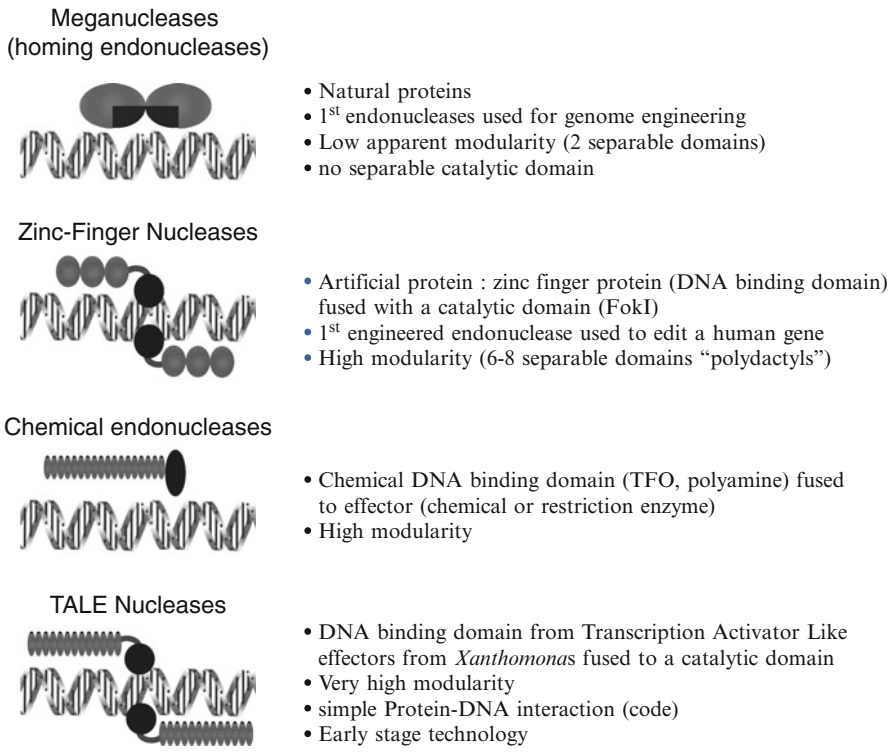


Fig. 6.1 Different classes of endonucleases used for genome engineering purposes

sequencing technology, complete genome sequences of many microbes, plants, animals and humans are now available, and reverse genetics is entering its golden age. Thus new tools to introduce genetic modifications in complex organisms have become a necessity. Although modifying genomes remains an arduous task, the past 10 years have seen the emergence of powerful technologies based on double-strand break (DSB)-induced recombination which seem to fulfill this need. Many demonstrations throughout the literature have shown the robustness of this targeted approach. The concept of targeted genomic modification has paved the way for building custom made molecules able to induce specific double-strand breaks triggering homologous recombination or non-homologous end joining (NHEJ) at the desired locus. Several families of proteins including natural endonucleases as well as artificial enzymes have been studied and developed to reach this goal. Endonucleases for genome engineering can be divided into four groups (Fig. 6.1): chemical nucleases, zinc finger nucleases (ZFNs), meganucleases and the recently added transcription activator like nucleases (TALEN's).

Chemical endonucleases resulting from the fusion of DNA-reactive agents to a DNA-binding polymer such as polyamide or triplex-forming oligonucleotides (TFO) provide the simplest solution (for review see Schleifman et al. 2008). Because the sequence specificity of such polymers can be designed *a priori*, it eliminates the extensive screening process usually necessary for the identification of new peptidic DNA-binders. A variety of active domains such as topoisomerase inhibitors (camptothecin) (Arimondo et al. 2006), psoralen (Majumdar et al. 2008), bypiridine (Simon et al. 2008) or restriction enzymes (Eisenschmidt et al. 2005) in fusion with TFOs have been described as potential sequence-specific tools for genome engineering. However, efficiency of endogenous gene modifications, safety issues as well as intracellular delivery still need to be addressed for therapeutic applications. Zinc finger nucleases are the most widely used tool in the field of genome engineering but will not be addressed here as a special chapter of this book describes in detail this technology (see Chap. 5). This chapter will focus on meganucleases, providing the historical landmarks at the origin of the targeted approach for genome engineering as well as recent advances made in meganuclease engineering and their use. Finally, we will discuss the recent data available in the literature regarding the TALEN technology which holds the potential to revolutionize the field of genome manipulation.

6.2 Historical Considerations

At the end of the last century (late 1970s, early 1980s), the baker's yeast *Saccharomyces cerevisiae* has been an essential tool for studying homologous recombination (HR). The basic knowledge obtained using this model organism has led to a better understanding of the mechanisms of genetic recombination and DNA damage repair. From studies on mating type switching (Haber 1998), the reassortment of alleles in meiosis (Roeder 1997), DNA repair (Paques and Haber 1999; Rothstein 1983), and "homing" of classI introns and inteins to novel alleles (Kostriken et al. 1983; Jacquier and Dujon 1985), it has been shown that HR is a conserved DNA maintenance pathway. Integrity of the genetic material in a cell after a double-strand break or other DNA lesion is maintained by a specific machinery that uses DNA sequences sharing homology to repair the lesions (Paques and Haber 1999; Rothstein 1983). These findings provided the basis for the first gene targeting (GT) experiments, allowing the modification of a chosen locus instead of relying on random insertion/modification and thus opening the field of genome engineering (Hinnen et al. 1978; Orr-Weaver et al. 1981, 1983; Rothstein 1983; Szostak et al. 1983). If HR uses endogenous sequences that share homology to promote the exchange of genetic information; GT will "introduce" exogenous DNA material to promote exchange between an endogenous chromosomal sequence and donor DNA. By designing special DNA matrices, genes could be knocked-out or knocked-in, sequences could be replaced, corrected or mutated in a rational, precise and efficient manner. It has been shown that only a few hundred base pairs of homology between the targeting

construct and the targeted locus were needed for the process to occur (Hinnen et al. 1978), and that significant stimulation occurred by free DNA ends in the targeting construct (Orr-Weaver et al. 1981; Szostak et al. 1983; Paques and Haber 1999). However, the desired event was not always easy to obtain, due to the low frequencies with which the targeted events occur.

Once the first demonstrations in a lower eukaryote cell like yeast were obtained with a high level of success, the same kind of experiments were performed in mammalian cells over the next 10 years (Doetschman et al. 1987; Koller and Smithies 1989; Mansour et al. 1988; Thomas and Capecchi 1987, 1990). However, a major difference in efficacy was immediately observed. While 100% of selected yeast transformants show targeted events, only 10^{-2} to 10^{-3} of the transformants in ES cells displayed the desired genotype, or basically 10^{-6} to 10^{-9} of treated mammalian cells (Fig. 6.2) (Doetschman et al. 1987; Koller and Smithies 1989; Mansour et al. 1988; Thomas and Capecchi 1987, 1990). Gene targeting is very efficient in *Saccharomyces cerevisiae*, however, it has proven to be relatively inefficient in most organisms or cell types. Only *Trypanosoma brucei* (Blundell et al. 1996; ten Asbroek et al. 1990), the moss *Physcomitrella patens* (Schaefer and Zryd 1997; Schaefer 2001), the DT40 avian lymphoid cell line (Bezzubova et al. 1997; Buerstedde and Takeda 1991), and a few modified *E. coli* strains (Murphy 1998; Poteete 2001), display GT proficiencies very similar to that observed in *S. cerevisiae*.

To overcome this biological barrier, people used selection methods or screening procedures to retrieve rare targeted recombination events among the large majority of non-targeted integrations (Bronson and Smithies 1994). This approach initiated a new era in mouse genetics and allowed a large number of mouse gene knock-outs to be obtained by GT in ES cells (Bronson and Smithies 1994; Capecchi 2001; Koller and Smithies 1989; Smithies 2001).

These findings were the source of further developments that define the field of genome engineering by GT as we use it today. The term gene targeting includes alternatives for inserting, deleting or modifying chosen sequences at a very precise locus on a chromosome of any living cell type. These events can occur by spontaneous HR or by "induced gene targeting", meaning that the desired HR event is triggered by a molecule that promotes natural cellular HR mechanisms (Fig. 6.2). Sequence-specific endonucleases induce DNA breaks at a particular locus, thus triggering HR at that exact position. Various proteins are designed for such purposes, allowing downstream targeted gene insertion, gene deletion or gene inactivation and modification, with a wide range of applications (see below).

6.3 The I-SceI Success Story

In the 1980s, it was the discovery of two yeast mobile elements that started the (meganuclease) endonuclease story. Namely, HO, which is responsible for mating type switching (Kostriken et al. 1983) and I-SceI, which is encoded by the mitochondrial genome (Jacquier and Dujon 1985). Both proteins have an endonuclease

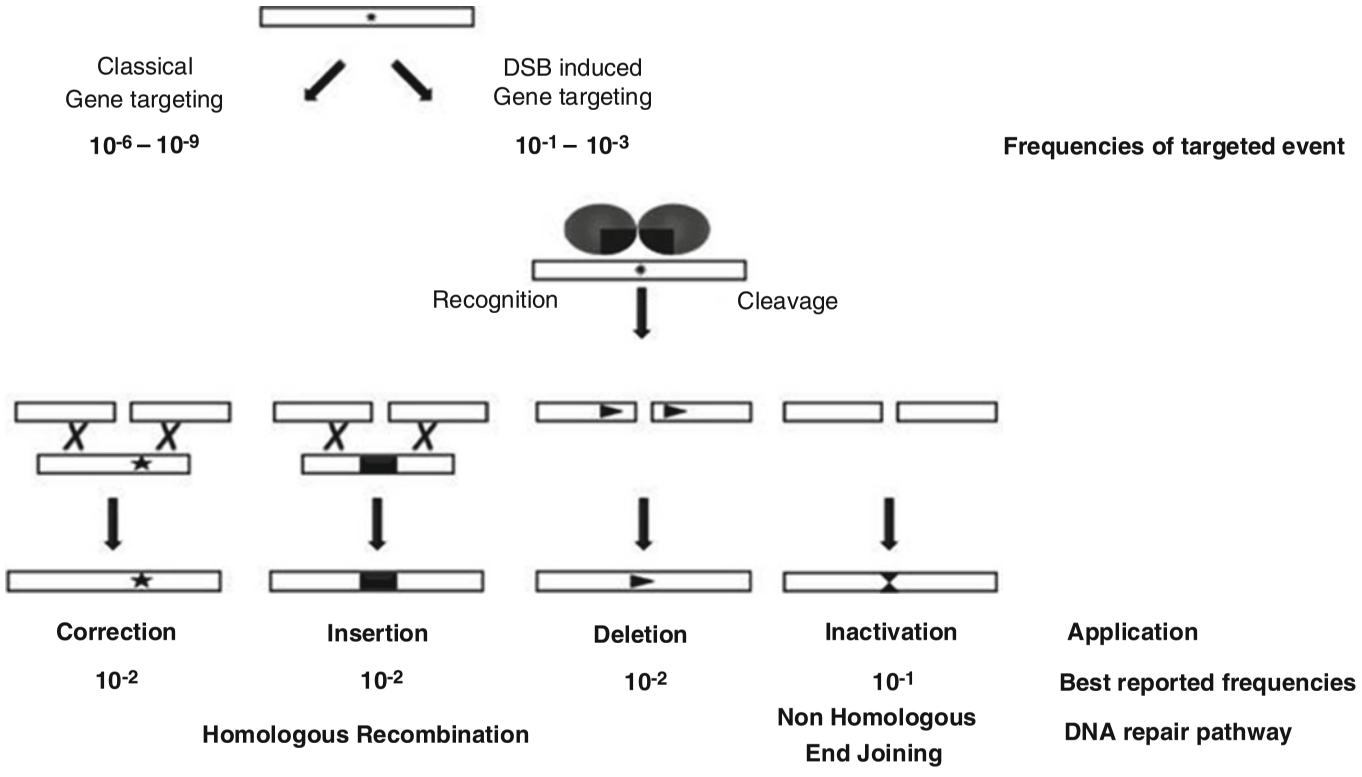


Fig. 6.2 Gene targeting and double-strand break induced gene targeting: Introducing a modification at a particular locus in a genome can be achieved by classical gene targeting (*upper left panel*) with a pretty low chance of success. However, introducing a double-strand break at that locus increases the frequencies by a 2–3 fold factor (*upper right panel*). After the cleavage, DNA repair mechanisms can produce various outcomes. Gene correction or insertion can be obtained by using a DNA repair matrix containing sequence homologies surrounding the targeted region. If the break occurs between tandem repeats, a deletion of one repeat and the intervening sequences can occur. Finally, without homologous sequences, error-prone non-homologous end joining can result in deletions or even insertions, leading to gene inactivation. Some frequencies are also reported and indicate some of the best results published in the literature (see Table 6.1)

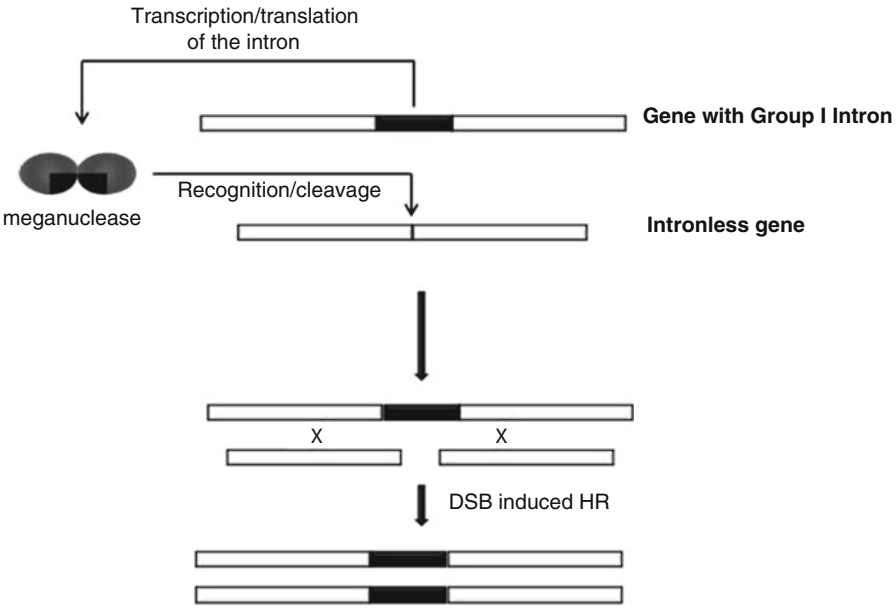


Fig. 6.3 Homing mechanism: the group I intron of the upper gene encodes a meganuclease, when the same gene of another organel is intronless. The lateral transfer of the mobile intron to the second allele occurs when two strains mate. The intron's ORF is transcribed and translated into a meganuclease whose cleavage site represents the exon's flanking sequences. Upon cleavage, the double-strand break (*DSB*) induces a homologous recombination (*HR*) event during which the intervening sequence (*the intron*) is copied into the intronless allele, thus self propagating its own sequence

activity that is capable of initiating homologous recombination events at very precise loci. In the case of HO, the cleavage at the MAT locus induces homologous recombination and mating type switching by a gene conversion event (Haber 1998). In contrast, I-SceI is encoded by a group I mitochondrial intron and is responsible for spreading this intron into intronless copies of the same gene (Jacquier and Dujon 1985). This process has since been known as "homing" (Fig. 6.3). The recombinogenic properties of DSBs had been established shortly before (Orr-Weaver et al. 1981), but the fact that nature had evolved sequence-specific endonucleases to trigger recombination was new. The other important finding was that these proteins had a very high specificity toward large single recognition/binding/cleavage sequences within the yeast nuclear or mitochondrial genomes. Later analysis resulted in the identification of two 18pb sequences as the smallest recognition sites for HO (Nickoloff et al. 1986) and I-SceI (Colleaux et al. 1988). Compared to other classical restriction enzymes, this sequence size makes it theoretically unique in any genome (one chance out of 4^{18} bp or 6.8 10^{10} bp). It then becomes obvious, that such proteins would be the perfect tool to introduce DSBs and initiate HR events at a unique site in the genome of any cell type, especially if one was capable of designing a custom enzyme for a specific locus of interest (see later).

When the biological function of I-SceI was discovered, its distinct properties were exploited as a molecular tool in the lab. People started thinking of various experiments that could take advantage of the length of the recognition site (18pb) and of the ability of the protein to induce HR events. At the beginning, one of the early successful uses of I-SceI was its "rare cutter" restriction enzyme quality that allowed it to be used for physical mapping of large DNA fragments cloned in YACs (Colleaux et al. 1993). It was also utilized as a restriction enzyme for molecular biology applications. Learning from the natural homing process, it soon became clear that GT could benefit from triggering HR with such enzymes. I-SceI was used to provide proof of concept that a meganuclease was able to induce the insertion of an exogenous sequence at a desired locus in a chromosome. The first examples, in NIH3T3 or mouse ES cells, used a chromosomal "broken" neomycin resistance reporter gene, in which the open reading frame was interrupted by an I-SceI cleavage site. The authors measured the correction of such a construct after co-transfecting an I-SceI expression vector together with a donor DNA template (Rouet et al. 1994; Smih et al. 1995). They were able to monitor targeted correction events with frequencies ranging from 3×10^{-5} in ES cells to 4×10^{-4} in NIH3T3 and showed that the correction eliminated the 18pb I-SceI recognition site. Other similar co-transfection experiments demonstrated that the insertion of a larger neomycin marker could occur in NIH3T3 and PCC7-S mouse cells with homologous recombination frequencies of 1.8×10^{-4} to 4×10^{-4} (Choulika et al. 1995). Each time, about 10^6 cells were tested and no event was detected without I-SceI, proving that HR was stimulated by at least a 100 fold factor. Later on, higher percentages of GT were published: 1% of recombination in ES cells (Donoho et al. 1998) or 10% in 293 cells (Szczepek et al. 2007). However, one has to remember that the initial recombination frequencies mentioned here represent the absolute rate of targeted events per total number of cells that were subjected to transformation, and thus underestimate the real number of events per transfected cells (the number that is usually used today).

Another value that can be estimated which reflects the efficiency of homologous recombination : the ratio of targeted versus non targeted events. The DNA donor template very often contains a selectable marker to isolate clones that were modified by knock-in or knock-out events. Among the possible events (site-directed and random), targeted modification usually represents a minority since both are selected. Gene targeting is proven by molecular analysis. Thus, the ratio of targeted versus non-targeted is a necessary notion since it directly impacts the size of the screen that has to be performed to retrieve the clones of interest. During KI experiments described above, it has been shown that this ratio climbed to 1:10, compared to $1:10^{-2}$ and $1:10^{-3}$ for classical non-induced gene targeting (Choulika et al. 1995). Concerning gene correction experiments, this value is much more difficult to determine since random integrations are usually not detected. Nevertheless, comparing random insertion of selectable genes in control experiments, it can be estimated that gene correction represents about 10% of random insertion events (Rouet et al. 1994; Smith et al. 1995). One additional important parameter is the molecular characterization of the GT event. In early gene correction studies in NIH3T3 with I-SceI, up to two thirds of the targeted clones had additional rearrangements on one

Table 6.1 Meganuclease induced gene targeting in different cell types (frequencies are maximal reported frequencies)

Meganuclease	Cell type	Locus	GT frequency	Reference
I-CreI engineered	293 H, MRC5	Endogenous	$4.4\ 10^{-2}$	Munoz et al. (2011)
I-CreI Engineered	COS-7	HSV1 DNA	$6.6\ 10^{-1}$	Grosse et al. (2011)
I-CreI Engineered	Maize	In situ chromosome deletions	$3\ 10^{-2}$	Gao et al. (2010)
I-AniI wt and Engineered	HEK293	Integrated reporter	$3\ 10^{-3}$	Metzger et al. (2011)
DmoCre hybrid Engineered	CHO-K1	Integrated reporter	$8.7\ 10^{-1}$	Grizot et al. (2010)
I-CreI wt	CHO-K1	Integrated reporter	$7.5\ 10^{-1}$	Cabaniols et al. (2010)
I-CreI Engineered	293 H	Endogenous	$6.1\ 10^{-2}$	Grizot et al. (2009)
I-AniI wt and Engineered	293 T	Integrated reporter	$6\ 10^{-4}$	McConnell Smith et al. (2009)
I-CreI, I-MsoI	293 T	Extrachromosomal (repeats)	$3\ 10^{-1}$	Li et al. (2009)
I-CreI Engineered	CHO-K1	Integrated reporter	$5\ 10^{-3}$	Redondo et al. (2008)
I-CreI Engineered	CHO-K1	Integrated reporter	$1.7\ 10^{-3}$	Arnould et al. (2007)
I-CreI Engineered	CHO	Extrachromosomal	$1.8\ 10^{-2}$	Arnould et al. (2006)
DmoCre hybrid	COS-7	Extrachromosomal (repeats)	nd	Epinat et al. (2003)
I-CreI	Drosophila	Integrated reporter	$5.7\ 10^{-1}$	Rong et al. (2002)

side of the break (Rouet et al. 1994). This same observation was made in ES cells, however this type of targeted insertion accounted for only 21% of the events (Smih et al. 1995). Similarly, data obtained by classical (non-meganuclease induced) GT indicated that targeted insertion could display "one-sided" events where one side of the break had a homologous junction and the other one had a non-homologous junction (Smih et al. 1995). It thus seems important to take this fact into account in any gene targeting approach.

As soon as the meganuclease-induced GT frequencies broke the % barrier, the technology invaded the laboratories and has been rapidly adopted worldwide. In today's literature, reports of 10% or more of nuclease induced GT events have been made (Table 6.1). These techniques, not only paved the way for genome engineering, but also allowed a much better understanding of DNA repair mechanisms in a variety of cell types and experimental conditions (Silva et al. 2011).

6.3.1 Further Developments with I-SceI

Soon after the publication of the pioneering results described above, several other academic laboratories confirmed that I-SceI was indeed efficient at triggering site-directed HR events in mammalian cells (Table 6.2) (Rouet et al. 1994; Smih et al. 1995;

Table 6.2 I-SceI induced gene targeting in different cell types (frequencies are maximal reported frequencies)

Cell type	Locus	GT frequency	Reference
Trypanosome	Integrated reporter	$8.7\ 10^{-1}$	Glover (2008)
CHO	Extrachromosomal	$3\ 10^{-3}$	Arnould et al. (2007)
transgenic mouse	Integrated reporter (repeats)	$1.3\ 10^{-2}$	Gouble et al. (2006)
HEK293	Integrated reporter	$3\ 10^{-3}$	Radecke et al. (2006)
HEK293	Integrated reporter	$1.1\ 10^{-2}$	Porteus (2006)
HEK293	Integrated reporter	$1\ 10^{-1}$	Alwin et al. (2005)
HEK293	Integrated reporter	$5\ 10^{-2}$	Porteus and Baltimore (2003)
HT-1080	Modified retrovirus DNA	$1\ 10^{-2}$	Miller et al. (2003)
Zebrafish	Extrachromosomal	$3\ 10^{-1}$	Thermes et al. (2002)
Mouse ES cells	Integrated reporter	$4.7\ 10^{-4}$	Elliott et al. (1998)
Mouse ES cells	Integrated reporter	$6\ 10^{-6}$	Cohen-Tannoudji et al. (1998)
Mouse ES cells	Integrated reporter	$1\ 10^{-2}$	Donoho et al. (1998)
Mouse ES cells	Integrated reporter	$3\ 10^{-4}$	Richardson et al. (1998)
Mouse ES cells	Integrated reporter	$1.1\ 10^{-5}$	Moynahan and Jasin (1997)
Tobacco seeds	Extrachromosomal	$2\ 10^{-3}$	Puchta et al. (1996)
Mouse ES cells	Integrated reporter	$3\ 10^{-5}$	Smih et al. (1995)
NIH3T3 and PCC7-S	Integrated reporter	$4\ 10^{-4}$	Choulika et al. (1995)
NIH3T3	Integrated reporter	$4\ 10^{-4}$	Rouet et al. (1994)

Choulika et al. 1995, Elliott et al. 1998; Cohen-Tannoudji et al. 1998; Donoho et al. 1998; Porteus and Baltimore 2003; Porteus et al. 2003; Miller et al. 2003; Alwin et al. 2005; Porteus 2006; Radecke et al. 2006). The most striking observation from these data is that targeting frequencies range from 10^{-1} to $6x10^{-6}$. However, besides the fact that the I-SceI target site is the same and has been introduced into the genome before performing the actual experiment, all other parameters are different. The insertion locus of the target site (also referred as "landing pad"), reporter system, cell type, selection method, expression system, repair matrix or correction/insertion scheme; were different. It is thus a hard task to understand which of those factors impacted most the final outcome. Nevertheless, despite the difference in the frequencies observed, all of these publications describe significantly increased GT events after meganuclease induction.

Once promising recombination frequencies could be obtained in mammalian cells, the question of I-SceI activity in differentiated tissues could be addressed. It was almost 20 years after the first experiments in cell lines that the first use of a meganuclease was reported in a mammalian animal model. A beta-galactosidase reporter cassette containing two tandem repeats separated by an I-SceI cleavage site was introduced into the mouse genome (Gouble et al. 2006). This non-active construct was stable over several generations. High levels of I-SceI induced recombination could be observed in liver (up to 1.3% of hepatocytes) upon tail-vein injection of an adenovirus expressing the protein. Even if the single-strand annealing (SSA) pathway for

HR that is used to recombine tandem repeats in a cell differs from the gene conversion mechanism (used to resolve gene insertion experiments for example), this study demonstrated the potential of the technology. The efficiency of induced events was in the range of what one would expect for therapeutic applications.

A large number of publications also refer to I-SceI-induced recombination events as a robust and efficient tool in non-mammalian organisms as well. These events include various modifications such as mutagenesis, recombination between repeats or GT in bacteria (Suzuki et al. 2005; Cox et al. 2007; Flannagan et al. 2008; Horzempa et al. 2010; Posfai et al. 1999), GT and various other DSB-induced recombination events in plants (Puchta 1999a, b, 2002; Puchta et al. 1996; Siebert and Puchta 2002; Yang et al. 2009), in mosquito (Nolan et al. 2011; Windbichler et al. 2007, 2011), and in Drosophila (Rong and Golic 2000, 2001, 2003; Rong et al. 2002; Maggert et al. 2008). The plants studies were particularly important for the community since GT in plants was considered to be up to 1,000 times less efficient than in mammalian cells and percentages of GT events per transformed cell were reported with I-SceI (Puchta et al. 1996). These studies also demonstrated the usefulness of I-SceI mediated SSA in vivo for marker removal (Puchta 2002).

I-SceI has also been used to improve transgenesis efficiencies in various organisms. In fish, to improve the mosaic distribution of transgenic cells and the subsequent low frequency of transgenic animals, people flanked the transgene of interest by two I-SceI cleavage sites. Co-injection of the transgene construct together with the I-SceI enzyme into medaka embryos at the one-cell stage increased the number of F0 embryos with uniform promoter-dependent expression from 26% in control experiments (without I-SceI) to 76%. The transgenesis frequency was increased to 30.5%, while the germline transmission rate jumped from a few percent to 50% in most lines after meganuclease co-injection (Grabher and Wittbrodt 2008; Thermes et al. 2002). Similarly, in frogs, 30% (in Xenopus tropicalis) or 20% (in Xenopus laevis) of injected embryos exhibited non-mosaic, promoter-dependent transgene expression and transgenes from the founder animals were efficiently transmitted to offspring (Ogino et al. 2006a, b; Loeber et al. 2009). This approach has also permitted efficient transgeneis in sea urchin (Ochiai et al. 2008) and the first report of germline transmission for sea anemone (Renfer et al. 2010). In fly, HR is used to overcome the "size" limitation of transposon-mediated transgenesis. It has been shown that I-SceI induced gap repair permits the efficient introduction of a 30-kb DNA segment into a transgene and demonstrates that HR may be coupled to NHEJ (Takeuchi et al. 2007). Finally, other meganucleases, like HO and I-CreI, were also used to induce intrachromosomal recombination of inverted repeats in Arabidopsis (Chiurazzi et al. 1996) and of direct repeats in Drosophila (Rong et al. 2002) and mosquito (Traver et al. 2009).

The extensive results that have been published since the early eighties illustrate the potential of meganucleases for genome engineering. I-SceI, as the prototype of such enzymes, was the real pioneer for all the applications that have been developed so far in genome engineering, whatever the type of nuclease used to generate the "magic" cleavage that triggers gene targeted homologous recombination at the desired locus. From those first studies, all the main strategies used today were founded:

sequence insertion (Choulika et al. 1995, Donoho et al. 1998; Pierce et al. 2001), sequence correction (Rouet et al. 1994; Taghian and Nickoloff 1997), sequence excision as well as gene inactivation (Liang et al. 1998). The two homologous recombination mechanisms (gene conversion (GC) versus single-strand annealing (SSA)) were better understood. The SSA mechanism is known to be highly efficient in mammalian cells to induce recombination between direct or inverted repeats. The later phenomenon is widely used today for maker excision or gene removal (Schmidt-Puchta et al. 2004; Traver et al. 2009; Rong et al. 2002). Alternatively, gene inactivation by single or multiple cleavages can be achieved by a NHEJ mechanism. Thus, knock-in or knock-out experiments, but also tandem repeat excision have benefited from better designs to be more efficient.

6.3.2 Meganucleases

The term meganuclease was coined to designate sequence specific endonucleases recognizing large DNA targets (>12 bp) (Thierry and Dujon 1992). While this very broad definition based on the nature of the substrate includes artificial endonucleases such as zinc-finger nucleases (ZFNs) and TALE nucleases (TALENs), it has progressively become synonymous in the literature with naturally occurring homing endonucleases (HEs). This section presents an overview of how the use of natural and engineered HEs has evolved, highlighting the current successes and future prospects of meganucleases.

6.3.2.1 Natural Homing Endonucleases

In the case of many group I and group II introns, and several archaeal introns, the introns contain an open reading frame (ORF) that encodes a protein that functions to promote the mobility of the genetic element. This type of mobility, termed "homing", involves the transfer of the intron to a cognate intronless allele that contains the "homing site". The ORF-encoded protein, or "homing endonuclease", stimulates intron mobility by initiating a double-strand break in the intronless allele at or near the position of intron insertion. By analogy to exons and introns, protein sequences called exteins can contain intervening protein sequences called inteins that are auto-catalytically excised and the exteins spliced during post-translational maturation (Perler et al. 1994). Surprisingly, many inteins harbor an in-frame homing endonuclease that confers homing properties to the intein element.

The parallel existence of related homing endonuclease genes in a multitude of genetic niches (i.e. introns, inteins and intergenic regions) supports the theory that the endonuclease itself is an invasive mobile element (Loizos et al. 1996; Gimble 2000; Belfort et al. 2002). Splicing of introns and inteins can be catalyzed independently of the presence of the endonuclease ORF or its activity. As such, it is thought

that the intron or intein provides a safe haven for the invading endonuclease ORF by minimizing the impact on the host genome. This creates a symbiotic relationship between the endonuclease and the intron/intein: whereas insertion into an intron/intein ensures expression of the endonuclease, the endonuclease confers mobility to the entire intron/intein to ensure its propagation. It is believed that in some cases the ORF protein has even co-evolved the ability to act as an RNA maturase to assist in the intron splicing reaction (Belfort 2003; Lambowitz et al. 1999; Bolduc et al. 2003). Nevertheless, these mobile genetic elements are generally considered "selfish" entities since the majority identified thus far provide no apparent benefit to the host organism (Cavalier-Smith 1985; Belfort et al. 2002). Over the past 20 years, the number of identified HEs has grown to include several hundred members, found in eukaryotes, bacteria and Archaea, but with the notable exception of metazoans (Stoddard 2005). They can be divided into five families based on key sequence and structure motifs: LAGLIDADG, GIY-YIG, HNH, His-Cys box and PD-(D/E)XK (Zhao et al. 2007). Structures of specific endonucleases are available for all families as well as for non-specific nuclease members of the HNH family (Heath et al. 1997; Duan et al. 1997; Flick et al. 1998; Silva et al. 1999; van Roey et al. 2001, 2002; Bolduc et al. 2003; Shen et al. 2004; Zhao et al. 2007; Marcaida et al. 2010). Despite the differences in protein fold, HEs share in the ability to recognize long DNA targets in a sequence tolerant manner to effect a double-strand break. Yet based on this lack of structure and sequence conservation, combined with diversity in the catalytic mechanisms, it is believed that the families arose independently. Interestingly, the HNH and His-Cys box proteins share a common active site configuration regardless of the differing global protein fold; hence, it has been proposed that these families be merged into the ββα-Me family to reflect the secondary structure elements and metal ion defining the motif (Friedhoff et al. 1999; Kuhlmann et al. 1999).

The most well studied family is that of the LAGLIDADG proteins, with a considerable body of biochemical, genetic and structural work having established that these endonucleases could be used as molecular tools (Sethuraman et al. 2009; Stoddard et al. 2007; Arnould et al. 2011). LAGLIDADG homing endonucleases (LHEs) have become the scaffold of choice for designing meganucleases with novel tailored specificities for inducing double-strand breaks at individual chromosomal loci.

6.3.2.2 The LAGLIDADG Family

The majority of information regarding LAGLIDADG proteins is derived from the endonuclease members of this family. It had long been known that the defining sequence motif, LAGLIDADG, represented an essential element for enzymatic activity. Some proteins contained only one such motif, while others contained two; in both cases the motifs were followed by ~75–200 amino acid residues having little to no sequence similarity with other family members. In 1997, two groups shed

light on the architecture of LAGLIDADG proteins with the structure of both a single-motif protein (I-CreI) (Heath et al. 1997) and an intein-encoded double-motif protein (PI-SceI) (Duan et al. 1997). The endonuclease domain in each structure adopted a similar αββαββα fold. The LAGLIDADG motif comprises the terminal region of the first helix and not only contributes to a bipartite catalytic center but also forms the core subunit/subdomain interaction. Two such α/β domains assemble to form the functional protein, with the β-strands in each creating a saddle-shaped DNA binding region. These studies, in combination with the structure of the compact monomeric I-DmoI (Silva et al. 1999), definitively established that two α/β subunits were needed for function, as exemplified by I-CreI, having an overall architecture similar to that of the double-motif endonuclease domain of PI-SceI and the monomeric I-DmoI. The α/β fold is a hallmark of LAGLIDADG proteins and has since allowed for better homology modeling and structure/function prediction for proteins lacking structural data.

It is estimated that LHEs recognize comparatively long DNA target sites, ranging from 14 to 40 base pairs in length. Despite the ability to bind related sequences with the same range of affinity, LHEs exhibit relatively high DNA specificity in that the number of cleavable sites is much smaller than those bound (Gimble et al. 2003; Scalley-Kim et al. 2007; Thyme et al. 2009; Jarjour et al. 2009). Such sequence specificity is required to generate unique chromosomal breaks *in vivo* (Choulika et al. 1995). By considering the affinity at individual bases and the interdependence between consecutive bases, it has been estimated that the overall specificity of a meganuclease would be 1 in 10^9 (Argast et al. 1998; Gimble et al. 2003; Ashworth et al. 2010). The availability of co-crystal structures not only provided the first glimpse of LHE-DNA contacts (Jurica et al. 1998) but also enabled theories about the nature of catalysis (Chevalier and Stoddard 2001; Chevalier et al. 2001). The overall picture of LHE-DNA interactions can be summarized as (Chevalier and Stoddard 2001; Chevalier et al. 2001, 2003; Moure et al. 2002, 2003; Bolduc et al. 2003; Spiegel et al. 2006; Marcaida et al. 2008; Redondo et al. 2008): (i) specificity contacts arise from the burial of the extended β-strands into the major groove of the DNA, with the DNA binding saddle having a pitch and contour mimicking the helical twist of the DNA; (ii) the complete complement of hydrogen bonding potential between the protein and DNA is never fully realized, with many contacts being water-mediated; (iii) cleavage to generate the characteristic cohesive 4 nucleotide 3′-OH overhangs occurs across the minor groove, wherein the scissile phosphate bonds are brought closer to the protein catalytic core by a distortion of the DNA in the central "4-base" region; (iv) cleavage occurs via a proposed two-metal mechanism, sometimes involving a unique "metal sharing" paradigm; (v) and finally, additional affinity and/or specificity contacts can arise from "adapted" scaffolds, in regions outside the core α/β fold. Nevertheless, despite these generalities later studies have shown that related LHEs can use different subsets of residues to recognize similar DNA (Chevalier et al. 2003; Lucas et al. 2001). As discussed below, this "relaxed" property of meganucleases has proved to be a double-edge sword, making it at once both easy and problematical to re-engineer specificity.

6.4 The Challenge of Specificity Toward a Chosen Sequence

The field of meganuclease engineering has recently drawn the attention of a vast community, with complementary studies showing that meganuclease engineering can succeed at creating new molecules based on modifying and/or mixing existing enzymes. The spatial separation of the catalytic center with regions directly interacting with the DNA has allowed for specificity re-engineering either via a "bulk" approach wherein entire subunits from different proteins are combined to generate novel hybrid specificity (Epinat et al. 2003; Chevalier et al. 2002; Silva et al. 2006) or by using targeted approaches affecting key DNA-contacting residues (Seligman et al. 2002; Sussman et al. 2004; Arnould et al. 2006; Doyon et al. 2006; Rosen et al. 2006; Smith et al. 2006; Arnould et al. 2007). Figure 6.4 illustrates the evolution of engineering strategies.

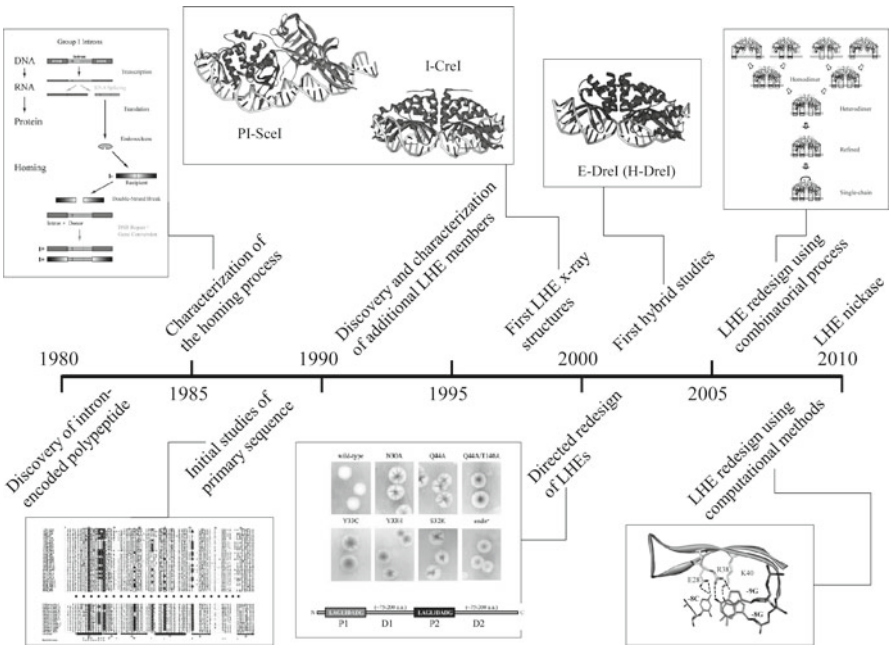


Fig. 6.4 Time-line illustrating key events in the development of engineering strategies. Upon discovery of homing endonucleases, initial studies focused on classical genetic methods to ascertain functional details. X-ray crystal structures subsequently corroborated these genetic analyses and provided a more direct method by which protein engineering could be targeted. Whereas the first hybrids validated the conceptual separation of protein monomers into semi-independent subdomains, the first true "whole protein" re-engineering was achieved via a combinatorial approach. Recently, computational methods have further advanced the targeted engineering of protein/DNA contacts

6.4.1 Hybrids

Several groups tested the concept of subunit interchangeability by fusing both similar and disparate α/β domains from distinct unrelated LAGLIDADG endonucleases. They succeeded at giving rise to chimeric proteins having hybrid specificity derived from each half of the parental protein targets (Chevalier et al. 2002; Epinat et al. 2003; Silva et al. 2006). Initial reports by two separate groups involved using I-CreI and I-DmoI as the building blocks. Whereas Chevalier and coworkers used extensive computational analysis to re-model the protein/protein interface, Epinat and colleagues used only structure modeling to create their I-DmoI/I-CreI hybrid protein. The majority of changes at the heterodimer interface involved amino acid residues not part of the LAGLIDADG region. The resulting artificial, highly specific DNA endonucleases were created by fusing the N-terminal domain of I-DmoI to the I-CreI monomer. Along the same lines, Silva and co-workers used the N-terminal domain of I-DmoI to not only illustrate homodimerization and activity of an H-DmoA "monomer" but also to create a novel, highly specific monomeric endonuclease. While active as is, H-DmoA was re-engineered and fused to itself to generate H-DmoC2, which both retains the thermostable nature of its parent protein and acquires a comparatively increased specificity for its target DNA.

More recently, the DmoCre hybrid created by Epinat and colleagues was further engineered to recognize modified DNA targets by independently altering the specificity of each starting domain (Grizot et al. 2010). This was achieved via a combinatorial approach (see below) using preexisting data available for the I-CreI protein as well as new specificity data for I-DmoI.

These proof-of-concept studies not only addressed the evolutionary concept that double-motif LAGLIDADG proteins arose from a gene duplication event but also demonstrated that distinct domains and subunits of such enzymes were responsible for specific recognition and binding of each DNA half-target. They furthermore highlighted the possibility of modifying the DNA specificity of a meganuclease by tackling the problem on each half of the molecule independently and then combining both "half molecules" into a single engineered protein. Gene fusion alleviates the restriction imposed by homodimers of having symmetric or pseudo-symmetric targets by allowing for asymmetric mutations in each subunit and thus recognition of asymmetric DNA targets. This model has undeniably proven valuable in paving the way for selection and screening studies based on the independence of DNA recognition by individual α/β subunits.

6.4.2 Directed Redesign of LHEs

Early strategies for engineering meganucleases involved a semi-rational approach in which specific residues are mutated on the basis of prior functional or structural knowledge to create libraries with limited diversity. Techniques include using

variations of a bacterial two-hybrid screening protocol (Gimble et al. 2003) or cleavage-induced inactivation of a reporter gene to identify mutations that altered DNA specificity (Seligman et al. 2002). Initial successes with the I-CreI (Seligman et al. 2002; Sussman et al. 2004; Rosen et al. 2006) and I-SceI scaffolds (Doyon et al. 2006; Chen et al. 2009) resulted in the identification of only a few mutants due to the limited screening technologies. A significant advance to these early efforts arose when a selection and/or functional screening approach in yeast was developed that mimicked the entire double-strand break-induced homologous recombination activity of the meganuclease (Chames et al. 2005). As noted below, this new technique paved the way for improved "cluster-based" engineering that is able to process significant mutant numbers at the functional level (i.e. altered specificity mutants having activity *in vivo*).

6.4.3 HE Redesign Using a Combinatorial Process

To analyze a much larger number of mutant/target combinations, an automated high-throughput screening (HTS) platform was developed for the I-CreI scaffold (Arnould et al. 2006). This engineering process relies on generating I-CreI mutants with modified specificity toward nucleotide triplets of the I-CreI DNA target (for detailed nomenclature, see Arnould et al. 2006). These mutants, termed building blocks or modules, are then combined in a second step to target a modified "half-site" DNA sequence. Finally, two such mutant I-CreI monomers are further combined to create the final heterodimer (or fused monomer) recognizing an asymmetric target sequence. Series of modules were also constructed to focus on base pairs at positions ±7 and ±6 (Grizot et al. 2011). It is estimated that by using such engineered I-CreI mutants, one 22 bp sequence every 300 bp can be targeted with a 40% success rate (Galetto et al. 2009; Paques and Duchateau 2007). Since a target site is composed of sub-sites that are cleaved by the building blocks, this figure can be refined depending on the desired site one would like to cleave. Given the qualification of those blocks, "classes" of targets can thus be defined with a corresponding predicted sequence space, success rate and time of engineering.

Combined with semi-rational techniques, HTS has been successfully and extensively used to change the entire specificity of a meganuclease for a chosen DNA sequence (Arnould et al. 2006, 2007; Chames et al. 2005; Smith et al. 2006; Grizot et al. 2009). Such engineered I-CreI variants were shown to successfully induce gene targeting events in mammalian cells (Redondo et al. 2008; Prieto et al. 2008; Grizot et al. 2009). These mutants were also shown to maintain the essential properties of the initial scaffold (i.e. proper folding and stability, cleavage efficiency and a narrow specificity). Although as mentioned above the sequence space that can be targeted by meganucleases is a crucial parameter for such approaches, the design of I-CreI variants appears more precise and faster every day. A dramatic shortcut was recently implemented to allow the creation and testing of "final" redesigned I-CreI variants in only a few weeks (our unpublished data). This process takes advantage

of a large database of biological material and results that include the sequence and cleavage profiles of more than 50,000 proteins. This so-called Omegbase contains data not only for building blocks but also for all homodimers and heterodimers derived from such blocks. Using this resource a process termed the Quick Meganuclease Access Program (QMAP) was recently implemented wherein meganucleases can be rapidly synthesized and validated with success rates close to 80% for targets statistically found every 300 kb. QMAP meganucleases have already been generated and proven efficient in yeast, mammalian cells, and plant cells (our unpublished data).

6.4.4 Computational Redesign

Recent advances in computational redesign software have made the future of meganuclease engineering quite promising. Computational studies have given rise to not only specificity re-engineering (Ashworth et al. 2006), but have also addressed the dimerization issue via targeting protein-protein interactions of the subunits (Fajardo-Sanchez et al. 2008) as well as through mimicking nature with single-chain constructs (Grizot et al. 2009, Li et al. 2009) . Exhaustive analysis of the I-AniI scaffold (Jarjour et al. 2009; Thyme et al. 2009) has led to better computational methods that can exploit binding energy for designs (Thyme et al. 2009). Through both computational (Ashworth et al. 2010) and structural (Redondo et al. 2008) analysis of novel structures of engineered proteins in complex with DNA, insight has been gained into the intricacies of the protein-DNA contacts that appear to involve substantial conformational changes in both the protein and/or DNA. Indeed, these data suggest that specificity engineering is more than a point-by-point process, and that altering specificity and/or activity is highly context dependant. The notion of using "residue clusters" to modify specificity, first successfully exploited via the HTS methods discussed above, is now being applied computationally to meganucleases (Ashworth et al. 2010; Ulge et al. 2011). While the pseudo-uniform nature by which LHEs contact DNA (i.e. using a subset of residues from the extended β-sheet surface) makes this approach promising, evidence suggests that a combination of HTS data and computational modeling could be sufficient for reliable *in silico* prediction and re-engineering of LHE specificity.

6.5 Cleavage Versus Nickase Activity

All known LHEs analyzed to date act as "cleavases" to cut both strands of the target DNA, with one catalytic center providing the two "semi-independent" active sites to individually cleave the respective strands. The notion of one vs. two active sites in the catalytic center is complicated by the symmetric (e.g. I-CreI and other homodimeric LHEs) or pseudo-symmetric (e.g. I-SceI and other monomeric LHEs) nature of the protein scaffold. Initial studies of the homodimeric I-CreI supported a type of

"inseparable" catalytic center wherein two active sites share a metal ion necessary for catalysis (Chevalier et al. 2001). However, later studies of the monomeric I-SceI illustrated an asymmetric active center (Moure et al. 2003), hinting that although LHEs share a common core architecture the catalytic mechanism can vary. As such, recent progress has been made in generating "mega-nickases" that specifically cleave only one strand of the DNA target (Niu et al. 2008; McConnell Smith et al. 2009). The key to generating meganickases from meganucleases lies in residues thought to indirectly support DNA catalysis: with few exceptions, a semi-symmetric pair of lysine residues flank the catalytic core. Nui and colleagues (Niu et al. 2008) demonstrated that by changing the nature of these residues in the monomeric I-SceI, it was possible to shift the preference in DNA cleavage to favor either the top or bottom strand (with the eventual second-strand break arising after prolonged incubations). Along the same lines, McConnell Smith and co-workers (McConnell Smith et al. 2009) used I-AniI to create a meganickase, however, in their case one variant displayed nearly complete preference for single-strand cleavage. Such enzymes can in principle provide similar levels of targeted induced HR with a minimization in the frequency of NHEJ (Metzger et al. 2011).

6.6 Customized Meganucleases for Genome Engineering

Homing endonucleases have the ability to efficiently induce a unique targeted double-strand break within the context of a complex genome. As double-strand breaks have been shown to be efficient inducers of DNA repair and recombination this property has been used to greatly facilitate many different types of genome engineering processes. As described earlier in this chapter, the natural homing endonuclease I-SceI has been successfully used in several different organisms to promote a wide variety of events including transgenesis (Ochiai et al. 2008; Ogino et al. 2006; Pan et al. 2006; Renfer et al. 2010; Thermes et al. 2002), gene correction (Rouet et al. 1994; Taghian and Nickoloff 1997), gene insertion (Choulika et al. 1995, Donoho et al. 1998; Pierce et al. 2001), targeted mutagenesis through a NHEJ mechanism (Liang et al. 1998) as well as the deletion of DNA sequences between direct repeats by a SSA mechanism (Siebert and Puchta 2002; Perez et al. 2005; Liang et al. 1998; Donoho et al. 1998). However, the use of I-SceI or other natural meganucleases presents an obvious drawback which is the need to first introduce the cleavage site into the region of interest. This can be a limiting step for certain applications and/or organisms. Thus, engineered meganucleases, designed to target pre-exisiting sites, represent the ideal solution. Extensive characterization of an I-CreI derived meganuclease cleaving the RAG1 gene has demonstrated that a fully designed meganuclease can be used to induce 6% targeted homologous recombination in HEK293 cells with a relative specificity that was comparable to I-SceI (Grizot et al. 2009). Since this initial publication a large number of engineered meganucleases have been successfully generated for specific genomic loci and have been shown to result in similar high levels of targeted recombination (1–12%) in

several different immortalized cell lines (Daboussi et al. 2012 and unpublished results). Thus, engineered meganucleases represent an effective as well as versatile alternative to natural meganucleases for genome engineering purposes.

6.6.1 Gene Inactivation

Genome engineering techniques applicable to reverse genetics represent indespensible tools in the post-genomic era. Among these techniques, gene inactivation represents one of the most important applications. Classically gene inactivation is achieved through the insertion of a transgene coding for a selectable marker by spontaneous homologous recombination. This approach is routinely used in mouse ES cells where spontaneous homologous recombination is relatively efficient, however it has been particularly difficult to use in ES cells from other organisms as well as other types of stem cells. The use of sequence specific nucleases in these cell types has greatly facilitated gene targeting. Lombardo et al. were the first to demonstrate that efficient targeting could be obtained in CD34+ cord blood progenitor cells (Lombardo et al. 2007). Using ZFNs and non-integrative lentiviral vectors as a repair matrix, frequencies of up to 0.1% were observed with GFP while the use of a selection strategy (PuroR) permitted targeting of up to 90% of colony-forming cells after differentiation. A similar strategy has also been successfully used in iPS cells (Hockemeyer et al. 2009; Zou et al. t, 2011). As in CD34+ cells, the absolute targeting frequency when measured, was relatively low, 10^{-5} at the PIG-A locus (Zou et al. 2009), however, when a selection strategy was used, up to 75% of clones correctly targeted could be obtained.

Targeted mutagenesis by a NHEJ mechanism represents an alternative to gene inactivation based on homologous recombination. In this approach, meganucleases can be used to generate site-specific DSBs that can then be rejoined in a seamless matter by simple re-ligation of the cohesive ends or, alternatively, deleterious insertions or deletions of various sizes can occur at the breaks, eventually resulting in gene inactivation. This approach may be simpler than those based on homologous recombination since there is no need for a repair plasmid and efficacy is likely less cell-type dependent since NHEJ appears to be active throughout the cell cycle (Aten et al. 2004). In immortalized mammalian cells targeted mutagenesis based on NHEJ has been used to inactivate single or even multiple genes (Liu et al. 2010; Cost et al. 2010). Meganucleases targeting the *LIGUELESS1* locus in maize were shown to be able to generate indels at a frequency of 3%, with both monoallelic and biallelic mutations being detected (Gao et al. 2010). Similarly, using ZFNs, mutations frequencies of greater than 4% were detected at the endogenous SuRA locus in tobacco (Maeder et al. 2008). Gene inactivation by targeted mutagenesis has also been successful in rats (Geurts et al. 2009; Mashimo et al. 2010), fish (Meng et al. 2008; Doyon et al. 2008), frogs (Young et al. 2011), pigs (Hauschild et al. 2011) and rabbits (Flisikowska et al. 2011), thus providing new perspectives for organisms in which classical HR-based methods have proven ineffective.

Targeted mutagenesis has also been successfully used to inactivate viral genomes. An engineered ZFN has been shown to specifically cleave a target plasmid that

contains the hepatitis B virus (HBV) genome (Cradick et al. 2010). Analysis of the DNA target reveals misrepair of affected HBV sequences within cultured cells coupled with a reduction in pre-genomic viral RNA levels. Similarly, a meganuclease has been shown to efficiently target an integrated reporter lentivirus suggesting that this technology may allow the inactivation of integrated HIV provirus within latently infected cells (Aubert et al. 2011). Finally, an activity with wild-type replicating virus has been demonstrated with meganucleases targeting the herpes simplex virus 1 (HSV1) genome, reductions in the levels of viral genomic DNA of up to 70% were observed at low and moderate multiplicities of infection (Grosse et al. 2011).

6.6.2 Gene Modification by Site-Directed Mutagenesis

Although gene inactivation is the most common approach to study gene function, certain studies may require the introduction of point mutations or small alterations. This is often the case in cancer biology where the phenotype of interest is the consequence of a point mutation. Alternatively this approach may be used to create a disease model or inactivate specific functional motifs. Engineered meganucleases can be efficiently used to introduce small genetic alterations directly at the cleavage site via homologous recombination. The RAG1 meganuclease has been shown to promote efficient insertion of 10 bp of exogenous DNA containing a HindIII site; up to 4% of 293 cells were effectively targeted (Grizot et al. 2009). Similarly ZFNs have been used to insert a restriction site polymorphism at the IL2RG locus; up to 20% of the IL2RG alleles could be targeted in a pool of K562 cells (Urnov et al. 2005). More recently, point mutations associated with Parkinson's disease have been inserted at or directly adjacent to a ZFN induced DSB in both hES and iPS cells at frequencies approaching 1% of transfected cells (Soldner et al. 2011).

However, the introduction of mutations that are not directly adjacent to the DSB may prove more difficult since for the double-strand break repair (DSBR) pathway the conversion efficiency of sequences adjacent to the DNA break decreases as the distance from the DSB increases. Elliott and colleagues observed a very sharp decrease of the conversion efficiency with I-SceI: a polymorphism located at a distance of 100 bp from the DSB was corrected in only 13–16% of the repair events, and by 400 bp correction dropped to only 3%. (Elliott et al. 1998). Donoho and coworkers, however, reported longer conversion tracts, with up to 13% correction at a distance of about 4,000 bp from the DSB (Donoho et al. 1998). Experiments in the human cell line HEK293 using the RAG1 meganuclease have shown that a mutation located 265 bp from the DSB could be efficiently (66%) corrected (Munoz et al. 2011). In plants, a study in tobacco using ZFNs indicate that frequencies of greater than 2% could be obtained for the insertion of a point mutation at a distance of more than 1.3 kb from the cleavage site (Townsend et al. 2009). Despite the discrepancies in these results, likely arising in part from experimental design differences, the trend in conversion efficiency observed as a function of distance remains an important consideration to keep in mind for gene targeting approaches aimed at the introduction of small modifications.

6.6.3 Gene Insertion

Meganucleases can also be used for targeted gene insertion. Various modifications of endogenous genes can be greatly facilitated by gene targeting. For example, marker genes can be inserted under the control of endogenous promoters to study expression patterns during development or to create cell-type specific lineage reporters. Alternatively, endogenous genes can be tagged with fluorescent markers and thus allow protein localization studies. Finally, meganucleases can be used to generate stable cell lines in which a gene of interest has been integrated at a predefined locus. This is particularly useful for generating isogenic cell lines that can be used for comparative studies or for drug screening. Cabaniols et al. have shown that targeted gene insertion by meganuclease-induced recombination of five different genes resulted in reproducible levels of expression between the genes with stable levels of expression over a 23 week period (Cabaniols et al. 2009, 2010). In addition, the use of engineered meganucleases allows the locus to be targeted to be pre-defined, which can be of interest for protein production where loci permitting high levels of expression can be targeted (Cabaniols and Paques 2008; Cabaniols et al. 2009).

Finally, for all strategies of genome engineering one has to keep in mind that the locus may have important effects on the efficiency of genome modification (Fig. 6.5). As described in this review, very different efficacies have been reported by different authors in reports using identical or related nucleases (I-SceI, ZFNs, I-CreI derivatives) in different conditions.

6.7 Overview on TAL Effector Nuclease (TALEN)

Transcription activator-like effector (TAL effectors) proteins have emerged recently as an alternative tool for genome modifications. Despite the fact that meganucleases and zinc-finger proteins have proven to be efficient tools for precise manipulation of the genome, one of the major limitations of these technologies is the difficulty and cost involved in their engineering. The promise of TAL effector scaffolds resides in the simplicity of the interactions between the protein and its DNA binding site, bringing this technology within reach of any laboratory. Although it is too early to appreciate the real benefit of the TALENs, this technology has the potential of revolutionizing the field of genome modification. In this section we will provide an overview of this promising technology as well as summarize the most relevant data regarding its application to the genome engineering field.

Natural TAL effectors are produced by phytopathogenic bacteria and function upon infection as transcription activators of plant genes (for review see (Bogdanove et al. 2010)). Since the isolation of the first TAL effector gene (Bonas et al. 1989), the significance of nearly identical repeat motifs within the central domain has been questioned. Even though this central domain was thought to be the key for DNA recognition, it was only recently that the link between the repeats and the DNA-binding site was fully described (Boch et al. 2009; Moscou and Bogdanove 2009), revealing a new type of DNA-binding domain (Fig. 6.6a). In general, each repeat contains 33 or 34 amino

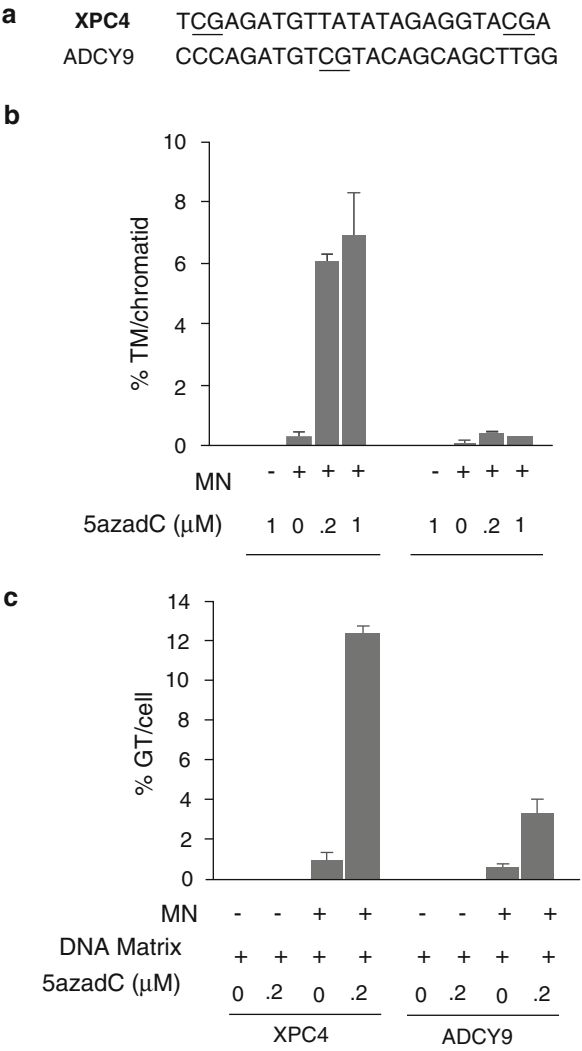
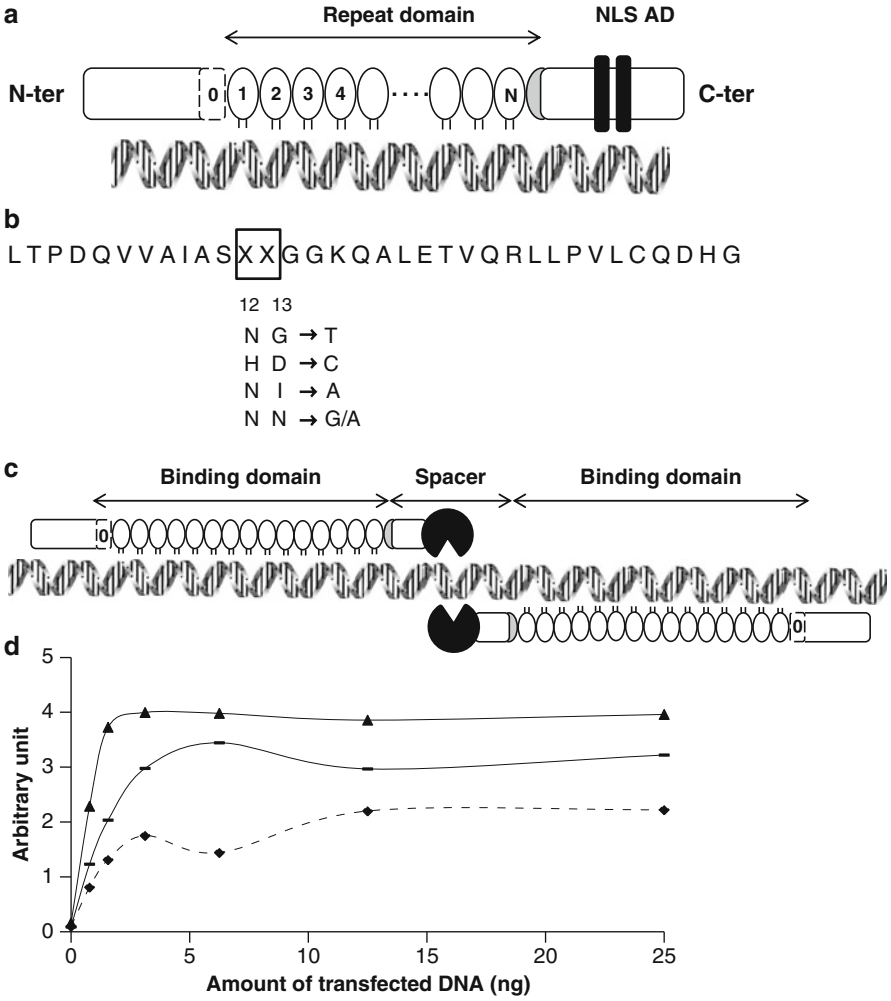


Fig. 6.5 Influence of methylation on meganuclease efficacy. (**a**) Sequence of XPC4 and ADCY9 target sequences with methylated CpGs underlined. Methylation of these sequences in HEK293 cells was determined by bisulfite sequencing to be approximately 100%. (**b**) Targeted mutagenesis (*TM*) in the presence and absence of 5aza-dC. (**c**) Gene targeting (*GT*) frequencies in the presence and absence of 5aza-dC

Fig. 6.6 (continued) Upon cleavage of the target site, tandem repeat recombination by single-strand annealing (*SSA*) restores a functional β-galactosidase gene, which can be monitored by standard assays. For each TALEN, a dose response is performed. The meganuclease I-SceI was used as positive control (*dashed line*). TALEN-induced mutagenesis or homologous gene targeting was performed. TALEN expressing vectors are transfected into HEK293 cells; the targeted locus is amplified by PCR and 10,000 molecules (on average) are sequenced to detect InDels. To monitor TALEN-induced homologous gene targeting (*HGT*) efficacy, TALENs expressing vectors are co-transfected into HEK293 cells together with a repair matrix; cells, seeded at 10 cells/well are grown in 96 well plates, and targeted alleles are detected by PCR



Locus	promoter	Mutagenesis (% Indel)	Gene insertion (%)
RAG	pEF1α	**18**	**8**
DMD	pEF1α	**15**	**7**

Fig. 6.6 TALEN architecture and activity. (**a**) Schematic of Transcription Activator Like effectors from *Xanthomona*s. The central domain is constituted of direct repeats numbered from 1 to n. This tandem repeat array is flanked by a half repeat (*grey*) and a repeat-like domain (numbered 0) mediating the specificity towards Thymidine. Two NLS domains as well as the transcription activation domain are located in the C-terminal domain. (**b**) Representative amino acid sequence of one repeat. XX shows the repeat variable diresidues (RVD) located at position 12 and 13. The relationship between RVD and nucleotide interaction is shown. (**c**) TALEN architecture. Two TAL effectors fused to catalytic domains (Fok1) bind to two nearby DNA sequences. The two binding sites are separated by a sequence spacer. (**d**) Characterization of TALEN in extrachromosomal and chromosomal assays. TALENs were designed to target specific sequences within hRAG1 (*black lines*) and hDMD genes (*black triangle*). Cleavage efficiency was tested in CHO-K1 using an extrachromosomal assay. TALEN and reporter expressing vector are co-transfected into CHO-KI cells. A specific target DNA sequence was inserted within the reporter gene and is flanked by two direct repeats.

acids and mediates the recognition of 1 nucleotide of the DNA target through two critical amino acids located at positions 12 and 13 in each repeat. These two hypervariable positions are referred to as "repeat-variable di-residue" (RVD). More than 15 different RVDs have been described today, however, HD, NG, NI, NN and NK are the most prevalent RVDs and are preferentially associated with the nucleotides C, T, A, G/A, and G respectively. Thus, the discovery of this simple code (Fig. 6.6b), where one RVD preferentially binds to one nucleotide and does not seem to be influenced by its neighboring repeat, allowed the *a priori* design of new sequential RVDs with novel DNA binding specificity (Boch et al. 2009). This finding triggered the interest of the scientific community for the TAL-effector proteins as a potential tool for genome engineering, and it didn't take long before the first TAL Nuclease (TALEN) consisting of natural or custom TAL effectors fused to the nuclease catalytic domain of the Fok1 nuclease were made (Christian et al. 2010; Miller et al. 2010; Cermak et al. 2011; Li et al. 2011a, b) (Fig. 6.6c).

However, despite the major advantage that the code brings to the engineering of a DNA-binding protein, an important drawback is that the assembly of identical repeat arrays still represents a challenge using conventional cloning techniques. Thus, new approaches had to be developed to bring RVD assembly within reach of all laboratories.

6.7.1 Methods of Assembly

In nature, transcription activator-like effector proteins contain a variable number of repeats. Arrays consisting of up to 33.5 RVDs have been described (Boch and Bonas). However, from the first reported engineered TALENs, it seems that a minimum of 10 RVDs is necessary to allow detectable cleavage activity (Christian et al. 2010). It is too early to define precisely the optimum number of RVDs needed to obtain highly active TALENs and it is likely that the number of RVDs necessary will depend on the DNA sequence one wants to cleave. At a first approximation, engineered TALENs consisting of 15–20 RVDs arrays appear to be a good starting point when one wants to synthesize a custom TALEN (Miller et al. 2010; Christian et al. 2010; Cermak et al. 2011; Li et al. 2011a, b). Even though RVD arrays could be assembled in a predetermined order using classical molecular biology techniques, the full potential of this technology would only be fulfilled by the development of an easy, cheap and fast method to assemble RVDs.

Several laboratories have reported efficient methods to assemble RVD arrays (Zhang et al. 2011; Li et al. 2011b; Morbitzer et al. 2011; Cermak et al. 2011; Miller et al. 2010; Weber et al. 2011 Geissler et al. 2011). Except for the PCR-based gene assembly method described by Miller et al. (2010), they are all based on the same approach using Type IIS restriction enzymes. These enzymes have the particularity of cleaving dsDNA outside of their binding site and leaving a 4 bp overhang. Thus, it is possible after a digestion-ligation step to remove the original binding site from the cloned DNA fragment. Most of the methods described today use this property in

order to perform restriction and ligation in a single reaction. This strategy, now referred to as the 'Golden Gate' cloning technique (Spear 2000; Engler et al. 2008) allows the assembly of up to 12 RVDS in a single restriction-ligation step (Zhang et al. 2011). In contrast, in order to facilitate high throughput methods, we have developed a sequential solid-phase cloning strategy where RVDs or stretches of RVDs are assembled one by one (unpublished data). Each method shows advantages and drawbacks and it is not our intention here to discuss in detail each of them as it would be out of the scope of this chapter. Nevertheless, all of these different cloning strategies bring the TAL technology within reach of any laboratory, allowing an engineered TALEN consisting of two custom-TAL effectors fused to a catalytic domain to be synthesized within a week.

6.7.2 Rules and Performances

Shortly after the first demonstration that targets of new TAL effectors could be predicted and custom TALEs could function as transcription factors (Romer et al. 2009) (Boch et al. 2009; Moscou and Bogdanove 2009), the first study on custom-TALENs as a reagent for genome engineering was reported (Christian et al. 2010). Using the molecular strategy used for ZFNs, i.e. a pair of TAL effectors fused to Fok1 nuclease catalytic domains, these authors could show that (i) specific custom-made endonucleases could be produced and, (ii) using extra chromosomal assays, that they were efficient to induce homologous recombination between direct repeats. The size of the spacer length between the 2 DNA binding domains was also partly addressed in this pioneer work although later studies (Miller et al. 2010) including our own unpublished data brought more light on that issue. After analysis of 20 previously reported TAL effectors (Moscou and Bogdanove 2009) and their own data; the same team also published a guideline to help design *de novo* TALEN (Cermak et al. 2011): the targeted sequence must start by a T, referred as T0, then T and A should be disfavored at position +1 and +2 respectively as G should be at position N-1 and N of each recognition site. They also found a strong bias towards the RVD-NG at the last position of the repeat arrays. Finally, the target DNA should have a low G content ($9 \pm 8\%$). The robustness of these rules is not yet established. While Miller et al. (2010) have also designed TALEN against poor G targets, we have been successful in engineering TALEN against DNA sequences containing up to 40% of G. So far, it seems that the most established limitation when one wants to design a TALEN against a chosen sequence is the presence of a T at the beginning of each target DNA sequence (T0). Actually, this essential T is not imposed by a specific RVD as it interacts directly with the N-terminal domain of the protein, and thus is not governed by the so called RVD code.

ZFNs are classically heterodimeric proteins that bind two DNA sequences separated by 6 bp. The TALENs described so far were all designed following the same architecture as ZFNs i.e. TALENs act as heterodimeric proteins in which the nuclease catalytic domain such as Fok1 is fused to the TALE C-terminal region. Thus, the optimal length of the DNA sequence separating the two binding domains

had to be determined. While ZFN DNA targets contain almost exclusively 6 bp intervening sequences, TALENs appear to tolerate a wider range of DNA lengths and as expected, appear to be dependent on the TALE scaffold used. As mentioned above, natural TAL effector proteins are made of RVD arrays that confer specificity, flanked by an N-terminal peptide sequence involved in cellular trafficking and a C-terminal domain that contains the trans-activator domain and nuclear localization sequences (NLS). Early work on TALEs has already demonstrated that 152 amino acids could be deleted from the N-terminal domain without affecting the protein activity (Gurlebeck et al. 2005). Obviously, for nuclease purposes, the trans-activation domain appears superfluous. Early studies performed by Christian et al. (2010) used two BamHI restriction sites located in the N- and C-terminal domains to truncate the protein. Without further investigation the authors were able to show that this design was effective to generate active TALEN. Furthermore using this scaffold, a spacer of 15 nucleotides was optimum, although 18 or 24 bp was also possible. In a yeast assay, this design could achieve an activity similar to that observed with ZFNs. An extensive analysis on scaffold optimization was performed by Miller et al. (2010). In this study, TAL effectors lacking the first 152 amino acids were tested in combination with C-terminal truncations on homodimeric targets bearing various spacer lengths (from 2 to 24 bps). A spacer length below 10 nucleotides did not allow efficient cleavage *in vitro* as did the C-terminal truncation bearing the 95 first amino acids of the C-terminal domain. Moreover, TALEN bearing the shortest truncation (+28 residues) showed a very good nuclease activity *in vitro* when tested on targets containing spacers from 10 to 24 bps, with a maximal activity for a spacer of 12–13 bps. Sequences that were smaller than 8 bp did not allow significant cleavage activity. It is difficult to infer any strict guidelines in view of this study. However, we performed an alternative study in vivo and found a close relationship between the spacer length of the target and the length of the C-terminal domain. Progressive truncations up to the complete deletion of the C-terminal domain of the TALE were tested for their impact on cleavage activity *in vivo* using an extra-chromosomal assay. The absence of the C-terminal domain does not inhibit cleavage activity and shifts the target's spacer preference towards short sequences while progressive lengthening favors longer spacer sequences. Thus, the DNA target sequence one wants to cleave will determine the particular C-terminal truncation to be used for TALEN design. While the TALE C-terminal domain can tolerate various truncations, the N-terminal domain of the protein seems to be much less malleable as the only N-ter truncation reported that does not affect significantly TALEN activity is the deletion of the first 152 amino acids described 10 years ago.

To summarize, as the TALEN technology is still in its infancy, it is difficult to establish rigid rules regarding TALEN designs. Even though some guide lines have been described (Cermak et al. 2011) for designing active TALEN, too few data are currently available to confirm their benefits. Without a doubt, the months to come will be rich in information about TALEN design and performance. Today, one can consider that candidate cleavage sites can be found on average every 35 bp (Cermak et al. 2011). Thus, virtually any DNA sequence can be targeted as long as a T is present at each 5′-end of the DNA recognition site.

Meganucleases have a long history in the field of genome engineering and have proven to be efficient. Despite its youth, TALEN have attracted great interest due to their enormous potential as a genome modification tool. The proof of principle of the TALEN technology was made by Christian et al. (Christian et al. 2010). They were able engineer active TALEN, comprising 12–13 RVDs that cleaved gene sequences from Arabidopsis and Zebrafish using an extra-chromosomal assay in yeast. Concomitant studies were reported confirming the potential of the technology. Similar data were obtained by Mahfouz et al. using a transient assay in tobacco (Mahfouz et al. 2011). In yeast, using selection procedures, TALENs targeting endogenous *URA3*, *LYS2* and *ADE2* genes were able to induce gene disruption via NHEJ events, although the rate was low likely due to yeast's tendency to repair DSB through accurate NHEJ events. Moreover, in the presence of a DNA repair matrix, up to 34% of gene conversion events could be detected among the transformed cells (Li et al. 2011b). Finally, the authors took advantage of the relatively small yeast genome (~12 Mb) to assess TALEN specificity through whole genome sequencing. Expected targeted mutations ascribable to NHEJ events were the only mutations that could be detected.

A new step in the validation of this technology was achieved when TALENs were successfully used in higher eukaryotic cells. Human NTF3 and CCR5 genes could be modified at a high rate in K562 cell lines (Miller et al. 2010). Depending on the TALEN scaffolds used in this work, up to 25% of mutated alleles could be detected by a surveyor nuclease assay (Transgenomic, Inc., USA) upon transfection of TALEN expressing plasmids. Similarly, up to 16% of gene conversion events were detected among the hCCR5 locus-specific amplicons when cells were simultaneously transfected with TALEN expressing plasmids and a DNA repair matrix. Seventeen and 14% of hCCR5 and hIL2RG mutated alleles respectively were also recently reported after transfection of HEK293T cells with specific TALEN expressing vectors (Mussolino et al. 2011). Our own unpublished data confirm the effectiveness of this new tool as a high rate of gene modification could be achieved in the HEK293 cell line. After transfection of the cells with plasmids expressing engineered TALENs targeting either hRAG1 or hDMD genes, up to 18% of the alleles analyzed carried a mutagenic event while co-transfection with the appropriate DNA repair matrix allowed up to 8% of targeted gene insertion events (Fig. 6.6d). More recently, successful genome modifications were reported in primary cells such as human pluripotent (ESC and iPS) cells (Hockemeyer et al. 2011) and somatic zebrafish cells (Sander et al. 2011). TALENs were also successful in generating knock-out rats (Tesson et al. 2011) or zebrafish (Huang et al. 2011) by embryo microinjection. An important issue when considering DSB-induced genome modification technology is the specificity of the reagent used to produce DNA DSBs. Regarding TALEN, although no adverse effects has yet been reported, there is no clear data about their specificity. At the very least, TALEN seem to be specific enough to avoid any detectable cellular toxicity. In yeast, the sequencing of the entire genome of 5 TALEN-treated strains did not reveal any off-site mutations attributable to NHEJ events (Li et al. 2011b). However, the issue is still open as RVD specificity is not strictly restrained to one nucleotide. For example, The RVD-"HD" used to recognize a cytosine within the target DNA

sequence can also bind an adenine (Boch and Bonas 2010). Thus theoretically, TALENs should tolerate a certain level of degeneracy within their cognate target sequence. Moreover, in some rare cases, probably because of the influence of its neighboring repeat, RVD specificity may be shifted towards an unexpected nucleotide (Miller et al. 2010). Similarly, a non matching RVD, depending on its nature and position within the repeat array, may or may not have an impact on DNA binding (Romer et al. 2009). Thus, it has been postulated, as with oligonucleotides, that interactions between a TAL repeat domain and DNA may be dependent on total length as well as the number and position of non matching RVDs (Scholze and Boch 2010) within the repeat array. Moreover, despite the specificity brought by the RVD array, the critical T located at the position −1 of each recognition site (Boch et al. 2009; Romer et al. 2009) could also represent an opportune safeguard to avoid adverse off-site target effects. Obviously, TALEN technology provides an attractive solution for scientists who are involved in genome engineering. Although important milestones have already been reached, researchers are still facing several important challenges. The specificity issue is one of them. Even though the code allows the *a priori* prediction of the repeat arrays needed to target a chosen DNA target, nothing is known about potential off-site target effects. There are already some examples of unexpected behavior of the code thus the molecular basis of the TALE-DNA interactions need to be better understood in order to reach the level of safety necessary for human applications. However, given (i) the exquisite property of this DNA-binding protein and (ii) the ability to fuse the DNA-binding domain with various catalytic domains, we can already envision many applications in biotechnology that could revolutionize modern biology (not only can DSBs be precisely targeted in a chosen gene, but any enzyme involving DNA such as epigenetic modifying enzymes or enzymes from the DNA repair machinery, etc.).

6.8 Conclusions

Starting in the 1980s with I-SceI, the field of endonuclease driven genome engineering is now wide open. The use of the cellular homologous recombination and non-homologous end joining mechanisms allow virtually any kind of modification at any desired locus on a chromosome in any cell type. Over the years, meganucleases have proven to be a highly effective tool to induce genome modifications including transgenesis in many organisms such as bacteria, fly, plants, frogs, etc.. However the spread of the meganuclease technology was for a long time tempered by the difficulty to engineer new specificities. Today, tremendous advances have been made. We estimate that potential targetable recognition sites can be found every 300 bp within the human genome sequence. However, as for the ZFN technology, the lack of a robust code driving the interactions between the protein and the DNA often implies steps of optimization after the initial design. However, the marriage of computational design with High Throughput screening is starting to bring new perspectives for meganuclease engineering.

When considering DSB-induced gene modification, today's researchers have several tools at their disposal such as meganucleases, zinc finger nucleases (ZFN) and the recent TALE nucleases (TALEN). All three platforms may have pros and cons. ZFNs and TALENs appear to be very modular with a common architecture where engineerable DNA binding domains are fused to a catalytic head. Engineered ZFN proteins have been relatively easy to produce, (this is probably the reason for the world wide spread of this technology) and are by far the most popular tools for genome engineering. However optimization steps are most of the time necessary. Despite the youth of TALEN technology, the extreme modularity of the DNA binding domain and the straight forward 1 to 1 code (i.e. the parity 1 RVD : 1 nucleotide) make it today very attractive. In contrast to precedent nucleases, meganucleases which are small molecules bear both binding specificity and catalytic activities within the same compact structure assure a better control of DNA cleavage.

It is too early to conclude on the limits of each technology which may depend on the type of modification, the locus one wants to target, the cell type, the absolute specificity required and the vectorization procedures. Nevertheless the growing data reported in the literature as well as the new technologic developments available today show how the field of genome engineering has reached a mature stage and is no longer restricted to a handful of specialists.

