



GENE TARGETING AND GENOME EDITING

The delivery systems used in gene therapy are non-specific, infecting more than one cell type. In *ex vivo* or *in situ* manipulation this is not a serious problem. However, if *in vivo* therapy is to be developed, then cell specificity becomes desirable. In such cases, the gene carriers can be injected into the bloodstream much like administering many drugs.

20.1 Recombination

Gene targeting is a technology based on homologous recombination, a biological process occurred widely in prokaryotic cells and less frequently in eukaryotes. In homologous recombination, two double-stranded DNA molecules with a region of homologous sequence, line up adjacent to one another, and through a series of complex steps, exchange the two identical DNA segments. This type of homologous recombination, involving a swap of two homologous sequences, is known as reciprocal exchange or conserved exchange. In some cases, exchange of nucleotides in the homologous sequence may also be unidirectional. This type of exchange is non-reciprocal or non-conservative. It is also referred to as gene conversion, because a portion of the recipient sequence is converted to the incoming sequence.

Homologous recombination provides a unique way to introduce foreign DNA into a specific location or to engineer genes *in situ* at their natural loci in the genome. Most gene targeting involves engineering alterations to a chosen gene for the purpose of studying gene structure/function. Targeted alterations of a chosen gene are called “gene knockout”. The approach, however, is most appealing as a potential application in gene therapy. In the gene therapy protocols commonly in use, the introduction of a therapeutic gene integrates into the genome randomly, and thus requires transcriptional and translational regulatory elements in the gene construct. This is a complementation process in which a defective gene is augmented by introducing a functional gene. In contrast, gene

targeting enables a direct replacement of a defective gene. The sequence carrying the mutation is replaced by the therapeutic gene sequence. The regulatory region of the gene may not need to be considered in the operation. Gene targeting, of course, has wide use and implication in other areas as well, such as in the production of transgenic plants and animals.

20.2 Replacement Targeting Vectors

There are several methods of constructing a vector for various selection purposes. In one of the procedures, the engineered gene is constructed, so that it is interrupted by a selectable marker (such as the *neo* gene), and flanked by short sequences homologous to the sequence in the genomic loci targeted for replacement. A second selectable marker (for example, the thymidine kinase (*tk*) gene) is placed downstream of the gene and the homologous region. The two markers are known as positive and negative selectable markers, respectively. The entire construct (the gene plus homologous sequences at each end plus selectable markers) is a replacement targeting vector (Fig. 20.1).

The vector is introduced into suitable host cells by various methods, for example, microinjection, calcium phosphate precipitation, etc. Since the vector carries sequence homologous to the targeted gene in the genome, homologous recombination occurs replacing the genomic gene with the vector sequence. In homologous recombination, the vector aligns with the gene in the chromosome. The segment of the vector carrying the engineered gene and the *neo* gene will replace the targeted gene, while the *tk* gene lying outside of the homologous sequence, will not be included in the replacement. At the same time, the majority of the recombination occurs in a non-homologous way, resulting in random insertion. In this case, the entire vector DNA (the replacement gene + *neo* gene + *tk* gene) will be incorporated into the cell chromosome at random.

The final step is to select cells containing the target replacement. This is achieved by a double selection by growing all the cells in a medium containing G418 and ganciclovir. Non-transformants will not survive because they do not carry the *neo* gene, and therefore sensitive to G418 (a neomycin analog). Cells resulting from non-homologous recombination carry the herpes virus *tk* gene and will be sensitive to the nucleoside analog, ganciclovir. The only cells that can grow in the medium are the ones generated by homologous recombination.

The procedure has been incorporated with the use of embryonic stem (ES) cells for potential gene replacement in live animals. The targeting vector is introduced into mouse embryonic stem cells in culture via homologous recombination as described. Stem cells are undifferentiated cells in the early stage of an embryo that gives rise to various cell types during development (see Sect. 23.1). The ES cells from the selection step are introduced into the embryo at the blastocyst stage. Since ES cells are capable of developing into many cell types,

