DE GRUYTER
OPEN

CRISPR/Cas in genome defense and gene editing

Svetlana Kryštofová

The Institute of Biochemistry and Microbiology, Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava
svetlana.krystofova@stuba.sk

Abstract: Targeted genome editing using engineered nucleases such as ZFNs and TALENs has been rapidly replaced by the CRISPR/Cas9 (clustered, regulatory interspaced, short palindromic/ CRISPR-associated nuclease) system. CRISPR/Cas9 technology represents a significant improvement enabling a new level of targeting, efficiency and simplicity. Gene editing mediated by CRISPR/Cas9 has been recently used not only in bacteria but in many eukaryotic cells and organisms, from yeasts to mammals. Other modifications of the CRISPR-Cas9 system have been used to introduce heterologous domains to regulate gene expressions or label specific loci in various cell types. The review focuses not only on native CRISPR/Cas systems which evolved in prokaryotes as an endogenous adaptive defense mechanism against foreign DNA attacks, but also on the CRISPR/Cas9 adoption as a powerful tool for site-specific gene modifications in fungi, plants and mammals.

Keywords: genome, CRISPR, Cas9, single guide RNA

Introduction

Technologies for manipulating, modifying and synthesizing DNA have expanded throughout all fields of biology over the past decades. At the very beginning was the DNA structure discovery and later rapid evolution of DNA/RNA manipulations including PCR, DNA cloning and specific gene targeting. The advances in whole-genome sequencing technologies along with hardware and software boom paved the way toward many a fascinating discovery in biology, biotechnology, agriculture, pharmacy and medicine. Recently, the RNA guided nuclease Cas9, which is a part of the CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated) adaptive bacterial defense system, is revolutionizing biology by providing an efficient and accurate genome engineering tool for editing both prokaryotic and eukaryotic organisms. This remarkable revolution has unraveled in the span of less than 4 years since the initial publication in 2012 (Wiedenheft et al., 2012; Jinek et al., 2012).

CRISPR-Cas is one of three genome editing methods which are currently in use. The basic strategy of the genome editing methods is based on three preconditions: sequence recognition, double strand cleavage and cellular DNA repair mechanisms (Gaj et al., 2013). The sequence recognition and DNA digest can be achieved by chimeric zinc finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs) (Urnov et al., 2010; Caroll, 2011; Christian et al., 2010; Reyon et al., 2012). ZFNs are composed of a zinc finger domain for specific DNA recognition and a nuclease domain (such as FokI) to cleave the specific DNA site. These chimeric nucleases induce targeted DNA double-strand breaks (DSBs) that initiate the cellular DNA repair mechanisms, including error-prone non-homologous end joining (NHEJ) and homology-directed repair (HDR) (Wyman and Kanaar, 2006). This process enables introduction of genetic modifications to specific loci. Although the system utilizing the ZFNs was found to be potent at inducing genome sequence changes in *Drosophila,* mammalian cells and plants (Bibikova et al., 2002, 2003; Zhang et al., 2010), it was not generally employed because of the difficulties in designing and validating ZFNs for some specific DNA loci of interest. The transcription activator-like effectors (TALEs), which are naturally present in plant pathogenic bacteria, could be used in similar matter to ZFNs for site-directed genome editing (Boch et al., 2009; Moscou et al., 2009). Other effector domains such as nucleases, transcriptional activators and site-specific recombinases have been fused to TALE repeats to enable targeted genetic modifications (Miller et al., 2011; Zhang et al., 2011; Mercer et al., 2012). Although TALE nucleases (TALENs) were better than ZFNs to produce and validate, but the requirements of specific enzyme engineering for various targets in transcription activator-like effector nucleases limit their applications.

CRISPR/Cas in prokaryotes

Bacteria possess multiple ways to regulate foreign DNA invasions and resist phage attacks. These mechanisms could be carried out by the mutation or masking of cell surface receptors, restriction modi-

Acta Chimica Slovaca, Vol. 9, No. 1, 2016, pp. 68—74, DOI: 10.1515/acs-2016-0012

fication systems, abortive infection and the CRISPR systems (Richter et al., 2012). CRISPR systems are a widespread mechanism among prokaryotes that provides bacteria with an "adaptive immune system" that carries a genetic memory of past genetic invasions (Wiedenheft et al., 2012).