References

Alwin S, Gere MB, Guhl E, Effertz K, Barbas CF 3rd, Segal DJ, Weitzman MD, Cathomen T (2005) Custom zinc-finger nucleases for use in human cells. Mol Ther 12:610–617

Argast GM, Stephens KM, Emond MJ, Monnat RJ Jr (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. J Mol Biol 280:345–353

Arimondo PB, Thomas CJ, Oussedik K, Baldeyrou B, Mahieu C, Halby L, Guianvarc'h D, Lansiaux A, Hecht SM, Bailly C, Giovannangeli C (2006) Exploring the cellular activity of camptothecin-triple-helix-forming oligonucleotide conjugates. Mol Cell Biol 26:324–333

Arnould S, Chames P, Perez C, Lacroix E, Duclert A, Epinat JC, Stricher F, Petit AS, Patin A, Guillier S, Rolland S, Prieto J, Blanco FJ, Bravo J, Montoya G, Serrano L, Duchateau P, Paques F (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. J Mol Biol 355:443–458

Arnould S, Perez C, Cabaniols JP, Smith J, Gouble A, Grizot S, Epinat JC, Duclert A, Duchateau P, Paques F (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. J Mol Biol 371:49–65

Arnould S, Delenda C, Grizot S, Desseaux C, Paques F, Silva GH, Smith J (2011) The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. Protein Eng Des Sel PEDS 24:27–31

Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat RJ Jr, Stoddard BL, Baker D (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. Nature 441:656–659

Ashworth J, Taylor GK, Havranek JJ, Quadri SA, Stoddard BL, Baker D (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. Nucleic Acids Res 38:5601–5608

Aten JA, Stap J, Krawczyk PM, Van Oven CH, Hoebe RA, Essers J, Kanaar R (2004) Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. Science 303:92–95

Aubert M, Ryu BY, Banks L, Rawlings DJ, Scharenberg AM, Jerome KR (2011) Successful targeting and disruption of an integrated reporter lentivirus using the engineered homing endonuclease Y2 I-AniI. PLoS One 6:e16825

Belfort M (2003) Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Genes Dev 17:2860–2863

Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM (2002) Mobile introns: pathways and proteins. In: Craig N, Craigie R, Gellert M, Lambowitz A (eds) Mobile DNA II. ASM Press, Washington, DC, pp 761–783

Bezzubova O, Silbergleit A, Yamaguchi-Iwai Y, Takeda S, Buerstedde JM (1997) Reduced X-ray resistance and homologous recombination frequencies in a RAD54−/− mutant of the chicken DT40 cell line. Cell 89:185–193

Blundell PA, Rudenko G, Borst P (1996) Targeting of exogenous DNA into Trypanosoma brucei requires a high degree of homology between donor and target DNA. Mol Biochem Parasitol 76:215–229

Boch J, Bonas U (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol 48:419–436

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512

Bogdanove AJ, Schornack S, Lahaye T (2010) TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol 13:394–401

Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring RB, Stoddard BL (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. Genes Dev 17:2875–2888

Bonas U, Stall RE, Staskawicz B (1989) Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria. Mol Gen Genet 218:127–136

Bronson SK, Smithies O (1994) Altering mice by homologous recombination using embryonic stem cells. J Biol Chem 269:27155–27158

Buerstedde JM, Takeda S (1991) Increased ratio of targeted to random integration after transfection of chicken B cell lines. Cell 67:179–188

Cabaniols JP, Paques F (2008) Robust cell line development using meganucleases. Methods Mol Biol 435:31–45

Cabaniols JP, Mathis L, Delenda C (2009) Targeted gene modifications in drug discovery and development. Curr Opin Pharmacol 9:657–663

Cabaniols JP, Ouvry C, Lamamy V, Fery I, Craplet ML, Moulharat N, Guenin SP, Bedut S, Nosjean O, Ferry G, Devavry S, Jacqmarcq C, Lebuhotel C, Mathis L, Delenda C, Boutin JA, Duchateau P, Coge F, Paques F (2010) Meganuclease-driven targeted integration in CHO-K1 cells for the fast generation of HTS-compatible cell-based assays. J Biomol Screen 15:956–967

Cavalier-Smith T (1985) Selfish DNA and the origin of introns. Nature 315(6017):283–284

Capecchi MR (2) Generating mice with targeted mutations. Nat Med 7:1086–1090

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39(12):e82

Chames P, Epinat JC, Guillier S, Patin A, Lacroix E, Paques F (2005) In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. Nucleic Acids Res 33:e178

Chen Z, Wen F, Sun N, Zhao H (2009) Directed evolution of homing endonuclease I-SceI with altered sequence specificity gzp [pii] 10.1093/protein/gzp. Protein Eng Des Sel PEDS 22:249–256

Chevalier BS, Stoddard BL (2) Homing endonucleases: structural and functional insight into the catalysis of intron/intein mobility. Nucleic Acids Res 29:3757–3774

Chevalier BS, Monnat RJ Jr, Stoddard BL (2) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. Nat Struct Biol 8:312–316

Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, Stoddard BL (2002) Design, activity, and structure of a highly specific artificial endonuclease. Mol Cell 10:895–905

Chevalier B, Turmel M, Lemieux C, Monnat RJ Jr, Stoddard BL (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. J Mol Biol 329:253–269

Chiurazzi M, Ray A, Viret JF, Perera R, Wang XH, Lloyd AM, Signer ER (1996) Enhancement of somatic intrachromosomal homologous recombination in Arabidopsis by the HO endonuclease. Plant Cell 8:2057–2066

Choulika A, Perrin A, Dujon B, Nicolas JF (1995) Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. Mol Cell Biol 15:1968–1973

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186:757–761

Cohen-Tannoudji M, Robine S, Choulika A, Pinto D, El Marjou F, Babinet C, Louvard D, Jaisser F (1998) I-SceI-induced gene replacement at a natural locus in embryonic stem cells. Mol Cell Biol 18:1444–1448

Colleaux L, D'auriol L, Galibert F, Dujon B (1988) Recognition and cleavage site of the intron-encoded omega transposase. Proc Natl Acad Sci U S A 85:6022–6026

Colleaux L, Rougeulle C, Avner P, Dujon B (1993) Rapid physical mapping of YAC inserts by random integration of I-Sce I sites. Hum Mol Genet 2:265–271

Cost GJ, Freyvert Y, Vafiadis A, Santiago Y, Miller JC, Rebar E, Collingwood TN, Snowden A, Gregory PD (2010) BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. Biotechnol Bioeng 105:330–340

Cox MM, Layton SL, Jiang T, Cole K, Hargis BM, Berghman LR, Bottje WG, Kwon YM (2007) Scarless and site-directed mutagenesis in Salmonella enteritidis chromosome. BMC Biotechnol 7:59

Cradick TJ, Keck K, Bradshaw S, Jamieson AC, Mccaffrey AP (2010) Zinc-finger nucleases as a novel therapeutic strategy for targeting hepatitis B virus DNAs. Mol Ther 18:947–954

Daboussi F, Zaslavskiy M, Poirot L, Loperfido M, Gouble A, Guyot V, Leduc S, Galetto R, Grizot S, Oficjalska D, Perez C, Delacote F, Dupuy A, Chion-Sotinel I, Le Clerre D, Lebuhotel C, Danos O, Lemaire F, Oussedik K, Cedrone F, Epinat JC, Smith J, Dickson G, Popplewell L, Koo T, Vandendriessche T, Chuah MK, Duclert A, Duchateau P, Paques F (2012) Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. Nucleic Acids Res. Epub Mar 29

Doetschman T, Gregg RG, Maeda N, Hooper ML, Melton DW, Thompson S, Smithies O (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature 330:576–578

Donoho G, Jasin M, Berg P (1998) Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. Mol Cell Biol 18:4070–4078

Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. J Am Chem Soc 128:2477–2484

Doyon Y, Mccammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, Hocking TD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol 26:702–708

Duan X, Gimble FS, Quiocho FA (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. Cell 89:555–564

Eisenschmidt K, Lanio T, Simoncsits A, Jeltsch A, Pingoud V, Wende W, Pingoud A (2005) Developing a programmed restriction endonuclease for highly specific DNA cleavage. Nucleic Acids Res 33:7039–7047

Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M (1998) Gene conversion tracts from double-strand break repair in mammalian cells. Mol Cell Biol 18:93–101

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:e3647

Epinat JC, Arnould S, Chames P, Rochaix P, Desfontaines D, Puzin C, Patin A, Zanghellini A, Paques F, Lacroix E (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. Nucleic Acids Res 31:2952–2962

Fajardo-Sanchez E, Stricher F, Paques F, Isalan M, Serrano L (2008) Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences Nucleic Acids Res 36:2163–2173

Flannagan RS, Linn T, Valvano MA (2008) A system for the construction of targeted unmarked gene deletions in the genus Burkholderia. Environ Microbiol 10:1652–1660

Flick KE, Jurica MS, Monnat RJ Jr, Stoddard BL (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. Nature 394:96–101

Flisikowska T, Thorey IS, Offner S, Ros F, Lifke V, Zeitler B, Rottmann O, Vincent A, Zhang L, Jenkins S, Niersbach H, Kind AJ, Gregory PD, Schnieke AE, Platzer J (2011) Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. PLoS One 6:e21045

Friedhoff P, Franke I, Meiss G, Wende W, Krause KL, Pingoud A (1999) A similar active site for non-specific and specific endonucleases. Nat Struct Biol 6:112–113

Galetto R, Duchateau P, Paques F (2009) Targeted approaches for gene therapy and the emergence of engineered meganucleases. Expert Opin Biol Ther 9:1289–1303

Gao H, Smith J, Yang M, Jones S, Djukanovic V, Nicholson MG, West A, Bidney D, Falco SC, Jantz D, Lyznik LA (2010) Heritable targeted mutagenesis in maize using a designed endonuclease. Plant J 61:176–187

Geissler R, Scholze H, Hahn S, Streubel J, Bonas U, Behrens SE, Boch J (2011) Transcriptionnal Activators of Human genes with progammable DNA-specificity. PLoS ONE 6:e19509

Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Menoret S, Anegon I, Davis GD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jacob HJ, Buelow R (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325:433

Gimble FS (2000) Invasion of a multitude of genetic niches by mobile endonuclease genes. FEMS Microbiol Lett 185:99–107

Gimble FS, Moure CM, Posey KL (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. J Mol Biol 334:993–1008

Glover L, Mcculloch R, Horn D (2008) Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. Nucleic Acids Res 36:2608–2618

Gouble A, Smith J, Bruneau S, Perez C, Guyot V, Cabaniols JP, Leduc S, Fiette L, Ave P, Micheau B, Duchateau P, Paques F (2006) Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. J Gene Med 8:616–622

Grabher C, Wittbrodt J (2008) Recent advances in meganuclease-and transposon-mediated transgenesis of medaka and zebrafish. Methods Mol Biol 461:521–539

Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, Villate M, Thomas S, Lemaire L, Montoya G, Blanco FJ, Paques F, Duchateau P (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res 37:5405–5419

Grizot S, Epinat JC, Thomas S, Duclert A, Rolland S, Paques F, Duchateau P (2010) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. Nucleic Acids Res 38:2006–2018

Grizot S, Duclert A, Thomas S, Duchateau P, Paques F (2011) Context dependence between subdomains in the DNA binding interface of the I-CreI homing endonuclease. Nucleic Acids Res 39:6124–6136

Grosse S, Huot N, Mahiet C, Arnould S, Barradeau S, Clerre DL, Chion-Sotinel I, Jacqmarcq C, Chapellier B, Ergani A, Desseaux C, Cedrone F, Conseiller E, Paques F, Labetoulle M, Smith

J (2011) Meganuclease-mediated inhibition of HSV1 infection in cultured cells. Mol Ther 19:694–702

Gurlebeck D, Szurek B, Bonas U (2005) Dimerization of the bacterial effector protein AvrBs3 in the plant cell cytoplasm prior to nuclear import. Plant J 42:175–187

Hauschild J, Petersen B, Santiago Y, Queisser AL, Carnwath JW, Lucas-Hahn A, Zhang L, Meng X, Gregory PD, Schwinzer R, Cost GJ, Niemann H (2011) Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. Proc Natl Acad Sci U S A 108:12013–12017

Hauschild J, Petersen B, Santiago Y, Queisser AL, Carnwath JW, Lucas-Hahn A, Zhang L, Meng X, Gregory PD, Schwinzer R, Cost GJ, Niemann H (2011) Proc Natl Acad Sci USA 108:12013–12017

Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL (1997) The structure of I-CreI, a group I intron-encoded homing endonuclease. Nat Struct Biol 4:468–476

Hinnen A, Hicks JB, Fink GR (1978) Transformation of yeast. Proc Natl Acad Sci U S A 75:1929–1933

Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, Dekelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng X, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol 27:851–857

Hockemeyer DWH, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, Zeitler B, Cherone JM, Meng X, Hinkley SJ, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2011) Genetic engineering of human pluripotent cells using TALE nucleases. Nature Biotechnol 29:731–734

Horzempa J, Shanks RM, Brown MJ, Russo BC, O'Dee DM, Nau GJ (2010) Utilization of an unstable plasmid and the I-SceI endonuclease to generate routine markerless deletion mutants in Francisella tularensis. J Microbiol Methods 80:106–108

Huang P, Xiao A, Zhou M, Zhu Z, Lin S, Zhang B (2011) Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol 29:699–700

Jacquier A, Dujon B (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41:383–394

Jarjour J, West-Foyle H, Certo MT, Hubert CG, Doyle L, Getz MM, Stoddard BL, Scharenberg AM (2009) High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. Nucleic Acids Res 37:6871–6880

Jurica MS, Monnat RJ Jr, Stoddard BL (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. Mol Cell 2:469–476

Koller BH, Smithies O (1989) Inactivating the beta 2-microglobulin locus in mouse embryonic stem cells by homologous recombination. Proc Natl Acad Sci U S A 86:8932–8935

Kostriken R, Strathern JN, Klar AJ, Hicks JB, Heffron F (1983) A site-specific endonuclease essential for mating-type switching in Saccharomyces cerevisiae. Cell 35:167–174

Kuhlmann UC, Moore GR, James R, Kleanthous C, Hemmings AM (1999) Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? FEBS Lett 463:1–2

Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS (1999) Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In: The RNA world, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Li H, Pellenz S, Ulge U, Stoddard BL, Monnat RJ Jr (2009) Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. Nucleic Acids Res 37:1650–1662

Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP, Yang B (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res 39:359–372

Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res 39(14):6315–6325

Liang F, Han M, Romanienko PJ, Jasin M (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci U S A 95:5172–5177

Liu PQ, Chan EM, Cost GJ, Zhang L, Wang J, Miller JC, Guschin DY, Reik A, Holmes MC, Mott JE, Collingwood TN, Gregory PD (2010) Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. Biotechnol Bioeng 106:97–105

Loeber J, Pan FC, Pieler T (2009) Generation of transgenic frogs. Methods Mol Biol 561:65–72

Loizos N, Silva GH, Belfort M (1996) Intron-encoded endonuclease I-TevII binds across the minor groove and induces two distinct conformational changes in its DNA substrate. J Mol Biol 255:412–424

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25:1298–1306

Lucas P, Otis C, Mercier JP, Turmel M, Lemieux C (2) Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases. Nucleic Acids Res 29:960–969

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Muller-Lerch F, Fu F, Pearlberg J, Gobel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, Mccray PB Jr, Cathomen T, Voytas DF, Joung JK (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31:294–301

Maggert KA, Gong WJ, Golic KG (2008) Methods for homologous recombination in Drosophila. Methods Mol Biol 420:155–174

Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A 108:2623–2628

Majumdar A, Muniandy PA, Liu J, Liu JL, Liu ST, Cuenoud B, Seidman MM (2008) Targeted gene knock in and sequence modulation mediated by a psoralen-linked triplex-forming oligonucleotide. J Biol Chem 283:11244–11252

Mansour SL, Thomas KR, Capecchi MR (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature 336:348–352

Marcaida MJ, Prieto J, Redondo P, Nadra AD, Alibes A, Serrano L, Grizot S, Duchateau P, Paques F, Blanco FJ, Montoya G (2008) Crystal structure of I-DmoI in complex with its target DNA provides new insights into meganuclease engineering. Proc Natl Acad Sci U S A 105:16888–16893

Marcaida MJ, Munoz IG, Blanco FJ, Prieto J, Montoya G (2010) Homing endonucleases: from basics to therapeutic applications. Cell Mol Life Sci 67:727–748

Mashimo T, Takizawa A, Voigt B, Yoshimi K, Hiai H, Kuramoto T, Serikawa T (2010) Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. PLoS One 5:e8870

Mcconnell Smith A, Takeuchi R, Pellenz S, Davis L, Maizels N, Monnat RJ Jr, Stoddard BL (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. Proc Natl Acad Sci U S A 106:5099–5104

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol 26:695–701

Metzger MJ, Mcconnell-Smith A, Stoddard BL, Miller AD (2011) Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. Nucleic Acids Res 39:926–935

Miller DG, Petek LM, Russell DW (2003) Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. Mol Cell Biol 23:3550–3557

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ (2010) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29:143–148

Morbitzer R, Elsaesser J, Hausner J, Lahaye T (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res 39(13):5790–5799

Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501

Moure CM, Gimble FS, Quiocho FA (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. Nat Struct Biol 9:764–770

Moure CM, Gimble FS, Quiocho FA (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. J Mol Biol 334:685–695

Moynahan ME, Jasin M (1997) Loss of heterozygosity induced by a chromosomal double-strand break. Proc Natl Acad Sci U S A 94:8988–8993

Munoz IG, Prieto J, Subramanian S, Coloma J, Redondo P, Villate M, Merino N, Marenchino M, D'Abramo M, Gervasio FL, Grizot S, Daboussi F, Smith J, Chion-Sotinel I, Paques F, Duchateau P, Alibes A, Stricher F, Serrano L, Blanco FJ, Montoya G (2011) Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. Nucleic Acids Res 39: 729–743

Murphy KC (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. J Bacteriol 180:2063–2071

Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res 39:9283–9293

Nickoloff JA, Chen EY, Heffron F (1986) A 24-base-pair DNA sequence from the MAT locus stimulates intergenic recombination in yeast. Proc Natl Acad Sci U S A 83:7831–7835

Niu Y, Tenney K, Li H, Gimble FS (2008) Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. J Mol Biol 382:188–202

Nolan T, Papathanos P, Windbichler N, Magnusson K, Benton J, Catteruccia F, Crisanti A (2011) Developing transgenic anopheles mosquitoes for the sterile insect technique. Genetica 139:33–39

Ochiai H, Sakamoto N, Suzuki K, Akasaka K, Yamamoto T (2008) The Ars insulator facilitates I-SceI meganuclease-mediated transgenesis in the sea urchin embryo. Dev Dyn 237: 2475–2482

Ogino H, Mcconnell WB, Grainger RM (2006) Highly efficient transgenesis in Xenopus tropicalis using I-SceI meganuclease. Mech Dev 123:103–113

Ogino H, Mcconnell WB, Grainger RM (2006) High-throughput transgenesis in Xenopus using I-SceI meganuclease. Nat Protoc 1:1703–1710

Orr-Weaver TL, Szostak JW, Rothstein RJ (1981) Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci U S A 78:6354–6358

Orr-Weaver TL, Szostak JW, Rothstein RJ (1983) Genetic applications of yeast transformation with linear and gapped plasmids. Methods Enzymol 101:228–245

Pan FC, Chen Y, Loeber J, Henningfeld K, Pieler T (2006) I-SceI meganuclease-mediated transgenesis in Xenopus. Dev Dyn 235:247–252

Paques F, Duchateau P (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gen Ther 7:49–66

Paques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 63:349–404

Perez C, Guyot V, Cabaniols JP, Gouble A, Micheaux B, Smith J, Leduc S, Paques F, Duchateau P (2005) Factors affecting double-strand break-induced homologous recombination in mammalian cells. Biotechniques 39:109–115

Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M (1994) Protein splicing elements: inteins and exteins–a definition of terms and recommended nomenclature. Nucleic Acids Res 22:1125–1127

Pierce AJ, Hu P, Han M, Ellis N, Jasin M (2001) Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. Genes Dev 15:3237–3242

Porteus MH (2006) Mammalian gene targeting with designed zinc finger nucleases. Mol Ther 13:438–446

Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763

Porteus MH, Cathomen T, Weitzman MD, Baltimore D (2003) Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. Mol Cell Biol 23:3558–3565

Posfai G, Kolisnychenko V, Bereczki Z, Blattner FR (1999) Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. Nucleic Acids Res 27:4409–4415

Poteete AR (2001) What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. FEMS Microbiol Lett 201:9–14

Prieto J, Epinat JC, Redondo P, Ramos E, Padro D, Cedrone F, Montoya G, Paques F, Blanco FJ (2008) Generation and analysis of mesophilic variants of the thermostable archaeal I-DmoI homing endonuclease. J Biol Chem 283:4364–4374

Puchta H (1999) Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. Genetics 152:1173–1181

Puchta H (1999) Use of I-Sce I to induce DNA double-strand breaks in Nicotiana. Methods Mol Biol 113:447–451

Puchta H (2002) Gene replacement by homologous recombination in plants. Plant Mol Biol 48:173–182

Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci U S A 93:5055–5060

Radecke F, Peter I, Radecke S, Gellhaus K, Schwarz K, Cathomen T (2006) Targeted chromosomal gene modification in human cells by single-stranded oligodeoxynucleotides in the presence of a DNA double-strand break. Mol Ther 14:798–808

Redondo P, Prieto J, Munoz IG, Alibes A, Stricher F, Serrano L, Cabaniols P, Daboussi F, Arnould S, Perez C, Duchateau P, Paques F, Blanco FJ, Montoya G (2008) Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. Nature 456:107–111

Renfer E, Amon-Hassenzahl A, Steinmetz PR, Technau U (2010) A muscle-specific transgenic reporter line of the sea anemone, Nematostella vectensis. Proc Natl Acad Sci U S A 107: 104–108

Richardson C, Moynahan ME, Jasin M (1998) Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. Genes Dev 12:3831–3842

Roeder GS (1997) Meiotic chromosomes: it takes two to tango. Genes Dev 11:2600–2621

Romer P, Recht S, Lahaye T (2009) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. Proc Natl Acad Sci U S A 106:20526–20531

Rong YS, Golic KG (2000) Gene targeting by homologous recombination in Drosophila. Science 288:2013–2018

Rong YS, Golic KG (2001) A targeted gene knockout in Drosophila. Genetics 157:1307–1312

Rong YS, Golic KG (2003) The homologous chromosome is an effective template for the repair of mitotic DNA double-strand breaks in Drosophila. Genetics 165:1831–1842

Rong YS, Titen SW, Xie HB, Golic MM, Bastiani M, Bandyopadhyay P, Olivera BM, Brodsky M, Rubin GM, Golic KG (2002) Targeted mutagenesis by homologous recombination in D. melanogaster. Genes Dev 16:1568–1581

Rosen LE, Morrison HA, Masri S, Brown MJ, Springstubb B, Sussman D, Stoddard BL, Seligman LM (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. Nucleic Acids Res 34:4791–4800

Rouet P, Smih F, Jasin M (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol 14:8096–8106

Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK, Yeh JR (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat Biotechnol 29:697–698

Scalley-Kim M, Mcconnell-Smith A, Stoddard BL (2007) Coevolution of a homing endonuclease and its host target sequence. J Mol Biol 372:1305–1319

Schaefer DG (2001) Gene targeting in Physcomitrella patens. Curr Opin Plant Biol 4:143–150
Schaefer DG, Zryd JP (1997) Efficient gene targeting in the moss Physcomitrella patens. Plant J 11:1195–1206
Schleifman EB, Chin JY, Glazer PM (2008) Triplex-mediated gene modification. Methods Mol Biol 435:175–190
Schmidt-Puchta W, Orel N, Kyryk A, Puchta H (2004) Intrachromosomal homologous recombination in Arabidopsis thaliana. Methods Mol Biol 262:25–34
Scholze H, Boch J (2010) TAL effector-DNA specificity. Virulence 1:428–432
Seligman LM, Chisholm KM, Chevalier BS, Chadsey MS, Edwards ST, Savage JH, Veillet AL (2002) Mutations altering the cleavage specificity of a homing endonuclease. Nucleic Acids Res 30:3870–3879
Sethuraman J, Majer A, Friedrich NC, Edgell DR, Hausner G (2009) Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa. Mol Biol Evol 26:2299–2315
Shen BW, Landthaler M, Shub DA, Stoddard BL (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. J Mol Biol 342:43–56
Siebert R, Puchta H (2002) Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. Plant Cell 14:1121–1131
Silva G, Poirot L, Galetto R, Smith J, Montoya G, Duchateau P, Paques F (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. Curr Gene Ther 11:11–27
Silva GH, Belfort M, Wende W, Pingoud A (2006) From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. J Mol Biol 361:744–754
Silva GH, Dalgaard JZ, Belfort M, van Roey P (1999) Crystal structure of the thermostable archaeal intron-encoded endonuclease I-DmoI. J Mol Biol 286:1123–1136
Simon P, Cannata F, Perrouault L, Halby L, Concordet JP, Boutorine A, Ryabinin V, Sinyakov A, Giovannangeli C (2008) Sequence-specific DNA cleavage mediated by bipyridine polyamide conjugates. Nucleic Acids Res 36:3531–3538
Smih F, Rouet P, Romanienko PJ, Jasin M (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. Nucleic Acids Res 23:5012–5019
Smith GR, Amundsen SK, Dabert P, Taylor AF (1995) The initiation and control of homologous recombination in Escherichia coli. Philos Trans R Soc Lond B Biol Sci 347:13–20
Smith J, Grizot S, Arnould S, Duclert A, Epinat JC, Chames P, Prieto J, Redondo P, Blanco FJ, Bravo J, Montoya G, Paques F, Duchateau P (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 34:e149
Smithies O (2001) Forty years with homologous recombination. Nat Med 7:1083–1086
Soldner F, Laganiere J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, Khurana V, Golbe LI, Myers RH, Lindquist S, Zhang L, Guschin D, Fong LK, Vu BJ, Meng X, Urnov FD, Rebar EJ, Gregory PD, Zhang HS, Jaenisch R (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146:318–331
Spear MA (2000) Efficient DNA subcloning through selective restriction endonuclease digestion. Biotechniques 28:660–662, 664, 666 passim
Spiegel PC, Chevalier B, Sussman D, Turmel M, Lemieux C, Stoddard BL (2006) The structure of I-CeuI homing endonuclease: evolving asymmetric DNA recognition from a symmetric protein scaffold. Structure 14:869–880
Stoddard BL (2005) Homing endonuclease structure and function. Q Rev Biophys 38:49–95
Stoddard BL, Scharenberg AM, Monnat RJ (2007) Advances in engineering homing endonucleases for gene targeting: ten years after structures. In: Bertolotti R, Ozawa K (eds) Progress in gene therapy: autologous and cancer stem cell gene therapy. World Scientific Publishing Co. Pte. Ltd., pp 135–168
Sussman D, Chadsey M, Fauce S, Engel A, Bruett A, Monnat R Jr, Stoddard BL, Seligman LM (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. J Mol Biol 342:31–41

Suzuki N, Nonaka H, Tsuge Y, Okayama S, Inui M, Yukawa H (2005) Multiple large segment deletion method for *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 69:151–161

Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat* Biotechnol 25:786–793

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25–35

Taghian DG, Nickoloff JA (1997) Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells. Mol Cell Biol 17:6386–6393

Takeuchi H, Georgiev O, Fetchko M, Kappeler M, Schaffner W, Egli D (2007) In vivo construction of transgenes in Drosophila. Genetics 175:2019–2028

Ten Asbroek AL, Ouellette M, Borst P (1990) Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei.* Nature 348:174–175

Tesson L, Usal C, Menoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, Gregory PD, Anegon I, Cost GJ (2011) Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol 29:695–696

Thermes V, Grabher C, Ristoratore F, Bourrat F, Choulika A, Wittbrodt J, Joly JS (2002) I-SceI meganuclease mediates highly efficient transgenesis in fish. Mech Dev 118:91–98

Thierry A, Dujon B (1992) Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes. Nucleic Acids Res 20:5625–5631

Thomas KR, Capecchi MR (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell 51:503–512

Thomas KR, Capecchi MR (1990) Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. Nature 346:847–850

Thyme SB, Jarjour J, Takeuchi R, Havranek JJ, Ashworth J, Scharenberg AM, Stoddard BL, Baker D (2009) Exploitation of binding energy for catalysis and design. Nature 461:1300–1304

Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK, Voytas DF (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459:442–445

Traver BE, Anderson MA, Adelman ZN (2009) Homing endonucleases catalyze double-stranded DNA breaks and somatic transgene excision in Aedes aegypti. Insect Mol Biol 18:623–633

Ulge UY, Baker DA, Monnat RJ Jr (2011) Comprehensive computational design of mCreI homing endonuclease cleavage specificity for genome engineering. Nucleic Acids Research 39:4330–4339

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435:646–651

van Roey P, Meehan L, Kowalski JC, Belfort M, Derbyshire V (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. Nat Struct Biol 9:806–811

van Roey P, Waddling CA, Fox KM, Belfort M, Derbyshire V (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. EMBO J 20:3631–3637

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6:e16765

Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat RJ Jr, Burt A, Crisanti A (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature 473:212–215

Windbichler N, Papathanos PA, Catteruccia F, Ranson H, Burt A, Crisanti A (2007) Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. Nucleic Acids Res 35:5922–5933

Yang M, Djukanovic V, Stagg J, Lenderts B, Bidney D, Falco SC, Lyznik LA (2009) Targeted mutagenesis in the progeny of maize transgenic plants. Plant Mol Biol 70:669–679

Young JJ, Cherone JM, Doyon Y, Ankoudinova I, Faraji FM, Lee AH, Ngo C, Guschin DY, Paschon DE, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Harland RM, Zeaitler B (2011) Efficient targeted gene disruption in the soma and germ line of the frog Xenopus tropicalis using engineered zinc-finger nucleases. Proc Natl Acad Sci U S A 108:7052–7057

Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29:149–153

Zhao L, Bonocora RP, Shub DA, Stoddard BL (2007) The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. EMBO J 26:2432–2442

Zou J, Maeder ML, Mali P, Pruett-Miller SM, Thibodeau-Beganny S, Chou BK, Chen G, Ye Z, Park IH, Daley GQ, Porteus MH, Joung JK, Cheng L (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell 5:97–110

Zou J, Sweeney CL, Chou BK, Choi U, Pan J, Wang H, Dowey SN, Cheng L, Malech HL (2011) Oxidase deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease mediated safe harbor targeting. Blood 117:5561–5572

Part III
Integration Based on Site-Specific Recombination

Chapter 7
Cre/loxP, Flp/FRT Systems and Pluripotent Stem Cell Lines

Candice G.T. Tahimic, Kenji Sakurai, Kazuhiro Aiba, and Norio Nakatsuji

Abstract Cre and Flp recombinases are the most widely used site-specific recombinases in genome engineering. Both are members of the tyrosine class of recombinases and catalyze the reversible, site-specific recombination between two identical sequences of 34 bp length in the absence of accessory factors. The substrate sequences for Cre and Flp recombinases are called loxP and FRT sites respectively. Cre recombinase was discovered in the *E. coli* bacteriophage P1 where it plays a crucial role in the life cycle of P1 while Flp recombinase was originally derived from the 2μ circle of *Saccharomyces cerevisiae* and catalyzes recombination between inverted repeats within the 2μ plasmid. This chapter shall provide a brief historical perspective on the discovery and early development of

C.G.T. Tahimic
Endocrine Research Unit, San Francisco Veterans Affairs Medical Center, Building 6 Room 312, 4150 Clement St., 94121 San Francisco, California, USA
e-mail: candice.tahimic@ucsf.edu

K. Sakurai
Institute for Frontier Medical Sciences, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan
e-mail: kenji.sakurai@kyowa-kirin.co.jp

K. Aiba (✉)
Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Yoshida-Ushinomiyacho, Sakyo-ku, Kyoto 606-8501, Japan
e-mail: kaiba@icems.kyoto-u.ac.jp

N. Nakatsuji (✉)
Institute for Frontier Medical Sciences, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Yoshida-Ushinomiyacho, Sakyo-ku, Kyoto 606-8501, Japan
e-mail: nnakatsu@icems.kyoto-u.ac.jp

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*, Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_7,

the Cre/loxP and Flp/FRT systems citing key studies that paved the way for the application of these site-specific recombination technologies to the engineering of mammalian genomes. Also included are discussions on the mechanisms of Cre/loxP and Flp/FRT systems and application of these site-specific recombinase technologies to the introduction of transgenes in human pluripotent stem cells. Certain studies using mouse pluripotent stem cells will also be discussed in order to highlight the possibility of adopting the same strategy in their human counterparts. Lastly, future prospects for these two site-specific recombinases will be presented.

Keywords Cre/loxP system • Flp/FRT system • Pluripotent stem cells • Site-specific recombination • Site-specific transgene integration

7.1 A Historical Perspective on Cre/loxP and Flp/FRT Systems

7.1.1 Cre/loxP

In 1981, Sternberg and Hamilton reported a high frequency of recABC-independent site-specific recombination in the *E. coli* bacteriophage P1 (Sternberg and Hamilton 1981). In this study, they identified two crucial components for recombination found in phage DNA, a site which they called loxP (for locus of crossing over (x), Pl), and a trans-acting gene called Cre (cyclization recombination). In another report that year, Sternberg and colleagues showed that P1 phage can integrate its DNA at low frequencies into *E. coli* chromosome (Sternberg et al. 1981). The site on the bacterial chromosome into which the P1 plasmid integrates was called loxB. They also showed that integration between loxP and loxB sites generated two hybrid sites which they called LoxL and LoxR. From 1982, major strides towards application of Cre/loxP technology to the engineering of mammalian genomes were made, starting from that year's publication by Hoess and colleagues that reported the nucleotide sequences of all four sites, loxP, loxB, loxL and loxR (Hoess et al. 1982). In their study, they showed that loxP has a perfect 13 bp inverted repeat sequence flanking each side of an asymmetric 8 bp spacer region (Fig. 7.1a). The next year, Abremski and colleagues defined the requirements for loxP/loxP recombinations in terms of co-factors and substrates (Abremski et al. 1983). They reported that no external energy cofactors are required for recombination and that *in vitro* recombination between loxP sites can occur whether the molecule containing them is a supercoiled, nicked-circle or linear DNA. Another major stride towards Cre/loxP-based engineering of mammalian genomes was the 1987 publication of Sauer demonstrating that Cre was functional in the more structurally complex genome of the eukaryote, *Saccharomyces cerevisiae* (Sauer 1987). In this study, Sauer showed that Cre recombinase can catalyze recombination between two loxP

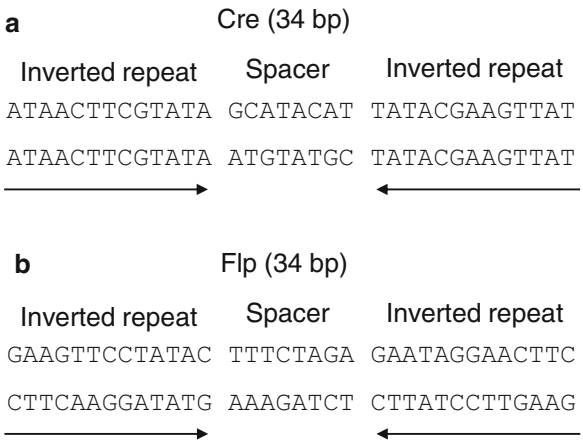


Fig. 7.1 LoxP (**a**) and Flp (**b**) sequences showing positions of inverted repeats and nonpalindromic spacer region

sites in the genome of *S. cerevisiae*. This remarkable property of Cre recombinase may be fully appreciated if one considers that in order to conduct more complex transactions on a eukaryotic chromosome, a prokaryotic protein must not only recognize DNA but must also be capable of managing synapsis, DNA cleavage, strand exchange, and religation (Sauer 1987). The year after, Sauer and Henderson further demonstrated that the Cre/loxP system was also functional in mammalian cells (Sauer and Henderson 1988). In their study, they showed that Cre recombinase can mediate the excision of DNA sequences flanking two loxP sites in the genome of the murine cell line C-127. In 1990, they reported the first targeted insertion of exogenous DNA into both yeast and mammalian cells. They demonstrated that Cre-mediated recombination can direct the site-specific integration of a loxP-containing circular DNA into a chromosomal loxP site and that the resulting integrants are predominantly simple single-copy insertions. In 1993, the use of the Cre/loxP system to manipulate the genome of mouse ES cells (ESCs) was reported by Gu and colleagues (Gu et al. 1993). In their study, the Cre/loxP system was used to study Ig isotype switch recombination in a murine system. They described an approach wherein gene targeting was used to introduce two loxP sites flanking the J_H-E_μ sequences of the IgH locus of mouse ESCs. Next, transient expression of Cre recombinase via an expression vector was performed, resulting in the excision of the J_H-E_μ sequences. ESC clones carrying the IgH deletion were then introduced into blastocysts, thereby generating IgH mutant mice. Their study was a significant step towards the widespread use of the Cre/loxP system in the creation of null phenotypes in mouse. The significant achievements of their study were the demonstration that combined used of the Cre/loxP system and conventional gene targeting can efficiently generate deletions of defined length in mouse ESCs and that the combinatorial method allowed germline transmission.

The first report for the application of the Cre/loxP system in human ESCs came in 2007 (Irion et al. 2007) where it was used to introduce a fluorescent reporter gene in the newly discovered human homolog of the ROSA26 locus. The recombination strategies used in this study will be further discussed in one section of this chapter.

7.1.2 Flp/FRT

In 1978, Beggs reported the high recombination frequency in the 2μ plasmid of *S. cerevisiae* (Beggs 1978) and observed that plasmid sequences recombined intramolecularly during propagation in this yeast. Two years after, Broach and Hicks reported that the recombination proficiency of the 2μ plasmid was linked to an inverted repeat of 599 bp in length and a gene product which they called Flp, in reference to the ability of the gene product to promote recombination between two 2μ plasmids (Broach and Hicks 1980). By 1982, Broach and colleagues delineated the site for recombination to 65 bp within the inverted repeat and also showed that Flp was sufficient to induce intermolecular recombination between two inverted repeats (Broach et al. 1982). The next year, Vetter and colleagues demonstrated that the Flp gene product can catalyze *in vitro* recombination between two inverted repeats and that this process required the presence of cations. They also reported that Flp works efficiently on supercoiled, relaxed circular, or linear DNA and noted that Flp possessed certain similarities with Cre recombinase (Vetter et al. 1983). In 1985, Andrews and colleagues, using an in vitro DNA protection experiment, defined the Flp binding site to an 8-bp spacer flanked on both sides by a 13 bp inverted repeat (Andrews et al. 1985) (Fig. 7.1b). The year after, McLeod and colleagues coined the word "FRT" to refer to the Flp recognition sequence. They also determined the actual recombination crossover point *in vivo* and confirmed that the essential sequences for the recombination event are limited to the 8-base-pair core sequence and the two 13-base-pair repeated units immediately flanking it (McLeod et al. 1986). Three years after, it was shown in *S. cerevisiae* that Flp can catalyze the recombination between FRT sites separated by a distance of up to 13.3 kb (Rank et al. 1989). In 1992, Dang and Perrimon described the use of the Flp/FRT system to create *Drosophila* embryonic mosaics (Dang and Perrimon 1992). In their work, they placed Flp under the control of a heat shock protein promoter and used this inducible expression system to promote mitotic exchange between homologous chromosomes containing FRT sequences. The next year, Fiering and colleagues demonstrated the utility of Flp-mediated site-specific recombination to study complex cis-regulatory DNA elements within their normal chromosomal context. In their study, Flp/FRT system was used to delete a selectable marker that they inserted into the locus control region (LCR) of the murine beta globin locus of a mouse erythroleukemia cell somatic hybrid (Fiering et al. 1993). As a follow-up publication, Fiering and colleagues showed that the system can also be applied to mouse ESCs. They also highlighted that the introduction of a selection marker to facilitate selection of correctly targeted clones may interfere in the analysis of complex regulatory elements. Hence, they suggested that the selectable marker be removed to avoid influencing the phenotype of the mutation (Fiering et al. 1995). For this purpose, the addition of FRT sites on both sides of the transgene was shown to be useful. The application of the Flp/FRT system in human ESCs was first reported by Patsch and colleagues (Patsch et al. 2010). Their study described the use of a cell-permeant Flp protein that can mediate enhanced site-specific recombination

in human ESCs. In their system, they also used Cre and Flp as switches to regulate the expression of lacZ and the caudal-related homeobox protein CDX2 transgenes in human ESCs.

7.2 Mechanisms of Cre/loxP and Flp/FRT Recombination

Cre and Flp belong to the tyrosine class of site-specific recombinases (SSRs). Tyrosine recombinases mediate site-specific recombination through nucleophilic attack on the phosphodiester backbone of DNA by a tyrosine hydroxyl group resulting in a covalent protein-DNA intermediate (Grindley et al. 2006; Anastassiadis et al. 2009). Cre is a 38.5 kDa protein product of the bacteriophage P1 cre gene that plays critical roles in the propagation and infectivity of P1 phage. Firstly, it provides an alternative mechanism for the circularization of the P1 genome after infection (Segev and Cohen 1981; Hochman et al. 1983) and secondly, it ensures proper partitioning of the P1 plasmid during bacterial cell division through resolution of plasmid dimers (Austin et al. 1981; Sauer 1987). Flp is a 48.6 kDa protein originally found in the 2 μ plasmid of *S. cerevisiae* and functions in the amplification of the plasmid in yeast. Both Cre and Flp exist as homotetramers in the presence of DNA (Guo et al. 1997; Chen et al. 2000).

Cre recognizes a 34 bp sequence called loxP, composed of an 8 bp non-palindromic spacer region flanked on each side by 13 bp inverted repeats. On the other hand, the full-length substrate of Flp is a 48 bp FRT site consisting of two inverted 13 bp repeats around an 8 bp spacer, and an additional 13 bp direct repeat. However, Flp can also recognize a truncated 34 bp version of FRT which lacks the third 13 bp repeat. This 34 bp FRT sequence appears to be a more favorable substrate for the excision reaction. It has been suggested that the use of full length FRT sites may increase occurrence of the integration over the excision reaction (Baer and Bode 2001). For both loxP and FRT, the nonpalindromic nature of the spacer sequence provides directionality to recombination.

Cre and Flp both catalyze reciprocal conservative DNA recombination between recognition sequences. The process involves the following steps, (1) sequence recognition and pairing of recombination sequences, (2) cleavage, (3) strand exchange, (4) ligation of DNA and (5) release of the recombinant DNA products (Fig. 7.2). The recombinase initially binds to each of the recombination sites. Cleavage of the DNA strand is induced when the hydroxyl group of the tyrosine residue attacks a phosphate group in the spacer region. This conserves the energy of the phosphodiester bond thereby allowing reversibility of the reaction without requiring a high-energy cofactor. Intermediate products are generated, namely free hydroxyl groups at the 5′ ends of the DNA and covalent DNA-protein phosphotyrosines link at the 3′ ends. The next step is an exchange reaction where the free 5′ ends attack the 3′ phosphotyrosines of the opposing DNA substrates to generate a Holliday junction. The attack of a nucleophilic tyrosine on a phosphate group is repeated for the intact strand and the new 5′ ends migrate and attack the 3′ phosphotyrosine linkage of the

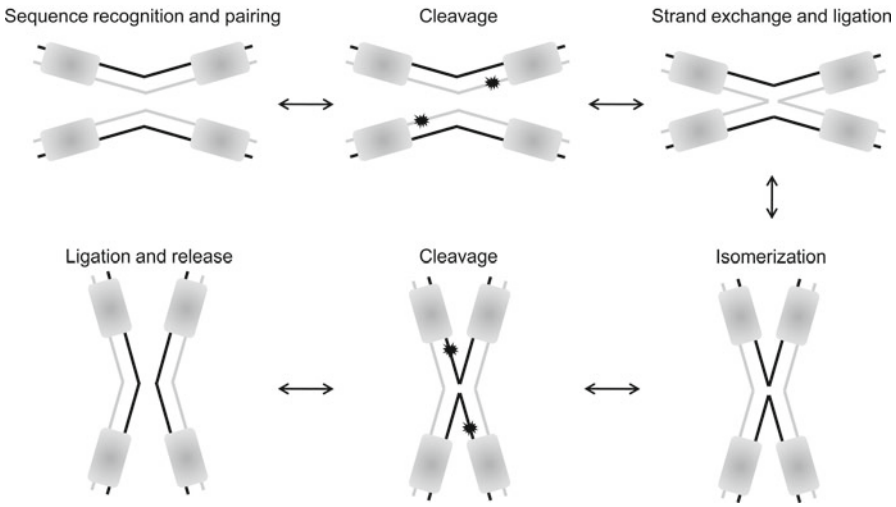


Fig. 7.2 Diagram of tyrosine recombinase-mediated DNA recombination. Two DNA duplexes (*gray and black lines*) are bound by four recombinase molecules (*gray rounded rectangles*). The *black star* represents DNA cleavage. One strand of each DNA duplex is cleaved (*gray line*), exchanged and then ligated to form a Holliday structure. Isomerization allows cleavage, exchange and ligation of the other strands (*black lines*). Finally, the recombined products are released

other DNA duplex. This leads to the resolution of the Holliday junction (Hoess and Abremski 1985; Grindley et al. 2006).

Despite sharing a mechanism common to tyrosine recombinases, some differences exist between Flp and Cre. Flp does not display a significant preference for one or the other end of the spacer for initial cleavage. In contrast, Cre preferentially cleaves at the GC end of the spacer rather than at the AT end (Ghosh et al. 2005; Grindley et al. 2006). Kinetic studies have also shown that Flp has a lower binding affinity for FRT than Cre has for loxP. Cre dimers bind the two halves of their target sites cooperatively at approximately threefold higher cooperativity than Flp (Ringrose et al. 1998; Patsch et al. 2010) and recombination efficiency of Flp is 10-fold lower than that of Cre (Ringrose et al. 1998). Moreover, Flp is known to be less thermostable than Cre with its activity significantly reduced at 37°C (Bucholz et al. 1996). This is partly due to the difference in optimal growth temperatures between the original hosts *S. cerevisiae* and *E. coli*. While *E. coli* grows well in 37°C, *S. cerevisiae* prefers temperatures lower than 37°C. In an attempt to circumvent the lower activity of Flp, molecular evolution was used to generate FLPe, a modified version of Flp exhibiting improved performance at 37°C. Despite the fivefold increase in activity of FLPe compared to the wild type version, FLPe has been shown to possess only 25% the activity of Cre (Buchholz et al. 1998).

Depending on the position and orientation of the two loxP sites relative to each other, Cre activity will result in different post-recombination outcomes (Fig. 7.3). Recombination between two loxP sites of similar orientation located within the same molecule will result in the excision of DNA sequences between them. On the

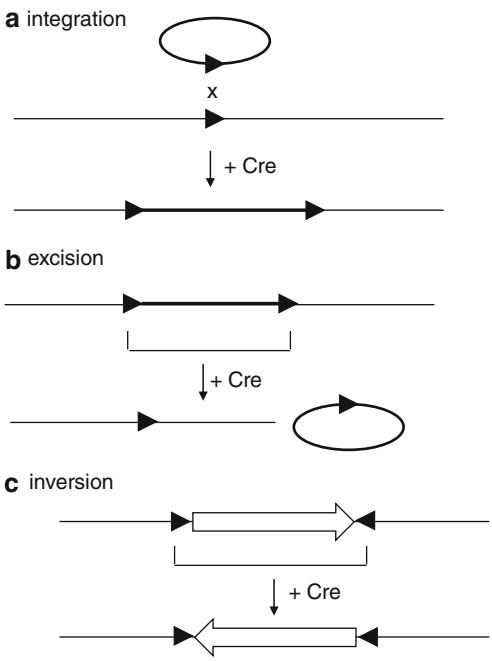


Fig. 7.3 Types of recombination activities mediated by Cre or Flp recombinase. The three basic outcomes are (**a**) integration, (**b**) excision and (**c**) inversion

other hand, Cre activity will result in the inversion of DNA sequences between a pair of loxP sites of opposite orientation. In scenarios wherein one loxP site is situated in two separate DNA molecules, Cre-mediated recombination will result in the integration of DNA sequences. Cre activity can also induce translocations between two chromosome segments that each carry a loxP site (Wu et al. 2007).

7.3 Mutant lox and FRT Sites

Studies using mutant lox and FRT sites have revealed the major requirements for efficient recombination. The altered specificities of these mutant recombinase recognition sites have been exploited to create more sophisticated recombination strategies and to restrict the recombination reaction to one direction.

Homology within the central spacer sequence is necessary for efficient recombination between two loxP sites. It has been demonstrated that a single nucleotide substitution within the 8 bp spacer unit renders a loxP site incapable of recombining with the wild type sequence. Interestingly, this mutant loxP site can efficiently recombine with itself (Hoess et al. 1986). Examples of modified loxP sequences carrying mutations within the non-palidoromic spacer region include lox511 (Hoess et al. 1986), lox2272 and lox5171 (Lee and Saito 1998).

Mutant loxP sites have also generated through the introduction of nucleotide substitutions in the left or right inverted repeat elements and are referred to as LE

Table 7.1 List of selected mutant lox and FRT sequences

Name	Sequence	References
Wild type loxP	ATAACTTCGTATA GCATACAT TATACGAAGTTAT	Hoess et al. (1982)
lox2272[a]	ATAACTTCGTATA GgATACtT TATACGAAGTTAT	Lee and Saito (1998)
lox5171[a]	ATAACTTCGTATA GtAcACAT TATACGAAGTTAT	Lee and Saito (1998)
lox511[b]	ATAACTTCGTATA GtAtACAT TATACGAAGTTAT	Hoess et al. (1986)
lox71[c]	taccgTTCGTATA GCATACAT TATACGAAGTTAT	Albert et al. (1995)
lox66[c]	ATAACTTCGTATA GCATACAT TATACGAAcggta	Albert et al. (1995)
Wild type FRT	GAAGTTCCTATAC TTTCTAGA GAATAGGAACTTC	Broach et al. (1982) and Andrews et al. (1985)
F3 mutant[a]	GAAGTTCCTATAC TTcaaAtA GAATAGGAACTTC	Schlake and Bode (1994)
F5 mutant[a]	GAAGTTCCTATAC TTcaaAag GAATAGGAACTTC	Schlake and Bode (1994)

The wild type sequences are provided as a reference. Small caps indicate differences with corresponding wild type sequence

[a]Recombines only with a mutant site having the same sequence

[b]May have significant cross-reaction with loxP

[c]lox71 and lox66 recombine to generate incompatible wild type loxP and lox78 sites

or RE lox. LE/RE lox recombination favors integration over excision (Araki et al. 2010). This is because recombination between LE and RE sites results in a wild type loxP and a double mutant lox site that is an inefficient substrate of Cre recombinase. Due to its unidirectionality, the LE/RE lox system has been incorporated in Cre-mediated recombination strategies to minimize the possibility of transgene excision after the initial integration reaction.

Homology within the central spacer sequence is also necessary for efficient recombination of FRT sites (Senecoff and Cox 1986; Bode et al. 2000). However, the spacer sequence itself is not so important provided that homology is maintained between the two target spacer sequences (Senecoff and Cox 1986; Lee and Saito 1998).

Table 7.1 shows some of the currently available mutant loxP and FRT. As will be further discussed in the succeeding sections, these additional sequences have increased the precision of both Cre and FRT-based site-specific recombination and spurned the development of more sophisticated genetic manipulation strategies.

7.4 Pluripotent Stem Cells, an Overview

In 1981, Evans and Kaufman reported the derivation of murine embryonic stem cells (ESCs) from the inner cell mass of murine preimplantation blastocyst embryos (Evans and Kaufman 1981; Martin 1981). ESCs have two characteristic properties,

self renewal which pertains to their ability to expand indefinitely, and pluripotency defined as the capacity to differentiate into derivatives of all three germ layers. More than a decade after, the derivation of ESCs from *Rhesus* macaque (Thomson et al. 1995), and the common marmoset *Callithrix jacchus* (Thomson et al. 1996) were reported. Human ESC lines were established in 1998 (Thomson et al. 1998) and a variety human and non-human primate ESC lines were generated by other groups thereafter (see Nakatsuji and Suemori 2002; Guhr et al. 2006 for reviews). Adult cell-derived pluripotent stem cells (PSCs) also became available with the discovery that differentiated cells can revert to an embryonic state via forced expression of transcription factors such as Oct3/4, Sox2, Nanog and c-Myc (Takahashi and Yamanaka 2006; Takahashi et al. 2007; Okita et al. 2007). These cells are referred to as induced pluripotent stem cells (iPSCs) and a number of mouse and human-derived lines are currently available.

Unlike mouse ESCs and iPSCs, their human counterparts cannot be routinely passaged as single cells. This greatly impedes the creation of human ES and iPS reporter cell lines as genetic modification is usually followed by clonal selection. Moreover, human ESCs and iPSCs exhibit lower gene targeting efficiencies of approximately 10^{-6} compared to values of 10^{-2} or 10^{-3} observed in murine counterparts. A proposed explanation for such differences is that hESCs are actually more similar in properties to mouse epiblast stem cells. Both cell types exhibit dependence on bFGF supplementation, flattened colony morphology, low tolerance for single cell disruption, and slower growth rate compared to mouse ESCs (Brons et al. 2007; Tesar et al. 2007). A study has strengthened this theory by demonstrating that human iPSCs that mimic morphological, molecular, and functional properties of murine ESCs can be generated by performing the reprogramming process in the presence of leukemia inhibitory factor (LIF) (Buecker et al. 2010). These mESC-like human iPSCs, referred to as hLR5 iPSCs, have been shown to possess gene targeting rates similar to that found in mESCs. However, it should be noted that this particular cell line exists in a metastable state, meaning that LIF withdrawal and transcriptional inactivation of reprogramming factors will result in conversion into typical hESC behavior. Nevertheless, the study is a clear demonstration that low gene targeting rates is not necessarily an intrinsic property of human PSCs, but rather, is a consequence of an incomplete understanding of factors required for the establishment and maintenance of pluripotency. Further improvements in methodologies to derive PSCs should result in cell lines that are more amenable to genetic manipulation and downstream applications.

An ever-expanding list of PSC lines is currently available to researchers. Naturally, the choice of cell line will be dictated by the objectives of the study and other practical considerations. In studies that use a single hESC line to integrate various versions of the transgene at the same locus, the Cre/loxP and Flp/FRT systems may provide the added benefit of more robust recombination rates. For example, loxP or FRT sites can be introduced via conventional gene targeting and subsequent transgene integrations can be performed via the relatively more efficient Cre or Flp systems. This strategy will be further discussed in the next section.

7.5 Basic Strategies for Site-Specific Integration of Transgenes via Cre/loxP and Flp/FRT Systems

Although literature search will reveal numerous approaches to site-specific integration of transgenes using Cre/loxP and Flp/FRT systems, these strategies are combinations of three basic approaches which will be categorized as follows: (1) integration, (2) excision, (3) inversion, and (4) recombination-mediated cassette exchange (RMCE). For brevity, the Cre/loxP system will be used as an example to describe these approaches in detail. In the simple integration approach (Fig. 7.4a), gene targeting is initially performed to introduce a single loxP site. Correctly targeted clones are then subjected to the next round of genetic manipulation involving co-transfection of a Cre expression vector containing a loxP site and the gene of interest. Cre activity results in the intermolecular recombination between the two loxP sites and the subsequent integration of the transgene into the chromosome of the host cell.

The excision approach (Fig. 7.4b) makes use of the excision capabilities of Cre. The initial gene targeting step involves the use of a vector containing two loxP sites that flank a gene of interest (GOI). Recombinant clones are then co-transfected with Cre expression plasmid and a plasmid carrying a single loxP site and another GOI. Intermolecular recombination between the two loxP sites in the chromosomal DNA will result in the excision of the sequences between these two sites, leaving one chromosomal loxP site.

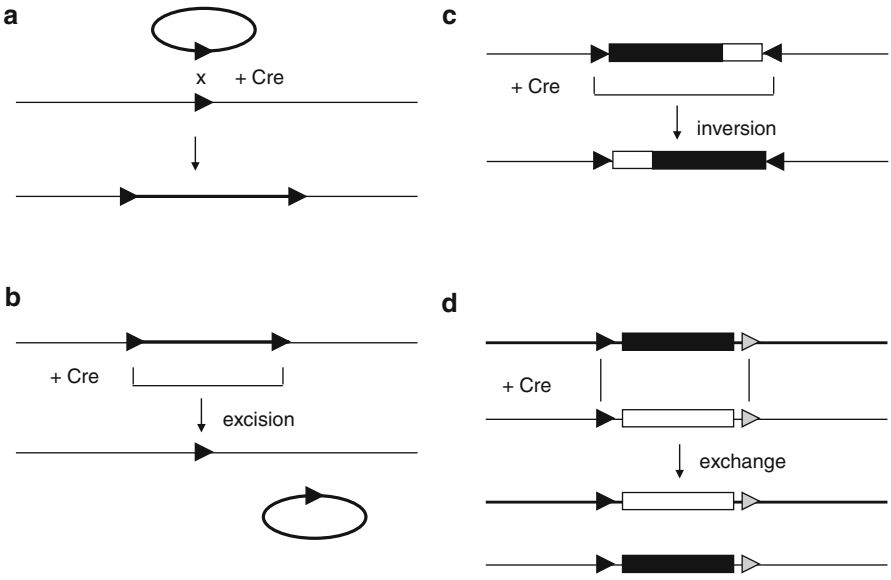


Fig. 7.4 Basic strategies for site-specific integration of transgenes. (**a**) integration, (**b**) excision, (**c**) inversion, and (**d**) recombination-mediated cassette exchange (*RMCE*)

The inversion approach (Fig. 7.4c) begins with a gene targeting step to introduce a GOI flanked by a pair of anti-parallel loxP sites. Cre recombinase recombines the two loxP sites and in the process inverts the transgene. In most applications, Cre-mediated inversion brings about reconstitution of transgene expression. In a straightforward application of this approach, no transgene is co-transfected with the Cre-expressing vector as the transgene is already present in the post-targeted allele.

RMCE (Fig. 7.4d) is a more sophisticated approach that involves an initial targeting step to introduce a GOI flanked by two mismatched (heterotypic) lox sites. Recombinant clones are then co-transfected with Cre expression vector and a plasmid carrying two heterotypic lox sites similar to that used in the targeting vector. A transgene is usually present between the heterotypic lox sites of the plasmid vector. Cre activity will result in pairing of the corresponding heterotypic sites in the chromosomal and plasmid DNA. The result is an equivalent exchange of sequences flanked by the hetorotypic lox sites. This approach minimizes undesirable post-recombination outcomes since only the matching lox sites can effectively recombine.

7.6 Delivery of Recombinases into PSCs

Conventionally, Cre or Flp-expressing plasmids are introduced into PSCs via electroporation. In an ideal situation, the transient burst of recombinase expression results in the desired recombination event and the recombinase-expressing plasmid is subsequently lost as PSCs undergo cell division. Nevertheless, as in any gene delivery attempt involving DNA, there is always a risk of plasmid integration into the host genome. An alternative delivery method such as protein transduction can be employed to eliminate this risk. This method makes use of short peptides called cell penetrating peptides (CPPs) or protein transduction domains to carry macromolecules into cells (Gump and Dowdy 2007). Peitz and colleagues have developed a cell-permeable Cre protein called HTN-Cre (HTNC) (Peitz et al. 2002). HTNC is composed of an N-terminal cell permeation domain derived from the human immunodeficiency virus (HIV) transactivator of transcription (TAT) protein fused to a nuclear localization sequence (NLS). They showed that high recombination efficiencies can be achieved in fibroblasts and mouse ESCs using this method. As mentioned earlier, a cell-permeable version of the Flp protein has been generated by Patsch and colleagues and has been successfully used to perform genome manipulations in both mouse and human ESCs (Patsch et al. 2010).

7.7 Applications of Cre/loxP and Flp/FRT Systems in the Introduction of Transgenes to PSCs

At the time of writing, there is less extensive history on the genetic modification of human ES for the creation of iPS reporter cell lines and disease models compared to efforts in mouse. This may be attributed to the relatively recent discovery of human

ESCs and to technical challenges in their manipulation. Nevertheless, a number of elegant Cre/loxP and Flp/FRT-based strategies for transgene integrations have been reported and will be the focus of this section. Wherever necessary, studies using mouse PSCs will be described.

7.7.1 Creation of Reporter Cell Lines

In its simplest sense, a reporter gene should allow easy identification of a cell population or event of interest. However, in more demanding applications such as the study of developmental genes and cell lineage specification, it is critical that reporter genes faithfully mimic the spatiotemporal patterns and expression levels of their physiological counterparts. Positional effects are always a concern in expression data obtained via the random integration of transgenes. For example, random integration at hypermethylated regions of the genome may result in the transgene silencing. The opposite, but equally unwanted outcome is a loss of spatial and temporal control of gene expression when the transgene randomly inserts into an area of constitutive expression.

The site specificity of Cre and Flp recombination systems make them superior to random integration-based methods for the creation of reporter cell lines. However, it should be noted that the risk of undesirable integrations cannot be completely eliminated even with site-specific integration approaches. Moreover, some studies have reported the presence of pseudo loxP sites in mammalian genomes (Thyagarajan et al. 2000; Schmidt et al. 2000) that may potentially allow illegitimate Cre and Flp-mediated recombination. Although most of the reported pseudo sites recombine with legitimate loxP sequences at a lower efficiency compared to legitimate loxP pairings, a determination of transgene copy number must be performed for recombinant cell lines if the requirement of the study is single copy expression of a transgene.

7.7.2 Turning on Transgenes via Recombinase-Mediated Inversions

An early proof-of-principle on the use of the Cre/lox system to generate a reporter hESC line was performed by Irion and colleagues (Irion et al. 2007). This study made use of a combination of Cre-mediated excision and inversion reactions to reconstitute expression of a transgene (Fig. 7.5). First, gene targeting introduced a viral splice acceptor followed by a loxP-flanked promoterless neomycin-resistance gene (neomycin^r). The neomycin gene was followed by an inverted tandem RFP variant, tdRFP which in turn was flanked by mutant lox2272 sites. Cre-mediated recombination may first induce inversion of the intervening DNA at either the loxP

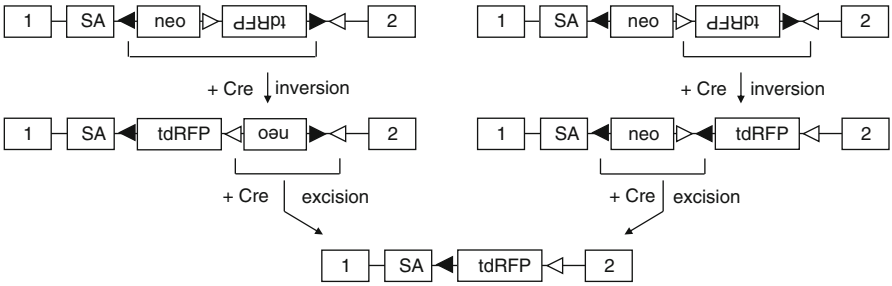


Fig. 7.5 Cre recombinase-mediated excision-inversion reaction to reconstitute expression of a fluorescent transgene. Diagram shows targeting vector between exon 1 and 2 of the human homolog of the ROSA26 locus. Lox P: *filled triangles*, lox2272: *white triangles*, SA: splice acceptor. Two scenarios are presented. Cre-mediated activity catalyzes inversion of sequences between two loxP (*left panel*) or two lox2272 (*right panel*) sites. As a second Cre-mediated event, two lox2272 (*left panel*) or two loxP (*right panel*) sites recombine resulting in excision of the neomycin resistance marker. In both scenarios, expression of the tdRFP transgene is reconstituted

or the lox2272 sites. This will cause inversion of the promoterless tdRFP, with its ATG codon now at the same orientation as the endogenous promoter. Further Cre-mediated activity will then excise neomycin^r located between the two loxP or lox2272 sites, finally leaving two mismatched lox sites which are incapable of further recombination. This is indeed an elegant way of minimizing the potential for unwanted post-Cre recombination outcomes. However, a word of caution is given to those who intend to eliminate a potentially recombinogenic lox site flanked by matching lox variants as one study has reported that the distance between the compatible lox sites should be at least 82 base pairs to ensure efficient recombination (Hoess et al. 1985).

The report by Irion and colleagues also demonstrated how heterotypic lox sites could be strategically positioned to limit the unwanted excision of the GOI, in this case, tdRFP (Irion et al. 2007). The other findings of this study mirror what has been established in the murine, that is, transgenes integrated into the ROSA26 locus retain their expression throughout differentiation into lineages of the three germ layers and does not adversely affect the pluripotency of the recombinant cell lines (Zambrowicz et al. 1997; Soriano 1999). Hence, the human homolog of ROSA26 is considered a candidate "safe-harbor" site for the introduction of transgenes into human PSCs.

7.7.3 Cre/loxP, Flp/FRT and Chemically Inducible Transgene Expression Systems

The Cre and Flp systems may also be used to circumvent bottlenecks in the study of gene function. For example, expression of certain mutant forms of a gene may result in early lethality, making it impossible to generate viable cells past a certain developmental stage. This can be circumvented by placing the transgene under the control of

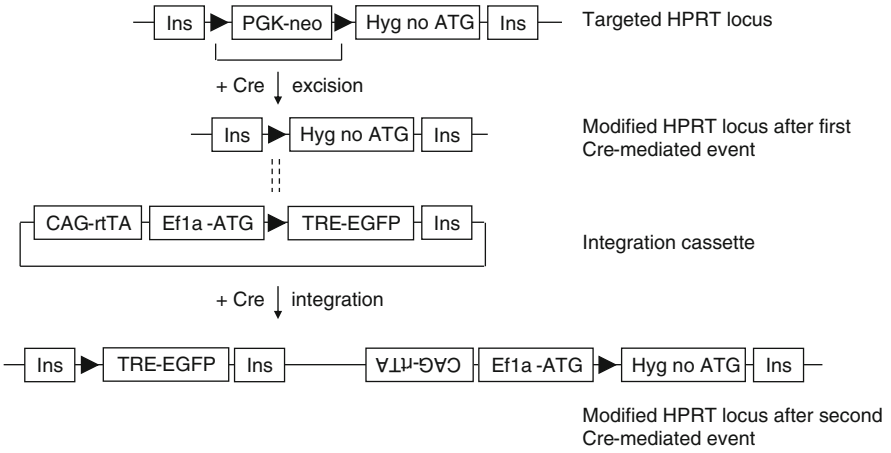


Fig. 7.6 Use of Cre recombinase for site-specific introduction of EGFP transgene under the control of tetracycline regulatory elements. Figure shows targeted HPRT locus and the integration cassette. Cre activity will result in excision of the PGK promoter-driven neomycin resistance marker and the integration of a cassette carrying elements of the tetracycline inducible system. Integration of the cassette also results in reconstitution of hygromycin expression

inducible systems such as the tetracycline (Gossen and Bujard 1992) or tamoxifen (Littlewood et al. 1995) systems. We have shown proof of principle for this approach in one of our publications (Sakurai et al. 2010) where we performed gene targeting at the HPRT locus of a human female ESC line and subsequently introduced the tetracycline inducible elements via Cre-mediated recombination. (Fig. 7.6). We first generated a mutant HPRT allele with a loxP site 5′ to a neomycin expression cassette. Another loxP site was located 5′ to a promoterless hygromycin resistance gene lacking the ATG codon, with each end of the targeting construct flanked by an insulator sequence. After gene targeting, we performed co-transfection of a Cre-expression vector with a plasmid carrying an EF1α promoter, a Kozak sequence, an ATG codon, a loxP site and an EGFP transgene under the control of tetracycline responsive element (TRE). Also in this plasmid is the "Tet-on" tetracycline transactivator, rtTA under the control of CAGGS promoter (fusion of cytomegalovirus immediate early enhancer/chicken beta-actin promoter). Subsequent Cre-mediated recombination results in the excision of the neomycin expression cassette and the introduction of the plasmid carrying the EGFP transgene and the promoterless hygromycin. This also results in repositioning of the promoterless hygromycin gene 3' to the EF1α promoter and the start codon, thereby reconstituting hygromycin expression. In our hands, all of the hygromycin-resistant clones that were picked underwent the desired recombination event. Hence, this strategy can be employed to significantly minimize the number of clones to be picked and screened.

However, in our hands, we observed that the response to doxycycline induction varied from clone to clone. Hence, we recommend testing a number of clones to verify levels of transgene induction as they go through continuous passage and directed differentiation.

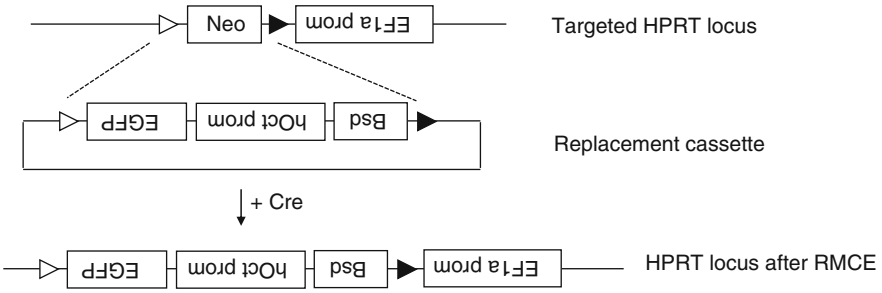


Fig. 7.7 RMCE approach to generate human ESC reporter line. LoxP (*black triangles*) and lox2272 (*white triangles*) in the targeted HPRT locus pair with their corresponding partners in the exchange cassette to facilitate transgene swapping. As the end result of RMCE, a human Oct4 (*hOct4*) promoter-driven EGFP transgene is introduced and expression of blastocidin (*bsd*) resistance marker is reconstituted. EF1a prom: elongation factor 1 alpha promoter, neo: neomycin resistance marker

7.7.4 Recombination-Mediated Cassette Exchange (RMCE)

The RMCE approach involves an exchange between a transgene flanked by mismatched lox sites and a corresponding plasmid carrying a transgenes flanked by the same types of mismatched lox sites found in the chromosomal locus. Di Domenico and colleagues replaced a chromosomal EF1α promoter-driven neomycin resistance transgene with a human Oct4 promoter-driven EGFP transgene using loxP and lox2272 sites (Di Domenico et al. 2008) (Fig. 7.7). Clones that underwent the desired recombination event become resistant to blasticidin through Cre-mediated reconstitution of this selection marker.

It is important that the choice of heterotypic lox sites take into consideration the potential for illegitimate recombination between non-target lox sites. Certain mismatched lox sites have a greater tendency to illegitimately recombine more than others. In fact, one study has demonstrated cassette replacement via recombination between loxP and lox511 sites (Soukharev et al. 1999). Siegel and colleagues reported the rates of illegitimate recombination between certain mutant lox sites (Siegel et al. 2001).

7.7.5 Site-Specific Integration of BACs

The use of bacterial artificial chromosomes (BACs) in the creation of reporter cell lines is expected to confer faithful transgene expression compared to the use of shorter promoter constructs. This is due to the ability of BACs to house longer DNA sequences that retain full length gene regulatory elements. Placantonakis and colleagues provided a proof of principle for the use of BACs in the creation of hESC reporter lines (Placantonakis et al. 2009). In their study, BACs carrying an EGFP transgene under the control of full-length regulatory elements of neuron-specific

genes were introduced to hESCs via random integration. Using this approach, they were able to isolate clones that maintained proper spatio-temporal expression of EGFP as differentiation progressed towards the neuronal lineage. It will be worthwhile to examine whether introduction of BACs via site-specific integration into "safe harbor" sites will increase the efficiency of generating high fidelity human reporter cell lines.

However, a number of concerns have been raised in recombinase-mediated integration of BACs. Due to their larger size, BACs may be less efficiently introduced into chromosomal regions via Cre. The use of mammalian codon-optimized Flp (Raymond and Soriano 2007) and Cre sequences (Shimshek et al. 2002) or the introduction of cell-permeant versions (Nolden et al. 2006; Patsch et al. 2010) may potentially help offset the decreased efficiency of these recombinases on large constructs such as BACs. BACs may also contain too many unwanted prokaryotic genes or regulatory elements. It is therefore necessary to carefully consider intermediate steps in BAC construction to generate final constructs that are devoid of unwanted sequences or perhaps, employ a strategy to eliminate these sequences via a secondary recombination event in the PSC line.

To date, a number of human BAC libraries have been generated and are available to researchers (BACPAC Resources 2012; Invitrogen 2012; see references for websites). Specifically, BAC libraries have been created for the H1 and H9 hESC lines (BACPAC Resources 2012; see reference for website). These two libraries were generated in a transposon-free *E. coli* host to minimize the possibility of sequence disruption caused by random insertions of prokaryotic transposon elements.

7.8 Technological Improvements and Future Outlook

Site-directed integration of transgenes is usually preceded by a gene targeting step. Hence, Cre and Flp-mediated recombination is expected to gain more widespread application in the creation of human PSC reporter lines as better gene targeting methods and more robust human PSC lines are developed. Consequently, there will be a need to expand the current list of well-characterized loci available for site-specific integration. HPRT locus has been extensively studied and has been favored in many genetic manipulation attempts for both mouse and human PSCs due to the low risk of transgene silencing after differentiation. However, it is also known that null mutations in this gene result in Lesch-Nyhan syndrome. Hence, the utility of HPRT as an integration site may be limited to cell types that are not substantially affected by HPRT depletion.

At present, a number of candidate "safe-harbor" sites exist such as the adeno-associated virus integration site 1 (AAVS1) in chromosome 19 and the previously discussed ROSA26 locus. One group has also shown that integration of an EGFP transgene between the thrombopoietin (*TPO* or *THPO*) and solute carrier 25 (*SC25*) genes of human chromosome 12 does not disrupt any known gene and can facilitate reliable reporter gene expression in derivatives of the three germ layers (Costa et al.

2005). Another study has suggested that the region between IGFBP7 and LOC255130 in human chromosome 4 and that flanked by CLDN3 and CLDN4 may serve as potential integration sites that have a low risk of transgene silencing after differentiation (Du et al. 2009). However, further studies in human cells are needed to confirm that no refractory phenotypes will result from disruptions of these regions.

Any technology is always at risk of being overtaken by newer methodologies. As expected, the discovery of the PhiC31 recombination system and the development of zinc finger nuclease-assisted genetic manipulations, both discussed in other chapters of this book, have provided alternative approaches to site-directed transgene integration. Nevertheless, the relative simplicity, accessibility to researchers, the low cost and the availability of a number of pre-existing lox and FRT-carrying PSC lines will continue to make the Cre and Flp systems a practical method of choice for the site-directed integration of transgenes.

References

Abremski K, Hoess R, Sternberg N (1983) Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. Cell 32(4):1301–1311, PubMed PMID: 6220808

Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J 7(4):649–659, PubMed PMID: 7742860

Anastassiadis K, Fu J, Patsch C, Hu S, Weidlich S, Duerschke K, Buchholz F, Edenhofer F, Stewart AF (2009) Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli*, mammalian cells and mice. Dis Model Mech 2(9–10):508–515, Epub 2009 Aug 19. PubMed PMID: 19692579

Andrews BJ, Proteau GA, Beatty LG, Sadowski PD (1985) The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences. Cell 40(4):795–803, PubMed PMID: 3879971

Araki K, Okada Y, Araki M, Yamamura K (2010) Comparative analysis of right element mutant lox sites on recombination efficiency in embryonic stem cells. BMC Biotechnol 10:29, PubMed PMID: 20356367; PubMed Central PMCID: PMC2865440

Austin S, Ziese M, Sternberg N (1981) A novel role for site-specific recombination in maintenance of bacterial replicons. Cell 25(3):729–736, PubMed PMID: 7026049

BACPAC Resources (2012) website http://bacpac.chori.org

Baer A, Bode J (2001) Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. Curr Opin Biotechnol 12(5):473–480, Review. PubMed PMID: 11604323

Beggs JD (1978) Transformation of yeast by a replicating hybrid plasmid. Nature 275(5676):104–109, PubMed PMID: 357984

Bode J, Schlake T, Iber M, Schübeler D, Seibler J, Snezhkov E, Nikolaev L (2000) The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. Biol Chem 381(9–10):801–813, Review. PubMed PMID: 11076013

Broach JR, Hicks JB (1980) Replication and recombination functions associated with the yeast plasmid, 2 mu circle. Cell 21(2):501–508, PubMed PMID: 7407923

Broach JR, Guarascio VR, Jayaram M (1982) Recombination within the yeast plasmid 2mu circle is site-specific. Cell 29(1):227–234, PubMed PMID: 6286142

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA, Vallier L (2007) Derivation of pluripotent

epiblast stem cells from mammalian embryos. Nature 448(7150):191–195, Epub 2007 Jun 27. PubMed PMID: 17597762

Buchholz F, Ringrose L, Angrand PO, Rossi F, Stewart AF (1996) Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. Nucleic Acids Res. 1996 Nov 1; 24(21):4256–62, PubMed PMID: 8932381; PubMed Central PMCID: PMC146240

Buchholz F, Angrand PO, Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. Nat Biotechnol 16(7):657–662, PubMed PMID: 9661200

Buecker C, Chen HH, Polo JM, Daheron L, Bu L, Barakat TS, Okwieka P, Porter A, Gribnau J, Hochedlinger K, Geijsen N (2010) A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. Cell Stem Cell 6(6):535–546, PubMed PMID: 20569691

Chen Y, Narendra U, Iype LE, Cox MM, Rice PA (2000) Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. Mol Cell 6(4):885–897, PubMed PMID: 11090626

Costa M, Dottori M, Ng E, Hawes SM, Sourris K, Jamshidi P, Pera MF, Elefanty AG, Stanley EG (2005) The hESC line Envy expresses high levels of GFP in all differentiated progeny. Nat Methods 2(4):259–260, Epub 2005 Mar 23. PubMed PMID: 15782217

Dang DT, Perrimon N (1992) Use of a yeast site-specific recombinase to generate embryonic mosaics in Drosophila. Dev Genet 13(5):367–375

Di Domenico AI, Christodoulou I, Pells SC, McWhir J, Thomson AJ (2008) Sequential genetic modification of the hprt locus in human ESCs combining gene targeting and recombinase-mediated cassette exchange. Cloning Stem Cells 10(2):217–230, PubMed PMID: 18386992

Du ZW, Hu BY, Ayala M, Sauer B, Zhang SC (2009) Cre recombination-mediated cassette exchange for building versatile transgenic human embryonic stem cells lines. Stem Cells 27(5):1032–1041, PubMed PMID: 19415769; PubMed Central PMCID: PMC2801346

Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. Nature 292(5819):154–156, PubMed PMID: 7242681

Fiering S, Kim CG, Epner EM, Groudine M (1993) An "in-out" strategy using gene targeting and FLP recombinase for the functional dissection of complex DNA regulatory elements: analysis of the beta-globin locus control region. Proc Natl Acad Sci USA 90(18):8469–8473, PubMed PMID: 8378321; PubMed Central PMCID: PMC47378

Fiering S, Epner E, Robinson K, Zhuang Y, Telling A, Hu M, Martin DI, Enver T, Ley TJ, Groudine M (1995) Targeted deletion of 5′HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. Genes Dev 9(18):2203–2213, PubMed PMID: 7557375

Ghosh K, Lau CK, Gupta K, van Duyne GD, Ghosh K, Lau CK, Gupta K, van Duyne GD (2005) Preferential synapsis of loxP sites drives ordered strand exchange in Cre-loxP site-specific recombination. Nat Chem Biol 1(5):275–282, Epub 2005 Sep 11. PubMed PMID: 16408057

Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci USA 89(12):5547–5551, PubMed PMID: 1319065; PubMed Central PMCID: PMC49329

Grindley ND, Whiteson KL, Rice PA, Grindley ND, Whiteson KL, Rice PA (2006) Mechanisms of site-specific recombination. Annu Rev Biochem 75:567–605, Review. PubMed PMID: 16756503

Gu H, Zou YR, Rajewsky K (1993) Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Cell 73(6):1155–1164, PubMed PMID: 8513499

Guhr A, Kurtz A, Friedgen K, Löser P (2006) Current state of human embryonic stem cell research: an overview of cell lines and their use in experimental work. Stem Cells 24(10):2187–2191, Epub 2006 Jun 15. Review. PubMed PMID: 16778154

Gump JM, Dowdy SF (2007) TAT transduction: the molecular mechanism and therapeutic prospects. Trends Mol Med 13(10):443–448, Review. PubMed PMID: 17913584

Guo F, Gopaul DN, van Duyne GD (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature 389(6646):40–46, PubMed PMID: 9288963

Hochman L, Segev N, Sternberg N, Cohen G (1983) Site-specific recombinational circularization of bacteriophage P1 DNA. Virology 131(1):11–17, PubMed PMID: 6228057

Hoess RH, Abremski K (1985) Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. J Mol Biol 181(3):351–362, PubMed PMID: 3856690

Hoess RH, Ziese M, Sternberg N (1982) P1 site-specific recombination: nucleotide sequence of the recombining sites. Proc Natl Acad Sci USA 79(11):3398–3402, PubMed PMID: 6954485; PubMed Central PMCID: PMC346427

Hoess R, Wierzbicki A, Abremski K (1985) Formation of small circular DNA molecules via an in vitro site-specific recombination system. Gene 40(2–3):325–329

Hoess RH, Wierzbicki A, Abremski K (1986) The role of the loxP spacer region in P1 site-specific recombination. Nucleic Acids Res 14(5):2287–2300, PubMed PMID: 3457367; PubMed Central PMCID: PMC339658

Invitrogen (2012) website http://www.invitrogen.com

Irion S, Luche H, Gadue P, Fehling HJ, Kennedy M, Keller G (2007) Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nat Biotechnol 25(12):1477–1482, Epub 2007 Nov 25. PubMed PMID: 18037879

Lee G, Saito I (1998) Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. Gene 216(1):55–65, PubMed PMID: 9714735

Littlewood TD, Hancock DC, Danielian PS, Parker MG, Evan GI, Littlewood TD, Hancock DC, Danielian PS, Parker MG, Evan GI (1995) A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. Nucleic Acids Res 23(10):1686–1690, PubMed PMID: 7784172; PubMed Central PMCID: PMC306922

Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci USA 78(12):7634–7638, PubMed PMID: 6950406; PubMed Central PMCID: PMC349323

McLeod M, Craft S, Broach JR (1986) Identification of the crossover site during FLP-mediated recombination in the *Saccharomyces cerevisiae* plasmid 2 microns circle. Mol Cell Biol 6(10):3357–3367, PubMed PMID: 3540590; PubMed Central PMCID: PMC367081

Nakatsuji N, Suemori H (2002) Embryonic stem cell lines of nonhuman primates. ScientificWorldJournal 2:1762–1773, Review. PubMed PMID: 12806169

Nolden L, Edenhofer F, Haupt S, Koch P, Wunderlich FT, Siemen H, Brüstle O (2006) Site-specific recombination in human embryonic stem cells induced by cell-permeant Cre recombinase. Nat Methods 3(6):461–467, PubMed PMID: 16721380

Okita K, Ichisaka T, Yamanaka S (2007) Generation of germline-competent induced pluripotent stem cells. Nature 448(7151):313–317, Epub 2007 Jun 6. PubMed PMID: 17554338

Patsch C, Peitz M, Otte DM, Kesseler D, Jungverdorben J, Wunderlich FT, Brüstle O, Zimmer A, Edenhofer F (2010) Engineering cell-permeant FLP recombinase for tightly controlled inducible and reversible overexpression in embryonic stem cells. Stem Cells 28(5):894–902, PubMed PMID: 20333748

Peitz M, Pfannkuche K, Rajewsky K, Edenhofer F (2002) Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. Proc Natl Acad Sci USA 99(7):4489–4494, Epub 2002 Mar 19. PubMed PMID: 11904364; PubMed Central PMCID: PMC123675

Placantonakis DG, Tomishima MJ, Lafaille F, Desbordes SC, Jia F, Socci ND, Viale A, Lee H, Harrison N, Tabar V, Studer L (2009) BAC transgenesis in human embryonic stem cells as a novel tool to define the human neural lineage. Stem Cells 27(3):521–532, PubMed PMID: 19074416

Rank GH, Arndt GM, Xiao W (1989) FLP-FRT mediated intrachromosomal recombination on a tandemly duplicated YEp integrant at the ILV2 locus of chromosome XIII in *Saccharomyces cerevisiae*. Curr Genet 15(2):107–112, PubMed PMID:2663188

Raymond CS, Soriano P (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. PLoS One 2(1):e162, PubMed PMID: 17225864; PubMed Central PMCID: PMC1764711

Ringrose L, Lounnas V, Ehrlich L, Buchholz F, Wade R, Stewart AF (1998) Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. J Mol Biol 284(2):363–384, PubMed PMID: 9813124

Sakurai K, Shimoji M, Tahimic CG, Aiba K, Kawase E, Hasegawa K, Amagai Y, Suemori H, Nakatsuji N (2010) Efficient integration of transgenes into a defined locus in human embryonic stem cells. Nucleic Acids Res 38(7):e96, Epub 2010 Jan 13. PubMed PMID: 20071742; PubMed Central PMCID: PMC2853137

Sauer B (1987) Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. Mol Cell Biol 7(6):2087–2096, PubMed PMID: 3037344; PubMed Central PMCID: PMC365329

Sauer B, Henderson N (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc Natl Acad Sci USA 85(14):5166–5170, PubMed PMID: 2839833; PubMed Central PMCID: PMC281709

Schlake T, Bode J (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. Biochemistry 33(43):12746–12751, PubMed PMID: 7947678

Schmidt EE, Taylor DS, Prigge JR, Barnett S, Capecchi MR (2000) Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. Proc Natl Acad Sci USA 97(25):13702–13707, PubMed PMID: 11087830; PubMed Central PMCID: PMC17639

Segev N, Cohen G (1981) Control of circularization of bacteriophage P1 DNA in *Escherichia coli*. Virology 114(2):333–342, PubMed PMID: 7027600

Senecoff JF, Cox MM (1986) Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. J Biol Chem 261(16):7380–7386, PubMed PMID: 3711092

Shimshek DR, Kim J, Hübner MR, Spergel DJ, Buchholz F, Casanova E, Stewart AF, Seeburg PH, Sprengel R (2002) Codon-improved Cre recombinase (iCre) expression in the mouse. Genesis 32(1):19–26, PubMed PMID: 11835670

Siegel RW, Jain R, Bradbury A (2001) Using an in vivo phagemid system to identify non-compatible loxP sequences. FEBS Lett 499(1–2):147–153, Corrected and republished in: FEBS Lett. 2001 Sep 21;505(3):467–73. PubMed PMID: 11418130

Soriano P (1999) Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet 21(1):70–71, PubMed PMID: 9916792

Soukharev S, Miller JL, Sauer B (1999) Segmental genomic replacement in embryonic stem cells by double lox targeting. Nucleic Acids Res 27(18):e21, PubMed PMID: 10471751; PubMed Central PMCID: PMC148613

Sternberg N, Hamilton D (1981) Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. J Mol Biol 150(4):467–486, PubMed PMID: 6276557

Sternberg N, Hamilton D, Hoess R (1981) Bacteriophage P1 site-specific recombination. II. Recombination between loxP and the bacterial chromosome. J Mol Biol 150(4):487–507, PubMed PMID: 6276558

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126(4):663–676, Epub 2006 Aug 10. PubMed PMID: 16904174

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131(5):861–872, PubMed PMID: 18035408

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448(7150):196–199, Epub 2007 Jun 27. PubMed PMID: 17597760

Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Becker RA, Hearn JP (1995) Isolation of a primate embryonic stem cell line. Proc Natl Acad Sci USA 92(17):7844–7848, PubMed PMID: 7544005; PubMed Central PMCID: PMC41242

Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Hearn JP (1996) Pluripotent cell lines derived from common marmoset (*Callithrix jacchus*) blastocysts. Biol Reprod 55(2):254–259, PubMed PMID: 8828827

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. Science 282(5391):1145–1147, Erratum in: Science 1998 Dec 4;282(5395):1827. PubMed PMID: 9804556

Thyagarajan B, Guimarães MJ, Groth AC, Calos MP (2000) Mammalian genomes contain active recombinase recognition sites. Gene 244(1–2):47–54, PubMed PMID: 10689186

Vetter D, Andrews BJ, Roberts-Beatty L, Sadowski PD (1983) Site-specific recombination of yeast 2-micron DNA in vitro. Proc Natl Acad Sci USA 80(23):7284–7288, PubMed PMID: 6316354; PubMed Central PMCID: PMC390039

Wu S, Ying G, Wu Q, Capecchi MR (2007) Toward simpler and faster genome-wide mutagenesis in mice. Nat Genet 39(7):922–930, Epub 2007 Jun 17. PubMed PMID: 17572674

Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P (1997) Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. Proc Natl Acad Sci USA 94(8):3789–3794, PubMed PMID: 9108056; PubMed Central PMCID: PMC20519

Chapter 8
Site-Specific Recombination Using PhiC31 Integrase

Jonathan M. Geisinger and Michele P. Calos

Abstract The integrase from phage ϕC31 of *Streptomyces* bacteria is an attractive recombinase for use in generating transgenic organisms and developing gene and cell therapeutic strategies. In nature, ϕC31 integrase mediates integration by interacting with specific sites in the phage and bacterial genomes. When applied to eukaryotes, ϕC31 integrase provides efficient unidirectional recombination between its own *attB* and *attP* sites or between an *attB* site on an incoming plasmid and a native genomic pseudo *attP* site that resembles *attP*. To date, the ϕC31 system has been used to generate stable transgenic organisms from multiple species, including plants, insects, and vertebrates. The features of the ϕC31 system also make it particularly amenable to therapeutic strategies. ϕC31 integrase has been used in potential therapies for numerous genetic diseases including hemophila, muscular dystrophy, and skin disorders. Additionally, the ϕC31 system has recently been used to modify human embryonic stem cells and to generate induced pluripotent stem cells. The ϕC31 system can also be combined with other recombinases to create advanced genome engineering strategies. In the future, the use of ϕC31 integrase may facilitate the development of new gene and cell therapies, as well as the generation of targeted transgenic organisms.

