

Claims

1. A method for generating a fusion protein of DNA recombinases for genome editing, preferably for recombination, more preferably for inversion of DNA sequences on genomic level in a cell, wherein said method comprises the steps of:
- i. identifying a nucleic acid sequence that is a potential target site for DNA-recombining enzymes that are capable to induce a site-specific DNA recombination of a sequence of interest in a genome, wherein said potential target site comprises two asymmetric recombinase target sites;
 - 10 ii. providing a nucleic acid molecule encoding a first recombinase enzyme, wherein said first recombinase enzyme has been evolved by directed evolution or rational design to specifically recognize a first half-site of a recombinase target site;
 - 15 iii. providing a nucleic acid molecule encoding a second recombinase enzyme, wherein said second recombinase enzyme has been evolved by directed evolution or rational design to specifically recognize a second half-site of a recombinase target site;
 - iv. providing a nucleic acid molecule encoding a linker peptide, wherein said nucleic acid encodes a linker oligopeptide comprising 6 to 30 amino acids;
 - v. creating an expression vector by cloning the nucleic acid molecule encoding a first recombinase enzyme, the nucleic acid molecule encoding a second recombinase enzyme and the nucleic acid molecule encoding a linker peptide into an expression vector;
 - 20 vi. transfecting a cell, which comprises a DNA sequence, which is to be recombined, preferably inverted, with the expression vector of step v);
 - 25 vii. expressing the fusion protein comprising a first recombinase enzyme, a second recombinase enzyme and a linker peptide;
 - viii. analyzing, whether the fusion protein expressed in step vii. is capable of inverting the DNA sequence on a human chromosome in said cell; and
 - 30 ix. selecting a fusion protein which is capable of inverting of DNA sequence on a human chromosome in said cell according to step viii.
2. The method of claim 1, further comprising the steps of:
- x. Identifying potential off-target sites of the desired fusion protein;

- xi. Analyzing the recombinase activity of the desired fusion protein on the off-target sites identified in step x.; and
 - xii. Selecting fusion proteins that do not show recombinase activity on at least one off-target site.
- 5 3. The method of claim 1 or 2, wherein in the fusion protein expressed in step vii., the C-terminus of the said first recombinase enzyme is connected to the N-terminus of said second recombinase enzyme via the linker peptide; or wherein the C-terminus of the said second recombinase enzyme is connected to the N-terminus of said first recombinase enzyme via the linker peptide.
- 10 4. The method according to any one of claims 1 to 3, wherein step i. of the method of claim 1 comprises the substeps of:
 - a) Screening the genome or a part thereof comprising the sequence of interest for two sequences that are potential spacer sequences, with a length of at least 5 and up to 12 bp, wherein one of the potential spacer sequence lies upstream of the sequence of
15 interest and the other potential spacer sequence lies downstream of the sequence of interest and wherein the two sequences have a maximum distance of 2 megabases and a minimum distance of 150 bp,
 - b) Identifying potential target sites by determining for each potential spacer sequence the neighboring nucleotides, preferably 10 to 20 nucleotides, more preferably 12 to 15
20 nucleotides, most preferably 13 nucleotides on one side thereof form the potential first half site and the neighboring nucleotides, preferably 10 to 20 nucleotides, more preferably 12 to 15 nucleotides, most preferably 13 nucleotides, on the other side form the potential second half site, whereas both potential half sites and the spacer sequence in between form a potential target site,
 - 25 c) The potential target sites identified in step b) are further screened to select for potential target sequences that do not occur (elsewhere) in the genome of the host to ensure a sequence specific recombination, preferably inversion.
- 5. The method according to any one of claims 1 to 4, wherein first recombinase enzyme and the second recombinase enzyme according to steps ii. and iii. are evolved by substrate linked
30 directed evolution (SLiDE), wherein said substrate linked directed evolution comprises the steps of:

5 a) Selecting a nucleotide sequence upstream of the nucleotide sequence to be altered as first target site and a nucleotide sequence downstream of the nucleotide sequence to be altered as second target site, whereas the sequences of the target sites are preferably not identical, wherein each target site comprises a first half site and a second half site with each 10 to 20 nucleotides separated by a spacer sequence with 5 to 12 nucleotides,

b) Applying molecular directed evolution on at least one library of DNA-recombining enzymes using a vector comprising the first target site and the second target site as selected in a) as substrate,
until at least one first designer DNA-recombining enzyme is obtained that is active on the first
10 target site and at least one second designer DNA-recombining enzyme is obtained that is active on the second target site as selected in a).