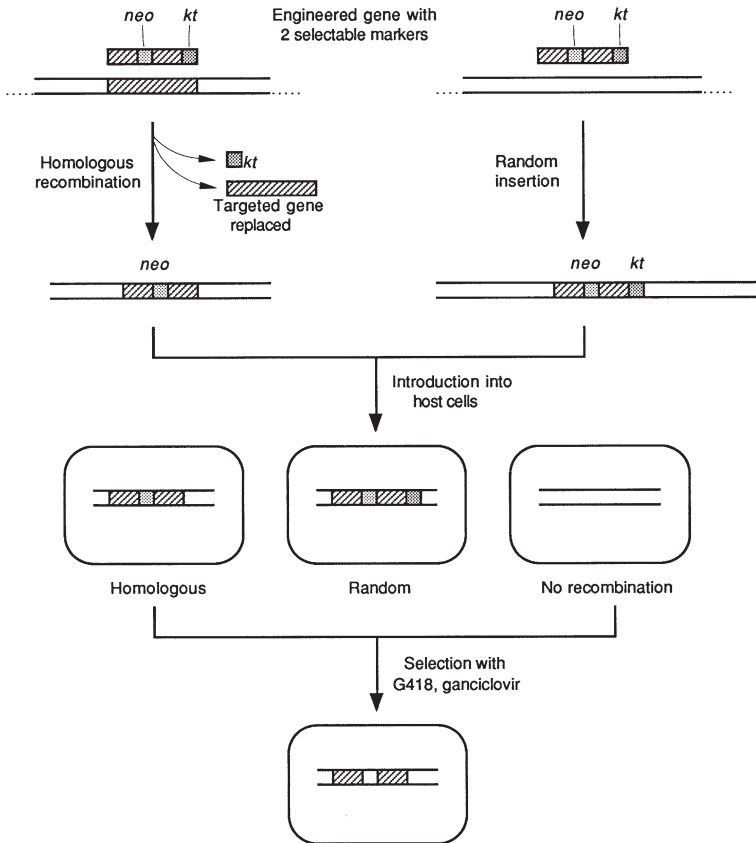


Fig. 20.1. Strategy in the use of a replacement targeting vector

the resulting mouse will carry the mutation in various tissue cells, including germ cells. However, germ line transmission from transgenesis (generated by injection of ES-like cells into blastocysts) has not been demonstrated so far for any species except mice.

20.3 Gene Targeting Without Selectable Markers

The insertion of a selectable marker in a gene for targeting is not desirable for two reasons. It causes the inactivation of the gene, which is fine for knockout experiments, but unsuitable for functional gene replacement purposes. In addition, a genetic marker that includes promoter/enhancer elements may run the risk of interfering transcription of neighboring genes. Strategies have been derived to introduce gene mutations by homologous recombination, without retaining the selectable markers in the targeted loci.

20.3.1 The PCR Method

Strategies have been derived to identify cells carrying the replacement gene, without the use of selectable markers. The detection method is based on the selective amplification of the recombinant DNA by PCR. In the case of targeting specific mutation to a gene, DNA from cells is amplified by PCR using two primers: primer 1 is identical to the mutation sequence, and primer 2 binds to an upstream sequence. Both primers will be used in PCR amplification if the cell DNA contains modified recombinant sequence. Double-stranded recombinant fragments will be generated in an exponential fashion. However, if homologous recombination has not occurred, the cell DNA will contain no binding site for primer 1, and PCR amplification yields ssDNA fragments non-exponentially. Modified cells are selected by analysis of the PCR products (Fig. 20.2).

20.3.2 The Double-Hit Method

In the double-hit gene replacement approach (also known as “tag and exchange”), two replacement type homologous recombination events are used. The first replacement vector is used to tag the gene by replacing part of the gene using positive (*neo* gene) and negative (*kt* gene) selectable markers. The resulting clones are subjected to positive selection (i.e. neomycin-resistance) to enrich for the replacement. In the second step, a replacement vector containing the gene with the mutation of interest is used to replace the selectable markers (*neo* and *kt*), and the clones that harbor the mutation can then be enriched by negative selection (Fig. 20.3).