Keywords Gene therapy • Induced pluripotent stem cells • Phage integrase • Recombinases • Site-specific integration • Transgenic organisms

J.M. Geisinger • M.P. Calos (✉)
Department of Genetics, Stanford University School of Medicine,
300 Pasteur Drive, Stanford, CA 94305-5120, USA
e-mail: calos@stanford.edu

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_8,

8.1 Introduction

The ability to generate transgenic cells and organisms by genomic integration is useful for the dual purposes of basic research and therapeutic applications. However, random integration is associated with significant caveats: There is little or no control over the location and number of integrations, resulting in inconsistent expression and the risk of insertional mutagenesis. Insertional mutagenesis is a concern because of possible integration near an oncogene or tumor suppressor gene, which could cause cancer. These caveats can be mitigated through the use of integration directed by site-specific recombinases. The majority of site-specific recombinases are found in bacteria and their viruses, known as phages, but a small number are eukaryotic in origin (Brown et al. 2011). Site-specific recombinases are useful because many of them function outside their organism of origin (Brown et al. 2011). The site-specific recombinases are divided into two separate families, based on evolutionary history and integration mechanism. According to which amino acid residue catalyzes the integration reaction, the families are known as the tyrosine and the serine recombinases (Brown et al. 2011).

The tyrosine recombinase family is found mainly in prokaryotes, but members have also been observed in archaea, fungi, ciliates, and in certain families of retrotransposons (Grindley et al. 2006). Members of this family include λ phage integrase, Cre resolvase, Flp invertase, and many others (Grindley et al. 2006). In nature, λ integrase is involved in both integration and excision reactions, whereas Cre performs excision, and Flp facilitates inversions (Grindley et al. 2006). The use of Cre and Flp in site-specific recombination is covered elsewhere in this volume. Despite the tyrosine recombinases being site-specific, their tendency to be able to catalyze both integration and excision, as well as the requirement of co-factors for some members such as λ integrase, makes them less than optimal for stable and efficient DNA integration (Brown et al. 2011).

The serine recombinases, in contrast to the tyrosine recombinases, are more heterogeneous both in their size range (from 180 to 800 amino acids) and in their domain organization (Grindley et al. 2006). Members of the serine recombinase family include the γδ and Tn*3* resolvases, the integrase from the mycobacteriophage Bxb1, and the integrase from the *Streptomyces* phage ϕC31 (Brown et al. 2011). Within this group, many of the serine integrases do not have species-specific co-factor requirements and thus have the ability to function in organisms other than their organism of origin (Keravala et al. 2006a). Furthermore, integration mediated by these enzymes is unidirectional, rather than reversible (Brown et al. 2011; Calos 2006). These properties make the serine integrases attractive for use in site-specific recombination systems.

In this chapter, we will focus on the site-specific recombination system based on the serine recombinase ϕC31 integrase. In recent years, this system has been utilized for a variety of applications, ranging from generating transgenic insects and vertebrates to developing therapies for monogenic diseases and to generating induced pluripotent stem cells. We will examine how ϕC31 integrase facilitates integration

of the phage genome and transgenes. The advantages of ϕC31 integrase over other serine recombinases and the methods used to examine the site-specificity of integration events will be discussed. The varied uses of the ϕC31 integrase system to date will be considered, including the organisms where it has been employed, its utility for generating transgenics, and its role in developing therapeutics. Lastly, we will discuss the potential of combining the ϕC31 integrase system with other recombinases, for the purpose of advanced genome engineering.

8.2 ϕC31 Integrase: Mechanism of Integration and Optimization for Heterologous Organisms

The ϕC31 integrase is a large serine integrase derived from phage ϕC31, which infects many *Streptomyces* soil bacteria (Kuhstoss and Rao 1991; Rausch and Lehmann 1991). In its natural setting, ϕC31 integrase catalyzes the integration of the phage genome into the genome of the host bacterium. An attachment site in the phage genome, known as the *attP* site, is brought into close proximity with an attachment site in the host *Streptomyces* genome, known as the *attB* site, by the integrase. The enzyme catalyzes double-strand breaks in the *attP* and *attB* sites, facilitating strand exchange between the cleaved sites (Grindley et al. 2006; Kuhstoss and Rao 1991). The resulting hybrid *attP-attB* sites are known as *attL* and *attR* sites. Notably, ϕC31 integrase on its own cannot recombine these sites (Thorpe and Smith 1998). This recombination process is depicted in Fig. 8.1a. The unidirectionality of ϕC31 integrase, combined with its ability to facilitate integration without *Streptomyces*-specific co-factors (Thorpe and Smith 1998), makes it an appealing enzyme to use for site-specific recombination in other organisms.

Because the minimal sizes of *attP* and *attB* sites are 39 and 34 base pairs, respectively, it was hypothesized that degenerate sequences that may be functionally similar to the wild-type ϕC31 *attP* and *attB* sites should exist at relatively low frequency in eukaryotic genomes, compared to the frequencies of viral and transposon integration sites (Groth et al. 2000). However, whether ϕC31 integrase could function in eukaryotic cells was unknown. Thus, experiments were carried out in human cells to investigate the ability of ϕC31 integrase to mediate recombination between its *attB* and *attP* sites on extrachromosomal vectors and to facilitate integration between a pre-integrated, chromosomal *attP* site and an *attB* site present on a plasmid (Groth et al. 2000; Thyagarajan et al. 2001). These experiments showed that ϕC31 integrase was capable of both recombining extrachromosomal vectors and facilitating integration into a pre-placed chromosomal *attP* site in eukaryotic cells. Interestingly, the converse reaction, testing the ability of ϕC31 integrase to facilitate integration of an *attP*-containing plasmid into a pre-placed chromosomal *attB* site, was found to be less efficient (Thyagarajan et al. 2001).

These results were extended in experiments testing the ability of ϕC31 integrase to integrate plasmids into native human and mouse chromosomal sequences that resemble *att* sites, termed pseudo *att* sites. Use of ϕC31 integrase facilitated

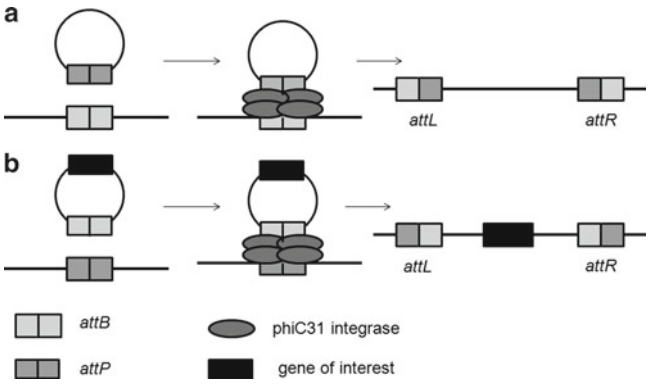


Fig. 8.1 Mechanism of ϕC31-mediated integration. (**a**) Mechanism of integration into bacterial host. In *Streptomyces* bacteria, the *attP* site of the phage genome and *attB* site of the bacterial genome are bound and brought into close proximity by ϕC31 integrase. The integrase then facilitates integration of the phage genome into the bacterial genome, generating the flanking *attL* and *attR* sites in the process. (**b**) Mechanism of integration into heterologous organism. A donor plasmid containing an *attB* site and the gene of interest is brought into close proximity with a pseudo *attP* site in the genome of a eukaryotic organism by ϕC31 integrase. The integrase facilitates integration of the donor plasmid into the pseudo *attP* site, generating *attL* and *attR* sites in the process

integration of an *attB*-containing plasmid into the genome at roughly 5–10-fold above background, indicating that there were pseudo *attP* sites in mammalian genomes and that ϕC31 integrase could recognize these sites (Thyagarajan et al. 2001). This process is illustrated in Fig. 8.1b. However, there was little integration of an *attP*-containing plasmid into pseudo *attB* sites (Thyagarajan et al. 2001). The reason for the discrepancy in integration efficiency between pseudo *attB* and pseudo *attP* sites is not understood.

After it was observed that pseudo *attP* sites exist in mammalian genomes, the specificity of ϕC31 integrase-mediated integration into the human genome was explored. A potential hotspot for ϕC31-mediated integration was identified at the human chromosomal locus 8p22 in 293 cells (Thyagarajan et al. 2001). However, this apparent hotspot was later suggested to be the result of bias introduced by the use of pooled plasmid rescue, because integration at 8p22 was only detected once among 117 independent clones (Chalberg et al. 2006). The latter integration specificity study used three different cell lines, D407, a retinal pigment epithelium line; HepG2, a liver cell line; and 293, an embryonic kidney cell line, to determine whether different cell types would utilize different sets of integration sites (Chalberg et al. 2006). In this study, ϕC31 integrase was found to have a ranked order of preferred integration sites, with some of these sites being cell line specific (Chalberg et al. 2006). Because the cell lines originated from different tissues, the observation of cell line specificity was interpreted as possible tissue specificity. 101 unique integration sites were observed among the three cell lines, and these sites were found to share a consensus sequence of ~30 bp through MEME motif analysis (Chalberg et al. 2006). However, this consensus sequence could not be used reliably to predict

the locations of pseudo *attP* sites, presumably due to contributions to specificity mediated by chromatin context.

A recent study using HeLaS3 cells examined the capability of the ϕC31 system to facilitate simultaneous, stable expression of two to four cDNA constructs within a single cell (Nishiumi et al. 2009). This study first showed that integrating a donor plasmid containing two or three tandem cDNA cassettes resulted in expression of all the cassettes. Additionally, single cells were obtained expressing two sets of two transgenes, each from different integrated donor plasmids. Placing the cHS4 insulator element from the chicken β-globin cluster between expression cassettes enhanced and stabilized transgene expression (Nishiumi et al. 2009). Similarly, another recent study has illustrated that the *Ars* insulator derived from the sea urchin is capable of promoting and stabilizing transgene expression when flanking the transgene of the donor plasmid in the ϕC31 system (Watanabe et al. 2010).

Recently, several modifications have been made to the ϕC31integrase system in an attempt to increase its efficiency in eukaryotic cells. The modifications include codon optimization of the ϕC31 integrase coding sequence, addition of a nuclear localization signal (NLS), and optimization of the *attB*-containing donor plasmid. Codon optimization of the ϕC31 integrase coding sequence was performed by applying mouse codon usage, avoiding various known deleterious sequence motifs, and reducing the number of CpG dinucleotides to reduce the risk of gene silencing (Raymond and Soriano 2007). In mouse embryonic stem (ES) cells, the codon-optimized ϕC31 integrase displayed significantly higher recombination activity compared to wild-type integrase in an excision assay between *attB* and *attP* sites pre-placed at the ROSA26 locus (Raymond and Soriano 2007). To date, this higher recombination has not been observed with regard to integration into pseudo *attP* sites.

A hypothesis to explain the relatively low recombination activity of ϕC31 integrase in mammalian cells is that only a small fraction of integrase molecules may be able to enter the nucleus and interact with genomic DNA. In an attempt to address this potential problem, ϕC31 integrase was modified to possess either an N-terminal or a C-terminal NLS, and its recombination activity was measured in extrachromosomal and stable-integrant excision assays (Andreas et al. 2002). The N-terminal NLS-ϕC31 integrase was found to perform worse than wild-type, whereas C-terminal ϕC31 integrase-NLS performed better than wild-type (Andreas et al. 2002). However, this increased recombination activity has not been replicated. Rather, two recent studies have concluded that adding an NLS to the C-terminal end of ϕC31 integrase does not significantly increase activity, and instead causes a decrease in activity (Aneja et al. 2009; Woodard et al. 2010a).

Another modification made to the ϕC31 integrase system is optimization of the donor plasmid. A concern that led to this modification was the possibility that the presence of CpG nucleotides and bacterial sequences in the donor plasmid backbone could lead to post-integrative silencing of the transgene. A recent study attempted to address post-integrative gene silencing by reducing the number of CpG dinucleotides in the plasmid backbone and by using cellular promoters to drive transgene expression, rather than the viral cytomegalovirus promoter (Aneja et al. 2009).

Reducing the number of CpG dinucleotides resulted in significantly reduced post-integrative silencing in mouse cells, but not in human cells, suggesting that species-specific differences may contribute to silencing of ϕC31-integrated transgenes. Using cellular promoters, however, resulted in stable, long-term expression of the integrated transgene (Aneja et al. 2009). Thus, whereas it may not be useful to optimize the coding sequence of ϕC31 integrase or to add a NLS sequence to either end of the protein, it may be advantageous to optimize the donor plasmid by using a cellular promoter for transgene expression. These studies underscore the observation that modifying ϕC31 integrase itself to obtain an enzyme that is better than wild-type integrase appears to be difficult.

Additionally, an attempt has been made to generate inducible versions of ϕC31 integrase. Two inducible versions of the integrase have been constructed: (1) a mifepristone-inducible version, with the hormone-binding domain of the mutated human progesterone receptor fused to the C-terminus, and (2) a 4-hydroxytamoxifen-inducible version, with the ER^{T2} domain fused to the C-terminus (Sharma et al. 2008). Whereas both inducible proteins had no activity above background in the absence of their respective inducers, in the presence of their inducers, the mifepristone-inducible ϕC31 had 40% of wild-type activity and the 4-hydroxytamoxifen-inducible ϕC31 had 75% of wild-type activity.

8.3 Use of ϕC31 Integrase in Heterologous Organisms

Following the observation that ϕC31 integrase functioned in cultured murine and human cells (Groth et al. 2000; Thyagarajan et al. 2001), the ϕC31 system has been used as a trangenesis method in many eukaryotic organisms, including plants, insects, and vertebrates. Table 8.1 lists the organisms covered and representative papers.

8.3.1 Use of ϕC31 Integrase in Plants

The ϕC31 system has recently begun to be used in plants, currently including *Nicotiana tabacum* (tobacco), *Triticum aestivum* (wheat), and *Arabidopsis thaliana*. Tobacco was the first plant model in which the ϕC31 system was utilized, albeit in a different way than it is typically used. In tobacco, ϕC31 integrase was employed to modify the plastid genome, rather than the nuclear genome (Lutz et al. 2004). An *attB* site was inserted into the plastid genome by homologous recombination, and an *attP*-containing donor plasmid was integrated into this site through the use of ϕC31 integrase (Lutz et al. 2004). Stable integration was assayed by antibiotic resistance, and ϕC31 integrase was found to be capable of facilitating integration into the plastid genome of tobacco.

In wheat, the first use of ϕC31 integrase was to generate lines stably expressing ϕC31 integrase through biolistic particle bombardment (Rubtsova et al. 2008). Integrase activity was assayed through co-bombardment with a green fluorescent

Table 8.1 Summary of eukaryotic organisms where ϕC31 integrase has been used

Organism	Representative reference
Tobacco	Lutz et al. (2004)
Arabidopsis thaliana	Gils et al. (2008)
Wheat	Rubtsova et al. (2008)
Mosquito	Meredith et al. (2011)
Fruit fly	Pfeiffer et al. (2008)
Silkworm	Nakayama et al. (2006)
Parhyale hawaiensis	Kontarakis et al. (2011)
Zebrafish	Lister (2010)
Xenopus laevis	Allen and Weeks (2005)
Chinese Hamster	Thyagarajan and Calos (2005)
Cow	Ma et al. (2006)
Mouse	Keravala et al. (2011)
Rat	Chalberg et al. (2005)
Human	Ortiz-Urda et al. (2002)

protein (GFP)-reporter plasmid that would express only if ϕC31 integrase recombined the *attP* and *attB* sites present on the plasmid. Stably-expressing lines were obtained and did not display any phenotypic abnormalities in growth or fertility (Rubtsova et al. 2008). In a subsequent study, the integrase-expressing lines were crossed with plants that contained a transgenic sequence possessing *attP* and *attB* sites, and the resulting progeny were assayed for excision (Kempe et al. 2010). Integrase-mediated excision was observed in the progeny and was found to be transmitted through the germ line (Kempe et al. 2010). These two studies were the first applications of serine-recombinase-mediated genome engineering in a monocotyledonous plant species.

Because *Arabidopsis thaliana* is an important model organism, a site-specific method of transgenesis would be useful. In one study, ϕC31 integrase was used in *Arabidopsis* as part of a proof-of-principle experiment to create a widely applicable hybrid seed system (Gils et al. 2008). An *Agrobacterium*-transformed line containing two transgenes, coding for male sterility and herbicide resistance, with each gene flanked by an *attP* site and an *attB* site, was crossed with a ϕC31-integrase-expressing line to generate mosaic plants. Offspring of these plants were crossed to generate heterozygous, male-sterile, herbicide-resistant plants (Gils et al. 2008). These plants were then used to generate hybrid seeds. In this study, ϕC31 integrase was used to excise either of the two transgenes. More recently, the ability of ϕC31-mediated transgene excision to be germinally transmitted in offspring lacking the ϕC31 integrase gene was examined (Thomson et al. 2010). The results of this experiment showed that the excision reaction is not newly generated in each generation and that ϕC31-integrase-mediated excision is heritable in the absence of the integrase gene. As the use of ϕC31 integrase becomes more widespread in plants, it is likely to be employed for additional applications, beyond excision of transgenes.

8.3.2 Use of ϕC31 Integrase in Invertebrates

Among invertebrates, use of the ϕC31 system has been extensive in insects, particularly in the fruit fly and in mosquitoes. Use of ϕC31 integrase in the fruit fly *Drosophila melanogaster* is covered in greater depth elsewhere in this volume and will be only briefly summarized here. Transgenesis in flies previously depended on the P-element transposase, which has the limitation of random integration, giving rise to position effects on transgene expression (Groth et al. 2004). The ϕC31 system was developed as a method to integrate any gene efficiently into a specific location. In this method, *attP*-carrying lines were generated by embryonic co-injection of an *attP*- and P-element-containing plasmid with a transposase-expressing plasmid (Groth et al. 2004). Site-specific transgenic flies were then generated by injecting the *attP*-containing lines with ϕC31 integrase mRNA and a donor plasmid carrying *attB* and the transgene of interest (Groth et al. 2004). A general protocol for this method was published, providing detailed instructions on how to perform the injections (Fish et al. 2007). Flies are excellent models for how integrase can be used to generate transgenic animals.

Generating transgenic *Drosophila* using ϕC31 has become widespread and varied in its applications. For example, this method has been used to generate many transgenic *D. melanogaster* lines with reproducible transgene expression limited to small groups of adult brain cells (Pfeiffer et al. 2008). The results obtained from expression studies in these lines suggested that there are more than 50,000 transcriptional enhancers in the *D. melanogaster* genome. ϕC31-mediated transgenesis has also been combined with homologous recombination to yield a robust method for targeted mutagenesis in *Drosophila* (Gao et al. 2008). This method places an *attP* site near the target locus through homologous recombination, followed by ϕC31-mediated integration of a donor plasmid carrying a modified version of the target gene. Flies carrying the integrated gene are then crossed to flies that express the rare-cutting endonuclease I-*Cre*I to induce a double-strand break and generate flies that are homozygous for the modified gene (Gao et al. 2008).

ϕC31-mediated transgenesis has also been adapted for transgenic RNA interference in *Drosophila*. The vector for transgenic RNA interference contains an *attB* site and two multiple cloning sites, which allow for the insertion and subsequent generation of a double-stranded RNA hairpin structure that initiates post-transcriptional silencing of the target gene and induces the interference response (Ni et al. 2008). ϕC31 integrase is particularly useful for this type of manipulation, because its utilization overcomes the variability in expression associated with random integration of P-element-based RNA interference constructs. Additionally, flies successfully transformed with ϕC31 integrase are recovered at a frequency 5-fold higher than that of P-element-transformed flies (Ni et al. 2008). This method of generating transgenic RNA interference lines has been recently used to create a resource of 2282 constructs covering 2043 genes, for the purpose of studying neurogenetics in *Drosophila* (Ni et al. 2009).

Drosophila melanogaster is not the only fly species where ϕC31 integrase has been used. In the Mediterranean fruit fly *Ceratitis capitata*, transgenic flies have been generated using a similar system, involving ϕC31-mediated integration of a

donor plasmid into a *piggyBac*-placed *attP* site (Schetelig et al. 2009). Having a method to generate efficient and stable integration was deemed important in this species, because *C. capitata* is an economically important pest.

In the case of mosquitoes, ϕC31 integrase has been used in *Aedes aegypti*, *Aedes albopictus, and Anopheles gambiae*. In all three species, the use of integrase has been conceived as a method for controlling the spread of human diseases. In a study involving *A. aegypti*, the integration efficiency of ϕC31 integrase was compared to the Cre/*lox* system in the generation of stable germline transformants (Nimmo et al. 2006). Using ϕC31 integrase increased integration efficiency up to 7.9-fold higher than Cre (Nimmo et al. 2006).

A. albopictus is a vector for dengue and for the chikungunya virus (Labbé et al. 2010). Similarly to *A. aegypti*, ϕC31 integrase has been used in *A. albopictus* to demonstrate the possibility of germline transformation. This transformation was achieved by placing an *attP* site in embryos with the use of the *piggyBac* transposon, obtaining transgenic lines carrying the *attP* site, and injecting three of the lines with donor plasmid carrying an *attB* site and a DsRed marker (Labbé et al. 2010). In all three of the ϕC31-injected lines, successful integration was observed at the pre-placed *attP* site.

In *A. gambiae*, an *attP* site was also placed in the genome through the use of the *piggyBac* transposon (Meredith et al. 2011). A donor plasmid expressing a synthetic anti-*Plasmodium* peptide called Vida3 was integrated into the pre-placed *attP*. Female transgenic mosquitoes then underwent an infection challenge with two *Plasmodium* species, including one capable of causing malaria in humans, *P. falciparum* (Meredith et al. 2011). For both *Plasmodium* species, transgenic female mosquitoes carrying the Vida3 gene had a significantly reduced parasite load compared to females without Vida3. This study marked the first time ϕC31 integrase was used in a non-human organism to combat the cause of a human disease, malaria in this case.

Additionally, the ϕC31 system has been used in the silkworm *Bombyx mori*, an interesting model organism because of its potential to produce large amounts of exogeneous proteins and its utility as a model for genetic, biochemical, and physiological studies on Lepidoptera (Nakayama et al. 2006). A disadvantage of the silkworm has been the lack of an efficient, site-specific germline transformation system. The methods that exist, nuclear polyhedrosis virus and the *piggyBac* transposon, result in random integration (Nakayama et al. 2006). To address the inefficiency and randomness of the current transformation systems, the ϕC31 system was tested in the silkworm cultured cell line *Bm*N4 and found to be capable of extrachromosomal recombination in *Bm*N4 cells (Nakayama et al. 2006). To determine if the ϕC31 system was capable of facilitating integration into the genome in *Bm*N4 cells, *piggyBac* transposase was used to introduce a ϕC31 *attB* site into the genome of these cells. Then, an *attP*-containing donor plasmid expressing DsRed and an integrase-expressing plasmid were co-transfected into the *Bm*N4-*attB* cells (Nakayama et al. 2006). Surprisingly, DsRed-positive cells were only detected when ϕC31 integrase carried a nuclear localization signal. This result may have been observed due to the use of *attP* on the donor plasmid, rather than the preferred *attB*.

More recently, the ϕC31 system has been adapted for gene trapping in arthropods. This specific gene trapping approach, called integrase-mediated trap conversion (iTRAC), constitutes a generalized strategy for use in existing and emerging model organisms (Kontarakis et al. 2011). To illustrate this point, iTRAC was demonstrated in the emerging model crustacean *Parhyale hawaiensis*. iTRAC uses a *Minos*-transposon-based primary exon-trapping vector that additionally carries a ϕC31 *attP* site (Kontarakis et al. 2011). This *attP* site allows for a secondary, *attB*-containing construct to be placed into the trap. In this study, the trapping portion of the iTRAC approach was found to be capable of marking organs and populations of cells, assisting in the imaging of cardiac function and limb regeneration, and facilitating the cloning of trapped genes in *Parhyale* (Kontarakis et al. 2011). Additionally, an *attB*-containing plasmid expressing a heat-inducible EGFP gene integrated with high efficiency into the *attP* site of the trap, illustrating that the integrase portion of the iTRAC approach worked. The iTRAC approach is thus a feasible strategy for gene trapping in novel model organisms.

8.3.3 Use of ϕC31 Integrase in Vertebrates

8.3.3.1 Fish and Amphibians

The ϕC31 system has also been used extensively in transgenesis studies in vertebrates, including fish, amphibians, birds, and mammals. For example, ϕC31 integrase has recently been employed in zebrafish, where it was utilized to excise transgenes (Lister 2010; Lu et al. 2011). While these studies have not yet taken full advantage of the ϕC31 system, they have made the important observations that ϕC31 integrase functions in zebrafish cells and in the germline, and can mediate efficient excision (Lister 2010; Lu et al. 2011). These studies thus represent the initial steps in developing a site-specific trangenesis system in zebrafish.

The African clawed frog *Xenopus laevis* is a useful model for studying embryonic development. However, there were no easy-to-use transgenesis techniques for this amphibian. Thus, the potential of ϕC31 integrase to generate transgenic *X. laevis* was investigated. A donor plasmid expressing green fluorescent protein under the control of the cytomegalovirus promoter, together with ϕC31 integrase mRNA, were injected into single-cell *Xenopus* embryos, resulting in genomic integration in a site-specific manner (Allen and Weeks 2005). This study also found that position effects could be overcome by using insulators, specifically the chicken *β-globin* 5' HS4 insulator, and that tissue-specific promoters, examined through the use of a lens-specific promoter, maintained their tissue specificity when integrated. The protocol used to generate these transgenic frogs is technically simple, the most difficult step being microinjection of the plasmid DNA into the single-cell embryo (Allen and Weeks 2006, 2009).

8.3.3.2 Birds

The use of the ϕC31 integrase system in birds to date has been limited to chicken primordial germ cells. These cells were co-transfected with an integrase-expressing plasmid, together with an *attB*-containing donor plasmid expressing GFP under the control of the β-actin promoter and flanked by the cHS4 insulator element (Leighton et al. 2008). When the cHS4 insulator was not included on the donor plasmid, widespread silencing of the transgene construct most likely occurred. GFP-positive primordial germ cell colonies were obtained at a yield 20-fold higher than background integration, and the integration sites were analyzed in 12 stably-integrated primordial germ cell lines (Leighton et al. 2008). Seven out of the 11 integration events that could be mapped were found to be in repetitive sequences that were classified as the PO41 repeat sequence. This repeat sequence is thought to be present in 259 different locations in the chicken genome, with each location several kb in size (Leighton et al. 2008). This study demonstrated that ϕC31 integrase was functional in chicken primordial germ cells and may have a preference for pseudo *attP* sites in a repetitive sequence.

8.3.3.3 Mammals

Among mammals, the ϕC31 integrase system has been used in at least a half dozen species. While many studies have been focused toward therapeutic applications and will be discussed in the next section, several studies were directed toward generating transgenic mammals or laying the groundwork for doing so. A general strategy for producing transgenic mammals can be seen in Fig. 8.2a. For example, the ϕC31 system is capable of aiding in the generation of transgenic mice. This application is useful, because the traditional method, involving pronuclei injection and random integration, has the limitations of lack of control of integration site, transgene integrity, and transgene copy number (Tasic et al. 2011). The rationale for introducing the ϕC31 system into the standard transgenesis method is that using ϕC31 integrase provides considerable control over transgene integration site, copy number, and integrity, a combination that had not been achieved with other methods.

The first attempt at using ϕC31 integrase for creating transgenic mice involved injection of donor plasmid and integrase mRNA into the pronucleus of mouse single-cell embryos (Hollis et al. 2003). Transgenic early embryos and a mid-gestation mouse were reported. Genomic integration at a known integrase hotspot on chromosome 2 was documented, as well as integration at a site on chromosome 4. Full-term transgenic mice were not reported in this study (Hollis et al. 2003).

In follow-up experiments, homologous recombination was used to insert a single large *attP* fragment or three tandem, smaller *attP* sites (*attPx3*) into the *Rosa26* or the *Hipp11* loci in mouse embryonic stem cells (Tasic et al. 2011). The *Hipp11* locus was used for most of the experiments in the study, because homozygous insertions in this locus were predicted not to disrupt any genes (Tasic et al. 2011). The capability of ϕC31 integrase integrase to mediate integration into the preplaced *attP*

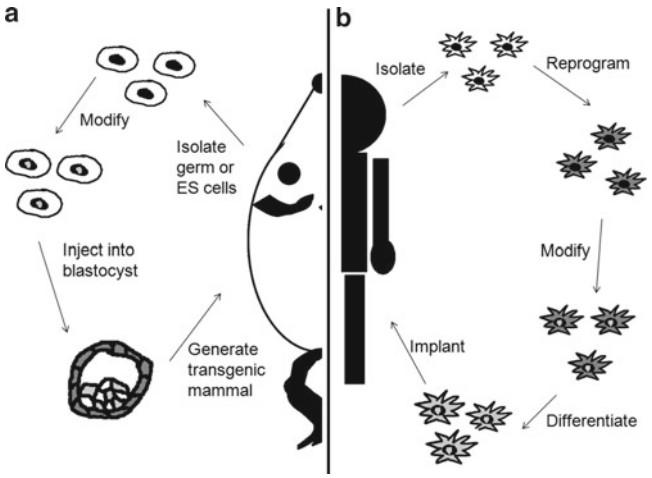


Fig. 8.2 Strategies for ϕC31-mediated modifications. (**a**) Using ϕC31integrase to generate transgenic mammals. Fertilized oocytes or embryonic stem cells are isolated, modified with ϕC31 integrase and a gene of interest, and then injected into blastocysts, which are used to generate chimeric animals that can in turn generate full transgenics. To generate transgenic insects, integrase mRNA and a donor plasmid are injected into early embryos. (**b**) Using ϕC31integrase as a strategy for cell therapy. Cells are isolated from a patient and undergo ϕC31-mediated reprogramming to become iPSC. The iPSC are then modified with another integrase and the gene of interest to generate corrected cells. The corrected iPSC are differentiated to the appropriate cell type and implanted into the patient

sites was evaluated by injecting ϕC31 integrase mRNA and a donor plasmid. One donor plasmid contained a GFP expression cassette flanked by two *attB* sites, to mediate integration into *attPx3*, while a standard donor *attB* plasmid expressing GFP was used to mediate integration into the single *attP*. These integrations were observed and were heritable, resulting in broad, robust GFP expression in transgenic mice (Tasic et al. 2011).

To determine if ϕC31 integrase could mediate tissue-specific expression using this method, the motor neuron-specific promoter of *Hb9* was used to generate mice expressing GFP (Tasic et al. 2011). GFP expression in these transgenic mice was detected in the motor neurons of the ventral spinal cord and the tail tip, indicating that tissue-specific expression was possible. Additionally, there was no statistically significant difference observed in integration efficiency between the single *attP* site and *attPx3*. Integration into the *Rosa26* locus was also successful, indicating that the use of ϕC31 integrase was not limited to just one locus for this method (Tasic et al. 2011). Some integrations did occur at locations other than the pre-placed *attP* sites. By pre-placing the target *attP* site in a transcriptionally favored location, this study introduced a modification to the standard method of generating transgenic mouse strains that may greatly aid in increasing the efficiency of the process.

The groundwork has also been laid for using ϕC31 integrase-mediated transgensis in larger mammals. The first study utilizing ϕC31 integrase in cow cells established that the integrase is functional in primary cow fibroblasts and kidney

cells and that there were at least two pseudo *attP* sites in the bovine genome (Ma et al. 2006). A second study observed two additional pseudo *attP* sites and examined whether any of the pseudo sites were potential hotspots, i.e., highly preferred integration sites (Ou et al. 2009). Out of the four known pseudo *attP* sites, only one was identified as a likely hotspot, based on observed frequency of integration in bovine fibroblasts. Additionally, the expression of GFP from the integrated donor plasmid at this potential hotspot was observed by ELISA to be greater than twofold higher than expression from the other newly observed pseudo site (Ou et al. 2009). These studies may eventually lead to the generation of transgenic cattle using the ϕC31 integrase system.

The ϕC31 integrase system has also been used in Chinese hamster ovary cells, which are often employed to produce recombinant proteins. The ϕC31 system was combined with Chinese hamster ovary cells in an effort to generate a recombinant-protein production method that was site-specific and allowed for sustained, high-level expression of the transgene (Thyagarajan and Calos 2005). This method used the standard approach of co-transfecting an integrase-expressing plasmid and a donor plasmid carrying the gene of interest and a drug resistance gene. When luciferase was used as a reporter gene, in the presence of the integrase plasmid and the selective drug, expression after ~30 days was 10,000% of the expression on day 3 after transfection. By contrast, luciferase expression in the absence of the integrase plasmid after ~30 days of selection was only 10% of day 3 levels (Thyagarajan and Calos 2005). These results suggested that the ϕC31 integrase system may stably place the reporter gene in transcriptionally favorable locations and may thus be useful as a robust method for producing recombinant proteins.

While the ϕC31 integrase system is clearly functional in many organisms, most studies address its functionality either in the context of a specific cell type or at the organismal level. There may be cell types in which ϕC31 integrase has reduced functionality. For example, in one study, ϕC31 integrase activity levels in human T lymphocytes and CD34-positive hematopoietic stem cells were compared to its activity in mesenchymal stem cells and lung, liver, and cervical carcinoma-derived cell lines (Maucksch et al. 2008). In the T-cell-derived lines, long-term expression of an integrated transgene was observed in the presence of high amounts of integrase mRNA, but not in the presence of an integrase-expressing plasmid. By contrast, co-tranfection of a donor plasmid with an integrase-expressing plasmid resulted in moderate to high levels of integration efficiency in the mesenchymal stem cells and the non-T cell-derived lines (Maucksch et al. 2008). Thus, there appeared to be some cell types in which the ϕC31 integrase system did not work well, which should be taken into consideration when planning experiments.

8.4 Therapeutic Uses of ϕC31 Integrase

Features of the ϕC31 integrase system, such as site-specificity and unidirectional recombination, render the use of ϕC31 integrase especially amenable to gene and cell therapy applications. In this section, we will cover the use of ϕC31 integrase in

developing gene therapy approaches for genetic diseases, use of the integrase system in stem cells, and the safety of ϕC31 integrase in mammalian cells.

The ϕC31 system has been used to create gene therapy strategies for single-gene disorders such as hemophila B. Hemophila B is caused by having deficient circulating levels of coagulation factor IX (FIX) (Keravala et al. 2011). The first study using the ϕC31 system to address hemophilia B examined the ability of ϕC31 integrase to function in the mouse liver, analyzing the expression levels produced by integration of a human FIX (hFIX)-expressing donor plasmid (Olivares et al. 2002). DNA delivery was achieved by hydrodynamic tail vein injection of the donor plasmid, with or without an integrase-expressing plasmid. C57BL/6J mice were injected with a liver-specific hFIX expression plasmid, with or without the integrase plasmid. After 100 days, mice that received both plasmids were producing 12-fold more hFIX than mice that received only the donor plasmid. hFIX levels remained stable in the mice that received both plasmids, even after a partial hepatectomy, whereas hFIX levels in mice receiving only the donor plasmid dropped significantly (Olivares et al. 2002). Inverse polymerase chain reaction revealed two mouse pseudo *attP* sites, mpsL1 and mpsL2, with mpsL1 appearing to be the greatly preferred integration site in mouse liver.

A more recent study examined the longevity of expression of the hFIX gene integrated in the liver with ϕC31 and whether the level of expression was therapeutic (Keravala et al. 2011). This study utilized $FIX^{-/-}$ knockout mice as a disease model. The mice were hydrodynamically injected via the tail vein with an optimized hFIX-expressing donor plasmid, with or without an integrase-expressing plasmid. After 168 days, mice that received the donor plasmid and a functional integrase plasmid had hFIX expression levels of 1,022 ± 199 ng/ml, compared to roughly an order of magnitude lower levels in mice that received a nonfunctional integrase plasmid or the donor plasmid alone. The level of circulating hFIX stabilized at roughly 10% of normal levels in the mice that received the hFIX plasmid and the integrase plasmid, which is considered to be clinically therapeutic for hemophila B. Animals that did not receive functional integrase had activity levels of 3% or less (Keravala et al. 2011). When the mice were subjected to two 10-min bleeding challenges, all of the mice that received both the integrase and hFIX plasmids were indistinguishable from wild-type mice. The previously characterized pseudo *attP* hotspot, mpsL1, was confirmed via polymerase chain reaction to be a frequent site of integration in this study (Keravala et al. 2011). These two studies demonstrated that the ϕC31system could be used to facilitate long-term, therapeutic levels of hFIX expression *in vivo* for hemophilia B. Similar studies have recently been completed in a mouse model of hemophilia A, involving ϕC31 integrase-mediated transfer of the human factor VIII gene (Chavez and Calos 2011).

A further study utilizing ϕC31 integrase-mediated gene transfer in mouse liver addressed correction of murine hereditary tyrosinemia type I (Held et al. 2005). This disease results from a deficiency of fumarylacetoacetate hydrolase (FAH) in the liver, which catalyzes the last step of tyrosine degradation. Corrected hepatocytes can readily be identified, counted, and histologically evaluated, due to the growth advantage that FAH^+ hepatocytes have over their FAH^- counterparts. In this

study, mice were hydrodynamically injected via tail vein with a donor plasmid expressing FAH and either an integrase-expressing plasmid or an empty vector, and livers were harvested after 25 days (Held et al. 2005). The frequency of stable integration in mice receiving the integrase plasmid was greater than that in mice that received the empty vector. Some of the FAH^+ hepatocytes displayed a transient dysplastic phenotype at a higher frequency than controls. This phenotype did not persist in the long term, and only normal, healthy FAH^+ hepatocytes were present after 90 days (Held et al. 2005). Analysis of the integration sites revealed seven pseudo *attP* sites, including mpsL1 (Held et al. 2005). Thus, this study illustrated that ϕC31 integrase was effective in reversing the FAH^- phenotype by mediating stable integration and expression of the FAH gene. Integrase appeared to stimulate the frequency of cells that showed an abnormal dysplastic morphology, but the abnormal phenotype was transient and had no adverse consequences. The dysplastic phenotype has not been seen in other cells exposed to ϕC31 integrase and is probably related to the inherent pathology of the FAH^- cells.

To further explore the suitability of using ϕC31 integrase in the liver, the longevity of integrase expression was examined in time course experiments in mouse liver. Integrase was found to be expressed by 2–3 h after plasmid transfection, when genomic integration could also be detected. Integrase levels declined to background within 24–48 h (Chavez et al. 2010) due to integrase plasmid loss and a short half-life of the integrase protein. The rapid expression and degradation of integrase were considered favorable to avoid undesired recombination or an immune response. Experiments in cultured cells indicated that integrase was degraded by the 26S proteasome (Chavez et al. 2010).

In contrast to *in vivo* use of integrase in the liver, the ϕC31 system is also well adapted for use in isolated cells. The blistering skin disorder known as recessive dystrophic epidermolysis bullosa (RDEB) was an appealing target for an *ex vivo* therapeutic strategy involving the ϕC31 system, because RDEB is caused by mutations in the large Type VII collagen (*COL7A1)* gene, whose cDNA is 8.9-kilobases long. Because the integrase system lacks a size limit for the donor plasmid, it is especially attractive for transfer of larger genes. Primary human RDEB keratinocytes were isolated from patients and transfected with a donor plasmid expressing COL7A1, with or without a plasmid expressing ϕC31 integrase (Ortiz-Urda et al. 2002). After 14 days, cells that received both plasmids displayed expression of full-length COL7A1, whereas cells that did not receive the integrase plasmid had no expression. The majority of integration sites identified through plasmid rescue corresponded to a previously identified pseudo *attP* hotspot in human cells (Ortiz-Urda et al. 2002). Cells that received both plasmids were engrafted into immune-deficient mice, to determine if correction of the RDEB phenotype occurred *in vivo*. Skin regenerated from control RDEB cells showed blistering and a lack of COL7A1 expression, while skin regenerated from the treated cells was phenotypically normal and displayed correctly localized expression of COL7A1 (Ortiz-Urda et al. 2002). A similar study demonstated the correction of junctional EB in patient keratinocyes by addition of the laminin-5 gene with ϕC31 integrase (Ortiz-Urda et al. 2003). These studies demonstrated

that the ϕC31 integrase system could be used in an *ex vivo* gene therapeutic strategy and could address disorders involving large genes.

Degeneration of the retina can be caused by a variety of genetic mutations, making it potentially amenable to treatment through gene therapy. Whereas there has been some gene therapy success recently with viral vectors, the potential drawbacks of immunogenicity, toxicity, and limited carrying capacity led to investigation of the use of ϕC31 integrase in the rat retina (Chalberg et al. 2005). Subretinal injection followed by *in vivo* electroporation was used to transfer a donor plasmid expressing an EGFP-luciferase fusion protein and a ϕC31 integrase-expressing plasmid to retinal pigment epithelial cells in the rat retina. When long-term expression was analyzed, luciferase expression in rats that received both plasmids was 85-fold higher than in rats that only received the donor plasmid. Expression was still present four and a half months later in rats that also received the integrase plasmid (Chalberg et al. 2005). An assay for integration into the rat genome was developed in Rat2 cells, resulting in the identification of three potential pseudo *att* sites for integration in the rat genome. When the injected retinas were assayed for evidence of integration, integration at one of the hotspots was observed, indicating that ϕC31 integrase can mediate integration at that site *in vivo* in rat retinal pigment epithelium cells (Chalberg et al. 2005). This study laid the groundwork for future studies directed at developing a non-viral, retina-specific, long-term gene therapy system.

Another genetic disease where ϕC31 integrase may be therapeutically useful is Duchenne muscular dystrophy (DMD). DMD is an X-linked disorder caused by mutations in the large gene encoding dystrophin, resulting in an absence of functional dystrophin in muscle (Quenneville et al. 2004). Because the dystrophin transcript is 13.7 kb in size, most viral vectors have insufficient capacity to carry the full-length cDNA. By contrast, because the ϕC31 integrase system does not have a size limit, it is able to transfer the full-length dystrophin transcript. In the first study employing the ϕC31 system in DMD, the ability of this integrase to function in mouse muscle-derived stem cells and human myoblasts was examined (Quenneville et al. 2004). The Amaxa nucleofector was used to transfect a GFP-expressing donor plasmid into mouse MD1 cells, isolated from the *mdx* mouse model of DMD, with or without an integrase-expressing plasmid. Parallel experiments were done in human primary myoblasts. After a week-long selection and two additional weeks in culture, threefold more resistant, GFP-expressing colonies were obtained from MD1 cells transfected with both plasmids compared to those receiving the donor plasmid alone. Fifteenfold more resistant, GFP-expressing colonies were obtained from primary human myoblasts that received both plasmids, compared to those that received only the donor plasmid (Quenneville et al. 2004).

After this confirmation that ϕC31 integrase was functional in human primary myoblasts, these cells were then transfected with a donor plasmid expressing a full-length dystrophin-GFP fusion protein and an integrase-expressing plasmid. When the cells were transfected without the integrase plasmid, fluorescence was undetectable after 1 week (Quenneville et al. 2004). In cells that received both plasmids, some fluorescent myogenic cells survived early culturing and a week of selection. These cells were then allowed to grow to confluence, at which time they

fused to become myotubes (Quenneville et al. 2004). Western blotting confirmed the presence of the fusion protein in these myotubes, and polymerase chain reaction illustrated that integration events occurred at relatively high frequency at a previously known human pseudo *attP* hotspot.

A follow-up study (Quenneville et al. 2007) examined the ability of engineered muscle precursor cells to form corrected muscle fibers after transplantation. MD1 cells were co-transfected with an integrase-expressing plasmid and either a shortened or a full-length GFP-dystrophin fusion donor plasmid, both under the control of a muscle-specific promoter (Quenneville et al. 2007). After transplanting one million corrected cells from these two groups into irradiated tibialis anterior muscles of *mdx* mice, immunofluorescence revealed that 3–6 weeks later, 250 muscle fibers per cross-section were positive for the shortened fusion protein and fewer fibers were positive for the full-length fusion. The restoration of the dystrophin complex was confirmed via co-localization of the dystrophin-associated proteins α-sarcoglycan and β-dystroglycan, which were found to be correctly localized at the membrane of the muscle fibers (Quenneville et al. 2007). Additionally, human muscle precursor cells tranfected with a GFP-expressing donor plasmid and an integrase plasmid were observed to express GFP and grow for at least 35 days *in vitro,* following 1 week of selection. These cells were then transplanted into the tibialis anterior muscles of SCID mice, which can tolerate xenografts (Quenneville et al. 2007). After 1 month, the muscles were harvested, and the human cells were observed to have fused with mouse muscle fibers and to express GFP. These two studies have laid the groundwork for a cell-based therapeutic strategy for DMD utilizing the ϕC31 integrase system.

An alternative to a cell-based strategy for DMD is direct delivery of therapeutic plasmid DNA to muscle. In a study of ϕC31-mediated integration for treating DMD, muscles of wild-type mice were injected with a donor plasmid expressing luciferase under the control of a muscle-specific promoter, with or without a functional integrase plasmid, followed by electroporation of the muscle to stimulate DNA transfer (Bertoni et al. 2006). Mice that received the functional integrase plasmid had higher levels of luciferase expression as early as 4 days after injection, which persisted for 6 months or more with an average enhancement of 5- to 10-fold. Plasmid levels were found to decline in muscles that received a nonfunctional integrase plasmid, but not in the muscles that received the functional integrase plasmid (Bertoni et al. 2006). The muscles of *mdx* mice were then injected with a donor plasmid expressing dystrophin, with or without a functional integrase plasmid, followed by electroporation of the muscle (Bertoni et al. 2006). After 2 weeks, mice that received the functional integrase plasmid were found to have roughly two times more dystrophin-positive muscle fibers than mice that received the nonfunctional integrase plasmid.

After 6 months, dystrophin-positive fibers were lost in both groups, but the mice receiving the functional integrase plasmid had significantly more positive fibers at this later time point (Bertoni et al. 2006). There was still a gradual decline in dystrophin expression over the course of the experiment, presumably due to fiber turnover. At 6 months post-injection, the muscles were examined for longitudinal distribution of dystrophin and staining with Evans blue dye, which

is taken up by degenerating fibers (Bertoni et al. 2006). In muscles that received the nonfunctional integrase plasmid, many fibers apparently received a subtherapeutic level of dystrophin, as indicated by being positive for both dystrophin and dye uptake. In muscles that received the functional integrase plasmid, significantly more fibers were dystrophin-positive, and significantly fewer dystrophin-positive fibers were positive for dye uptake (Bertoni et al. 2006). This study illustrated that ϕC31-mediated integration could enhance the effect of plasmid-based gene therapy in DMD. However, it would be challenging to extend this DNA delivery method to extensive areas of muscle. In addition, the gradual decline of dystrophin expression that was observed suggests that a stem cell-based therapy for DMD may be optimal for long-term benefit.

Use of ϕC31 integrase in muscle has also been evaluated as a potential therapeutic strategy for addressing peripheral vascular disease. It was hypothesized that ϕC31-mediated integration of transgenic vascular endothelial growth factor (VEGF)-A in muscle would allow expression to persist for at least 30 days, preventing the regression of newly formed blood vessels (Portlock et al. 2006). To test this hypothesis, the tibialis anterior muscles of C57BL/6 mice were injected with a donor plasmid expressing mouse VEGF and a ϕC31 integrase-expressing plasmid, then electroporated to boost gene transfer. VEGF protein levels were assayed after 40 days and were observed to be highest in muscles that received both the donor plasmid and the integrase plasmid (Portlock et al. 2006). The level of expression in these muscles was 1.4-fold higher than basal levels, which was previously known to be angiogenic in ischemic muscle. Evidence of multiple, independent integrations at the mpsL1 hotspot for ϕC31 integration on mouse chromosome 2 was also observed (Portlock et al. 2006). This study demonstrated that the ϕC31 system was capable of facilitating expression levels of mouse VEGF that may be therapeutic over an extended time span.

Another study examined the potential of ϕC31 integrase to address rheumatoid arthritis and other joint disorders. Rabbit and human synovial cells *in vitro* were transfected with a donor plasmid expressing GFP and kanamycin resistance and a plasmid expressing ϕC31 integrase (Keravala et al. 2006b). In the rabbit synovial cells, the presence of ϕC31 integrase resulted in a 10-fold increase in GFP expression, compared to expression in the absence of integrase. In the human cells, the presence of ϕC31 integrase resulted in a 35-fold increase. Three pseudo *attP* sites in the rabbit genome were identified from these cells, and two previously known integration sites were observed in the human cells. A donor plasmid expressing a *lacZ* or luciferase cassette, together with a plasmid expressing ϕC31 integrase, were then injected into the hind limb joints of rabbits to confirm functionality *in vivo* (Keravala et al. 2006b). Only those joints injected with both the *attB*-containing donor plasmid and the integrase plasmid showed X-gal staining in histological sections, reflecting expression of *lacZ*. There was also a fivefold increase in luciferase expression, compared to joints that did not receive the integrase plasmid. This study demonstrated the suitability of the ϕC31 integrase system as a potential non-viral gene therapy platform for joint disorders.

Another study evaluated the functionality of ϕC31 integrase in murine lung. Cells of the murine alveolar epithelial cell line MLE12 were transfected with a donor plasmid expressing a GFP-luciferase fusion protein and a plasmid expressing ϕC31 integrase (Aneja et al. 2007). Four weeks after transfection, luciferase expression levels in the cells that received both plasmids were 100-fold higher than levels in cells that did not receive the integrase plasmid. However, these levels were roughly 1000-fold lower than the expression levels 2 days after transfection. This study also examined the activity of ϕC31 integrase in the lungs of BALB/c mice through intravenous injection of the donor and integrase plasmids complexed with the polycation polyethylenimine. This compound is known to be capable of facilitating mouse lung transfection after tail-vein injection (Aneja et al. 2007). After 2 weeks, luciferase expression levels in the lungs of mice that did not receive the integrase plasmid had declined to background levels, whereas the expression level in mice receiving both plasmids was consistently detected to be 2-fold above background. This level of expression was three orders of magnitude below that observed on the first day after injection, and increasing the amount of integrase plasmid injected or administering a second injection had no effect on luciferase expression levels (Aneja et al. 2007). In the mice that received both plasmids, roughly half of them had an integration event at the mpsL1 pseudo *attP* site, a known ϕC31 hotspot, as determined by polymerase chain reaction. This study demonstrated that while ϕC31 integrase was functional in the mouse lung, the ϕC31 system and delivery method would likely need to be optimized before potential use as a therapeutic strategy for lung diseases.

Another example of a single-gene disorder where the ϕC31 system has been employed to develop a therapeutic strategy is X-linked severe combined immunodeficiency (X-SCID). X-SCID is caused by a mutation in the common cytokine receptor γ chain (γc) gene (Ishikawa et al. 2006). Thus, a study was performed to examine the potential of ϕC31 integrase to correct the disease. The functionality of ϕC31 integrase in human hematopoietic cells was evaluated by transfecting cells from the Jurkat human T-cell line (Ishikawa et al. 2006). The cells that were transfected with both a ϕC31 integrase-expressing plasmid and a donor plasmid expressing luciferase resulted in higher levels of luciferase expression over 12 days, compared to cells that did not receive the integrase plasmid or an *attB*-containing donor plasmid. Clonal analysis confirmed ϕC31-mediated genomic integration and revealed that no integration events occurred in or close to known proto-oncogenes. The integration sites did not completely correspond to previously identified pseudo *attP* hotspots in other cell types, suggesting a degree of cell-type specificity in integration sites (Ishikawa et al. 2006). To assess the suitability of the ϕC31 system for addressing X-SCID, a plasmid expressing the γc gene and the integrase plasmid were co-transfected into a human T-cell line that was defective for γc expression, ED40515(−). After 3 months of culture, flow cytometric analysis demonstrated γc expression in these cells. Moreover, stimulation with IL-2 resulted in phosphorylation of STAT5 in the corrected cells, but not in non-transfected cells (Ishikawa et al. 2006). Thus, this study showed that ϕC31

integrase was functional in some human T-cells *in vitro* and may be suitable as a therapeutic strategy for X-SCID.

Because neural stem cells are considered to be a potential therapeutic strategy for addressing neural diseases, another study evaluated the activity of ϕC31 integrase in mouse neural progenitor cells. In this study, mouse neural progenitor cells were transfected with a donor plasmid expressing luciferase and kanamycin resistance and an integrase-expressing plasmid (Keravala et al. 2008). After 8 weeks of culture in selective media, the cells that received the integrase plasmid had 39-fold higher levels of luciferase expression than the cells that did not receive integrase. Selective culture with G418 resulted in 201 resistant neurospheres in cells that received the integrase plasmid, compared to five resistant neurospheres in cells that did not (Keravala et al. 2008). Integration was confirmed via polymerase chain reaction to have occurred at the pseudo *attP* hotspot mpsL1, and cells containing integration events mediated by ϕC31 integrase were observed to retain both their proliferative and differentiation capacities. This study illustrated that ϕC31 integrase could be used to modify neural progenitor cells, with no gross impact on their ability to proliferate and differentiate.

8.5 Use of ϕC31 Integrase in Pluripotent Stem Cells

Because of their ability to differentiate into any cell type, pluripotent stem cells, including embryonic stem cells and induced pluripotent stem cells, are of great interest as disease models and in the development of therapeutic strategies. The ϕC31 integrase system has been used to generate engineered human embryonic stem cell (hESC) lines, under the rationale that being able to manipulate the human genome to create disease models will offer greater insights into human disease. In the initial study, hESC from the BG01v cell line were transfected with an integrase-expressing plasmid and a hygromycin-resistant donor plasmid (Thyagarajan et al. 2008). After 2 weeks of selection, cells that received both plasmids showed a 1.4-fold increase in colony formation compared to the background of random integration. Pseudo *attP* sites in these colonies were analyzed by plasmid rescue followed by polymerase chain reaction (Thyagarajan et al. 2008). Twenty-three integration sites were identified, most of which had not been observed before and many of which were located in introns of genes. GFP-expressing hESC lines were then generated via transfection with an integrase-expressing plasmid and a donor plasmid expressing GFP under the control of either the Oct4 or the EF1α promoter (Thyagarajan et al. 2008). When these lines were analyzed, they showed no significant reduction in GFP expression after 4–5 weeks in culture and were capable of embryoid body formation. Following differentiation, GFP expression decreased in the Oct4-GFP lines, but not in the EF1α-GFP lines, as expected.

After demonstrating that the ϕC31 system was functional in hESC, the next step was to generate a platform for ease of targeting transgenes to an optimal locus. Having such a platform was important, to avoid silencing of transgenes through

position effects. This problem is particularly relevant for stem cells, because new position effects may arise through differentiation (Liu et al. 2009b). To determine if the ϕC31 system was suitable to construct such a platform, hESC from the H9 and BG01V lines were transfected via electroporation with an integrase-expressing plasmid and a donor plasmid containing the *attP* site of another serine recombinase, the integrase from *Streptomyces* bacteriophage R4 (Olivares et al. 2001). The R4 *attP* site formed the basis of the transgene targeting system. Clones were isolated, and their potential for silencing was evaluated after differentiation through use of a microarray database. From this analysis, only a clone containing an integration event at a pseudo *attP* hotspot at the chromosome 13q32 locus was capable of being retargeted via transfection with an R4 *attB*-containing donor plasmid expressing GFP under the control of the EF1α promoter and carrying zeocin resistance and R4 integrase (Liu et al. 2009b).

On average, 17 colonies per one million transfected cells were obtained with zeocin selection. Out of 22 clones analyzed, all except one had a single integration at the R4 *attP* retargeting locus and maintained hESC morphology and pluripotentcy marker expression. Greater than 95% of these cells expressed GFP after 2.5 months in culture. Additionally, neural lineage cells expressing GFP were obtained from these clones via directed differentiation (Liu et al. 2009b). The chromosome 13 locus was also assessed for its capability to facilitate proper transcriptional regulation through retargeting with an R4 donor plasmid expressing GFP under the control of the Oct4 promoter. After differentiation, the retargeted clone showed a decrease in GFP expression, as expected (Liu et al. 2009b). Thus, this study demonstrated that it was possible to use the ϕC31 system to generate a platform for targeted transgenesis in hESC.

Following the generation of a retargeting platform, another study examined the suitability of the platform for drug discovery. In this study, the platform was first extended to the human embryonic kidney cell line HEK293 and the Chinese hamster ovary cell line CHO-S by transfecting these cells with a plasmid expressing ϕC31-integrase and an R4 *attP*-containing donor plasmid (Lieu et al. 2009). The hESC BG01V platform cell line derived in the previous study was also used. The retargeting plasmid consisted of GFP-c-Jun, CCKAR, or TRPM8 under the control of the EF1α promoter and was tranfected into the platform cell lines, along with a plasmid expressing R4 integrase (Lieu et al. 2009). These three genes were chosen because they represent pathways that have been previously explored for drug discovery using other methods; results obtained using the platform lines could thus be compared to existing methods (Lieu et al. 2009). After generating the retargeted lines, the CCKAR and TRPM8 lines were tested for ligand-induced activation, and the GFP-c-Jun lines were tested for ligand-induced phosphorylation. Responses across the three cell types were consistent with the expected pharmacological results, demonstrating that ϕC31 integrase was capable of facilitating the development of a drug discovery platform that involved site-specific targeting of transgenes and was capable of generating cell lines more quickly than previous methods (Lieu et al. 2009).

Since the discovery that induced pluripotent stem cells (iPSC) could be generated from differentiated cells, there has been much discussion of using iPSC as

models for human diseases, as well as for therapeutic purposes. A major caveat of iPSC lies with the most commonly used method of generating them: integrating multiple copies of retroviral vectors. This approach suffers from some problems, including poor control over the number of integrations and lack of site specificity, with the attendant insertional mutagenesis risks. In an attempt to overcome these issues, two recent studies have generated iPSC utilizing the ϕC31 integrase system. In the first study, the donor plasmid carried an OCT3/4-SOX2-KLF4-cMYC cassette. These reprogramming factors were expressed from a tetracycline response element, utilizing the Tet reverse transactivator under control of the CAG promoter for inducible expression of the reprogramming factors (Ye et al. 2010). Previous work had shown that the Tet-inducible expression system is tightly controlled when combined with the ϕC31 system (Inoue et al. 2009). This reprogramming vector was then electroporated into mouse embryonic fibroblasts (MEFs) along with an integrase-expressing plasmid, resulting in the appearance of iPSC colonies after 20 days. 70% of the colonies were observed to contain single integrations of the donor plasmid via Southern blot analysis (Ye et al. 2010). Integration site analysis revealed that there did not appear to be a preferred pseudo *attP* site in MEFs.

The authors then used the donor plasmid to reprogram human cells isolated from amniotic fluid obtained for prenatal diagnosis (Ye et al. 2010). Interestingly, a ratio of 1:20 of donor plasmid to integrase plasmid was used. By day 9, iPS-like colonies appeared, which were isolated for analysis between day 35 and 40 (Ye et al. 2010). Cells from these colonies expressed pluripotency markers, possessed self-renewal properties, and generally met the accepted criteria for pluripotency. One chromosomal abnormality was observed, but the authors noted that it was not known to be the result of an integrase-mediated integration (Ye et al. 2010). Analysis of the integration events revealed that the majority were single copy integrants, without a preferred pseudo *attP* site. This study was important in demonstrating the viability of generating iPSCs with the ϕC31 system.

The study by Ye and colleagues was useful as a proof-of-principle demonstration of the ability of ϕC31 integrase to generate iPSCs, but some issues were not addressed. These issues included removal of the reprogramming cassette, targeting of a therapeutic gene to the integration locus, and the requirement for tetracycline or a derivative to achieve reprogramming. In another recent study, these issues were addressed. The donor plasmid employed contained the same four reprogramming factors used by Ye and colleagues, with the following differences: the ϕC31 *attB* site was flanked by *loxP* sites, the plasmid additionally contained an R4 integrase *attP* site, and the tetracycline inducible system was absent (Karow et al. 2011). MEFs and mouse adipose-derived mesenchymal stem cells (AD-MSC) were reprogrammed by nucleofection with the donor plasmid and an integrase-expressing plasmid. iPSC-like colonies were obtained, expanded in culture, and integration site analysis was carried out via a combination of Southern blot and linker-mediated polymerase chain reaction (Karow et al. 2011). Integration site analysis revealed that 37% of the MEF-derived iPSC clones and 31% of the AD-MSC-derived iPSC clones contained a single integration event. More than three integration events per single clone were not observed (Karow et al. 2011). One clone from each population

of single integrants was chosen for additional analysis. The MEF-derived clone contained an integration in an intronic region of the gene *Ptpn1*, whereas the AD-MSC-derived clone possessed an integration in an intergenic region of chromosome 1 (Karow et al. 2011). Further analysis of 14 single-integrant clones revealed that six clones possessed integrations in intergenic regions, and that two of these clones represented safe integration sites by stringent criteria.

Cre recombinase was then introduced into the clones by transfecton of a Cre-expressing plasmid. Cre was effective in excising the reprogramming cassette, which was confirmed by Southern blotting. Before and after the excision of the reprogramming cassette, both clones satisfied conventional criteria for pluripotency (Karow et al. 2011). Additionally, both clones were capable of contributing to the generation of chimeric mice, which was not addressed by Ye and colleagues. The inclusion of an R4 *attB* site in the donor plasmid, while not utilized in this study, was intended to facilitate the subsequent site-specific integration of a therapeutic gene into the iPSCs (Karow et al. 2011). This study introduced the concept of combining ϕC31 with other recombinases to excise the reprogramming cassette and to introduce therapeutic genes in site-specific manner. The work also demonstrated that iPSC can be generated from mouse cells using the ϕC31 system without the tetracycline-inducible system. The overarching goal of these studies is ultimately to develop a generalized gene/cell therapy strategy, such as the one depicted in Fig. 8.2b.

8.6 Safety of the ϕC31 Integrase System

When introducing a recombinase into cells, the possibility of unwanted, adventitious events must be evaluated. Several studies documented that a fraction (~10%) of ϕC31-mediated integration events in cultured cells were aberrant, in terms of containing chromosome rearrangments or large deletions (Chalberg et al. 2006; Ehrhart et al. 2006; Liu et al. 2006, 2009a, b). Most of the events appeared to be non-random and targeted to ϕC31 pseudo-sites (Chalberg et al. 2006). Such events have never been documented in *in vivo* studies with ϕC31 integrase, nor have any tumors or other adverse events ever been documented in any type of organism treated with ϕC31 integrase (Calos 2006).

A recent study assessed the safety of using ϕC31 integrase in human umbilical cord-lining cells. Cord-lining epithelial cells (CLECs) were transfected with a GFP-expressing, kanamycin-resistant donor plasmid, either with or without an integrase-expressing plasmid (Sivalingam et al. 2010). From 5,000 GFP-positive cells, a significantly higher number of GFP-expressing, kanamycin-resistant colonies were obtained from cells that received both plasmids, compared to the GFP donor plasmid only. From 2,000 GFP-positive cells, a significantly higher number of stable integrants were obtained in the presence of integrase than in its absence. 44 independent integration events were analyzed, most of which were intronic (Sivalingam et al. 2010). More than 70% of the integrations were greater than 50 kb away from a

transcriptional start site. The effect of integration on endogenous gene expression was measured by comparing the transcriptomes of unmodified CLECs with a polyclonal population of stable integrants (Sivalingam et al. 2010). More than 96.5% of transcripts showed no difference in expression levels. Of those that did show expression differences, three were tumor suppressor genes that were upregulated (promoting cell death over proliferation) and 12 others were involved in cell cycle regulation or cell adhesion. Additionally, minimal effects on copy number variation were observed in the stably integrated CLECs, and the variations observed were more consistent with background copy number variation than the copy number variation observed in cancer cells.

To determine whether ϕC31 integrase had induced chromosomal rearrangements, spectral karyotyping was performed on unmodified and stably integrated CLECs (Sivalingam et al. 2010). There were no abnormalities detected in the unmodified CLECs, but four out of 90 metaphase spreads from modified CLECs displayed translocations. Further analysis of the modified cells showed no abnormalities in more than 210 karyotypes, suggesting that the observed nonrecurrent abnormalities were consistent with background levels of abnormalities in human somatic cells (Sivalingam et al. 2010). Additionally, transplantation of modified CLECs into NOD-SCID mice did not result in the development of tumors. Thus, based on the lack of tumorigenicity, the background level of nonrecurrent chromosomal aberrations, and low impact on transcription, this study concluded that the ϕC31 system may be safe to use in human cells and recommended increasing safety even more by using clonal populations for implantation (Sivalingam et al. 2010). In particular, we note that when using ϕC31 integrase in pluripotent stem cells, it is desirable to work with clonal populations of cells in which a correct karyotype has been carefully documented.

A further study focused on whether hepatic exposure to ϕC31 integrase increased the rate of liver cancer *in vivo*. A transgenic mouse model of hepatocellular carcinoma was utilized in which human C-MYC expression was induced by withdrawal of tetracycline, resulting in liver cancer. These mice were withdrawn from tetracycline at 7–8 weeks of age, injected 1 week later with various plasmids, as detailed below, and sacrificed when moribund (Woodard et al. 2010b). The mice were injected by hydrodynamic tail vein injection with either saline alone, or a luciferase-expressing donor plasmid and an empty vector, or the luciferase donor plasmid with either an inactive integrase plasmid, a functional integrase plasmid, the luciferase donor plasmid alone, or the functional integrase plasmid alone. In the mice that received only a saline injection, tumors were found to appear significantly earlier than in mice that did not receive any hydrodynamic injection, suggesting a tumorigenic effect of hydrodynamic injection itself. The inclusion of DNA in the injection, as well as expression of luciferase, resulted in an average tumor latency that was not significantly different from that of saline alone (Woodard et al. 2010b). Surprisingly, the presence of either inactive or functional integrase resulted in tumors forming at the same rate as they did in non-injected mice. These results suggested that ϕC31 integrase was non-tumorigenic and may also mitigate the tumorigenic effects of hydrodynamic injection in this model. Additionally, in the mice that were treated

with integrase, the tumors that did form did not display luciferase activity or possess ϕC31 integrase-mediated integrations (Woodard et al. 2010b). Thus, this findings of this study were consistent with the lack of tumors observed to date in all organisms treated ϕC31 integrase and suggested that this prokaryotic recombinase, like Cre recombinase, may be largely benign in eukaryotic cells.

8.7 The Future of ϕC31 Integrase: Combinatorial Usage with Other Recombinases

The value of ϕC31 integrase can often be amplied by using it in combination with other site-specific recombinases. ϕC31 was first used in combination with another serine integrase, R4, in experiments in which ϕC31 integrase was used to place an R4 *att* site into the genome, to achieve more specific and efficient targeting by R4 integrase (Olivares et al. 2001). The combination of ϕC31 and R4 integrases has also been employed in studies in hESC, as described above (Liu et al. 2009b). The availability of multiple serine integrases that function in mammalian cells affords many such opportunities (Keravala et al. 2006a).

Cre recombinase can also be used with ϕC31 integrase to good effect. For example, a multi-step approach was employed in which an alternating use of ϕC31 integrase and Cre facilitated the integration of complex, 150 kb-long arrays of vertebrate centromeric sequences in chicken DT40 cells (Dafhnis-Calas et al. 2005). This approach was called iterative site-specific integration and was useful, because it offered a strategy to integrate larger, more complex transgenic arrays into genomes than was possible with single-recombinase strategies. The combination of ϕC31 integrase and Cre was also useful for excising the reprogramming genes after generating iPSC with ϕC31 integrase (Karow et al. 2011).

In another strategy using multiple recombinases, a drug resistance cassette that carried a FRT site and a ϕC31 *attP* site was introduced into the genome of mouse embryonic stem cells by homologous recombination (Monetti et al. 2011). Colonies were obtained and analysed via Southern blotting in order to obtain single integrants. The donor plasmid contained the ϕC31 *attB* site, a gene of interest, another FRT site, and the thymidine kinase gene for selection. Once stable integrants were obtained, the drug resistance genes were excised with FLP recombinase, leaving the *attR* site, the gene of interest, and a FRT site (Monetti et al. 2011). This method was useful for removal of selection genes.

In conclusion, the ϕC31 system has been shown to result in unidirectional, sequence-specific integration into the genomes of a wide variety of plants and animals. This system has the potential to be useful in generating transgenic organisms and facilitating the development of cell-based therapies. In order to take full advantage of the genome engineering power of ϕC31 integrase, it may be combined with other site-specific recombinase systems, including other serine integrases or tyrosine recombinases such as Cre and FLP. The use of such combinatorial systems allows for even more opportunities for site-specific genome engineering.

Acknowledgements JMG was supported by a graduate fellowship from the National Science Foundation. Work in the Calos lab is supported by grants to MPC from the National Institutes of Health, the California Institute for Regenerative Medicine, the Muscular Dystrophy Association, and the Jain Foundation. MPC is an inventor on Stanford-owned patents covering phage integrases.

