Takashi Yamamoto *Editor*

Targeted Genome Editing Using Site-Specific Nucleases

ZFNs, TALENs, and the CRISPR/Cas9 System

Springer

Takashi Yamamoto
Editor

Targeted Genome Editing Using Site-Specific Nucleases

ZFNs, TALENs, and the CRISPR/Cas9 System

Springer

Editor
Takashi Yamamoto
Department of Mathematical and Life Sciences
Hiroshima University
Higashi-Hiroshima, Hiroshima, Japan

ISBN 978-4-431-55226-0 ISBN 978-4-431-55227-7 (eBook)
DOI 10.1007/978-4-431-55227-7
Springer Tokyo Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014954587

Printed on acid-free paper

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

Targeted Genome Editing Using Site-Specific Nucleases

Preface

Genome editing using targetable nucleases is an emerging technology for the precise genome modification of organisms ranging from bacteria to plants to animals. Its attraction is that it can be used for almost all organisms in which targeted genome modification has not been possible. To date, various modifications have been reported in cultured cells and many organisms, including gene knockout and knock-in experiments with targeting vectors, and chromosomal editing. Indeed, we are constantly surprised at the rapid progress and possibilities suggested by this technology. The development of TALENs in 2010 has enabled the genome modification of non-model organisms, while the emergence of the CRISPR/Cas9 system in 2013 has allowed us to anticipate the forthcoming new era of genome editing research.

Compared with other biotechnologies, the development of genome editing has been unusually rapid. Therefore, some researchers are unaware of its existence, while many do not fully understand the basics or how to apply the technology to their particular organism. The purpose of this book is therefore twofold. The first part sets out to introduce the history of the development of genome editing tools such as ZFN, TALEN, and CRISPR/Cas9 to beginners. The second part discusses the applications of genome editing to different organisms. However, because it is not certain how long the various methods introduced in this book will be in use, and because the applications of genome editing will undoubtedly progress, it is recommended that researchers continuously update their knowledge based on the latest genome editing technology.

In the near future, we anticipate that more precise and complex modifications such as single nucleotide substitution at multiple loci, and chromosomal editing will be possible in cultured cells and organisms. Moreover, the progress of such a convenient and safe method will undoubtedly have a high utility value in medical research. Soon, it is likely that the development of therapeutics by genome editing may no longer be just a dream. Genome editing is also hoped to be available for use in the generation of crops and livestock with useful traits. Considering these possibilities, it appears certain that genome editing will become the next generation of biotechnology.

Although the availability of genome editing is dependent on species, it is neither a difficult nor an expensive technology. Thus, I hope that young scientists will adopt this technology in their own research, and I am certain that this book will be helpful for such a purpose. Finally, I thank all authors for their contributions towards establishing this book.

Higashi-Hiroshima, Japan Takashi Yamamoto

CPSIA information can be obtained at www.ICGtesting.com
Printed in the USA
LVOW02*1847080115
422030LV00001B/9/P

Contents

Part I
Basics of Genome Editing

Chapter 1
Genome Editing Using Zinc-Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs)

Hiroshi Ochiai and Takashi Yamamoto

Abstract Targetable nucleases, including zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas), induce DNA double-strand breaks (DSBs) into user-defined sites. DSBs are immediately repaired through the evolutionarily conserved pathways of error-prone non-homologous end joining (NHEJ) or homology-directed repair (HDR). With the utilization of these repair processes, researchers have been able to disrupt specific genes, add exogenous DNA elements into intended genomic sites, introduce single-nucleotide substitutions, and perform many other applications. Consequently, this "genome editing" technology has revolutionized the life science field. In addition, this technology has the potential to improve agricultural products and be applicable to therapeutic use.

Here, we will introduce a brief history of targetable nuclease-mediated genome editing and the applications of the tools that the technology provides. In this chapter, we will primarily focus on ZFNs and TALENs, which are artificial proteins composed of a specific DNA-binding domain and a restriction enzyme FokI DNA-cleavage domain. We will also review the properties and construction methods of these nucleases.