CRISPR/Cas is comprised of small non-coding RNAs for defense and Cas proteins required for transcription and processing small crRNAs (CRISPR RNA) and foreign DNA degradation (Fig. 1). The CRISPRs are composed of a leader sequence, short repeats and intervening sequences originated from foreign invasions. The CRISPR sequences are usually termed as the CRISPR arrays. Upon infection with a foreign element (e.g., phages or plasmids), part of the foreign DNA is typically attached to the leader end of the CRISPR array and the endogenous repeat is duplicated. The CRISPR arrays are adjacent to a cluster of *cas* genes (Richter et al., 2012; Rath et al., 2015; Doudna and Charpentier, 2014).

The CRISPR/Cas interference consists of three phases (Fig. 1): a) the integration of short sequences from foreign genetic elements termed spacers into the CRISPR array, b) CRISPR array transcription and processing into crRNAs by Cas proteins, c) targeting of the foreign DNA by a Cas protein complex that is guided by crRNAs. interferes in a sequence specific manner with the foreign DNA. Within arrays, the repeats are typically identical in terms of length and sequence but the repeat at the end of an array is often truncated or deviates from the consensus sequence (Horvath et al., 2008). Repeats are usually between 23 and 47 bp in length. In contrast, the spacer sequences are usually unique within a genome. Many spacers carry sequences similarities to extra-chromosomal sources such as phage or plasmids (Mojica et al., 2005; Bolotin et al., 2005). The spacers confer the sequence-specific immunity against those extra-chromosomal agents. The sequences in the foreign DNA from which spacers are evolved are labeled as protospacers. The third component of the CRISPR array is the leader sequence, which is positioned upstream of the first repeat (Richter et al., 2012; Rath et al., 2015; Jansen et al., 2002). The leader sequence is about 200—500 bp long and includes the promoter required for transcription of the array. The leader region plays an important role in the spacer acquisition (Pougach et al., 2010; Yosef et al., 2012).

Cas proteins supply the enzymatic machinery required for the acquisition of new spacers from invading foreign DNA as well as targeting and eliminating them (Fig. 2, Fig. 3). CRISPR/Cas systems are currently classified into type I, II and III based on the phylogenic analysis and presence of particular Cas proteins (Fig. 3). Cas1 and Cas2 are present in every Cas operon of the three main types, and Cas1 is considered the universal marker of CRISPR/Cas systems (Makarova et al., 2011). Other Cas proteins such as Cas3, Cas9, and Cas10 are signature proteins for type I, II, and III, respectively. Cas1 and Cas2 are important for spacer incorporation from plasmids and phages