References

Allen BG, Weeks DL (2005) Transgenic *Xenopus laevis* embryos can be generated using ϕC31 integrase. Nat Methods 2(12):975–979

Allen BG, Weeks DL (2006) Using ϕC31 integrase to make transgenic *Xenopus laevis* Embryos. Nat Protoc 1(3):1248–1257

Allen BG, Weeks DL (2009) Bacteriophage ϕC31 integrase mediated transgenesis in *Xenopus laevis* for protein expression at endogenous levels. In: Carrol DJ (ed) Microinjection: methods and applications, vol 518. Humana, New York

Andreas S, Schwenk F, Küter-Luks B, Faust N, Kühn R (2002) Enhanced efficiency through nuclear localization signal fusion on phage ϕC31 -integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. Nucleic Acids Res 30(11):2299–2306

Aneja MK, Imker R, Rudolph C (2007) Phage phiC31 integrase-mediated genomic integration and long-term gene expression in the lung after nonviral gene delivery. J Gene Med 9:967–975

Aneja MK, Geiger J, Imker R, Üzgün S, Kormann M, Hasenpusch G, Maucksch C, Rudolph C (2009) Optimization of *Streptomyces* bacteriophage ϕC31 integrase system to prevent post integrative gene silencing in pulmonary type II cells. Exp Mol Med 41(12):919–934

Bertoni C, Jarrahian S, Wheeler TM, Li Y, Olivares EC, Calos MP, Rando TA (2006) Enhancement of plasmid-mediated gene therapy for muscular dystrophy by directed plasmid integration. Proc Natl Acad Sci 103(2):419–424

Brown WRA, Lee NCO, Xu Z, Smith MCM (2011) Serine recombinases as tools for genome engineering. Methods 53(4):372–379

Calos MP (2006) The ϕC31 integrase system for gene therapy. Curr Gene Ther 6:633–645

Chalberg TW, Genise HL, Vollrath D, Calos MP (2005) ϕC31 integrase confers genomic integration and long-term transgene expression in rat retina. Invest Opthalmol Vis Sci 46(6):2140–2146

Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT, Hoelters J, Calos MP (2006) Integration specificity of phage ϕC31 integrase in the human genome. J Mol Biol 357:28–48

Chavez CL, Calos MP (2011) Therapeutic applications of the phiC31 integrase system. Curr Gene Ther 11(5):375–381

Chavez CL, Keravala A, Woodard LE, Hillman RT, Stowe TR, Chu JN, Calos MP (2010) Kinetics and longevity of ϕC31 integrase in mouse liver and cultured cells. Hum Gene Ther 21:1287–1297

Dafhnis-Calas F, Xu Z, Haines S, Malla SK, Smith MCM, Brown WRA (2005) Iterative *in vivo* assembly of large and complex transgenes by combining the activities of ϕC31 integrase and Cre recombinase. Nucleic Acids Res 33(22):e189

Ehrhart A, Engler JA, Xu H, Cherry AM, Kay MA (2006) Molecular analysis of chromosomal rearragements in mammalian cells after phiC31-mediated integration. Hum Gene Ther 17:1077–1094

Fish MP, Groth AC, Calos MP, Nusse R (2007) Creating transgenic *Drosophila* by microinjecting the site-specific ϕC31 integrase mRNA and a transgene-containing donor plasmid. Nat Protoc 2(10):2325–2331

Gao G, McMahon C, Chen J, Rong YS (2008) A powerful method combining homologous recombination and site-specific recombination for targeted mutagenesis in *Drosophila*. Proc Natl Acad Sci 105(37):13999–14004

Gils M, Marillonnet S, Werner S, Grützner R, Giritch A, Engler C, Schachschneider R, Klimyuk V, Gleba Y (2008) A novel hybrid seed system for plants. Plant Biotechnol J 6:226–235

Grindley NDF, Whiteson KL, Rice PA (2006) Mechanisms of site-specific recombination. Annu Rev Biochem 75:567–605

Groth AC, Olivares EC, Thyagarajan B, Calos MP (2000) A phage integrase directs efficient site-specific integration in human cells. Proc Natl Acad Sci 97(11):5995–6000

Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage ΦC31. Genetics 166:1775–1782

Held PK, Olivares EC, Aguilar CP, Finegold M, Calos MP, Grompe M (2005) *In vivo* correction of murine hereditary tyrosinemia type I by ϕC31 integrase-mediated gene delivery. Mol Ther 11(3):399–408

Hollis RP, Stoll SM, Sclimenti CR, Lin J, Chen-Tasi Y, Calos MP (2003) Phage integrases for the construction and manipulation of transgenic mammals. Reprod Biol Endocrinol 1:79

Inoue K, Sone T, Oneyama C, Nishiumi F, Kishine H, Sasaki Y, Andoh T, Okada M, Chesnut JD, Imamoto F (2009) A versatile nonviral vector system for tetracycline-dependent one-step conditional induction of transgene expression. Gene Ther 16:1383–1394

Ishikawa Y, Tanaka N, Uchiyama T, Kumaki S, Tsuchiya S, Kugoh H, Oshimura M, Calos MP, Sugamura K (2006) Phage ϕC31 integrase-mediated genomic integration of the common cytokine receptor gamma chain in human T-cell lines. J Gene Med 8:646–653

Karow M, Chavez CL, Farruggio AP, Geisinger JM, Keravala A, Jung WE, Lan F, Wu JC, Chen-Tsai Y, Calos MP (2011) Site-specific recombinase strategy to create iPS cells efficiently with plasmid DNA. Stem Cells 29(11):1692–1704

Kempe K, Rubtsova M, Berger C, Kumlehn J, Schollmeier C, Gils M (2010) Transgene excision from wheat chromosomes by phage ϕC31 integrase. Plant Mol Biol 72:673–687

Keravala A, Groth AC, Jarrahian S, Thyagarajan B, Hoyt JJ, Kirby P, Calos MP (2006a) 'A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. Mol Genet Genomics 276:135–146

Keravala A, Portlock JL, Nash JA, Vitrant DG, Robbins PD, Calos MP (2006b) PhiC31 integrase mediates integration in cultured synovial cells and enhances gene expression in rabbit joints. J Gene Med 8:1008–1017

Keravala A, Ormerod BK, Palmer TD, Calos MP (2008) Long-term transgene expression in mouse neural progenitor cells modified with ϕC31 integrase. J Neurosci Methods 173:299–305

Keravala A, Chavez CL, Hu G, Woodard LE, Monahan PE, Calos MP (2011) Long-term phenotypic correction in factor IX knockout mice by using phiC31 integrase-mediated gene therapy. Gene Ther 18:842–848

Kontarakis Z, Pavlopoulos A, Kiupakis A, Konstantinides N, Douris V, Averof M (2011) A versatile strategy for gene trapping and trap conversion in emerging model organisms. Development 138:2625–2630

Kuhstoss S, Rao RN (1991) Analysis of the integration function of the Streptomycete bacteriophage ϕC31. J Mol Biol 222:897–908

Labbé GMC, Nimmo DD, Alphey L (2010) *piggybac*- and phiC31-mediated genetic transformation of the Asian tiger mosquito, *Aedes albopictus* (Skuse). PLoS Negl Trop Dis 4(8):e788

Leighton PA, Van de Lavior M-C, Diamond JH, Xia C, Etches RJ (2008) Genetic modification of primordial germ cells by gene trapping, gene targeting, and ϕC31 integrase. Mol Reprod Dev 75:1163–1175

Lieu PT, Machleidt T, Thyagarajan B, Fontes A, Frey E, Fuerstenau-Sharp M, Thompson DV, Swamilingiah GM, Derebail SS, Piper D, Chesnut JD (2009) Generation of site-specific retargeting platform cell lines for drug discovery using phiC31 and R4 integrases. J Biomol Screen 14:1207–1215

Lister JA (2010) Transgene excision in zebrafish using the phiC31 integrase. Genesis 48:137–143

Liu J, Jeppesen I, Nielsen K, Jensen TG (2006) PhiC31 integrase induces chromosomal aberrations in primary human fibroblasts. Gene Ther 13:1188–1190

Liu J, Skjørringe T, Gjetting T, Jensen TG (2009a) PhiC31 integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. BMC Biotechnol 9:31–38

Liu Y, Thyagarajan B, Lakshmipathy U, Xue H, Lieu P, Fontes A, MacArthur CC, Scheyhing K, Rao MS, Chesnut JD (2009b) Generation of platform human embryonic stem cell lines that allow efficient targeting at a predetermined genomic location. Stem Cells Dev 18(10):1459–1471

Lu J, Maddison LA, Chen W (2011) PhiC31 integrase induces efficient site-specific excision in zebrafish. Transgenic Res 20(1):183–189

Lutz KA, Corneille S, Azhagiri AK, Svab Z, Maliga P (2004) A novel approach to plastid transformation utilizes the phiC31 phage integrase. Plant J 37:906–913

Ma Q-W, Sheng H-Q, Yan J-B, Cheng S, Huang Y, Chen-Tsai Y, Ren Z-R, Huang S-Z, Zeng Y-T (2006) Identification of pseudo *attP* sites for phage ϕC31 integrase in bovine genome. Biochem Biophys Res Commun 345:984–988

Maucksch C, Aneja MK, Hennen E, Bohla A, Hoffmann F, Elfinger M, Rosenecker J, Rudolph C (2008) Cell type differences in activity of the *Streptomyces* bacteriophage ϕC31 integrase. Nucleic Acids Res 36(17):5462–5471

Meredith JM, Basu S, Nimmo DD, Larget-Thiery I, Warr EL, Underhill A, McArthur CC, Carter V, Hurd H, Bourgouin C, Eggleston P (2011) Site-specific integration and expression of an anti-malarial gene in transgenic *Anopheles gambiae* significantly reduces *Plasmodium* infections. PLoS One 6(1):e14587

Monetti C, Nishino K, Biechele S, Zhang P, Baba T, Woltjen K, Nagy A (2011) PhiC31 integrase facilitates genetic approaches combining multiple recombinases. Methods 53:380–385

Nakayama G, Kawaguchi Y, Koga K, Kusakabe T (2006) Site-specific gene integration in cultured silkworm cells mediated by ϕC31 integrase. Mol Genet Genomics 275:1–8

Ni J-Q, Markstein M, Binari R, Pfeiffer B, Liu L-P, Villalta C, Booker M, Perkins E, Perrimon N (2008) Vector and parameters for targeted transgenic RNA interference in *Drosophila melanogaster*. Nat Methods 5(1):49–51

Ni J-Q, Liu L-P, Binari R, Hardy R, Shim H-S, Cavallaro A, Booker M, Pfeiffer BD, Markstein M, Wang H, Villalta C, Laverty TR, Perkins LA, Perrimon N (2009) A *Drosophila* resource of transgenic RNAi lines for neurogenetics. Genetics 182:1089–1100

Nimmo DD, Alphey L, Meredith JM, Eggleston P (2006) High efficiency site-specific genetic engineering of the mosquito genome. Insect Mol Biol 15(2):129–136

Nishiumi F, Sone T, Kishine H, Thyagarajan B, Kogure T, Miyawaki A, Chesnut JD, Imamoto F (2009) Simultaneous single cell stable expression of 2–4 cDNAs in HeLaS3 using ϕC31 integrase system. Cell Struct Funct 34:47–59

Olivares EC, Hollis RP, Calos MP (2001) Phage R4 integrase mediates site-specific integration in human cells. Gene 278:167–176

Olivares EC, Hollis RP, Chalberg TW, Meuse L, Kay MA, Calos MP (2002) Site specific genomic integration produces therapeutic Factor IX levels in mice. Nat Biotechnol 20:1124–1128

Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q, Fang M, Calos MP, Khavari PA (2002) Stable nonviral genetic correction of inherited human skin disease. Nat Med 8(10):1166–1170

Ortiz-Urda S, Thyagarajan B, Keene D, Lin Q, Calos MP, Khavari PA (2003) ϕC31 integrase-mediated nonviral genetic correction of junctional epidermolysis bullosa. Hum Gene Ther 14:923–928

Ou H-L, Huang Y, Qu L-J, Xu M, Yan J-B, Ren Z-R, Huang S-Z, Zeng Y-T (2009) A ϕC31 integrase-mediated integration hotspot in favor of transgene expression exists in the bovine genome. FEBS J 276:155–163

Pfeiffer BD, Jenett A, Hammonds AS, Ngo T-TB, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Laverty TR, Mungall C, Svirskas R, Kadonga JT, Doe CQ, Eisen MB, Celniker SE, Rubin GM (2008) Tools for neuroanatomy and neurogenetics in *Drosophila*. Proc Natl Acad Sci 105(28):9715–9720

Portlock JL, Keravala A, Bertoni C, Lee S, Rando TA, Calos MP (2006) Long-term increase in $mVEGF_{164}$ in mouse hindlimb muscle mediated by phage ϕC31 integrase after nonviral DNA delivery. Hum Gene Ther 17:871–876

Quenneville SP, Chapdelaine P, Rousseau J, Beaulieu J, Caron NJ, Skuk D, Mills P, Olivares EC, Calos MP, Tremblay JP (2004) Nucleofection of muscle-derived stem cells and myoblasts with ϕC31 integrase: stable expression of a full-length-dystrophin fusion gene by human myoblasts. Mol Ther 10(4):679–687

Quenneville SP, Chapdelaine RJ, Tremblay JP (2007) Dystrophin expression in host muscle following transplantation of muscle precursor cells modified with the phiC31 integrase. Gene Ther 14:514–522

Rausch H, Lehmann M (1991) Structural analysis of the actinophage ϕC31 attachment site. Nucleic Acids Res 19(19):5187–5189

Raymond CS, Soriano P (2007) High-efficiency FLP and ϕC31 site-specific recombination in mammalian cells. PLoS One 2(1):e162

Rubtsova M, Kempe K, Gils A, Ismagul A, Weyen J, Gils M (2008) Expression of active *Streptomyces* phage ϕC31 integrase in transgenic wheat plants. Plant Cell Rep 27:1821–1831

Schetelig MF, Scolari F, Handler AM, Kittelmann S, Gasperi G, Wimmer EA (2009) Site-specific recombination for the modification of transgenic strains of the Mediterranean fruit fly *Ceratitis capitata*. Proc Natl Acad Sci 106(43):18171–18176

Sharma N, Moldt B, Dalsgaard T, Jensen TG, Mikkelsen JG (2008) Regulated gene insertion by steroid-induced ϕC31 integrase. Nucleic Acids Res 36(11):e67

Sivalingam J, Krishnan S, Ng WH, Lee SS, Phan TT, Kon OL (2010) Biosafety assessment of site-directed transgene integration in human umbilical cord-lining cells. Mol Ther 18(7):1346–1356

Tasic B, Hippenmeyer S, Wang C, Gamboa M, Zong H, Chen-Tsai Y, Luo L (2011) Site-specific integrase-mediated transgenesis in mice via pronuclear injection. Proc Natl Acad Sci 108(19):7902–7907

Thomson JG, Chan R, Thilmony R, Yau Y-Y, Ow DW (2010) PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the *Arabidopsis* genome. BMC Biotechnol 10:17

Thorpe HM, Smith MCM (1998) *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc Natl Acad Sci 95:5505–5510

Thyagarajan B, Calos MP (2005) Site-specific integration for high-level protein production in mammalian cells. In: Smales CM, James DC (eds) Methods in molecular biology, vol 308, Therapeutic proteins: methods and protocols. Humana, Totowa

Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP (2001) Site-specific genomic integration in mammalian cells mediated by phage ϕC31 integrase. Mol Cell Biol 21(12):3926–3934

Thyagarajan B, Liu Y, Shin S, Lakshmipathy U, Scheyhing K, Xue H, Ellerström C, Strehl HJ, Rao MS, Chesnut JD (2008) Creation of engineered human embryonic stem cell lines using phiC31 integrase. Stem Cells 26:119–126

Watanabe S, Nakamura S, Sakurai T, Akasaka K, Sato M (2010) Improvement of a phiC31 integrase-based gene delivery system that confers high and continuous transgene expression. New Biotechnol 28(4):312–319

Woodard LE, Hillman RT, Keravala A, Lee S, Calos MP (2010a) Effect of nuclear localization and hydrodynamic delivery-induced cell division on ϕC31 integrase activity. Gene Ther 17:217–226

Woodard LE, Keravala A, Jung WE, Wapinsky OL, Yang Q, Felsher DW, Calos MP (2010b) Impact of hydrodynamic injection and phiC31 integrase on tumor latency in mouse model of MYC-induced hepatocellular carcinoma. PLoS One 5(6):e11367

Ye L, Chang JC, Lin C, Qi Z, Yu J, Kan YW (2010) Generation of induced pluripotent stem cells using site-specific integration with phage integrase. Proc Natl Acad Sci 107(45):19467–19472

Chapter 9
Modified Transposases for Site-Directed Insertion of Transgenes

Sean Colloms and Sylvaine Renault

Abstract Cut and paste DNA transposons are widely used to stably integrate DNA into a wide variety of organisms. They can integrate DNA with high efficiency, provide long lasting expression of inserted transgenes, and avoid some of the safety concerns of viral gene delivery systems. One of the chief disadvantages of transposons for gene therapy and other gene delivery applications is that the site of insertion cannot be chosen. This can lead to poor expression of the integrated gene if it is inserted into a region of heterochromatin, or to undesirable insertional mutagenesis of the host cell if it integrates in an important gene. Three main strategies have been used to direct transposon insertions to specific locations: (1) modifying the transposase by fusing it to a DNA binding domain, (2) tethering the transposase protein to the desired target site, or (3) tethering the transposon DNA to the target site using appropriate fusion proteins. Here we review progress with these strategies in bacteria, fish, insect and mammalian systems.

Keywords Transposition • Targeting • Transgenesis • DNA binding domain

S. Colloms
Institute of Molecular Cell and Systems Biology, University of Glasgow,
Bower Building, University Ave, Glasgow G12 8QQ, Scotland, UK
e-mail: Sean.Colloms@glasgow.ac.uk

S. Renault (✉)
EA 6306 Innovation moléculaire et thérapeutique, Université François Rabelais,
UFR des Sciences Pharmaceutiques, 31 Avenue Monge, 37200 Tours, France
e-mail: sylvaine.renault@univ-tours.fr

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_9,

9.1 Introduction

Transposons are widely used for insertional mutagenesis and as vectors for genomic integration of transgenes in a wide variety of organisms, ranging from bacteria to mammals. Transposons are small segments of DNA that have the capabilities to move autonomously from one locus to another in the genome. DNA transposons (or type II transposons, hereafter simply referred to as transposons) are characterized by a DNA transposition intermediate, in contrast to retrotransposons which move *via* an RNA intermediate. Transposons have a simple structure, in most cases consisting of two inverted repeat (IR) sequences, surrounding a single ORF encoding the transposase protein (Fig. 9.1a). The transposase binds specifically to the transposon IRs and cleaves both DNA strands at the outside extremities of the IR sequences, excising the transposon from its genomic location. The transposon-transposase complex then finds a genomic target sequence, cleaves this target site and simultaneously inserts the transposon into it (Fig. 9.1a). The two ends of the transposon are inserted at staggered positions on the two strands of the target site, leaving single stranded regions that are repaired to form target duplications of characteristic length for different transposons.

The ability of transposases to carry out "cut and paste" transposition has been useful for integration of transgenes into the genomes of a wide range of organisms (Fig. 9.1b) (Izsvak et al. 2010). To perform transgenesis, two plasmids are transfected into the target cells. One plasmid carries a modified transposon, consisting simply of the transgene to be integrated flanked by two copies of the IR sequence. The other plasmid expresses the transposase protein in *trans*. Almost any DNA sequence (subject to certain length constraints) can be integrated, so long as it is flanked by two copies of the IR sequence. The transposase is expressed, cuts at both IRs, and then integrates the modified transposon along with its transgene "cargo" into a new target site (Fig. 9.1b).

A small number of transposons have been extensively used to perform transgenesis or mutagenesis. They are used according to their host range. Tn5 from *Escherichia coli,* and *Himar* from the fly *Haematobia irritans* are used for genome-wide mutagenesis in various species of bacteria (Reznikoff 2006; Wong et al. 2011). P elements have been used extensively for mutagenesis and transgenesis in the fruit fly *Drosophila melanogaster* (Ryder and Russell 2003). The mariner-like element Mos1, first identified in *Drosophila mauritiana* is used for transgenesis in insects and the nematode *Caenorhabditis elegans* (Lidholm et al. 1993; Bessereau et al. 2001). Sleeping Beauty (SB), a reconstructed Tc1-like element (belonging to the same Tc1/mariner family as Mos1) from fish is able to transpose efficiently in mammals (Ivics et al. 1997). Tol2, a natural element isolated from medaka fish and a member of the hAT super family of transposons can be used in fish and in mammalian cells (Suster et al. 2009). Finally, *piggyBac* from the cabbage looper moth *Trichoplusia ni* transposes efficiently in insects and mammalian cells, including human cells (Wu et al. 2006).

Some transposons insert at specific sequences, whereas others show only mild target preferences. For instance, sequence analysis of over 500 Tn5 insertions showed that

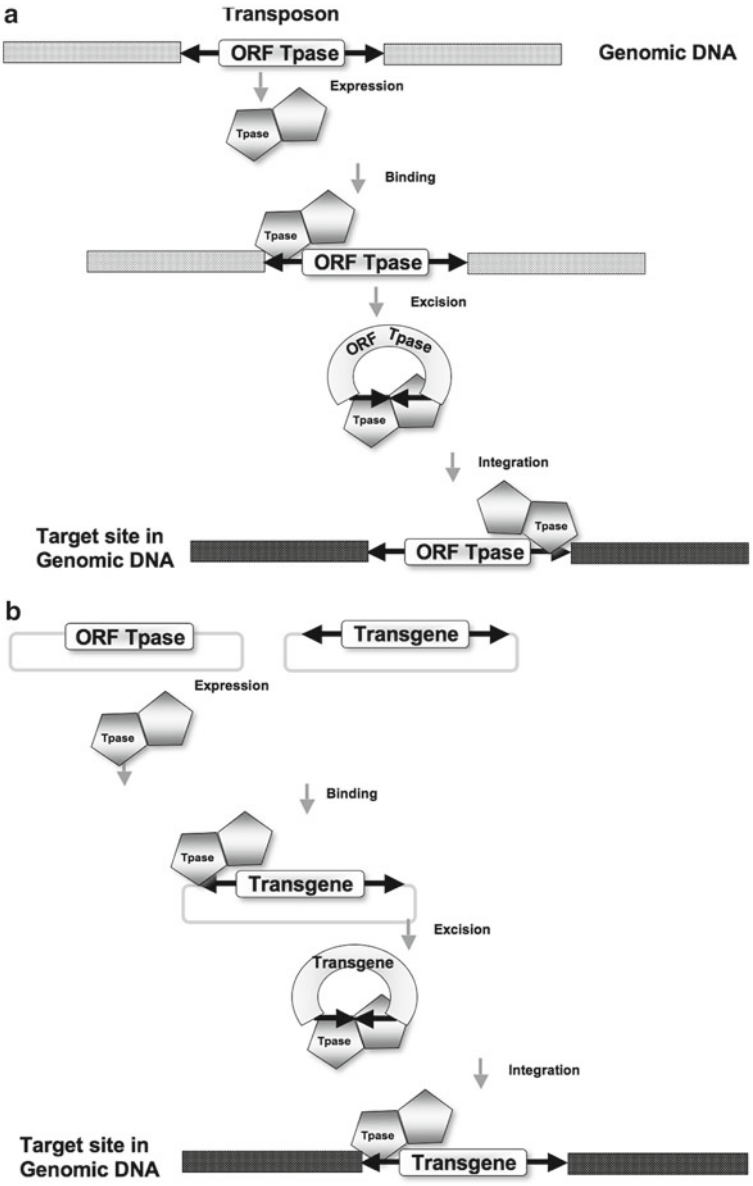


Fig. 9.1 Transposition cycle and transposase-mediated transgenesis. (**a**) Transposons consist of two inverted repeats (*IR*) (*black arrows*) surrounding an ORF encoding the transposase, and are inserted in genomic DNA (*grey box*). The transposase ORF (*pentagon*) is expressed and binds specifically to its IR, bringing together two IRs in a synaptic complex containing a transposase dimer in the case of Mos1. The transposase cleaves both strands at each IR, leading to excision of the transposon. Strand transfer leads to integration of the transposon at a new site (*black box*). (**b**) Transposase-mediated transgenesis is performed by replacing the transposase ORF on a transposon by the gene to be inserted (the transgene). Co-transfection of plasmids, one expressing the transposase and the other harbouring the transgene-containing transposon, is carried out in cultured cells. Transposase performs all of the steps as described in **a** to integrate the transgene at a genomic locus

while it will insert into many different sequences, it displays a preference for the consensus sequence A-**GNTYWRANC-**T (where the 9 bp target duplication is shown in bold) (Goryshin and Reznikoff 1998). Tol2 inserts into a weakly conserved AT-rich consensus sequence creating an 8 bp target duplication (Kondrychyn et al. 2009), while *piggyBac* inserts only into targets with the sequence TTAA (Fraser et al. 1995). All members of the Tc1/mariner/IS630 family of transposons, including Mos1, Himar, SB and ISY100 insert exclusively into TA dinucleotides and duplicate this TA upon insertion. However, different T/A-rich consensus sequences have been reported outside the conserved TA dinucleotide for each of these elements (Préclin et al. 2003; van Luenen and Plasterk 1994; Feng et al. 2010; Vigdal et al. 2002).

In addition to these sequence preferences, many transposons show other regional target choice preferences. *SB* is believed to integrate more frequently into expressed genes (Vigdal et al. 2002; Walisko et al. 2006), while *Mos1* demonstrates a preference for rRNA coding genes in *C. elegans* (Granger et al. 2004) and the *cat* gene in *E. coli* (Crénès et al. 2009). In addition, local hopping is frequently observed, resulting in integrations in regions close to the donor site (Guimond et al. 2003; Keng et al. 2005; Kondrychyn et al. 2009). Despite this only quasi-random insertion specificity, transposons integrate at many different locations throughout the genome and can be used in random mutagenesis approaches to obtain insertions in essentially all genes. This has been successfully applied to identify oncogenes in mice using Sleeping Beauty (Copeland and Jenkins 2010) or for the saturation mutagenesis of Mycobacterium with Himar (Rubin et al. 1999; Sassetti et al. 2001).

However, many approaches in the post-genomic era require efficient targeted transposition. For instance, if transposons could be targeted to a chosen site, any desired gene could be inactivated or modified. Transposon-based gene delivery for gene therapy would also benefit greatly from targeted transposition. Transposons have many advantages over viral vectors for gene therapy (VandenDriessche et al. 2009). However, random insertions cause insertional mutagenesis, inactivating essential host genes or inappropriately activating harmful genes such as oncogenes. In addition, insertion into regions of heterochromatin can lead to poor expression of the transgene (Chap. 1). If transposition could be targeted to specific sites in the genome, the drawbacks of random insertion by natural transposons could be avoided. Strategies have been developed to target transposition to specific sites using DNA binding domains (DBDs) which recognize specific sequences and tether either the transposase protein or the transposon DNA to a specific chosen genomic target.

In the first strategy (Fig. 9.2a), target site specificity is grafted onto the transposase protein itself by adding an extra DBD. The transposon-based vector is then expected to insert into (or near) the sequence recognized by the DBD. This type of strategy has already proved effective in a number of non-transposase based systems. For instance, fusing HIV-1 integrase to the E2C-DBD (a zinc finger (ZF) DBD which binds the utrophin gene promoter) gave a protein that integrated HIV termini into a plasmid carrying the E2C target site *in vitro* (Tan et al. 2004). However, no *in vivo* targeting was detected, due to the poor infectivity of the modified HIV (Tan et al. 2006). The targeting specificity of other retroviral integrases has been similarly

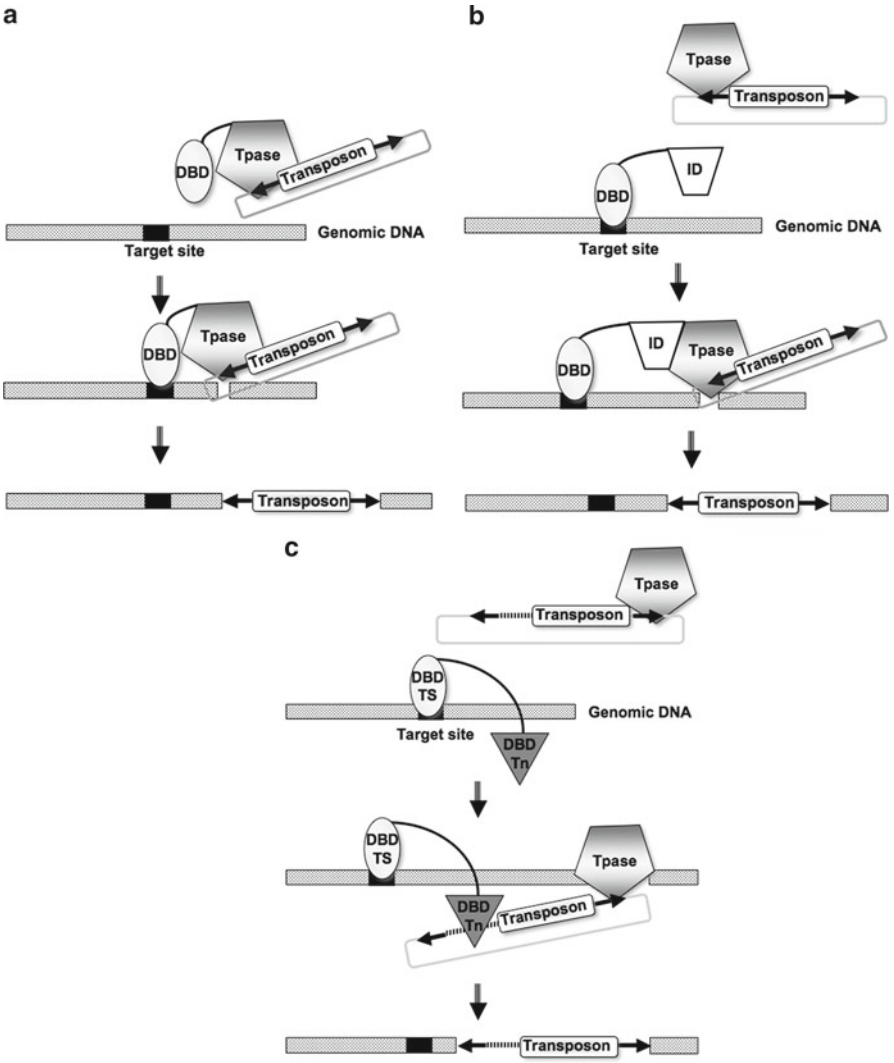


Fig. 9.2 Transposase and transposon targeting strategies. (**a**) The transposase (*pentagon*) linked to a DBD (*oval*) binds specifically to a chosen target site (*black box*) in the genomic DNA (*grey box*). The DBD is bound to the chosen target site, whereas the transposase is expected to promote excision and integration of its cognate transposon (*white box* surrounded by two *black arrows*) near the DBD target site. (**b**) The DBD (specifically bound to its target site in the genomic DNA) is linked to a protein (ID, *trapezoid*), which interacts specifically with the transposase, thus bringing the transposase to the chosen target site. (**c**) The DBD (specifically bound to its target site in the genomic DNA) is linked to another DBD (DBD-Tn, *triangle*), which binds to a specific sequence (*dashed line*) located in the transposon, thus bringing the transposon near the chosen target. The transposase (*pentagon*) then performs transposition. *DBD* DNA binding domain, *Tpase* Transposase, *ID* interacting domain, *Tn* transposon

modified (Bushman 1994; Goulaouic and Chow 1996; Katz et al. 1996). In a related approach, attaching the catalytic domain of the restriction endonuclease *Fok*I to ZF-DBDs, has been used to create targeted ZF-endonucleases. Specific cleavage by the ZF-endonuclease stimulates repair using a homologous DNA template, giving efficient site-specific gene correction (Porteus and Carroll 2005).

These results show that DBD-fusion technology is capable of delivering a protein to a specific genomic site. In an attempt to design site-directed transposases, this technology has therefore been applied to several transposases of bacterial (IS30, ISY100) or eukaryotic origin (*SB*, Mos1, *piggyBac, Tol2*) fused to a number of different DBDs (λcI, Gli, Zif268, Sp1, E2C, Gal4, Jazz) (Szabo et al. 2003; Feng et al. 2010; Wilson et al. 2005; Yant et al. 2007; Maragathavally et al. 2006; Ivics et al. 2007; Wu et al. 2006).

A second strategy involves indirectly targeting the transposase protein to the chosen genomic site using protein-protein interactions to a DNA-bound protein (Fig. 9.2b). Neither the transposase nor the transposon is modified. Instead, an additional protein consisting of a sequence specific DBD fused to a transposase-binding protein is used. This strategy has been successfully demonstrated for Sleeping Beauty (Ivics et al. 2007). Indeed, many naturally occurring transposable elements utilize protein–protein interactions for targeted insertion into defined sites. This is the case for Ty5 (an endogenous retrotransposon of *S. cerevisiae*) that is naturally targeted into heterochromatin, due to a specific interaction between Sir4P (a component of the heterochromatin) and the Ty5 integrase (Zhu et al. 2003). The specificity of Ty5 integration can be altered by fusing the LexA DBD (a component of *E. coli* SOS damage response) to the C-terminal part of Sir4P. Under these experimental conditions, inter-plasmid targeting assays revealed targeted integrations near LexA binding sites (Zhu et al. 2003). Similarly, the integration specificity of HIV-1 integrase has been modified by fusing the integrase binding domain of the LEDGF protein to a bacterial DBD (Ciuffi et al. 2006), and targeting of the *Mu* transposon to a specific locus has been achieved using a fusion between MuB and the Arc DBD (Schweidenback and Baker 2008).

A third strategy is to tether the transposon DNA itself to the chosen target site. This is achieved by creating a fusion protein containing two different DBDs (Fig. 9.2c). The first DBD binds to a DNA sequence present in the transposon, and the second DBD binds to the chosen genomic integration site (Ivics et al. 2007). This idea is based on data from *P*-element transposition in *D. melanogaster*. It was reported as early as 1990 (Hama et al. 1990) that *P*-element derivatives containing 3.4-kb of upstream regulatory DNA from the segmentation gene *engrailed* insert at a very high frequency near the endogenous *engrailed* gene. A few years later, another study showed that these insertion events do not occur at specific DNA sites, but rather in "regions" of the chromosome (Kassis et al. 1992). A model was therefore proposed in which one or more protein(s) bound to the *engrailed* fragment within the *P*-element brings it to a particular region of the genome as a result of protein-protein or protein-DNA interactions. This phenomenon is now known as "transposon homing" (Guimond et al. 2003).

In the following sections, the use of these different strategies will be described in four different sections, according to the different classes of host organisms used: bacteria, insects, fishes and mammalian cells. It should be noted that no transposon-based targeted insertion strategy has yet been developed for plants.

9.2 Targeted Transposition in Bacteria

Although homologous recombination is probably the preferred method for targeted gene delivery in many bacteria, this strategy is not applicable to all bacterial species. Modified transposases are therefore being developed in bacteria for gene delivery to chosen genomic locations, for efficient gene-specific mutagenesis, and as model systems for targeted transposition in more complex organisms. The strategy used for targeting transposons in bacteria is to create chimaeric transposase proteins by fusing specific DBDs to the transposase (Fig. 9.2a). This strategy has been tested with the bacterial transposases of IS30 and ISY100 and the insect transposase Mos1 (Szabo et al. 2003; Feng et al. 2010; Demattéi et al. 2010).

The first proof-of-principle for specific targeting with modified transposase was with the bacterial transposon IS*30* (Szabo et al. 2003). The authors tested whether the addition of a DBD (the cI repressor of bacteriophage λ) to the C-terminus of IS30 transposase could modify the insertion specificity of IS30. Inter-plasmid transposition was performed in *E. coli* using two plasmids: a transposon donor plasmid containing two abutted IRs separated by 2 bp ($(IS30)_2$, a natural intermediate in IS30 transposition) and a target plasmid containing the λ operator O_R. The chimaeric transposase retained full transposition activity into a consensus IS30 hot site, showing that the addition of an extra DBD does not abolish transposition activity. The cI domain was also functional as the fusion protein conferred immunity to infection by λ. 53 independent insertions into the target plasmid containing the cI binding site (O_R) were examined. In 30 out of these 53 transposition events, the transposon had inserted close to (within 400 bp of) the target sequence, and in one case it had inserted directly into O_R. The specific targeting depended on the presence of both the cI DBD and the O_R binding site. No insertions were obtained into the target plasmid with wild-type transposase or when the target plasmid lacked O_R. This was the first demonstration that target specificity could be modified by fusing a DBD to a transposase.

A similar approach with IS*30* transposase fused to a DBD was recently applied to target integration specifically into the flagellin operon of *Salmonella* Enteridis (Imre et al. 2011). The *fliC* gene encodes flagellin H1 and its expression is controlled by the FljA repressor which binds an operator sequence between the divergently transcribed *fliC* and *fliD* genes. Mutants in *fliC* have reduced motility, and might therefore be used to generate attenuated live oral vaccines (Parker and Guard-Petter 2001). Transposon insertions were produced in *S.* Enteridis using a plasmid expressing IS30 transposase fused to the FljA repressor, and a donor plasmid containing the transposition intermediate $(IS30)_2$ (Imre et al. 2011). The transposase-FljA

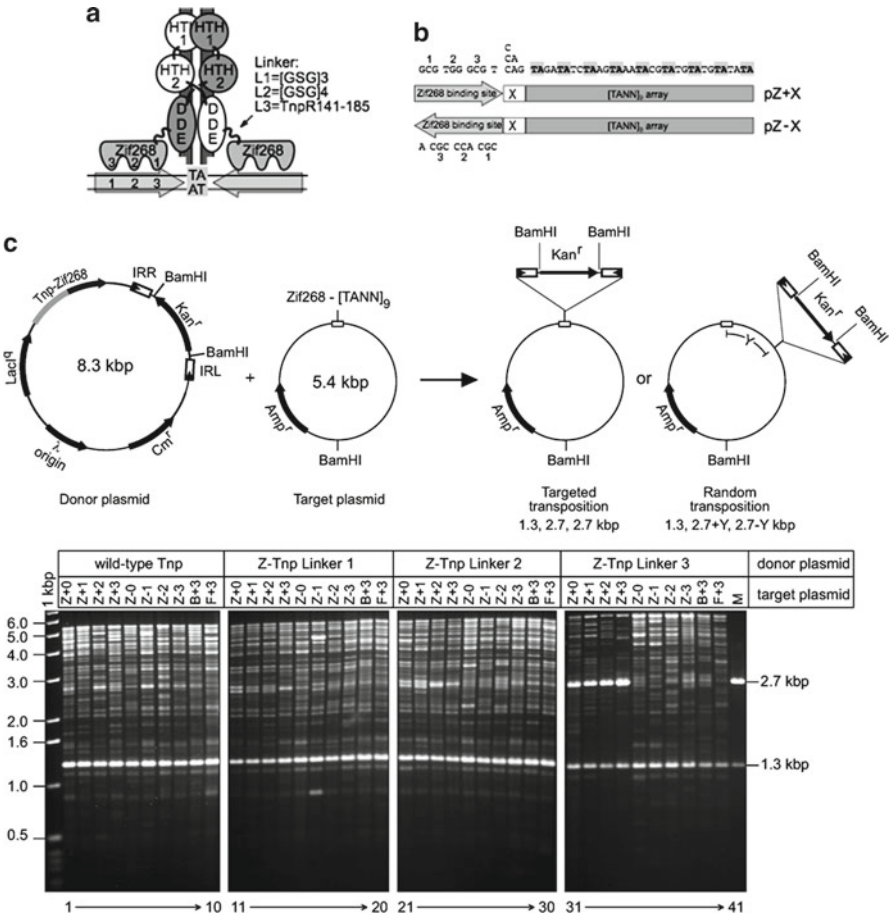


Fig. 9.3 Targeted integrations of a transgene with DBD-transposase chimaeras in bacteria and insects. (**a**) ISY100 transposase is expected to act as a dimer (*white* and *light black monomers*). ISY100 transposase contains two N-terminal helix–turn–helix domains (HTH1 and HTH2), which recognize DNA sequences within the transposon IRs (*black arrows*) and a catalytic DDE domain (DDE) which integrates the IRs into a target TA. The C-terminus of transposase was fused to the DNA-binding domain of Zif268 via three different linkers: L1, L2 or L3. The three zinc-fingers of Zif268 are numbered one to three, as are the three triplets of its recognition sequence. Note that finger 1 recognizes triplet 3, and finger 3 recognizes triplet 1. Target sites contain a single Zif268 binding site (*grey arrow*) adjacent to an array of target TAs. The dimeric nature of Z-transposase suggests how a second Zif268 binding site (*grey arrow*) could enhance the specificity or efficiency of targeting (Feng et al. 2010) (with permission from *Nucleic Acids Res*). (**b**) Target sites for chimaeric Z-transposase consist of a 10 bp Zif268 binding site separated from a [TANN]9 array by 0, 1, 2 or 3 bp (Feng et al. 2010) (with permission from *Nucleic Acids Res*). (**c**) Targeted transposition of ISY100 by a chimaeric Z-transposase. *Upper panel*. Structure of donor and target plasmids. Donor plasmids carry a mini-ISY100 transposon encoding resistance to kanamycin, and express transposase from a hybrid trp-lac promoter under the control of the plasmid-encoded lacIq gene on a λdv replicon. Target plasmids carry a [TANN]9 array adjacent to a binding site for Zif268.

fusion protein was expressed and retained both transposition activity, and the ability to repress *fliC* expression. Insertion mutants produced using fusion and wild-type transposases were then screened for loss of motility. Four out of 600 mutants obtained with the fusion protein were completely non-motile (0.67%), whereas none of the 600 mutants produced with wild-type transposase were non-motile (<0.16%). Three of the non-motile mutants obtained with the fusion protein contained insertions in the *fliD* gene, all at exactly the same location close to the binding site of FljA. The fourth non-motile mutant contained an insertion in a gene of unknown function *yjjY* located in a distant region of the genome.

Analysis of the insertion sites showed that the transposase-FljA fusion protein retained quite similar sequence preference to wild-type transposase. However, the fusion protein had four favoured insertion sites in the genome not observed for the wild-type transposase, in particular within the *fliD* gene (3 of 600 insertions) and within the *yjjY* gene at a different position from the insertion that caused complete loss of motility (11/600). Insertions in and around y*jjY* appear to cause partial or complete loss of motility, and the authors speculate that there might be a nearby unidentified FljA binding site (Imre et al. 2011). Although only a relatively small proportion of inserts were targeted, these experiments show the potential of IS30-based site-directed mutagenesis to obtain mutants in a chosen gene in organisms where other strategies may not be applicable.

Recently, the target specificity of ISY100 transposase from the Cyanobacteria *Synechocystis* sp. has also been successfully modified by fusion to a specific DBD. ISY100 transposase was fused at its C-terminus to the DBD of the eukaryotic zinc finger transcription factor Zif268. This chimaeric "Z-transposase" catalysed transposition at high frequency in close proximity to the Zif268 binding site (Feng et al. 2010). ISY100 belongs to the widespread Tc1/mariner/IS630 family of mobile elements. Like all other members of this family, ISY100 transposes specifically into TA target sites. In addition ISY100 has a preference for target sequences of the form ADW**TA**WHT (with the central target TA in bold, and where W = A or T, D = not C and H = not G) (Feng et al. 2010). Inter-plasmid transposition assays were carried out in *E. coli* using two plasmids: a donor plasmid encoding the transposase and carrying a mini-transposon, and a target plasmid harbouring the 9 bp Zif268 binding site adjacent to an array of nine copies of the sequence TANN. By placing this TANN array 0, 1, 2 or 3 nucleotides from the ZBS, with the ZBS in both possible ("+" or "−") orientations, potential TA targets were put at all possible distances from 1 to 36 bp to either side of the ZBS (Fig. 9.3b). Three different Z-transposases were tested, differing only in length

Fig. 9.3 (contiuned) On digestion with BamHI, transposition products give a 1.3 kb kanamycin resistance fragment and two other fragments that add up to 5.4 kb. If the transposon is in the [TANN]9 target, a 2.7 kb doublet is produced. *Lower panel*. Agarose gel showing BamHI-digested pooled transposition products from assays using the indicated donor and target plasmids. The sizes of the Kan^r fragment (1.3 kb) and the doublet indicative of targeted transposition (2.7 kb) are shown. Lane 41 (M) contains BamHI-digested DNA from an isolated targeted transposition product (Feng et al. 2010) (with permission from *Nucleic Acids Res*)

and sequence of the peptide linker (L1, L2 or L3) between the C-terminus of transposase and the N-terminus of the Zif268 DBD (Fig. 9.3a). All three Z-transposases retained transposition activity, though 3.5- to 9-fold lower than wild-type. Targeting specificity was analysed by restriction enzyme digestion of pooled transposition products (Fig. 9.3c) This analysis showed that only Z-transposase with the longer L3 linker gave targeted integration, and only if the ZBS was in the "+" orientation (Fig. 9.3c). With this combination of transposase and targets, approximately 50% of the insertions were adjacent to the ZBS in the TANN array, the most efficient demonstration of targeted integration described to date. All others combinations of wild-type and Z-transposases, with "+" and "–" orientations of the ZBS, gave random integration throughout the target plasmid (Fig. 9.3c). Zif268 is a modular zinc finger DNA binding protein, and much work has gone into the creation of variants that bind specifically to different specific DNA sequences (reviewed in Klug 2010). Two such variants of Zif268 (A and B) were fused to ISY100 transposase and their target specificity was tested with the expected result: A- and B-transposases had modified target specificity, and preferentially catalysed transposition into their cognate target sites (Feng et al. 2010). Efficient targeted transposition was also obtained into a single TA target flanked by two Zif268 binding sites. This could lead to the production of heterodimeric Z-transposases that could target transposition specifically into sequence flanked by two specific 9 bp ZBS, long enough to be unique in a complex eukaryotic genome. This study validates the efficiency of targeted integration with modified transposases in inter-plasmid assays in *E. coli*. For future development, it has yet to be proved that ISY100 can catalyse transposition in other bacteria and even eukaryotes. Considering that ISY100 was isolated from *Synechocystis* sp., and is able to transpose in *E. coli*, it seems likely that ISY100 will at least be active in a wide range of other bacterial species, as already demonstrated for another Tc1/mariner element *Himar* (Rubin et al. 1999).

Mos1 isolated from the insect *D. mauritiana* is also able to transpose in *E. coli* (Augé-Gouillou et al. 2001). Attempts to target transposition in *E. coli* were made by fusing the N-terminus of Mos1 transposase to a four zinc finger DBD (ZF) (Demattéi et al. 2010) that recognizes the *bcr-abl* fusion oncogene sequence (ZBS12; McNamara and Ford 2000). The transposition activity of ZF-Mos1 was evaluated in *E. coli*, and was shown to be at least 100-fold lower than the wild-type transposase. Biochemical analyses with purified ZF-Mos1 show that the fusion protein is defective in both ITR-binding and ZBS12-binding activities, as confirmed by EMSA and was unable to promote *in vitro* transposition. The fact that transposition is observed in *E. coli* and not *in vitro* is probably due to the presence of some truncated DBD-Transposase in *E. coli*, giving a low level of transposition activity. The *in vitro* transposition assay was performed with purified full-length DBD-transposase and as no transposition was observed, it can be concluded that this fusion protein is not active.

9.3 Targeted Transposition in Insects

Transposons, especially the P element, have long been used to integrate transgenes into the genome of the fruit fly *Drosophila melanogaster* (Ryder and Russell 2003). More recently, similar techniques using transposons such as *piggyBac*, *Mos1*, *Hermes* and *Minos*, have been developed for gene delivery in other important insect species such as the mosquito *Aedes aegypti* (O'Brochta and Handler 2008). However, a major problem with the use of transposons for gene delivery is their integration at random sites, which can lead to reduced fitness of their host or poor transgene expression. Therefore attempts have been made to target *piggyBac* and Mos1 transposition in embryos of *Aedes aegypti*, (Maragathavally et al. 2006) and *piggyBac* transposition in cell lines derived from the silkworm *Bombyx mori* and *D. melanogaster* (Wang et al. 2010). The sequence specific DBD of the yeast transcription factor Gal4 was fused to the N-termini of both *piggyBac* and Mos1 transposases using a flexible linker (Fig. 9.2a). Inter-plasmid transposition assays were carried out using a plasmid harbouring a mini-transposon (two IRs surrounding a bacterial kanamycin resistance gene and a plasmid replication origin) and a chloramphenicol resistant target plasmid containing the DNA binding site for Gal4 (the Gal4 upstream activating site or UAS). These plasmids were introduced into embryos or cell lines in conjunction with a plasmid expressing the Gal4-transposase fusion. After allowing time for transposition to occur, DNA was isolated and transposition products were selected in *E. coli* by virtue of the transposon-borne kanamycin resistance and plasmid replication origin, and the target plasmid chloramphenicol resistance gene.

Addition of the Gal4 DBD to Mos1 and *piggyBac* transposases appeared to have no adverse effect on transposition. Mos1 and Gal4-Mos1 catalysed transposition into target plasmids lacking the UAS at similar frequencies (Maragathavally et al. 2006). The Gal4 *piggyBac* fusion also catalysed transposition but this protein was not compared to the wild-type transposase (Maragathavally et al. 2006). Transposition in *Ae. aegypti* embryos, and *B. mori* and *D. melanogaster* cell lines produced the expected target site duplications of TTAA for *piggyBac* and TA for Mos1.

Transposition by Gal4-Mos1 transposase in mosquito embryos was 13-fold more efficient when the target plasmid contained the UAS, whereas transposition by the Gal4-*piggyBac* fusion was 4–12-fold more efficient when the UAS was present. In all cases, hot spots for integration were reported about 1 kbp away from the UAS in the 2.7 kbp target plasmid. In *Ae. egypti* embryos, 51 out of 53 insertions catalysed by Gal4-Mos1 were in a single TA target site, 954 bp from the UAS, whereas 45 out of 67 (67%) insertions produced by the Gal4-*piggyBac* were in a single TTAA 912 bp from the UAS (Maragathavally et al. 2006). The distribution of insertions made by Gal4-*piggyBac* in insect cell lines also appeared to be non-random, but there appeared to be several hot sites that were different in *B. mori* and *D. melanogaster* cell lines, and all of these were different from the preferred integration site in *Ae. egypti* embryos (Wang et al. 2010). No inserts were obtained on either side immediately adjacent to the UAS. It is not clear what mechanism could target transposition

specifically to a site 1 kbp from the Gal4 binding site, and interpretation of the data is further complicated by the fact that no inserts immediately adjacent to the UAS on one side can be isolated with this strategy, as they would inactivate the chloramphenicol resistance gene on the target plasmid. Use of a target plasmid with the UAS in both orientations at a site distant from the resistance gene is needed to clarify the results. Whether a similar approach, perhaps with zinc-finger transposase fusion proteins, could be used to target transposition to chosen sites in insect genomes has yet to be seen.

9.4 Targeted Transposition in Zebrafish

Targeted transposition has been reported in the zebrafish (*Danio rerio*) with the bacterial IS30 transposase (Szabo et al. 2003). As well as being the only report of targeted transposition in a fish, this report is also notable as one of only a few studies where a transposase of bacterial origin is shown to be active in eukaryotes (both HeLa cells and zebrafish embryos). The only other report known to these authors of a bacterial transposase catalysing transposition in a eukaryote is for bacteriophage Mu transposase, which was shown to be active in yeast, mouse and human cells (Paatero et al. 2008).

To obtain targeted integration, the DBD of the Gli1 transcription factor was fused to the C-terminus of IS30 transposase and the target binding site of Gli1 (*gli*) was inserted into the first intron of the zebrafish sonic hedgehog (*shh*) gene. The donor transposon carries a copy of green fluorescent protein (GFP) gene, such that transposition into *shh* intron 1 will place GFP expression under the control of the *ssh* regulatory sequences leading to specific expression in the notochord (Szabo et al. 2003). The donor transposon and target plasmids were injected into zebrafish embryos simultaneously with IS30 transposase-Gli1DBD mRNA. GFP expression was observed only when the fusion transposase and the target plasmid carrying the *gli* target site were present. Expression of GFP in notochord cells was observed in 3 out of 104 embryos when the target carried the *gli* site whereas none out of 102 embryos expressed *gfp* in the notochord when the *gli* site was absent from the *ssh* intron in the target plasmid (Szabo et al. 2003). Junctions between the donor and target plasmid in the region of the Gli1 binding site were amplified from pooled embros by PCR and analysed by DNA sequencing. Six out of 12 different insertion sites were within 100 bp of the Gli1 binding site. However the majority of insertions appeared to have occurred by a transposase dependent illegitimate recombination event, as the expected IS30 IR-target junction was only observed at one out of 12 insertion sites. These results suggest that the IS30 fusion protein initiates transposition near to the Gli1 binding site but that the reaction is blocked at some stage producing an intermediate that is subject to a host repair process. Nevertheless, it appears that fusing DBDs to IS30 holds promise as a means to deliver transgenes to specific genomic locations in fish or other eukaryotes, and that this technique could also be used to identify genomic locations of transcription factor binding sites. So far only inter-plasmid

targeted transposition has been reported and it will be important to show that genomic insertions can be obtained. However, it is surprising that no other publications have made use of this potentially powerful system in zebrafish.

9.5 Targeted Insertions in Mammalian Cells

The three strategies to target integration of transposons shown in Fig. 9.2 have all been attempted in mammalian cells. The most thoroughly tested strategy is to fuse a DBD to the transposase protein, and SB and *piggyBac* have been used extensively for these studies (Wilson et al. 2005; Yant et al. 2007; Ivics et al. 2007; Wu et al. 2006; Kettlun et al. 2011). One of the largest barriers to this strategy is that fusion of a DBD to transposases substantially reduces, or even completely abolishes their catalytic activity. In this section we first review the effects on transposition activity of DBD fusions, and then look at the success of this strategy to bring about targeted integration. Finally we examine strategies that do not require altering the transposase protein.

The first reported use of a transposase-DBD fusion in a mammalian system was with C-terminal and N-terminal fusions of the zinc finger DBD Sp1 to SB transposase (Wilson et al. 2005). Transposition from a plasmid to the genome of HEK293 cells was undetectable using the C-terminal fusion, and activity could not be restored using hyperactive transposase mutants, a longer linker between the transposase and the DBD, or the use of a hyperactive transposon mutant. Indeed these authors report that the addition of other DBDs, or even a short peptide tag to the C-terminus of SB transposase abolished its activity, even though expression could be detected by Western blots. Yant and colleagues (2007) also tested C-terminal fusions to SB transposase. These authors used the synthetic polydactyl zinc finger protein E2C fused to the C-terminus of a hyperactive SB transposase HSB5, as they had previously observed that even the addition of a 6 × His tag to the N- or C-terminus of the less active SB10 transposase resulted in substantial loss of activity. However, even with this hyperactive mutant, and using flexible linkers consisting of one, two or three copies of a glycine-glycine-serine (GGS) tripeptide, fusions of the E2C DBD to the C-terminus of HSB5 transposase were inactive for both excision and transposition (Yant et al. 2007). A similar lack of activity was also reported for the E2C DBD fused to the C-terminus of SB transposase by a linker consisting of ten glycine residues (Ivics et al. 2007).

Much more success has been had with fusions of DBDs to the N-terminus of SB transposase. SB with an N-terminal Sp1 domain retained 17% of native enzyme activity (Wilson et al. 2005). Addition of an N-terminal nuclear localisation signal (NLS) or increasing the length of the linker between the DBD and transposase did not increase the efficiency of transposase, but introducing activating mutations in the transposase (SB12) and using hyperactive versions of the transposon IRs increased transposition up to 85% of the native SB level (Wilson et al. 2005). SB and other Tc1/mariner family enzymes show a phenomenon known as overproduction

inhibition, in which an excess of transposase inhibits transposition. Interestingly, Wilson et al. (2005) noted that fusion of the Sp1 DBD to the N-terminus of SB appeared to alleviate this overproduction inhibition. A protein consisting of six zinc fingers from a different ZF protein (ZF202) fused to the N-terminus of SB12 also retained a high level of activity and did not display overproduction inhibition (Wilson et al. 2005).

Yant et al. (2007) tested a series of N-terminal fusions of the E2C ZF-DBD to a hyperactive SB transposase (HSB5), using from zero to seven copies of the flexible GGS linker to join the two proteins. Transposition in HeLa cells ranged from 3- to 8-fold higher than the background obtained with no transposase, and was most efficient with the linker containing five copies of the GGS sequence. The level of transposition obtained with this E2C-$(GGS)_5$-HSB5 was only about 10% of the level observed with the non-fusion HSB5 protein. It was noted that the E2C fusion proteins were poorly expressed in human cells and that expression and the transposition rate could be improved 2–3-fold (to about 23% of the native HSB5 level) by optimising the codon bias of the E2C domain and $(GGS)_5$ linker for expression in human cells. A fusion of the yeast Gal4 DBD to the N-terminus of HSB5 using the $(GGS)_5$ linker also catalysed transposition at about 25% of the native HSB5 level, though this protein left unusual footprints at the repaired donor sites, suggesting an aberrant excision reaction of some sort (Yant et al. 2007).

In another study, transposition by *piggyBac*, Mos1, Tol2 and a first generation hyperactive SB transposase SB11 was compared in four different types of mammalian cells (HeLa, H1299, Hek293 and CHO cells) (Wu et al. 2006). Mos1 gave no detectable transposition in any of these cell lines, whereas *piggyBac* gave the highest level of transposition in all four cell lines. These authors fused the Gal4 DBD to the N-termini of *piggyBac*, SB and Tol2 transposases, and found that the *piggyBac* fusion retained near wild-type levels of transposition in HEK293 cells, while SB11 and Tol2 transposases were completely inactivated (Wu et al. 2006). This result disagrees with that of Yant et al. (2007) who found that a Gal4-SB fusion retained a high level of activity (see above). The most likely explanation for this discrepancy is that Yant et al. (2007) used an optimised $(GGS)_5$ linker to fuse the two domains whereas Wu et al. (2006) used an 18–21 amino acid linker of unpublished sequence. Other differences between the experiments, the cell types used (HeLa vs HEK293) and the two different hyperactive versions of transposase (HSB5 vs SB11), might also account for the different results reported.

Ivics et al. (2007) tested a number of N-terminal fusions of DBDs to SB transposase. Fusion of the bacterial tetracycline repressor, or the E2C zinc finger DBD to the N-terminus of SB totally inactivated it for excision and transposition, whereas a fusion to a different zinc finger DBD (Jazz) gave excision and transposition activity at about 10–15% of the level of wild-type SB. These authors used a ten glycine sequence to link the different domains, and again this may account for the different results obtained for E2C N-terminal fusions compared to the activity observed for a similar fusion by Yant et al. (2007).

The target specificities of active modified transposases described in the studies above were tested in mammalian cells, using either inter-plasmid or genomic integration transposition assays. Inter-plasmid assays in mammalian cells are similar to

those developed in insects and zebrafish, and use a target plasmid carrying binding sites for the DBD, a donor plasmid carrying a mini-transposon, and a transposase expression plasmid. E2C-SB and Gal4-SB fusion proteins with $(GGS)_5$ linkers were tested in HeLa cells using such an assay (Yant et al. 2007). Two days after introducing donor, target, and transposase expression vectors into HeLa cells by co-transfection, plasmid DNA was recovered and plasmids resulting from transposition were selected in *E. coli* (Fig. 9.4c). Insertion sites in the target plasmid were localized by restriction enzyme digestion and by sequencing (Fig. 9.4c, d). Transposition into a 443 bp fragment containing five repeats of the E2C binding site was increased 8-fold when E2C-$(GGS)_5$-HSB5 was used compared to the non-fusion transposase, and transposition in an equivalent region adjacent to a UAS was increased 11-fold when GAL4-$(GGS)_5$-HSB5 was used (Yant et al. 2007) (Fig. 9.4e). The Gal4 fusion catalysed integration adjacent to an E2C binding site at the same frequency as wild-type transposase (Fig. 9.4e), showing that targeting is DBD and target-site specific. Furthermore, the DBD had to be fused to the transposase to have an effect. When E2C was expressed separately along with non-fusion SB transposase, binding by E2C prevented any integration near to the E2C binding site. Curiously, targeting was more efficient with E2C-$(GGS)_5$-HSB5 when the target site was mutated so that only three of the six zinc fingers of E2C could bind. The authors speculated that tight binding by all six fingers of the E2C DBD might somehow prevent the transposition reaction reaching completion, and that flexibility of the DNA binding domain adjacent to the N-terminus of SB transposase might be required for some conformational change in the transposition reaction (Yant et al. 2007).

The location of genomic insertions in mammalian cells was investigated with three of the SB-based chimaeric transposases: Sp1-SB (Wilson et al. 2005), E2C-$(GGS)_5$-HSB5 (Yant et al. 2007) and Jazz-SB (Ivics et al. 2007). Results with Sp1-SB could not be interpreted due to the high frequency of predicted binding sites for Sp1 in human genomic sequences. Putative Sp1 binding sites were found at high frequency within 2 kb windows around insertions mediated by wild-type SB as well as in randomly selected 2 kb regions in the human genome (Wilson et al. 2005). In contrast, E2C (an artificial six finger DBD) has a single binding site in the human genome, in the erb2 gene on chromosome 17. Integration sites were mapped for 67 insertions by E2C-$(GGS)_5$-HSB5 and 55 insertions by the non fusion HSB5. There was no increase in the number of insertions on chromosome 17, and none of the inserts were near to the endogenous E2C binding site in the erb2 gene (Yant et al. 2007). Similar results were obtained with the Jazz-SB chimaera, where no insertions were observed near to the targeted utrophin locus, and no copies of the 9 bp Jazz binding site were detected within 1 kb of any of the insertions sites (Ivics et al. 2007). One possible explanation for the lack of targeted transposition in these two studies is that not enough insertion sites were analysed. Only 67 integration sites were analysed with E2C-$(GGS)_5$-HSB5 (Yant et al. 2007), whereas the number of insertions characterised for Jazz-SB was not reported (Ivics et al. 2007). It is worth noting that in their bacterial targeting experiment, Imre et al. (2011) characterised 600 insertions to isolate 3 targeted events. In a mammalian genome, the fraction of targeted insertions might be much lower due to the larger genome size, and many more insertions would have to be studied to detect a low level of targeted integration.

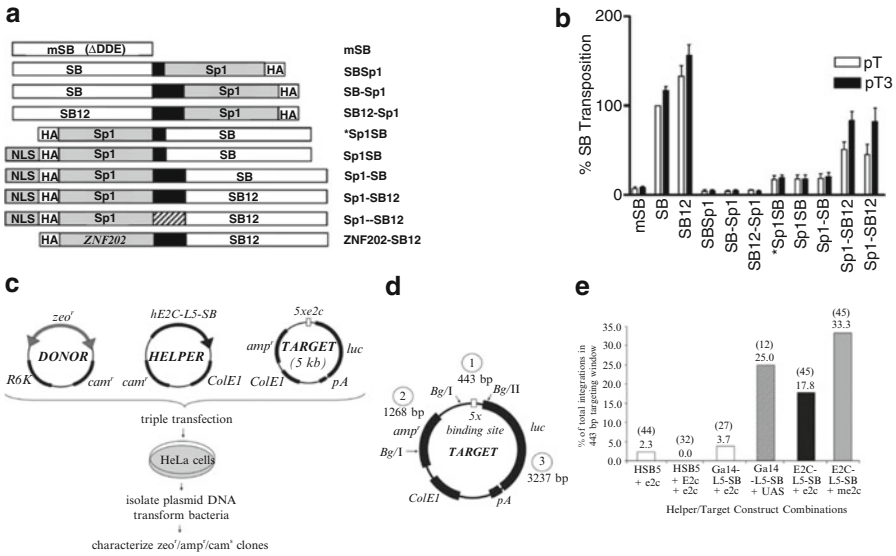


Fig. 9.4 Targeted integrations of transgene with DBD-transposase chimaeric proteins in mammalian cells. (**a**) Schematic representations of recombinant chimaeric SB transposases. Each *horizontal bar* represents a specific recombinant protein annotated with the following abbreviations: *mSB (DDDE)* catalytically inactive SB transposase, *SB12* hyperactive SB transposase, *Sp1* Sp1 DNA binding domain fused to either the N-terminus (Sp1SB) or C-terminus (SBSp1) of SB transposase, *Sp1SB* denotes a construct with a minimal peptide linker sequence of Gy-Thr-Gly between protein domains, *Sp1-SB* denotes chimaera constructed using the linker $GTG(SGGS)_3(SGT)_2G$, *Sp1--SB* denotes chimaera constructed using the linker GTTVDRSSDPTSQTSGTG, * denotes absence of additional nuclear localization signal, *NLS* denotes additional N-terminal nuclear localization signal derived from the SV40 large T antigen, *HA* epitope tag, *ZNF202* ZNF202 DBD(Wilson et al. 2005) (with permission from *FEBS Lett*). (**b**) Transposition assays. HEK293 cells were transfected with a transposon containing a neomycin resistance cassette and the recombinant chimaeric SB transposase as indicated in the figure. Transposition activity was quantified by counting colonies after 2 weeks of selection under G418. Abbreviations are as follows: *pT* native transposon, *pT3* hyperactive transposon. Data shown are mean values with S.D. (N = 3–6) (Wilson et al. 2005) (with permission from *FEBS Lett*). (**c**) Schematic overview of plasmid-based assay for investigating site-directed transposition. Transposition was initiated by transfecting human HeLa cells with a plasmid encoding chimaeric or unfused HSB5 transposase, together with a donor plasmid encoding a bleomycin-marked (zeo^r) transposon and a counter-selectable chloramphenicol-resistance (cam^r) gene. These plasmids were co-delivered with an ampicillin-resistant (amp^r) target plasmid containing five tandem DBD recognition sequences and allowed to undergo transposition. Low-molecular weight plasmid DNA fractions were isolated 2 days later and transformed into DH10B *E. coli*. Replication of the R6K origin-containing donor plasmid is strictly dependent on expression of the pir1 gene product, which is absent in this bacterial strain. Amp^r/zeo^r bacterial colonies were patched onto LB-chloramphenico plates to screen for inter-plasmid transposition events specific for the target plasmid (i.e. cam^s). Both pooled and clonal $amp^r/zeo^r/cam^s$ populations of bacteria were amplified, plasmid DNA isolated and the locations of transposon insertions relative to the target sites determined by restriction site analysis and DNA sequence analysis, respectively (Yant et al. 2007) (with permission from *Nucleic Acids Res*). (**d**) Target plasmid features. Positions of BglI and BglII restriction endonuclease recognition sites are shown, as are the sizes for each resulting DNA fragment (Yant et al. 2007) (with permission from *Nucleic Acids Res*). (**e**) Targeted transposition frequencies. Recombinant target DNA molecules were isolated from individual $amp^r/zeo^r/cam^s$ colonies and sequenced using an internal transposon-specific primer. Bars denote the percentage of total integrations occurring within the 443-bp targeting window, whereas the numbers in parentheses denote the actual number of integrations analyzed in each group (Yant et al. 2007) (with permission from *Nucleic Acids Res*)

Although appropriate plasmids were available to investigate targeting in inter-plasmid transposition with Gal4-*piggyBac* transposase, and this fusion transposase carried out transposition at rates similar to the native transposase, the results of any such experiment are not available (Wu et al. 2006). The lack of established mammalian cell lines containing an appropriate genomic UAS also prevented these authors from investigating targeted transposition into the genome (Wu et al. 2006).

A more recent study, using a zinc finger DNA binding domain fused to the N-terminus of *piggyBac* transposase, investigated transposition into both plasmid and genomic target sites (Kettlun et al. 2011). These authors fused an artificial zinc finger DBD that binds tightly and specifically to a single 18 bp sequence upstream of a checkpoint kinase gene in the human genome to create a fusion protein referred to as ZFP-*piggyBac*. Consistent with earlier results of Wu et al. (2006), this fusion protein retained 100% of its transposition activity (Kettlun et al. 2011). Chromatin immuno-precipitation assays demonstrated that ZFP-*piggyBac* recognised its target site in the human genome, but no inserts were detected close to this genomic target, perhaps due to a lack of nearby TTAA target sequences. An artificial target was therefore constructed with 14 TTAA sequences within 250 bp on either side of the ZFP binding site. In an inter-plasmid transposition assay, transposition into a plasmid carrying this target was 3.5-fold higher when ZFP-*piggyBac* was used than when native *piggyBac* was used. This increased was not seen when the ZFP binding site was scrambled in an otherwise identical target plasmid. There was a modest increase in insertions close to the ZFP binding site when ZFP-*piggyBac* was used in these inter-plasmid assays; 74% of insertions were within 250 bp of the ZFP binding site with ZFP-*piggyBac* but only 50% with wild-type *piggyBac*. A copy of the target construct was then integrated into the genome, and targeting was investigated in HEK293 cells. Cells containing genomic insertions were isolated by selection for drug resistance, and PCR was used to amplify any insertions close to the ZFR binding site. With wild-type *piggyBac* transposase, 23% of colonies contained insertions close to the ZFP binding site, and this increased to 44% when ZFP-*piggyBac* was used. Evidently, the target construct is a hotspot for insertions by the native *piggyBac*, probably due to the high concentration of TTAA sequences in a short stretch of DNA, and targeting was increased only slightly by the addition of the specific DBD to transposase. Quantitative PCR was used to show that there were 26–32 copies of *piggyBac* per cell in the drug resistant colonies obtained in these genomic transposition assays. The drug resistant colonies are probably heterogeneous, with individual different cells within the colony containing different complements of insertions. Thus the overall frequency of insertions that are targeted is much lower than the 40% of colonies that contain a targeted insertion, and every cell containing a targeted insertion will almost certainly contain multiple other insertions at different sites.

In general, it seems that it has been much easier to detect target integration in inter-plasmid assays than in assays for genomic insertions. This discrepancy between plasmid and genomic integration could result from differences in the number of target sites available. There are only two copies of a single genomic target in a eukaryotic cell, whereas up to 30,000 plasmid copies are present by 24 h post-transfection

(Carapuca et al. 2007). This should make targeting transposition to a plasmid target much easier than targeting to a genomic site. Furthermore it should be noted that a powerful bacterial selection scheme is used to isolate insertions in the target plasmid, and this makes it easy to observe even a modest increase in transposition in one region of the target plasmid. A much higher targeting specificity will be required to get insertion at a chosen genomic site at high frequency without a background of off target insertions.

Due to the loss of activity and poor targeting encountered with DBD-transposase fusion, Ivics et al. (2007) developed two alternative strategies that elegantly circumvent the need for transposase fusion proteins. Using SB, they targeted either the transposase or the transposon to chosen genomic sites. In the first strategy, SB transposase was targeted to a specific genomic location using a fusion between a transposase-binding domain and a DBD (Fig. 9.2b). The transposase binding domain consisted of the first 57 amino-acids of the SB transposase itself (N57), which is known to bind specifically to transposase, apparently without interfering with the transposition reaction catalysed by the full length protein (Izsvak et al. 2002). SB N57 was fused to a bacterial DNA binding protein, the tetracycline repressor (TetR) together with a nuclear localization signal (NLS). The hope was that this N57-NLS-TetR fusion protein would bind to SB transposase and tether it to a TetR binding site (TRE) thus directing transposition adjacent to the TRE. HeLa cells with a genomic copy of TRE were co-transfected with three plasmids: a SB transposase expression plasmid, a plasmid expressing TetR-NLS-N57, and a donor plasmid carrying a zeomycin resistant mini-transposon (Fig. 9.5a). About 400 colonies resistant to zeomycin, resulting from transposition of the mini-transposon into the genome, were pooled and their genomic DNA was analysed by PCR to detect integration adjacent to the TRE. Twelve insertion sites in the vicinity of the TRE were recovered (Fig. 9.5a), and about 10% of colonies contained an insertion that could be detected by PCR near the TRE (Ivics et al. 2007). Targeting required the N57 domain to be fused to TetR, and also required binding of TetR to the TRE. Controls with TetR fused to a different domain did not give any targeted insertions, and addition of doxycycline, which prevents TetR from binding to the TRE, also abolished targeted transposition.

The other strategy explored by Ivics and colleagues was to tether the transposon DNA to a specific genomic site using a bifunctional DNA binding protein (Fig. 9.2c). A binding site for the bacterial DNA-binding protein LexA was introduced into the SB mini-transposon, and targeting was carried out using fusions between LexA and a second DBD. This second DBD was either the scaffold attachment factor (SAF) which binds to matrix attachment regions (MARs) that are found throughout the human genome, or the tetracycline repressor TetR which binds to a single genomic copy of the TRE as described above (Ivics et al. 2007). The presence of the LexA-SAF fusion protein led to a small but statistically significant increase in the number of insertions within 1 kb of genomic MAR sequences, and a corresponding decrease in the number of insertions further than 5 kb from a MAR. However, most of the insertions detected were remote from a MAR. When the LexA-TetR fusion protein was used to target transposition, two insertions close to the genomic TetR binding site were identified in about 400 colonies containing genomic insertions. When the *SB*-based mini-transposon did not contain a LexA binding site, no insertions close

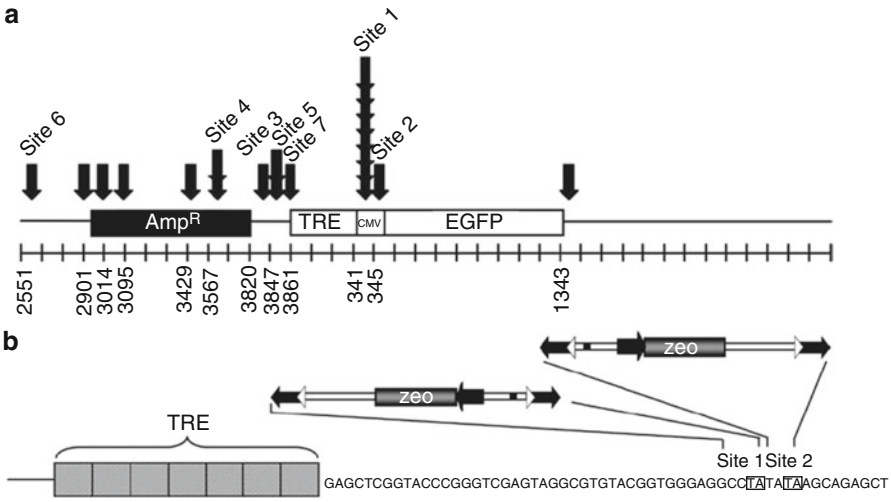


Fig. 9.5 Indirect site-directed integration of transgenes by targeting transposase (**a**) or transposon (**b**) to a chosen target genomic site in HeLa cells. (**a**) Mapping of targeted SB insertions (*arrows*) with respect to the TRE-EGFP target isolated from six independent experiments is shown. *Multiple arrows* represent independent insertions into the same site. Positions of the insertions are indicated below; the numbers correspond to the base pair numbering of the pTRE-d2EGFP plasmid (Clontech) (Ivics et al. 2007) (with permission from Mol. Therapy).(**b**) Targeted transposition close to the TRE. Two transposon insertions are in close proximity of the targeted TRE region, in the two possible orientations, within two TA dinucleotides of the CMV promoter TATA-box (Sites 1 and 2) (Ivics et al. 2007) (with permission from *Mol Ther*)

to the TRE were identified in a similar number of colonies (Fig. 9.5b). However, more colonies would need to be tested to check the significance of this difference.

The number of different insertions present in each colony is not reported in these experiments so it is not possible to estimate the fraction of actual insertions that are targeted using these two strategies. As argued above for targeted *piggyBac* insertions, the fraction of colonies containing insertions close to the TRE that can be detected by PCR is probably a large overestimate of the fraction of individual transposon insertions that are targeted. Nevertheless, strategies involving the use of a DBD fused to either a transposase-binding or transposon-binding domain to target insertions near a selected locus are promising, because (1) they do not measurably interfere with the transposition process; and (2) they could easily be developed for any different transposase.

9.6 Conclusion and Discussion

Transposons have proved their usefulness for transgenesis and mutagenesis in a large number of different systems, including for example Tn5 and Himar in bacteria, and SB in mammalian cells. Transposons insert with only limited sequence specificity, leading

to insertion at many different chromosomal sites. This property is advantageous when transposons are being used for random mutagenesis, but has several drawbacks for their use to stably integrate transgenes. Insertions in some genomic regions can lead to poor transgene expression, while insertion close to or within important genes can lead to genotoxic effects. An important challenge for the use of transposons as gene delivery vectors is therefore to modify them so that they can be targeted to chosen genomic sites without reducing their efficiency. One approach to this has been to modify the transposase proteins by fusing them to an additional DNA binding domain so that it binds specifically to new target sequences. However, this has proved problematic because the fusion proteins often have greatly reduced transposition activity. This has been partly overcome by elegant strategies developed in the Ivics/ Izsvák laboratory, escaping the need to modify transposases.

Fusing a specific DBD to transposase to obtain targeted integration has many similarities to the approach taken with zinc finger nucleases (reviewed in Urnov et al. 2010) and zinc finger recombinases (reviewed in Akopian and Stark 2005). The biggest difference between these two systems and chimaeric zinc finger (or other DBD) transposases, is that the natural DBD of the nuclease or recombinase is removed completely and replaced with a new DBD to change the DNA binding specificity of the protein. In contrast, a zinc finger transposase must retain the ability to recognize the transposon IRs and must gain an additional specificity for its desired target. Thus the IR specific DBD has to be retained in the chimaeric transposase and the additional target specific DBD must not interfere with IR binding or catalysis of strand cleavage and strand transfer.

Targeted transposition with transposase-DBD fusions has been most clearly demonstrated with the prokaryotic transposases IS30 and ISY100 in bacteria. Because of their efficiency, and because they can be studied both *in vitro* and using powerful bacterial genetic techniques, IS30 and ISY100 modified transposases are promising tools for future development. IS30 has been shown to transpose in eukaryotes (HeLa cells and zebrafish) (Szabo et al. 2003), whereas ISY100 has yet to be tested in a eukaryotic system. ISY100 belongs to the Tc1/mariner family of transposons, members of which have been shown to be active in multiple species. For example SB from fish functions well in mammalian cells, while Mos1 from insects function in the nematode *C. elegans* (Ivics et al. 1997; Bessereau et al. 2001). ISY100 functions in bacteria without the need for any host specific co-factors and the purified protein can carry out a full transposition reaction *in vitro*, so there is no *a priori* reason to believe that it will not function in mammalian or other eukaryotic systems. However, this needs to be tested, as for unknown reasons some transposons (e.g. *piggyBac* and SB) function well while others (e.g. Mos1) function poorly in mammalian cells.

One of the major problems encountered with the development of targeted chimaeric transposases is their loss of activity upon fusion to a DBD. C-terminal fusions are in general more severely affected than N-terminal fusions. This is not specific for DBDs, as fusion of even a small epitope tag to the C-terminus of SB or Mos1 transposases results in loss of activity. In contrast, fusions to the N-terminus of SB, Tol2 and Mos1 retained 10–20% of their transposition activity, while N-terminal fusions

to *piggyBac* transposase retained full activity. SB and Mos1 transposase are both members of the Tc1/mariner family of transposases and their sequences can be aligned convincingly along their full length. The crystal structure of Mos1 transposase has recently been solved bound in an active dimeric complex containing two transposon IR sequences (Richardson et al. 2009). Examination of this structure shows that the C-terminus is close to a region important for IR binding and dimerisation of the catalytic domain, explaining why addition of sequences to the C-terminus results in loss of activity. It is somewhat surprisingly at first sight therefore, that ISY100 transposase from this same family tolerates fusions to its C-terminus so well, retaining 18–31% of wild-type activity (Feng et al. 2010). However, ISY100 transposase lacks the most C-terminal ~32 amino acids present in SB and Mos1, placing its C-terminus and any peptide fused to it at a less critical position in the structure. ISY100 transposase fused to Zif268 gave insertions in a very narrow window 7–17 bp from the Zif268 binding site, as expected for tethering of the C-terminal catalytic domain at the target site. In contrast, N-terminal fusions to *piggyBac*, SB and Mos1 gave targeted insertions in a much wider window from 100 to 1,000 bp from the DBD binding site. The N-terminus of Mos1 transposase (and probably most other transposases) is distant from the target binding site in the catalytic domain, and the target DNA would have to form a large loop to bind simultaneously to the DBD and the catalytic domain of the transposase. This DNA loop could be of variable size, explaining the large targeting window. Given the crystal structure of Mos1 transposase, it might now be possible to design SB or Mos1 transposases that contain DBDs inserted at non critical positions within the catalytic domain. These could potentially retain full activity and carry out precise targeted integration close to the DBD binding site.

Even for those transposases where fusions have retained activity, the level of targeted transposition has been somewhat disappointing. In inter-plasmid transposition assays, where genetic techniques are used to isolate only those insertions that occurred in the desired target plasmid, the number of insertions that are close to the DBD binding site is typically increased by only 10–20 fold. Results for genomic targeted integration also seem broadly consistent with an increase of targeted transposition of this order of magnitude. The vast majority of insertions are at non-specific sites, especially when attempts are made to target a site that is present only once in the genome. This might be improved by increasing the binding affinity or specificity of the DBDs used, but the real problem appears to be that the fusion proteins retain the ability to bind tightly to target sequences using their non-specific catalytic domains. One approach to improving targeting by chimaeric transposases might therefore be to ablate the target binding activity of the catalytic domain so that the enzyme becomes dependent on its fused DBD for target binding. It should be possible to obtain mutants with reduced target binding activity, and these could be tested to see if the ratio of on-target to off-target transposition is improved.

Another possibility that might improve targeting and the overall efficiency of gene delivery using transposase is to use purified transposase protein to pre-form an excised transposon-transposase complex that is ready to find a target site and catalyse integration. Such a strategy has proved useful for highly efficient Tn5 delivery to

bacterial cells (Goryshin et al. 2000), though it does rely on the availability of purified protein. One problem for the use of this strategy in non-bacterial systems might be delivering "transpososome" complexes to the nucleus of eukaryotic cells in an active form. However, Tn5 transpososomes have successfully been used to integrate foreign DNA in fungi and trypanosomes (Goryshin et al. 2000; Feng et al. 2008 ; Shi et al. 2002)

One challenge for every targeting system is to choose a suitable genomic site for integration. Many of the benefits of targeted integration are canceled out if the target site is not naturally present and has to be integrated into the genome before site-directed transgenesis can be performed. At this point in time, the most promising DBDs for targeting any chosen site in a complex genome are zinc finger DBDs (reviewed in Klug 2010), and more recently, transcription activator-like effector DBDs (TALEs) (Cermak et al. 2011). Great progress has been made in the production of DBDs of both these families that bind tightly and specifically to chosen DNA sequences. Although patents might limit their use, these two families of DBD remain the most promising for targeting transposon insertions whatever strategy is used.

Directing insertion to specific target sites using a transposase-binding protein has proven a successful survival strategy in nature for retroviral integrases (Bushman 2003). HIV is targeted by LEDGF/p75 interactions (Ciuffi et al. 2005), while Ty5 interacts with Sir4p (Zhu et al. 2003). This same strategy has been exploited for SB using an N-terminal deletion derivative of transposase (N57) that binds to full-length transposase without reducing the efficiency of transposition (Ivics et al. 2007). The fusion of a DBD (TetR) to N57 SB gave detectable targeting of a transgene close to the TetR target site (TRE) in about 10% of transfected colonies. The TRE used contained seven repeats of the binding site for TetR, and it would be interesting to see if targeting could be obtained with fewer copies of this sequence. It would also be worth investigating whether a zinc finger or TALE DBD fused to N57 could be used to target integration to a specific single-copy sequence in a eukaryotic genome. Generalisation of this strategy to other transposons, such as *piggyBac*, will depend on the availability of suitable transposase binding proteins.

In choosing suitable integration sites for long term expression of transgenes, there are several constraints. The integrated gene must be correctly regulated and expressed long-term, and integration must not cause any deleterious effects for the cells (Palazzoli et al. 2008). One strategy that might comply with these constraints, and help to overcome the difficulties associated with integration into a single copy genomic site, would be to use a non-heterochromatic repeated sequence. Several repeat sequences might be suitable. For instance, some retro-transposons are known to integrate into tRNA genes without disturbing cell development and allowing efficient transgene expression (Aye et al. 2004). Alternatively, successful integration and expression of transgenes has been obtained by recombination into rDNA sequences, and these might also constitute good sites for targeted integrations (Liu et al. 2007).

Finally, what is the minimal rate of targeted integration to be useful, and has any one succeeded in obtaining it. At the moment, it seems that no one has convincingly demonstrated that targeted transposition can be used to isolate a clonal population of cells containing a single insertion at a single genomic site, and this remains the

ultimate goal for gene therapy applications. However, there are many avenues available to improve the efficiency of targeted transposition and this remains a promising area for further research.