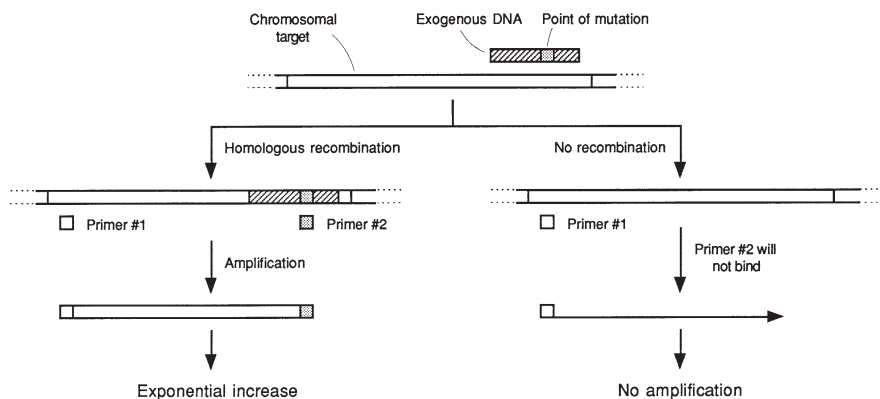


Fig. 20.2. Gene targeting without using selectable markers by PCR method

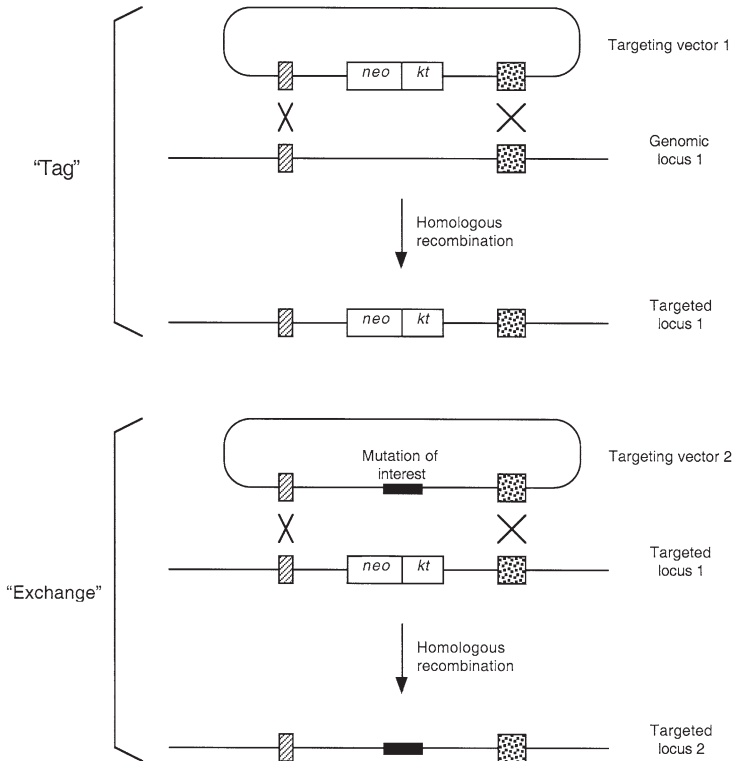


Fig. 20.3. Double-hit replacement

20.3.3 The *Cre/loxP* Recombination

Another versatile strategy to introduce mutations is based on the *Cre/loxP* recombination system. The enzyme, Cre recombinase, recombines DNA at a specific DNA site containing a 34 bp sequence. This *loxP* site has two inverted 13 bp repeats separated by an 8 bp spacer. The enzyme catalyzes recombination resulting in the inversion of the intervening sequence when two *loxP* sites are arranged in opposite orientation. The enzyme also catalyzes excision and recirculation of the intervening sequence when the two *loxP* sites are in the same orientation. In a general scheme, a replacement vector consisting of both positive and negative selectable markers flanked by two *loxP* sites and the desired mutation is inserted into the genomic locus of interest. In the second step, Cre recombinase is introduced to mediate excision of the markers, leaving one *loxP* site in the genome. The resulting clones that contain the introduced mutation can be enriched by negative selection (Fig. 20.4).