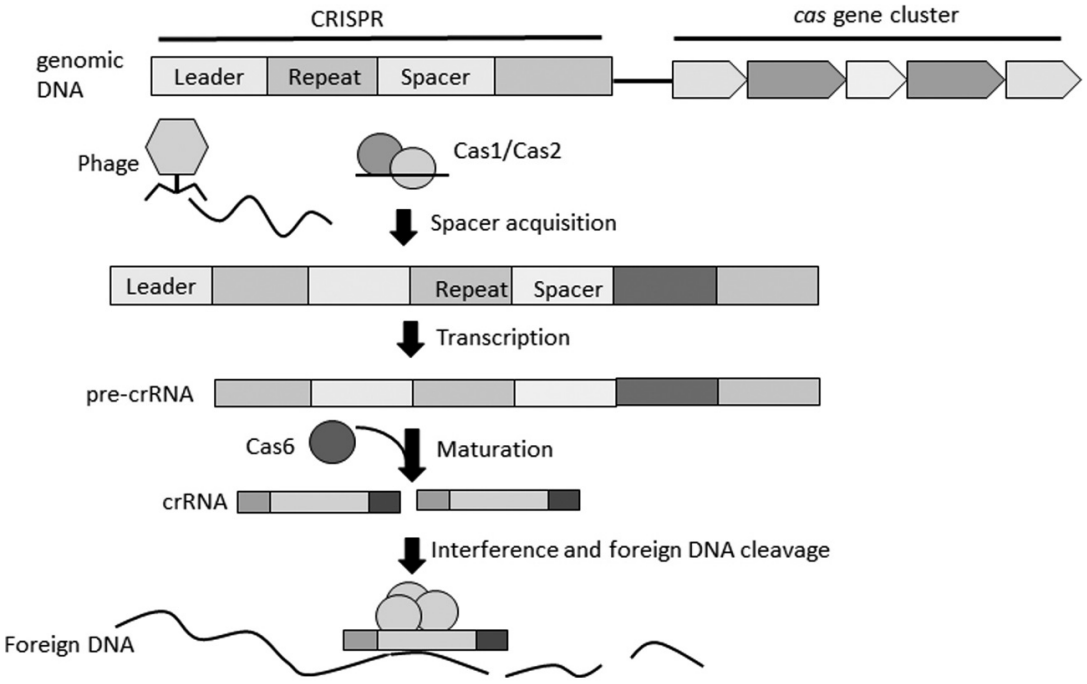

Fig. 1. Overview of CRISPR/Cas mechanism: spacer acquisition, pre-crRNA transcription and crRNA maturation, interference and foreign DNA digest (modified according to Richter et al., 2012).

(Yosef et al., 2012). Cas1 and Cas2 are magnesium dependent nucleases but their exact function in the spacer acquisition remains unclear. Both nucleases form a dimer and mutations in their sequences interfere with the foreign DNA acquisition. There are two types of spacer acquisition; naive, in which the invader has not been previously encountered, and primed, in which there is a pre-existing sequence of the invader in the CRISPR (Fig. 2). In naive spacer acquisition, spacer selection is guided by certain sequence elements in the target DNA. Analysis of target sequences has revealed a short motif next to the protospacer sequence termed PAM (protospacer adjacent motif) that is crucial for identification of self and non-self-sequences (Mojica et al., 2009). PAM is important for both foreign DNA interference as well as spacer acquisition (Fig. 2). In both the type I-E and type II-A systems, it was demonstrated that parts of the leader and one repeat are required for spacer integration. Moreover, the first repeat serves as template for synthesis of the new repeat (Yosef et al., 2012) (Fig. 2). The protospacer introduced into the CRISPR array contains the last nt of the PAM motif, which becomes the final nt of