References

Akopian A, Stark W (2005) Site-specific DNA recombinases as instruments for genomic surgery. Adv Genet 55:1–23

Augé-Gouillou C, Hamelin MH, Demattéi MV, Périquet M, Bigot Y (2001) The ITR binding domain of the *Mariner Mos1* transposase. Mol Genet Genomics 265:58–65

Aye M, Irwin B, Beliakova-Bethell N, Chen E, Garrus J, Sandmeyer S (2004) Host factors that affect Ty3 retrotransposition in *Saccharomyces cerevisiae*. Genetics 168:1159–1176

Bessereau JL, Wright A, Williams DC, Schuske K, Davis MW, Jorgensen EM (2001) Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line. Nature 413:70–74

Bushman FD (1994) Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc Natl Acad Sci USA 91:9233–9237

Bushman FD (2003) Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. Cell 115:135–138

Carapuça E, Azzoni AR, Prazeres DM, Monteiro GA, Mergulhão FJ (2007) Time-course determination of plasmid content in eukaryotic and prokaryotic cells using real-time PCR. Mol Biotechnol 37:120–126

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39:e82

Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR, Bushman F (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11:1287–1289

Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD (2006) Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor. Hum Gene Ther 17:960–967

Copeland NG, Jenkins NA (2010) Harnessing transposons for cancer gene discovery. Nat Rev Cancer 10:696–706

Crénès G, Ivo D, Hérisson J, Dion S, Renault S, Bigot Y, Petit A (2009) The bacterial Tn9 chloramphenicol resistance gene: an attractive DNA segment for *Mos*1 mariner insertions. Mol Genet Genomics 281:315–328

Demattéi MV, Thomas X, Carnus E, Augé-Gouillou C, Renault S (2010) Site-directed integration of transgenes: transposons revisited using DNA-binding-domain technologies. Genetica 138:531–540

Feng M, Zhou T, Wei W, Song Y, Tan R (2008) Tn5 transposase-assisted high-efficiency transformation of filamentous fungus Phoma herbarum YS4108. Appl Microbiol Biotechnol 80:937–944

Feng X, Bednarz AL, Colloms SD (2010) Precise targeted integration by a chimaeric transposase zinc-finger fusion protein. Nucleic Acids Res 38:1204–1216

Fraser MJ, Cary L, Boonvisudhi K, Wang HG (1995) Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. Virology 211:397–407

Goryshin IY, Reznikoff WS (1998) Tn5 in vitro transposition. J Biol Chem 273:7367–7374

Goryshin IY, Jendrisak J, Hoffman LM, Meis R, Reznikoff WS (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat Biotechnol 18:97–100

Goulaouic H, Chow SA (1996) Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli LexA protein. J Virol 70:37–46

Granger L, Martin E, Ségalat L (2004) *Mos* as a tool for genome-wide insertional mutagenesis in *Caenorhabditis elegans*: results of a pilot study. Nucleic Acids Res 32:e117

Guimond N, Bideshi DK, Pinkerton AC, Atkinson PW, O'Brochta DA (2003) Patterns of *Hermes* transposition in *Drosophila melanogaster*. Mol Genet Genomics 268:779–790

Hama C, Ali Z, Kornberg TB (1990) Region-specific recombination and expression are directed by portions of the *Drosophila engrailed* promoter. Genes Dev 4:1079–1093

Imre A, Olasz F, Nagy B (2011) Site-directed (IS30-FljA) transposon mutagenesis system to produce nonflagellated mutants of *Salmonella Enteritidis*. FEMS Microbiol Lett 317:52–59

Ivics Z, Hackett PB, Plasterk RH, Izsvák Z (1997) Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon from fish, and its transposition in human cells. Cell 14:501–510

Ivics Z, Katzer A, Stüwe EE, Fiedler D, Knespel S, Izsvák Z (2007) Targeted *Sleeping Beauty* transposition in human cells. Mol Ther 15:1137–1144

Izsvák Z, Khare D, Behlke J, Heinemann U, Plasterk RH, Ivics Z (2002) Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition. J Biol Chem 277:34581–34588

Izsvák Z, Hackett PB, Cooper LJ, Ivics Z (2010) Translating Sleeping Beauty transposition into cellular therapies: victories and challenges. Bioessays 3:756–767

Kassis JA, Noll E, VanSickle EP, Odenwald WF, Perrimon N (1992) Altering the insertional specificity of a *Drosophila* transposable element. Proc Natl Acad Sci USA 89:1919–1923

Katz RA, Merkel G, Skalka AM (1996) Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles. Virology 217:178–190

Keng VW, Yae K, Hayakawa T, Mizuno S, Uno Y, Yusa K, Kokubu C, Kinoshita T, Akagi K, Jenkins NA, Copeland NG, Horie K, Takeda J (2005) Region-specific saturation germline mutagenesis in mice using the *Sleeping Beauty* transposon system. Nat Methods 2:763–769

Kettlun C, Galvan DL, George AL Jr, Kaja A, Wilson MH (2011) Manipulating *piggyBac* transposon chromosomal integration site selection in human cells. Mol Ther 19:1636–1644

Klug A (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. Q Rev Biophys 43:1–21

Kondrychyn I, Garcia-Lecea M, Emelyanov A, Parinov S, Korzh V (2009) Genome-wide analysis of Tol2 transposon reintegration in zebrafish. BMC Genomics 10:418–431

Lidholm D-A, Lohe AR, DL DL (1993) The transposable element *mariner* mediates germline transformation in *Drosophila melanogaster*. Genetics 134:859–868

Liu X, Liu M, Xue Z, Pan Q, Wu L, Long Z, Xia K, Liang D, Xia J (2007) Non-viral ex vivo transduction of human hepatocyte cells to express factor VIII using a human ribosomal DNA-targeting vector. J Thromb Haemost 5:347–351

Maragathavally KJ, Kaminski JM, Coates CJ (2006) Chimaeric *Mos1* and *piggyBac* transposases result in site-directed integration. FASEB J 20:1880–1882

McNamara AR, Ford KG (2000) A novel four zinc-finger protein targeted against p190(BcrAbl) fusion oncogene cDNA: utilisation of zinc-finger recognition codes. Nucleic Acids Res 28:4865–4872

O'Brochta DA, Handler AM (2008) Perspectives on the state of insect transgenics. Adv Exp Med Biol 627:1–18

Paatero AO, Turakainen H, Happonen LJ, Olsson C, Palomaki T, Pajunen MI, Meng X, Otonkoski T, Tuuri T, Berry C, Malani N, Frilander MJ, Bushman FD, Savilahti H (2008) Bacteriophage Mu integration in yeast and mammalian genomes. Nucleic Acids Res 36:e148

Palazzoli F, Carnus E, Wells DJ, Bigot Y (2008) Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. Curr Gene Ther 8:367–390

Parker CT, Guard-Petter J (2001) Contribution of flagella and invasion proteins to pathogenesis of Salmonella enterica serovar enteritidis in chicks. FEMS Microbiol Lett 204:287–291

Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. Nat Biotechnol 23:967–973

Préclin V, Martin E, Segalat L (2003) Target sequences of Tc1, Tc3 and Tc5 transposons of Caenorhabditis elegans. Genet Res 82:85–88

Reznikoff WS (2006) Tn5 transposition: a molecular tool for studying protein structure-function. Biochem Soc Trans 34:320–323
Richardson JM, Colloms SD, Finnegan DJ, Walkinshaw MD (2009) Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. Cell 138:1096–1108
Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, Mekalanos JJ (1999) *In vivo* transposition of mariner-based elements in enteric bacteria and mycobacteria. Proc Natl Acad Sci USA 96:1645–1650
Ryder E, Russell S (2003) Transposable elements as tools for genomics and genetics in Drosophila. Brief Funct Genomic Proteomic 2:57–71
Sassetti CM, Boyd DH, Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci USA 98:12712–12717
Schweidenback CT, Baker TA (2008) Dissecting the roles of MuB in Mu transposition: ATP regulation of DNA binding is not essential for target delivery. Proc Natl Acad Sci USA 105:12101–12107
Shi H, Wormsley S, Tschudi C, Ullu E (2002) Efficient transposition of preformed synaptic Tn5 complexes in Trypanosoma brucei. Mol Biochem Parasitol 121:141–144
Suster ML, Sumiyama K, Kawakami K (2009) Transposon-mediated BAC transgenesis in zebrafish and mice. BMC Genomics 10:477
Szabó M, Müller F, Kiss J, Balduf C, Strähle U, Olasz F (2003) Transposition and targeting of the prokaryotic mobile element IS30 in zebrafish. FEBS Lett 550:46–50
Tan W, Zhu K, Segal DJ, Barbas CF 3rd, Chow SA (2004) Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. J Virol 78:1301–1313
Tan W, Dong Z, Wilkinson TA, Barbas CF 3rd, Chow SA (2006) Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. J Virol 80:1939–1948
Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646
van Luenen HG, Plasterk RH (1994) Target site choice of the related transposable elements Tc1 and Tc3 of Caenorhabditis elegans. Nucleic Acids Res 22:262–269
VandenDriessche T, Ivics Z, Izsvák Z, Chuah MK (2009) Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. Blood 114:1461–1468
Vigdal TJ, Kaufman CD, Izsvák Z, Voytas DF, Ivics Z (2002) Common physical properties of DNA affecting target site selection of *Sleeping beauty* and other *Tc1/mariner* transposable elements. J Mol Biol 323:441–452
Walisko O, Schorn A, Rolfs F, Devaraj A, Miskey C, Izsvák Z, Ivics Z (2006) Transcriptional activities of the *Sleeping Beauty* transposon and shielding its genetic cargo with insulators. Mol Ther 16:359–369
Wang N, Jiang CY, Jiang MX, Zhang CX, Cheng JA (2010) Using chimaeric *piggyBac* transposase to achieve directed interplasmid transposition in silkworm Bombyx mori and fruit fly Drosophila cells. J Zhejiang Univ Sci B 11:728–734
Wilson MH, Kaminski JM, George AL Jr (2005) Functional zinc finger/sleeping beauty transposase chimaeras exhibit attenuated overproduction inhibition. FEBS Lett 579:6205–6209
Wong SM, Gawronski JD, Lapointe D, Akerley BJ (2011) High-throughput insertion tracking by deep sequencing for the analysis of bacterial pathogens. Methods Mol Biol 733:209–222
Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM (2006) *piggyBac* is a flexible and highly active transposon as compared to *Sleeping beauty*, *Tol2*, and *Mos1* in mammalian cells. Proc Natl Acad Sci USA 103:15008–15013
Yant SR, Huang Y, Akache B, Kay MA (2007) Site-directed transposon integration in human cells. Nucleic Acids Res 35:1–13
Zhu Y, Dai J, Fuerst PG, Voytas DF (2003) Controlling integration specificity of a yeast retrotransposon. Proc Natl Acad Sci USA 100:5891–5895

Chapter 10
Targeted Plasmid Integration into the Human Genome by Engineered Recombinases

Charles A. Gersbach and Carlos F. Barbas III

Abstract The targeted integration of transgenes into cellular genomes is central to numerous applications in biotechnology, basic science, and medicine. In recent years, a variety of advances have improved upon conventional methods for site-specific transgene integration. Most of these methods involve nucleases that cleave DNA to activate DNA repair pathways including homologous recombination or integrases that fully catalyze the integration reaction but are limited in their capacity to target new sites in the genome. Recently, zinc-finger recombinases have emerged as a class of engineered enzymes that combines the strengths of both of these previous methods. Zinc-finger recombinases can fully and autonomously catalyze plasmid integration into the genome of mammalian cells without creating free DNA breaks. In addition, they can be engineered to target new genomic recognition sites by exchanging the modular and programmable DNA-binding domain and through directed evolution of the serine recombinase catalytic domain. This chapter reviews the development of the zinc-finger recombinase technology, including discussions of its strengths and weaknesses and the future directions necessary to translate this technology into routine use for transgene integration into cellular genomes.

Keywords Genome engineering • Protein engineering • Synthetic biology • Gene targeting

C.A. Gersbach (✉)
Department of Biomedical Engineering, Duke University,
Box 90281, 136 Hudson Hall, Durham, NC 27708, USA
e-mail: charles.gersbach@duke.edu

C.F. Barbas III
The Skaggs Institute for Chemical Biology and the Departments of Molecular Biology and Chemistry, The Scripps Research Institute,
10550 North Torrey Pines Rd, La Jolla, CA 92037, USA
e-mail: carlos@scripps.edu

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_10,

10.1 Introduction

The addition of new gene sequences to mammalian cells is central to numerous applications in biotechnology, basic science, and medicine. For many applications, it is necessary that these new gene sequences be stably expressed for extended periods of time. This typically requires integration of the new genes into chromosomal DNA. The permanent incorporation of new genes into the genome ensures that they will be copied with each cell cycle along with the native chromosomal DNA to prevent dilution over many cell divisions. Traditionally, the field of genetic engineering has been severely limited in the diversity of tools available for integrating genes into cellular genomes. As discussed throughout this book, there has recently been a rapid expansion in technologies for adding genes to genomes that address many of the limitations of conventional methods.

Mammalian genomes are large and complex, and gene expression is highly regulated. The haploid human genome contains over three billion base pairs and encodes tens of thousands of genes and many more regulatory RNAs (Birney et al. 2007; Lander et al. 2001; Venter et al. 2001). In any particular cell type, some of these genes will be highly expressed, while others will be tightly repressed. The vast majority of the human genome is not actively transcribed, and roughly 98% of the genome is incorporated into heterochromatin, in which the DNA strand is tightly wound with histone proteins into nucleosomes and made inaccessible to DNA-binding proteins and gene delivery vehicles (Boyle et al. 2008). Consequently, there are relatively few locations in mammalian genomes that are suitable sites for hosting a transgene that must be robustly expressed (Papapetrou et al. 2011). Therefore it is highly desirable to be able to add genes to specific sites within the genomic DNA. Site-specific incorporation helps to ensure (1) that the new gene sequence will function predictably and reproducibly and (2) that the addition of the new gene sequence will not have any unintended effects on the host cell.

Applications that could benefit from the targeted integration of genes into locations within the genome are numerous and diverse. For example, gene addition to mammalian cells is commonly used for studies related to functional genomics, cell biology, proteomics, cell-based drug discovery, and many other applications in biotechnology and basic science. Consistent expression of the transgene is essential to ensure reproducible results and accurate data interpretation. Biopharmaceuticals, such as monoclonal antibodies, cytokines, and growth factors, are often industrially produced by mammalian cells at large scales (Wurm 2004). To achieve the most cost-effective production of these drugs it is necessary that the desired protein is produced at optimal levels. Precise and reproducible integration of the transgene ensures that each cell produces protein at uniformly high levels. Applications such as synthetic biology and metabolic engineering, in which many transgenes encoding a complete gene circuit or enzymatic pathway must be introduced into cells and finely balanced to optimize pathway flux, are even more sensitive to levels of gene expression (Purnick and Weiss 2009).

The field of gene therapy has achieved many notable clinical successes in recent years, including the apparent cure of several genetic diseases (Aiuti et al. 2009; Boztug et al. 2010; Cartier et al. 2009; Cavazzana-Calvo et al. 2010; Hacein-Bey-Abina et al. 2010; Ott et al. 2006) and positive results in clinical trials for the treatment of HIV (Mitsuyasu et al. 2009) and rheumatoid arthritis (Wehling et al. 2009). Importantly, all of these clinical trials used retroviral or lentiviral vectors to permanently integrate genes into random locations in the chromosomes of patients' cells. Unfortunately, these successes have been accompanied by adverse side effects related to the uncontrolled integration of foreign DNA into the genome. In particular, the integration of retroviral DNA near an endogenous protooncogene led to leukemia in 5 of the 20 patients treated for SCID-X1 (Fischer et al. 2010). Similar clonal expansions have been observed in other stem cell-based gene therapy clinical trials using retrovirus or lentivirus vectors (Cavazzana-Calvo et al. 2010; Ott et al. 2006). The cancerous side effects of viral gene therapy have been attributed to the preferential integration of vectors into promoter or intragenic regions of actively transcribed genes, which are subsequently overexpressed or deregulated in response to the strong viral promoters (Bushman et al. 2005; Fischer et al. 2010; Wu et al. 2003). This phenomenon, known as insertional mutagenesis, has lead to the widespread investigation of alternative methods to integrate therapeutic transgenes into the genome.

10.2 Nonviral Methods for Transgene Integration

10.2.1 Random Integration by Transposases

To address the semi-random integration of genes by viral vectors, enzymes that cut and paste DNA, including transposases, integrases, and recombinases, have been developed to insert gene sequences from transfected plasmid DNA into the genome. The *piggyBac* and *Sleeping Beauty* transposon systems, isolated from the moth and fish genomes, respectively (Ding et al. 2005; Ivics et al. 1997), have both been used to integrate genes into mammalian genomes *in vitro* and *in vivo* (Mates et al. 2009; VandenDriessche et al. 2009; Wilson et al. 2007). These enzymes are routinely used for genetic engineering of model organisms. A clinical trial is being planned with the *Sleeping Beauty* system to genetically modify patient T cells for cancer immunotherapy (Hackett et al. 2010). Because the transposon systems do not contain strong viral promoters, it is anticipated that activation of nearby oncogenes will be unlikely. Additionally, *Sleeping Beauty*, and to a lesser extent *piggyBac,* appear to integrate more randomly into the genome than γ-retroviral vectors and do not show as strong a preference for integration into genes (Vigdal et al. 2002; Wilson et al. 2007; Yant et al. 2005). However, transposition into an oncogene or tumor suppressor and subsequent insertional mutagenesis has been demonstrated in genetic screens that are designed to favor these events (Collier et al. 2005; Rad et al. 2010).

Although a few studies have attempted to direct gene transposition by these enzymes to genomic target sites by the fusion of sequence-specific DNA-binding proteins, the majority of gene integration events in these systems are random (Ivics et al. 2007; Yant et al. 2007).

10.2.2 Site-Directed Integration by Phage Integrases

The phage-derived integrase phiC31 directs the integration of a transfected plasmid into pseudo-recognition sites in the human genome (Thyagarajan et al. 2001). This approach is widely used for transgene addition in basic research and has been successfully applied in preclinical studies of gene therapies for a variety of diseases (Calos 2006). A thorough characterization of the pseudo-recognition sites in human cells identified over 100 distinct genomic integration sites with a slight preference for integration into intragenic regions (Chalberg et al. 2006). Therefore this enzyme does not target a specific site in the genome and the potential for insertional mutagenesis remains. Additionally, chromosomal rearrangements in human cells following phiC31 expression occur relatively frequently (Ehrhardt et al. 2006; Liu et al. 2006). Although recent *in vivo* safety studies indicate that phiC31-based gene therapies should be pursued (Woodard et al. 2010), there remains a clear need for enzymes with strict DNA-binding domains that recognize unique sites within the human genome.

10.2.3 Targeted Integration by Homologous Recombination and Zinc Finger Nucleases

Homologous recombination is a mechanism by which precise changes to defined genomic sequences can be introduced (Capecchi 2005; Koller and Smithies 1992). Gene targeting by homologous recombination is highly efficient and commonly used in mouse embryonic stem cells to generate transgenic mice (Capecchi 2005; Koller and Smithies 1992). Unfortunately, conventional homologous recombination methods are not readily transferable to gene targeting in other cell types from the mouse or in cells from other species. Two discoveries have recently made high rates of homologous recombination in mammalian somatic cells possible. The first is the finding that the introduction of a double-strand break at a genomic target site increases the rate of homologous recombination at that site by several orders of magnitude (Rouet et al. 1994). The second is the development of engineered zinc-finger nucleases (ZFNs) that generate site-specific double-strand breaks at targeted genome sequences (Bibikova et al. 2003, 2001; Porteus and Baltimore 2003; Smith et al. 2000).

The Cys2-His2 zinc finger domain is the most common DNA-binding motif in the human proteome (Tupler et al. 2001). It has a ββα configuration where the α-helix projects into the major groove of DNA (Pavletich and Pabo 1991). A single

zinc finger contains approximately 30 amino acids and the domain typically functions by binding three consecutive base pairs of DNA via interactions of a single amino acid side chain per base pair (Elrod-Erickson et al. 1996; Pavletich and Pabo 1991). Early structure-function studies of zinc finger proteins have aided in defining residues that participate in DNA recognition (Desjarlais and Berg 1992; Nardelli et al. 1992). Later studies involving phage display selections of combinatorial libraries identified synthetic zinc finger domains with defined DNA target site specificity (Choo and Klug 1994; Greisman and Pabo 1997; Jamieson et al. 1994; Maeder et al. 2008; Rebar and Pabo 1994; Wu et al. 1995). Through this work, synthetic zinc finger domains have been established that bind to almost all of the 64 possible nucleotide triplets (Dreier et al. 2001, 2005, 2000; Segal et al. 1999). Significantly, the modular structure of the zinc finger motif permits the conjunction of several domains in series, allowing for the recognition and targeting of extended sequences in multiples of three nucleotides (Segal et al. 2003). It is now possible to design synthetic zinc finger proteins to bind practically any target site in the human genome wherein specificity is provided using proteins that bind up to 18 bp of DNA sequence (Maeder et al. 2008; Mandell and Barbas 2006; Sander et al. 2011).

This approach was first applied to specifically regulate the endogenous human genes erbB2 and erbB3 with polydactyl zinc finger domains fused to transcriptional activation domains or repression domains in order to stimulate or repress transcription of the target gene, respectively (Beerli et al. 2000, 1998). Artificial zinc finger transcription factors have since been applied to many genes and in a variety of organisms (Blancafort et al. 2004). Although this approach facilitates gene-specific regulation and provided the first demonstration of protein targeting to specific endogenous genomic sequences, it suffers from the same limitations of gene addition and gene delivery described above. However, successful gene-specific regulation by zinc finger transcription factors provided the critical evidence of genome-wide specificity and targeting that encouraged the adaptation of this technology for solutions to gene targeting.

Towards this end, synthetic ZFNs have emerged as powerful new tools for genome engineering. ZFNs consist of zinc finger domains fused to the catalytic domain of the FokI restriction endonuclease (Urnov et al. 2010). When expressed within mammalian cells, these ZFNs cleave DNA to create a double-strand break at a targeted genomic locus (Porteus and Baltimore 2003; Urnov et al. 2005, 2010). This DNA cleavage stimulates DNA repair pathways and increases the efficiency of homologous recombination at the site by several orders of magnitude, which otherwise occurs below background levels of random plasmid integration in human cells. This method has been used to incorporate gene sequences at specific locations in the genomes of human cell lines and embryonic and adult stem cells (Urnov et al. 2010). However, the potential for off-target DNA cleavage, the induction of the DNA-damage response pathway, and the associated genotoxicity that has been routinely observed with these enzymes remain concerns as this technology progresses in clinical studies (Cornu et al. 2008; Gabriel et al. 2011; Pattanayak et al. 2011; Urnov et al. 2010).

10.3 Transgene Integration by Site-Specific Recombinases

10.3.1 Mechanism of Site-Specific Recombination

The process of site-specific DNA recombination between two recombination sites occurs through a series of distinct molecular events (Fig. 10.1). The complete recombination event involves a tetrameric complex of four recombinase molecules. First, the recombinases bind as dimers to each target recognition sequence in the recombination sites. The dimers then bind to form the tetrameric synaptic complex. The recombinase enzymes then cleave each strand of DNA, rotate 180°, and ligate the new partner pairs. Finally, the recombinases release the DNA and recombination is complete.

This mechanism of site-specific recombination has several important implications. First, the outcome of the recombination reaction is dependent on the initial relationship of the recombination sites (Fig. 10.2). If one site is on an episomal plasmid and the other site is in the genomic DNA, then the reaction results in the site-specific integration of the complete plasmid at that site. This approach is most relevant to the site-specific integration of transgenes. Importantly, the plasmid can integrate in either the sense or the antisense orientation. Alternatively, if the recombination sites are present on the same strand of DNA, then the reaction can result in either the excision of the intervening segment (the reverse of the integration reaction, also referred to as the resolution reaction) or the inversion of this segment. Consequently, recombination can be used to direct a variety of genomic rearrangements. A second implication of this mechanism is that the reaction is autonomously performed by the recombinase enzymes: There is no requirement for DNA synthesis or high-energy cofactors. Although cofactors may be incorporated into the synaptic complex in natural systems for structural or regulatory functions, they do not play a direct role in catalysis. The cellular DNA repair machinery is also not necessary for genomic recombination. This is in contrast to transgene integration strategies that rely on homologous recombination or engineered nucleases to generate double-strand breaks in the DNA. Finally, the requirement for two recombination sites and four

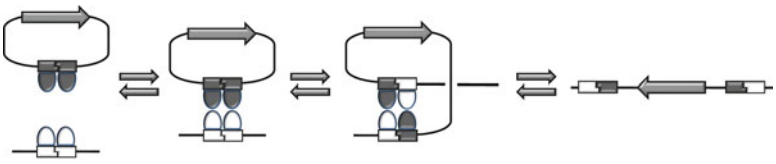


Fig. 10.1 Mechanism of targeted plasmid integration by site-specific recombinases. Site-specific recombinases bind as dimers to both the target site in the episomal plasmid (*shaded*) and the genomic target site (*open*). The dimers interact to form a tetramer; this species cleaves the DNA strands, and covalent bonds are formed between the DNA and the recombinase. A 180° turn by the recombinase results in exchange of the DNA strands, and ligation leads to complete integration of the plasmid DNA and any gene or regulatory sequences it contains (*arrow*)

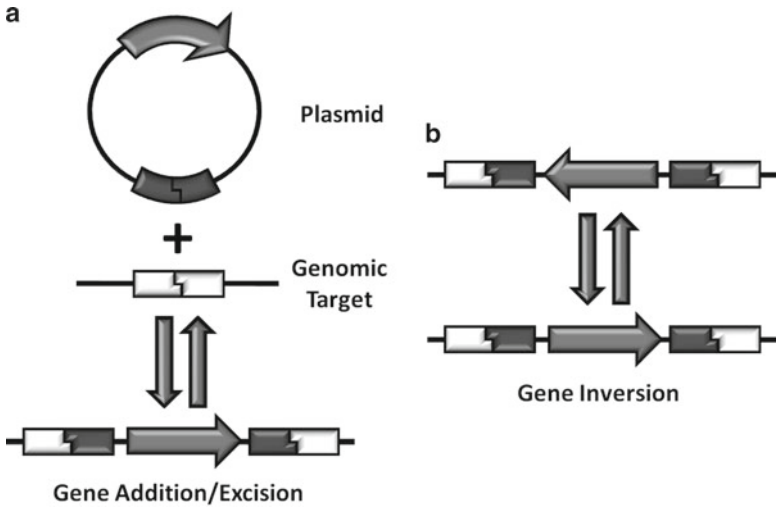


Fig. 10.2 Mechanisms of genomic rearrangement by site-specific recombinases. (**a**) The recombination reaction can direct plasmid integration into the genome as detailed in Fig. 10.1. However, this reaction is also reversible, which leads to gene excision. (**b**) Additionally, the recombination reaction between two sites on the same DNA strand can lead to the inversion of the intervening sequence

recombinase molecules restricts the flexibility of engineered recombination systems (Fig. 10.3). If both recombination sites are identical and each site consists of the recombinase recognition sequence as an inverted repeat, then all four recombinase molecules can be identical (Fig. 10.3a). However, if the recombination sequences are distinct or are not structured as inverted repeats, then two or four unique recombinase enzymes may be necessary to bind to the DNA targets and catalyze the reaction (Fig. 10.3b and 10.3c). The sequence specificity of the recombinase is a primary determinant of the flexibility of target site sequence and structure.

10.3.2 Serine and Tyrosine Recombinases

Almost all known recombinases can be categorized into one of two families based on the amino acid that forms the covalent bond with the DNA strand. Bond formation occurs through a transesterification reaction in which a phosphodiester bond is formed between the sugar-phosphate backbone of the DNA and a side chain of the recombinase. For all characterized recombinases, this side chain is either serine or tyrosine. Interestingly, the serine and the tyrosine families of recombinases differ significantly in structure and function, implying distinct evolutionary paths for proteins that perform very similar functions. The differences in these families also significantly affect how they have been used for integrating transgenes into the genomes of mammalian cells. The tyrosine recombinases, also known as the λ

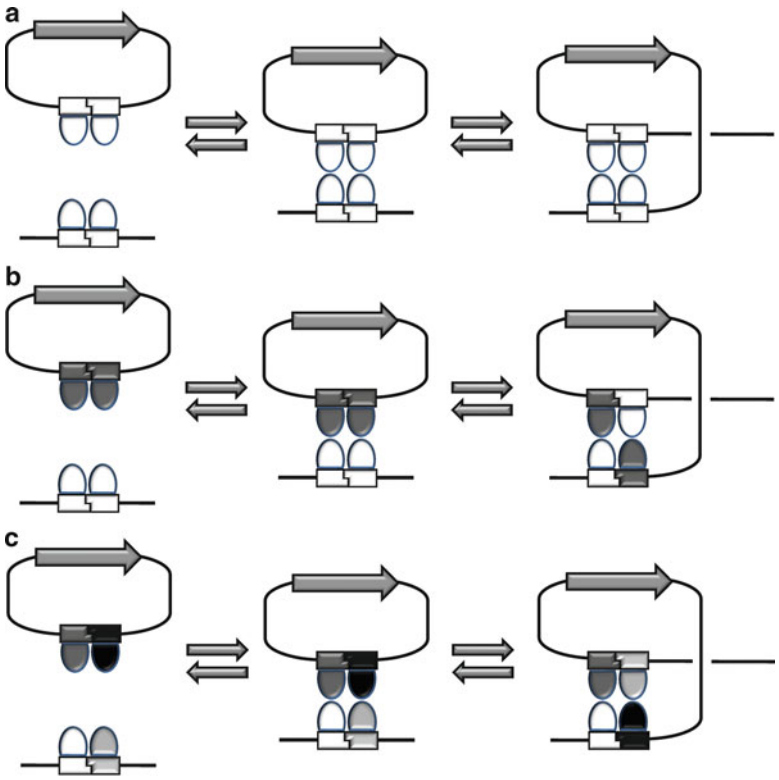


Fig. 10.3 Modes of recombination between various target sites. (**a**) For a symmetrical target site, in which the DNA binding sites for the recombinase are perfect inverted repeats and the target sites the plasmid and in the genome are identical, a single recombinase (*open*) can complete the full recombination reaction. (**b**) If the recombinase target sites on the plasmid and in the genome are different, then distinct recombinases are necessary to recognize each of these targets (*open* and *shaded*). (**c**) If the target sites are asymmetrical then distinct recombinases are necessary to bind each side of the target. Therefore the choice of target sites and design of the plasmid target sequence greatly affects the complexity of the engineered system

integrase family, are best known in the context of mammalian genome engineering by the Cre and Flp recombinases derived from bacteriophage and yeast, respectively. In contrast to the random addition of genes by viral vectors and transposases, Cre and Flp catalyze the exchange of DNA strands between loxP and FRT sequences, respectively (O'Gorman et al. 1991; Sauer and Henderson 1988). Cre and Flp have both been used to target plasmid integration into loxP or FRT sites that have been pre-introduced into mammalian cell genomes (O'Gorman et al. 1991; Sauer and Henderson 1990). The efficiency of Flp-mediated plasmid integration is comparable to random plasmid integration (Logie and Stewart 1995; O'Gorman et al. 1991). Therefore additional selection or screening steps are necessary to ensure site-specific integration. Cre expression can be toxic to mammalian cells (Loonstra et al. 2001;

Naiche and Papaioannou 2007) and can lead to chromosomal rearrangements by reacting with off-target pseudo-loxP sites present in the human genome (Schmidt et al. 2000; Thyagarajan et al. 2000).

In both cases, the strict sequence specificity of these enzymes for their target sequences in their natural genomes in phage or yeast generally limits their uses to cell lines or transgenic organisms that have been previously engineered to contain these sequences. Several studies have attempted to modify the target site specificity of these enzymes through directed evolution (Bolusani et al. 2006; Buchholz et al. 1998; Buchholz and Stewart 2001; Santoro and Schultz 2002; Sarkar et al. 2007; Voziyanov et al. 2003). However, these efforts have been limited by the difficulty in isolating enzyme variants that retain high levels of activity on their new target sequences. For example, one study required 126 rounds of evolution to identity Cre variants with new specificity for a sequence in the long terminal repeats of HIV-1 although the sequence had 50% sequence identity to loxP (Sarkar et al. 2007). The challenge in altering the sequence specificity of tyrosine recombinases can be explained by the observation that many of the residues that determine DNA-binding specificity are also involved in enzyme catalysis (Akopian and Marshall Stark 2005). Consequently it is difficult to alter enzyme specificity without decreasing enzymatic activity.

The serine recombinases are also known as the resolvase family for the natural function performed by the putative members of this family, Tn3 and γδ. The vast majority of structural information for this class of proteins has been obtained through X-ray crystallography studies of the γδ resolvase (Grindley et al. 2006). In contrast to the tyrosine recombinases, the serine recombinases are modular in both structure and function (Fig. 10.4a). The catalytic domain of γδ consists of approximately 100 amino acids at the N-terminus of the protein, followed by a long α-helix and a 10-amino acid linker to a canonical helix-turn-helix DNA-binding domain at the C-terminus (Yang and Steitz 1995). The DNA-binding domain recognizes a short sequence of about 10 base pairs and appears to play no other critical role as it can be replaced by alternative DNA-binding domains, as described below. Consequently, the specificity of the natural enzymes is largely determined by interactions with co-factors and topological constraints (Grindley et al. 2006). Many other recombinases in this family, including Tn3, Gin, Hin, Sin, and Beta have similar size and structure, although other members vary dramatically.

10.3.3 Retargeting Modular Recombinases with New DNA-Binding Domains

Early demonstrations that the sequence specificity of serine recombinases could be redesigned by altering the DNA-binding domain suggested that this class of enzymes may allow for truly modular programming of designer recombinases (Ackroyd et al. 1990; Avila et al. 1990; Schneider et al. 2000). This hypothesis was further supported by the elucidation of the γδ structure, which showed that the DNA-binding domain was loosely coupled to the catalytic domain by a flexible

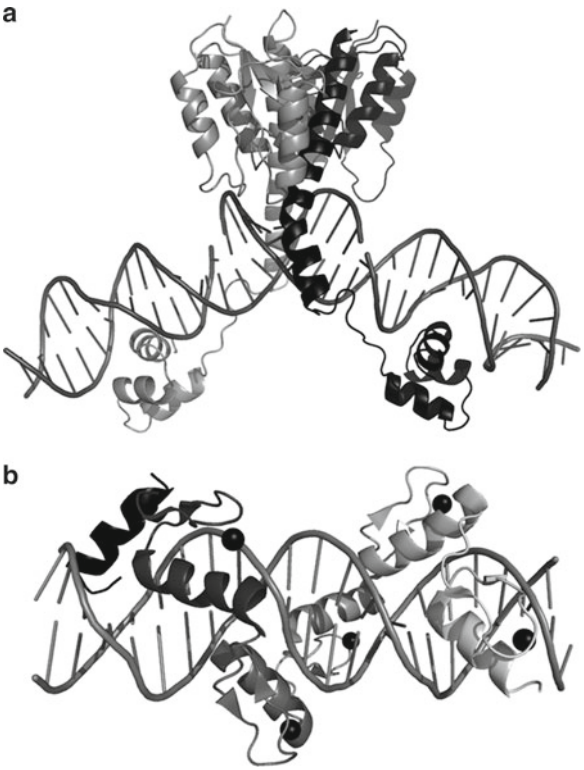


Fig. 10.4 The structure of a characteristic serine recombinase and synthetic zinc finger protein. (**a**) The structure of the dimer of γδ resolvase (Yang and Steitz 1995) (PDB file 1GDT). Monomers are represented in *blue* and *green*. The catalytic domains (*top*, *center*) are connected to the small DNA binding domains (*bottom left* and *right*) by a short flexible linker. (**b**) The structure of the synthetic zinc finger protein Aart (Segal et al. 2006) (PDB file 2I13). The protein contains six zinc fingers. Zinc atoms are shown as spheres

linker (Yang and Steitz 1995) (Fig. 10.4a). Furthermore, recombinase mutants are known that do not require accessory cofactors for recombination activity (Burke et al. 2004; Klippel et al. 1988; Olorunniji et al. 2008). Akopian et al. extended the studies of Schneider et al. 2000 by replacing the DNA-binding domain of the Tn3 resolvase with the zinc finger DNA-recognition domain of the natural mouse transcription factor Zif268 (Akopian et al. 2003). The recombination activity of these fusion proteins was directed to target sites flanked by Zif268 recognition sequences, as demonstrated in a plasmid-based excision assay in *E. coli*. This work suggested, like the work of Schneider et al. before them, the possibility that recombinase activity could be programmed to new target sites with synthetic zinc finger proteins engineered to react with any target site.

This possibility was realized by Gordley et al. in studies that replaced the DNA-binding domains of the Gin, Hin and Tn3 recombinases with engineered zinc finger proteins (Fig. 10.4b) designed to target novel sequences (Gordley et al. 2007). A structural model was used to predict the appropriate length of a flexible linker between the catalytic and DNA-binding domains. When expressed in *E. coli*, these zinc finger recombinases (ZFRs) catalyzed excision and inversion reactions within a plasmid and integration reactions between plasmids. The spacer length of the DNA sequence between zinc finger recognition sequences was also optimized. In order to demonstrate the activity of these enzymes in mammalian cells, a single GFP expression cassette flanked by recombination sites was integrated into the genome of human HEK293 cells. Introduction of the ZFR led to excision of the GFP cassette in 17% of the cells. This study was the first to show the activity of ZFRs in human cells (Gordley et al. 2007).

These analyses revealed a critical characteristic of these proteins: although diverse DNA-binding sequences could be accommodated in the target site by reengineering the zinc finger proteins, the catalytic domain was only reactive with spacer sequences homologous to their natural target site. This is in contrast to the FokI nuclease, which reacts with DNA independently of sequence. In order to address this potential restriction of target site sequences for these engineered recombinases, a directed evolution system was developed to alter the sequence specificity of the catalytic domain, as described below.

10.3.4 Targeted Integration with ZFRs

The successful engineering of ZFRs that catalyze precise recombination reactions in human cells provided a potential nonviral method for integration of transgenes at specific sites in the genome. To explore this approach, a single ZFR target site was integrated into an accessible chromosomal locus in the genome of the human HEK293 cell line (Gordley et al. 2009). A transgene-encoding plasmid that was modified to contain a ZFR target site was co-transfected into these cells with or without a ZFR expression vector. The addition of the ZFR expression vector increased the frequency of plasmid integration into the genome ~10-fold (Gordley et al. 2009). Moreover, two independent assays showed that >95% of these integration events were at the predicted target site in the human genome. To date, this represents the greatest accuracy of autonomous targeted integration achieved in the absence of homologous recombination and native cellular DNA repair pathways.

This approach was later expanded by randomly distributing ZFR target sites throughout the genome of multiple mammalian cell lines with the *piggyBac* transposase (Gersbach et al. 2011). The ZFRs efficiently and specifically integrated plasmids into these target sites in several cell types. Notably, the ZFR was able to catalyze the integration of the donor plasmid into multiple target sites in a single cell. This study further supported the use of the ZFR technology for diverse applications in different cell types and at many chromosomal loci.

10.3.5 Directed Evolution of ZFRs

As discussed above, the engineered zinc finger protein that constitutes the DNA-binding domain of the ZFR (Fig. 10.4b) can be modified to target the ZFR to almost any DNA sequence in a cellular genome. However, the serine recombinase catalytic domain reacts only when spacer sequences have very high sequence homology to the natural target site of the serine recombinase from which it was derived (Gordley et al. 2009, 2007). As a result, ZFRs with these natural catalytic domains cannot readily be retargeted to integrate transgenes into endogenous sequences in mammalian genomes. To overcome this limitation, directed evolution has been used to alter the sequence specificity of the ZFR catalytic domain. Systems for directed evolution of recombinases based on substrate-linked protein evolution (SLiPE) were originally developed to alter the target site specificity of the tyrosine recombinases Flp and Cre (Buchholz et al. 1998; Buchholz and Stewart 2001; Sarkar et al. 2007). In this system, the recombination product also carries the gene sequence of the recombinase that catalyzed the reaction, such that the active recombinase sequence can be recovered by polymerase chain reaction (PCR). Gordley et al. adapted this method to the directed evolution of ZFRs (Gordley et al. 2007). After eight rounds of selection and additional DNA shuffling, ZFR catalytic domains were obtained that dramatically increased activity at target sites with diverse sequences. These evolved domains were able to perform precise recombination reactions in the genomes of mammalian cells.

Despite the significant successes of using SLiPE to evolve new recombinase variants, the dependence of this method on PCR for recovery of active recombinases with novel mutations presents several challenges. First, PCR amplification of a DNA target has the potential to incorporate mutations. For directed evolution studies, these mutations may counteract or obscure the desired mutations that led to selection from a mutant library. Second, nonspecific primer binding can lead to confounding PCR artifacts. Third, the PCR amplification of rare active mutants from large pools of inactive mutants can be challenging when attempting to substantially alter catalytic activity. Therefore a modified version of this system was created for ZFR evolution in which recovery of active mutants was based on the reassembly of a gene sequence that confers antibiotic resistance (Gersbach et al. 2010). With this system, the selection stringency could be tuned by varying the duration of culture time before introducing the selective antibiotic or by altering the length of the DNA binding domain. Active ZFR variants were recovered from libraries in the presence of $>10^6$ inactive variants. Novel ZFR variants were identified that had >1,000-fold increased activity on spacer sequences from the human genome relative to the parent enzyme. Finally, mutations that were repeatedly identified in these evolutions suggested residues critical to the determination of substrate specificity (Gersbach et al. 2010). A colorimetric assay of plasmid recombination in bacteria has also been used to screen ZFR mutants to obtain those with broadened sequence specificities (Proudfoot et al. 2011). Collectively, these efforts for directed evolution of ZFR specificity have increased the flexibility of the serine recombinase

catalytic domain to react with diverse sequences (Gersbach et al. 2010; Gordley et al. 2007; Proudfoot et al. 2011). Theoretically, this should allow programming of ZFR activity to chromosomal loci by the targeted zinc finger protein as has been done for ZFNs (Rahman et al. 2011; Urnov et al. 2010).

The incorporation of sequence specificity into the catalytic domain may provide an additional layer of strict target site selectivity and eliminate off-target effects observed with ZFNs in mammalian cells (Cornu et al. 2008; Miller et al. 2007; Szczepek et al. 2007). The original studies of ZFR evolution used random mutagenesis across the complete catalytic domain and surprisingly the same mutations were found to broaden sequence specificity when selection was performed using a variety of sequences (Gersbach et al. 2010; Gordley et al. 2007). To avoid these mutations that cause promiscuous enzyme activity, Gaj et al. used targeted mutagenesis based on structural predictions of residues that made the specificity-determining contacts with nearby nucleotides (Gaj et al. 2011). By using saturation mutagenesis at these targeted residues with the selection system described above (Gersbach et al. 2010), the specificities of the Gin and Tn3 catalytic domain were shifted $>10^4$- and $\sim 10^3$-fold, respectively, to react with each other's natural target sequences. The new variants of the Gin and Tn3 catalytic domains showed negligible activity on their natural target sites and had activities on their new target sites that were similar to levels of the parental enzyme on its natural target. These results showed that structure-guided reprogramming of recombinase activity can yield highly active novel enzymes targeted to new sequences without the need to generate promiscuous catalytic domains with relaxed sequence specificities (Gaj et al. 2011). Collectively, this work is moving the field towards a method for enzyme-mediated transgene integration that addresses many of the safety concerns of current approaches.

10.4 Conclusions, Challenges, and Future Directions

Technologies for integrating transgenes into cellular genomes have the potential to revolutionize genetics, agriculture, pharmaceutical discovery and production, and gene therapy. However, much work remains to be done in order to establish methods that are safe, efficient, and robust. The ZFR technology described here is promising as it addresses the lack of site specificity that has hindered the broad adoption of genetic engineering and gene therapy in modern medicine and biotechnology. However, much is still unknown and unproven. First, the ability to target transgene integration into any endogenous genomic target sites remains to be demonstrated. The studies described in this chapter provide a promising blueprint for this technical advance. Second, the non-specific effects of ZFRs on off-target sites in the genome remain to be characterized. Recent studies of off-target effects of zinc finger nucleases demonstrate this is a reasonable endeavor (Gabriel et al. 2011; Pattanayak et al. 2011). Biochemical characterization of ZFR activity will facilitate more robust recombinase engineering, and recent studies of *in vitro* ZFR activity are a promising advance in this area (Prorocic et al. 2011). Finally, the recent development of

alternative programmable DNA-binding proteins based on the TAL effector scaffold may provide additional, and perhaps more effective, methods for engineering serine recombinases for targeted transgene integration (Boch et al. 2009; Miller et al. 2011; Moscou and Bogdanove 2009). Ultimately, we expect that modular recombinases will be used to integrate transgenes into cellular genomes in a way that is safer, more efficient, and more broadly applicable than conventional methods.

References

Ackroyd AJ, Avila P, Parker CN, Halford SE (1990) Site-specific recombination by mutants of Tn21 resolvase with DNA recognition functions from Tn3 resolvase. J Mol Biol 216:633–643

Aiuti A, Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, Scaramuzza S, Andolfi G, Mirolo M, Brigida I et al (2009) Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med 360:447–458

Akopian A, Marshall Stark W (2005) Site-specific DNA recombinases as instruments for genomic surgery. Adv Genet 55:1–23

Akopian A, He J, Boocock MR, Stark WM (2003) Chimeric recombinases with designed DNA sequence recognition. Proc Natl Acad Sci USA 100:8688–8691

Avila P, Ackroyd AJ, Halford SE (1990) DNA binding by mutants of Tn21 resolvase with DNA recognition functions from Tn3 resolvase. J Mol Biol 216:645–655

Beerli RR, Segal DJ, Dreier B, Barbas CF 3rd (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci USA 95:14628–14633

Beerli RR, Dreier B, Barbas CF 3rd (2000) Positive and negative regulation of endogenous genes by designed transcription factors. Proc Natl Acad Sci USA 97:1495–1500

Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, Chandrasegaran S (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol 21:289–297

Bibikova M, Beumer K, Trautman JK, Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. Science 300:764

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816

Blancafort P, Segal DJ, Barbas CF 3rd (2004) Designing transcription factor architectures for drug discovery. Mol Pharmacol 66:1361–1371

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512

Bolusani S, Ma CH, Paek A, Konieczka JH, Jayaram M, Voziyanov Y (2006) Evolution of variants of yeast site-specific recombinase Flp that utilize native genomic sequences as recombination target sites. Nucleic Acids Res 34:5259–5269

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132:311–322

Boztug K, Schmidt M, Schwarzer A, Banerjee PP, Diez IA, Dewey RA, Bohm M, Nowrouzi A, Ball CR, Glimm H et al (2010) Stem-cell gene therapy for the Wiskott-Aldrich syndrome. N Engl J Med 363:1918–1927

Buchholz F, Stewart AF (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. Nat Biotechnol 19:1047–1052

Buchholz F, Angrand PO, Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. Nat Biotechnol 16:657–662

Burke ME, Arnold PH, He J, Wenwieser SV, Rowland SJ, Boocock MR, Stark WM (2004) Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. Mol Microbiol 51:937–948

Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S, Hoffmann C (2005) Genome-wide analysis of retroviral DNA integration. Nat Rev Microbiol 3:848–858

Calos MP (2006) The phiC31 integrase system for gene therapy. Curr Gene Ther 6:633–645

Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6:507–512

Cartier N, Hacein-Bey-Abina S, Bartholomae CC, Veres G, Schmidt M, Kutschera I, Vidaud M, Abel U, Dal-Cortivo L, Caccavelli L et al (2009) Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science 326:818–823

Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, Down J, Denaro M, Brady T, Westerman K et al (2010) Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. Nature 467:318–322

Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT, Hoelters J, Calos MP (2006) Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol 357:28–48

Choo Y, Klug A (1994) Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. Proc Natl Acad Sci USA 91:11163–11167

Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA (2005) Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature 436:272–276

Cornu TI, Thibodeau-Beganny S, Guhl E, Alwin S, Eichtinger M, Joung JK, Cathomen T (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. Mol Ther 16:352–358

Desjarlais JR, Berg JM (1992) Toward rules relating zinc finger protein sequences and DNA binding site preferences. Proc Natl Acad Sci USA 89:7345–7349

Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 122:473–483

Dreier B, Segal DJ, Barbas CF 3rd (2000) Insights into the molecular recognition of the 5′-GNN-3′ family of DNA sequences by zinc finger domains. J Mol Biol 303:489–502

Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF 3rd (2001) Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 276:29466–29478

Dreier B, Fuller RP, Segal DJ, Lund CV, Blancafort P, Huber A, Koksch B, Barbas CF 3rd (2005) Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 280:35588–35597

Ehrhardt A, Engler JA, Xu H, Cherry AM, Kay MA (2006) Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. Hum Gene Ther 17:1077–1094

Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. Structure 4:1171–1180

Fischer A, Hacein-Bey-Abina S, Cavazzana-Calvo M (2010) 20 years of gene therapy for SCID. Nat Immunol 11:457–460

Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppel C, Nowrouzi A, Bartholomae CC, Wang J, Friedman G et al (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol 29(9):816–823

Gaj T, Mercer AC, Gersbach CA, Gordley RM, Barbas CF III (2011) Structure-guided reprogramming of serine recombinase DNA sequence specificity. Proc Natl Acad Sci USA 108:498–503

Gersbach CA, Gaj T, Gordley RM, Barbas CF 3rd (2010) Directed evolution of recombinase specificity by split gene reassembly. Nucleic Acids Res 38:4198–4206

Gersbach CA, Gaj T, Gordley RM, Mercer AC, Barbas CF 3rd (2011) Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. Nucleic Acids Res 39:7868–7878

Gordley RM, Smith JD, Graslund T, Barbas CF 3rd (2007) Evolution of programmable zinc finger-recombinases with activity in human cells. J Mol Biol 367:802–813

Gordley RM, Gersbach CA, Barbas CF 3rd (2009) Synthesis of programmable integrases. Proc Natl Acad Sci USA 106:5053–5058

Greisman HA, Pabo CO (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. Science 275:657–661

Grindley ND, Whiteson KL, Rice PA (2006) Mechanisms of site-specific recombination. Annu Rev Biochem 75:567–605

Hacein-Bey-Abina S, Hauer J, Lim A, Picard C, Wang GP, Berry CC, Martinache C, Rieux-Laucat F, Latour S, Belohradsky BH et al (2010) Efficacy of gene therapy for X-linked severe combined immunodeficiency. N Engl J Med 363:355–364

Hackett PB, Largaespada DA, Cooper LJ (2010) A transposon and transposase system for human application. Mol Ther 18:674–683

Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91:501–510

Ivics Z, Katzer A, Stuwe EE, Fiedler D, Knespel S, Izsvak Z (2007) Targeted Sleeping Beauty transposition in human cells. Mol Ther 15:1137–1144

Jamieson AC, Kim SH, Wells JA (1994) In vitro selection of zinc fingers with altered DNA-binding specificity. Biochemistry 33:5689–5695

Klippel A, Cloppenborg K, Kahmann R (1988) Isolation and characterization of unusual gin mutants. EMBO J 7:3983–3989

Koller BH, Smithies O (1992) Altering genes in animals by gene targeting. Annu Rev Immunol 10:705–730

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Liu J, Jeppesen I, Nielsen K, Jensen TG (2006) Phi c31 integrase induces chromosomal aberrations in primary human fibroblasts. Gene Ther 13:1188–1190

Logie C, Stewart AF (1995) Ligand-regulated site-specific recombination. Proc Natl Acad Sci USA 92:5940–5944

Loonstra A, Vooijs M, Beverloo HB, Allak BA, van Drunen E, Kanaar R, Berns A, Jonkers J (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. Proc Natl Acad Sci USA 98:9209–9214

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA et al (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31:294–301

Mandell JG, Barbas CF 3rd (2006) Zinc finger tools: custom DNA-binding domains for transcription factors and nucleases. Nucleic Acids Res 34:W516–W523

Mates L, Chuah MK, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, Grzela DP, Schmitt A, Becker K, Matrai J et al (2009) Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat Genet 41:753–761

Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA et al (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25:778–785

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ et al (2011) A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29:143–148

Mitsuyasu RT, Merigan TC, Carr A, Zack JA, Winters MA, Workman C, Bloch M, Lalezari J, Becker S, Thornton L et al (2009) Phase 2 gene therapy trial of an anti-HIV ribozyme in autologous CD34+ cells. Nat Med 15:285–292

Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501

Naiche LA, Papaioannou VE (2007) Cre activity causes widespread apoptosis and lethal anemia during embryonic development. Genesis 45:768–775

Nardelli J, Gibson T, Charnay P (1992) Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. Nucleic Acids Res 20:4137–4144

O'Gorman S, Fox DT, Wahl GM (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science 251:1351–1355

Olorunniji FJ, He J, Wenwieser SV, Boocock MR, Stark WM (2008) Synapsis and catalysis by activated Tn3 resolvase mutants. Nucleic Acids Res 36:7181–7191

Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kuhlcke K, Schilz A, Kunkel H et al (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med 12:401–409

Papapetrou EP, Lee G, Malani N, Setty M, Riviere I, Tirunagari LM, Kadota K, Roth SL, Giardina P, Viale A et al (2011) Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. Nat Biotechnol 29:73–78

Pattanayak V, Ramirez CL, Joung JK, Liu DR (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods 8(9):765–770

Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252:809–817

Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763

Prorocic MM, Wenlong D, Olorunniji FJ, Akopian A, Schloetel JG, Hannigan A, McPherson AL, Stark WM (2011) Zinc-finger recombinase activities in vitro. Nucleic Acids Res 39(21):9316–9328

Proudfoot C, McPherson AL, Kolb AF, Stark WM (2011) Zinc finger recombinases with adaptable DNA sequence specificity. PLoS One 6:e19537

Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10:410–422

Rad R, Rad L, Wang W, Cadinanos J, Vassiliou G, Rice S, Campos LS, Yusa K, Banerjee R, Li MA et al (2010) PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. Science 330:1104–1107

Rahman SH, Maeder ML, Joung JK, Cathomen T (2011) Zinc-finger nucleases for somatic gene therapy: the next frontier. Hum Gene Ther 22:925–933

Rebar EJ, Pabo CO (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science 263:671–673

Rouet P, Smih F, Jasin M (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci USA 91:6064–6068

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y et al (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8:67–69

Santoro SW, Schultz PG (2002) Directed evolution of the site specificity of Cre recombinase. Proc Natl Acad Sci USA 99:4185–4190

Sarkar I, Hauber I, Hauber J, Buchholz F (2007) HIV-1 proviral DNA excision using an evolved recombinase. Science 316:1912–1915

Sauer B, Henderson N (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc Natl Acad Sci USA 85:5166–5170

Sauer B, Henderson N (1990) Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. New Biol 2:441–449

Schmidt EE, Taylor DS, Prigge JR, Barnett S, Capecchi MR (2000) Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. Proc Natl Acad Sci USA 97:13702–13707

Schneider F, Schwikardi M, Muskhelishvili G, Droge P (2000) A DNA-binding domain swap converts the invertase gin into a resolvase. J Mol Biol 295:767–775

Segal DJ, Dreier B, Beerli RR, Barbas CF 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. Proc Natl Acad Sci USA 96:2758–2763

Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, Huber A, Koksch B, Lund CV, Magnenat L, Valente D et al (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry 42:2137–2148

Segal DJ, Crotty JW, Bhakta MS, Barbas CF 3rd, Horton NC (2006) Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. J Mol Biol 363:405–421

Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, Carroll D (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res 28:3361–3369

Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol 25:786–793

Thyagarajan B, Guimaraes MJ, Groth AC, Calos MP (2000) Mammalian genomes contain active recombinase recognition sites. Gene 244:47–54

Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol Cell Biol 21:3926–3934

Tupler R, Perini G, Green MR (2001) Expressing the human genome. Nature 409:832–833

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435:646–651

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646

VandenDriessche T, Ivics Z, Izsvak Z, Chuah MK (2009) Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. Blood 114:1461–1468

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al (2001) The sequence of the human genome. Science 291:1304–1351

Vigdal TJ, Kaufman CD, Izsvak Z, Voytas DF, Ivics Z (2002) Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. J Mol Biol 323:441–452

Voziyanov Y, Konieczka JH, Stewart AF, Jayaram M (2003) Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site. J Mol Biol 326:65–76

Wehling P, Reinecke J, Baltzer AW, Granrath M, Schulitz KP, Schultz C, Krauspe R, Whiteside TW, Elder E, Ghivizzani SC et al (2009) Clinical responses to gene therapy in joints of two subjects with rheumatoid arthritis. Hum Gene Ther 20:97–101

Wilson MH, Coates CJ, George AL Jr (2007) PiggyBac transposon-mediated gene transfer in human cells. Mol Ther 15:139–145

Woodard LE, Keravala A, Jung WE, Wapinski OL, Yang Q, Felsher DW, Calos MP (2010) Impact of hydrodynamic injection and phiC31 integrase on tumor latency in a mouse model of MYC-induced hepatocellular carcinoma. PLoS One 5:e11367

Wu H, Yang WP, Barbas CF 3rd (1995) Building zinc fingers by selection: toward a therapeutic application. Proc Natl Acad Sci USA 92:344–348

Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 300:1749–1751

Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22:1393–1398

Yang W, Steitz TA (1995) Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. Cell 82:193–207

Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA (2005) High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol 25:2085–2094

Yant SR, Huang Y, Akache B, Kay MA (2007) Site-directed transposon integration in human cells. Nucleic Acids Res 35:e50

Part IV
Applications

Chapter 11
Gene Site-Specific Insertion in Plants

Olivier Da Ines and Charles I. White

Abstract Gene site-specific insertion is an important tool to precisely integrate DNA molecules at a defined genomic location. Gene Targeting has many important implications in plant biotechnology such as creation of novel plant traits or the study of gene function. It requires the insertion of an extrachromosomal DNA into the plant cell that will be further integrated at a defined genomic locus through the use of the plant endogenous recombination machinery. Extensive efforts have been made to understand the mechanisms governing gene targeting and for establishing efficient system to achieve precise and efficient targeting. In particular the recent development of engineered nucleases represents an important step forward towards allowing the insertion of a transgene at any given locus in the genome.

Keywords Engineered nucleases • Gene targeting • Homologous recombination • Positive-negative selection • T-DNA integration

11.1 Introduction

The insertion of a foreign DNA molecule at a desired genome locus, a process termed gene targeting (GT), has been a major aim for plant researchers and biotechnologists for decades. However, efficient gene targeting through homologous recombination remains elusive in plants. Considerable effort has been made and promising progress has been achieved in increasing the efficiency of gene targeting in plants, with

O. Da Ines • C.I. White (✉)
Génétique, Reproduction et Développement, UMR CNRS 6293 – Clermont Université – INSERM U1103, Centre National de Recherche Scientifique/Université Blaise Pascal, 24 avenue des Landais, BP 80026, 63171 Aubière CEDEX, France
e-mail: chwhite@univ-bpclermont.fr

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_11,

studies reporting rates of gene targeting ranging from 10^{-3} to 10^{-6} (reviewed in (Mengiste and Paszkowski 1999; Iida and Terada 2005; Tzfira and White 2005)). Such engineering is particularly important for crop improvement but depends upon the control of the location and nature of the integration of extrachromosomal DNA into the host genome. The precise engineering of any genome is rendered possible through delivery of an extrachromosomal DNA molecule into a cell followed by search for the homologous target site and finally, integration at the define genome locus via a recombination mechanism. In this review we discuss the different approaches used, or in use, for plant gene targeting.

Before starting discussion of gene targeting in plants it is worth mentioning work carried out in moss. The moss Physcomitrella patens shows a high efficiency of GT, ranging from 10% to 90% (Schaefer and Zryd 1997; Hohe et al. 2004; Trouiller et al. 2006; Trouiller et al. 2007; Schaefer et al. 2010). In combination with facile transformation and a sequenced genome this makes P. patens a strong model system for plant genetics (Schaefer et al. 1991; Schaefer 2001; Schaefer and Zryd 2001; Schaefer 2002; Rensing et al. 2008). A number of moss genes have been disrupted for functional analyses (reviewed in (Schaefer 2001)) and interestingly, a recent comparative study in Ceratodon purpureus raises the possibility that high gene targeting efficiency is a conserved feature of Bryophyte transformation (Trouiller et al. 2007). Understanding the mechanisms by which mosses achieve such efficiencies of GT could provide valuable information for implementing better gene targeting strategies in higher plants. Some hints come from analyses of recombination defective mutants, with a slight decrease in gene targeting reported in a Physcomitrella patens msh2 mutant (Trouiller et al. 2006) and the abolishment of GT in rad51-1 rad51-2 double mutants of P. patens (Schaefer et al. 2010).

11.2 Approaches for the Delivery of a Foreign Molecule into Plant Cells

The integration of extrachromosomal DNA into plant cells is a prerequisite for gene targeting and it is therefore critical that delivery of DNA molecules to the nuclei of the target cells is not limiting. DNA molecules are transferred into plant cells either using direct gene transfer methods or, more frequently, using the Gram negative soil bacterium *Agrobacterium tumefaciens*, which allows the transfer of the DNA sequence of interest into the plant genome via integration of a Transfer-DNA (T-DNA). In this respect, improvements in methods of Agrobacterium-mediated transformation of Arabidopsis plants represented a considerable breakthrough allowing the generation of large numbers of transformed plants with greater efficiency (Bechtold et al. 1993; Clough and Bent 1998). *Agrobacterium tumefaciens*-mediated gene transfer also has the advantage of frequently generating simple, single-locus insertions, although a recent study shows that this is strongly dependent upon the particular target cell (De Buck et al. 2009).

11.2.1 Agrobacterium-Mediated T-DNA Transformation

T-DNA integration into the plant genome involves a complex machinery with contributions from both the Agrobacterium donor and the receiving plant cell. In the bacterial cell, the transfer or "T-DNA" segment defined by flanking Left Border (LB) and Right Border (RB) sequences is excised from the Tumour inducing (Ti) plasmid as a single-stranded DNA molecule. This T-DNA is transported into the host-cell nucleus, where it is integrated into the plant genome. Agrobacterium-mediated plant transformation has been the subject of a number of recent reviews (Lacroix et al. 2006; Tzfira and Citovsky 2006; Dafny-Yelin et al. 2008; Gelvin 2010; Pitzschke and Hirt 2010; Magori and Citovsky 2011).

In addition to the Agrobacterium Vir (virulence) proteins, host plant proteins make major contributions to the transformation process, and have been the subject of much interest (reviewed by Gelvin 2010). It has been shown that plants respond differently to Agrobacterium transformation (Nam et al. 1997) and transcriptome analyses revealed deregulation of many genes involved in plant defence, hormone signalling or encoding histone proteins (Deeken et al. 2006; Ditt et al. 2006; Kim et al. 2007; Lee et al. 2009). Other approaches such as genetic screens for the competence of plants for Agrobacterium-mediated transformation allowed identification of various Arabidopsis mutants, hypersensitive (*hat* mutants) or resistant to Agrobacterium transformation (*rat* mutants) (Nam et al. 1999; Zhu et al. 2003). These studies clearly demonstrate the importance of plant proteins for delivery and integration of extrachromosomal DNA into plant cells (for recent reviews see (Gelvin 2010; Magori and Citovsky 2011)).

Notwithstanding the importance of the final step of the T-DNA transformation process for gene targeting, understanding of the pathway(s) used for integration of T-DNA into plant genomes remains elusive. It has been suggested that integration occurs either via nonhomologous end-joining (NHEJ) at a double-strand break (DSB) site that has occurred in the genome or through a microhomology search mechanism (Gheysen et al. 1991; Mayerhofer et al. 1991). The former pathway has been the favoured hypothesis, and it has been assumed that specifically reducing or eliminating NHEJ recombination would favour homologous recombination and thus enhance gene site-specific interaction. As a consequence the relationship between plant proteins involved in recombination (and in particular in the NHEJ pathway) and their putative roles in T-DNA integration has raised a lot of interest (see below). Ten years ago the group of Barbara Hohn used an in vitro ligation-integration assay to show that ligation of the T-DNA-VirD2 complex to a target DNA requires a DNA ligase from the plant host cell (Ziemienowicz et al. 2000). However, the specific plant ligase responsible for this was not identified. Studies of Agrobacterium-mediated T-DNA integration in yeast demonstrated clearly that the recombination pathways of the yeast host cell determined the type and efficiency of integration, and that several NHEJ proteins were involved in T-DNA integration in the absence of homology between the T-DNA and the yeast genome (van Attikum et al. 2001). In particular, the authors showed that the yeast DNA Ligase 4 and Ku70 proteins are required for T-DNA integration. Similar data were

obtained for Mre11, Rad50 and Xrs2, the proteins of the MRX complex involved in the initial processing of double strand breaks (van Attikum et al. 2001).

In Arabidopsis, DNA ligase 4 is required for NHEJ recombination, however its precise role in T-DNA transformation of Arabidopsis remains subject to some confusion (Friesner and Britt 2003; van Attikum et al. 2003; Tanaka et al. 2010). Similarly, small increases and decreases in efficiencies of T-DNA integration in Arabidopsis *ku80* mutants were reported (Friesner and Britt 2003; Gallego et al. 2003). This confusion appears to be due to quantitative variability in the floral-dip transformation method used in these studies and the role of Ku80 was subsequently confirmed by the use of a root transformation assay which clearly showed that stable T-DNA integration was strongly repressed in a ku80 knockout mutant while overexpression of the Ku80 protein induced T-DNA integration (Li et al. 2005b). That the efficiency of stable, and not transient transformation was affected in the mutants strongly supports a direct role of Arabidopsis Ku80 in T-DNA integration (Li et al. 2005b) (see also discussion by Gelvin 2010). Concerning the influence of homologous recombination proteins in T-DNA integration, a role in T-DNA integration in yeast has been suggested for the MRX complex and the key recombination proteins RAD52 and RAD51 (van Attikum and Hooykaas 2003), but data are limited in plants. Expression of the heterologous yeast RAD54 protein in the plant has been shown to play an important role (Shaked et al. 2005), but whether the plant RAD54 fulfills the same function remains to be demonstrated (see below).

A number of other proteins, less directly involved in recombination, have been shown to be important for T-DNA integration. For instance mutation of the VirE2-interacting proteins VIP1 and VIP2 reduces stable transformation in Arabidopsis (Li et al. 2005a; Anand et al. 2007a). This is also the case for proteins important for chromatin organisation and modification, which have been shown to be altered in T-DNA integration (for recent reviews see (Gelvin 2010; Magori and Citovsky 2011)). For instance, several histones and histone modifying enzymes play significant roles in Agrobacterium transformation. The Arabidopsis Histone 2A mutant, *hta1*, displays reduced stable transformation efficiency, while transient transformation remains unaffected (Nam et al. 1999; Mysore et al. 2000; Zhu et al. 2003). Similarly, mutants of histones H2B, H3 and H4 are deficient in T-DNA integration (Zhu et al. 2003), and overexpression of these proteins increases Agrobacterium-mediated plant transformation (Mysore et al. 2000; Yi et al. 2006; Tenea et al. 2009; Zheng et al. 2009). In tobacco, histones also play a crucial role for Agrobacterium transformation and H2A and H3 silenced plants are recalcitrant to stable Agrobacterium-mediated transformation (Anand et al. 2007b). Histone-modifying enzymes such as HDAC deacetylases, HAT acetyltransferases and DNA methyltransferases also influence T-DNA integration in plants (Zhu et al. 2003; Crane and Gelvin 2007). Reduction of the expression of the histone H3 chaperone SGA1 in Arabidopsis renders the plants more resistant to Agrobacterium-mediated transformation (Zhu et al. 2003; Crane and Gelvin 2007) whereas overexpression of SGA1 increases T-DNA integration in Arabidopsis (Crane and Gelvin 2007; Gelvin 2010). On the contrary, mutation of CAF-1, the Chromatin assembly factor, stimulates T-DNA integration in Arabidopsis (Endo et al. 2006a).

Although much remains to be learned, there is now no doubt concerning the importance of host-cell factors in the process of T-DNA integration. The determination of the specific nature of these roles will be crucial for efforts to modulate host-cell recombination pathways and chromatin structure in studies of plant transformation and gene targeting.

11.2.2 Oligonucleotide-Directed Plant Gene Targeting

In addition to Agrobacterium-mediated transformation, plant cells can be transformed using direct gene transfer methods. These are often used for protoplast or root transformation and involve Ca/PEG precipitation, electroporation, liposome mediated introduction, microinjection and particle bombardment or biolistics (reviewed by (Barampuram and Zhang 2011)). Permitting great freedom in the choice of the type of nucleic acid used for transformation, such methods have proven very useful but they also have some limitations and in particular, the regeneration of transformed plants is often difficult, complicating further analysis of transformants.

Direct gene transfer has been used for targeted genetic modification of plants by chimeric RNA/DNA oligonucleotides, consisting of a short DNA sequence flanked by modified RNA regions of about ten nucleotides that are complementary to the target gene. There are few published studies employing the RNA/DNA oligonucleotide-directed gene targeting approach in plants (see reviews by Hohn and Puchta 1999; Oh and May 2001; Iida and Terada 2005). The method was first shown to be applicable in tobacco and maize (Beetham et al. 1999; Zhu et al. 1999). Tobacco suspension calli and Maize embryogenic calli were transformed by particle bombardment to deliver a chimeric RNA/DNA oligonucleotide designed to target the tobacco acetolactate synthase gene (ALS) (Beetham et al. 1999) or the Maize acetohydroxyacid synthase gene (AHAS) (Zhu et al. 1999). In both species, the mutation of one base in the gene causes a specific amino acid substitution and renders the plants resistant to sulphonylurea herbicides. Using this method the authors achieved a gene targeting efficiency around 10^{-4} that is in the range of what is reported for gene targeting by homologous recombination (see below and Tables 11.1 and 11.2). These authors also reported targeting of a modified, inactive GFP transgene and made it active by either site-specific base substitution (Zhu et al. 1999) or insertion to make the GFP active (Beetham et al. 1999). Importantly, Zhu and colleagues demonstrated that although the introduced chimeric oligonucleotides were degraded several hours after delivery within the cell, the mutation was stable and was transmitted to following generations in a Mendelian manner (Zhu et al. 1999; Zhu et al. 2000). These studies also described the occurrence of non-specific modifications adjacent to the target site (Beetham et al. 1999; Zhu et al. 1999; Zhu et al. 2000) and these non-specific alterations were further confirmed in vitro using plant cell-free extract systems. Interestingly, cell-free extracts prepared from tobacco showed non-specific base alterations, and extracts from banana and maize induced more precise, site-specific base changes (Rice et al. 2000). Kipp and coworkers used this

Table 11.1 Gene targeting experiments using target-gene-specific marker selection

Selection marker	Target locus	Transformation process	Plants	GT frequency	Reference
Kanamycin	APH(3')II (exogenous)	DNA mediated direct transformation of protoplasts	Tobacco	$0.5–4.2 \times 10^{-4}$	Paszkowski et al. (1988)
Kanamycin	NPTII (exogenous)	A. tumefaciens transformation of protoplasts	Tobacco	$10^{-4}–10^{-5}$	Offringa et al. (1990, 1993), Risseeuw et al. (1995)
Chlorsulfuron herbicide	Acetolactate synthase (ALS) (endogenous)	A. tumefaciens transformation of protoplasts	Tobacco	8.4×10^{-5}	Lee et al. (1990)
Hygromycin	Hygromycin phosphotransferase gene (hpt) (exogenous)	DNA mediated direct transformation of protoplasts	Arabidopsis thaliana	10^{-4}	Halfter et al. (1992)
Kanamycin	NPTII (exogenous)	DNA mediated direct transformation of leaf protoplasts	Tobacco	10^{-4}	Hrouda and Paszkowski (1994)
Kanamycin	TGA3 (endogenous)	Root explant transformation by A. tumefaciens	Arabidopsis thaliana	$4–8 \times 10^{-4}$	Miao and Lam (1995)
Kanamycin	AGL5 (endogenous)	Agrobacterium-mediated Arabidopsis floral dip transformation	Arabidopsis thaliana	10^{-3}	Kempin et al. (1997)
Kanamycin	NPTII (exogenous)	Leaf protoplast transformation with A. tumefaciens	Tobacco	$1–14 \times 10^{-4}$	Reiss et al. (2000)
Butafenacil	PPO (endogenous)	Agrobacterium-mediated Arabidopsis floral dip transformation	Arabidopsis thaliana	7.2×10^{-4}	Hanin et al. (2001)
Imazapyr	Acetolactate synthase (ALS) (endogenous)	Agrobacterium-mediated Arabidopsis floral dip transformation	Arabidopsis thaliana	3.2×10^{-4}	Endo et al. (2006b)

Bispyribac	Acetolactate synthase (ALS) (endogenous)	Agrobacterium-mediated embryonic Rice transformation	Rice	$3–5 \times 10^{-4}$	Endo et al. (2007)
Trp-analogue 5-methyl-Trp (5MT)	OASA2 (key enzyme of tryptophan biosynthesis)	Agrobacterium-mediated embryonic Rice transformation	Rice	$1.5–4 \times 10^{-3}$	Saika et al. (2011)

For each experiment, the selection marker, target locus, transformation process and plant species are indicated. Estimates of the frequencies of true gene targeting events are taken from the corresponding studies

Table 11.2 Gene targeting experiments using positive-negative selection

Selection Markers(+ve/−ve)	Target locus	Tissue	Plant	Estimated total transformants	PNS+	Gene targeting events	Reference
Kanamycin/codA	Glu1, Pzf	Hypocotyl callus	Lotus japonicus	18,974	185	0	Thykjaer et al. (1997)
Kanamycin/codA	CHS	Suspension callus	Arabidopsis	109,475	4,379	0	Gallego et al. (1999)
Kanamycin/codA	ADH	Root explant callus	Arabidopsis	6,250	39	1[a]	Xiaohui Wang (2001)
Hygromycin/ DT−A	waxy	Embryonic callus	Rice	9,269	638	6	Terada et al. (2002)
Hygromycin/ DT-A	ADH	Embryonic callus	Rice	Not given	468	9	Johzuka-Hisatomi et al. (2008)
Hygromycin/ DT-A	met1	Embryonic callus	Rice	Not given	283[b]	15	Yamauchi et al. (2009)

All studies used *A. tumefaciens*-mediated transformation of the indicated tissues and selection with the indicated marker genes. For each experiment, positive and negative selection marker, target locus, the tissue transformed and plant species are indicated. Estimated total numbers of individual transformed cells screened, numbers of clones surviving the positive-negative selection (PNS+) and numbers of gene targeted transformants are shown. (see text for details)

[a]Unfortunately this transformant was lost and molecular analysis was not completed

[b]The figure of 283 PNS + calli was calculated from the frequency given in the article by Yamauchi et al

approach to introduce site-specific point mutations in the targeted endogenous gene using self-complementary oligonucleotides containing a mutation to be introduced in the gene of interest (Kipp et al. 2000). The efficiency, mechanism and precision of this process is however still largely a matter of debate, although some information is available. Kochevenko and Willmitzer demonstrated that the target sequence has limited influence (Kochevenko and Willmitzer 2003). After mesophyll protoplast electroporation or leaf tissue particle bombardment, they obtained similar gene targeting efficiencies using two different chimeric oligonucleotides targeting two crucial positions in the ALS gene in Tobacco. In rice, three different oligonucleotides transferred into embryogenic calli by biolistics were successfully used to modify three different positions of the rice ALS gene (Okuzaki and Toriyama 2004). Surprisingly, the co-delivery of two different chimeric oligonucleotides induced modification of only one of the two targeted bases. In wheat, this strategy has been applied to cultured scutellum cells after delivery of mutated GFP constructs together with chimeric oligonucleotides by particle bombardment (Dong et al. 2006). The visualisation of GFP fluorescence was indicative of site-specific base changes at the target site. These studies have shown the effectiveness of oligonucleotide-directed gene targeting in plants, with an efficiency similar to that obtained with other gene targeting approaches. However, they all have involved use of either the endogenous ALS/AHAS gene, or of an artificial GFP reporter gene.

11.3 Selection Approaches

The generation of transgenic plants and, hence, the delivery of foreign DNA molecules to the cellular nuclei is now routine in a number of plant species. However, the integration of extrachromosomal DNA occurs at apparently random locations in the genome (Tissier et al. 1999; Brunaud et al. 2002; Sessions et al. 2002; Alonso et al. 2003; Rosso et al. 2003). Gene site-specific insertion is made possible through homologous recombination (HR), which in plants, has been assumed to be less favoured than non-homologous end-joining (NHEJ) as the pathway for the chromosomal integration of foreign DNA molecules. To deal with the relatively low efficiency of HR, researchers have used either specific selection systems to facilitate recovery of gene targeting events at a selectable target locus (Tables 11.1 and 11.2. Reviewed in (Iida and Terada 2005)), or the specific generation of a specific recombination site through induction of controlled double-strand breaks in the target DNA sequence.

11.3.1 Direct Gene-Specific Selection Procedures Using Selectable Markers

The first GT study reported in plants involved DNA-mediated direct transformation of transgenic tobacco protoplasts containing a non-functional unit of the APH(3')II gene. In this work the authors were able to reconstitute a functional unit of this gene

by integrating the missing part of the gene through homologous recombination (Paszkowski et al. 1988). The resulting Kanamycin resistance allowed recovery of several resistant clones with a gene targeting frequency of about 10^{-4} per transformed cell. A number of studies have since reported gene-targeting events using target locus selection systems and direct DNA transformation of protoplasts or Agrobacterium-mediated transformation in both tobacco (Lee et al. 1990; Offringa et al. 1990; Hrouda and Paszkowski 1994; Risseeuw et al. 1995; Reiss et al. 2000) and Arabidopsis (Halfter et al. 1992; Miao and Lam 1995; Kempin et al. 1997; Hanin et al. 2001; Endo et al. 2006b; Tanaka et al. 2010). Conversion of an introduced T-DNA molecule by a homologous chromosomal locus has also been described (Offringa et al. 1993). Moreover, using the same approach, it has been demonstrated that gene targeting seems independent of the genomic location of the target construct (Risseeuw et al. 1995). Most of these reports used direct DNA transfer into protoplasts possessing artificially modified exogenous target genes. Miao and Lam reported the first successful disruption of a non-selectable endogenous gene, the TGA3 locus, in *Arabidopsis thaliana* through targeted insertion of the bacterial neo gene conferring kanamycin resistance fused to the ß-glucuronidase (GUS) gene as a screenable marker to distinguish between homologous recombination events and ectopic insertions (Miao and Lam 1995). In this study, the extrachromosomal DNA was inserted into the plant genome via infection of Arabidopsis root explants by *Agrobacterium tumefaciens* and two homologous recombination events were isolated among 2,580 Kanamycin-resistant calli – giving a rate of gene targeting slightly higher than those previously reported (Table 11.1). Although successful, none of these studies allowed regeneration of entire plants and this therefore limited further characterisation. The development and improvement of intact plant transformation via *Agrobacterium tumefaciens* provided new opportunities and Kempin et al. used this approach to target the AGL5 MADS-box gene by homologous recombination for replacement by a Kanamycin-resistance gene (Kempin et al. 1997). Only one plant with a targeted insertion was isolated however and a precise frequency could not be determined. Similar data were obtained in another study targeting the Acetolactate Synthase (ALS) gene in Arabidopsis (Endo et al. 2006b) in which the authors obtained two targeting events, but only one was a true targeting event. To date, the most robust study of an endogenous target locus disruption has been described by Hanin et al. (2001), who introduced two mutations into the Arabidopsis protoporphyrinogen oxidase (PPO) gene, making the plant resistant to the herbicide butafenacil. They clearly differentiated true gene targeting and ectopic events (conversion of exogenous DNA by chromosomal PPO locus followed by ectopic insertion) and were able to characterise selfed-progeny of nine recombinant plants representing three true gene targeting and six ectopic events. Hanin et al. thus estimated the basal gene targeting frequency in their experiments to be 2.4×10^{-3} while true gene targeting frequency was estimated to be 0.72×10^{-3}.

Surprisingly, most GT studies have been done in dicot plants and monocotyledonous plants, including highly important crops such as cereals, have been little studied. This is changing however, with recent reports of successful gene targeting in rice using a gene-specific selection procedure based on mutation of the ALS gene

(Endo et al. 2007), as well as the positive-negative selection approach pioneered in rice by the Iida group (see below). Using an efficient Agrobacterium-mediated rice transformation, Endo et al. replaced the endogenous ALS gene by an exogenous copy harbouring two point mutations, rendering the rice tolerant to the herbicide bispyribac. Sixty-six independent GT rice plants were regenerated and further characterisation showed that most GT events were free of additional foreign DNA and rearrangements, leading the authors to argue that rice plants thus generated are equivalent to non-genetically modified tolerant plants, but with increased resistance (Endo et al. 2007). In a recent publication, this group has reported the successful engineering of rice plants which accumulate 230-fold higher levels of Tryptophan through gene targeting of the rice OASA2 gene (Saika et al. 2011).

11.3.2 Positive-Negative Selection

Gene-specific selection procedures have proven useful for studies of gene targeting in plants, however they require transformation and screening of very large numbers of plants and are often accompanied by undesirable ectopic insertions. Above all, this approach is restricted to specific loci at which the modification can be directly selected. To overcome these difficulties, researchers have followed the lead of the work in animals (Capecchi 1989), in using positive-negative selection (see Table 11.2). Vectors are designed such that positive-selection markers are placed within the sequence homologous to the target locus, and negative-selection markers are located outside of the homologous region (Fig. 11.1). In consequence, this approach permits selection of targeted insertions (which retain only the positive selection marker) and elimination of non-targeted insertions (which retain both selective markers). Two different negative selection markers have been used for this approach in plants: the DT-A gene encoding the diphteria toxin A-chain, and the cytosine deaminase (*codA*) gene from Escherichia coli.

The *codA* gene was first used as a negative selection marker in the legume *Lotus japonicus* (Thykjaer et al. 1997) and later in Arabidopsis (Gallego et al. 1999; Xiaohui Wang 2001) (Table 11.2). In lotus, a vector construct containing both an *nptII* and a *codA* cassette was used to target the *Gln1* or *Pzf* loci (Thykjaer et al. 1997) and although 185 double-resistant calli were obtained among 18,974 transformation events, no true targeting events were identified. Similarly in transformed Arabidopsis cell suspension, no targeting events were identified at the Chalcone Synthase locus (*CHS*) nor at an artificially created hpt locus, among 4,379 double-selected calli (from an estimated 110,000 transformants) (Gallego et al. 1999). Using a root explant transformation procedure with the same positive-negative selection markers, Xiaohui Wang et al. were able to recover one site specific insertion event among 6,250 transformants in a study targeting the Arabidopsis alcohol dehydrogenase gene (ADH) (Xiaohui Wang 2001).

The DT-A chain inhibits protein synthesis in the host cell by inhibiting ADP-ribosylation of elongation factor 2 (Pappenheimer 1977; Palmiter et al. 1987).