6. A fusion protein for genome editing, which has been obtained by the method according to any one of claims 1 to 5, wherein said fusion protein consists of a complex of recombinases,
15 wherein said complex consists of at least a first recombinase enzyme, a second recombinase enzyme and at least one linker, wherein said first recombinase enzyme and said second recombinase enzyme specifically recognize a first half-site and a second half-site of an upstream target site and/or a downstream target site of a recombinase; wherein said first recombinase enzyme and said second recombinase enzyme are interconnected via a linker; and
20 wherein said linker comprises or consists of an oligopeptide comprising 4 to 50 amino acids; wherein said fusion protein specifically recognizes the upstream recombinase target sequence of the *loxP8* target site, which has the nucleic acid sequence

ATAAATCTGTGGAAACGCTGCCACACAATCTTAG (SEQ ID NO: 17) or a
25 reverse complement sequence thereof;

and recognizes the downstream recombinase target sequence of the *loxP8* target site, which has the nucleic acid sequence

30 CTAAGATTGTGTGGCAGCGTTTCCACAGATTAT (SEQ ID NO: 18) or a reverse complement sequence thereof;

and which catalyzes the inversion of a gene sequence between the upstream recombinase target sequence of SEQ ID NO. 17 and the downstream recombinase target sequence of SEQ ID NO: 18 of the *loxP8* recombinase target site; and

wherein the capability to catalyzes the inversion of a gene sequence between the upstream recombinase target sequence of SEQ ID NO. 17 and the downstream recombinase target sequence of SEQ ID NO: 18 of the *loxP8* recombinase target site is tested by a method comprising the steps of:

- a) expressing the fusion protein comprising a first recombinase enzyme, a second recombinase enzyme and a linker peptide; and
- b) analyzing, whether the fusion protein expressed in step a) is capable of inverting of DNA sequence on a human chromosome in said cell.

7. The fusion protein of claim 6, wherein said linker

- comprises or consists of an oligopeptide with an amino acid sequence selected from formula 1, formula 2 and formula 3:

$X_1-X_2-X_3-X_4-X_5-X_6-(G_2S)_4-X_7-X_8-X_9-X_{10}-X_{11}-X_{12}$ (formula 1);

$(G_2S)_2-X_1-X_2-X_3-X_4-X_5-X_6-X_7-X_8-X_9-X_{10}-X_{11}-X_{12}-(G_2S)_2$ (formula 2); and

$X_1-X_2-X_3-X_4-X_5-X_6-X_7-X_8-X_9-X_{10}-X_{11}-X_{12}-X_{13}-X_{14}-X_{15}-X_{16}-X_{17}-X_{18}-X_{19}-X_{20}-X_{21}-X_{22}-X_{23}-X_{24}$ (formula 3);

wherein

25

G is glycine;

S is serine; and

each of X_1 to X_{24} is independently selected from the group consisting of alanine, arginine, asparagine, aspartic acid, glutamine, glycine, lysine, serine and threonine;

wherein said oligopeptide of formula 1, formula 2 or formula 3 does not consist of glycine and serine residues only.

- or comprises or consists of an oligopeptide selected from the group consisting of

	AEATSEGGSGGSGGSGGSGNGARRT	(SEQ ID NO: 8);
	AGTTARGGSGGSGGSGGSGRRGAK	(SEQ ID NO: 9);
	KNGRGRGGSGGSGGSGGSRKRET	(SEQ ID NO: 10);
5	GGSGGSTAAKEGAASSASGGSGGS	(SEQ ID NO: 11);
	GGSGGSNSRSNTENSDKGGGSGGS	(SEQ ID NO: 12);
	GGSGGSNGEEGTERGKATGGSGGS	(SEQ ID NO: 13);
	GGSGGSTTKANRAKGGRGGGSGGS	(SEQ ID NO: 14);
	GANEDTNTEAAGSEGNEKTGTNSA	(SEQ ID NO: 15); and
10	GESRAEDGAKGNRGKGEATAGAA	(SEQ ID NO: 16).