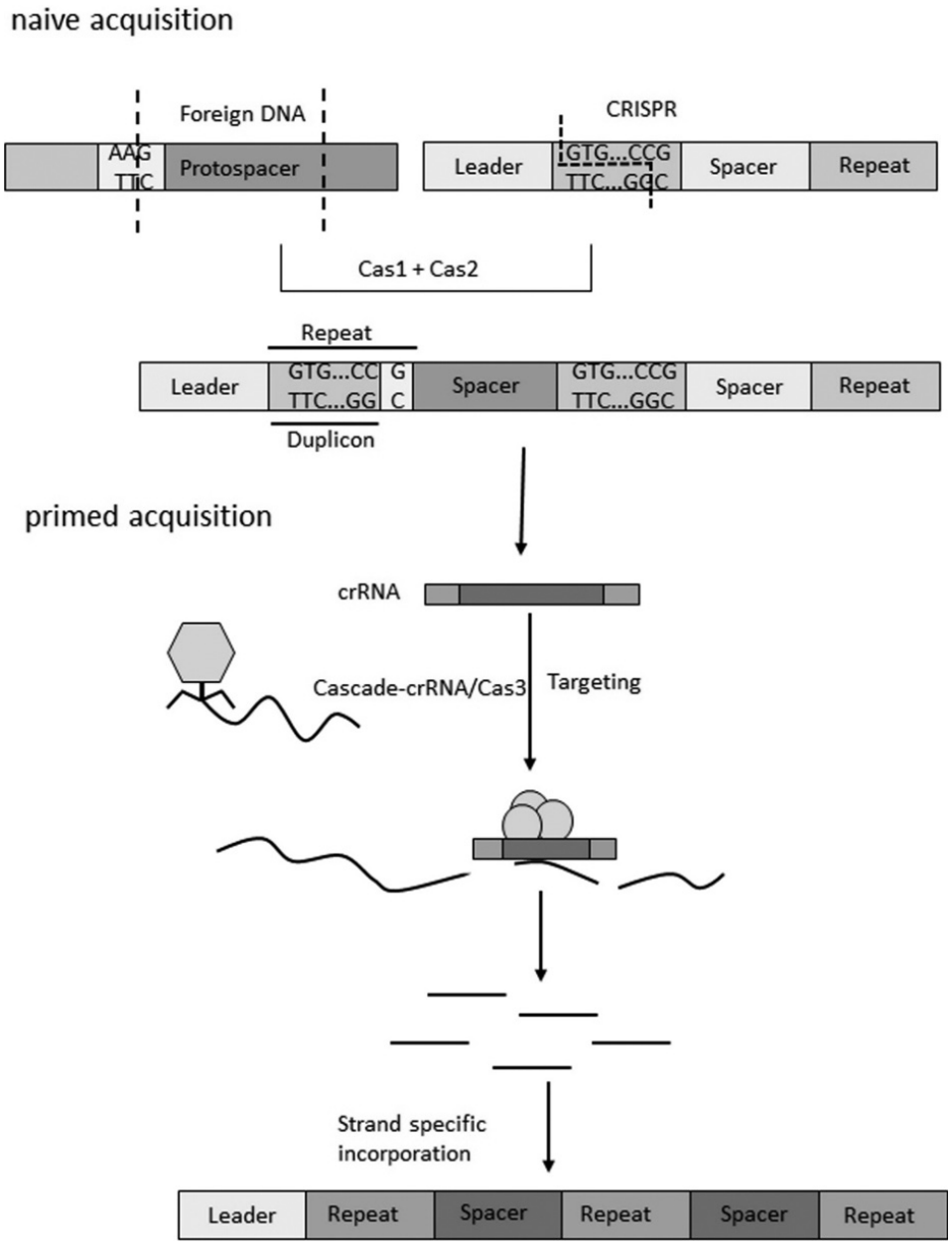

Fig. 2. CRISPR/Cas adaptation (modified according to Richter et al., 2012). The first repeat adjacent to the leader sequence of the CRISPR array is copied, and a new spacer sequence from a protospacer is incorporated. The last nucleotide of the repeat is not duplicated and the repeat connects to the PAM nucleotide of the new spacer. Primed acquisition requires crRNA, Cas3, Cas1 and Cas2 and results in incorporation of new spacers derived from the initial spacer.

the 5' repeat (Swarts et al., 2012). The spacer integration is directed not only by sequence but structure of the CRISPR as well. The palindromic nature of many CRISPR repeats is essential to determine the location and direction of spacer integration into the array (Nuñez et al., 2015). The adaptation can be accelerated by primed spacer acquisition, which occurs when the targeting spacer is already present in the CRISPR array (Fig. 2).

CRISPR transcription initiates in the leader region (Spilman et al., 2013). The leader contains promoter elements and binding sites for regulatory proteins. A long primary transcript for the pre-crRNA is generated and may contain secondary structures (hairpins). The pre-crRNA is subsequently processed into smaller units encoding a single spacer flanked by partial repeats (Fig. 3). The Cas proteins responsible for the processing vary with the subtype. Type I and III systems utilize a Cas6 protein to process pre-crRNA. The Cas6e and the crRNA (it has hairpin structure) are crucial components of the *E. coli* Cascade complex, which also contains one copy of CRISPR-associated protein Cse1, two copies of Cse2, one copy of Cas5e and six copies of Cas7 (Jore et al., 2011). In type III, the 5' end of the crRNA interacts with CRISPR typ III associated RAMP proteins Csm1-Csm4/CMR2-CMR3 and the 3' end with Csm5/CMR1-CMR6 (Spilman et al., 2013). Type II systems employ a very different strategy for crRNA generation where processing is dependent on host RNase III and a trans-encoded small RNA (tracrRNA) that base pairs with the pre-crRNA (Fig. 3). Type II processing also requires the Cas9 protein. Another distinct feature of the type II systems is the 5' processing of the crRNA by an unknown nuclease (Jinek et al., 2012; Deltcheva et al., 2011).