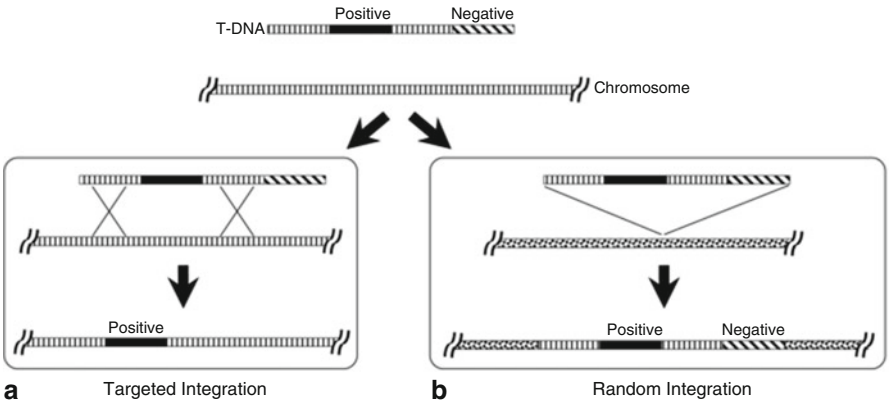


Fig. 11.1 Positive-negative selection. The positive-negative selection vector (T-DNA) is designed such that the positive-selection marker (*black fill*) is placed within the sequence homologous to the target locus (*vertical hatching*), and the negative-selection marker (*diagonal hatching*) is located outside of the homologous region. In consequence, this approach permits selection of targeted insertions into the homologous chromosomal locus which retain only the positive selection marker (**a**) and elimination of non-targeted insertions which retain both selective markers (**b**)

This very potent counter-selection marker has been shown to be highly toxic to plant cells (Czako and An 1991; Thorsness et al. 1991), although the toxicity of DT-A to plant cells is lower than to mammalian cells (Nilsson et al. 1998). The use of the DT-A gene rather than the complete toxin means that the active A-chain is not imported into adjacent, non-expressing cells and the selection is thus cell-autonomous. These characteristics make the DT-A gene particularly suitable for use in positive-negative selection in plants, as first described by the Iida group in rice. In 2002 Terada et al. reported the application of positive-negative selection to target the rice *waxy* gene, using hygromycin as a positive marker and two copies of the DT-A gene as negative markers (Terada et al. 2002). Homologous recombination between the *waxy* gene and the T-DNA construct resulted in precise gene targeting events in 1% of the 600 hygromycin resistant calli carried. GT plants were regenerated and further characterisation confirmed insertion at the *waxy* gene and that only one copy had been inserted. Although this is similar to the frequency reported for targeting the Arabidopsis PPO locus (Hanin et al. 2001), the absence of ectopic events following positive-negative selection holds great promise for gene site-specific insertion in plants. Terada et al. further developed a large-scale transformation procedure with the DT-A marker, transforming rice with high efficiency and facilitating selection of gene targeting events (Terada et al. 2004). Further work by Iida and colleagues demonstrated that the strategy can be extended to other endogenous targets (Iida and Terada 2005; Johzuka-Hisatomi et al. 2008; Yamauchi et al. 2009).

11.3.3 Gene Targeting Selection by Visual Screening

An alternative to positive-negative selection combines positive selection with screening for a visual marker. A first study using the GUS gene as a visual negative marker reported the isolation of one homologous recombination event in this way (Miao and Lam 1995) (Table 11.1). Visualisation of GFP expression in seeds was used by Shaked et al., who reported the reproducible modification of the Arabidopsis cruciferin gene (Shaked et al. 2005). Screening of seeds of transformed plants permitted identification of 19 GT events, all of which were found to contain the anticipated cruciferin gene structure with the GFP sequence integrated correctly into the gene by homologous recombination. The authors reported that overexpressing the yeast RAD54 protein in the target plants enhanced gene targeting from 10^{-4} to 10^{-3} in WT to 10^{-2} to 10^{-1} (see below).

11.4 Current Advances for Improving Targeting Efficiency

These methods developed to facilitate selection of GT events are not yet generally applicable, at least in part due to the low frequency of these events and thus the need to produce and screen very large numbers of transformants. To solve this, two general strategies are under development to increase the efficiency of GT: targeted induction of DSB to induce recombination at the target site and manipulation of the cellular recombination machinery to favour desired events. Advances in these two approaches will be discussed below and special attention given to the recent application of TAL effector nucleases for the induction of DSB at target loci, which is causing considerable excitement in the field.

11.4.1 Homology

The length of DNA sequence homology required for efficient gene targeting has been estimated as several hundred base pairs and several thousand base pairs in yeast and mouse, respectively (Rothstein 1991; Deng and Capecchi 1992). In less complex genomes such as that of *E. coli*, 20 bp appeared sufficient to promote homologous recombination (Watt et al. 1985). In plants, the length of homology used to achieve gene targeting differs with respect to the plant species (Schaefer 2001). In tobacco plants, it has been shown that at least a few hundred base pairs are sufficient to drive homologous recombination (Baur et al. 1990; Puchta and Hohn 1991). Using a transient enzyme assay, the authors monitored the homologous recombination by cotransfecting tobacco protoplasts with plasmids containing overlapping parts of the GUS gene. They examined the influence of the length of overlap (between 2 and 1,200 bp) on recombination rate of both supercoiled circular and linear plasmids. Interestingly, in both cases, the authors observed a positive correlation

between GUS activity and the length of overlapping homology with a strong drop occurring with overlaps of 456 bp or lower (Puchta and Hohn 1991). Similar data were obtained in Arabidopsis (Li et al. 2004). It is not currently known exactly what is the minimum length of homology required for homologous recombination. Puchta and Hohn observed homologous recombination with overlap as low as 222 bp (Puchta and Hohn 1991) while Li et al. were able to measure somatic homologous recombination with a 153 bp overlap. However, in both cases recombination frequencies were very low when using such small lengths of homology. In *Physcomitrella patens*, most reported gene targeting events used at least 1,000 base pairs of homology (Schaefer and Zryd 1997; Schaefer 2001) but a study investigating the minimal extent of sequence homology necessary for gene targeting suggested that stretches of sequence homology as low as 53 and 191 bp were sufficient for HR-mediated conversion of the apt gene (Houba-Hérin et al. 1997; Schaefer 2001). Interestingly, with such short homology the authors achieve GT only at low efficiency while optimal efficiency was obtained using 500–700 bp of homology (Houba-Hérin et al. 1997; Schaefer 2001).

Besides length, the degree of sequence similarity is important for homologous recombination but it is still unknown to what extent sequence can diverge and still allow homologous recombination. In Arabidopsis, it has been shown that about two percent divergence in a sequence is sufficient to trigger a 10-fold reduction in the frequency of recombination (Li et al. 2004, 2006). In fact, a single mismatch within a 618 bp of sequence homology has been reported to decrease homologous recombination 3-fold and the type or the position of the mismatch has no further effect on the recombination frequency (Opperman et al. 2004). The crucial role of the mismatch repair machinery in this process has been demonstrated by analyses of mutants of the mismatch repair protein MSH2, in which the reduction of homologous recombination induced by sequence divergence is abolished (Emmanuel et al. 2006; Li et al. 2006).

It is thus clear that the frequency of homologous recombination is to some extent dependent on the length and the degree of homology. However, further studies are needed to clearly characterise this relationship.

11.4.2 Modification of Proteins Involved in Homologous Recombination (HR) and Chromatin Remodelling

As mentioned above, transforming DNA is thought to be preferentially integrated in plants through non-homologous rather than homologous recombination. Gene site-specific insertion however involves homologous recombination and, in consequence, it is assumed that the non-homologous end-joining repair pathway negatively influences gene targeting efficiency. Understanding the molecular machinery involved in these processes and how the choice between these pathways is established will be helpful for implementing better strategies (for a recent review see Lieberman-Lazarovich and Levy 2011).

Several studies have attempted to modulate recombination and gene targeting efficiency *in planta* through increasing or reducing expression of specific recombination proteins. Studies on the effect of overexpression in plants has mostly concerned recombination genes from heterologous species. In tobacco, the overexpression of the *E. coli* RecA protein, a key protein in the homology search and strand exchange process, was associated with a significant increase in the rate of intrachromosomal homologous recombination and numbers of sister chromatid exchanges (Reiss et al. 1996, 2000). No effect of RecA expression on GT was however found by these authors (Reiss et al. 2000). Overexpression of the *E. coli* Holliday junction resolvase RuvC in tobacco has also been reported to stimulate homologous recombination (Shalev et al. 1999), with more than 10-fold stimulation of somatic crossovers and intrachromosomal recombination, and a 56-fold increase in extrachromosomal recombination reported. Overexpression of the (Arabidopsis) Ku80 protein in Arabidopsis led to increased resistance to DNA-damaging agents and enhanced the rate of extrachromosomal T-DNA to T-DNA recombination (Li et al. 2005b). Previous work in our lab suggested that inhibition of key recombination proteins could also affect gene targeting since homologous recombination in Arabidopsis is stimulated in the absence of Rad50 (Gherbi et al. 2001). It is thus clear that it is possible to modulate the rate of recombination in plants through engineering of the expression level of key proteins. However, as none of these studies reported an improvement in gene targeting efficiency, the effects of this on gene targeting remain to be established.

The effect of absence of recombination proteins on gene targeting efficiency in mutant plants has not been unequivocally demonstrated, although plant mutants have been shown to have altered rates of T-DNA integration (see above). An exception to this is the recent study of targeting the AGAMOUS gene in which a reduced transformation rate observed in the Arabidopsis ligase 4 mutant is accompanied with higher gene targeting efficiency (Tanaka et al. 2010). Quantitation is difficult in this work with biolistic transformation of callus *in vitro* and no Southern Hybridisation confirmation of the nature of the targeted events is presented. This perhaps explains the contrast between this result and earlier publications on T-DNA transformation of the ligase 4 mutant which found minor effects on transformability (Friesner and Britt 2003; van Attikum et al. 2003). This interesting result must however remain tentative in the absence of a full analysis of the GT events by southern hybridisation.

Another approach for modulation of homologous recombination is through chromatin remodelling. Several studies have demonstrated that chromatin state has a strong impact on homologous recombination efficiency. INO80, a member of the SWI2/SNF2 gene family involved in chromatin remodelling was suggested to be a positive regulator of homologous recombination (Fritsch et al. 2004). An Arabidopsis mutant affected in the expression of INO80 exhibited a strong reduction in homologous recombination, without showing effects on sensitivity to bleomycin and MMS, or T-DNA integration. Recombination is reduced in mutants with downregulation of other of SWI2/SNF2 gene family members (Shaked et al. 2006) and overexpression of the yeast Rad54 protein, another member of the SWI2/SNF2 gene family, has been shown to boost gene targeting in Arabidopsis by one to two orders of magnitude

(Shaked et al. 2005). Several models have been proposed to explain how yeast Rad54 expression could enhance gene targeting in Arabidopsis: RAD54 could directly favour DNA strand invasion through its chromatin remodelling activity or RAD54 could interact with RAD51, favouring homology search and strand exchange (Puchta and Hohn 2005; Shaked et al. 2005). An ortholog of RAD54 has been characterised in Arabidopsis (Osakabe et al. 2006; Shaked et al. 2006; Klutstein et al. 2008). AtRAD54 is involved in homologous recombination and is able to complement, at least partially, the yeast mutant. It has not however been demonstrated whether the plant RAD54 homolog is implicated in gene targeting. Unravelling the specific roles of these proteins will however be complicated by the fact that 40 members of the SWI2/SNF2 gene family have been identified in Arabidopsis (Shaked et al. 2006).

The idea that modifying DNA accessibility through alteration of chromatin structure could enhance homologous recombination is also supported by analyses of Chromatin assembly factor-1 (CAF-1) mutants (Endo et al. 2006a). Increased frequencies of somatic homologous recombination and T-DNA integration have been observed in the Arabidopsis CAF-1 mutants, fas1 and fas2. Also, other genes involved in chromatin dynamics (e.g. MIM, BRU1) (Hanin et al. 2000; Takeda et al. 2004) or in various developmental processes have been associated with homologous recombination (Molinier et al. 2004; Shaked et al. 2006; Yin et al. 2009).

These different studies of modulation of recombination *in planta* clearly confirm a complex interplay between non-homologous (NHR) and homologous recombination (HR). Understanding the regulation of this functional interplay will be of great importance in implementing better gene targeting strategies for plants. Recent work from our lab demonstrated a functional hierarchy of four NHR and HR pathways in Arabidopsis (Charbonnel et al. 2010; Charbonnel et al. 2011). Analyses of multiple mutants in Arabidopsis demonstrate that inhibiting an NHR pathway may enhance other NHR pathways and/or HR and vice versa (Charbonnel et al. 2011). Furthermore, in unpublished work we have shown that several of these multiple mutants were severely affected in T-DNA integration strongly suggesting that T-DNA is predominantly integrated into plant cells by NHR (F. Norre and C. White, unpublished data) and that multiple NHR pathways are involved. These results offer promising prospectives for the development of new approaches to gene targeting.

11.4.3 Targeting Using Engineered Nucleases

Double-strand DNA breaks (DSB) initiate homologous recombination in both somatic and meiotic cells. As a consequence, considerable effort is being invested into the use of endonucleases to enhance GT through favouring of homologous recombination at recombination site by generation of controlled DSBs. The meganuclease I-SceI has been the endonuclease of choice, although the use of meganucleases is being replaced more and more by engineered Zinc-Finger Nucleases (ZFN) or the recently described transcription activator-like effector nucleases (TALEN), which offer great promise for the creation of tools for stimulating targeting of any given genomic sequence.

11.4.3.1 I-SceI Meganuclease

I-SceI is a yeast intron-encoded endonuclease which drives the mobility of an optional intron of the 21S rRNA gene in yeast mitochondrial DNA (Jacquier and Dujon 1985; Dujon 1989). This rare-cutting, homing restriction enzyme has an 18 bp DNA recognition site and was originally shown to enhance homologous recombination in yeast cells (Plessis et al. 1992). In a pioneering study on extrachromosomal DNA recombination in tobacco protoplasts, Puchta et al. demonstrated that expression of the I-SceI enzyme in plant cells can enhance homologous recombination through induction of DSB at an I-SceI recognition site present within a recombination substrate (Puchta et al. 1993). The positive effect of induced DSBs on recombination in planta was confirmed using transgenic tobacco plants containing a target sequence containing the I-SceI recognition site. Transformation of the I-SceI expression construct and a vector carrying sequence homologies to the target into tobacco plants showed that intrachromosomal recombination at the target site could be increased by up to two orders of magnitude after DSB induction (Puchta et al. 1996). A similar approach was later used to show that DSB generated by I-SceI represent a preferred site for integration of T-DNA into the plant genome (Salomon and Puchta 1998). I-SceI induced recombination by one to two orders of magnitude in similar substrates in Arabidopsis (Orel et al. 2003), and as much as four orders of magnitude in a recent report (Wehrkamp-Richter et al. 2009). Similar results were obtained using a dexamethasone-inducible I-SceI enzyme and induction of I-SceI under the specific Apetala3 floral promoter specifically enhanced germinal recombination events up to 4-fold. The integration of extrachromosomal DNA through HR at a specific I-SceI target site has also been shown in Maize (D'Halluin et al. 2008). The authors optimised an I-SceI gene for expression in maize and showed induction of homologous recombination-mediated insertion at the desired site after DSB induction. Although these studies (and others) strongly underline the interest of I-SceI-induced DSBs, the need to have the recognition site for the endonuclease in the the target locus and the complexity of engineering these enzymes has led to searches for other means of inducing DNA breaks at specific genomic locations.

11.4.3.2 Targeting by Zinc Finger Nucleases

In recent years, considerable interest has surrounded the development of synthetic Zinc-Finger Nucleases (ZFN) as a means of building specific endonucleases for any given recognition sequence and thus permit generation of DSBs at defined genomic loci. ZFNs are chimeric nucleases comprising a synthetic DNA-recognition domain made of three (or more) C_2H_2 zinc fingers attached to the non-specific DNA cleavage domain of the restriction endonuclease FokI (Fig. 11.2). As each finger recognises a specific three nucleotide sequence, the synthetic 3-finger DNA-recognition domain has a nine-nucleotide sequence specificity. FokI functions as a dimer for DNA cleavage and the chimeric nuclease is thus composed of two synthetic DNA-recognition domains separated by a 5–7 bp spacer (Fig. 11.2), resulting

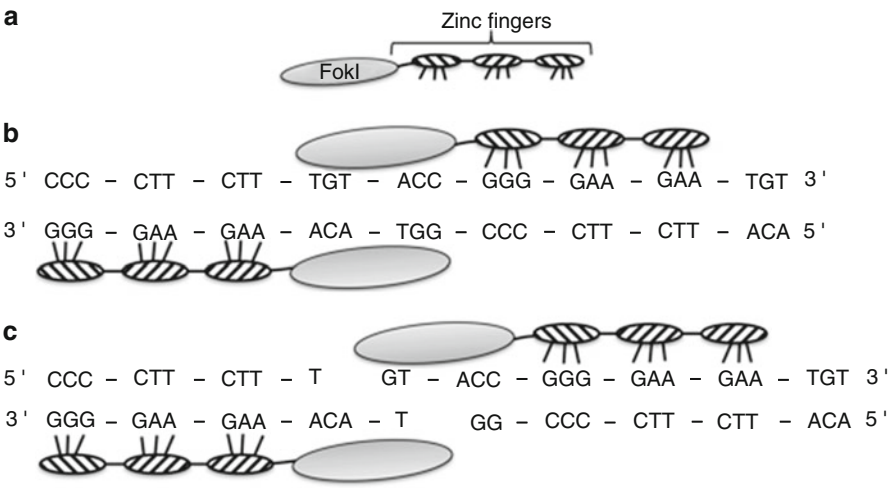


Fig. 11.2 Zinc finger nucleases. (**a**) ZFNs are chimeric nucleases comprising a synthetic DNA-recognition domain made of three (or more) C_2H_2 zinc fingers attached to the FokI type II restriction endonuclease. The synthetic 3-finger DNA-recognition domain has a nine-nucleotide sequence specificity. (**b**) FokI functions as a dimer for DNA cleavage. Two synthetic DNA-recognition domains separated by a 5–7 bp spacer will bind to their target DNA, resulting in a unique specificity of 18 nucleotides. (**c**) After DNA binding, the functional ZFN dimer can cleave the DNA. The DSBs provoked by the ZFN are primarily repaired by NHEJ and often result in deletion or insertion mutations of the target locus

in a unique specificity of 18 nucleotides. ZFN can theoretically be designed to recognise nearly all sequences of 18 nucleotides and by extension any gene could be targeted (for reviews see Durai et al. 2005; Kumar and Thompson 2009; Urnov et al. 2010; Weinthal et al. 2010). DSBs provoked by the ZFN are primarily repaired by NHEJ and thus result in site-specific mutagenesis of a target chromosomal locus. ZFNs have been used both for this and to enhance the frequency of GT by homologous recombination (reviewed by Urnov et al. 2010).

The targeted mutagenesis strategy was first developed in Drosophila (Bibikova et al. 2002) and enhancement of gene targeting has been reported for Drosophila and in mammalian cells (Bibikova et al. 2003; Porteus and Baltimore 2003). In plants, targeted mutagenesis was demonstrated in Arabidopsis plants expressing a ZFN gene driven by a heat-shock inducible promoter as well as the ZFN target (Lloyd et al. 2005). After induction of the ZFN through heat shock, they observed cleavage at the target site and showed that repair of the break resulted in mutations at the target site at a frequency of up to 0.2 mutations per target, with the mutations transmitted to subsequent generations in 10% of the cases. Sequence analysis of the mutations showed that they were mostly simple deletions (78%) of 1–52 bp and the rest were simple insertions (13%) or deletions accompanied by insertions (8%). The approach is now well established and the Zinc Finger Consortium has made publicly available resources for the design and construction of ZFNs (Wright et al. 2006; Maeder et al. 2008; Maeder et al. 2009; Kim et al. 2010; Sander et al. 2011).

Strategies such as modular assembly, Oligomerized Pool ENgineering (OPEN) and Context-Dependent Assembly (CoDA) are currently used for construction of ZFNs. The CoDA is the most recent platform and seems to be more rapid and easy to perform (Sander et al. 2011). Several optimised assays have been established and it is now possible to routinely test ZFN function at chromosomal targets in plant cells (Hoshaw et al. 2010; Zhang and Voytas 2011). The utility of ZFNs for gene targeting in plants has also been demonstrated. A defective GUS::NPTII reporter containing a ZFN restriction site was designed and introduced at different chromosomal location in tobacco (Wright et al. 2005). After electroporation of the plants with extrachromosomal DNA encoding the ZFN and a donor DNA for repair of the defective reporter gene, the authors observed that the reporter gene was repaired by homologous recombination in about 20% of the cases, clearly demonstrating that DSBs created by ZFNs strongly induce homologous recombination at the target site. Another study yielded a similar homologous targeting efficiency of up to 10% targeted integration at the cleavage site (Cai et al. 2009).

In these studies, the targets were either exogenous genes or engineered native loci introduced into the plant together with the ZFN enzyme. Two groups have recently extended these results to confirm that ZFNs are also very effective for directly targeting endogenous plant genes (Shukla et al. 2009; Townsend et al. 2009). The Voytas group targeted the tobacco acetohydroxyacid synthase SuRA and SuRB genes, specific mutations in which can confer resistance to imidazolinone and sulphonylurea herbicides (see above). They reported the introduction of mutations within the SuR genes by ZFN-mediated gene targeting at frequencies ranging from 0.2% to 4%. In more than 40% of the recombinant plants, mutations occurred in multiple SuR alleles although a SuRB-derived construct was used (Townsend et al. 2009). Surprisingly, they observed that in more than 2% of the cases, mutations occurred at a distance more than 1.3 kb from the cleavage site. Precise targeted genome modification of Maize has also been reported (Shukla et al. 2009). These authors targeted the IPK1 gene which encodes inositol-1,3,4,5,6-pentakisphosphat 2-kinase, a crucial enzyme of phytate biosynthesis in maize seeds that leads to easily recognisable phenotype when mutated. The targeting strategy used the herbicide tolerance gene PAT. Two different donor plasmids were designed containing either an autonomous or a non-autonomous PAT expression cassette (requiring precise integration within the target sequence for expression). After codelivery of one such cassette with a ZFN, the authors obtained multiple independent targeted plants and showed that targeted insertions were inherited. Two IPK genes with 98% sequence identity exist in maize (IPK1 and IPK2), and in contrast to the tobacco study, only targeted events at the IPK1 gene were detected with four different ZFNs, suggesting a high cleavage specificity of these enzymes. ZFN targeted mutagenesis has recently been demonstrated in Arabidopsis (Zhang et al. 2010) and in the polyploid legume Soybean (Curtin et al. 2011). Targeted mutagenesis of a transgene as well as nine endogenous genes encompassing individual and duplicate gene pairs in Soybean demonstrated that ZFNs are efficient for site-directed mutagenesis of single members of a large gene family and permit discrimination between homeologous genes in a polyploid genome.

The validity of the use of engineered zinc finger nucleases for enhancing GT is thus clearly established. ZFNs represents a considerable step forward in the gene targeting field and offer a wide range of applications for functional genomics of plants with complex genomes and in genome modifications of crop plants. The construction of ZFNs is however labor-intensive and not yet routinely accessible, and this has led to the search for alternative strategies for design and construction of specific endonucleases to cleave specified DNA sequences.

11.4.3.3 TAL Effector Nucleases

DSBs generated by meganucleases and ZFNs are valuable tools to enhance homologous recombination and gene targeting. However, the complicated and long engineering process for generating new DNA target specificities has thus far restricted the generalisation of their use. Exciting recent reports describe a new class of sequence-specific nucleases named Transcription Activator-Like Effector Nucleases (TALEN) (Christian et al. 2010). TALENs are formed by fusing the catalytic domain of the FokI endonuclease to Transcription activator-like effectors (TALEs), a recently described class of specific DNA binding proteins produced by plant pathogenic bacteria of the genus Xanthomonas (Boch et al. 2009; Moscou and Bogdanove 2009). During infection, TALEs are delivered into the plant cell through the bacterial type III secretion pathway, enter the nucleus, bind to effector-specific sequences in host gene promoters and activate gene expression (Bogdanove et al. 2010). These proteins contain a nuclear localisation domain (NLS), an N-terminal translocation signal and an acidic activation domain. They are characterised by a highly conserved region consisting of up to 30 tandem repeats of a 33–35 aminoacid peptide (Fig. 11.3). Each repeat contributes to the DNA binding specificity by recognising a single DNA base pair. The sequences of the repeats are nearly identical with the exception of two adjacent aminoacids in position 12 and 13, called repeat variable di-residues (RVDs). The unique DNA binding specificity of each repeat is thus determined by these RVDs and the code ruling this correspondence has been bioinformatically and experimentally determined (Boch et al. 2009; Moscou and Bogdanove 2009; Morbitzer et al. 2010). Since the arrangement of the repeat domains is flexible, TALEs offer an invaluable array of DNA binding specificity that holds great promise for targeting of any given genomic locus. Boch and colleagues showed that randomly associated repeat domains bind to the predicted DNA target and adopting the molecular architecture used for ZFNs, the Voytas group have tested whether fusing FokI endonuclease domain to TALEs could be effective in producing DSBs at specific sites in the genome (Christian et al. 2010). Using the two known TALEs, AvrBs3 and PthXo1, they tested cleavage activity of engineered TALENs in a yeast assay and determined the optimal length of the spacer between the two DNA recognition domain, to allow FokI monomers to dimerise and cleave. The enzymes exhibit optimal cleavage with spacers of 15 bp and 21 or 24 bp, respectively and no cleavage was observed for spacers of less than 13 bp. Similar results were obtained with TALENs designed to bind and cleave the Arabidopsis *ADH1* or the Zebrafish *gridlock* genes, confirming

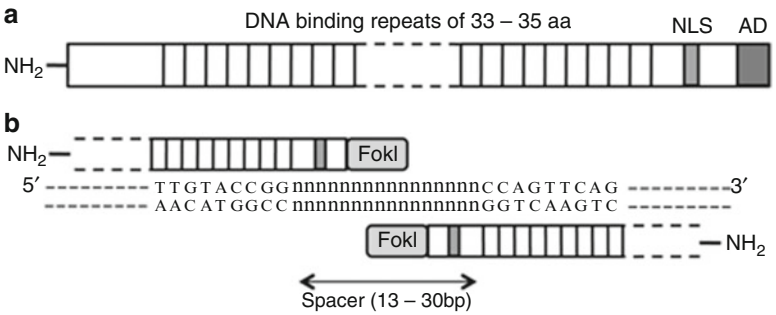


Fig. 11.3 TALE nucleases. (**a**) Transcription activator-like effectors (TALEs) are specific DNA binding proteins containing an N-terminal translocation signal, a nuclear localisation domain (NLS), and an acidic activation domain. They are characterised by DNA binding region consisting of up to 30 tandem repeats of a 33–35 aminoacid peptide. Each repeat differs only by two adjacent aminoacids in position 12 and 13, called repeat variable di-residues (RVDs), which specify the specific target nucleotide. (**b**) TALENs are formed by fusing the catalytic domain of the FokI type II restriction endonuclease to TALEs. Two chimeric TALEs will bind to their target DNA, separated by a spacer of about 15 bp, allowing FokI to dimerize and cleave the DNA target

that custom TALE-nucleases can direct DSBs to specific target sites (Christian et al. 2010). Another recent report has shown that the TALE *Hax3*, a member of the *AvrBs3* family produced by the Brassicaceae pathogen *Xanthomonas campestris*, can be used to generate a hybrid nuclease that efficiently activates target gene expression *in vivo* and efficiently cleaves target DNA *in vitro* and *in vivo* (Mahfouz et al. 2011).

These data thus confirm that TALEs are effective for enhancing gene expression and cleavage of DNA. Generalisation of these tools for gene targeting will however require engineering of TALEs allowing targeting of any given genomic sequence and extensive efforts are being invested in the conception and testing of efficient and rapid methods to assemble custom TALE repeat arrays. Zhang et al. developed a method facilitating synthesis of new TALE variants for modulating mammalian transcription (Zhang et al. 2011). They synthesised 17 TALEs and showed that they were all able to recognise, bind and modulate transcription of their target. In plants, a method to design and assemble custom TALENs has recently been provided and a complete plasmid set for making custom TALEs is now available through AddGene repository (Cermak et al. 2011). Using these methods, cleavage sites were found in nine plant genes and the activities of several engineered TALENs demonstrated by mutation of human *HPRT1* or the Arabidopsis *ADH1* genes.

The specificity of TALENs along with the relative rapidity and facility of engineering new enzymes makes them an important new addition to the arsenal of tools for targeted genome modification and potentially opens the way to the development of the long-awaited routine gene targeting methodology in plants.

Acknowledgements We would like to thank the members of the recombination group for fruitful discussions and critical reading of the manuscript and apologise to all colleagues whose valuable work was only cited through review. Work in our laboratory is supported by a European Union

research grant (FP7 KBBE-2008-227190, RECBREED), the Centre National de la Recherche Scientifique, the Université Blaise Pascal, the Université d' Auvergne and the Institut National de la Santé et de la Recherche Médicale.

References

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseeuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301:653–657

Anand A, Krichevsky A, Schornack S, Lahaye T, Tzfira T, Tang Y, Citovsky V, Mysore KS (2007a) Arabidopsis VIRE2 INTERACTING PROTEIN2 is required for Agrobacterium T-DNA integration in plants. Plant Cell 19:1695–1708

Anand A, Vaghchhipawala Z, Ryu CM, Kang L, Wang K, del-Pozo O, Martin GB, Mysore KS (2007b) Identification and characterization of plant genes involved in Agrobacterium-mediated plant transformation by virus-induced gene silencing. Mol Plant Microbe Interact 20:41–52

Barampuram S, Zhang ZJ (2011) Recent advances in plant transformation. Methods Mol Biol 701:1–35

Baur M, Potrykus I, Paszkowski J (1990) Intermolecular homologous recombination in plants. Mol Cell Biol 10:492–500

Bechtold N, Ellis J, Pelletier G (1993) In planta Agrobacterium-mediated gene transfer by infiltration of adult Arabidopsis thaliana plants. C R Acad Sci Ser III Sci Vie Life Sci 316:1194–1199

Beetham PR, Kipp PB, Sawycky XL, Arntzen CJ, May GD (1999) A tool for functional plant genomics: chimeric RNA/DNA oligonucleotides cause in vivo gene-specific mutations. Proc Natl Acad Sci U S A 96:8774–8778

Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161:1169–1175

Bibikova M, Beumer K, Trautman JK, Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. Science 300:764

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512

Bogdanove AJ, Schornack S, Lahaye T (2010) TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol 13:394–401

Brunaud V, Balzergue S, Dubreucq B, Aubourg S, Samson F, Chauvin S, Bechtold N, Cruaud C, DeRose R, Pelletier G, Lepiniec L, Caboche M, Lecharny A (2002) T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites. EMBO Rep 3:1152–1157

Cai CQ, Doyon Y, Ainley WM, Miller JC, Dekelver RC, Moehle EA, Rock JM, Lee YL, Garrison R, Schulenberg L, Blue R, Worden A, Baker L, Faraji F, Zhang L, Holmes MC, Rebar EJ, Collingwood TN, Rubin-Wilson B, Gregory PD, Urnov FD, Petolino JF (2009) Targeted transgene integration in plant cells using designed zinc finger nucleases. Plant Mol Biol 69:699–709

Capecchi MR (1989) Altering the genome by homologous recombination. Science 244:1288–1292

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39(12):e82

Charbonnel C, Gallego ME, White CI (2010) Xrcc1-dependent and Ku-dependent DNA double-strand break repair kinetics in Arabidopsis plants. Plant J 64:280–290

Charbonnel C, Allain E, Gallego ME, White CI (2011) Kinetic analysis of DNA double-strand break repair pathways in Arabidopsis. DNA Repair (Amst) 10(6):611–619

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186:757–761

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16:735–743

Crane YM, Gelvin SB (2007) RNAi-mediated gene silencing reveals involvement of Arabidopsis chromatin-related genes in Agrobacterium-mediated root transformation. Proc Natl Acad Sci U S A 104:15156–15161

Curtin SJ, Zhang F, Sander JD, Haun WJ, Starker C, Baltes NJ, Reyon D, Dahlborg EJ, Goodwin MJ, Coffman AP, Dobbs D, Joung JK, Voytas DF, Stupar RM (2011) Targeted mutagenesis of duplicated genes in soybean with zinc finger nucleases. Plant Physiol 156(2):466–473

Czako M, An G (1991) Expression of DNA coding for diphtheria toxin chain a is toxic to plant cells. Plant Physiol 95:687–692

D'Halluin K, Vanderstraeten C, Stals E, Cornelissen M, Ruiter R (2008) Homologous recombination: a basis for targeted genome optimization in crop species such as maize. Plant Biotechnol J 6:93–102

Dafny-Yelin M, Levy A, Tzfira T (2008) The ongoing saga of Agrobacterium-host interactions. Trends Plant Sci 13:102–105

De Buck S, Podevin N, Nolf J, Jacobs A, Depicker A (2009) The T-DNA integration pattern in Arabidopsis transformants is highly determined by the transformed target cell. Plant J 60:134–145

Deeken R, Engelmann JC, Efetova M, Czirjak T, Muller T, Kaiser WM, Tietz O, Krischke M, Mueller MJ, Palme K, Dandekar T, Hedrich R (2006) An integrated view of gene expression and solute profiles of Arabidopsis tumors: a genome-wide approach. Plant Cell 18:3617–3634

Deng C, Capecchi MR (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. Mol Cell Biol 12:3365–3371

Ditt RF, Kerr KF, de Figueiredo P, Delrow J, Comai L, Nester EW (2006) The Arabidopsis thaliana transcriptome in response to Agrobacterium tumefaciens. Mol Plant Microbe Interact 19:665–681

Dong C, Beetham P, Vincent K, Sharp P (2006) Oligonucleotide-directed gene repair in wheat using a transient plasmid gene repair assay system. Plant Cell Rep 25:457–465

Dujon B (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations–a review. Gene 82:91–114

Durai S, Mani M, Kandavelou K, Wu J, Porteus MH, Chandrasegaran S (2005) Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res 33:5978–5990

Emmanuel E, Yehuda E, Melamed-Bessudo C, Avivi-Ragolsky N, Levy AA (2006) The role of AtMSH2 in homologous recombination in Arabidopsis thaliana. EMBO Rep 7:100–105

Endo M, Ishikawa Y, Osakabe K, Nakayama S, Kaya H, Araki T, Shibahara K, Abe K, Ichikawa H, Valentine L, Hohn B, Toki S (2006a) Increased frequency of homologous recombination and T-DNA integration in Arabidopsis CAF-1 mutants. EMBO J 25:5579–5590

Endo M, Osakabe K, Ichikawa H, Toki S (2006b) Molecular characterization of true and ectopic gene targeting events at the acetolactate synthase gene in Arabidopsis. Plant Cell Physiol 47: 372–379

Endo M, Osakabe K, Ono K, Handa H, Shimizu T, Toki S (2007) Molecular breeding of a novel herbicide-tolerant rice by gene targeting. Plant J 52:157–166

Friesner J, Britt AB (2003) Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. Plant J 34:427–440

Fritsch O, Benvenuto G, Bowler C, Molinier J, Hohn B (2004) The INO80 protein controls homologous recombination in Arabidopsis thaliana. Mol Cell 16:479–485

Gallego ME, Sirand-Pugnet P, White CI (1999) Positive-negative selection and T-DNA stability in Arabidopsis transformation. Plant Mol Biol 39:83–93

Gallego ME, Bleuyard JY, Daoudal-Cotterell S, Jallut N, White CI (2003) Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in Arabidopsis. Plant J 35:557–565

Gelvin SB (2010) Plant proteins involved in Agrobacterium-mediated genetic transformation. Annu Rev Phytopathol 48:45–68
Gherbi H, Gallego ME, Jalut N, Lucht JM, Hohn B, White CI (2001) Homologous recombination in planta is stimulated in the absence of Rad50. EMBO Rep 2:287–291
Gheysen G, Villarroel R, Van Montagu M (1991) Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev 5:287–297
Halfter U, Morris PC, Willmitzer L (1992) Gene targeting in Arabidopsis thaliana. Mol Gen Genet 231:186–193
Hanin M, Mengiste T, Bogucki A, Paszkowski J (2000) Elevated levels of intrachromosomal homologous recombination in Arabidopsis overexpressing the MIM gene. Plant J 24:183–189
Hanin M, Volrath S, Bogucki A, Briker M, Ward E, Paszkowski J (2001) Gene targeting in Arabidopsis. Plant J 28:671–677
Hohe A, Egener T, Lucht JM, Holtorf H, Reinhard C, Schween G, Reski R (2004) An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, Physcomitrella patens. Curr Genet 44:339–347
Hohn B, Puchta H (1999) Gene therapy in plants. Proc Natl Acad Sci U S A 96:8321–8323
Hoshaw JP, Unger-Wallace E, Zhang F, Voytas DF (2010) A transient assay for monitoring zinc finger nuclease activity at endogenous plant gene targets. Methods Mol Biol 649:299–313
Houba-Hérin N, Reynolds S, Schaefer D, von Schwartzenberg K, Laloue M (1997) Molecular characterisation of homologous recombination events in the moss Physcomitrella patens. In: Petch JC, Latché A, Bouzayen M (eds) Plant Sciences, 3e Colloque général de la Société Française de Physiologie Végétale. SFPV-INRA, Toulouse, pp 22–23
Hrouda M, Paszkowski J (1994) High fidelity extrachromosomal recombination and gene targeting in plants. Mol Gen Genet 243:106–111
Iida S, Terada R (2005) Modification of endogenous natural genes by gene targeting in rice and other higher plants. Plant Mol Biol 59:205–219
Jacquier A, Dujon B (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41:383–394
Johzuka-Hisatomi Y, Terada R, Iida S (2008) Efficient transfer of base changes from a vector to the rice genome by homologous recombination: involvement of heteroduplex formation and mismatch correction. Nucleic Acids Res 36:4727–4735
Kempin SA, Liljegren SJ, Block LM, Rounsley SD, Yanofsky MF, Lam E (1997) Targeted disruption in Arabidopsis. Nature 389:802–803
Kim SI, Veena HJ, Gelvin SB (2007) Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. Plant J 51:779–91
Kim JS, Lee HJ, Carroll D (2010) Genome editing with modularly assembled zinc-finger nucleases. Nat Methods 7:91, author reply 91–2
Kipp PB, Van Eck JM, Beetham PR, May GD (2000) Gene targeting in plants via site-directed mutagenesis. Methods Mol Biol 133:213–221
Klutstein M, Shaked H, Sherman A, Avivi-Ragolsky N, Shema E, Zenvirth D, Levy AA, Simchen G (2008) Functional conservation of the yeast and Arabidopsis RAD54-like genes. Genetics 178:2389–2397
Kochevenko A, Willmitzer L (2003) Chimeric RNA/DNA oligonucleotide-based site-specific modification of the tobacco acetolactate syntase gene. Plant Physiol 132:174–184
Kumar S, Thompson WF (2009) Plant genome engineering using zinc finger nucleases. In: Jain SM, Brar DS (eds) Molecular Techniques in Crop Improvement. Springer, New York, pp 579–590
Lacroix B, Li J, Tzfira T, Citovsky V (2006) Will you let me use your nucleus? How Agrobacterium gets its T-DNA expressed in the host plant cell. Can J Physiol Pharmacol 84:333–345
Lee KY, Lund P, Lowe K, Dunsmuir P (1990) Homologous recombination in plant cells after Agrobacterium-mediated transformation. Plant Cell 2:415–425
Lee CW, Efetova M, Engelmann JC, Kramell R, Wasternack C, Ludwig-Muller J, Hedrich R, Deeken R (2009) Agrobacterium tumefaciens promotes tumor induction by modulating pathogen defense in Arabidopsis thaliana. Plant Cell 21:2948–2962

Li L, Santerre-Ayotte S, Boivin EB, Jean M, Belzile F (2004) A novel reporter for intrachromosomal homoeologous recombination in Arabidopsis thaliana. Plant J 40:1007–1015
Li J, Krichevsky A, Vaidya M, Tzfira T, Citovsky V (2005a) Uncoupling of the functions of the Arabidopsis VIP1 protein in transient and stable plant genetic transformation by Agrobacterium. Proc Natl Acad Sci U S A 102:5733–5738
Li J, Vaidya M, White C, Vainstein A, Citovsky V, Tzfira T (2005b) Involvement of KU80 in T-DNA integration in plant cells. Proc Natl Acad Sci U S A 102:19231–19236
Li L, Jean M, Belzile F (2006) The impact of sequence divergence and DNA mismatch repair on homeologous recombination in Arabidopsis. Plant J 45:908–916
Lieberman-Lazarovich M, Levy AA (2011) Homologous recombination in plants: an antireview. Methods Mol Biol 701:51–65
Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci U S A 102:2232–2237
Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Muller-Lerch F, Fu F, Pearlberg J, Gobel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, McCray PB Jr, Cathomen T, Voytas DF, Joung JK (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31: 294–301
Maeder ML, Thibodeau-Beganny S, Sander JD, Voytas DF, Joung JK (2009) Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. Nat Protoc 4:1471–1501
Magori S, Citovsky V (2011) Epigenetic control of Agrobacterium T-DNA integration. Biochim Biophys Acta 1809(8):388–394
Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A 108:2623–2628
Mayerhofer R, Koncz-Kalman Z, Nawrath C, Bakkeren G, Crameri A, Angelis K, Redei GP, Schell J, Hohn B, Koncz C (1991) T-DNA integration: a mode of illegitimate recombination in plants. EMBO J 10:697–704
Mengiste T, Paszkowski J (1999) Prospects for the precise engineering of plant genomes by homologous recombination. Biol Chem 380:749–758
Miao ZH, Lam E (1995) Targeted disruption of the TGA3 locus in Arabidopsis thaliana. Plant J 7:359–365
Molinier J, Ramos C, Fritsch O, Hohn B (2004) CENTRIN2 modulates homologous recombination and nucleotide excision repair in Arabidopsis. Plant Cell 16:1633–1643
Morbitzer R, Romer P, Boch J, Lahaye T (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A 107:21617–21622
Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501
Mysore KS, Nam J, Gelvin SB (2000) An Arabidopsis histone H2A mutant is deficient in Agrobacterium T-DNA integration. Proc Natl Acad Sci U S A 97:948–953
Nam J, Matthysse AG, Gelvin SB (1997) Differences in susceptibility of Arabidopsis ecotypes to crown gall disease may result from a deficiency in T-DNA integration. Plant Cell 9:317–333
Nam J, Mysore KS, Zheng C, Knue MK, Matthysse AG, Gelvin SB (1999) Identification of T-DNA tagged Arabidopsis mutants that are resistant to transformation by Agrobacterium. Mol Gen Genet 261:429–438
Nilsson O, Wu E, Wolfe DS, Weigel D (1998) Genetic ablation of flowers in transgenic Arabidopsis. Plant J 15:799–804
Offringa R, de Groot MJ, Haagsman HJ, Does MP, van den Elzen PJ, Hooykaas PJ (1990) Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation. EMBO J 9:3077–3084

Offringa R, Franke-van Dijk ME, De Groot MJ, van den Elzen PJ, Hooykaas PJ (1993) Nonreciprocal homologous recombination between Agrobacterium transferred DNA and a plant chromosomal locus. Proc Natl Acad Sci U S A 90:7346–7350

Oh TJ, May GD (2001) Oligonucleotide-directed plant gene targeting. Curr Opin Biotechnol 12:169–172

Okuzaki A, Toriyama K (2004) Chimeric RNA/DNA oligonucleotide-directed gene targeting in rice. Plant Cell Rep 22:509–512

Opperman R, Emmanuel E, Levy AA (2004) The effect of sequence divergence on recombination between direct repeats in Arabidopsis. Genetics 168:2207–2215

Orel N, Kyryk A, Puchta H (2003) Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. Plant J 35:604–612

Osakabe K, Abe K, Yoshioka T, Osakabe Y, Todoriki S, Ichikawa H, Hohn B, Toki S (2006) Isolation and characterization of the RAD54 gene from Arabidopsis thaliana. Plant J 48:827–842

Palmiter RD, Behringer RR, Quaife CJ, Maxwell F, Maxwell IH, Brinster RL (1987) Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene. Cell 50:435–443

Pappenheimer AM Jr (1977) Diphtheria toxin. Annu Rev Biochem 46:69–94

Paszkowski J, Baur M, Bogucki A, Potrykus I (1988) Gene targeting in plants. EMBO J 7:4021–4026

Pitzschke A, Hirt H (2010) New insights into an old story: Agrobacterium-induced tumour formation in plants by plant transformation. EMBO J 29:1021–1032

Plessis A, Perrin A, Haber JE, Dujon B (1992) Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. Genetics 130:451–460

Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763

Puchta H, Hohn B (1991) A transient assay in plant cells reveals a positive correlation between extrachromosomal recombination rates and length of homologous overlap. Nucleic Acids Res 19:2693–2700

Puchta H, Hohn B (2005) Green light for gene targeting in plants. Proc Natl Acad Sci U S A 102:11961–11962

Puchta H, Dujon B, Hohn B (1993) Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. Nucleic Acids Res 21:5034–5040

Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci U S A 93:5055–5060

Reiss B, Klemm M, Kosak H, Schell J (1996) RecA protein stimulates homologous recombination in plants. Proc Natl Acad Sci U S A 93:3094–3098

Reiss B, Schubert I, Kopchen K, Wendeler E, Schell J, Puchta H (2000) RecA stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not gene targeting, in plants transformed by Agrobacterium. Proc Natl Acad Sci U S A 97:3358–3363

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin IT, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69

Rice MC, May GD, Kipp PB, Parekh H, Kmiec EB (2000) Genetic repair of mutations in plant cell-free extracts directed by specific chimeric oligonucleotides. Plant Physiol 123:427–438

Risseeuw E, Offringa R, Franke-van Dijk ME, Hooykaas PJ (1995) Targeted recombination in plants using Agrobacterium coincides with additional rearrangements at the target locus. Plant J 7:109–119

Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B (2003) An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol 53:247–259

Rothstein R (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol 194:281–301

Saika H, Oikawa A, Matsuda F, Onodera H, Saito K, Toki S (2011) Application of gene targeting to designed mutation breeding of high-tryptophan rice. Plant Physiol 156(3):1269–1277

Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J 17:6086–6095

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, Pierick CJ, Hoffman E, Maeder ML, Khayter C, Reyon D, Dobbs D, Langenau DM, Stupar RM, Giraldez AJ, Voytas DF, Peterson RT, Yeh JR, Joung JK (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8:67–69

Schaefer DG (2001) Gene targeting in Physcomitrella patens. Curr Opin Plant Biol 4:143–150

Schaefer DG (2002) A new moss genetics: targeted mutagenesis in Physcomitrella patens. Annu Rev Plant Biol 53:477–501

Schaefer DG, Zryd JP (1997) Efficient gene targeting in the moss Physcomitrella patens. Plant J 11:1195–1206

Schaefer DG, Zryd JP (2001) The moss Physcomitrella patens, now and then. Plant Physiol 127:1430–1438

Schaefer DG, Zryd JP, Knight CD, Cove DJ (1991) Stable transformation of the moss Physcomitrella patens. Mol Gen Genet 226:418–424

Schaefer DG, Delacote F, Charlot F, Vrielynck N, Guyon-Debast A, Le Guin S, Neuhaus JM, Doutriaux MP, Nogue F (2010) RAD51 loss of function abolishes gene targeting and de-represses illegitimate integration in the moss Physcomitrella patens. DNA Repair (Amst) 9:526–533

Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, Clarke JD, Cotton D, Bullis D, Snell J, Miguel T, Hutchison D, Kimmerly B, Mitzel T, Katagiri F, Glazebrook J, Law M, Goff SA (2002) A high-throughput Arabidopsis reverse genetics system. Plant Cell 14:2985–2994

Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. Proc Natl Acad Sci U S A 102:12265–12269

Shaked H, Avivi-Ragolsky N, Levy AA (2006) Involvement of the Arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. Genetics 173:985–994

Shalev G, Sitrit Y, Avivi-Ragolski N, Lichtenstein C, Levy AA (1999) Stimulation of homologous recombination in plants by expression of the bacterial resolvase ruvC. Proc Natl Acad Sci U S A 96:7398–7402

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi VM, Rock JM, Wu YY, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD, Urnov FD (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459:437–441

Takeda S, Tadele Z, Hofmann I, Probst AV, Angelis KJ, Kaya H, Araki T, Mengiste T, Mittelsten Scheid O, Shibahara K, Scheel D, Paszkowski J (2004) BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis. Genes Dev 18:782–793

Tanaka S, Ishii C, Hatakeyama S, Inoue H (2010) High efficient gene targeting on the AGAMOUS gene in an ArabidopsisAtLIG4 mutant. Biochem Biophys Res Commun 396:289–293

Tenea GN, Spantzel J, Lee LY, Zhu Y, Lin K, Johnson SJ, Gelvin SB (2009) Overexpression of several Arabidopsis histone genes increases agrobacterium-mediated transformation and transgene expression in plants. Plant Cell 21:3350–3367

Terada R, Urawa H, Inagaki Y, Tsugane K, Iida S (2002) Efficient gene targeting by homologous recombination in rice. Nat Biotechnol 20:1030–1034

Terada R, Asao H, Iida S (2004) A large-scale Agrobacterium-mediated transformation procedure with a strong positive-negative selection for gene targeting in rice (Oryza sativa L.). Plant Cell Rep 22:653–659

Thorsness MK, Kandasamy MK, Nasrallah ME, Nasrallah JB (1991) A Brassica S-locus gene promoter targets toxic gene expression and cell death to the pistil and pollen of transgenic Nicotiana. Dev Biol 143:173–184

Thykjaer T, Finnemann J, Schauser L, Christensen L, Poulsen C, Stougaard J (1997) Gene targeting approaches using positive-negative selection and large flanking regions. Plant Mol Biol 35: 523–530

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD (1999) Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell 11:1841–1852

Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK, Voytas DF (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459:442–445

Trouiller B, Schaefer DG, Charlot F, Nogue F (2006) MSH2 is essential for the preservation of genome integrity and prevents homeologous recombination in the moss Physcomitrella patens. Nucleic Acids Res 34:232–242

Trouiller B, Charlot F, Choinard S, Schaefer DG, Nogue F (2007) Comparison of gene targeting efficiencies in two mosses suggests that it is a conserved feature of Bryophyte transformation. Biotechnol Lett 29:1591–1598

Tzfira T, Citovsky V (2006) Agrobacterium-mediated genetic transformation of plants: biology and biotechnology. Curr Opin Biotechnol 17:147–154

Tzfira T, White C (2005) Towards targeted mutagenesis and gene replacement in plants. Trends Biotechnol 23:567–569

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646

van Attikum H, Hooykaas PJ (2003) Genetic requirements for the targeted integration of Agrobacterium T-DNA in Saccharomyces cerevisiae. Nucleic Acids Res 31:826–832

van Attikum H, Bundock P, Hooykaas PJ (2001) Non-homologous end-joining proteins are required for Agrobacterium T-DNA integration. EMBO J 20:6550–6558

van Attikum H, Bundock P, Overmeer RM, Lee LY, Gelvin SB, Hooykaas PJ (2003) The Arabidopsis AtLIG4 gene is required for the repair of DNA damage, but not for the integration of Agrobacterium T-DNA. Nucleic Acids Res 31:4247–4255

Watt VM, Ingles CJ, Urdea MS, Rutter WJ (1985) Homology requirements for recombination in Escherichia coli. Proc Natl Acad Sci U S A 82:4768–4772

Wehrkamp-Richter S, Degroote F, Laffaire JB, Paul W, Perez P, Picard G (2009) Characterisation of a new reporter system allowing high throughput in planta screening for recombination events before and after controlled DNA double strand break induction. Plant Physiol Biochem 47:248–255

Weinthal D, Tovkach A, Zeevi V, Tzfira T (2010) Genome editing in plant cells by zinc finger nucleases. Trends Plant Sci 15:308–321

Wright DA, Townsend JA, Winfrey RJ Jr, Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD, Voytas DF (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 44:693–705

Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF, Joung JK (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nat Protoc 1:1637–1652

Xiaohui Wang H (2001) Positive-Negative selection for homologous recombination in Arabidopsis. Gene 272:249–255

Yamauchi T, Johzuka-Hisatomi Y, Fukada-Tanaka S, Terada R, Nakamura I, Iida S (2009) Homologous recombination-mediated knock-in targeting of the MET1a gene for a maintenance DNA methyltransferase reproducibly reveals dosage-dependent spatiotemporal gene expression in rice. Plant J 60:386–396

Yi H, Sardesai N, Fujinuma T, Chan CW, Veena HJ, Gelvin SB (2006) Constitutive expression exposes functional redundancy between the Arabidopsis histone H2A gene HTA1 and other H2A gene family members. Plant Cell 18:1575–89

Yin H, Zhang X, Liu J, Wang Y, He J, Yang T, Hong X, Yang Q, Gong Z (2009) Epigenetic regulation, somatic homologous recombination, and abscisic acid signaling are influenced by DNA polymerase epsilon mutation in Arabidopsis. Plant Cell 21:386–402

Zhang F, Voytas DF (2011) Targeted mutagenesis in Arabidopsis using zinc-finger nucleases. Methods Mol Biol 701:167–177

Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, Joung JK, Voytas DF (2010) High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc Natl Acad Sci U S A 107:12028–12033

Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29:149–153

Zheng Y, He XW, Ying YH, Lu JF, Gelvin SB, Shou HX (2009) Expression of the Arabidopsis thaliana histone gene AtHTA1 enhances rice transformation efficiency. Mol Plant 2:832–837

Zhu T, Peterson DJ, Tagliani L, St Clair G, Baszczynski CL, Bowen B (1999) Targeted manipulation of maize genes in vivo using chimeric RNA/DNA oligonucleotides. Proc Natl Acad Sci U S A 96:8768–8773

Zhu T, Mettenburg K, Peterson DJ, Tagliani L, Baszczynski CL (2000) Engineering herbicide-resistant maize using chimeric RNA/DNA oligonucleotides. Nat Biotechnol 18:555–558

Zhu Y, Nam J, Humara JM, Mysore KS, Lee LY, Cao H, Valentine L, Li J, Kaiser AD, Kopecky AL, Hwang HH, Bhattacharjee S, Rao PK, Tzfira T, Rajagopal J, Yi H, Veena HJ, Yadav BS, Crane YM, Lin K, Larcher Y, Gelvin MJ, Knue M, Ramos C, Zhao X, Davis SJ, Kim SI, Ranjith-Kumar CT, Choi YJ, Hallan VK, Chattopadhyay S, Sui X, Ziemienowicz A, Matthysse AG, Citovsky V, Hohn B, Gelvin SB (2003) Identification of Arabidopsis rat mutants. Plant Physiol 132:494–505

Ziemienowicz A, Tinland B, Bryant J, Gloeckler V, Hohn B (2000) Plant enzymes but not Agrobacterium VirD2 mediate T-DNA ligation in vitro. Mol Cell Biol 20:6317–6322

Chapter 12
Site-Directed Insertion of Transgenes in *Drosophila*

Dana Carroll

Abstract The fruit fly, *Drosophila melanogaster*, has been a popular and productive model organism for more than a century. Methods for essentially random transgene introduction developed in the 1980s have been supplemented more recently with site-specific and targeted approaches. Integrated recognition sites for PhiC31 integrase allow reproducible locations for related transgenes. Targeting through homologous recombination has been achieved by *in situ* generation of linear donor DNA. Enhanced targeting frequencies have been achieved by making a double-strand break in the target with zinc-finger nucleases (ZFNs). Relative strengths and weaknesses of these approaches are considered.

Keywords *Drosophila* • P element • PhiC31 integrase • Zinc-finger nuclease (ZFN) • Gene targeting

12.1 Introduction

We now have genomic sequences of many organisms and many individuals. As the cost of high-throughput DNA sequencing continues to drop, many more genomes will be available, including those of particular mutants. This represents a treasure trove for geneticists: simple inspection reveals information about probable gene function, phylogenetic relationships and evolutionary origins. But geneticists are hard to satisfy. They want not only to know the sequences of their favorite genomic elements, but to change them in ways chosen by them, not by nature.

D. Carroll (✉)
Department of Biochemistry, University of Utah School of Medicine,
15 N. Medical Drive East, room 4100, Salt Lake City, UT 84112-5650, USA
e-mail: dana@biochem.utah.edu

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*,
Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_12,

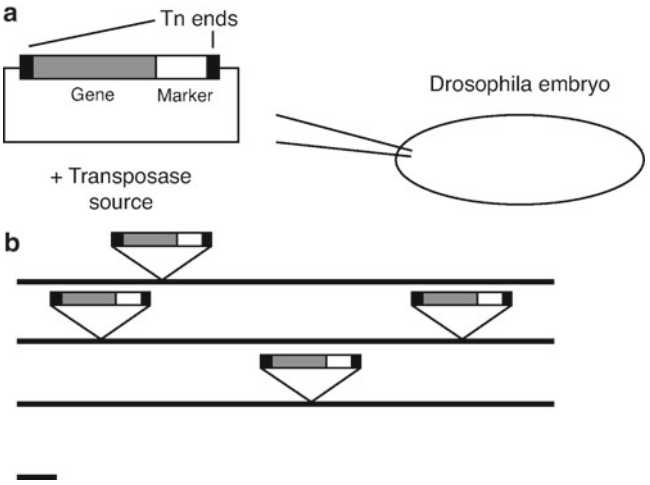


Fig. 12.1 Transgene insertion with P elements and other minimally specific transposons. (**a**) A plasmid DNA carries the gene of interest (*gray box*) between ends of the natural transposon (Tn ends) that will be recognized by the corresponding transposase. Typically there will also be a dominant marker (*white box*) inside the element to allow identification of successful transformants. The transposase source is typically a co-injected plasmid encoding the protein. The components are injected into the posterior pole of the embryo, at the site of the developing germ line. To ensure expression, the gene must carry its own regulatory sequences. (**b**) The *horizontal lines* represent the four *Drosophila* chromosomes. The various transgene locations are meant to illustrate the fact that insertion sites are essentially random and are very rare on the smallest fourth chromosome. Conditions are usually arranged so that only a single insertion is present in any one cell

The tools for directed genome manipulations in *Drosophila* originated with experiments in the early 1980s by Allan Spradling and Gerry Rubin. They showed that the natural transposable element called P could be induced to jump into the chromosomes of germ line cells by injecting cloned P DNA into early embryos (Spradling and Rubin 1982). The insertions were approximately random in the genome, and modified elements could carry a functional gene, thereby establishing a method of transgenesis that revolutionized fly genetics (Rubin and Spradling 1982) (Fig. 12.1). Among the applications that have been made of transformation by P elements, and more recently other transposons with even broader site selection, are: mutant validation by complementation, construction and analysis of new mutations, gene trapping, enhancer trapping, insertional mutagenesis, deletion creation, and many others.

An issue that complicated experiments based on nonspecific integration was a phenomenon known in *Drosophila* as position effect variegation (Levis et al. 1985; Girton and Johansen 2008). The same construct located at different positions within the genome is subject to regulation by nearby sequence elements, so transgene expression levels vary greatly depending on genomic context. This highlighted the need for methods that allow for targeted, or at least reproducible, integration.

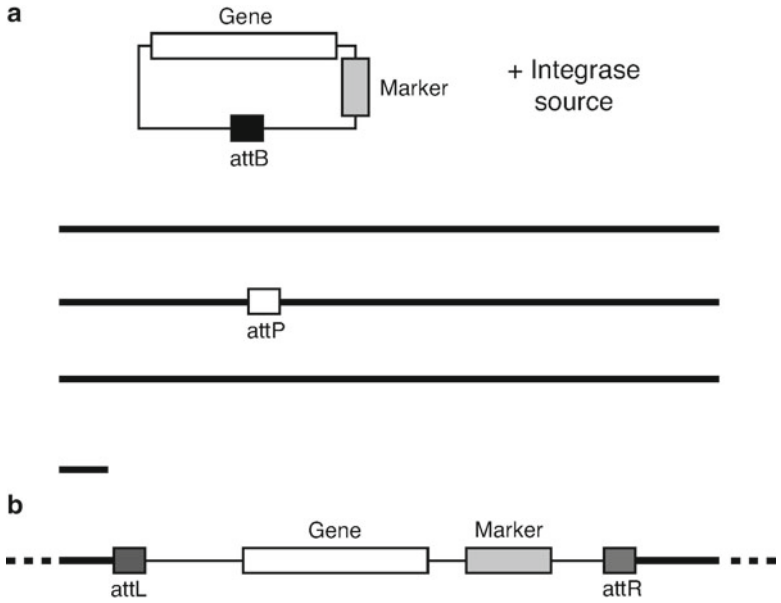


Fig. 12.2 PhiC31 integrase-mediated transgenesis. (**a**) The donor plasmid carries the gene of interest (including regulatory sequences ensuring its expression), a dominant marker gene, and the PhiC31 attB site. As in Fig. 12.1, this is injected into a recipient embryo along with a source of integrase that can be DNA, RNA or a transgene in the embryo. The recipient genome (four *horizontal lines* representing the *Drosophila* chromosomes) has a single, mapped attP site. (**b**) The product of integration has the gene of interest and the marker gene at the target site, flanked by the products of integrase mediated recombination, attL and attR

12.2 Recombinases and Integrases

Gene targeting by homologous recombination, a process in common use in fungi and in murine ES cells, was not effective in early attempts with *Drosophila*. An alternative was to insert into the genome recognition sites for site-specific recombinases, like Cre or Flp (Bischof and Basler 2008). Once a site was established, transgenes could be integrated there using appropriately designed vectors and recombinase expression. These two systems proved only modestly useful, since their recombination reactions are fully reversible, and equilibrium in most cases lies in favor of excision.

The bacteriophage PhiC31 integrase has proved much more effective (Venken and Bellen 2007). This enzyme catalyzes recombination between non-identical phage (attP) and bacterial (attB) sequences. The products of the reaction (attL and attR) are not substrates for the integrase, so excision is prevented (Fig. 12.2). Michele Calos and colleagues demonstrated that this system could be used effectively in flies (Groth et al. 2004). Vectors carrying attB sites integrate essentially exclusively at pre-introduced attP sites in the genome; there are no good pre-existing targets for the integrase.

The donor for PhiC31-mediated integration carries an attB site and a gene or construct of interest, along with a dominant marker, and is typically injected into

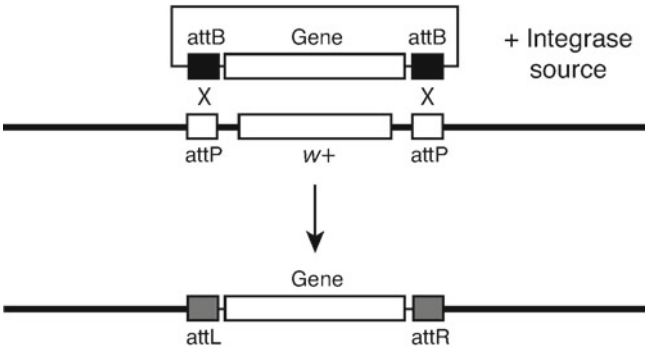


Fig. 12.3 RMCE. The genomic target is a pre-integrated element with two attP sites flanking a marker gene, in this case *w*+. The donor plasmid carries two attB sites surrounding the gene to be inserted. Two integrase-mediated recombination events lead to the replacement of the marker with the gene of interest and the conversion of the att sites to attL and attR. Successful replacements are scored by the loss of the marker. The same locus can be targeted repeatedly to introduce variants of the gene of interest at a consistent locus

embryos as done by Spradling and Rubin. The integrase protein can be expressed from a co-injected plasmid or mRNA, or can be expressed from a transgene in the recipient embryo (Bischof et al. 2007). Multiple stocks have been established with individual attP targets, or landing sites, at various positions in the genome (Bischof et al. 2007). This allows an investigator to integrate a novel transgene and its engineered variants at a unique genomic locus. Consistent position effects eliminate one source of variation in transgene expression. Interestingly, different genes may show optimal expression at different landing sites; but once a promising one has been found, it can be used repeatedly for related constructs.

An elaboration of the PhiC31 integrase approach called recombinase-mediated cassette exchange (RMCE) simplifies repeated utilization of a validated landing site (Bateman et al. 2006; Bateman and Wu 2008) (Fig. 12.3). The initial transgene is inserted on a "random" transposon, like P. It carries a marker gene – e.g., *w* + – between two attP sites. The donor element, provided along with PhiC31 integrase, has the gene of interest located between two attB sites. Two integrase-mediated recombination events cause the replacement of the marker with the gene of interest, and the desired products can be identified by loss of the dominant marker. This method has the advantage that no marker gene is left behind that might interfere with expression of the delivered construct.

12.3 Gene Targeting

Early attempts at homology-mediated gene targeting in *Drosophila* involved injection of donor DNA into embryos, but never led to the desired products. Rong and Golic devised a unique method of delivery – excision and linearization of the donor from a genomic site – that proved effective (Rong and Golic 2000) (Fig. 12.4).

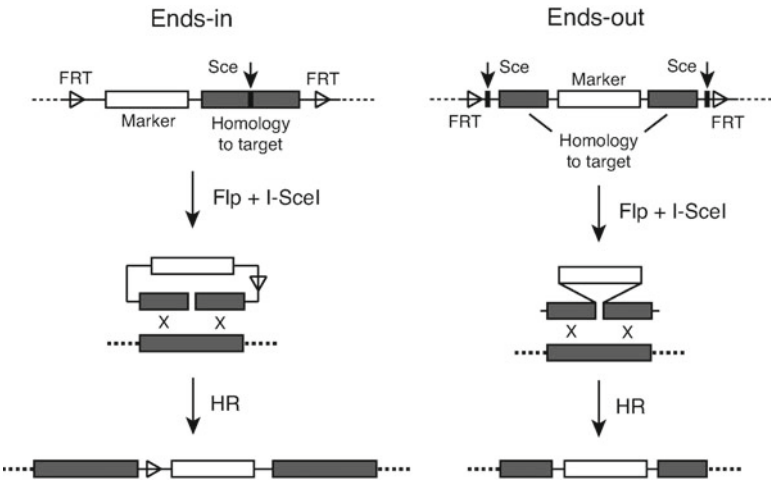


Fig. 12.4 Homologous gene targeting à la Golic. Ends-in. The original Rong and Golic procedure (Rong and Golic 2000) starts with a donor DNA integrated in the genome on a P element. Sequences homologous to the intended target (*gray box*) and a marker gene (*white box*) are flanked by recognition sites for the Flp recombinase (FRT). A recognition site for the meganuclease I-*Sce*I (Sce) is located within the target homology. Expression of Flp excises the donor as a circle, and I-*Sce* I generates a linear molecule with ends pointing in relative to the target. Homolgous recombination (HR) in both homologous segments leads to the incorporation of the donor, with the marker between duplicated copies of the homology present in the donor. Ends-out. In the variation of Gong and Golic (2003), the marker in the donor lies between sequences homologous to the target, and two Sce sites lie between those homologies and the FRTs. Excision by Flp and cleavage by I-*Sce*I generate a linear donor with ends pointing out relative to the target. HR produces a simple insertion of the marker at the target, between the homologies in the donor

The donor sequence was pre-integrated in a P element, where it was flanked by recognition sites for Flp and included a site for the rare-cutting endonuclease, I-SceI within sequences homologous to the target. Expression of the corresponding enzymes from heat-shock promoters induced excision and cleavage, producing a single copy of the linear donor in an ends-in configuration relative to the target. In initial experiments targeting the *yellow* (*y*) locus, correction of a genomic mutation was achieved in one in 500 female gametes (Rong and Golic 2000).

Ends-in targeting yields duplications as products (Fig. 12.4). These can be reduced to single copies by expression of a second endonuclease (e.g., I-*Cre*I), but this requires additional crosses and screening (Rong et al. 2002). A variation on the original technique provided the donor in an ends-out configuration relative to the target (Gong and Golic 2003) (Fig. 12.4). This yields gene replacements, rather than duplications. Using both ends-in and ends-out methods, many fly genes have been successfully targeted in many laboratories. The frequencies are low – up to 0.3%, but typically less, and in some cases not successful – and require inclusion of a dominant marker or very extensive screening to isolate the desired products (Rong et al. 2002). Remarkably, however, very large deletions (up to 47 kb) have been created without severe loss of efficiency (Gong and Golic 2004).

12.4 Break-Mediated Gene Targeting

The Rong and Golic procedure enhanced gene targeting by delivering the donor DNA in an effective format. Frequencies were still limited, however, by the inertness of the genomic target. Many studies in many organisms demonstrated that a double-strand break (DSB) stimulates recombination dramatically. The challenge then was to develop reagents that could be designed to cleave any desired sequence. Zinc-finger nucleases (ZFNs) have this property (Carroll 2011), and *Drosophila* was the first organism in which their capabilities were demonstrated (Bibikova et al. 2002, 2003).

The DNA-recognition domain of ZFNs consists of Cys_2His_2 zinc fingers that bind DNA in a very modular fashion. Combinations of fingers can be designed or selected to bind specific DNA sequences with good specificity. The cleavage domain, from the restriction enzyme *Fok*I, must dimerize to be active, so two ZFN monomers are required for one target (Fig. 12.5).

Initial experiments in *Drosophila* relied on heat shock-induced expression of the ZFNs from randomly integrated transgenes (Beumer et al. 2006; Bibikova et al. 2002, 2003). The ends-out donor was generated by the method of Rong and Golic. Several targets were hit successfully, with gene replacements representing as much as 15% of all gametes in the best case (Beumer et al. 2006). Subsequently, embryo injection of ZFN mRNAs with donor DNA was shown to be an equally effective and much simpler procedure (Beumer et al. 2008).

ZFN-induced gene targeting depends on cellular machinery for homologous repair of the DSB. A parallel pathway for DSB repair is nonhomologous end joining (NHEJ), in which ends at the break are rejoined without regard to homology, often accompanied by small insertions and deletions at the junction. NHEJ can be considered a competing pathway to the desired homologous recombination (HR), but in some circumstances, mutations at the target may be a very desirable outcome. Insertions and deletions in coding sequence often produce frameshifts and yield knockout mutations. This has been accomplished with ZFNs in genes that previously had no known null alleles (Beumer et al. 2008).

The balance between HR and NHEJ depends on a number of factors, including the activities required for each type of repair. In the absence of the NHEJ-specific DNA ligase IV, repair of ZFN-induced breaks is shifted strongly toward HR, while mutation of Rad51 has the opposite effect (Beumer et al. 2008; Bozas et al. 2009).

Other targetable DNA cleavage reagents should have similar utility in *Drosophila*. Meganucleases (also called homing endonucleases) have been engineered to novel specificities and used to make targeted genomic modifications in a variety of cells and organisms (Silva et al. 2011; Stoddard 2011). The newly discovered TALE DNA-recognition domain has been engineered into ZFN-like nucleases, called TALENs. The TALE building block is a 34-amino acid module that recognizes a single DNA base pair (Scholze and Boch 2011). Consequently, these domains are potentially more readily designable for new targets.

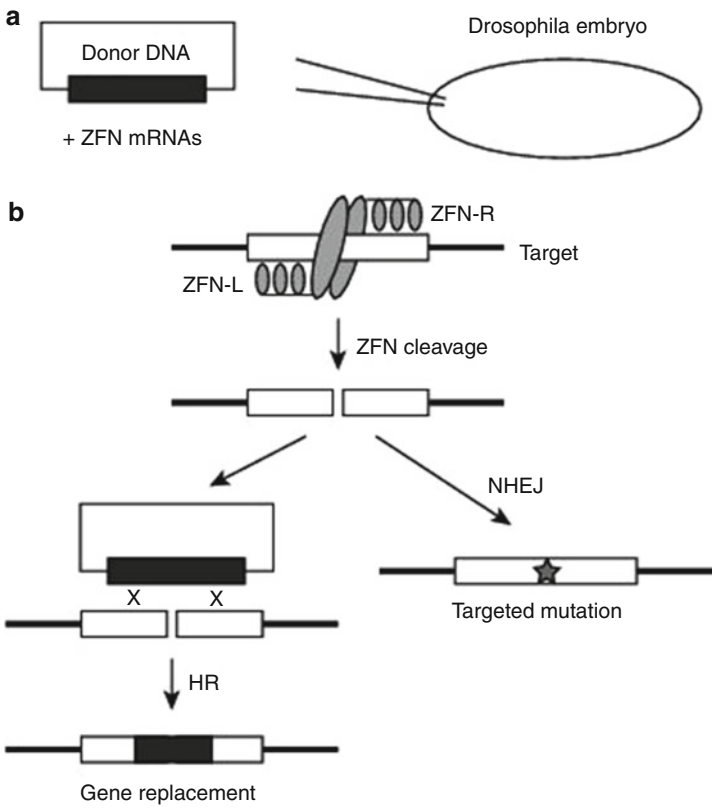


Fig. 12.5 Targeting with ZFNs. (**a**) A donor DNA with homology to the target is injected into embryos along with synthetic mRNAs for two ZFNs designed to cleave the target. (**b**) Each ZFN has three (or more) zinc fingers (*small gray ovals*) that bind neighboring sequences in the target. This brings the *Fok*I cleavage domains (*large gray oval*) linked to each set of fingers into proximity and allows dimerization and cleavage. Repair of the break by HR using the donor as template leads to replacement of target sequences with those from the donor. An alternative repair pathway (NHEJ) can produce mutations at the target

12.5 Comparison of Approaches

Homology-dependent gene targeting has the powerful advantage that the modified sequence lies in its natural genomic context. Both proximal and distant elements that influence the developmental timing and level of its expression are intact. In contrast, genes that are delivered by PhiC31 integrase must carry their own regulatory sequences in the donor construct. Typically this means that only proximal or artificial elements are included. The use of bacterial recombineering to generate large donor DNAs makes it possible to incorporate very extensive flanking regions and reporter constructs with the transgene (Venken et al. 2006).

Two groups have developed methods that combine the use of homologous recombination with PhiC31 integrase-mediated site-specific recombination (Gao et al. 2008; Huang et al. 2009). In the first step, a Golic targeting procedure (ends-in or ends-out) is used to establish an attP site within or in place of a chosen (rather than random) locus. That site is then used to insert a modified version of the target gene using an attB-carrying donor DNA (as in Fig. 12.2). The procedures are rather complicated, requiring elimination steps after each insertion, and have quite low overall efficiencies. Once the specific attP site is established, however, it can be used to introduce many different alleles of the target gene. These procedures have the advantage over the use of random attP landing sites that all the normal regulatory sequences are in place and need not be included in the donor. Use of break-induced targeting in the first step could potentially improve the efficiency of the process quite dramatically.

Break-mediated gene targeting is much more efficient than simply providing the donor DNA à la Golic. In many cases no marker gene or known phenotype is necessary; the frequency is high enough that the desired products can be identified and isolated by molecular analysis. The disadvantage of this method is that new cleavage reagents must be produced for each new target, and there is no guarantee that any pair of designed ZFNs, for example, will actually work. In our lab we have successfully targeted five of the ten genes we have attempted, but others report higher failure rates in other systems (Ramirez et al. 2008; Kim et al. 2010; Joung et al. 2010). It seems possible that TALEN designs will show a higher success rate.

Until better rules are defined for the design of cleavage reagents, break-mediated approaches may be most useful for two purposes: simple knockouts and repeated targeting of a single locus. The former requires no donor DNA and proceeds naturally during the repair of the break by NHEJ (Fig. 12.5). Once effective cleavage reagents have been made for a given target, they can be used repeatedly to introduce various sequence modifications. This allows genetic fine structure analysis by creation and analysis of mutations distributed throughout a locus. We are fortunate in *Drosophila* that many genes are limited in extent (few large introns) and conversion tracts (the extent of incorporation of sequences from the donor) are long (Gloor et al. 1991; Nassif et al. 1994) (K. Beumer and D. Carroll, unpublished results).

The Golic procedure requires only the construction of a suitable donor DNA for any new target. A drawback is that the donor must first be incorporated into the genome of flies carrying heat-inducible genes for Flp and I-*Sce*I. Because the absolute frequency of targeting is low, a dominant marker is required in the donor construct, unless the genomic product itself generates a predictable and scorable phenotype. The marker can be removed in subsequent steps, but this extends the time required to produce and verify the desired alteration. Break-mediated targeting can be accomplished with straightforward injection procedures.

Despite the introduction of many additional eukaryotic model organisms, *Drosophila* continues to be a very popular subject for the study of gene function. This is due in part to the adoption of the tools for directed genetic manipulation described here.

Note Added in Proof: As suggested in the text, TALENs are proving to be a very effective alternative to ZFNs in a number of systems, including *Drosophila* (unpublished results).

References

Bateman JR, Wu C-T (2008) A simple polymerase chain reaction-based method for the construction of recombinase-mediated cassette exchange donor vectors. Genetics 180:1763–1766

Bateman JR, Lee AM, Wu C-T (2006) Site-specific transformation of *Drosophila* via PhiC31 integrase-mediate cassette exchange. Genetics 173:769–777

Beumer K, Bhattacharyya G, Bibikova M, Trautman JK, Carroll D (2006) Efficient gene targeting in *Drosophila* with zinc finger nucleases. Genetics 172:2391–2403

Beumer KJ, Trautman JK, Bozas A, Liu J-L, Rutter J, Gall JG, Carroll D (2008) Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A 105:19821–19826

Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. Genetics 161:1169–1175

Bibikova M, Beumer K, Trautman JK, Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. Science 300:764

Bischof J, Basler K (2008) Recombinases and their use in gene activation, gene inactivation, and transgenesis. In: Dahmann C (ed) Drosophila: methods and protocols. Humana, Totowa, NJ

Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific PhiC31 integrases. Proc Natl Acad Sci U S A 104:3312–3317

Bozas A, Beumer KJ, Trautman JK, Carroll D (2009) Genetic analysis of zinc-finger nuclease-induced gene targeting in *Drosophila*. Genetics 182:641–651

Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics 188:773–782

Gao G, McMahon C, Chen J, Rong YS (2008) A powerful method combining homologous recombination and site-specific recombination for targeted mutagenesis in *Drosophila*. Proc Natl Acad Sci U S A 105:13999–14004

Girton JR, Johansen KM (2008) Chromatin structure and the regulation of gene expression: the lessons of PEV in *Drosophila*. Adv Genet 61:1–43

Gloor GB, Nassif NA, Johnson-Schlitz DM, Preston CR, Engels WR (1991) Targeted gene replacement in *Drosophila* via P element-induced gap repair. Science 253:1110–1117

Gong WJ, Golic KG (2003) Ends-out, or replacement, gene targeting in *Drosophila*. Proc Natl Acad Sci U S A 100:2556–2561

Gong WJ, Golic KG (2004) Genomic deletions of the *Drosophila melanogaster Hsp70* genes. Genetics 168:1467–1476

Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. Genetics 166:1775–1782

Huang J, Zhou W, Dong W, Watson AM, Hong Y (2009) Directed, efficient, and versatile modifications of the *Drosophila* genome by genomic engineering. Proc Natl Acad Sci U S A 106:8284–8289

Joung JK, Voytas DF, Cathomen T (2010) Reply to "Genome editing with modularly assembled zinc-finger nucleases". Nat Methods 7:91–92

Kim J-S, Lee HJ, Carroll D (2010) Genome editing with modularly assembled zinc-finger nucleases. Nat Methods 7:91

Levis R, Hazelrigg T, Rubin GM (1985) Effects of genomic position on the expression of transduced copies of the white gene of *Drosophila*. Science 229:558–561

Nassif NA, Penney J, Pal S, Engels WR, Gloor GB (1994) Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol 14:1613–1625

Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA, Cathomen T, Voytas DF, Joung JK (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods 5:374–375

Rong YS, Golic KG (2000) Gene targeting by homologous recombination in *Drosophila*. Science 288:2013–2018

Rong YS, Titen SW, Xie HB, Golic MM, Bastiani M, Bandyopadhyay P, Olivera BM, Brodsky M, Rubin GM, Golic KG (2002) Targeted mutagenesis by homologous recombination in *D. melanogaster*. Genes Dev 16:1568–1581

Rubin GM, Spradling AC (1982) Genetic transformation of *Drosophila* with transposable element vectors. Science 218:348–353

Scholze H, Boch J (2011) TAL effectors are remote controls for gene activation. Curr Opin Microbiol 14:47–53

Silva G, Poirot L, Galetto R, Smith J, Montoya G, Duchateau P, Pâques F (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. Curr Gene Ther 11:11–27

Spradling AC, Rubin GM (1982) Transposition of cloned P elements into *Drosophila* germ line chromosomes. Science 218:341–347

Stoddard BL (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modifications. Structure 19:7–15

Venken KJT, Bellen HJ (2007) Transgenesis upgrades for *Drosophila melanogaster*. Development 134:3571–3584

Venken KJ, He Y, Hoskins RA, Bellen HJ (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. Science 314:1747–1751

Chapter 13
Nuclease Mediated Targeted Genome Modification in Mammalian Cells

Jenny C. Barker[1], Richard A. Voit[1], and Matthew H. Porteus

Abstract The development of targeted genome modification by using engineered nucleases has been a rapidly developing field in the last few years. In this chapter we review the different types of nucleases that can be engineered for this purpose, the different types of genome modifications that have been created, and examples of targeted genome modification used for both research purposes and for potential gene therapy uses.

