Claims

1. Method for preparing an expression vector encoding a tailored recombinase, which tailored recombinase is capable of recombining asymmetric target sequences within the LTR of proviral DNA of a plurality of retrovirus strains, comprising the steps of
   (a) identifying sequences with a homology of at least 30 % to the left half-site sequence and the right half-site sequence of at least one known recombinase target site in the sequence of the LTR of proviral DNA of a plurality of retrovirus strains, wherein the homologous sequences are separated by a spacer of 5-12 nucleotides, and wherein the asymmetric target sequence is found in a plurality of retrovirus strains;
   (b) identifying two sequences, wherein the first sequence corresponds to the sequence of the asymmetric target sequence of step (a) homologous to the left half-site of said known target site and is referred to as "half-site sequence 1", and wherein the second sequence corresponds to the sequence of the asymmetric target sequence of step (a) homologous to the right half-site and is referred to as "half-site sequence 2";
   (c) determining the nucleotides within the sequences of step (b) deviating from the corresponding homologous left half-site and right half-site sequences of the at least one known homologous target site of step (a);
   (d) generating a first subset of two target nucleic acids comprising target sequences, wherein the first target sequence is designated subsite 1 and comprises, adjacent to each other and in 5' to 3' order, half-site sequence 1 of step (b), the spacer sequence of the asymmetric target sequence and an inverted repeat of half-site sequence 1, and wherein the second target sequence is designated subsite 2 and comprises, adjacent to each other and in 5' to 3' order, an inverted repeat of half-site sequence 2, the spacer sequence of the asymmetric target sequence and half-site sequence 2 of step (b) ;
   (e) generating a second subset of target nucleic acids comprising modified target sequences on the basis of the target sequences in the first subset of step (d),
   wherein, in sequences based on subsite 1, in the left half-site sequence, a portion of the nucleotides deviating from the corresponding homologous half-site sequence of the at least one known target-site of step (a) is replaced by the native nucleotides found in said known target-site, until said half-site sequence contains one, two or three nucleotides deviating from said known target site, wherein the right half-site of said modified target sequence is formed by an inverted repeat of said modified left

half-site sequence, which is separated from said modified left half-site sequence by the spacer sequence of the asymmetric target sequence, and

wherein, in sequences based on subsite 2, in the right half-site sequence, a portion of the nucleotides deviating from the corresponding homologous half-site sequence of the at least one known target-site of step (a) is replaced by the native nucleotides found in said known target-site, until said half-site sequence contains one, two or three nucleotides deviating from said known target site, wherein the left half-site of said modified target sequence is formed by an inverted repeat of said modified right half-site sequence, which is separated from said modified right half-site sequence by the spacer sequence of the asymmetric target sequence,

such that in all modified half-site sequences originating from one target sequence of the first subset of step (d) taken together, all deviating nucleotides can be found, whereas none of said modified half-site sequences alone comprises all deviating nucleotides,

(f) separately applying molecular directed evolution on at least one recombinase recognising a known homologous target site according to step (a) using each nucleic acid of the second subset obtained in step (e) as a substrate;

(g) shuffling the recombinase libraries evolved in step (f), wherein all recombinase libraries evolved on sequences based on subsite 1 are combined and shuffled, and wherein all recombinase libraries evolved on sequences based on subsite 2 are combined and shuffled;

(h) applying molecular directed evolution on the shuffled libraries obtained in step (g) using each nucleic acid of the subset according to step (d) as a substrate;

(i) shuffling the recombinase libraries evolved in step (h);

(j) applying molecular directed evolution on the shuffled library obtained in step (g) using a nucleic acid comprising the asymmetric target sequence of step (a) as a substrate, until at least one recombinase is obtained that is active on the asymmetric target sequence within the LTR of the retrovirus DNA of step (a);

(k) isolating the nucleic acid encoding the least one recombinase obtained in step (j) from the library; and

(l) cloning the nucleic acid obtained in step (k) into a suitable expression vector.