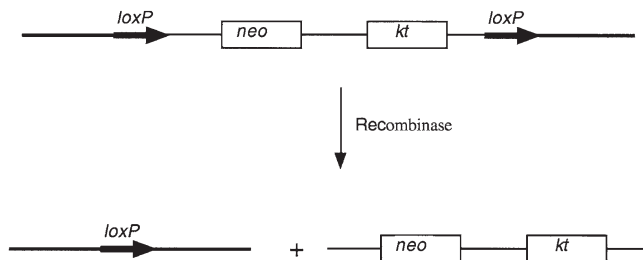


Fig. 20.4 The *Cre/loxP* recombination

20.4 Gene Targeting for Xenotransplants

Transplanting animal organs and tissue into humans (xenotransplantation) has created much promise as a potential solution to deal with the severe shortage of human organs. Pigs are considered one of the better donors of organs, because they can be raised easily and their organs are similar in size and nature to those of humans. The major hurdle of using xenografts, however, is the development of hyperacute rejection and acute vascular rejection, resulting in the destruction of the grafts. The rejection is triggered by the binding of anti-donor antibodies in the recipient patient to the galactose- α 1,3-galactose (α 1,3-Gal), a common carbohydrate moiety on the cell surface glycoproteins of almost all mammals, except humans, apes, and Old World monkeys. Since the key step in the synthesis of the α 1,3-Gal epitope requires the enzyme α 1,3-galactosyltransferase (α 1,3GT), one of the approaches to eliminate the rejections is by knocking out the α 1,3GT gene in the pig (Dai et al. 2002. *Nature Biotechnology* 20, 251–255).

In the approach, a 6.4 kb α 1,3GT genomic segment which expands most of exons 8 and 9 was generated by PCR from genomic DNA purified from porcine fetal fibroblast cells. The coding region of the pig α 1,3GT gene is located in exon 9, and the gene is known to be expressed well in fetal fibroblasts. To create a targeting vector for the knockout of the α 1,3GT gene, a 1.8 kb IRES-*neo*-polyA sequence was inserted into the 5' end of exon 9. The internal ribosome entry site (IRES) functions as the translation initiation site for the *neo* gene (which expresses the neomycin phosphotransferase protein as a G418 resistance marker). The *neo* gene has dual purposes of (1) disrupting the α 1,3GT gene sequence and function, and also (2) providing a convenient screening strategy for positive clones based on G418 resistance (Fig. 20.5).

The vector thus constructed was used to infect cell lines derived from porcine fetal fibroblasts. Homologous recombination resulting in a knock-out α 1,3GT gene was screened by recovering colonies that are resistant to G418. The insertion (knock-out) was further confirmed by PCR. In one of the trans-

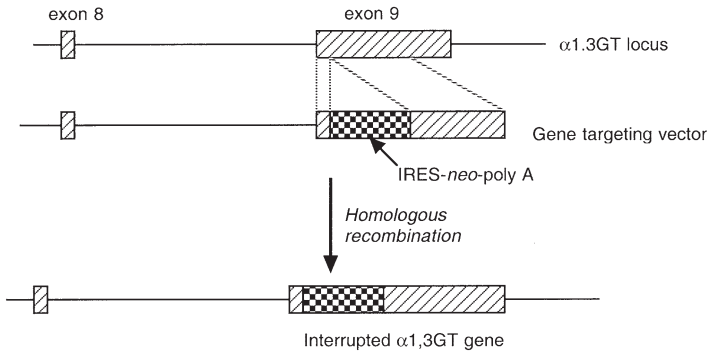


Fig. 20.5. Lockout of the $\alpha 1,3\text{GT}$ gene

ected cell lines, 599 colonies were G418 resistant, 69 were confirmed by 3' PCR, and 18 were confirmed by long-range PCR. The 18 colonies were then subjected to southern blot to yield 14 positive colonies. Seven of the 18 Southern blot-confirmed $\alpha 1,3\text{GT}$ knockout single colonies were used for nuclear transfer experiments to produce 5 female piglets of normal size and weight, all containing one disrupted pig $\alpha 1,3\text{GT}$ allele. Starting with fibroblast cell cultures from such heterozygous animals, cells were selected in which the second allele of the gene was also mutated.

20.5 Engineered Nucleases: ZFN, TALEN, CRISPR

Homologous recombination in ES cell-based gene targeting is rather inefficient. The success has been mostly on yeast and mice models, which seem to have particularly active homologous recombination systems. It remains a challenge for the technique to be applied in a wider range of cells and species. In addition, the method is time-consuming involving laborious vector construction, selection and screening,

Recent advances have introduced strategies for efficiently inducing precise, targeted genome alterations in a broad range of organisms and cell types. Editing plant, animal, or human genome has become a reality due to the ability to engineering precise DNA insertion, deletion, or replacement in the genome using custom-designed "engineered" nucleases. Fundamental to the use of nucleases in genome editing is the key step of inducing site-specific double-stranded DNA breakages (DSB) at desired locations in the genome. These engineered enzymes consist of (1) a DNA-binding domain designed to target a sequence site in the genome and (2) a *FokI* endonuclease domain. The DNA-binding domain is derived from zinc finger transcription factor or

transcription activator-like effector proteins. The corresponding chimeric nucleases (the DNA-binding domain plus the *FokI* cleavage domain) are known as zinc-finger nuclease (ZFN) and transcription activator-like effector nuclease (TALEN), respectively. The new technology based on these nucleases can manipulate the genome DNA in a diverse range of cell types and organisms (Gaj et al. 2013. *Trends Biotechnol.* 31, 397–405; Joung and Sander 2013. *Nat. Rev. Mol. Cell Biol.* 14, 49–55; Nemudry et al. 2014. *Acta Nature* 6, 19–40; Tan et al. 2016. *Transgenic Res.* 25, 273–287).