Keywords Genome Engineering • Targeted Genome Modification • Gene Therapy • Homologous Recombination • Non-Homologous End-Joining • Homing Endonuclease • Meganuclease • TAL Effector Nuclease • Zinc Finger Nuclease

13.1 Introduction

The fundamental basis of medicine is to provide therapy for human diseases and a major aspect of therapy is to provide small molecule compounds to do this. These small molecule based therapies have had a dramatic impact on the human condition and provide curative therapy for many diseases. Nonetheless, for the preponderance of diseases, small molecule based therapy provides symptomatic rather than curative relief. In particular, for diseases caused by genetic mutations, small molecule based

[1]co-first authors

J.C. Barker • R.A. Voit
Medical Scientist Training Program, UT Southwestern Medical Center, Dallas, TX, USA

M.H. Porteus (✉)
Department of Pediatrics, Stanford University, Stanford, CA, USA
e-mail: mporteus@stanford.edu

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*, Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_13,

therapy provides treatment for the consequences of the underlying genetic defect rather than directly addressing the root cause. In the last decades, however, new classes of therapeutics have either been developed or are being developed that attempt to treat genetic diseases at their root cause including the use of protein replacement therapy, bone marrow transplantation, and gene therapy.

Enzyme replacement therapy for the lysosomal storage diseases and for severe combined immunodeficiency caused by mutations in the adenosine deaminase (ADA) gene, and clotting factor replacement for hemophilia are examples of protein therapeutics in which protein is administered directly to replace a missing protein. Protein replacement therapy provides life-altering treatment for patients for which it is available, but it does not cure the disease as the underlying genetic defect remains and is also limited in both cost and effectiveness. The high cost of protein therapy (hundreds of thousands of dollars per year per patient) means this type of therapy is not available to patients from resource poor parts of the world. Finally, despite attempts at prolonging the half-life of therapeutic proteins, protein replacement therapy requires life-long repeated infusions at daily, weekly, or monthly intervals.

Bone marrow transplantation in which the patient's bone marrow is replaced with the bone marrow from a person who does not have the disease is a highly effective therapy for patients with genetic diseases of the blood such as sickle cell disease, thalassemia and SCID and is even effective in non-blood genetic diseases such as adrenoleukodystrophy and Hurler's syndrome. Bone marrow transplantation replaces the hematopoietic system containing the disease-causing mutation with a hematopoietic system free of the genetic defect. Bone marrow transplantation remains the only broadly used stem cell therapy to treat patients. However, the limitations of stem cell therapy by bone marrow transplantation are several-fold. First, like protein therapeutics, bone marrow transplantation is a highly technical and expensive procedure and so is not available in many parts of the world. Second, many patients do not have a suitable immunologically matched donor. Third, there can be severe long-term side-effects such as graft versus host disease and the consequences of myeloablative conditioning therapy.

An alternative to protein replacement therapy or allogeneic stem cell therapy is to use genome engineering to modify the patient's own cells and use those gene modified cells as treatment for the disease ("gene therapy"). The modification of the patient's cells can be done *in vivo* in which the modification takes place without the cells leaving body or *ex vivo* in which the cellular modification takes place outside the body and then the cells are transplanted back into the patient. Unlike small molecule therapy or protein replacement therapy, this strategy is directed at the root cause of the disease and if successful, cure could be achieved with a single intervention. Unlike allogeneic bone marrow transplantation, this approach uses the patient's own cells and so the problems of finding matched donors and graft versus host disease are solved.

There have been several gene therapy clinical trials for genetic diseases of the blood including for SCID-X1, ADA-SCID, Wiskot-Aldrich syndrome, chronic granulomatous disease, and thalassemia (Abuljadayel et al. 2006; Aiuti et al. 2009; Bank et al. 2005; Boztug et al. 2010; Cavazzana-Calvo et al. 2000, 2005; Gaspar et al. 2006;

Kang et al. 2010; Malech et al. 1997). In each of these trials, the patient's bone marrow cells have been removed, transduced with a retroviral or lentiviral vector that carries a wild-type copy of the mutated gene and then transplanted back into the patient. Tens of patients of worldwide have benefited from this therapy. A fundamental feature of this viral based strategy is that the transgene integrates in an uncontrolled fashion. An unfortunate consequence of this uncontrolled integration is that the integration can activate a latent oncogene and this has occurred in a number of patients, causing frank leukemia, myelodysplasia, or a clonal proliferation (Hacein-Bey-Abina et al. 2003, 2008; Ott et al. 2006). Overall, the frequency of this event seems to be greater than 10%, although there are vector and disease specific differences, and this relatively high frequency may make this an unsustainable approach for the treatment of genetic diseases of the blood.

Targeted genome engineering, a technology that enables site-specific, controlled insertion of a transgene, allows for the benefits of transgene technology while avoiding the complications of random integration. Targeted genome engineering strategies in human cells have been developed to stimulate sequence specific DNA modifications to achieve one of three therapeutic endpoints. First, the targeted disruption of an endogenous locus can provide a rapid, efficient means for researchers to achieve permanent gene knockouts in human cells, a powerful research tool that has heretofore been restricted to model organisms. In addition, the ability to disrupt a gene at its endogenous locus provides a new therapeutic approach in which targeted gene mutation can be used to treat a disease. The second use of targeted genome engineering is the direct correction of disease-causing mutations at endogenous loci. Third, targeted genome modification can be used to target transgene addition to precise locations in the genome, including "safe harbors." Such targeted transgene addition has important ramifications both therapeutically, to eliminate insertional oncogenesis, and as a scientific tool. With such potentially precise methods of genomic manipulation available, researchers now have the ability to begin to address the root cause of many of the more than 10,000 monogenic diseases.

13.2 Methods for Targeted Genome Engineering in Human Cells

While AAV vectors can be used for targeted gene addition and gene correction (Khan et al., 2010, 2011), the focus of this chapter will be on using site-specific nucleases to perform targeted gene mutation, gene correction, and targeted gene addition. Central to targeted genome engineering is the ability to design highly specific nucleases. There are now three general classes of sequence specific nucleases: (1) homing endonucleases; (2) zinc finger nucleases (ZFNs) and (3) TAL effector nucleases (TALENs). ZFNs and TALENs share a common general structure in which an engineered DNA binding domain is fused to the nuclease domain derived from the FokI restriction enzyme. The genome specificity of ZFNs and TALENs derives from both the sequence specificity of the DNA binding domain and the need for the

FokI nuclease domain to dimerize (Bitinaite et al. 1998) in order to create a DNA double-strand break (DSB) at a specific target site. The DSB is the initiating event in generating targeted genome modifications.

Most natural DSBs are the result of ongoing environmental stresses such as ionizing radiation and metabolically derived reactive oxygen species. Unrepaired DSBs lead to cell death and multiple, redundant cellular DNA repair pathways have evolved to repair these important genetic lesions. The two most important DSB repair mechanisms are non-homologous end joining (NHEJ) and homologous recombination (HR) (Fig. 13.1). In NHEJ, the DNA ends created by a DSB are ligated back to one another. Often, this process occurs without altering the DNA sequence near the break site, but sometimes repair by NHEJ introduces insertions or deletions at the site of the break ("mutagenic NHEJ"). These insertions or deletions, if they occur in the coding region of a gene, can result in frameshift mutations creating a non-functional gene product. If the DSB is created by an engineered nuclease and that DSB is repaired by mutagenic NHEJ, inactivating gene specific frameshift mutations can be generated. Thus, by combining the engineering of gene specific nuclease with the intrinsic mutagenic property of NHEJ, researchers can create targeted gene mutations at endogenous loci. In this method of targeted genome modification, the site of the mutation is controlled by the specificity of the nuclease but the specific genomic modification at that site is uncontrolled. Depending on the activity and expression level of the nuclease, targeted gene mutation frequencies of greater than 50% can be generated (Perez et al. 2008).

HR is a higher-fidelity repair pathway in which regions of homologous DNA serve as a template to accurately repair DNA damage. In a normal setting in dividing cells, repair of a DSB proceeds using the sister chromatid as the homologous template, in a process that fixes the break without altering the genomic sequence. In the experimental setting, if an extra-chromosomal piece of donor DNA with homology to the break site is supplied, the same HR machinery can use the exogenous

Fig. 13.1 (continued) large chromosomal rearrangements (**c**) by ligating two distal chromosomal segments with simultaneous DSB formation. The pathway mediators for NHEJ include Ku70 and Ku80 which promote end binding, DNA-PKcs that promote end juxtaposition and cell signaling and the nuclease, Artemis, that is responsible for end processing. One of a few polymerases, including TdT, pol mu or pol lambda participate in end processing as well, and finally, DNA Ligase IV, XRCC4 and XLF complete re-ligation of the break. Homologous recombination (HR) is a second pathway by which DSBs can be repaired. This pathway relies upon the use of homologous DNA as a template for re-synthesis of the damaged strand. Typically, the sister chromatid is the homologous template utilized and this results in non-mutagenic repair (**d**). However, for the purposes of genome engineering, an exogenously introduced plasmid with regions of homology can also be used as a template for repair. This can result in correction of a gene (**e**) if a "donor plasmid" contains the sequence necessary to correct an endogenous point mutation. Alternatively, an entirely new gene can be inserted into the genome in a site-specific manner (**f**), so long as the gene of interest is included between regions of homology to a DSB target locus. The pathway of HR involves (1) binding of a DSB (2) end resection that leaves 3′ overhangs, both of which are mediated by the MRN complex (Rad50/Mre11/Nbs1) (3) homology search and strand invasion mediated by Rad51 (4) D-loop formation and DNA synthesis from the homologous template (5) resolution of the junction, gap filling and ligation

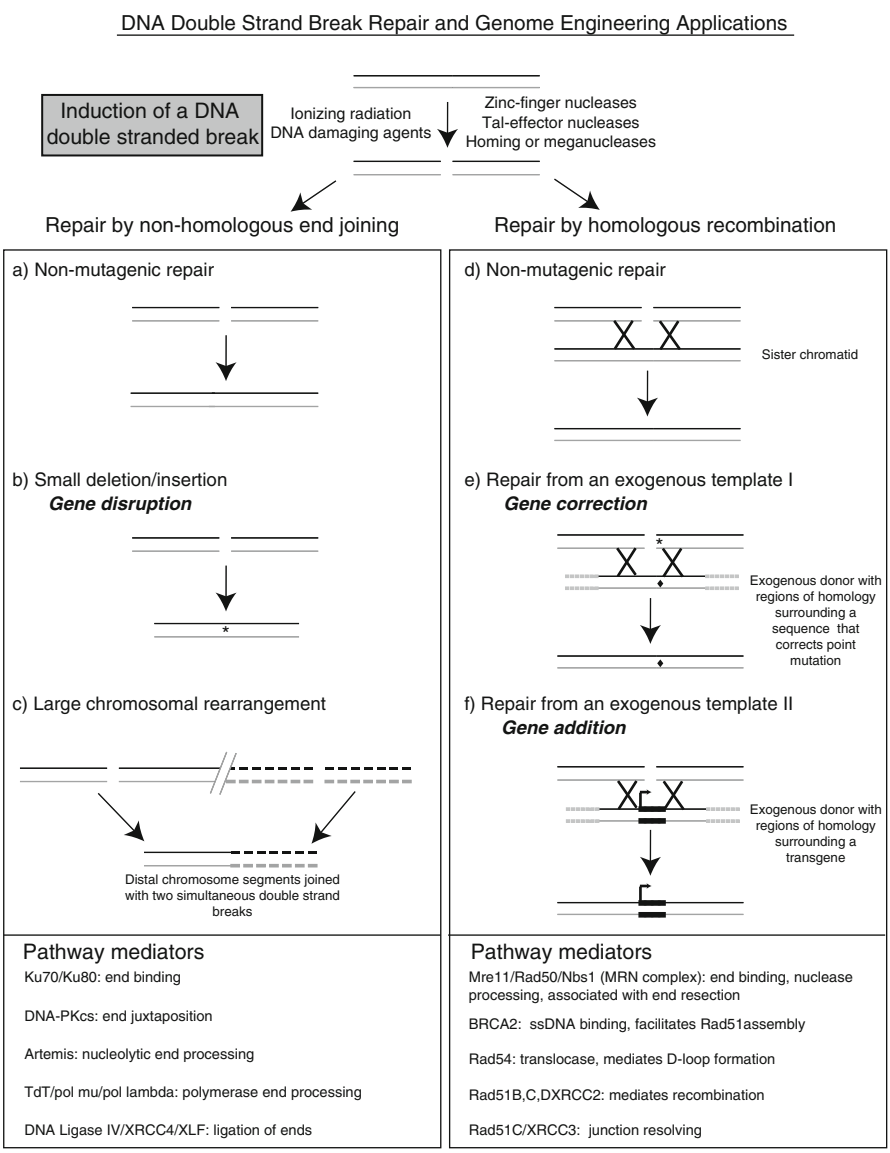


Fig. 13.1 DNA double strand break repair and genome engineering applications**.** Genome engineering makes use of a cell's endogenous DNA repair machinery to effect site-specific genomic modifications. DNA double strand breaks (DSBs) can be induced by ionizing radiation or DNA damaging agents in a natural setting, and also by artificial nucleases, such as zinc-finger nucleases, TAL-effector nucleases or meganucleases (homing endonucleases) for the purposes of genome engineering. After a break is created, repair of the break can proceed down one of two pathways. Non-homologous end-joining (NHEJ) re-ligates DNA ends in either a non-mutagenic fashion (**a**) or in a manner that creates small deletions (**b**). By harnessing the ability of NHEJ to create small deletions, disruption of a gene can be intentionally achieved. It has also been demonstrated that NHEJ is capable of inducing

DNA to repair the DSB. Through careful design of the exogenous template, a disease-causing point mutation can be changed to the wild type base, leading to reversion of the mutation when the donor template is used in HR ("gene correction"). Using an alternative donor template design in which a single gene or even cassette of genes is flanked by homology arms on each side of the DSB, gene targeting by HR can occur in which the entire cassette between the arms is inserted precisely into the genome at the site of the nuclease induced DSB ("targeted gene addition"). In gene targeting by HR, not only is the site of the genome modification determined but by harnessing HR to repair the induced DSB using an experimentally provided donor template, the exact sequence of the genome modification is also determined. Depending on the activity of the engineered nuclease, gene correction rates and targeted gene addition rates of greater than 20% can be achieved (Benabdallah et al. 2010; Hockemeyer et al. 2009; Urnov et al. 2005).

13.2.1 Zinc-Finger Nucleases

Zinc-finger nucleases consist of a zinc finger DNA binding domain fused to the non-specific nuclease domain derived from the type IIS FokI restriction enzyme (Bitinaite et al. 1998; Kim and Chandrasegaran 1994). Zinc fingers have a conserved ββα structure and the α helix fits into the major groove of DNA, interacting with three nucleotides. The first six amino acids of the α helix, along with the immediately preceding amino acid, give the zinc finger its nucleotide specificity. Thus, the overall DNA recognition motif for a zinc finger domain is seven amino acids in total. By stringing multiple zinc fingers in series, proteins can be designed with 9- or 12-base pair (or longer) recognition sequences. When these zinc finger motifs are fused at their C-termini to the FokI nuclease domain, the resulting chimeric protein is known as a zinc finger nuclease (ZFN) (Fig. 13.2). Study of the activity of the wild type FokI endonuclease revealed that efficient cleavage is mediated by the dimerization of nuclease domains (Bitinaite et al. 1998). The dimerization requirement for efficient activity is maintained for chimeric nucleases including ZFNs. For ZFNs, dimerization is achieved by having two different ZFNs bind in an inverted orientation on opposite DNA strands thereby allowing the nuclease domain to dimerize and cut the DNA in the spacer region between the two ZFN binding sites (Bibikova et al. 2001). Each individual ZFN has a 9- or 12- target binding site, a length not sufficient to confer genome specificity, but a pair of ZFNs has an 18- or 24-base pair binding site, a length that is sufficient to confer genome specificity.

There are several approaches to the design of ZFNs. The simplest is to use modular-assembly in which individual fingers are strung together in an array of three or more. This approach is relatively straightforward and has been useful in several different studies (Beerli et al. 2000; Dreier et al. 2001, 2005). The major disadvantage is that it has a low success rate when target sites do not fit a certain type of sequence and even with potentially ideal target sites, the zinc finger nucleases made using this strategy have lower activity and higher off-target effects than proteins made using

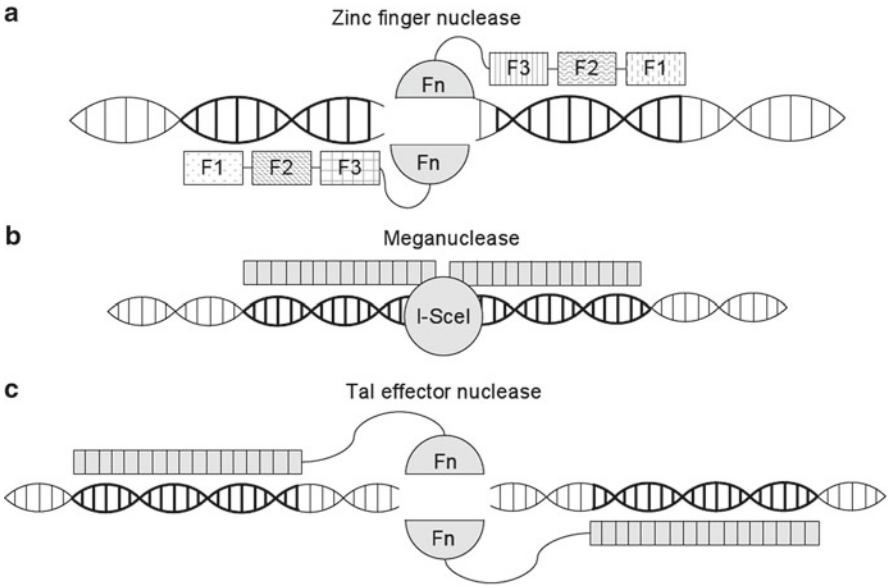


Fig. 13.2 Induction of DNA double strand breaks by three classes of chimeric nucleases. (**a**) The DNA binding domain of zinc finger nucleases (ZFNs) is discrete from the nuclease domain and consists of arrays of three or four zinc fingers, each of which binds three nucleotides. When two ZFNs bind recognition sequences separated by 5–6 nucleotides, the nuclease domains dimerize to form an active enzyme and stimulate DNA cleavage. (**b**) Meganucleses such as I-SceI have long (12–40 base pair) DNA recognition sequences and bind and cleave DNA as monomers. (**c**) Tal effector nucleases (TALENs) bind DNA using a series of 34 amino acid repeats, each containing two key residues that mediate binding to one nucleotide. Arrays of 11–17 (or more) repeats are fused to a nuclease domain which must dimerize with another TALEN separated by a spacer of 12–23 nucleotides

other strategies (Pruett-Miller et al. 2008; Ramirez et al. 2008). A similar strategy is to use modular-assembly to string together two-finger modules into an array (Perez et al. 2008; Urnov et al. 2005). This strategy was first developed by Gendaq, then bought by Sangamo Biosciences, and now sold to the research community through Sigma-Aldrich Pharmaceuticals. Because the method has not been publicly described, the overall success rate has not been published but the strategy has had a high success rate for targeting multiple genes. This method is based on a proprietary archive of two-finger modules and thus is not freely available to researchers. A derivative of this approach is to assemble a ZFN using two-finger modules and then to subsequently refine the original protein for improved properties ("Sangamo approach"). It is through this method that the best published ZFNs from Sangamo Biosciences have been made but is also not an approach that is freely available to others. An alternative approach to assembling pre-determined modules, is to use selection strategies, either by phage display or using bacterial based methods (including the OPEN method), to select for zinc finger domains that bind the desired target sequence out of a randomized library of proteins (Maeder et al. 2008; Meng et al.

2008; Durai et al. 2006; Rebar et al. 1996). These selection approaches are difficult to use but do result in nucleases with properties of high activity and lower off-target effects. Interestingly, ZFNs made using the Sangamo approach and using the OPEN approach have been compared for their off-target effects and both methods seem to generate relatively specific ZFNs but they achieve the specificity through different mechanisms (Pattanayak et al. 2011; Gabriel et al. 2011). Finally, there are two derivatives of the OPEN approach to making ZFNs. In context dependent assembly (CoDA), modular assembly is used to assemble ZFNs using zinc finger modules that have been previously identified using the OPEN system (Sander et al. 2011). In a hybrid approach, individual modules can be combined with the OPEN system to generate proteins that have high activity and relatively low toxicity (Wilson and Porteus, manuscript submitted).

13.2.2 Homing Endonucleases and TAL Effector Nucleases (TALENs)

Zinc finger nucleases have had the widest application in targeted genome modifications, but other classes of nucleases are also being developed to stimulate targeted DNA double strand breaks to achieve targeted genome modification. The first of these classes includes meganucleases, also called homing endonucleases. The major family of meganucleases, called LAGLIDADG endonucleases because of the presence of a conserved amino acid motif, are derived from the mitochondria and chloroplasts of eukaryotic unicellular organisms such as yeast (Paques and Duchateau 2007). As their name suggests, meganucleases have very long DNA recognition sites (between 12 and 40 base pairs), allowing them to bind DNA with very high specificity. The first example of a meganuclease used in gene targeting is I-SceI. In 1994, Maria Jasin and colleagues stably integrated two I-SceI sites into mouse chromosomes and demonstrated cleavage at one or both of those sites by transient expression of I-SceI (Rouet et al. 1994). To use meganucleases in targeted genome modification, however, the nuclease must be re-engineered to recognize a new target site. Several different academic groups and companies (Cellectis and Precision Biosciences) have had some success with the re-engineering of meganucleases to recognize new target sites for the purposes of genome engineering using targeted genome modification (Arnould et al. 2006, 2011; Grizot et al. 2010; Smith et al. 2006). The re-engineering of meganucleases to new target sites is a challenging endeavor and has not been widely adopted, however.

Recently, another DNA-binding domain has been identified as a potential motif for use in chimeric nucleases. Transcription activator-like (TAL) effectors are virulence factors in the phytopathogenic bacteria *Xanthomonas* and bind to DNA through a series of nearly identical repeats. Each repeat of 34 amino acids has two key residues that recognize and bind one nucleotide, and proteins with 11–17 repeats can be designed *de novo* to bind very specific DNA sequences (Boch et al. 2009; Moscou and Bogdanove 2009). TAL effector repeats, like zinc finger proteins, can

be fused to the FokI nuclease domain to generate TAL effector nucleases (TALENs) which dimerize and effectively cleave expected DNA sequences in yeast and plants (Cermak et al. 2011; Christian et al. 2010; Mahfouz et al. 2011). Notably, TALENs have also been developed which stimulate homologous recombination at an endogenous human locus at a rate of 16% (Miller et al. 2010). A significant advantage of TALENs compared to ZFNs is their seemingly total modularity; the same repeat recognizes and binds the same nucleotide regardless of the context of its neighboring repeats. This modularity eliminates most of the screening steps required in the synthesis of other chimeric nucleases. In fact, multiple groups have established a rapid method to design and synthesize novel TALENs and have made the necessary reagents publicly available (Cermak et al. 2011; Reyon et al. 2012; Sanjana et al. 2012). Just as ZFNs can be purchased from Sigma-Aldrich Biosciences, TALENs can be purchased from Cellectis or Life Sciences if investigators do not wish to invest resources in making their own. While the best designed ZFNs seem to have comparable activity to the best TALENs in general, in early comparisons, TALENs seem to have fewer off-target effects than ZFNs (Mussolino et al. 2011). Thus, it is likely that ZFNs and TALENs will both be useful nuclease platforms for targeted genome modifications in the future.

13.2.3 Targeted Genome Modification Using Non-nuclease Approaches

In addition to nuclease-mediated genome editing strategies, adeno-associated virus (AAV) has been used for targeted genome modifications (Khan et al. 2011). Through mechanisms still not clear, at high multiplicities of infection, AAV can stimulate targeted transgene integration or gene correction at frequencies up to 1% under certain experimental conditions. While this frequency is a log lower than what can be achieved using nuclease mediated genome modification it does have the advantage of not needing to engineer site-specific nucleases.

Papapetrou et al. have described another site-specific viral targeting approach in induced pluripotent stem cells (iPS). This lentiviral-mediated approach depends upon retrospective analysis and selection of optimal integration events at a "safe harbor" locus (Papapetrou et al. 2011). The authors define a safe harbor locus as a locus that meets the following citeria: a locus that is 50 kb from the 5′ end of any gene, 300 kb from any cancer-related gene, 300 kb from any microRNA and one that is located outside transcription units or "ultraconserved" regions. The iPS cells for this experiment were derived from patients with β -thalassemia and consequently did not express functional β -globin when differentiated into the erythroid lineage. The iPS cells were infected at a low multiplicity of infection with a β-globin expressing lentivirus. Single copy integrant clones were then selected and screened for the genomic location of integration. After screening, the authors found that they could isolate an iPS clone expressing β -globin efficiently from a locus that met the "safe harbor" criteria (Papapetrou et al. 2011). Although this retrospective strategy differs

from many of the prospective targeting approaches, the end result is similar and perhaps provides a plausible alternative for targeted genome engineering in cell types where individual clones can be isolated and expanded.

13.3 Genetic Modifications Achieved with Targeted Genome Engineering

In the following we will describe the various examples of using ZFNs to perform targeted genome modifications. These modifications can be classified into three major categories: (1) endogenous gene disruption (2) endogenous gene correction (3) safe harbor gene addition (Fig. 13.1). Table 13.1 summarizes the work in using nucleases for targeted genome modification in human cells.

13.3.1 Gene Disruption

The simplest method of targeted genome modification is gene disruption by mutagenic NHEJ after the induction of a DSB by the engineered nuclease. This method has been used for a wide variety of experimental purposes including the creation of knockout human cell lines as well as primary cells including stem cells (Alwin et al. 2005; DeKelver et al. 2010; Hockemeyer et al. 2009; Kim et al. 2009; Perez et al. 2008; Porteus and Baltimore 2003; Urnov et al. 2005; Zou et al. 2009). Further, whole animal knockout has been achieved by this method in model organisms that include flies (Beumer et al. 2006, Beumer et al., 2008; Bibikova et al. 2002; Bozas et al. 2009), zebrafish (Doyon et al. 2008; Foley et al. 2009; Meng et al. 2008), mice (Carbery et al. 2010; Cui et al. 2011; Meyer et al. 2010), rats (Geurts et al. 2009; Mashimo et al. 2010), rabbits (Flisikowska et al. 2011), frogs (Young et al. 2011), sea urchins (Ochiai et al. 2010), plants (Lloyd et al. 2005; Wright et al. 2005), nematodes (Morton et al. 2006) and worms (Takasu et al. 2010). As one might expect, by the time this review is published, this list of organisms will be incomplete. To create gene specific knockouts, the ZFNs have been introduced as either DNA expression plasmids or as mRNA through a variety of delivery techniques including standard transfection, nucleofection, adenoviral infection, integration-defective lentiviral infection, or by glass-needle microinjection. From a clinical gene therapy perspective the most important use of this approach has been to create HIV resistant cells through the ZFN induced mutation in the CCR5 gene. CCR5 is a necessary co-receptor for infection of CD4+ T-cells by HIV. People with homozygous Δ32 mutations in CCR5, which results in non-functional CCR5 protein, are resistant to HIV infection but are otherwise phenotypically normal except for an increased risk of infection by West Nile Virus and *Listeria monocytogenes*. The hypothesis based on these findings is that by creating an immune system that is CCR5 mutated in a patient who was already infected with HIV, one could alter the course of the disease. In an exciting

Table 13.1 Human genes modified by targeted genome engineering

Cell type	Gene	Frequency of modification	References
Gene disruption			
CD4+/HSC	CCR5	54%/17%	Perez et al. (2008), Holt et al. (2010)
HT1080	CAG repeats	>0.006%	Mittleman et al. (2009)
293	HOXB13	9.60%	Maeder et al. (2008)
293	CFTR	1.20%	Maeder et al. (2008)
Gene correction			
K562/CD4+	IL2Ry	20%/5%	Urnov et al. (2005), Lombardo et al. (2007)
K562	VEGF	7.7%	Maeder et al. (2008)
CHO	XPC	0.20%	Arnould et al. (2007)
293	RAG1	6%	Grizot et al. (2009)
Gene addition			
K562	IL2Ry	15%	Moehle et al. (2007), Lombardo et al. (2007)
hESC/iPSC	AAVS1	56%/75% (selected)	Hockemeyer et al. (2009), Zou et al. (2011)
MSC/hESC	CCR5	40%/5.3%	Benabdallah et al. (2010), Lombardo et al. (2007)
hESC/iPSC	PIGA	0.24%	Zou et al. (2009)
hESC	OCT4	94% (selected)	Hockemeyer et al. (2009)
hESC	PITX3	11% (selected)	Hockemeyer et al. (2009)

HSC hematopoietic stem cell, *HT1080* human fibrosarcoma cell line, *293* human embryonic kidney cell line, *K562* human erythroleukemia cell line, *CHO* Chinese hamster ovary cell line, *hESC* human embryonic stem cell, *iPS* induced pluripotent stem cell, *MSC* mesenchymal stem cell

proof-of-principle experiment, an HIV infected patient underwent a bone marrow transplant for acute myelogenous leukemia using donor marrow that was homozygous CCR5 Δ32. Remarkably, after the transplant the patient's HIV viral load has become undetectable and his donor CD4 count, derived from the CCR5 Δ32 donor, has risen (Allers et al. 2011; Hutter et al. 2009). This human case study demonstrated that it might be possible to engineer the immune system to become HIV resistant with a beneficial clinical outcome. Unfortunately, bone marrow transplantation using homozygous Δ32 donors is not a broadly feasible strategy for the treatment of HIV because most patients will not have an HLA matched homozygous Δ32 donor and because of the general toxicity from allogeneic bone marrow transplantation.

An alternative strategy would be to engineer the patient's own immune system to become HIV resistant by creating ZFN mediated knockout of the CCR5 gene in autologous cells, either CD4+ T-cells or hematopoietic stem cells. In pre-clinical studies, Carl June and co-workers have demonstrated that using CCR5 directed ZFNs, primary T-cell populations can be generated in which >50% of the CCR5 alleles have been mutated and that these T-cells are resistant to a laboratory CCR5 tropic strain of HIV (Perez et al. 2008). Similarly, Paula Cannon and her co-workers have demonstrated that the same CCR5 ZFNs can cause inactivation of the CCR5 gene in human CD34+ cells and that T-cells derived from these gene modified cells

are resistant to HIV (Holt et al. 2010). Based on these pre-clinical findings, clinical trials have opened based on this strategy. In the first clinical trials to open, patient derived CD4+ T-cells were infected with an adenovirus that expresses the CCR5 ZFNs causing CCR5 gene modification. The population of autologous T-cells was then expanded *in vitro* and transplanted back into the patient. These trials have enrolled a number of patients and the results, including whether the modified T-cells persist, are able to confer functional clinical protection against HIV and can provide a functional immune system that protects the patient from developing AIDS, are pending.

Of the handful of published ZFN pairs designed to disrupt human genes, the CCR5 ZFNs have had the most pre-clinical and clinical success. Others, however, are being developed that are perhaps promising therapeutic tools. One such pair of ZFNs has been developed to recognize CAG triplet repeat tracts. CAG tracts are physiologically normal in multiple genes when the number of repeats is fewer than 30, however expansion of repeat number leads to debilitating diseases such as myotonic dystrophy, Huntigton's disease and spinocerebellar ataxia. ZFNs that bind these repeats were designed by modular assembly, and demonstrate the ability to contract CAG repeats in a reporter assay in human cells (Mittelman et al. 2009).

A third pair of ZFNs used to disrupt an endogenous human gene has been developed against the *PIG-A* gene, located on the X-chromosome Xp22.1. Mutations in this gene cause paroxysmal nocturnal hemoglobinuria (PNH). The *PIG-A* gene encodes a phosphatidylinositol glycan class A protein that functions in a complex required for production of many glycosyl-phosphatidyl-inositol anchored proteins (GPI-APs). Among these anchor proteins are a few involved in complement regulation, without which cells are inappropriately recognized and lysed by the complement system. For patients with PNH, a somatic mutation within the hematopoietic progenitor population results in pools of abnormal red blood cells. Their lysis by the complement system results in a hemolytic anemia and hemoglobinuria. Zou et al. demonstrated that the PIG-A ZFNs could target and knock out the *PIG-A* gene in both human embryonic stem cells (hESC) and induced pluripotent stem (iPS) cells by either mutagenic NHEJ-mediated repair of the DSB or by targeting a hygromycin resistance cassette flanked by homology arms to the locus. PIG-A knock-out was phenotypically characterized by resistance to aerolysin-mediated cell killing and by cell surface absence of GPI-APs. The authors confirmed that restoration of a PIG-A transgene could rescue these phenotypes (Zou et al. 2009). Though this work did not yet demonstrate targeting of the PIG-A gene in a therapeutically relevant manner for PNH patients, the PIG-A ZFNs could hold potential for a ZFN-mediated gene correction strategy in HSCs.

Though endogenous gene disruption is clinically relevant in perhaps only a few scenarios, an important point to consider is the impact that endogenous gene disruption could have for studying the biology of disease. As demonstrated, highly efficient targeting strategies now allow for robust and precise genetic manipulation in previously inaccessible cells types such as hESC/iPSC. Moreover, ZFNs have been used to create targeted gene mutations to create genetically modified organisms in a wide variety of species for which it previously was not possible to perform targeted gene modification. The ability to create precise genetically modified lines in species

other than mouse is certainly going to result in improved models of human disease, improved species for agricultural purposes, and in the long-term general improvements in human health as a result.

13.3.2 Gene Correction

By harnessing the capability of nucleases to stimulate gene targeting through homologous recombination, small, precise changes can be introduced into the genome including the correction of disease causing mutations. To achieve gene correction at an endogenous locus, site-specific nucleases must be designed to a sequence within the gene through one of the strategies described above. The donor DNA template must also be designed that is homologous to the endogenous gene and contains the small changes that are to be introduced into the genome. The donor DNA template is used by the homologous recombination machinery as the substrate to synthesize new DNA that will then be used repair the DSB. Finally, both components (the nucleases and the donor DNA template) must be able to be delivered to a cell-type that is capable of reconstituting the diseased tissue or organ. Translation of this therapeutic paradigm to the clinic has not yet been accomplished, but progress has been made *in vitro* towards its end, and work is currently underway to demonstrate the feasibility of the strategy in animal models.

The first examples of nuclease mediated gene correction of a chromosomally integrated gene in human cells were described in 2003, 2005 and 2006 (Porteus and Baltimore 2003; Urnov et al. 2005; Porteus 2006). In these first examples of ZFN mediated gene correction, a mutated GFP gene was introduced as a single copy into the genome of cells to create a reporter line. The frequency of gene correction was measured by determining the frequency that cells became GFP positive. When the donor DNA template was introduced without nucleases, the frequency of gene correction was on the order of 0.0001% (10^{-6}). However, when ZFNs that target the integrated reporter were co-transfected with the donor DNA template, the gene correction frequency increased to 0.5-2.2% in HEK-293 cells. In HEK-293 cells, the frequency of gene correction, however, when ZFNs that target the integrated reporter gene were co-transfected with the donor DNA template, increased to 0.5–2.2% (Porteus and Baltimore 2003; Urnov et al. 2005). This frequency increased to ~10% if the cells were transiently arrested in G2/M by vinblastine (Potts et al. 2006; Urnov et al. 2005). A similar stimulation in gene correction using the GFP reporter system was also found in human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hIPSCs) (Zou et al. 2009). Importantly, the gene corrected hESCs and hIPSCs retained their pluripotency and did not develop any overt karyotypic abnormalities. Finally, gene correction rates of ~2% were obtained in primary mouse fibroblasts using a transgenic gene targeting mouse reporter line in which the mutated GFP gene was "knocked-in" to the ROSA26 locus (Connelly et al. 2010).

Urnov et al. published the first example of gene correction in an endogenous human gene (Urnov et al. 2005). Mutations in the IL2Rγ gene cause SCID-X1, the

most common form of severe combined immunodeficiency (SCID). ZFNs to exon 5 were engineered using the Sangamo strategy. Transfection of these ZFNs along with a gene correction DNA donor template, demonstrated gene correction efficiencies of ~20% in K562 cells and ~4% in primary T-cells. Lombardo et al. built on this work and demonstrated that using integration defective lentivirus rather than standard transfection techniques, similar rates of gene targeting could be obtained in HEK-293 cells, K562 cells, and an EBV transformed lymphoblastoid cell line (Lombardo et al. 2007).

In addition to ZFNs, homing endonucleases have also been designed to stimulate HR at chromosomal loci of human diseases. The first example of this is an I-CreI derived homing endonuclease that targets XPC, a gene involved in the skin disorder Xeroderma Pigmentosum (Arnould et al. 2007). The endonuclease-binding site from XPC was cloned into the lacZ gene, disrupting expression of lacZ. Then the mutated lacZ construct was integrated into the genome of CHO-K1 cells, transfected with the designed homing endonuclease and a lacZ repair template, and shown to successfully stimulate HR. Another gene relevant to human disease that has been targeted with a homing endonuclease is RAG1. RAG1 is essential for VDJ recombination in B- and T-cells, and mutations in RAG1 cause Omenn syndrome, a SCID variant. RAG1 homing endonucleases stimulate HR at the endogenous locus in HEK-293 cells at rates of up to 6% (Grizot et al. 2009; Munoz et al. 2011). Although the results of these two studies with homing endonucleases show promising initial targeting results, there have not yet been reports of the correction of disease-causing mutations in primary human cells.

Endogenous gene correction has several advantageous features. First, with modification of the endogenous gene, the resultant transgene product remains under the control of endogenous regulatory elements at its correct chromosomal location. This is crucial for maintaining the biological optimum of protein expression that may be lost with ectopically located transgene integrations. Second, direct gene correction rather than virally mediated transgene addition eliminates the concern regarding insertional oncogenesis from uncontrolled viral integration sites. Third, gene correction, in contrast to gene addition, could be used to treat dominant diseases or diseases in which the mutant protein has dominant features. An example of the latter is sickle cell disease which is caused by a point mutation in the beta-globin gene. In a gene addition strategy, the transgene will have to out-compete the two mutant genes that continue to express large amounts of mutant beta-globin. In contrast, after gene correction, one would be turning the homozygous state into a heterozygous state, a non-disease causing condition, and issues of competition would be avoided.

The major challenges to translating nuclease mediated gene correction into therapy include developing high quality nucleases to disease causing genes, demonstrating that clinically relevant levels of gene correction can occur in the appropriate cell type and finally demonstration that gene corrected cells can correct the disease phenotype.

13.3.3 Safe Harbor Gene Addition

An alternative to direct gene correction is to use homologous recombination to target transgene addition to a safe harbor in the genome. While the defining features of a genetic safe harbor are still being established, in principle a safe-harbor locus can be defined as one in which the insertion of a transgene has no aberrant physiologic consequences. Another feature of a safe harbor locus, in addition to its safety, is it must be a location in which a transgene can be expressed at the levels necessary to achieve therapeutic efficacy. A robust example of safe harbor utilization is the frequent targeting of the murine ROSA26 locus for knock-in mouse technology.

Targeted gene addition by homologous recombination can be achieved using nucleases engineered to recognize the specific safe harbor site. In addition, a donor DNA plasmid is engineered by inserting a transgene cassette (driving the expression of one or several genes) between arms of homology. The arms of homology can be as short as 1 kilobase or shorter in length and flank the site of the nuclease site. Targeted gene addition occurs when the homologous recombination machinery uses the donor DNA template to repair the nuclease induced DSB. When the homologous recombination machinery uses the donor DNA template, it results in the insertion of the transgene cassette at the site of the break. Moehle et al. first described this technology using the previously described IL2Rγ ZFNs. The authors began by inserting a very small, in-frame tag of only four amino acids (RAKR, furin cleavage site) into the IL2Rγ gene. They found that they could achieve targeting up to 15% by this strategy in K562 cells. Then, to investigate the full potential of gene addition, the authors targeted a 900 base pair GFP cassette, an 1,100 base pair GFP-PolyA cassette and a full transcriptionally active unit of a promoter-GFP that measured 1,500 base pairs. The authors demonstrated targeting frequencies of 6%, 3% and 6% for these three constructs, respectively. Finally, the authors generated a donor plasmid with a very large insert (8 kb) between homology arms that contained three independent, transcriptionally active units. Remarkably, the authors found by Southern blot analysis that 5–8% of the IL2Rγ alleles had been targeted (Moehle et al. 2007). These data highlight the potential for nuclease-mediated gene addition strategies by suggesting that even very large payloads can be efficiently and precisely targeted to safe harbor loci.

Lombardo et al. described gene addition using the IL2Rγ ZFNs in a panel of human cell types using integration deficient lentivirus (IDLV) delivery of targeting components. First, the author attempted targeting with a donor construct where GFP was flanked by IL2Rγ homology arms. They reported targeting efficiencies of up to 6.6% and up to 2.4% in K562s and in an Epstein Barr virus transformed lymphoblastoid cell line, respectively. Of note, in the latter cell line, gamma chain expression was lost in GFP + cells. Because SCID can be caused by many mutations in the IL2Rγ, the author thought a potential treatment for the disease could be to target normal IL2Rγ cDNA to the endogenous locus, as a one shot approach to cover all of the different mutations (instead of designing novel ZFNs to target each mutated

exon separately). The authors found that they could target IL2R-g cDNA to the endogenous IL2Rγ locus at frequency of up to 6% in their lymphoblastoid cell line and that IL2Rγ expression was then restored (Lombardo et al. 2007).

One site being developed as a potential safe harbor is the AAVS1 integration site on human chromosome 19. The AAVS1 locus encodes the ubiquitously expressed *PPP1R12C* gene. Because AAV integration at this site does not seem to be deleterious and because the AAVS1 site is ubiquitously expressed, it has characteristics that may make it a good safe harbor. To this end, Sangamo Biosciences developed a pair of zinc-finger nucleases that recognized the AAVS1 locus and found that they could target several transgenes to the AAVS1 locus in various cell types, including hESCs (DeKelver et al. 2010; Hockemeyer et al. 2009). In Hockemeyer et al., the authors demonstrated that they could achieve three methods of transgene expression within this locus. First, they could successfully target a puromycin resistance cassette driven by the endogenous *PPP1R12C* promoter through a splice acceptor. Second, they could achieve constitutively expressed eGFP by targeting an eGFP cassette with cytomegalovirus (CMV) promoter with a chicken β-actin enhancer. Lastly, the authors targeted the eGFP gene with a tetracycline response element and demonstrated that eGFP expression could be induced by administration of doxycycline (Hockemeyer et al. 2009). A theoretical limitation of this site as a safe harbor is that *PPP1R12C* codes for a protein phosphatase and since phosphatases are known to be important regulators of cell proliferation it is possible that using this gene as a safe harbor might increase the probability of cell transformation.

Recently, Zou et al. has described an AAVS1 ZFN-mediated targeting strategy in iPS cells derived from a patient with X-linked chronic granulomatous disease (CGD). Patients with CGD have a deficiency of the p91-phox gene product and lack reactive oxygen species (ROS)-mediated microbicidal activity in neutrophils. As a result, these patients present with severe, recurrent, and atypical infections. In this paper, the authors described iPS cell derivation from patients with CGD, that, when differentiated into neutrophils, expressed no functional p91-phox protein and stained for reduced ROS production, compared to normal iPS cells. The authors targeted a p91-phox minigene ($gp91^{phox}$) to the AAVS1 locus in the iPS cells using mRNA for the AAVS1 ZFNs and achieved targeted transgene insertion in 75% (15/20) of the iPS clones they examined. The authors then excluded clones with targeting at more than one AAVS1 allele and any clones that had experienced small NHEJ-mediated deletions at the other, non-targeted AAVS1 allele. Of the three (3/20) remaining clones, all retained normal iPS morphology, stable expression of the $gp91^{phox}$ transgene (for more than 4 months) and pluripotency (as measured by embryoid body and teratoma formation assays). Importantly, when differentiated into neutrophils, oxidase activity was restored, correcting the original disease phenotype. The authors also performed an interesting and important control for this experiment. Using the same CGD iPS cells, the authors infected a $gp91^{phox}$ expressing lentivirus into the cells in an attempt to correct the disease phenotype as might have been attempted as an earlier non-targeted gene therapy strategy. Though $gp91^{phox}$ was initially expressed at very high levels, only 4% of the resultant differentiated neutrophils were weakly oxidase positive at the end of the experiment.

The authors suggested that this was because "a significant proportion of lentivector genomic inserts are subject to accelerated silencing" when the iPS cells were differentiated into neutrophils (Zou et al. 2011). This data highlights the importance of moving towards precise, site-specific genome engineering and away from random integration and virally-mediated genome engineering.

Another example of a potential safe harbor locus is CCR5. In addition to using ZFNs to create mutations in the CCR5 gene, there have also been reports using those ZFNs to target transgenes to that locus. One example shows the successful gene addition of erythropoietin (Epo) to CCR5 in up to 40% of human mesenchymal stromal cells (MSCs). Epo is a cytokine that drives the maturation of red blood cells in the bone marrow, and loss of Epo leads to severe anemia. Injection of the MSCs into immunodeficient mice resulted in increased levels of circulating Epo and higher hematocrit compared to controls (Benabdallah et al. 2010). These results indicate that ZFN-mediated targeting of plasma-soluble factors to a safe harbor locus may be a viable therapeutic option.

Though endogenous gene correction could be considered the ultimate and paramount goal for gene therapy for loss-of-function genetic diseases, safe harbor gene addition provides a significant amount of flexibility when compared to gene correction. For safe harbor targeting through nuclease-mediated strategies, only one pair of ZFNs need be designed and optimized to achieve targeting of any number of therapeutic transgenes (as opposed to the individual tailoring that is required for endogenous gene correction). As briefly discussed, the effort put forth to bringing a pair of ZFNs to clinical trials are substantial and a strategy that necessitates only one pair could be advantageous. Secondly, as highlighted by Benabdallah et al. (2010), alternative cell types (such as MSCs) can be manipulated to secrete ectopic transgene products (such as Epo) and this strategy might only be available through gene addition. Lastly, gene addition provides the option to express transgenes through any number of mechanisms, including splice acceptor/endogenous promoter approaches, constitutive expression or response-element regulatable expression, as demonstrated by Hockemeyer et al. (2009).

13.4 Non-clinical Applications for Targeted Genome Engineering

Site-specific genome engineering holds promise for use as a clinical therapeutic for many monogenic diseases. However, precise, genetic modification of human cells can also be developed as a tool to better understand more basic biological questions. One such example was published by DeKelver et al. in 2010. For this study, the authors utilized ZFNs directed to the AAVS1 locus to generate a panel of stable human cell lines. Typically, stable cell lines are generated by DNA transfection or viral infection of a gene of interest along with a selectable marker (usually drug selection). The gene randomly integrates into the genome, integrants are selected for

using the marker, and a clonal population is isolated. Unfortunately, many confounding factors can arise from the random integration and clonal selection that may alter the true phenotype that would have resulted from the genetic manipulation. Further, the transgene expression can be unstable due to silencing effects over time, lending to unpredictable results. DeKelver et al. postulated that targeting trangenes to a safe harbor locus would circumvent these complications and allow for generation of more predictable and "isogenic" stable cell lines. To this end, the authors demonstrated targeting at the AAVS1 locus in a panel of cell lines (including K562, Hep3B, HEK293 and U2OS) and in two primary human cell types, fibroblasts and hESC. The authors targeted a variety of constructs to the AAVS1 locus that included (1) a GFP reporter in K562, Hep3B, and HEK293 cells where targeting frequencies ranged from 3% to 10% (2) glucocorticoid receptor response element luciferase reporter constructs in U2OS and (3) shRNA cassettes targeted to CD58 and to components in the mTOR pathway in K562, HEK293 and hESC. Though the scientific potential for targeted "isogenic" cell lines is evident, and the ability to target a variety of constructs to many cell types was thoroughly demonstrated, a comparison between targeted "isogenic" stable cells and stable cell lines generated by conventional means was lacking in this report.

Similar to the generation of targeted stable cells lines is the application of genome targeting for protein tagging. Classically, cellular trafficking studies have been accomplished by overexpression of a tagged fusion protein from plasmid DNA. Often, however, protein overexpression can perturb the cell's normal physiology and confound experimental results. A recent example of targeted protein tagging demonstrated that clarithin-mediated endocytosis could better be studied by the ZFN-mediated generation of fluorescent fusion proteins at their endogenous loci (Doyon et al. 2011a). Other studies have also demonstrated that ZFN-mediated gene targeting can be used to create tagged proteins that allow for lineage tracing. One example involved tagging Oct4 with eGFP as a reporter to monitor the pluripotent state of hESCs. In this same publication, the authors also targeted a non-hESC expressed locus (PITX3) with eGFP in hESC and iPS cells. They suggested that this strategy could be used as a reporter when the hESC or iPS cells were differentiated into neurons (where PITX3 is expressed), though this data was not reported (Hockemeyer et al. 2009).

Another non-therapeutic role of ZFNs is the demonstration that two distinct pairs of nucleases can stimulate chromosomal translocations when administered to cells simultaneously. Using first the nuclease combination of I-SceI and ZFNs to the IL2Rγ locus and then the combination of the IL2Rγ and AAVS1 ZFNs, Brunet et al. demonstrated the translocation of large chromosomal fragments. In the IL2Rγ /AAVS1 experiments, reciprocal translocations between the X-chromosome (site of IL2Rγ) and chromosome 19 (AAVS1 site) were observed at a frequency of 10^{-4} in HEK-293 cells and 2×10^{-6} in human embryonic stem cells (Brunet et al. 2009). Although these rates are several orders of magnitude lower than rates of gene targeting at a single locus, this strategy provides a framework to study the mechanisms of chromosomal rearrangement and the factors required for this process. Additionally, developing ZFNs to chromosomal sites that have been implicated in oncogenic

translocations would provide a way to study the expression and regulation of the resulting fusion proteins from their endogenous loci. The demonstration that ZFNs can induce translocations, albeit at very low frequencies, also has implication for how to evaluate the safety of ZFNs as they proceed into human clinical trials.

Another application of ZFNs for chromosomal manipulation is the use of two distinct pairs that recognize sites on the same chromosome to induce deletions. Using ZFNs designed to the adjacent genes CCR2 and CCR5, Lee et al. used a PCR approach to detect deletions of up to 15 kbp within the gene cluster. Further, when the authors utilized ZFNs upstream to the CCR5 locus, they were able to induce large chromosomal deletions up to 15 Mbp (Lee et al. 2010). Similarly, Sollu et al. described chromosomal deletions at a frequency of 10% using two pairs of nucleases designed to different sites within the HOXB13 locus (Sollu et al. 2010). In this way, it may now be possible to achieve the targeted deletion of entire gene clusters, introns or even specific exons in human cells. While the ability to create specific chromosomal rearrangements by ZFNs has important research implications, it also has important implications for their translation into a clinical therapeutic as the induction of unintended gross chromosomal rearrangements would be a serious adverse event.

13.5 Obstacles and Safety Concerns for Targeted Genome Engineering in Human Cells

Since the first report on using ZFNs to create targeted genome modifications, it has been evident that expression of ZFNs in mammalian cells can result in reduced cell viability (Porteus and Baltimore 2003). The reduced cell viability is directly correlated with the expression level of the ZFN as transfecting more ZFN expression plasmids results in a higher degree of cell toxicity (Pruett-Miller et al. 2008). The mechanism of reduced cell viability is directly related to the creation of off-target DNA double-strand breaks. A point mutation in the nuclease domain that inactivates its catalytic activity, for example, eliminates the cytotoxicity of ZFNs (Beumer et al. 2006). Moreover, staining for 53BP1 foci, a marker for double-strand breaks, demonstrates that increased toxicity is directly correlated with increased numbers of 53BP1 foci (Pruett-Miller et al. 2008). Elimination of increased 53BP1 foci also eliminates the reduced cell viability associated with ZFN cell expression (Pruett-Miller et al. 2009). Intriguingly, TALENs seem to have less cytotoxicity than ZFNs when expressed in cells although they have not been studied to the same depth as ZFNs (Mussolino et al. 2011). Nonetheless, the apparent decreased cell toxicity of TALENs is a promising feature of this alternative nuclease platform.

The identification of the sites of the ZFN off-target double-strand breaks has been a challenging problem. Gabriel et al. have used the capture of extra-chromosomal DNA at sites of double-strand breaks in a non-homologous fashion to identify such sites and identified a number of genomic targets for a specific ZFN pair that had not been previously recognized. The nature of their approach, however, clearly biased against identifying rare off-target sites and underestimated the number of off-target

sites (Gabriel et al. 2011). Pattanayak et al. used an entirely different strategy to identify off-target sites for the same set of ZFNs studied by Gabriel et al. in the same cell type (Pattanayak et al. 2011). Interestingly, while both groups confirmed a number of new off-target loci, the set of sites were completely non-overlapping from the two different approaches. Thus, the full set of off-target sites, even for this highly studied nuclease pair, has not been described and it is likely that further improvements will be needed to generate a comprehensive or near comprehensive list. Both of these studies evaluated the CCR5 ZFN pair that is now in clinical trials. Prior to entering clinical trials, this pair was introduced into large numbers of mouse hematopoietic progenitor cells, the ZFN exposed cells were then transplanted into mice and the transplanted mice were observed for tumors. There was no increase in tumor formation in the transplanted mice but it is difficult to assess the significance of this heroic study as there are no positive controls for such studies to validate how this assay would relate to the probability of tumor formation in human cells transplanted after exposure to ZFNs.

The reduced cell viability after ZFN exposure has prompted an investigation into strategies to reduce this effect. Two major strategies have been developed. The first is to modify the nuclease domain such that a nuclease pair can only cleave DNA as heterodimers ("obligate heterodimer" nuclease domains). While theoretically and now confirmed experimentally, this reduces the number of off-target sites by ~50%. The functional result is a marked improvement in cell viability when using nuclease pairs that contain the obligate heterodimer nuclease (Doyon et al. 2011b; Miller et al. 2010; Pruett-Miller et al. 2008; Szczepek et al. 2007). In some cases, however, the use of the obligate heterodimer nuclease domains also results in reduced on-target activity (Wilson and Porteus manuscript submitted). An alternative strategy is to regulate the expression level of the ZFNs. Since the targeted genome modification is a "hit and run" strategy, sustained nuclease expression is not needed and, in fact, sustained nuclease expression can result in increased cell toxicity. Pruett-Miller et al. demonstrated that using small molecule regulation of ZFN protein level that on-target activity can be maintained while decreasing the cell toxicity and off-target foci formation to background levels (Pruett-Miller et al. 2009). The importance of regulating ZFN protein level both in amount and over time is confirmed by the off-target studies of Pattanayak et al. (2011).

13.6 Conclusion

The ability to precisely modify the genome of human cells has the potential to shift the paradigm of how to treat monogenic and other human diseases. Instead of repeated infusions of deficient proteins like Factor VIII to treat hemophilia or ADA to treat one form of SCID, which are very laborious and expensive regimens, gene therapy holds the promise of a robust, long-term cure. Additionally, gene therapy can address the root cause of other monogenic diseases like sickle cell disease and X-linked SCID, in which the deficiency is not a secreted protein, cannot be treated

by protein infusion, and can currently only be cured through allogeneic stem cell transplants. Targeted modification of a patient's own stem cells avoids the side-effects and reliance on immunosuppressive drugs associated with allogeneic transplants and is now under investigation as a way to cure HIV infection. The targeted therapy approaches highlighted in this chapter address and improve the problems associated with classic retroviral-mediated transgene addition. Targeted methods do not rely on random integration that can lead to oncogenesis, and they are more broadly applicable because they can be used not only for gene addition, but also for gene correction and gene disruption. As the specificity of these methods continues to be improved and more clinical data from patient trials becomes available, targeted genome engineering will perhaps one day fulfill the promise of a cure for monogenic diseases.

Acknowledgements The authors thank the UT Southwestern Medical Scientist Training Program for the support of JB and RV. The work in the Porteus lab is supported by grants from the Lucille Packard Foundation for Children's Health, the Amon Carter Foundation, and the NIH Director's Common Fund (PN2EY018244).

References

Abuljadayel IS, Quereshi H, Ahsan T, Rizvi S, Ahmed T, Khan SM, Akhtar J, Dhoot G (2006) Infusion of autologous retrodifferentiated stem cells into patients with beta-thalassemia. ScientificWorldJournal 6:1278–1297

Aiuti A, Brigida I, Ferrua F, Cappelli B, Chiesa R, Marktel S, Roncarolo MG (2009) Hematopoietic stem cell gene therapy for adenosine deaminase deficient-SCID. Immunol Res 44:150–159

Allers K, Hutter G, Hofmann J, Loddenkemper C, Rieger K, Thiel E, Schneider T (2011) Evidence for the cure of HIV infection by CCR5Delta32/Delta32 stem cell transplantation. Blood 117:2791–2799

Alwin S, Gere MB, Guhl E, Effertz K, Barbas CF 3rd, Segal DJ, Weitzman MD, Cathomen T (2005) Custom zinc-finger nucleases for use in human cells. Mol Ther 12:610–617

Arnould S, Chames P, Perez C, Lacroix E, Duclert A, Epinat JC, Stricher F, Petit AS, Patin A, Guillier S, Rolland S, Prieto J, Blanco FJ, Bravo J, Montoya G, Serrano L, Duchateau P, Paques F (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. J Mol Biol 355:443–458

Arnould S, Perez C, Cabaniols JP, Smith J, Gouble A, Grizot S, Epinat JC, Duclert A, Duchateau P, Paques F (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. J Mol Biol 371:49–65

Arnould S, Delenda C, Grizot S, Desseaux C, Paques F, Silva GH, Smith J (2011) The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. Protein Eng Des Sel 24:27–31

Bank A, Dorazio R, Leboulch P (2005) A phase I/II clinical trial of beta-globin gene therapy for beta-thalassemia. Ann N Y Acad Sci 1054:308–316

Beerli RR, Dreier B, Barbas CF 3rd (2000) Positive and negative regulation of endogenous genes by designed transcription factors. Proc Natl Acad Sci U S A 97:1495–1500

Benabdallah BF, Allard E, Yao S, Friedman G, Gregory PD, Eliopoulos N, Fradette J, Spees JL, Haddad E, Holmes MC, Beausejour CM (2010) Targeted gene addition to human mesenchymal stromal cells as a cell-based plasma-soluble protein delivery platform. Cytotherapy 12:394–399

Beumer K, Bhattacharyya G, Bibikova M, Trautman JK, Carroll D (2006) Efficient gene targeting in Drosophila with zinc-finger nucleases. Genetics 172:2391–2403

Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, Gall JG, Carroll D (2008) Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A 105:19821–19826

Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, Chandrasegaran S (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol 21:289–297

Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161:1169–1175

Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A 95:10570–10575

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512

Bozas A, Beumer KJ, Trautman JK, Carroll D (2009) Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila. Genetics 182:641–651

Boztug K, Schmidt M, Schwarzer A, Banerjee PP, Diez IA, Dewey RA, Bohm M, Nowrouzi A, Ball CR, Glimm H, Naundorf S, Kuhlcke K, Blasczyk R, Kondratenko I, Marodi L, Orange JS, von Kalle C, Klein C (2010) Stem-cell gene therapy for the Wiskott-Aldrich syndrome. N Engl J Med 363:1918–1927

Brunet E, Simsek D, Tomishima M, Dekelver R, Choi VM, Gregory P, Urnov F, Weinstock DM, Jasin M (2009) Chromosomal translocations induced at specified loci in human stem cells. Proc Natl Acad Sci U S A 106:10620–10625

Carbery ID, Ji D, Harrington A, Brown V, Weinstein EJ, Liaw L, Cui X (2010) Targeted genome modification in mice using zinc-finger nucleases. Genetics 186:451–459

Cavazzana-Calvo M, Hacein-Bey S, De Saint Basile G, Gross F, Yvon E, Nusbaum P, Selz F, Hue C, Certain S, Casanova JL, Bousso P, Deist FL, Fischer A (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science 288:669–672

Cavazzana-Calvo M, Lagresle C, Hacein-Bey-abina S, Fischer A (2005) Gene therapy for severe combined immunodeficiency. Annu Rev Med 56:585–602

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39:e82

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186:757–761

Connelly JP, Barker JC, Pruett-Miller S, Porteus MH (2010) Gene correction by homologous recombination with zinc finger nucleases in primary cells from a mouse model of a generic recessive genetic disease. Mol Ther 18:1103–1110

Cui X, Ji D, Fisher DA, Wu Y, Briner DM, Weinstein EJ (2011) Targeted integration in rat and mouse embryos with zinc-finger nucleases. Nat Biotechnol 29:64–67

Dekelver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijsing SH, Sancak Y, Cui X, Steine EJ, Miller JC, Tam P, Bartsevich VV, Meng X, Rupniewski I, Gopalan SM, Sun HC, Pitz KJ, Rock JM, Zhang L, Davis GD, Rebar EJ, Cheeseman IM, Yamamoto KR, Sabatini DM, Jaenisch R, Gregory PD, Urnov FD (2010) Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res 20:1133–1142

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, Hocking TD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol 26:702–708

Doyon JB, Zeitler B, Cheng J, Cheng AT, Cherone JM, Santiago Y, Lee AH, Vo TD, Doyon Y, Miller JC, Paschon DE, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Drubin DG (2011a) Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. Nat Cell Biol 13:331–337

Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, Xia DF, Miller JC, Urnov FD, Gregory PD, Holmes MC (2011b) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods 8:74–79

Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF 3rd (2001) Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 276:29466–29478

Dreier B, Fuller RP, Segal DJ, Lund CV, Blancafort P, Huber A, Koksch B, Barbas CF 3rd (2005) Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 280:35588–35597

Durai S, Bosley A, Abulencia AB, Chandrasegaran S, Ostermeier M (2006) A bacterial one-hybrid selection system for interrogating zinc finger-DNA interactions. Comb Chem High Throughput Screen 9:301–311

Flisikowska T, Thorey IS, Offner S, Ros F, Lifke V, Zeitler B, Rottmann O, Vincent A, Zhang L, Jenkins S, Niersbach H, Kind AJ, Gregory PD, Schnieke AE, Platzer J (2011) Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. PLoS One 6:e21045

Foley JE, Yeh JR, Maeder ML, Reyon D, Sander JD, Peterson RT, Joung JK (2009) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). PLoS One 4:e4348

Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppel C, Nowrouzi A, Bartholomae CC, Wang J, Friedman G, Holmes MC, Gregory PD, Glimm H, Schmidt M, Naldini L, von Kalle C (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol 29:816–823

Gaspar HB, Bjorkegren E, Parsley K, Gilmour KC, King D, Sinclair J, Zhang F, Giannakopoulos A, Adams S, Fairbanks LD, Gaspar J, Henderson L, Xu-Bayford JH, Davies EG, Veys PA, Kinnon C, Thrasher AJ (2006) Successful reconstitution of immunity in ADA-SCID by stem cell gene therapy following cessation of PEG-ADA and use of mild preconditioning. Mol Ther 14:505–513

Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Menoret S, Anegon I, Davis GD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jacob HJ, Buelow R (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325:433

Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, Villate M, Thomas S, Lemaire L, Montoya G, Blanco FJ, Paques F, Duchateau P (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res 37:5405–5419

Grizot S, Epinat JC, Thomas S, Duclert A, Rolland S, Paques F, Duchateau P (2010) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. Nucleic Acids Res 38:2006–2018

Hacein-Bey-Abina S, Von Kalle C, Schmidt M, Mccormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JI, De Saint Basile G, Alexander I, Wintergerst U, Frebourg T, Aurias A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE, Wissler M, Prinz C, Rabbitts TH, Le Deist F, Fischer A, Cavazzana-Calvo M (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 302:415–419

Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, Clappier E, Caccavelli L, Delabesse E, Beldjord K, Asnafi V, Macintyre E, Dal Cortivo L, Radford I, Brousse N, Sigaux F, Moshous D, Hauer J, Borkhardt A, Belohradsky BH, Wintergerst U, Velez MC, Leiva L, Sorensen R, Wulffraat N, Blanche S, Bushman FD, Fischer A, Cavazzana-Calvo M (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest 118:3132–3142

Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, Dekelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng X, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol 27:851–857

Holt N, Wang J, Kim K, Friedman G, Wang X, Taupin V, Crooks GM, Kohn DB, Gregory PD, Holmes MC, Cannon PM (2010) Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nat Biotechnol 28:839–847

Hutter G, Nowak D, Mossner M, Ganepola S, Mussig A, Allers K, Schneider T, Hofmann J, Kucherer C, Blau O, Blau IW, Hofmann WK, Thiel E (2009) Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med 360:692–698

Kang EM, Choi U, Theobald N, Linton G, Long Priel DA, Kuhns D, Malech HL (2010) Retrovirus gene therapy for X-linked chronic granulomatous disease can achieve stable long-term correction of oxidase activity in peripheral blood neutrophils. Blood 115:783–791

Khan IF, Hirata RK, Wang PR, Li Y, Kho J, Nelson A, Huo Y, Zavaljevski M, Ware C, Russell DW (2010) Engineering of human pluripotent stem cells by AAV-mediated gene targeting. Mol Ther 18:1192–1199

Khan IF, Hirata RK, Russell DW (2011) AAV-mediated gene targeting methods for human cells. Nat Protoc 6:482–501

Kim YG, Chandrasegaran S (1994) Chimeric restriction endonuclease. Proc Natl Acad Sci U S A 91:883–887

Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res 19:1279–1288

Lee HJ, Kim E, Kim JS (2010) Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res 20:81–89

Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci U S A 102:2232–2237

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25:1298–1306

Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Muller-Lerch F, Fu F, Pearlberg J, Gobel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbs D, Mccray PB Jr, Cathomen T, Voytas DF, Joung JK (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell 31:294–301

Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A 108:2623–2628

Malech HL, Maples PB, Whiting-Theobald N, Linton GF, Sekhsaria S, Vowells SJ, Li F, Miller JA, Decarlo E, Holland SM, Leitman SF, Carter CS, Butz RE, Read EJ, Fleisher TA, Schneiderman RD, van Epps DE, Spratt SK, Maack CA, Rokovich JA, Cohen LK, Gallin JI (1997) Prolonged production of NADPH oxidase-corrected granulocytes after gene therapy of chronic granulomatous disease. Proc Natl Acad Sci U S A 94:12133–12138

Mashimo T, Takizawa A, Voigt B, Yoshimi K, Hiai H, Kuramoto T, Serikawa T (2010) Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. PLoS One 5:e8870

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol 26:695–701

Meyer M, de Angelis MH, Wurst W, Kuhn R (2010) Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. Proc Natl Acad Sci U S A 107:15022–15026

Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, Gregory PD, Pabo CO, Rebar EJ (2010) An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25:778–785

Mittelman D, Moye C, Morton J, Sykoudis K, Lin Y, Carroll D, Wilson JH (2009) Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. Proc Natl Acad Sci U S A 106:9607–9612

Moehle EA, Rock JM, Lee YL, Jouvenot Y, Dekelver RC, Gregory PD, Urnov FD, Holmes MC (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. Proc Natl Acad Sci U S A 104:3055–3060

Morton J, Davis MW, Jorgensen EM, Carroll D (2006) Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. Proc Natl Acad Sci U S A 103:16370–16375

Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501

Munoz IG, Prieto J, Subramanian S, Coloma J, Redondo P, Villate M, Merino N, Marenchino M, D'Abramo M, Gervasio FL, Grizot S, Daboussi F, Smith J, Chion-Sotinel I, Paques F, Duchateau P, Alibes A, Stricher F, Serrano L, Blanco FJ, Montoya G (2011) Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. Nucleic Acids Res 39:729–743

Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res 39(21):9283–9293

Ochiai H, Fujita K, Suzuki K, Nishikawa M, Shibata T, Sakamoto N, Yamamoto T (2010) Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. Genes Cells 15:875–885

Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kuhlcke K, Schilz A, Kunkel H, Naundorf S, Brinkmann A, Deichmann A, Fischer M, Ball C, Pilz I, Dunbar C, Du Y, Jenkins NA, Copeland NG, Luthi U, Hassan M, Thrasher AJ, Hoelzer D, von Kalle C, Seger R, Grez M (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med 12:401–409

Papapetrou EP, Lee G, Malani N, Setty M, Riviere I, Tirunagari LM, Kadota K, Roth SL, Giardina P, Viale A, Leslie C, Bushman FD, Studer L, Sadelain M (2011) Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. Nat Biotechnol 29:73–78

Paques F, Duchateau P (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gene Ther 7:49–66

Pattanayak V, Ramirez CL, Joung JK, Liu DR (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods 8:765–770

Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, Wang N, Lee G, Bartsevich VV, Lee YL, Guschin DY, Rupniewski I, Waite AJ, Carpenito C, Carroll RG, Orange JS, Urnov FD, Rebar EJ, Ando D, Gregory PD, Riley JL, Holmes MC, June CH (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26:808–816

Porteus MH (2006) Mammalian gene targeting with designed zinc finger nucleases. Mol Ther 13:438–446

Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763

Potts PR, Porteus MH, Yu H (2006) Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. EMBO J 25:3377–3388

Pruett-Miller SM, Connelly JP, Maeder ML, Joung JK, Porteus MH (2008) Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. Mol Ther 16:707–717

Pruett-Miller SM, Reading DW, Porter SN, Porteus MH (2009) Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. PLoS Genet 5:e1000376

Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA, Cathomen T, Voytas DF, Joung JK (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods 5:374–375

Rebar EJ, Greisman HA, Pabo CO (1996) Phage display methods for selecting zinc finger proteins with novel DNA-binding specificities. Methods Enzymol 267:129–149

Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK (2012) FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol. Apr 8. doi: 10.1038/nbt.2170

Rouet P, Smih F, Jasin M (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol 14:8096–8106

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, Curtin SJ, Blackburn JS, Thibodeau-Beganny S, Qi Y, Pierick CJ, Hoffman E, Maeder ML, Khayter C, Reyon D, Dobbs D, Langenau DM, Stupar RM, Giraldez AJ, Voytas DF, Peterson RT, Yeh JR, Joung JK (2011)

Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8:67–69

Sanjana NE, Cong L, Zhou Y, Cunniff MM, Feng G, Zhang F (2012) A transcription activator-like effector toolbox for genome engineering. Nat Protoc. Jan 5;7(1):171–92. doi: 10.1038/nprot.2011.431

Smith J, Grizot S, Arnould S, Duclert A, Epinat JC, Chames P, Prieto J, Redondo P, Blanco FJ, Bravo J, Montoya G, Paques F, Duchateau P (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 34:e149

Sollu C, Pars K, Cornu TI, Thibodeau-Beganny S, Maeder ML, Joung JK, Heilbronn R, Cathomen T (2010) Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. Nucleic Acids Res 38:8269–8276

Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol 25:786–793

Takasu Y, Kobayashi I, Beumer K, Uchino K, Sezutsu H, Sajwan S, Carroll D, Tamura T, Zurovec M (2010) Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection. Insect Biochem Mol Biol 40:759–765

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435:646–651

Wright DA, Townsend JA, Winfrey RJ Jr, Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD, Voytas DF (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 44:693–705

Young JJ, Cherone JM, Doyon Y, Ankoudinova I, Faraji FM, Lee AH, Ngo C, Guschin DY, Paschon DE, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Harland RM, Zeitler B (2011) Efficient targeted gene disruption in the soma and germ line of the frog Xenopus tropicalis using engineered zinc-finger nucleases. Proc Natl Acad Sci U S A 108:7052–7057

Zou J, Maeder ML, Mali P, Pruett-Miller SM, Thibodeau-Beganny S, Chou BK, Chen G, Ye Z, Park IH, Daley GQ, Porteus MH, Joung JK, Cheng L (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell 5:97–110

Zou J, Sweeney CL, Chou BK, Choi U, Pan J, Wang H, Dowey SN, Cheng L, Malech HL (2011) Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. Blood 117:5561–5572

Chapter 14
Bio-applications Derived from Site-Directed Genome Modification Technologies

Christophe Delenda, Sébastien Paris, Sylvain Arnould, Edward Balbirnie, and Jean-Pierre Cabaniols

Abstract This review summarizes the overall downstream applications of currently available genome customization systems, based on cell-based assays (*in cellulo*) or animal models (*in vivo*). These include (i) functional genomics, (ii) drug discovery, (iii) bioproduction, (iv) cell transformation (i.e. immortalization, reprogrammation and differentiation), as well as (v) molecular biology and microbiology tools. The different enzymatic systems that exist to specifically modify (*insertional* or *site-directed mutagenesis*), modulate (*knock-down*) and disrupt (*constitutive or conditional knock-out*) cellular genes, or to integrate transgenic elements (*knock-in*) at specific chromosomal loci will be discussed herein.

Keywords Meganucleases • Transcription activator-like effector nucleases (TALENs) • Zinc finger nucleases (ZFNs) • Recombinases • Integrases • AAV vector • Triplex-forming oligonucleotides (TFOs)

C. Delenda (✉) • S. Paris • J.-P. Cabaniols
Cellectis Bioresearch, Biopark, 8 rue de la Croix de Jarry, 75013 Paris, France
e-mail: christophe.delenda@cellectis.com; sebastien.paris@cellectis.com; cabaniols@cellectis.com

S. Arnould
Cellectis Therapeutics, Biopark, 8 rue de la Croix de Jarry, 75013 Paris, France
e-mail: sylvain.arnould@cellectis.com

E. Balbirnie
Ectycell, 5 rue Henri Desruères, Genopole, 91030 Evry, France
e-mail: edward.balbirnie@cellectis.com

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*, Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5_14,

14.1 Introduction

Many different enzymatic systems exist to specifically modify (*insertional* or *site-directed mutagenesis*), modulate (*knock-down*) and disrupt (*constitutive or conditional knock-out*) cellular genes, or to integrate transgenic elements (*knock-in*) at specific chromosomal loci. As an introduction to what will be discussed in this review, details for each genome personalization system are given in Table 14.1, wherein each of these systems are referenced based on the specificity of their target site. Today, three systems based on DNA-binding enzymes exist which can be used for genome modifications:

- those of natural origin that require the pre-integration of exogenous target sites, such as recombinases [e.g. Cre (Sauer and Henderson 1989) and FLP (Umlauf and Cox 1988)], or meganucleases, also known as homing endonucleases [e.g. I-*SceI* (Kuhlman and Cox 2010; Cabaniols and Paques 2008) and I-*CreI* (Cabaniols et al. 2009)] ;
- those of natural origin that can address various fixed endogenous target sites, such as integrases [PhiC31 (Belteki et al. 2003), R4 (Olivares et al. 2001), TP901-1 (Stoll et al. 2002) and Bxb1 (Kim et al. 2003)] ;
- those that can be engineered to recognize unique endogenous target sites, such as zinc finger nucleases (ZFNs) (Durai et al. 2005), meganucleases (Arnould et al. 2007), and very recently transcription activator-like effector nucleases (TALENs) (Christian et al. 2010).

In addition to these enzymatic systems, nucleotide-based technologies have also been used for genome modifications:

- adeno-associated (AAV) vectors, which can exert homologous recombination (HR) when homology arms are included in the backbone of the viral genome (Vasileva and Jessberger 2005);
- triplex-forming oligonucleotides (TFOs) (Postel et al. 1991). In this specific class of nucleotide-based methods, small DNA fragments are used to obtain small fragment homologous replacement (SFHR) (Colosimo et al. 2001).

This review summarizes the overall downstream applications of currently available genome customization systems, based on cell-based assays (*in cellulo*) or animal models (*in vivo*). These include (i) functional genomics, (ii) drug discovery, (iii) bioproduction, (iv) cell transformation (i.e. immortalization, reprogrammation and differentiation), as well as (v) molecular biology and microbiology tools.

Issues related to plant applications will not be treated herein since they are discussed in another review associated within this book.

Table 14.1 Description of site-directed genome modification systems. The nature for each system refers to enzymes (integrases, meganucleases, recombinases, TALENs, transposases or TALENs) or nucleotide-based elements (AAV vectors, TFOs). The origin, as well as the target site, for each dedicated system are also indicated

Nature	Name	Target site	Origin
Adeno-associated (AAV) vectors	Custom	Varying	Found as episomes in various animal species
Integrases	PhiC31	AttB (34 bp), AttP (39 bp) Fixed and multiple pseudo-attP sites (varying lengths) in a given genome	*Streptomyces* temperate phage
	R4	AttB (50 bp), AttP (49 bp) Fixed and multiple pseudo-attP sites (varying lengths) in a given genome	*Streptomyces parvulus* actinophage R4
	TP901-1	AttB (43 bp), AttP (56 bp) Fixed and multiple pseudo-attP sites (varying lengths) in a given genome	*Lactococcus lactis subsp* phage
Meganucleases	I-*CreI*	CAAAACGTCGTGAGACAGTTTG (22 bp)	*Clamydomonas reinhardtii*
	I-*SceI*	TAGGGATAACAGGGTAAT (18 bp)	*Saccharomyces cerevisiae*
	Custom	Varying (22 bp)	Derived from I-*CreI* or other meganuclease backbone
Recombinases	Cre	*loxP* (34 bp) ATAACTTCGTATAGCATACATTATACGAAGTTAT consisting of two 13 bp inverted repeats flanking an 8 bp asymmetric spacer	Bacteriophage P1
	Flp	FRT (34 bp) GAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC consisting of two 13 bp inverted repeats flanking an 8 bp asymmetric spacer	*Saccharomyces cerevisiae* 2 μ plasmid

(continued)

Table 14.1 (continued)

Nature	Name	Target site	Origin
TALENs	Custom	Varying (24–26 bp) Separated by 12–30 base pairs to allow formation of a catalytically active nucleasic domain	Derived from *Xanthomonas* proteins linked to a non-specific DNA-cleavage domain
Transposases	*Sleeping beauty*	TA	Tc1/*mariner*-like reconstructed from fish *in vitro*
	Piggyback	TTAA	Found in the genome of various animal species
	Harbinger3_DR	AAACACCWGGTCTTT	
Triplex-forming oligonucleotides (TFOs)	Custom	Varying	Synthetic
Zinc finger nucleases	Custom	Varying, 12–18 bp separated by 4–7 base pairs to allow formation of the catalytically active *FokI* dimer	DNA-binding domain composed of three Cys_2His_2 zinc fingers linked to the non-specific DNA-cleavage domain from *FokI*

14.2 Functional Genomics

Four types of gene modifications are commonly used to better understand the function(s) of a gene: overexpression (*knock-in*), disruption (*knock-out*), mutation (*site-directed mutagenesis*) or modulation of its expression (*knock-down*). These various genome modifications mechanisms have been performed either *in cellulo* or *in vivo* for generation of cell lines and animal models, respectively.

14.2.1 Transgenic Overexpression Strategies

To better understand gene functionality, the classic approach is to over-express the gene of interest (*knock-in*). This approach contrasts with the 'forward genetics' strategy where the genotype is identified from the phenotype by positional cloning experiments. This 'reverse genetics' approach determines the function(s) of a gene, therefore allowing phenotype prediction of a cell, tissue or organism.

As opposed to standard transgenesis techniques such as transient or stable transfections, the benefits of using methods based on targeted integration allow better mimicking of the physiological state of expression of the transgene of interest. It has now become possible to specifically target a transgene expression cassette at a unique chromosomal locus for which transcriptional activity has been demonstrated. As mentioned above, two different genomic regions can be targeted either at pre-integrated molecular platforms (also called "landing pads") or at endogenous loci (which can only now be targeted by custom-made DNA nucleases).

14.2.1.1 By Targeting a Pre-integrated Molecular Platform

Many laboratories working on functional genomics have benefited from using natural enzymes that facilitate homologous recombination (HR) to generate cell-based assays wherein a pre-integrated DNA fragment containing the target recognition sites has been randomly integrated into the cellular genome. This technology, referred to as 'recombinase-mediated cassette exchange' (RMCE), includes *loxP* sites [Cre recombinase (Liu et al. 2006a; Sandhu et al. 2011)] or FRT sites [Flp recombinase (Turan et al. 2010; Roebroek et al. 2011), as in the Flp-in™ system developed by Invitrogen, http://products.invitrogen.com/ivgn/product/K601001]. To be effective, these recombinases must be used at very high concentrations, which, in the case of Cre, triggers endonucleolytic activity and therefore cellular toxicity. Very recently, Turan et al. (2011) have modified the recombination route of Flp and Cre recombinases, by changing the structure of the synaptic complex based on improved integrases and associated target site variants (Turan et al. 2011). While the performance of Flp-RMCE has been optimized by such *trans*- (optimized Flp variants) and *cis*-active (multiple sets of functional FRT target sites) mutations, the Cre system still suffers from the infidelity of its target site mutants.

By adapting molecular platforms to dedicated selection processes, Cellectis bioresearch has used target sites recognized by wild-type meganucleases (I-*CreI* or I-*SceI*) to develop a cellular genome positioning system (cGPS®) in several cell lines (Cabaniols and Paques 2008; Cabaniols et al. 2009). The pre-engineering of cell lines offers many advantages when several different genes need to be evaluated, since robust, fast and reproducible protocols are available for each of these targeted integration platforms (http://www.cellectis-bioresearch.com/genome-customization/targeted-integration/cgps). In contrast to the Flp-in™ system, the cGPS® technology maintains sustained transgenic expression, without any observed mosaic pattern of expression.

Because stable integration of a landing pad has been based upon random transgenesis processes, procedures for selecting such pre-engineered cell lines are often difficult or cumbersome to define because of the limitations inherent to this type of process (i.e. stable transfection using plasmid-based elements encoding a resistant cell marker gene), such as the requirement to isolate monocopy integration events at chromosomal regions that need to be accessible for downstream HR and active for transgene expression. Cellectis bioresearch has overcome these limitations by integrating the landing pad element by HR in endogenous chromosomal regions for which meganuclease and transgene activities have been demonstrated (See following section).