2. Method according to claim 1, wherein the retrovirus is HIV, preferably HIV-1.

3. Method according to claim 2, wherein the retrovirus is HIV-1 and the asymmetric target sequence has the sequence set forth as SEQ ID NO:1 or SEQ ID NO:2.

4. Method according to any one of claims 1 to 3, wherein the at least one known recombinase whose target sequence is used in step (a) and upon which molecular directed evolution is applied in step (f) is part of a recombinase library.

5. Method according to any one of claims 1 to 4, wherein the molecular directed evolution employed is substrate-linked protein evolution.

6. Method according to any one of claims 1 to 5, wherein the expression vector in step (l) is selected from the group consisting of retroviral vectors, lentiviral vectors, spumavirus vectors and adenoviral vectors.

7. Method for preparing a tailored recombinase, wherein said method comprises the method according to any one of claims 1 to 6 and the further step of expressing the tailored recombinase from the nucleic acid encoding the recombinase inserted into the expression vector in a suitable host cell, wherein the recombinase is optionally expressed as a fusion polypeptide comprising the amino acid sequence of the tailored recombinase.

8. Method for preparing a transformed cell, wherein said method comprises the method according to any one of claims 1 to 6 and the further step of introducing the expression vector *in vitro* into a cell, wherein the cell preferably is an adult stem cell.

9. Method according to any one of claims 1 to 8, further comprising the step of preparing the expression vector, the recombinase or cell as a pharmaceutical composition.

10. Nucleic acid encoding a tailored recombinase, which tailored recombinase is capable of recombining asymmetric target sequences within the LTR of proviral DNA of a plurality of retrovirus strains, wherein the amino acid sequence of the tailored recombinase has at least 97% sequence identity to a sequence according to SEQ ID NO: 38, wherein the nucleic acid optionally is an expression vector.

11. Nucleic acid as obtainable from step (k) of the method according to any one of claims 1 to 6 or expression vector as obtainable from the method according to any one of claims 1 to 6, wherein the amino acid sequence of the tailored recombinase optionally has at least 97% sequence identity to a sequence according to SEQ ID NO: 38 and/or wherein the tai-

lored recombinase encoded by the nucleic acid optionally comprises an amino acid sequence according to SEQ ID NO: 37.

12. Nucleic acid according to any of claims 10 or 11, wherein the tailored recombinase comprises an amino acid sequence according to SEQ ID NO: 37, wherein said tailored recombinase comprises at least one, preferably, 2, 3, 4, 5, 6, or 7 of the defined amino acid exchanges as compared to the Cre sequence (SEQ ID NO: 36) selected from the group consisting of P12S, P15L, M44V, K86N, G93A, A175S, and P307A.

13. Nucleic acid according to any of claims 10 to 12, wherein the amino acid sequence of the tailored recombinase comprises the sequence according to SEQ ID NO: 38, wherein the amino acid sequence of the tailored recombinase optionally comprises the sequence according to SEQ ID NO: 39, and/or wherein the amino acid sequence of the tailored recombinase is optionally selected from the group consisting of SEQ ID NO: 40-64 and a combination of sequences from the group consisting of SEQ ID NO: 40-64.

14. Tailored recombinase encoded by the nucleic acid of any of claims 10 to 13, which is optionally expressed as a fusion protein.

15. Transformed cell comprising a nucleic acid of any of claims 10 to 13, wherein said cell preferably is a stem cell from the hematopoietic lineage.

16. Pharmaceutical composition comprising a nucleic acid of any of claims 10 to 13, a tailored recombinase of claim 14 and/or a transformed cell according to claim 15.

17. Pharmaceutical composition according to claim 16, wherein the pharmaceutical composition is for use in treatment or prevention of retrovirus infection in a subject, wherein the retrovirus preferably is HIV, in particular HIV-1, wherein the pharmaceutical composition is optionally formulated for administration to a subject, if proviral DNA found in a sample obtained from the subject comprises the asymmetric target sequence identified in step (a) on which the recombinase has been selected.