20.5.1 Zinc-Finger Nucleases

The DNA binding domain consists of zinc fingers, which are eukaryotic transcription factors. Each zinc finger consists of 30 amino acids in a conserved $\beta\beta\alpha$ configuration. Each finger recognizes 3 bases of DNA sequence. The zinc finger also contains conserved Cys and His residues that form complexes with the zinc ion. By mixing and linking several selected zinc fingers together, it is possible to create zinc finger modules to recognize 18 bp (or more) to target a single locus in the human genome with high specificity. The endonuclease (cleavage) domain consists of the *FokI*, which is a type II non-specific restriction enzyme. In standard ZFN molecules, *FokI* is fused to the C-terminus of the zinc finger domain. The *FokI* cleavage domain must dimerize for the catalytic cleavage of DNA. The two individual ZFN molecules bind to opposite strands of the DNA with their C-terminal distance apart by a 5–7 bp space sequence to be recognized by the *FokI* cleavage domain (Fig. 20.6).

20.5.2 Transcription Activator-Like Effector Nucleases

In TALEN, the DNA-binding domain is the transcription activator-like effector protein derived from the plant pathogenic bacteria of the *Xanthomonas* genus. The protein is composed of a series of 33–35 amino acid repeats, each recognizing a single base pair. These repeats can be custom designed and

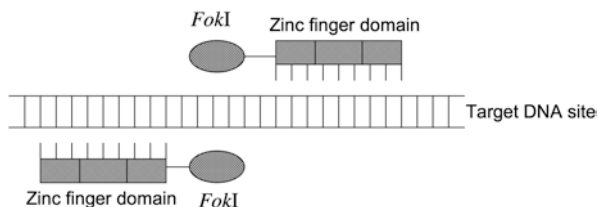


Fig. 20.6. Illustration of ZFN dimer bound to target DNA introduces double-stranded breaks into the site

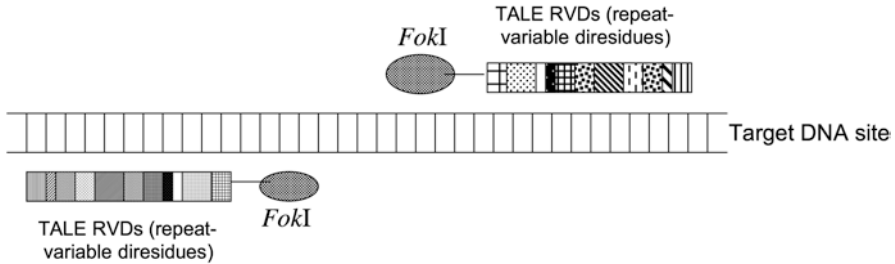


Fig. 20.7. TALEN dimer in complex with target DNA introduces double-stranded breaks into the site

assembled to recognize any sequences in the genome. The *FokI* cleavage domain is fused to the C-terminus of the TALE protein. Similar to ZFN, the TALEN dimer binds to the DNA sites left and right, separated by a spacer sequence of 12–20 bp (Fig. 20.7).

20.5.3 The CRISPR/Cas System

The CRISPR (clustered regularly interspaced short palindromic repeat) Type II system is derived from bacterial immune system, adapted for genome engineering. The system contains of two components: (1) a guide RNA (gRNA), and (2) a non-specific CRISPR-associated endonuclease (Cas9) (Sander and Joung 2014. *Nature Biotechnol.* 32, 347–365; Nemudry et al. 2014. *Acta Nature* 6, 19–40; Barrangou and Doudna 2016. *Nature Biotechnol.* 34, 933–941).