14.2.1.2 By Targeting Unique and Modular Endogenous Chromosomal Loci

With regard to engineered DNA nucleases, targeted chromosomal regions can be chosen based on their absence of transcriptional activity, whether they are derived from intergenic sequences or from gene regions that are transcriptionally inactive in the cell type. This approach specifically targets inactive regions and has the advantage that no promoter interference between endogenous and exogenous elements occurs, which could be detrimental to the induction of the transgene and endogenous expression, resulting in possible changes to the metabolic pattern of targeted cells.

The “Safe Harbor” (SH) approach is based on the introduction of genetic material into a pre-defined locus unrelated to a disease-causing gene. The targeted locus should allow efficient and stable expression, to permit functional correction and at the same time remain “safe”, i.e. not interfere with normal cell function and gene regulation. Potential safe harbors include the AAVS1 locus, the most common integration site for AAV, or alternatively loci distant from functional genes. ZFNs have been developed for AAVS1 (DeKelver et al. 2010) and meganucleases are currently being developed for SH regions located at least 200 kb from surrounding genes.

Transcriptionally inactive regions have also been considered by specifically targeting the RAG1 gene using an engineered meganuclease in non-T cell lines, such as HEK-293, HCT116, HepG2 and MRC5. The cGPS® Custom system (http://www.cellectis-bioresearch.com/genome-customization/targeted-integration/cgps-custom) developed by Cellectis bioresearch has also been applied to some hematopoietic cell lines derived from T cell (Jurkat) or chronic myeloid (K562)

leukemia. This targeted integration is possible because the RAG1-based meganuclease reserves most double-stranded breaks (DSB) to one allele in these dedicated cell lines, whilst maintaining a basal RAG1 endogenous activity from the untargeted allele.

Cellectis bioresearch has also designed target sites within genes that are ubiquitously expressed, allowing development of cell based assays in as many cell lines as possible. Two main strategies have been developed for the selection of meganuclease *in silico* hits, and for which the read-out system is equivalent (i.e. mRNA microarrays): one resulted from the meta-analysis of the 60 cell lines derived from the NCI60 reference list (available via http://biogps.gnf.org), and the other one from the selection of 575 house-keeping genes (Eisenberg and Levanon 2003). In addition to the benefit of allowing normal transgene expression, this type of targeting design should also help chromosomal meganucleasic activity since transcriptionally active regions are mainly restricted to decondensed euchromatin regions.

To assist in the development of cell line models that will better mimic the physiological state of gene expression for which its functions are unknown, Cellectis bioresearch is evaluating, at each given targeted locus, the transcriptional level of transgene expression. In Fig. 14.1a the luciferase expression values obtained from various isogenic cell lines targeted at a given chromosomal locus (i.e. in the RAG1 gene) by an integration matrix that contains a CMV-luciferase expression cassette are summarized. Whereas homogeneous expression was obtained from three independent cell clones generated for each individual set of integration, variable levels of luciferase expression were measured in all tested cell lines. Additionally, by changing the type of exogenous promoter [i.e. promoting sequences from the cytomegalovirus (pCMV), the SV40 virus (pSV40), the elongation factor 1α gene from its long (pEF1l) or short versions (pEF1s), or the growth arrest-specific 5 gene (pGAS5)] that drives the transgene expression at a given locus [i.e. in the dystrophin, muscular dystrophy (DMD) gene] and for a given cell line (i.e. HEK-293), variable levels of luciferase expression have been obtained (Fig. 14.1b), suggesting that modulation of transgene expression is possible.

For *in cellulo* protein-protein interaction studies, it is interesting to study the simultaneous functionality of two genes integrated at separate chromosomal locations. By applying specific selection procedures, Cellectis bioresearch has recently developed such dual targeted integration kits (namely cGPS® Custom Duo systems), with the example given in Fig. 14.2 wherein two different reporter genes have been iteratively targeted at the RAG1 (luciferase gene) and at the DMD (lacZ gene) chromosomal loci of HEK-293 cells.

Finally, despite of the emerging use of engineered DNA nucleases for their *knock-in* applications in genomics studies, direct gene targeting has not been addressed to date, with the exception of gene repair demonstrations. It is now possible, to fuse the protein of a known cell/organ with a fluorescent protein or luciferase markers dedicated to *in cellulo* or *in vivo* imaging studies. In the field of *in cellulo*-based imaging applications, Cellectis bioresearch is currently generating dedicated cell lines wherein specific organelle markers fused with living color genes (from the license acquired from Evrogen, http://www.evrogen.com/products/Organelle-

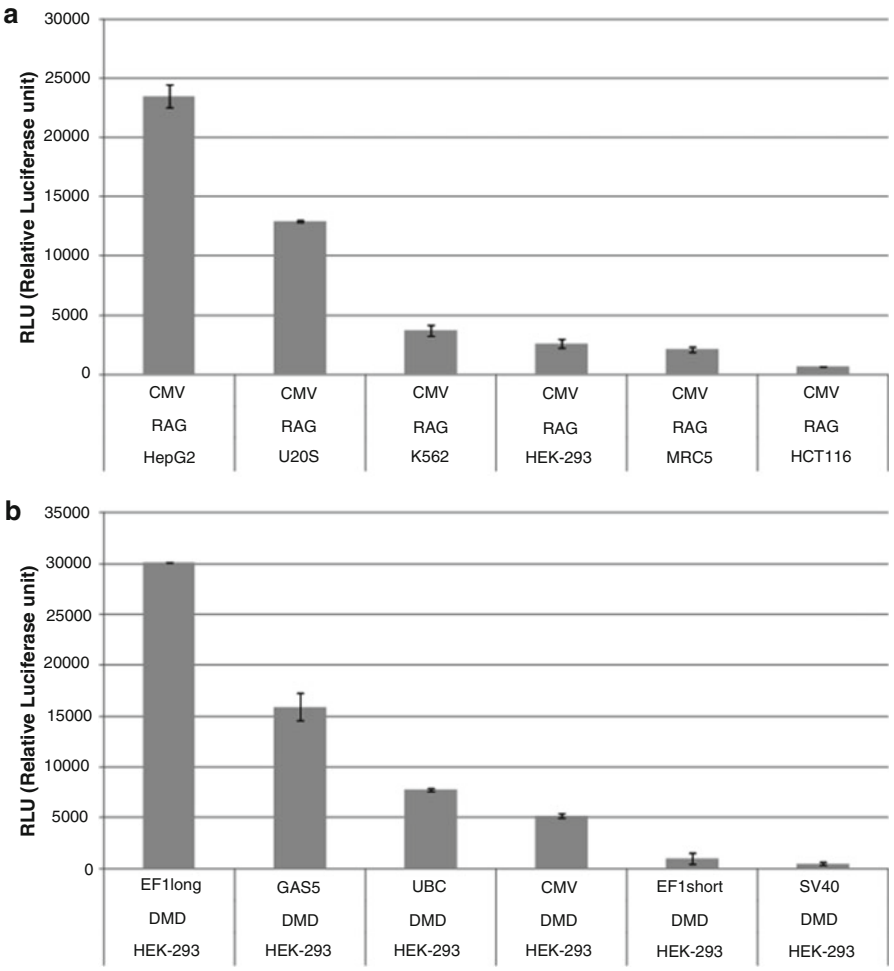


Fig. 14.1 Evaluation of the levels of transgene expression from meganuclease-driven targeted isogenic integration, depending on various parameters, i.e. cell line, chromosomal locus and exogenous promoter driving the transgenic (luciferase) transcription. Targeted integration has been given according to the settlement of specific selection procedures, wherein the cGPS® Custom technology has been applied and from which positive and negative selection schemes have been specifically addressed for each given cell type and for each targeted locus. Three different selected cell clones (triplicates) for each set of targeted integration have been analyzed. Luciferase expression is given as relative luciferase units (RLU). (**a**) Comparison of various human cell lines targeted at the same chromosomal locus (RAG1) and for which the luciferase expression is driven by the same exogenous promoter (pCMV). (**b**) Comparison of various exogenous promoters (pCMV, pEF1long, pEF1short, pGAS5, pUBc and pSV40) targeted at the same chromosomal locus (RAG1) of HEK-293 cells

Fig. 14.2 (continued) at the RAG1 locus) was subsequently submitted to a second integration process (neomycin-LacZ at the DMD locus). (**a**) Eighteen cell clones were analyzed by Southern blotting (hygromycin and neomycin probes) to confirm the genetic pattern of isogenic integration. (**b**) Four final independent isogenic cell clones were tested to evaluate their stability of expression over time (until passage P22) for both reporter genes

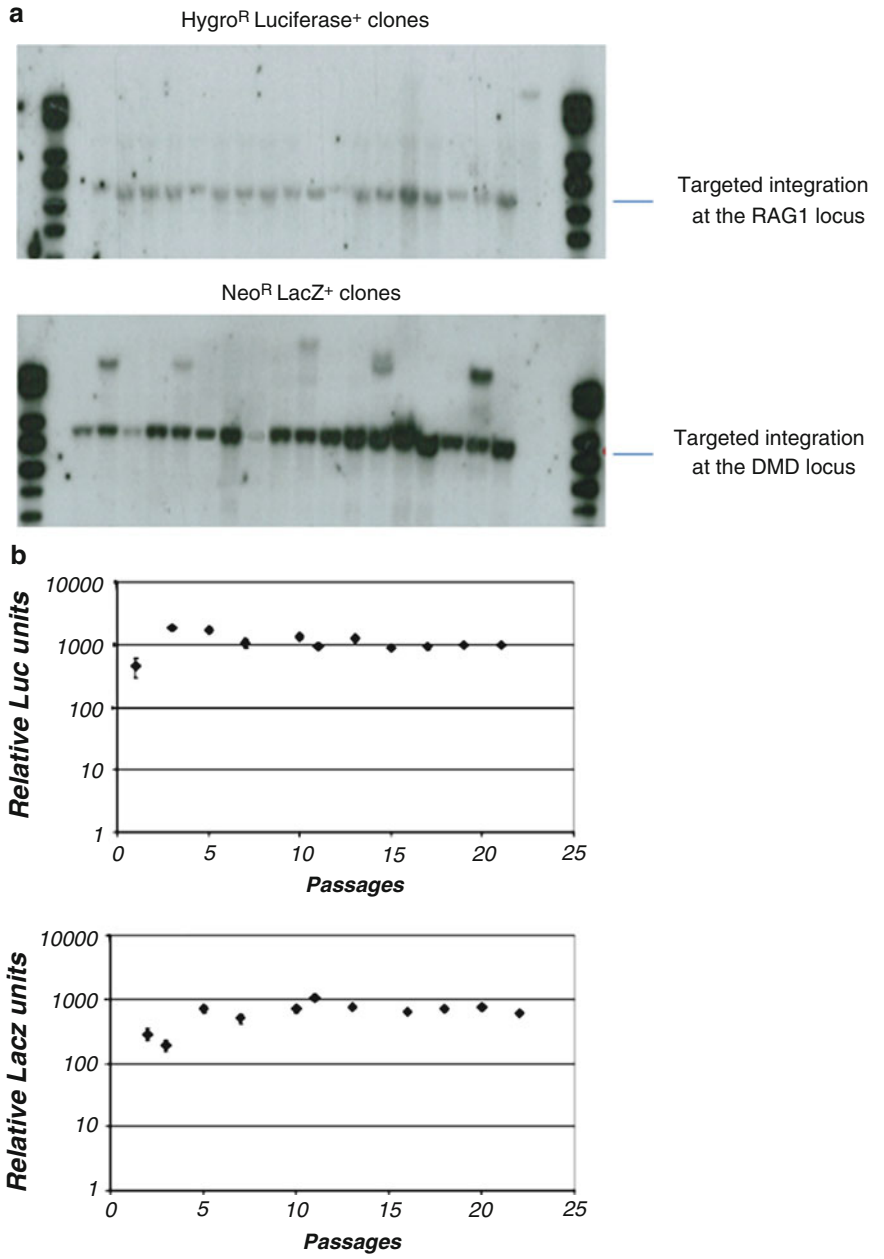


Fig. 14.2 Targeted integration at two independent chromosomal loci using engineered meganucleases. By applying specific selection conditions (cGPS® Custom systems), two different integration matrices have been used for stable integration of two reporter gene expression cassettes at two separate endogenous chromosomal regions (i.e. at RAG1 and DMD gene loci) of HEK-293 cells. Each integration matrix contains the 5′ and 3′ homology arms (of around 1 kb) dedicated for each chromosomal locus. Two gene expression cassettes were introduced into the recombinogenic element, one encoding the reporter gene and the other one driving the expression of a selectable marker (hygromycin-luciferase and neomycin-LacZ for the integration at the RAG1 and DMD loci, respectively). One isogenic cell clone derived from the first targeted integration (hygromycin-luciferase

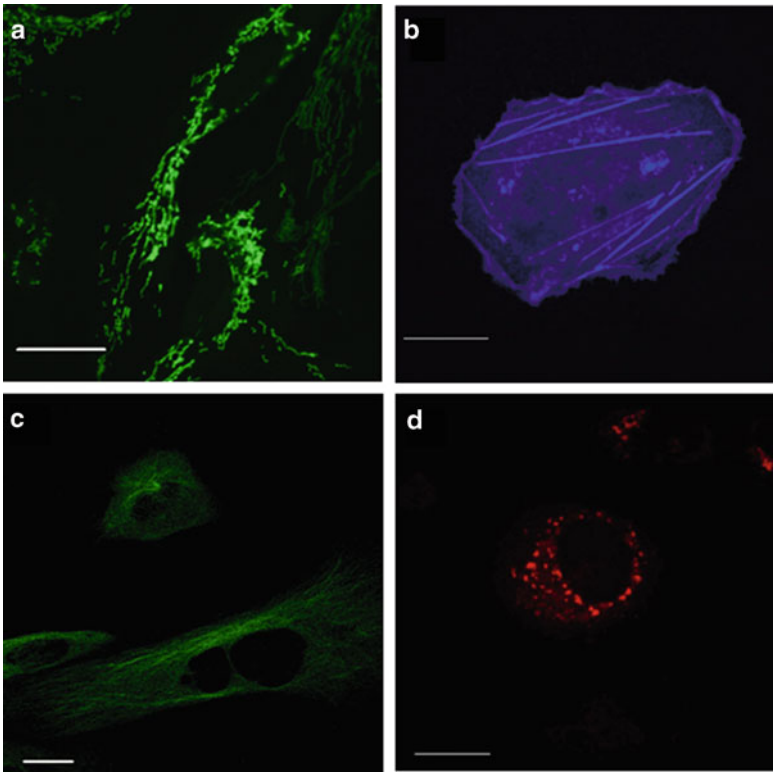


Fig. 14.3 Meganuclease-mediated targeted integration of fluorescent organelle markers and confocal analysis of such selected cell clones. Independent integrations of four different marker genes targeted the DMD locus of U2OS cells: (**a**) GFP-mito, (**b**) BFP-actin, (**c**) GFP-tubulin, and (**d**) RFP-golgi. Scale bar: 20 μm

labeling.shtml) are being individually targeted to cGPS® Custom sites of various cell lines. In Fig. 14.3, organelle-specific image examples obtained from the isogenic targeted integration of two different markers are given, i.e. tubulin-GFP and actin-RFP expression cassettes stably integrated at the DMD locus of U2OS cells. By using the cGPS® Duo technology, it will be possible to target, at another chromosomal locus, the gene of interest to learn more about its susceptible co-localization with control organelle proteins. The next step will be to address the integration of fluorescent marker genes by fusing them to the transcriptional unit of endogenous genes, as it is currently developed by Sigma with their CompoZr® cytoskeletal marker cell lines (http://www.sigmaaldrich.com/life-science/cells-and-cell-based-assays/compozr-cytoskeletal-marker-cells.html).

14.2.1.3 By Targeting Multiple and Fixed Endogenous Chromosomal Loci

In contrast to the previously described technologies that have been developed to allow isogenic (monocopy) transgene integration, several different phage-derived serine integrase enzymes have been used to integrate transgenic elements at multiple and fixed integration sites (i.e. pseudo-attP sites) in the genome of several different mammalian cells. Unlike the better-known recombinases such as Cre and Flp, these integrases catalyze recombination between two non-identical sites which, combined with the lack of a corresponding excisionase enzyme, makes the integration events unidirectional and virtually irreversible. The number and location of such target sites varies depending on the integrase used and on the inherent genomic differences between species.

The ability of PhiC31 integrase to mediate transgene integration has been demonstrated in different cell types or organs e.g. by producing dystrophin in mouse muscle-derived stem cells, human myoblasts and mouse muscle (Quenneville et al. 2004; Bertoni et al. 2006; Oumard et al. 2006; Birling et al. 2009). Using the PhiC31 integrase model, Invitrogen has developed and commercialized its Jump-In™ kit for integrating a transgene expression cassette at various fixed pseudo-attP sites in mammalian cells (http://products.invitrogen.com/ivgn/product/A10893). However, because PhiC31 integrase recognizes 100–1000 sites in human and mouse genomes that possess partial sequence identity to attP or attB (Liu et al. 2009), this integrase system is not particularly relevant for genome engineering of specific genomic regions nor suitable for isogenic (monocopy) integrations.

Recent work performed on other types of serine-based integrases - from R4 (Olivares et al. 2001), TP901-1 (Stoll et al. 2002) and Bxb1 phages (Russell et al. 2006) demonstrate that these enzymes mediate HR at heterotypic attB/attP binding sites in mammalian cells. However, although a high recombination efficiency of the PhiC31 integrase has been noticed (87% when compared to the values obtained by the Cre recombinase or nuclease-based enzymes), the other integrases have demonstrated more limited success, and thus the broad utility of such recombinases as research tools still remains to be established.

Transposon elements have also been routinely used to target various chromosomal sites via their inverted terminal repeats (ITRs). Different transposons have been employed for gain-of-function or over-expression analyses in different animal species (Bernstein et al. 1995; Huth et al. 1997; Luo et al. 1998; Mates et al. 2007; Clark et al. 2007; Sauka-Spengler and Barembaum 2008; Ivics et al. 2009). Many studies have described the ability of some transposon elements and their related factors to generate new functional genetic elements, such as genes, exons, introns and regulatory sequences. The integration pattern of most transposons has been shown not to be random, with characteristic preferences for integration sites at the primary DNA sequence level, and where *hot spots* and *cold regions* have been observed on a genome-wide scale. Whilst the *Tol2* transposon does not appear to exhibit a pronounced preference for any integration site (Grabundzija et al. 2011), the *Harbinger3_DR* transposon element is highly specialized at integrating into the palindromic AAACACCWGGTCTTT consensus sequence (Sinzelle et al. 2008),

the *piggyback* targets the sequence TTAA (Galvan et al. 2009) and all Tc1/*mariner* transposons (i.e. *Sleeping Beauty*, *Hsmar1*, *Mboumar-9*…) are specifically integrated into TA dinucleotides (Liu et al. 2005 , Munoz-Lopez et al. 2008; Claeys Bouuaert and Chalmers 2010). A database exists that details the insertion sites of the *Sleeping beauty* transposon that were determined from both mouse cancer models and mouse germlines (http://transposon.abcc.ncifcrf.gov/index.html).

After evaluating 300,000 transposon elements insertions, Levy et al. (2010) have shown that certain transposon elements prefer to insert into specific regions and in specific orientations (Levy et al. 2010). Similarly, a recent publication noted that a transposon (i.e. *MosI*) behaved differently depending on whether it was applied *in vitro* or *in vivo* (Crenes et al. 2010). *In vitro*, as for other mariner-like transposons, no specific feature determining target choice other than the requirement for a TA dinucleotide was identified. However, some specific targeted integration sites are preferred *in vivo*, such as selected regions located in the bacterial *cat* gene and the *Caenorhabditis elegans* rDNA locus. Targeted integration has also been optimized, through the use of transposases coupled to defined DNA-binding domains, to favour integration events to regions recognized by the chosen DNA-binding domain (Wu et al. 2008).

Although the transgenic methods described in this section have demonstrated certain benefits for increasing the number of transgenic cell and animal models, since some of these targeted systems (e.g. transposons and serine recombinases) result in multiple and/or fixed integration sites, they do not represent accurately physiological expression conditions.

14.2.1.4 By Disposing of Facilitating Transgenesis Systems

Several advances have been made which have increased transgenesis frequencies in various animal models. This up-dated section will only comment on the emerging technologies (or optimized systems applied to currently available methods), for which site-specific activity or recombination has been involved.

In animal model systems currently being developed and generated, functional genomics is supported by the ability to develop congenic inbred lines, cloning and generating mutants by either deleting (cf. *knock-out* section, Sect. 2.) or by incorporating/substituting specific genes. The current method that has been mainly used for the generation of transgenic animals is based upon the use of microinjection/electroporation of large DNA inserts (BAC and YAC vectors) or plasmids that contain the transgene expression cassette flanked by dedicated homology arms. Although many animal models (mainly resumed for rodents, mice and rats) have been generated by this approach, this technique cannot be adequately applied to other types of animal models because of the very low efficiency of HR when parallel DNA nucleasic activity is not possible.

One interesting method for increasing transgenesis rates consists of *in vivo* linearization of the DNA insert that contains the transgenic cassette flanked by restriction sites. Mostly used for plants and fungi, this technique called restriction

enzyme mediated integration (REMI) has recently been applied to the generation of transgenic frogs (Ishibashi et al. 2007; Amaya and Kroll 2010). Because of the limited length of the restriction target site (8 bp) used, a large number of laboratories who are generating transgenic sea animals [zebrafish (Babaryka et al. 2009; Grabher et al. 2004), medaka (Thermes et al. 2002; Rembold et al. 2006), urchin (Ochiai et al. 2008)], amphibians [xenopus (Pan et al. 2006; Kashiwagi et al. 2010)] or insects (drosophila) are now using the I-*SceI* meganuclease (commercialized by NEB, http://www.neb.com/nebecomm/products/productR0694.asp or Fermentas/ThermoScientific, http://www.fermentas.com/en/products/all/conventional-restriction-enzymes/nicking-homing/er177), which targets a larger sequence (18 bp). One important feature of this technology is that the I-*SceI* recombinant protein is directly co-injected *in ovo* along with the transgenic DNA, thus reducing the time for enzymatic activity, when compared to expression signals obtained from mRNA- or more particularly from plasmid DNA-encoding elements. The improved performance achieved when I-*SceI* recombinant proteins has allowed a substantial decrease of mosaicism in animal models.

As for transposon-mediated genomics approaches, although both REMI and I-*SceI* technologies have been especially dedicated to insertional mutagenesis applications (cf *knock-out* next section, Sect. 1.2.), they still remain effective tools for randomly integrating transgenes or reporter genes into the genome of several animal species.

14.2.2 Gene Disruption and Site-Directed Mutagenesis

To obtain a better understanding of gene function null mutations (*knock-out*) or other mutations that disrupt specific domains or residues (*site-directed mutagenesis*) are required. Because some gene knock-outs are lethal for animal development, they do need controlled induction either at a development stage or at a restricted targeted tissue that does not suffer for downstream knock-out (*conditional knock-out*). For a better understanding of mutagenic outcomes, random integration at different sites of the genome is also routinely used (*insertional mutagenesis*), as well as chromosomal rearrangements induced by excision of pre-integrated molecular platforms (*excisional mutagenesis*).

14.2.2.1 By Homologous Recombination

Originally, these modifications were addressed by HR with dedicated DNA matrices, with the best example being the use of BAC or YAC vectors that either contain a large deletion in catalytic exons (*knock-out* by *knock-in*) or unique points mutations (*site-directed mutagenesis*). By co-transfecting the integration matrix with a vector encoding a site-specific DNA nuclease, the frequency of targeted integration in mammalian cells has increased to levels greater than 1%.

The use of DNA nucleases to facilitate HR has been reported using DNA matrices with viral vector origins to give a 60- to 100-fold enhancement of AAV HR when I-*SceI* was co-delivered (Miller et al. 2003), or when non-integrative lentiviral vectors were used to specifically integrate at desired chromosomal loci of human stem cells using the DSB activity of engineered ZFNs (Lombardo et al. 2007). The use of viral vector backbones may be particularly well suited to cells that are either difficult to transfect or at a quiescent growth state.

Gene targeting has also been demonstrated using TFOs to specifically inhibit gene expression, enhance homologous recombination, induce mutation, inhibit protein binding, and direct DNA damage (see section Sect. 1.2.2.), providing a tool for gene-specific manipulation of DNA. Gaddis et al. (2006) have developed a flexible web-based search engine (http://spi.mdanderson.org/tfo) to find and annotate TFO target sequences within human and mouse genomes (Gaddis et al. 2006).

Very recent studies have indicated that the DNA recognition domain of transcription activator-like (TAL) effectors can be combined, using the zing finger-mediated strategy, with a nuclease catalytic domain to generate TAL effector nucleases (TALENs) that, in pairs, bind adjacent DNA target sites and produce double-strand breaks between the target sequences, stimulating HR and non-homologous end-joining. Several assays, developed in yeast using a single-stranded annealing assay, have shown that TALENs stimulate high rates (up to 34%) of gene replacement by HR. Professor Dan Voytas, CSO of Cellectis plant sciences, has been one of the pioneer researchers developing these tools for genome customization purposes (Christian et al. 2010; Cermak et al. 2011). Cellectis bioresearch, the other subsidiary of Cellectis that is dedicated to *in cellulo* applications, is offering TALENs as HR tools at affordable prices to the life sciences community (http://www.cellectis-bioresearch.com/talen).

By comparing the engineering of meganucleases and ZFNs, the frequency of obtaining *in silico* TALEN hits seems much more important because of their high modularity. As an example of gene targeting, Cermak et al. (2011) have successfully engineered a TALEN that can target a sequence which is in very close proximity to the DF508 CFTR mutation of the gene (Cermak et al. 2011). This offers an advantage over ZFNs, which target sites are at least 120 bp from the target site, which could compromise gene targeting efficiency.

14.2.2.2 By Non-homologous End-Joining

In addition to facilitating HR using enzymes that can induce double-stranded breaks (DSB) in chromosomes, DNA nucleases will be of even more interest, as soon as they can induce at their site of cleavage non-homologous end-joining (NHEJ), causing site-directed mutations, insertions or deletions, and therefore allowing frameshifts in the coding sequence and leading to gene *knock-out* when both alleles are targeted. Three different types of DNA nucleases including meganucleases (Bennardo and Stark 2010), ZFNs (Kandavelou et al. 2009) and TALENs (Li et al. 2011) have already been engineered and proven to be good NHEJ inducers.

Cellectis bioresearch offers custom engineering of meganucleases (http://www.cellectis-bioresearch.com/store/meganucleases) and TALENs for the knock-out of any given gene. Via its CompoZr® knock-out technology, Sigma Life Science is offering engineered ZFNs to permanently disrupt genes (http://www.sigmaaldrich.com/life-science/zinc-finger-nuclease-technology/knockout.html), from which cancer-related human cell lines have been derived. CompoZr® ZFNs are designed to target a sequence within the first two-thirds of the open reading frame ensuring that the resulting gene disruption will lead to knock-out.

The University of Iowa has created a user-friendly interface (ZFNGenome, http://bindr.gdcb.iastate.edu/ZFNGenome) that identifies and provides quality scores for all potential ZFN target sites in the complete genomes of several model organisms, *S. cerevisiae, C. reinhardtii, D. melanogaster, D. rerio, C. elegans, and H. sapiens*.

In addition to these enzymatic systems, oligonucleotide-based approaches have also been considered for gene knock-out. Pyrimidine TFOs with 2′-O-aminoethyl substitutions have been shown to have enhanced kinetics and stability of triplex formation, therefore allowing TFOs to be more active in gene (i.e. HPRT) knock-out assays in mammalian cells (Puri et al. 2001; Seidman et al. 2005).

14.2.2.3 By Site-Directed Chemical Modifications

Other types of damaging agents have been coupled to TFOs, including photoactivable agents (Perrouault et al. 1990), metal complexes such as Fe-EDTA (Strobel et al. 1991), orthophenantroline (Francois et al. 1989) or metalloporphirines (Bigey et al. 1995) to obtain gene mutagenesis. The most widely used chemical agent is psoralen, however, it suffers from limited reactivity (5′-TA or -AT sites) (Kean and Miller 1993), which limits the number of available target sites for gene mutagenesis. Similarly, TFOs based on the incorporation of a nucleoside derivative (2-amino-6-vinyl purine) have been shown to exhibit a high selectivity towards cytosines at a GC target site, and have recently been used to achieve site-specific modifications in the SupF reporter gene (Nagatsugi et al. 2003) or in the HER-2-neu promoter sequence (Ziemba et al. 2001).

14.2.2.4 By Conditional Gene Excision

Because the knock-out of some genes can be lethal *in vivo*, conditional knock-out has been applied, in which a gene is knocked out in a specific tissue or at a specific time. By limiting the expression of Cre or Flp recombinases to a specific tissue or cell type, and/or for a given time, based on the use of an inducible expression system, it has been possible to control the inactivation of a gene (Miller 2011; Wang 2009; Bockamp et al. 2008). The common strategy for conditional knockout of alleles is to flank with *loxP* or FRT sites at a proximal exon within the target gene, leading to frameshift and null allele mutations following Cre-mediated recombination.

Gene Bridges has opened up possibilities for the engineering of such target constructs, by using Cre- or Flp-derived recombination kits dedicated to conditional gene knock-out in animals (http://www.genebridges.com/gb/animal_targeting.php). A database from the University of Georgia has compiled abstracts and/or full text from a wide range of publications in the site-specific recombinase literature also includes publications in which site specific recombinases are used as a research tool, such as mouse strains that express Cre recombinase (http://ssrc.genetics.uga.edu).

Two animal lines are required for conditional gene excision: (i) a conventional transgenic line with Flp or Cre targeted to a specific tissue, and (ii) another transgenic line that embodies a target gene flanked by two *loxP* or FRT sites in a direct orientation. This kind of genome engineering is long and cumbersome; however, it could be simplified by exerting conditional gene knock-out using the use of two specific DNA nucleases (i.e. meganucleases, ZFNs, TALENs of modified specificity) that would recognize and excise the expected exon sequence. By limiting the DNA nuclease expression to one tissue type, the requirement to pre-engineer a second line of animals for nuclease expression would be eliminated. Glover et al. (2007) have successfully constructed trypanosome strains in which a single telomere was removed by conditional I-*SceI* meganuclease cleavage (Glover et al. 2007). In the specific context of meganucleases, the DNA nuclease activity can be delivered *in situ* using a recombinant protein format, since meganucleases are known to be stable at high temperatures (see also section Sect. 5.1.).

14.2.2.5 By Insertional Mutagenesis

Insertional mutagenesis is a versatile strategy for generating mutations, having the advantage that the integrated DNA fragment can serve as a molecular tag for the identification of the mutated allele.

Large-scale transposon mutagenesis is being applied to identify the function of genes in various organisms (including vertebrate species) for which controlled transposition has been possible (Hamer et al. 2001). Transposition efficiency depends on the type of transposon elements used and on the organism to be targeted. Fischer et al. (2001) have shown that the transgenic mouse progeny derived from *Tc1*-treated germline contained no transposition events, whereas when using *Sleeping Beauty* transposons, 20% of the progeny had new chromosomal re-arrangements (Fischer et al. 2001).

Retroviral vectors have also been used to introduce mutagenic cassettes into the genome of various species, but their chromosomal integration bias (i.e. in transcriptionally active chromatin regions) did not allow full gene coverage (Hansen et al. 2008).

14.2.2.6 By Excisional Mutagenesis

The application of the Cre-*loxP* recombination system over large distances in ES cells has made it possible to engineer specific chromosomal re-arrangements, particularly

in the mouse genome. The technology based on the Cre-*loxP* system, is now commonly used in the scientific community to generate new chromosomes carrying deletions, duplications, inversions and translocations in targeted regions of interest. For instance, this advance has opened up new opportunities for modelling human diseases that are associated with chromosomal rearrangements. Chromosome rearrangements can be tagged with visible markers, facilitating strain maintenance (Zheng et al. 2001).

In the specific context of the generation of mouse models of human cancers, Forster et al. (2005) have not only induced *de novo* reciprocal chromosomal translocations via Cre-*loxP* recombination (here referred as a 'translocator' system), but also developed a different model (an 'invertor' system) in cases where the gene orientation is incompatible or when the translocator model cannot be applied (Forster et al. 2005). This method involves introducing an inverted cDNA cassette into the intron of a target gene and bringing the cassette into the correct transcriptional orientation by Cre-*loxP* recombination.

Cre-*loxP* mediated chromosomal engineering in ES cells has a variety of applications, including the creation of model systems for studying aneuploidy. Targeted meiotic recombination (TAMERE) has been proposed (Olson et al. 2005), in which Cre recombinase, under the control of the synaptonemal complex one promoter, is expressed during male meiosis in transgenic mice. Although TAMERE has been successfully used *in vitro* with *loxP* sites up to 100 kb apart, it was ineffective at generating either a deletion or translocation *in vivo* and may be of limited use for large genomic rearrangements.

Harnessing the activity of customized homing endonuclease genes (i.e. I-*SceI*) has also been proposed as a method for spreading deleterious mutations through populations. This strategy was recently applied to *Drosophila melanogaster* where it was shown that high rates of homing can be achieved within the spermatogonial population of the testis and in the female germline (Chan et al. 2011).

As a more innovative and extendable approach, engineerable DNA nucleases such as meganucleases, TALENs or ZFNs could also be used for excisional mutagenesis, by designing and selecting enzymes with less target specificity allowing multiple chromosomal rearrangements. A first demonstration of this approach has already been described by Söllü et al. (2010), where two autonomous ZFN pairs have been directed simultaneously to two different sites to induce a chromosomal deletion in approximately 10% of alleles (Sollu et al. 2010).

14.2.3 Modulation of Transcription

It is now well-established that TFO-mediated gene targeting leads to major changes in the functionality of DNA (Duca et al. 2008). Indeed, TFOs have been shown to alter gene transcription *in vitro* by interfering with the binding of transcription factors. Transcription inhibition by TFOs has been observed for various types of target genes, i.e. for plasmid-harbored genes, for exogenous transgene elements stably integrated in the cellular genome or in several endogenous genes.

As an alternative mechanism, it has also been possible to induce transcription using TFOs by targeting repressor sites, thus leading to transcriptional stimulation of the endogenous gene of interest. Another approach is to activate transcription. This has been achieved by linking a promoter-specific TFO through a phosphoroamidate bond to several minimal transcriptional activation domains derived from HSV protein 16 (VP16) (Duca et al. 2006).

14.3 Drug Discovery

Site-specific recombinase technology has been widely used for high throughput drug screening campaigns through the Flp-In system commercialized by Invitrogen (Castaneda and Kinne 2005; Kapitskaya et al. 2006; Nair et al. 2008). The integration of a single copy of the construct is more physiologically relevant than previous screens made on transiently transfected cells or transduced cells (by plasmids and episomal viral vectors, respectively). However, heterogeneity of expression from clone to clone has been reported (Castaneda and Kinne 2005), and stability of Flp-derived cell lines is often compromised when the selective drug is removed from the culture medium (Liu et al. 2006b).

Transposons have also routinely been used for drug discovery applications. Recently, Staunstrup et al. (2011) used a genomically integratable transposon system for the assessment of the pharmacokinetic and pharmadynamic properties of vitamin D3 analogues (Staunstrup et al. 2011). A tri-cistronic *Sleeping beauty* transposon was cloned from which three different proteins are encoded: a drug-sensoring protein, a reporter protein expressed from an activated sensor-responsive promoter and a resistance marker allowing clonal selection. In clones of human keratinocytes carrying single to numerous integrations of the vitamin D3 sensor, even low concentrations of vitamin D3 analogues have been detected. In comparative studies, this sensor system revealed superior potency for new candidate drugs when compared with analogues that are currently in clinical use.

The cGPS® targeted integration system marketed by Cellectis bioresearch has also been used to express several types of human targets relevant to the screening of therapeutic compounds, including two GPCRs (melatonin receptors, MT1 and MT2) and a secreted enzyme (autotoxin, ATX) (Cabaniols et al. 2010). This case study performed in association with Servier Laboratories in CHO-K1 cells was compared to the standard stable transfection process (with pcDNA3.1 plasmid encoding neomycin resistance) and clearly showed several important advantages, including homogeneity and stability of the transgene expression, reduced time (less than 24 days to generate stable clones) and inter-assay reproducibility. It was also found that the proteins expressed from the cGPS retargeting locus of CHO-K1 cells have enzymatic (ATX) constants (K_m and V_{max}), or GPCR (MT1 and MT2) pK_D and EC_{50} values, close to the published values for their wild-type endogenous counterparts. In the case of pharmaceutical targets, achieving a physiological level of expression compatible with the relevant functional assay is clearly important.

Using the same model for developing retargeting platforms, PhiC31 and R4 integrases have also been used for specific drug screening applications. The phiC31 integrase was initially used to integrate, into a pseudo-attP site of three different cell line models (HEK-293, CHO-S and the human embryonic cell line BGO1V), a landing pad that contains a target site for the R4 integrase (Lieu et al. 2009). The R4 integrase was then used to incorporate into the R4 target site, three different genes of interest, i.e. a G-protein-coupled receptor (cholecystokinin receptor A, CCKAR), an ion channel (the transient receptor potential cation channel, subfamily M, member 8, TRPM8), and a GFP-c-Jun(1–79) fusion protein. Interestingly, all these retargeted cell lines exhibited functional and pharmacological responses consistent with those reported in the literature.

As an alternative technology, Yan et al. (2009) applied a gene-targeting method by using a recombinant AAV vector to integrate a luciferase reporter gene into the exon 1 of the tumor necrosis factor-alpha (TNFα) gene in HeLa cells (Yan et al. 2009). By comparing this gene targeting approach to randomly integrated transcriptional reporters, they have demonstrated that TNFα–targeted reporter activity reflected endogenous TNFα mRNA expression, whereas randomly integrated TNFα reporter cell lines gave variable expression in response to transcriptional and epigenetic regulators, therefore allowing predictive indexing of gene transcription for drug discovery.

In a review focusing on controlled genetic modification of human embryonic stem cells, some of these targeting methods are discussed for the development of drug discovery tools, including site-specific integration, HR, transposons and zinc finger nuclease systems (Zeng and Rao 2008).

14.4 Bioproduction

To satisfy the requirement to obtain large quantities of recombinant proteins from stable producer cell clones, two classic over-expression strategies are used to improve transgenic expression in mammalian cells. While the more widely used approach consists of the selection of highly transcriptionally active chromosomal regions (i.e. “hot spots”, *knock-in*), the second over-expression system is based upon transgene amplification by the down-regulation (or *knock-out*) of specific gene markers [i.e. the dihydrofolate reductase (DHFR) and the glutamine synthetase (GS) genes]. Recent advances are discussed where laboratories have undertaken these two related strategies, by taking advantages of site-directed genome modification methods (targeted integration for *knock-in* and mutagenesis for *knock-out*).

Moreover, a parallel gene modification approach, based on engineered DNA nucleases, is being applied to increase production yields in mammalian cells, by knocking-out pro-apoptotic genes such as BAX or BAK. Of specific interest to monoclonal antibody production, the FUT8 gene encoding the α1–6 fucosyl transferase is another interesting gene to knock-out since the removal of the α1,6 fucose core from the glycans in the Fc region of IgG1 antibodies has been demonstrated to improve antibody-dependent cellular cytotoxicity activity (Shinkawa et al. 2003).

14.4.1 Hot Spot Retargeting

By combining a Cre-*loxP* targeting platform and the DHFR amplification system [by inhibition with the methotrexate (MTX) drug], Mitsubishi Pharma has constructed engineered CHO strains that can be used for high-level production of foreign proteins (Kito et al. 2002). After transfecting DHFR-deficient CHO cells with a plasmid carrying a *loxP*-green fluorescent protein (GFP) fusion gene and a DHFR gene, cell colonies were screened by fluorescent intensity and two clones showed clear increased GFP expression upon gene amplification (160 mg/l in 7 days in a low-protein medium in a spinner-flask). An integration vector, carrying a *loxP*-fused hygromycin-resistance gene and the transgene expression cassette, was constructed to target pre-integrated chromosomal *loxP* sites by Cre recombinase-mediated site-specific recombination.

Using a different methodology to select highly productive cell clones – via the human secreted alkaline phosphatase (hSEAP) read-out system -, Celonic has stably integrated into the genome of a CHO-S strain a meganuclease-based retargetable platform (via a cGPS-like system licensed from Cellectis). The desired characteristics of a producer cell line include the ability to quickly and reproducibly generate stable clones producing high titers of recombinant proteins or monoclonal antibodies (mAbs) with controlled post-translational modifications. Typically, these cell lines are generated by random integration of a gene of interest – a process that does not control the level and/or stability of gene expression. This strategy is tedious, expensive, and time-consuming, but the real bottleneck is the work associated with producing, maintaining, and characterizing the numerous individual clones developed during this process. Cellectis bioresearch co-developed with Celonic a targeted integration system in CHO-S cells (cGPS® CHO-Sa CEMAX®), based on the stimulation of HR using the natural homing endonuclease I-*SceI* (http://www.cellectis-bioresearch.com/store/full-kits/cgps-cho-sa-cemax-full-kit). Briefly, a construct containing the cleavage site for the I-*SceI* meganuclease was introduced in the genome of a CHO-K1-derived CHO-S cell line along with non functional selection markers. Many clones were screened and the best selected based on two major criteria: the highest level of hSEAP expression and the presence of a single integrated copy of the construct. The so-called cGPS® CHO-S CEMAX® cell line is then co-transfected with an expression vector for the I-*SceI* meganuclease and a DNA matrix containing (i) homologous sequences flanking the I-*SceI* cleavage sites, (ii) genetic elements allowing the expression of the functional selection markers, and (iii) the expression cassette for the gene of interest. Following transfection, cells are subjected to a double selection process according to an optimized protocol. Further characterization revealed that the cGPS® CHO-Sa CEMAX® system is more rapid (2-week protocol), more efficient (all selected clones are targeted and expressed the transgene), reproducible (no or little transgene expression level variation), and stable over time (no change in transgene expression after 15 weeks in culture) than classical random integration. In a non-optimized small scale fed-batch process, titers close to 150 mg/L of Fc-fusion protein or close to 1–2 g/L for monoclonal antibodies were achieved.

This fast and robust method creates an opportunity for generating large collections of cell lines expressing high level of recombinant proteins and mAbs.

Various laboratories working in the bioproduction field are now strongly considering the real benefits of retargeting chromosomal "hot spot" region(s) that they have selected from genomic sequence analysis of their most highly producing cell clone(s). At Cellectis bioresearch, direct retargeting in "hot spot" regions can be achieved via cGPS-based platforms, whether they are stably incorporated by random integration (e.g. the currently commercialized cGPS® CHO-S CEMAX® system) or by targeted integration in pre-selected "hot spot" regions (current CRO service).

14.4.2 Disruption of Specific Genes

Both DHFR and GS genes have been knocked-out by Liu et al. (2009) in a CHO cell line, by exerting DNA nuclease activity of engineered ZFNs and downstream NHEJ-mediated mutagenesis (Liu et al. 2009). The authors have shown that the frequency of mono-allelic mutations varied from 15% to 25%, and more than 1% of bi-allelic mutation events were observed for both ZFN treatments.

Cellectis bioresearch using its meganuclease-based technology has succeeded in performing GS gene knock-out of a GMP compliant CHO producer cell line (publication in preparation). This qualified genome customization service was facilitated by the previous GS gene knock-out that Cellectis bioresearch performed in standard CHO-K1 cells (Fig. 14.4). Up to 7% of NHEJ events were obtained by transiently expressing a GS-related meganuclease and by deep sequence analysis of gene modification patterns in PCR amplicons, i.e. mutations, deletions or insertions (Fig. 14.4a). The isolation of GS knock-out CHO-K1 cell clones is presented in Fig. 14.4b, c. These clones were obtained after two rounds of transfection with the GS meganuclease-encoding plasmid, given the fact that at least four target sites are present in transfected cells (two functional alleles and two pseudogene alleles). Clone 4 F5 shows a deletion of 20 bp in one functional allele – (it was not possible to identify the type of mutagenic event regarding the second functional allele, probably because our PCR read-out involved a small PCR amplicon and therefore, cannot identify large deletion events Additionally, two large insertions were identified for the pseudogene alleles (136 bp and 614 bp). Regarding clone 3A6, two deletion events were identified on the functional alleles (22 bp and 3 bp) and two more deletions on the pseudogenes alleles (39 bp and 28 bp). Although the 3 bp deletion of in a functional gene could give rise to a functional protein, western blot analysis (Fig. 14.4b) indicates a significant reduction in the levels of GS protein expression as compared to wild type CHO-K1 cells. By contrast, the GS protein was not detectable for clone 4 F5. Current growth analysis studies have indicated that both clones are not capable of growth in the absence of L-glutamine (data not shown), proving the functional knock-out. Strategies designed to improve the problem of apoptosis in cell culture have focused on the knockdown of pro-apoptotic genes and/or the over-expression of anti-apoptotic genes. Recently, the ZFN cleavage of BAX and

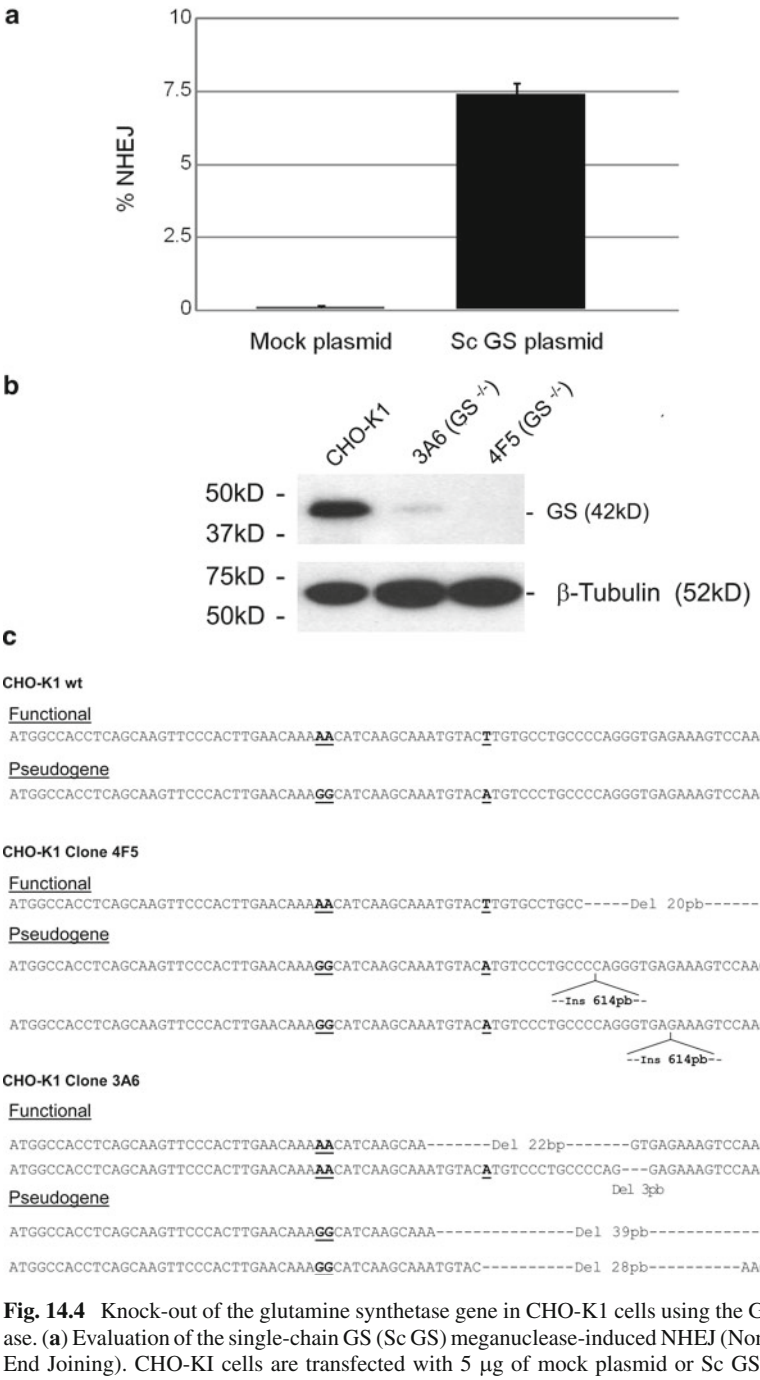


Fig. 14.4 Knock-out of the glutamine synthetase gene in CHO-K1 cells using the GS meganuclease. (**a**) Evaluation of the single-chain GS (Sc GS) meganuclease-induced NHEJ (Non Homologous End Joining). CHO-KI cells are transfected with 5 μg of mock plasmid or Sc GS plasmid. The percentage of NHEJ is measured by high throughput sequencing. (**b**) Western Blot. Total protein extracts from CHO-K1, GS KO clones 3A6 and 4 F5 are probed with anti GS and anti β-tubulin antibodies. (**c**) Sequences of the GS exon 2 (functional and pseudogene) from CHO-K1 and clones 4 F5 and 3A6. The GS meganuclease target site is underlined. Polymorphisms between the GS exon 2 functional and pseudogene are indicated. Regarding clone 4 F5, only one functional allele has been identified. Sizes of deletion (Del) or insertion (Ins) are indicated

BAK genes followed by inaccurate DNA repair has resulted in the knockout of both genes, with the generation of BAX- and BAK-deleted CHO cell clones capable of producing 2- to 5-fold more IgG than wild type CHO cells (Cost et al. 2010).

A FUT8 knockout CHO cell line has also been generated using the ZFN technology to produce afucosylated antibodies, where lower cell surface heparan sulfate levels were measured (Wong et al. 2010).

14.5 Cell Transformation

Various targeting systems have been demonstrated as useful methods to modify cell status, from primary cells to secondary cell lines (immortalization), from fibroblasts to induced pluripotent stem cells (iPS, reprogrammation) or from iPS to differentiated cells (differentiation).

14.5.1 Cell Immortalization

Cre-mediated reversible immortalization has been reported wherein immortalizing genes – i.e. encoding the human telomerase reverse transcriptase (hTERT) or the large T antigen from SV40 virus (SV40T) – flanked by *loxP* sites were successfully removed after cell line derivation by Cre recombinase expression.

This example was described by Kowolik et al. (2004) where primary human renal proximal tubule epithelial cells (RPTECs) were transduced with two lentiviral vectors each containing one of these "floxed" immortalizing genes (Kowolik et al. 2004). Transient Cre expression in such transduced cells led to efficient proviral deletion and decreased growth rates, therefore allowing reversible immortalization of human RPTECs and large-scale production of RPTECs that retain most tissue-specific properties.

In a similar fashion, human umbilical vein endothelial cells (HUVECs) and human liver sinusoidal endothelial cells (HLSECs) transduced with a retroviral vector containing a floxed SV40T gene yielded clones with greatly extended life spans after Cre recombinase expression (Noguchi et al. 2002).

14.5.2 Cell Reprogramming

Most of the studies describing the reprogramming of somatic cells to iPS cells are based upon the stable random integration of retroviral vector(s) encoding different reprogramming factors (i.e. Oct4, Klf4, Sox2, cMyc or Nanog) (Brambrink et al. 2008; Nakagawa et al. 2008; Sommer et al. 2009). A recent publication has indicated that gene activation or disruption via proviral integration sites may play a role in obtaining

the pluripotent phenotype (Winkler et al. 2010). An alternative random integration approach has also been given by using transposon elements (Yusa et al. 2009; Woltjen et al. 2009). Although successful reprogramming has been obtained by these gene transfer systems, such random-based integration events often accompanied with unwanted transcriptional silencing suffer of many drawbacks, such as the lack of robustness and reproducibility or the difficulty to specifically control the integration and expression of transgene cassettes.

The best gene transfer approach will allow standardization of iPS protocols by targeting a specific genomic location for which the challenges of both integration and sustained expression are addressed. Several laboratories are currently developing processes which utilize meganucleases (Ectycell), ZFNs (Sigma, Sangamo), PhiC31 and R4 recombinases (Invitrogen) for iPS cell reprogramming. The first published study to date referred to PhiC31 targeted integration in the genome of two separate somatic cell models, i.e. mouse embryonic fibroblasts and human amniotic fluid cells (Ye et al. 2010). Although this demonstration is the first evidence that iPS cell reprogramming is possible by targeted integration, the fact that PhiC31 integration sites are fixed and present in multiple copies could limit the opportunity to standardize this strategy.

14.5.3 Cell Differentiation

As for cell reprogramming, the current available procedures that have been established for converting iPS cells to differentiated cell types remain challenging due primarily to the poor cell homogeneity that can be currently achieved. Site-directed gene knock-in approaches to integrate at endogenous or safe harbor chromosomal loci could be used to circumvent these issues and facilitate future developments to standardize cell differentiation processes.

As first proof of concepts, engineered meganucleases and ZFNs are now being used for gene targeting within the genome of iPS cells. Using ZFN-mediated genome editing, Hockemeyer et al. (2009) have reported efficient transgene targeting at three separate chromosomal regions (Oct4 and Pitx3 genes, AAVS1 locus) of human iPS cells (Hockemeyer et al. 2009).

14.6 Molecular Biology and Microbiology Tools

14.6.1 Plasmid/Cosmid Cloning

Molecular cloning is fundamental in many areas of basic research in biology, drug screening, in the production of recombinant proteins, and even in medicine. One essential key lies in the ability of restriction enzymes (or restriction endonucleases) to cut double-stranded DNA at a very specific recognition nucleotide sequences

known as restriction sites. Subsequently, this fragment can be combined with vector DNA to generate a recombinant DNA molecule that could be transferred to cells or animal using various methods.

Like restriction enzymes used in molecular biology experiments (e.g. *Not*I), the capacity of meganucleases to accurately cleave double-stranded DNA has been exploited to generate restriction fragments for cloning purposes. Compared to standard restriction enzymes that usually target 4–8 bp sequences, the nucleic sequence recognized by meganucleases are very large (about 12–40 bp) and non-palindromic. Thus, the probability of finding meganuclease targets in artificial or native DNA sequences (e.g. plasmid) is virtually zero. This property makes meganucleases a powerful tool for many applications in molecular biology, like basic molecular cloning. Moreover, several homing endonucleases, such as I-*SceI* or I-*CeuI*, have been commercially available for several years.

In addition to the target specificity of natural homing endonucleases, Cellectis has developed numerous engineered meganucleases, each of them recognizing a specific nucleotide sequence (Arnould et al. 2006; Smith et al. 2006; Fajardo-Sanchez et al. 2008). This opens new perspectives for molecular biologists and recombinant DNA applications, because there is almost no limitation to the possible choice of target sequences. By contrast, the number of distinct sequences cleavable by standard commercially available restriction enzymes is limited to around 260 possibilities (Roberts et al. 2007). Another clear advantage of meganucleases given by their intrinsic biochemical properties is that they are highly stable. Figure 14.5 shows that *in vitro* DNA cleavage activity is not affected when I-*CreI* recombinant protein batches are stored at high temperatures (at least 3 weeks at 37°C) or maintained over a long time period (at least 54 months at −80°C).

Similarly, ZFNs have the potential to be used as artificial restriction enzymes. Indeed, Zeevi et al. (2008) demonstrated the potential of ZFNs for *in vitro* DNA cloning (Zeevi et al. 2008). They showed that ZFNs in combination with commercially available restriction enzymes (e.g. *Fok*I) can be used like basic restriction enzymes with designed specificity for DNA sequences. More recently, Tovkach et al. (2011) have demonstrated that the quality of ZFNs purified by nickel affinity could be sufficient for recombinant DNA applications such as cloning, and that the efficiency of ZFNs could be compared to that of PI-*PspI*, a commercially available meganuclease (Tovkach et al. 2011).

Another alternative to classic restriction enzymes for cloning is the Cre-*loxP* system. For example, Buchholz and Bishop (2001) have developed a way to use this system for an *in vitro* DNA cloning system, that they named "*loxP*-directed cloning" (Buchholz and Bishop 2001). Based on the results of three experiments, including the manipulation of BACs, the authors suggest that the 34 bp *loxP* site can serve as a universal target site for cloning a DNA restriction fragment. According to these authors, this technique could advantageously replace the need for compatible restriction sites between insert and cloning vectors by generating compatible 5′, 3′, or blunt ends in the desired orientation and in the reading frame.

In addition to *in vitro* cloning of BACs, both Cre-*loxP* and Flp-FRT recombinase systems have also largely been used as *in cellulo* targeting platforms for the generation

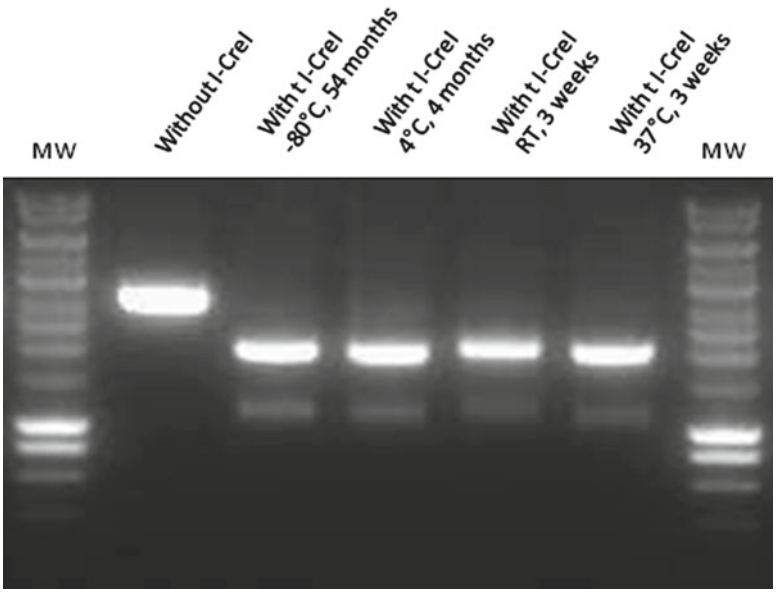


Fig. 14.5 *In vitro* DNA cleavage activity of a recombinant I-*CreI* protein stored at various temperatures and duration times. A recombinant I-*CreI* protein produced in E. *coli* and purified via a 6His-tag fused to the N-terminal end and stored at (i) −80°C for 54 months, (ii) 4°C for 4 months, (iii) room temperature (RT) and (iv) 37°C for 3 weeks. One μg of a DNA fragment containing the I-*CreI* target site was incubated at 37°C for 2 h with XXng of each I-CreI protein

of recombinant viruses or vectors, such as baculoviruses (Fitzgerald et al. 2006), adenoviruses (Nakano et al. 2005; Hillgenberg et al. 2006) and AAV (Mizukami et al. 2004).

14.6.2 Virus/Vector Production

The Cre-*loxP* excision system has also been used for the removal of viral sequences involved in virus and/or vector production schemes. By inserting *loxP*-flanked BAC cassettes into specific regions of herpes virus genomes several laboratories have described stable and infectious BAC virus clones that can be used to produce recombinant strains, i.e. human herpes simplex virus type 1 (HSV-1) (Tanaka et al. 2003), bovine herpes virus 4 (BoHV-4) (Gillet et al. 2005), wildebeest alcelaphine herpesvirus 1 (AlHV-1) (Dewals et al. 2006), equine herpesvirus 4 (EHV-4) (Azab et al. 2009), canine herpesvirus (CHV) (Arii et al. 2006), koi herpesvirus (KHV) (Costes et al. 2008). This insertion led to BAC plasmids that were stably maintained in bacteria and able to regenerate virions when permissive cells were transfected with each dedicated

plasmid. Reconstituted virions free of BAC cassette have been produced by the transfection of Cre recombinase-expressing cells with the BAC.

The Cre-*loxP*, and to a lesser extent the Flp-FRT, excision approach has also been applied to the removal of *cis*-active packaging signals of helper viruses used for viral vector production, i.e. HSV-1 amplicons (Logvinoff and Epstein 2001; Zaupa et al. 2003) and adenoviral vectors (Ng et al. 2001; Zhou et al. 2002).

14.6.3 Anti-microbial Applications

Very recently, both meganucleases and ZFNs have been used to directly target and mutate DNA-based viral genomes, an activity referred to as "virus clipping". Cellectis Therapeutics has demonstrated inhibition of infection by the herpes simplex virus 1 (HSV1) in cultured cells (up to 70%), at low and moderate multiplicities of infection. This activity was corroborated by a significant reduction of the viral load and a high rate of mutation (up to 16%) at the meganuclease cleavage site, consistent with a mechanism of action based on the cleavage of the viral genome (Grosse et al. 2011). Similarly, by using an indirect validation approach, Cradick et al. (2010) have shown that an engineered ZFN can result in specific cleavage of a target plasmid that contains the hepatitis B virus (HBV) genome (Cradick et al. 2010). After 3 days in culture, 10% of the target was cleaved and misjoined. Moreover, ZFN treatment decreased levels of the viral pregenomic RNA by 29%.

By taking advantage of the potency of AAV vectors to mediate HR, Carson et al. (2005) have developed a DNA recombination-based approach to prevent bovine papillomavirus (BPV) replication in infected cells through the conditional expression of a suicide marker gene, i.e. HSV1 thymidine kinase (TK) (Carson et al. 2005). In this anti-viral model, the transgene cassette was designed to express TK, through HR, only in cells that contain viral DNA. The pro-drug (ganciclovir, GCV) treatment of these cells demonstrated the elimination of over 95% of BPV-infected cells.

Various demonstrations of TFO-mediated mutagenesis have been described by targeting several different types of DNA viral genome intermediates, such as HBV (Guang et al. 2003) or human immunodeficiency virus (HIV) (Teulade-Fichou et al. 2001; Capobianco et al. 2005).

Indirect anti-microbial approaches have targeted the genome of host vectors. Homing endonucleases such as I-*PpoI*, I-*SceI*, I-*CreI* and I-*AniI* have been shown to induce dsDNA breaks in adult female *Aedes aegypti* chromosomes and to catalyze the somatic excision of a transgene (Traver et al. 2009). Similarly, a synthetic genetic element, consisting of mosquito regulatory regions and the homing endonuclease gene I-*SceI*, substantially increased its transmission to the progeny in transgenic mosquitoes of the human malaria vector *Anopheles gambiae*, and induced high rates of site-specific chromosomal cleavage and gene conversion (Windbichler et al. 2011).

14.7 Conclusion

Although Cre-*loxP* and FLP-FRT recombinase systems have widely been used as retargeting platforms, they cannot be applied for targeting specific endogenous chromosomal sites. It is the same situation for transposon elements where they are randomly transposed in the genome of mammalian cells, or have restricted target specificities. Multiple and fixed chromosomal sequences are also targeted by using serine recombinases (PhiC31, R4, TP901-1), therefore limiting their use to knock-in applications only.

The potency of AAV vectors to mediate HR in mammalian cells has allowed the generation of different *in cellulo* and *in vivo* models, whether genes have been repaired or mutated. The limited packaging capacity of AAV vectors (5 kb) is, however, an impediment for AAV-mediated gene transfer of larger genomes. This size limitation is, of course, much more relevant when using TFOs which are restricted to only a few nucleotides.

Engineered DNA nucleases, such as meganucleases and ZFNs, clearly offer many advantages, with many kinds of genome modifications already demonstrated. The emerging TALEN technology is receiving very strong interest because of their very high modularity for protein engineering.

References

Amaya E, Kroll K (2010) J Vis Exp. doi:10.3791/2010
Arii J, Hushur O, Kato K, Kawaguchi Y, Tohya Y, Akashi H (2006) Microbes Infect 8:1054–63
Arnould S, Chames P, Perez C, Lacroix E, Duclert A, Epinat JC, Stricher F, Petit AS, Patin A, Guillier S, Rolland S, Prieto J, Blanco FJ, Bravo J, Montoya G, Serrano L, Duchateau P, Paques F (2006) J Mol Biol 355:443–58
Arnould S, Perez C, Cabaniols JP, Smith J, Gouble A, Grizot S, Epinat JC, Duclert A, Duchateau P, Paques F (2007) J Mol Biol 371:49–65
Azab W, Kato K, Arii J, Tsujimura K, Yamane D, Tohya Y, Matsumura T, Akashi H (2009) Arch Virol 154:833–42
Babaryka A, Kuhn E, Koster RW (2009) J Fish Biol 74:452–7
Belteki G, Gertsenstein M, Ow DW, Nagy A (2003) Nat Biotechnol 21:321–4
Bennardo N, Stark JM (2010) PLoS Genet 6:e1001194
Bernstein M, Lersch RA, Subrahmanyan L, Cline TW (1995) Genetics 139:631–48
Bertoni C, Jarrahian S, Wheeler TM, Li Y, Olivares EC, Calos MP, Rando TA (2006) Proc Natl Acad Sci U S A 103:419–24
Bigey P, Pratviel G, Meunier B (1995) Nucleic Acids Res 23:3894–900
Birling MC, Gofflot F, Warot X (2009) Methods Mol Biol 561:245–63
Bockamp E, Sprengel R, Eshkind L, Lehmann T, Braun JM, Emmrich F, Hengstler JG (2008) Regen Med 3:217–35
Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, Jaenisch R (2008) Cell Stem Cell 2:151–9
Buchholz F, Bishop M (2001) Biotechniques 31: 906–908, 910, 912, 914, 916, 918
Cabaniols JP, Paques F (2008) Methods Mol Biol 435:31–45
Cabaniols JP, Mathis L, Delenda C (2009) Curr Opin Pharmacol 9:657–63

Cabaniols JP, Ouvry C, Lamamy V, Fery I, Craplet ML, Moulharat N, Guenin SP, Bedut S, Nosjean O, Ferry G, Devavry S, Jacqmarcq C, Lebuhotel C, Mathis L, Delenda C, Boutin JA, Duchateau P, Coge F, Paques F (2010) J Biomol Screen 15:956–67
Capobianco ML, De Champdore M, Arcamone F, Garbesi A, Guianvarc'h D, Arimondo PB (2005) Bioorg Med Chem 13:3209–18
Carson A, Wang Z, Xiao X, Khan SA (2005) Gene Ther 12:534–40
Castaneda F, Kinne RK (2005) Mol Cell Biochem 280:91–8
Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Nucleic Acids Res 39:e82
Chan YS, Naujoks DA, Huen DS, Russell S (2011) Genetics 188(1):33–44
Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Genetics 186:757–61
Claeys Bouuaert C, Chalmers R (2010) Nucleic Acids Res 38:190–202
Clark KJ, Carlson DF, Foster LK, Kong BW, Foster DN, Fahrenkrug SC (2007) BMC Biotechnol 7:42
Colosimo A, Goncz KK, Novelli G, Dallapiccola B, Gruenert DC (2001) Mol Ther 3:178–85
Cost GJ, Freyvert Y, Vafiadis A, Santiago Y, Miller JC, Rebar E, Collingwood TN, Snowden A, Gregory PD (2010) Biotechnol Bioeng 105:330–40
Costes B, Fournier G, Michel B, Delforge C, Raj VS, Dewals B, Gillet L, Drion P, Body A, Schynts F, Lieffrig F, Vanderplasschen A (2008) J Virol 82:4955–64
Cradick TJ, Keck K, Bradshaw S, Jamieson AC, McCaffrey AP (2010) Mol Ther 18:947–54
Crenes G, Moundras C, Demattei MV, Bigot Y, Petit A, Renault S (2010) Genetica 138:509–17
DeKelver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijsing SH, Sancak Y, Cui X, Steine EJ, Miller JC, Tam P, Bartsevich VV, Meng X, Rupniewski I, Gopalan SM, Sun HC, Pitz KJ, Rock JM, Zhang L, Davis GD, Rebar EJ, Cheeseman IM, Yamamoto KR, Sabatini DM, Jaenisch R, Gregory PD, Urnov FD (2010) Genome Res 20:1133–42
Dewals B, Boudry C, Gillet L, Markine-Goriaynoff N, de Leval L, Haig DM, Vanderplasschen A (2006) J Gen Virol 87:509–17
Duca M, Guianvarc'h D, Oussedik K, Halby L, Garbesi A, Dauzonne D, Monneret C, Osheroff N, Giovannangeli C, Arimondo PB (2006) Nucleic Acids Res 34:1900–11
Duca M, Vekhoff P, Oussedik K, Halby L, Arimondo PB (2008) Nucleic Acids Res 36:5123–38
Durai S, Mani M, Kandavelou K, Wu J, Porteus MH, Chandrasegaran S (2005) Nucleic Acids Res 33:5978–90
Eisenberg E, Levanon EY (2003) Trends Genet 19:362–5
Fajardo-Sanchez E, Stricher F, Paques F, Isalan M, Serrano L (2008) Nucleic Acids Res 36:2163–73
Fischer SE, Wienholds E, Plasterk RH (2001) Proc Natl Acad Sci U S A 98:6759–64
Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I (2006) Nat Methods 3:1021–32
Forster A, Pannell R, Drynan L, Cano F, Chan N, Codrington R, Daser A, Lobato N, Metzler M, Nam CH, Rodriguez S, Tanaka T, Rabbitts T (2005) Cold Spring Harb Symp Quant Biol 70:275–82
Francois JC, Saison-Behmoaras T, Barbier C, Chassignol M, Thuong NT, Helene C (1989) Proc Natl Acad Sci U S A 86:9702–6
Gaddis SS, Wu Q, Thames HD, DiGiovanni J, Walborg EF, MacLeod MC, Vasquez KM (2006) Oligonucleotides 16:196–201
Galvan DL, Nakazawa Y, Kaja A, Kettlun C, Cooper LJ, Rooney CM, Wilson MH (2009) J Immunother 32:837–44
Gillet L, Daix V, Donofrio G, Wagner M, Koszinowski UH, China B, Ackermann M, Markine-Goriaynoff N, Vanderplasschen A (2005) J Gen Virol 86:907–17
Glover L, Alsford S, Beattie C, Horn D (2007) Nucleic Acids Res 35:872–80
Grabher C, Joly JS, Wittbrodt J (2004) Methods Cell Biol 77:381–401
Grabundzija I, Izsvak Z, Ivics Z (2011) Methods Mol Biol 738:69–85
Grosse S, Huot N, Mahiet C, Arnould S, Barradeau S, Clerre DL, Chion-Sotinel I, Jacqmarcq C, Chapellier B, Ergani A, Desseaux C, Cedrone F, Conseiller E, Paques F, Labetoulle M, Smith J (2011) Mol Ther 19:694–702
Guang L, Yuan F, Xi M, Zhao C, Liu L, Wen E, Ai Y (2003) Chin Med J (Engl) 116:1248–52

Hamer L, DeZwaan TM, Montenegro-Chamorro MV, Frank SA, Hamer JE (2001) Curr Opin Chem Biol 5:67–73
Hansen GM, Markesich DC, Burnett MB, Zhu Q, Dionne KM, Richter LJ, Finnell RH, Sands AT, Zambrowicz BP, Abuin A (2008) Genome Res 18:1670–9
Hillgenberg M, Hofmann C, Stadler H, Loser P (2006) J Virol 80:5435–50
Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng X, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2009) Nat Biotechnol 27:851–7
Huth M, Schlake T, Boehm T (1997) Immunogenetics 45:282–3
Ishibashi S, Kroll KL, Amaya E (2007) CSH Protoc: pdb.prot4838. doi:10.1101/pdb.prot4840
Ivics Z, Li MA, Mates L, Boeke JD, Nagy A, Bradley A, Izsvak Z (2009) Nat Methods 6:415–22
Kandavelou K, Ramalingam S, London V, Mani M, Wu J, Alexeev V, Civin CI, Chandrasegaran S (2009) Biochem Biophys Res Commun 388:56–61
Kapitskaya M, Cunningham ME, Lacson R, Kornienko O, Bednar B, Petrukhin K (2006) Assay Drug Dev Technol 4:253–62
Kashiwagi K, Kashiwagi A, Kurabayashi A, Hanada H, Nakajima K, Okada M, Takase M, Yaoita Y (2010) Exp Anim 59:395–405
Kean JM, Miller PS (1993) Bioconjug Chem 4:184–7
Kim AI, Ghosh P, Aaron MA, Bibb LA, Jain S, Hatfull GF (2003) Mol Microbiol 50:463–73
Kito M, Itami S, Fukano Y, Yamana K, Shibui T (2002) Appl Microbiol Biotechnol 60:442–8
Kowolik CM, Liang S, Yu Y, Yee JK (2004) Oncogene 23:5950–7
Kuhlman TE, Cox EC (2010) Bioeng Bugs 1:296–299
Levy A, Schwartz S, Ast G (2010) Nucleic Acids Res 38:1515–30
Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B (2011) Nucleic Acids Res 39(14):6315–25
Lieu PT, Machleidt T, Thyagarajan B, Fontes A, Frey E, Fuerstenau-Sharp M, Thompson DV, Swamilingiah GM, Derebail SS, Piper D, Chesnut JD (2009) J Biomol Screen 14:1207–15
Liu G, Geurts AM, Yae K, Srinivasan AR, Fahrenkrug SC, Largaespada DA, Takeda J, Horie K, Olson WK, Hackett PB (2005) J Mol Biol 346:161–73
Liu K, Hipkens S, Yang T, Abraham R, Zhang W, Chopra N, Knollmann B, Magnuson MA, Roden DM (2006a) Genesis 44:556–64
Liu W, Xiong Y, Gossen M (2006b) J Mol Med 84:57–64
Liu J, Skjorringe T, Gjetting T, Jensen TG (2009) BMC Biotechnol 9:31
Logvinoff C, Epstein AL (2001) Hum Gene Ther 12:161–7
Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L (2007) Nat Biotechnol 25:1298–306
Luo G, Ivics Z, Izsvak Z, Bradley A (1998) Proc Natl Acad Sci U S A 95:10769–73
Mates L, Izsvak Z, Ivics Z (2007) Genome Biol 8(Suppl 1):S1
Miller RL (2011) Am J Physiol Renal Physiol 300:F291–300
Miller DG, Petek LM, Russell DW (2003) Mol Cell Biol 23:3550–7
Mizukami H, Okada T, Ogasawara Y, Matsushita T, Urabe M, Kume A, Ozawa K (2004) Mol Biotechnol 27:7–14
Munoz-Lopez M, Siddique A, Bischerour J, Lorite P, Chalmers R, Palomeque T (2008) J Mol Biol 382:567–72
Nagatsugi F, Suenaga Y, Sasaki S (2003) Nucleic Acids Res Suppl 3:155–6
Nair RR, Avila H, Ma X, Wang Z, Lennartz M, Darnay BG, Boyd DD, Yan C (2008) Mol Pharmacol 73:919–29
Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Nat Biotechnol 26:101–6
Nakano M, Odaka K, Takahashi Y, Ishimura M, Saito I, Kanegae Y (2005) Nucleic Acids Res 33:e76
Ng P, Beauchamp C, Evelegh C, Parks R, Graham FL (2001) Mol Ther 3:809–15
Noguchi H, Kobayashi N, Westerman KA, Sakaguchi M, Okitsu T, Totsugawa T, Watanabe T, Matsumura T, Fujiwara T, Ueda T, Miyazaki M, Tanaka N, Leboulch P (2002) Hum Gene Ther 13:321–34

Ochiai H, Sakamoto N, Suzuki K, Akasaka K, Yamamoto T (2008) Dev Dyn 237:2475–82
Olivares EC, Hollis RP, Calos MP (2001) Gene 278:167–76
Olson LE, Tien J, South S, Reeves RH (2005) Transgenic Res 14:325–32
Oumard A, Qiao J, Jostock T, Li J, Bode J (2006) Cytotechnology 50:93–108
Pan FC, Chen Y, Loeber J, Henningfeld K, Pieler T (2006) Dev Dyn 235:247–52
Perrouault L, Asseline U, Rivalle C, Thuong NT, Bisagni E, Giovannangeli C, Le Doan T, Helene C (1990) Nature 344:358–60
Postel EH, Flint SJ, Kessler DJ, Hogan ME (1991) Proc Natl Acad Sci U S A 88:8227–31
Puri N, Majumdar A, Cuenoud B, Natt F, Martin P, Boyd A, Miller PS, Seidman MM (2001) J Biol Chem 276:28991–8
Quenneville SP, Chapdelaine P, Rousseau J, Beaulieu J, Caron NJ, Skuk D, Mills P, Olivares EC, Calos MP, Tremblay JP (2004) Mol Ther 10:679–87
Rembold M, Lahiri K, Foulkes NS, Wittbrodt J (2006) Nat Protoc 1:1133–9
Roberts RJ, Vincze T, Posfai J, Macelis D (2007) Nucleic Acids Res 35:D269–70
Roebroek AJ, Gordts PL, Reekmans S (2011) Methods Mol Biol 693:277–81
Russell JP, Chang DW, Tretiakova A, Padidam M (2006) Biotechniques 40: 460, 462, 464
Sandhu U, Cebula M, Behme S, Riemer P, Wodarczyk C, Metzger D, Reimann J, Schirmbeck R, Hauser H, Wirth D (2011) Nucleic Acids Res 39:e1
Sauer B, Henderson N (1989) Nucleic Acids Res 17:147–61
Sauka-Spengler T, Barembaum M (2008) Methods Cell Biol 87:237–56
Seidman MM, Puri N, Majumdar A, Cuenoud B, Miller PS, Alam R (2005) Ann N Y Acad Sci 1058:119–27
Shinkawa T, Nakamura K, Yamane N, Shoji-Hosaka E, Kanda Y, Sakurada M, Uchida K, Anazawa H, Satoh M, Yamasaki M, Hanai N, Shitara K (2003) J Biol Chem 278:3466–73
Sinzelle L, Kapitonov VV, Grzela DP, Jursch T, Jurka J, Izsvak Z, Ivics Z (2008) Proc Natl Acad Sci U S A 105:4715–20
Smith J, Grizot S, Arnould S, Duclert A, Epinat JC, Chames P, Prieto J, Redondo P, Blanco FJ, Bravo J, Montoya G, Paques F, Duchateau P (2006) Nucleic Acids Res 34:e149
Sollu C, Pars K, Cornu TI, Thibodeau-Beganny S, Maeder ML, Joung JK, Heilbronn R, Cathomen T (2010) Nucleic Acids Res 38:8269–76
Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G (2009) Stem Cells 27:543–9
Staunstrup NH, Sharma N, Bak RO, Svensson L, Petersen TK, Aarenstrup L, Kristiansen K, Bolund L, Mikkelsen JG (2011) BMC Biotechnol 11:33
Stoll SM, Ginsburg DS, Calos MP (2002) J Bacteriol 184:3657–63
Strobel SA, Doucette-Stamm LA, Riba L, Housman DE, Dervan PB (1991) Science 254:1639–42
Tanaka M, Kagawa H, Yamanashi Y, Sata T, Kawaguchi Y (2003) J Virol 77:1382–91
Teulade-Fichou MP, Perrin D, Boutorine A, Polverari D, Vigneron JP, Lehn JM, Sun JS, Garestier T, Helene C (2001) J Am Chem Soc 123:9283–92
Thermes V, Grabher C, Ristoratore F, Bourrat F, Choulika A, Wittbrodt J, Joly JS (2002) Mech Dev 118:91–8
Tovkach A, Zeevi V, Tzfira T (2011) J Biotechnol 151:1–8
Traver BE, Anderson MA, Adelman ZN (2009) Insect Mol Biol 18:623–33
Turan S, Kuehle J, Schambach A, Baum C, Bode J (2010) J Mol Biol 402:52–69
Turan S, Galla M, Ernst E, Qiao J, Voelkel C, Schiedlmeier B, Zehe C, Bode J (2011) J Mol Biol 407:193–221
Umlauf SW, Cox MM (1988) EMBO J 7:1845–52
Vasileva A, Jessberger R (2005) Nat Rev Microbiol 3:837–47
Wang X (2009) Methods Mol Biol 561:265–73
Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat RJ Jr, Burt A, Crisanti A (2011) Nature 473:212–5
Winkler T, Cantilena A, Metais JY, Xu X, Nguyen AD, Borate B, Antosiewicz-Bourget JE, Wolfsberg TG, Thomson JA, Dunbar CE (2010) Stem Cells 28:687–94

Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hamalainen R, Cowling R, Wang W, Liu P, Gertsenstein M, Kaji K, Sung HK, Nagy A (2009) Nature 458:766–70
Wong AW, Baginski TK, Reilly DE (2010) Biotechnol Bioeng 106:751–63
Wu SC, Maragathavally KJ, Coates CJ, Kaminski JM (2008) Methods Mol Biol 435:139–51
Yan Z, Lei-Butters D, Engelhardt JF, Leno GH (2009) BMC Biol 7:8
Ye L, Chang JC, Lin C, Qi Z, Yu J, Kan YW (2010) Proc Natl Acad Sci U S A 107:19467–72
Yusa K, Rad R, Takeda J, Bradley A (2009) Nat Methods 6:363–9
Zaupa C, Revol-Guyot V, Epstein AL (2003) Hum Gene Ther 14:1049–63
Zeevi V, Tovkach A, Tzfira T (2008) Proc Natl Acad Sci U S A 105:12785–90
Zeng X, Rao MS (2008) Curr Opin Mol Ther 10:207–13
Zheng B, Mills AA, Bradley A (2001) Methods 24:81–94
Zhou HS, Zhao T, Rao XM, Beaudet AL (2002) J Gene Med 4:498–509
Ziemba A, Derosier LC, Methvin R, Song CY, Clary E, Kahn W, Milesi D, Gorn V, Reed M, Ebbinghaus S (2001) Nucleic Acids Res 29:4257–63

Index

S. Renault and P. Duchateau (eds.), *Site-directed Insertion of Transgenes*, Topics in Current Genetics 23, DOI 10.1007/978-94-007-4531-5,