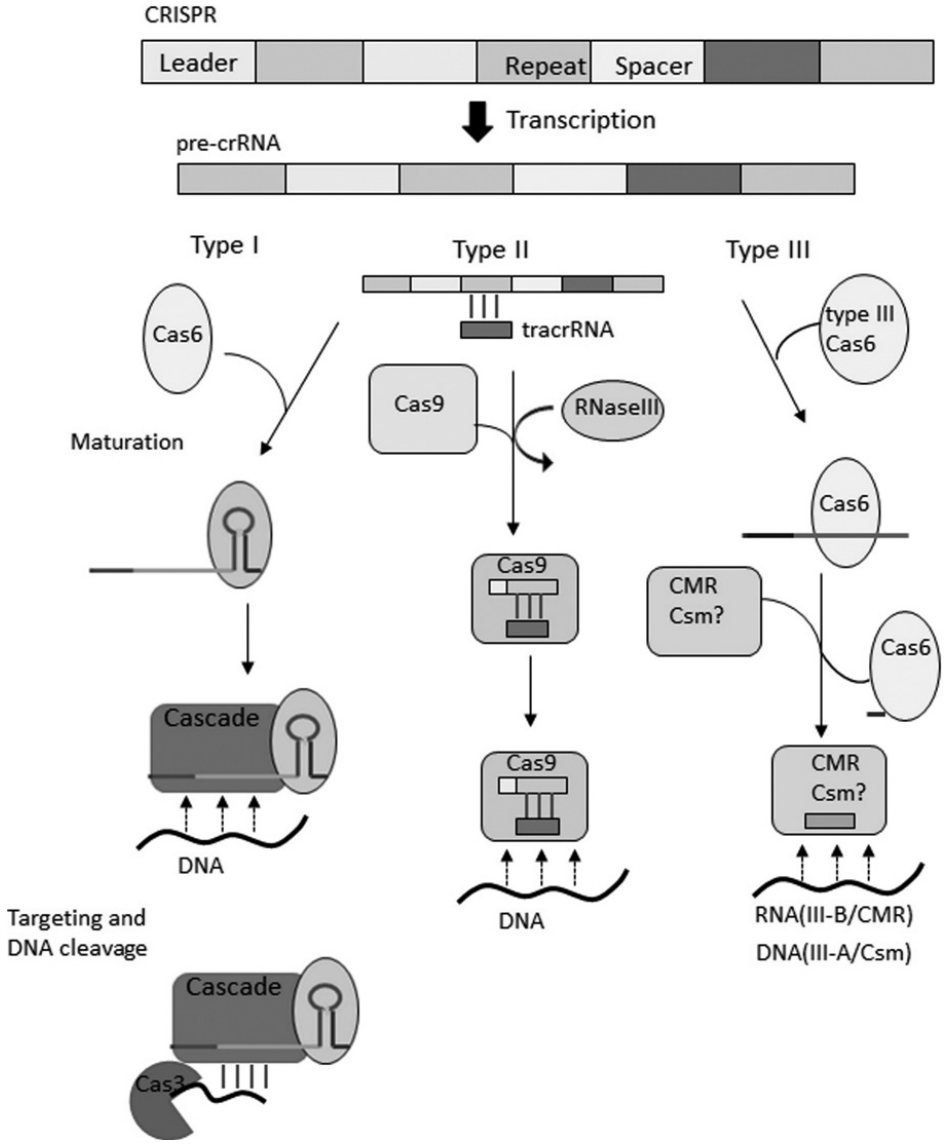

Fig. 3. Overview of crRNA generation and target interference in type I, II and III CRISPR/Cas systems (modified according to Richter et al., 2012).

The principle of target DNA interference by CRISPR-Cas systems is that crRNA bound to Cas protein(s) locates the corresponding protospacer to initiate degradation of the target (Fig. 3). The degradation is performed by specific Cas nucleases (Makarova et al., 2015). Type I and II systems interference requires the presence of a PAM sequence and perfect protospacer-crRNA complementarity in the so-called seed region, positioned adjacent to the PAM (Sternberg et al., 2015). The presence of a PAM triggers non-self-recognition, which prevents the systems from targeting its own CRISPR locus. In type I, the binding of the crRNA to the target causes conformational changes in the target DNA that could initiate Cas3 recruitment. Cas3 then cuts the target DNA and proceeds with degradation of the target (Westra et al., 2012). Type III-A typically contains 6 different proteins but the nuclease is not yet identified. Csm complexes target DNA (Marraffini and Sontheimer, 2008) and CMR complexes target RNA (Zebec et al., 2014), but targeting RNA and DNA by the same Cmr complex has been found in *S. islandicus* (Peng et al., 2015). Type II systems require the Cas9 protein, crRNA and also a tracrRNA bound to Cas9. The crRNA is responsible for target recognition and degradation (Deltcheva et al., 2011).