Keywords DSB • Gene targeting • Genome editing • TALEN • ZFN

H. Ochiai (✉)
Research Center for the Mathematics on Chromatin Live Dynamics,
Hiroshima University, Higashi-Hiroshima 739-8530, Japan
e-mail: ochiai@hiroshima-u.ac.jp

T. Yamamoto
Department of Mathematical and Life Sciences, Graduate School of Science,
Hiroshima University, Higashi-Hiroshima 739-8526, Japan
e-mail: tybig@hiroshima-u.ac.jp

T. Yamamoto (ed.), *Targeted Genome Editing Using Site-Specific Nucleases*,
DOI 10.1007/978-4-431-55227-7_1

1.1 Introduction

With the progress of DNA sequencing technology, the genomic sequences of many organisms have now been determined. Most researchers consider the blueprint of life to be written in the genomic DNA and that the elucidation of its code, especially the functions of genes, can directly lead to an understanding of living things. To explore the functions of genes, a large number of approaches have been developed to date. Among them, homologous recombination-based gene modification is the best way to precisely analyze gene functions, but the technique is only available in restricted organisms, such as yeasts and mice. The availability of this technique strongly drives detailed studies of genomic functions in yeasts and mice, compared with other organisms. However, individual model organisms each have certain strong points for studying specific research areas. For example, compared with mice, rats are the preferred animal for disease models because of their size and simplicity of manipulation (Aitman et al. 2008), while animals with small transparent embryos, such as sea urchins, ascidians, worms, fruit flies and zebrafish, are suitable for in vivo observations of morphogenesis at single-cell resolution (Megason and Fraser 2007). Thus, to understand living things, broad knowledge from researches on a wide range of organisms is required. To overcome the unavailability of homologous recombination-based genetic modification in many organisms, two approaches are performed. The first is random integration of exogenous DNA elements that express short hairpin RNAs (shRNAs) to inhibit the production of specific protein species or overexpress some proteins to gain insights about the protein function, and the second involves treatment with chemical compounds or short inhibitory RNAs (siRNAs) to inhibit the production or function of specific protein species. However, transgenes randomly integrated into genomic DNA are sometimes not stably expressed for a long time and may disrupt the proper expression or function of endogenous genes by their integration (Emery 2011). Conversely, in most of cases, the inhibition of protein function using chemical compounds or siRNAs is transient and imperfect (Subramanya et al. 2010). Other methods, such as random mutagenesis using chemical mutagens and transposons, so-called forward genetics, are also effective for gaining insights about gene functions (Ivics et al. 2009; Yates et al. 2009). However, these approaches are extremely labor-intensive, are time-consuming to obtain mutants of interest, and involve undetectable genetic changes other than specific mutations of interest.

The emergence of genome editing technology using targetable nucleases, such as zinc-finger nucleases (ZFNs) (Urnov et al. 2010), transcription activator-like effector nucleases (TALENs) (Joung and Sander 2013), and clustered regularly interspaced short palindromic repeat (CRISPR)/ CRISPR-associated (Cas) (Mali et al. 2013a) (see Chap. 2), has changed many aspects of the life sciences field. This technology employs two key players: targetable nucleases that introduce DNA double-strand breaks (DSBs) into user-defined sites and the DSB repair mechanisms in cells (Fig. 1.1). DSBs are mainly repaired through two pathways, non-homologous end joining (NHEJ) that is inherently error-prone and homology-directed repair (HDR)