8. The fusion protein according to claim 6 or 7, wherein in said fusion protein, the first recombinase enzyme, the second recombinase enzyme and the linker are interconnected such that the C-terminus of the second recombinase enzyme, which specifically recognizes a second half-site of a recombinase target site, is fused with the linker to the N-terminus of the first recombinase enzyme, which specifically recognizes a first half-site of a recombinase target site.;

or wherein in said fusion protein, the first recombinase enzyme, the second recombinase enzyme and the linker are interconnected such that the C-terminus of the first recombinase enzyme, which specifically recognizes a first half-site of a recombinase target site, is fused with the linker to the N-terminus of the second recombinase enzyme, which specifically recognizes a second half-site of a recombinase target site

9. The fusion protein according to any one of claims 6 to 8, wherein each recombinase enzyme of said heterodimer is a tyrosine site-specific recombinase, which has been evolved by directed evolution to independently recognize a first half-site and a second-half site of a recombinase target site of tyrosine site-specific recombinases.

10. The fusion protein according to claim 9, wherein said recombinase target site is a target site of a tyrosine site-specific recombinase is selected from the group consisting of Cre-, Dre-, VCre-, SCre-, Vika-, lambda-Int-, Flp-, R-, Kw-, Kd-, B2-, B3-, Nigri- and Panto-recombinase.

11. The fusion protein according to any one of claims 6 to 10, wherein said first recombinase enzyme is a protein, which has an amino acid sequence with at least 70 %, preferably 80 %, more preferably 90 %, sequence identity with a sequence according to SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 93 or SEQ ID NO: 99 and/or wherein said second recombinase enzyme
5 is a protein, which has an amino acid sequence with at least 70 %, preferably 80 %, more preferably 90 %, sequence identity with a sequence according to SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 94 or SEQ ID NO: 100;

or wherein said fusion protein has an amino acid sequence with at least 70 %, preferably 80 %, more preferably 90 %, sequence identity with a sequence according to SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 97 or SEQ ID NO: 103, wherein SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 97 and SEQ ID NO: 103 each include a linker with the amino acid sequence of SEQ ID NO: 14.
10

12. A fusion protein consisting of a first recombinase enzyme and a second recombinase enzyme, wherein said first recombinase enzyme and said second recombinase enzyme specifically recognize a first half-site and a second half-site of an upstream target site and/or a downstream target site of a recombinase, wherein said first recombinase enzyme is a polypeptide, which has an amino acid sequence with at least 70 %, preferably 80 %, more preferably 90 %, sequence identity with a sequence according to SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 93 or SEQ ID NO: 99 and/or wherein said second recombinase enzyme is a polypeptide, which has an amino acid sequence with at least 70 %, preferably 80 %, more preferably 90 %, sequence identity with a sequence according to SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 94 or SEQ ID NO: 100; wherein DNA recombinase heterodimer specifically recognizes the upstream recombinase target sequence of the *loxP8* target site, which has the nucleic acid
20
25 sequence

ATAAATCTGTGGAAACGCTGCCACACAATCTTAG (SEQ ID NO: 17), or a reverse complement sequence thereof;

30 and recognizes the downstream recombinase target sequence of the *loxP8* target site, which has the nucleic acid sequence

CTAAGATTGTGTGGCAGCGTTTCCACAGATTTAT (SEQ ID NO: 18), or a reverse complement sequence thereof;

and which catalyzes the inversion of a gene sequence between the upstream recombinase target sequence of SEQ ID NO. 17 and the downstream recombinase target sequence of SEQ ID NO: 18 of the *loxF8* recombinase target site;

wherein the capability to catalyzes the inversion of a gene sequence between the upstream recombinase target sequence of SEQ ID NO. 17 and the downstream recombinase target sequence of SEQ ID NO: 18 of the *loxF8* recombinase target site is tested by a method comprising the steps of:

- a) expressing the fusion protein comprising a first recombinase enzyme, a second recombinase enzyme and a linker peptide;
- b) analyzing, whether the fusion protein expressed in step a) is capable of inverting of DNA sequence on a human chromosome in said cell.