Gene editing with CRISPR technology in cells and model organisms

The CRISPR/Cas system appears the most popular genome-editing tool at the moment. Although other programmable editing tools, such as ZFNs

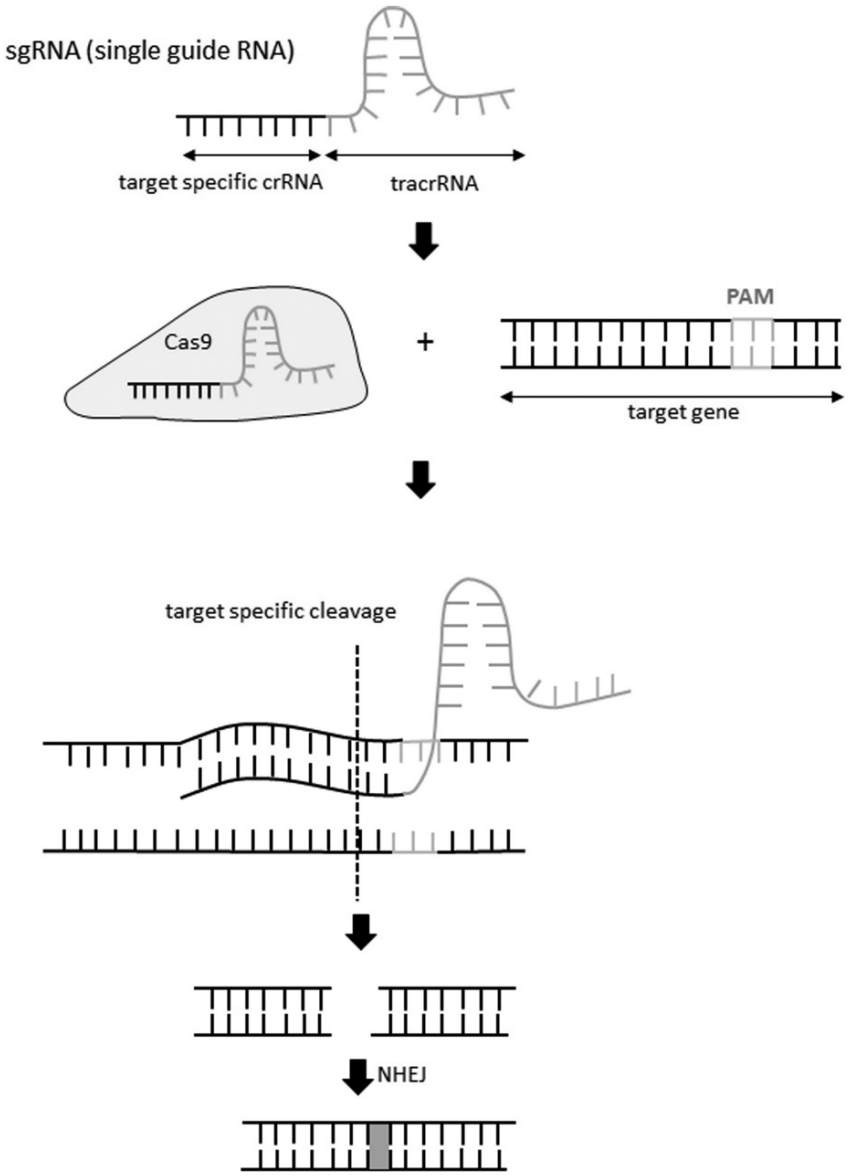

Fig. 4. Overview of genome editing using CRISPR/Cas9 system. The target gene is cleaved by a complex containing Cas9 and sgRNA composed of crRNA and tracrRNA sequences.

and TALENs have significantly improved the efficiency for precise genome modification, these techniques have some limitations. CRISPR/Cas9 technology represents a significant improvement enabling a new level of targeting, efficiency and simplicity (Cong et al., 2013). Over the past 2 years, many studies have presented the CRISPR/Cas9 system as a powerful genome-editing method that facilitates genetic alterations in genomes of different organisms (Jinek et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Feng et al., 2013; DiCarlo et al., 2013). A single multidomain Cas9 initiates a double-strand break (DSB) in the target DNA composed of a 20-bp sequence interacting with the protospacer of the single guide RNA (sgRNA) and a downstream PAM sequence, 5'-NGG (Fig. 4). The type II CRISPR/Cas system uses non-coding RNAs and is designed to create a simple, universal RNA-programmable method to facilitate genome editing in cells. It can be used to generate gene knock-outs (via insertion/deletion) or knock-ins (via HDR). To create gene disruptions, a single guide RNA (sgRNA) directs the Cas9 nuclease to DNA specific site where it catalysis cleavage. The DNA damage is subsequently repaired by either NHEJ or HDR pathways (Fig. 4).