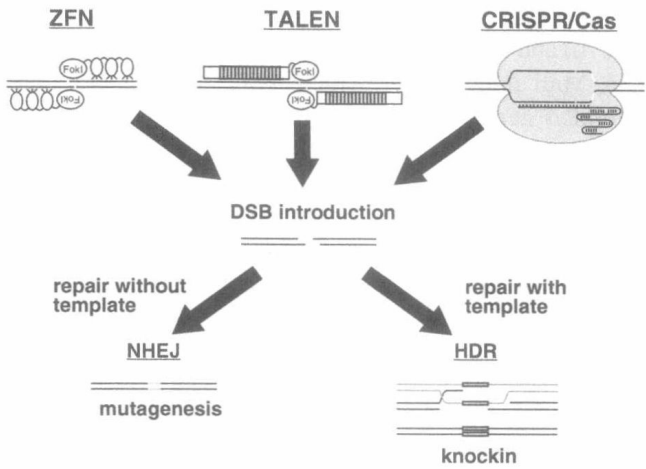

Fig. 1.1 Targetable nuclease-mediated genome editing. Targetable nucleases, such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeat (CRISPR)/ CRISPR-associated (Cas) introduce DNA double-strand breaks (DSBs) at user-intended sites. DSBs are mainly repaired through non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ repair does not require a repair template and is error-prone. Therefore, when DSBs are repaired through NHEJ, small insertion or deletion mutations are frequently introduced at the sites. On the other hand, HDR requires a repair template, which could be a sister chromatid, other homologous region, or exogenous repair donor. Thus, co-introduction of targetable nucleases and exogenous repair donors facilitates transgene integration through HDR repair

that is precise compared with NHEJ but requires a repair template. Therefore, targeted DSB introduction using targetable nucleases enables efficient introduction of mutations into the site of interest (Fig. 1.1). Additionally, co-introduction of targetable nucleases and exogenous repair donors (or targeting vectors) increases the HDR-mediated targeted transgene integration (or knockin) compared with conventional gene targeting (Fig. 1.1).

To date, the genomic sequences of many model organisms have been edited by using these technologies. The model systems have involved, cultured human cells including human embryonic stem (ES) and induced pluripotent stem (iPS) cells (see Chap. 3), nematodes (see Chap. 4), silkworms (see Chap. 5), sea urchins (see Chap. 6), ascidians (see Chap. 7), fish (see Chap. 8), amphibians (see Chap. 9), mice (see Chap. 10 for mouse editing with CRISPR/Cas and Chap. 11 mouse editing with TALENs), rats (see Chap. 12), and plants (see Chap. 13).

Although the technical difficulties associated with designing ZFNs that target specific sequences of interest have hindered the expansion of ZFN-mediated genome editing, the basis of the genome editing technology was built on studies using these

nucleases since 1996 (Kim et al. 1996). After the emergence of TALENs in 2010 (Christian et al. 2010) and the application of CRISPR/Cas9 for genome editing in 2013 (Cong et al. 2013; Mali et al. 2013b), which are easier to design for introduction of DSBs at the intended genomic sites than ZFNs, the technology has explosively expanded to a wide range of researchers. Consequently, we are definitely now in a genome editing era.

In this chapter, we will review the basics of the DSB repair and the molecular basis of ZFNs and TALENs. In particular, we will focus on the mechanisms of nucleotide sequence recognition and the techniques for designing sequence-specific nucleases. Finally, we will provide the examples of the genome editing using these nucleases.

1.2 DSB Repair

DSBs potentially induce loss of genetic material and oncogenic transformation through deletion and translocation mutations, and are therefore immediately repaired in cells. DSBs are mainly repaired through NHEJ or HDR (Fig. 1.1). Inherently error-prone NHEJ repair does not require a repair template and does not depend on the cell cycle. In most cases, small deletions and insertions are introduced at DSB sites through NHEJ repair. On the other hand, HDR requires a homologous repair template, such as a sister chromatid, other homologous region, or exogenous repair donor, and mainly occurs at the S/G2 phase of the cell cycle. In general, although HDR is considered to be a high-fidelity repair process, most DSBs are repaired through error-prone NHEJ. Therefore, introduction of DSBs in a gene of interest by targetable nucleases followed by NHEJ-mediated repair efficiently introduces insertion and/or deletion mutation, resulting in gene disruption (Fig. 1.1). Moreover, co-introduction of targetable nucleases and targeting vectors containing approximately 1-kb homologous regions from DSB sites facilitates HDR-mediated transgene introduction into the DSB site (Fig. 1.1). The DSB repair mechanisms are evolutionarily conserved among most of the organisms. Thus, targetable nuclease-mediated genome editing might be applicable to at least animals and plants.