13. A nucleic acid molecule encoding a fusion protein according to any one of claims 6 to 12.

14. The nucleic acid molecule of claim 13, which comprises or consists of

- i. nucleic acids with the sequence of SEQ ID NOs 34, 35 and 61; such as the nucleic acid of SEQ ID NO 64; or
- ii. nucleic acids with the sequence of SEQ ID NOs 36, 37 and 61; such as the nucleic acid of SEQ ID NO 65; or
- iii. nucleic acids with the sequence of SEQ ID NOs 95, 96 and 61; such as the nucleic acid of SEQ ID NO 98; or
- iv. nucleic acids with the sequence of SEQ ID NOs 101, 102 and 61; such as the nucleic acid of SEQ ID NO 104; or
- v. nucleic acids with the sequence of SEQ ID NOs 95 and 96; or
- vi. nucleic acids with the sequence of SEQ ID NOs 101 and 102;
- vii. nucleic acids with the sequence of SEQ ID NOs 34 and 35; or
- viii. nucleic acids with the sequence of SEQ ID NOs 36 and 37.

15. A polynucleotide molecule comprising the nucleic acid molecule according to claim 13 or 14 plus expression-controlling elements operably linked with said nucleic acid to drive expression thereof.

16. A mammalian, insect, plant or bacterial host cell comprising a nucleic acid molecule according to claim 13 or 14 or a recombinant polynucleotide molecule of claim 15 or an expression vector comprising the nucleic acid molecule of claim 13 or 14.
17. A pharmaceutical composition comprising a fusion protein according to any one of claims
5 6 to 12 or a nucleic acid molecule or recombinant polynucleotide according to any one of claims 13 to 15 optionally in combination with one or more therapeutically acceptable diluents or carriers.
18. A fusion protein according to any one of claims 1 to 12 or a nucleic acid molecule or recombinant polynucleotide according to any one of claims 13 to 15 or a pharmaceutical
10 composition according to claim 17 for use in the treatment of hemophilia A, in particular severe hemophilia A.
19. A method for inversion of a DNA sequence on genomic level in a cell *in vitro*, comprising a fusion protein for genome editing according to the invention, wherein said method comprises
15 the steps of:
- i. providing a nucleic acid molecule encoding a first recombinase enzyme, wherein said first recombinase enzyme has been evolved by directed evolution or rational design to specifically recognize a first half-site of a recombinase target site;
 - ii. providing a nucleic acid molecule encoding a second recombinase enzyme,
20 wherein said second recombinase enzyme has been evolved by directed evolution or rational design to specifically recognize a second half-site of a recombinase target site;
 - iii. providing a nucleic acid molecule encoding a linker peptide, wherein said nucleic acid encodes a linker oligopeptide comprising 6 to 30 amino acids;
 - 25 iv. creating an expression vector by cloning the nucleic acid molecule encoding a first recombinase enzyme, the nucleic acid molecule encoding a second recombinase enzyme and the nucleic acid molecule encoding a linker peptide into an expression vector;
 - v. delivering to a cell, which comprises a DNA sequence, which is to be inverted,
30 an expression vector of step iv), a RNA molecule encoding a fusion protein according to any one of claims 6 to 12 or a fusion protein according to any one of claims 6 to 12;
 - vi. inversion of the DNA sequence on a human chromosome in said cell.

20. The method of claim 19, wherein said method further comprises the step
- v. a) expressing the fusion protein comprising a first recombinase enzyme, a second recombinase enzyme and a linker peptide;
- 5 when in step v. of the method of claim 19 an expression vector of step iv) or a RNA molecule encoding a fusion protein according to any one of claims 6 to 12 is delivered to said cell, e.g. by transfection.
21. A *loxF8* recombinase target site comprising a 5' target sequence of SEQ ID NO: 17 and a
- 10 3' target sequence of SEQ ID NO: 18 or a reverse complement sequence thereof.