CRISPR/Cas9 found the most eager users in medical research. Its ability to introduce DSBs at defined sites has enabled to generate mammalian cell lines and cells carrying chromosomal translocations resembling those found in cancers such as lung cancer, acute myeloid leukemia (Choi and Meyerson, 2014; Chen et al., 2014). Other applications of CRISPR/Cas9 with relevance to human health include corrections of genetic mutations of inherited disorders such as cataracts and cystic fibrosis (Wu et al., 2013; Schwank et al., 2013).

Another example is the systematic analysis of gene functions in mammalian cells. A genome-scale lentiviral sgRNA library was created to generate a loss-of-function genetic screening approach designed for both positive and negative selection (Wang et al., 2014; Zhou et al., 2014). This method was also applied in identifications of genes essential for cell viability in cancer and stem cells (Shalem et al., 2014).

Until now, there have been only few reports on the CRISPR/Cas9 system or other genome editing methods in fungi despite the successful application of this technique in mammalian cells. CRISPR/Cas9 application in yeast strains lead to generation of single and multiple mutations with the aim to identify genes important for biosynthesis biotechnologically interesting products such as lactic acid, mevalonate (Jakočiunas et al., 2015; Stovicek et al., 2015). CRISPR/Cas9 system works efficiently in *Neurospora crassa*, in which the endogenous promoter of *clr-2* was replaced with the β-*tubulin* promoter and inserted a codon optimized fire fly luciferase under the control of the *gsy-1* promoter at the *csr-1* locus (Matsu-ura et al., 2015). CLR-2 is a transcription factor that regulates the expression of cellulases, and GSY-1 regulates the conversion of glucose into glycogen. The efficient CRISPR/Cas9 system in another filamentous fungus *Trichoderma reesei* confirmed itself as a powerful genome-manipulating tool other filamentous fungal species. In *T. reesei*, it was shown that the CRISPR/Cas9 system was controllable and conditional through inducible Cas9 expression (Liu et al., 2015).

Acknowledgement

This work was supported by courtesy of the Slovak Grant Agency VEGA 1/0870/14, the Slovak Research and Development Agency APVV-0719-12, APVV/0282-10.

References

Bibikova M, Beumer K, Trautman JK, Carroll D (2003) Science 300: 764.

Bibikova M, Golic M, Golic KG, Carroll D (2002) Genetics 161: 1169–1175.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Science 326: 1509–1512.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Microbiology 151: 2551–2561.

Carroll D (2011) Genetics 188: 773–782.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Science 339: 819–823.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) Nature 471: 602–607.

DiCarlo JE, Norville JE, Mali P, Rios X, John Aach GM (2013) Nucl. Acids Res. 41: 4336–4343.

Doudna JA, Charpentier E (2014) Science, 346 (6213): art. no. 1258096.

Feng Z, Zhang B, Ding W, Liu X, Yang D-L, Wei P, Cao F, Zhu S, Zhang F, Mao Y, Zhu J-K (2013) Cell Research 23: 1229–1232.

Gaj T, Gersbach CA, Barbas CF (2013) Trends Biotechnol 31: 397–405.

Horvath P, Romero DA, Côuté-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R (2008) J. Bacteriol. 190: 1401–1412.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh J-RJ, Joung JK, (2013) Nat. Biotechnol. 31: 227–229.

Chen C, Liu Y, Rappaport AR, Kitzing T, Schultz N, Zhao Z, Shroff AS, Dickins RA, Vakoc CR, Bradner JE, Stock W, LeBeau MM, Shannon KM, Kogan S, Zuber J, Lowe SW (2014) Cancer Cell 25: 652–665.

Choi PS, Meyerson M (2014) Nat. Commun. 5: 3728.

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Genetics 186: 757–761.