1.3 ZFNs

1.3.1 Zinc-Finger Domain

ZFNs are artificial proteins comprised of a C_2H_2 zinc-finger (ZF) DNA-binding domain and a restriction enzyme FokI nuclease domain (Fig. 1.1). The C_2H_2 ZF domain is comprised of about 30 amino acids, contains two antiparallel β-sheets and an α-helix, and is characterized by a zinc ion coordinated by two cysteine residues in the β-sheets and two histidine residues in the α-helix (Fig. 1.2).

In natural proteins, the C_2H_2 ZF domain is principally involved in DNA binding. Basically, each finger recognizes a three-nucleotide subsite (target triplet, Fig. 1.2)

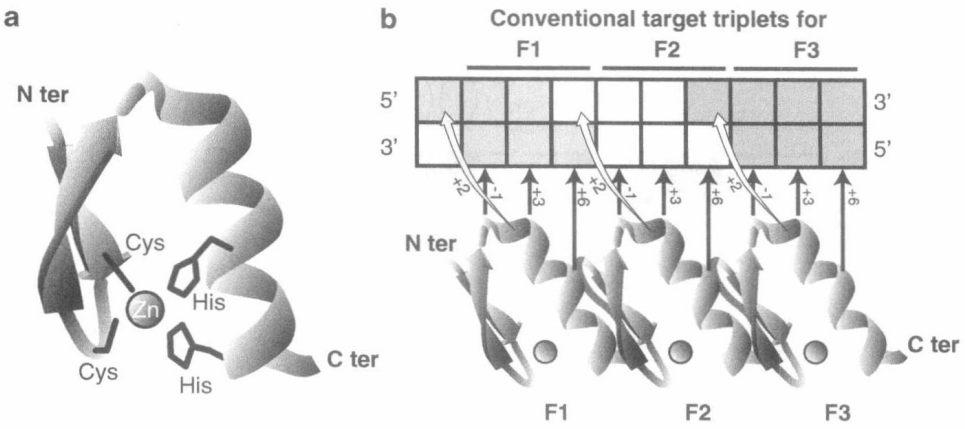

Fig. 1.2 Structure of a zinc-finger (ZF) domain. (**a**) Schematic representation of 3D structure of ZF domain. Two cysteine residues and two histidine residues coordinate a zinc ion. (**b**) Base recognition of three-finger ZFs. Fundamentally, three amino acids at helical positions −1, +3 and +6 in each ZF domain (F1, F2, and F3) directly recognize each target triplet, or subsite (*black arrows*). Sometimes, amino acid at helical position +2 is also involved in the recognition of a base in the antisense strand of an adjacent subsite (*white filled arrows*)

through its α-helix, the so-called recognition helix. It is considered that the amino acids at helical positions −1, +3 and +6 are involved in the recognition of and binding to only one strand of a target triplet. Consequently, a three-finger ZF domain recognizes a 9-bp sequence. Moreover, since ZFNs basically cleave DNA as a dimer (Figs. 1.1 and 1.3, and see below), the target of three-finger ZFNs becomes 18 bp in length, which is sufficiently long to specify a single target site in the human genome. In the helical residues, associations of specific amino acids with specific nucleotide bases are observed to some extent (Wolfe et al. 2000). However, it is considered that the base recognition is influenced by associations with other amino acids inside or outside of each finger and the nucleotide composition of the target DNA. Moreover, the residue at helical position +2 sometimes recognizes a nucleotide base in the anti-sense strand of the neighboring triplet (Fig. 1.2), resulting in an overlap between the recognition sequences of neighboring fingers. These complications make the de novo design of ZFNs, which recognize specific nucleotides, difficult. Furthermore, each finger shows strong preference for a GNN triplet for their recognition. This property strikingly restricts the target site repertoire for a pair of ZFNs (e.g. NNCNNCNNCNNNNNNGNNGNNGNN, underlines represent 6-bp spacer sequence, Fig. 1.3).