Guide RNA is synthesized modeling after the CRISPR TypeII system RNA hybrids. It contains a scaffold sequence that binds Cas9 and a 20-nucleotide user-defined spacer sequence that binds to a target DNA site in the genome. Cas9 is a non-specific endonuclease that in complex with gRNA, induces dsDNA cleavage at the target DNA site. Essential to the catalytic cleavage is the presence of a conserved motif called protospacer adjacent motifs (PAM) immediately downstream of the target site. For TypeII Cas9, the consensus sequence is 5'-NGG. There is an expected 160×10^6 NGG PAMs in the human genome, and one GG dinucleotides every 42 bases. In the gRNA:cas9 complex at the target DNA site, the PAM sequence is located on the non-complementary strand (the strand containing the same DNA sequences as the gRNA spacer sequence). The gRNA:Cas9 complex, after binding to the target DNA, induces DSB within the target DNA about 3–4 bp upstream of the PAM sequence.

Notice that both ZFN and TALEN involve protein-DNA interactions, and the construction of these nucleases requires protein engineering. The CRISPR-Cas system uses RNA-guided nucleases and depends on base-pairing interactions between an engineered RNA and the target DNA site. The latter is a straight forward and simpler system to work with (Fig. 20.8).

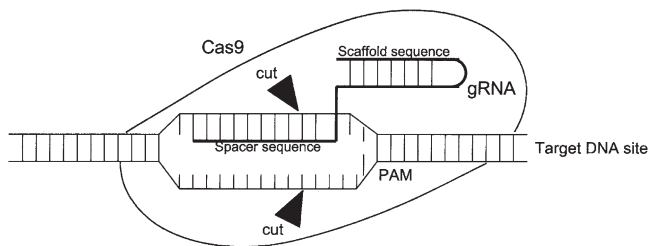


Fig. 20.8. A complex of gRNA and Cas9 introduces double-stranded breaks into DNA site

20.5.4 Nonhomologous End Joining and Homology-Directed Repair

In eukaryotes, after DSB, the cell will naturally repair the cut by joining the two ends of the DNA back together. This repair process, called non-homologous end joining (NHEJ), is error-prone with few bases added or lost around the site of repair, resulting in unintentional mutations. If the insertion/deletion occurs within the open reading frame, it may cause a frameshift and knock out the gene function.

However, another pathway known as homology-directed repair (HDR) is utilized. In this system, a DNA template (also called donor DNA) containing (1) the desired edit (base change) and (2) homologous sequence flanking the DSB location is added to the gRNA/Cas9 system. The DSB is sealed without missing the DNA sequence at the break point. Thus, by manipulate the DNA template with specific edits, HDR can be used to generate precise alterations from a single nucleotide change to large insertions or deletions. Homology-directed repair forms the basis of CRISPR genome editing.

20.5.5 Expressing Engineered Nucleases in Target Cells

In all three nucleases systems, the engineered enzymes need to be cloned as plasmids and delivered to the target cell. Focusing on the CRISPR system, the gRNA and the Cas9 sequence are cloned into a choice plasmid with proper promoters and selection markers, separately or together (i.e. all-in-one vector construct). Examples of Cas9 promoters include CMV (cytomegalovirus immediate early gene) and CBH (chicken β -actin), while U6 promoter is commonly used for gRNA (Ran et al. 2013 *Nature Protocols* 8, 2281–2308). The recombinant vector is delivered into the target cell line by lipofection or electroporation. Alternately, it can be introduced by viral transduction, which has higher efficiency and suitable for hard-to-transfect cell types. (Refer to Chap.

19: Gene Therapy.) However, the latter procedure is harder to perform and more time-consuming.

Review

1. What is homologous recombination? Reciprocal exchange? Nonreciprocal exchange?
2. What is “gene knockout”? What are the primary purposes of conducting such experiment?
3. How does a replacement targeting vector work?
4. What are the advantages and disadvantages of using selectable markers in gene targeting?
5. Describe one approach of gene targeting that does not require the use of selectable markers.
6. Describe one approach of gene targeting that does not retain the selectable marker.
7. In the knockout experiment, the *neo* gene was used to disrupt the $\alpha 1,3GT$ gene. Why was the *neo* gene used for the experiment? Could point mutations be introduced to achieve the same purpose? Explain your answers.
8. Describe the structural functions of the two engineered nucleases, ZFN and TALEN in the regulation and cleavage of DNA.
9. What are the two major components of the CRISPR system? Describe the structural functions of these components.
10. What is the major advantage of CRISPR in comparison to ZFN and TALEN?