Jakočiunas T, Bonde I, Herrgård M, Harrison SJ, Kristensen M, Pedersen LE, Jensen MK, Keasling JD (2015) Metabolic Engineering 28: 213—222.
Jansen R, van Embden JD, Gaastra W, Schouls LM (2002) OMICS 6: 23—33.
Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) Nat. Biotechnol. 31: 233—239.
Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J (2013) eLife 2: e00471.
Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) Science 337: 816—821.
Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul Ü, Wurm R, Wagner R et al. (2011) Nat. Struct. Mol. Biol. 18: 529—536.
Liu R, Chen L, Jiang Y Zhou Z, Zou G (2015) Cell Discovery 1: 15007.
Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF van der Oost J, Koonin EV (2011) Nat. Rev. Microbiol. 9: 467—477.
Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, Van Der Oost J, Backofen R, Koonin EV (2015) Nature Reviews Microbiology 13: 722—736.
Marraffini LA, Sontheimer EJ (2008) Science 322: 1843—1845.
Matsu-ura T, Baek M, Kwon J, Hong C (2015) Fungal Biol Biotechnol 2: 4—10.
Mercer AC, Gaj T, Fuller RP, Barbas CF (2012) Nucleic Acids Res. 40: 11163—11172.
Miller JC, Siyuan T, Guijuan Q, Kyle AB, Jianbin W, Danny FX, Xiangdong M, David EP, Elo L, Sarah JH, Gladys PD, Kevin LH, Irina A, Gregory JC, Fyodor DU, H Steve Z, Michael CH, Lei Z, Philip DG, Edward JR (2011) Nat. Biotechnol. 29: 143—148.
Mojica FJ, Díez-Villaseñor C, García-Martínez J, Soria E (2005) J. Mol. Evol. 60: 174—182.
Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Microbiology 155: 733—740.
Moscou MJ, Bogdanove AJ (2009) Science 326: 1501.
Nuñez JK, Lee ASY, Engelman A, Doudna JA (2015) Nature, 519: 193—198.
Peng W, Feng M, Feng X, Liang YX, She Q (2015) Nucleic Acids Research 43: 406—417.
Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K (2010) Mol. Microbiol.77: 1367—1379.
Rath D, Amlinger L, Rath A, Lundgren M (2015) Biochimie 117:119—128.
Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK (2012) Nat. Biotechnol 30: 460—465.
Richter C, Chang JT, Fineran PC (2012) Viruses 4: 2291—2311.
Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014) Science 343: 84—87.
Schwank G, Koo B-K, Sasselli V, Dekkers JF, Heo I, Demircan T, Sasaki N, Boymans S, Cuppen E, van der Ent CK, Nieuwenhuis EES, Beekman JM, Clevers H (2013) Cell Stem Cell 13: 653—658.
Spilman M, Cocozaki A, Hale C, Shao Y, Ramia N, Terns R, Terns M, Li H, Stagg S (2013) Molecular Cell 52: 146—152.
Sternberg SH, Lafrance B, Kaplan M, Doudna JA (2015) Nature 527: 110—113.
Stovicek V, Borodina I, Forster J (2015) Metabolic Engineering Communications 2: 13—22.
Swarts DC, Mosterd C, van Passel MW, Brouns SJ (2012) PLoS One 7: e35888.
Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Nat. Rev. Genet 11: 636—646.
Wang T, Wei JJ, Sabatini DM, Lander ES (2014) Science 343: 80—84.
Westra E, van Erp P, Künne T, Wong S, Staals R, Seegers C, Bollen S, Jore M, Semenova E, Severinov K, de Vos W, Dame R, de Vries R, Brouns SJJ, van der Oost J (2012) Molecular Cell 46: 595—605.
Wiedenheft B, Sternberg SH, Doudna JA (2012) Nature 482: 331—338.
Wu Y, Liang D, Wang Y, Bai M, Tang W, Bao S, Yan Z, Li D, Li J (2013) Cell Stem Cell 13: 659—662.
Wyman C, Kanaar R (2006) Annu. Rev. Genet. 40: 363—383.
Yosef I, Goren MG, Qimron U (2012) Nucleic Acids Res. 40: 5569—5576.
Zebec Z, Manica A, Zhang J, White MF, Schleper C (2014) Nucleic Acids Research 42: 5280—5288.
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) Proc. Natl. Acad. Sci. U. S. A. 107: 12028—12033.
Zhang, F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P (2011) Nat. Biotechnol. 29: 149—153.
Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, Wei W (2014) Nature 509: 487—491.