1.3.2 FokI Nuclease Domain

FokI is a type IIS restriction enzyme that cleaves DNA at a defined distance away from their recognition sites. Chandrasegaran and colleagues found that the DNA-binding and non-specific DNA-cleavage functions of FokI could be divided into

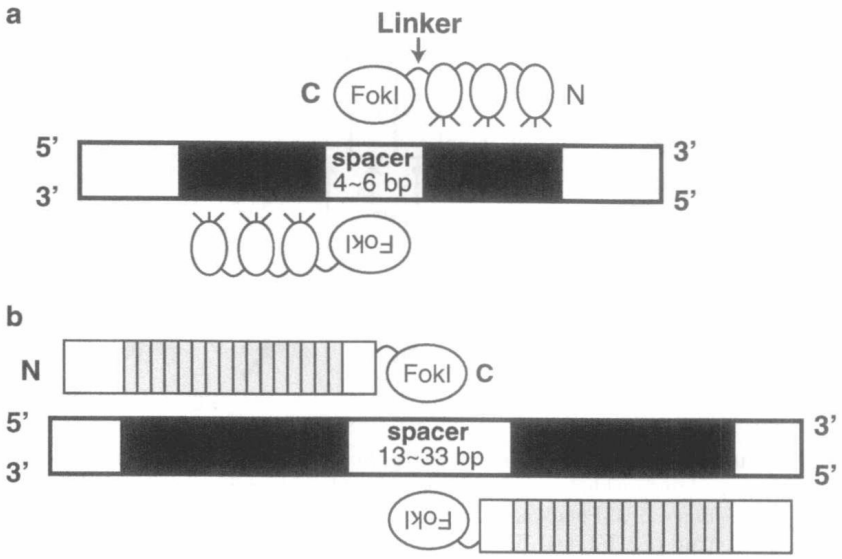

Fig. 1.3 Schematic representation of ZFNs (**a**) and TALENs (**b**). *Black boxes* represent target sequences. Nucleotide sequences between target nucleotides are referred to as spacer sequences

different domains (Li et al. 1992). They fused the FokI nuclease domain with ZFs and created artificial nucleases with novel specificities (Kim et al. 1996). This represented the first demonstration of ZFNs.

FokI needs to be dimerized to cut DNA. A full-length FokI bound to its target site can become dimerized with another unbound FokI in solution and subsequently cleave DNA (Pernstich and Halford 2012). Conversely, perhaps because of the low affinity, the nuclease domains in ZFNs only efficiently dimerize and cut DNA when both ZFNs bind to their target sites separated by the spacer sequence and face one another (Fig. 1.3) (Smith et al. 2000). Thus, ZFNs acquire high specificity from this attribute. However, if not only the target sites for a pair of ZFNs but also palindromes of the target site for one of the ZFNs exist in the genome, both sites could be cleaved by the ZFNs (Fig. 1.4). In the latter case, DSB introduction into unintended sites is called "off-target" cleavage. To avoid this incidence, obligate heterodimeric mutants of FokI nucleases were invented (Fig. 1.4) (Miller et al. 2007; Szczepek et al. 2007; Doyon et al. 2010). The homodimerization of these mutants is significantly inhibited, resulting in a drastic reduction of cytotoxicity (Miller et al. 2007; Szczepek et al. 2007; Doyon et al. 2010). A pair of ZFNs with wild-type (WT) and nuclease-dead mutant FokI nuclease domains cleaves only one strand of DNA (Ramirez et al. 2012; Kim et al. 2012; Wang et al. 2012). Such a single-strand break cannot be repaired through error-prone NHEJ, but is repaired by HDR. Therefore, these "ZFNickases" enable HDR-mediated genome editing almost completely without unwanted off-target effects (Ramirez et al. 2012; Kim et al. 2012; Wang et al. 2012). However, the efficacy of nick-induced HDR is extremely low. ZFs were fused with not only FokI nuclease domains but also other functional domains to generate site-specific activator (Liu et al. 1997), repressor (Liu et al. 1997), methyltransferase (Meister et al. 2010), and recombinase (Sirk et al. 2014).