CLAIMS:

1. Method for preparing an expression vector encoding a tailored recombinase, which tailored recombinase is capable of recombining asymmetric target sequences within the LTR of proviral DNA of a plurality of retrovirus strains which are HIV-1 strains, comprising the steps of

(a) identifying sequences with a homology of at least 30 % to the left half-site sequence and the right half-site sequence of at least one known recombinase target site in the sequence of the LTR of proviral DNA of a plurality of retrovirus strains, wherein the homologous sequences are separated by a spacer of 5-12 nucleotides, and wherein the asymmetric target sequence is found in a plurality of retrovirus strains, wherein the asymmetric target sequence has the sequence set forth as SEQ ID NO:1;

(b) identifying two sequences, wherein the first sequence corresponds to the sequence of the asymmetric target sequence of step (a) homologous to the left half-site of said known target site and is referred to as "half-site sequence 1", and wherein the second sequence corresponds to the sequence of the asymmetric target sequence of step (a) homologous to the right half-site and is referred to as "half-site sequence 2";

(c) determining the nucleotides within the sequences of step (b) deviating from the corresponding homologous left half-site and right half-site sequences of the at least one known homologous target site of step (a);

(d) generating a first subset of two target nucleic acids comprising target sequences, wherein the first target sequence is designated subsite 1 and comprises, adjacent to each other and in 5' to 3' order, half-site sequence 1 of step (b), the spacer sequence of the asymmetric target sequence and an inverted repeat of half-site sequence 1 , and wherein the second target sequence is designated subsite 2 and comprises, adjacent to each other and in 5' to 3' order, an inverted repeat of half-site sequence 2, the spacer sequence of the asymmetric target sequence and half-site sequence 2 of step (b) ;

(e) generating a second subset of target nucleic acids comprising modified target sequences on the basis of the target sequences in the first subset of step (d),

wherein, in sequences based on subsite 1 , in the left half-site sequence, a portion of the nucleotides deviating from the corresponding homologous half-site sequence of the at least one known target-site of step (a) is replaced by the native nucleotides found in said known target-site, until said half-site sequence contains one, two or three nucleotides deviating from said known target site, wherein the right half-site of said modified target sequence is formed by an inverted repeat of said modified left half-site sequence, which is separated from said modified left half-site sequence by the spacer sequence of the asymmetric target sequence, and

wherein, in sequences based on subsite 2, in the right half-site sequence, a portion of the nucleotides deviating from the corresponding homologous half-site sequence of the at least one known target-site of step (a) is replaced by the native nucleotides found in said known target-site, until said half-site sequence contains one, two or three nucleotides deviating from said known target site, wherein the left half-site of said modified target sequence is formed by an inverted repeat of said modified right half-site sequence, which is separated from said modified right half-site sequence by the spacer sequence of the asymmetric target sequence,

such that in all modified half-site sequences originating from one target sequence of the first subset of step (d) taken together, all deviating nucleotides can be found, whereas none of said modified half-site sequences alone comprises all deviating nucleotides,

(f) separately applying molecular directed evolution on at least one recombinase recognising a known homologous target site according to step (a) using each nucleic acid of the second subset obtained in step (e) as a substrate;

(g) shuffling the recombinase libraries evolved in step (f), wherein all recombinase libraries evolved on sequences based on subsite 1 are combined and shuffled, and wherein all recombinase libraries evolved on sequences based on subsite 2 are combined and shuffled;

(h) applying molecular directed evolution on the shuffled libraries obtained in step (g) using each nucleic acid of the subset according to step (d) as a substrate;

(i) shuffling the recombinase libraries evolved in step (h);

(j) applying substrate-linked protein evolution on the shuffled library obtained in step (g) using a nucleic acid comprising the asymmetric target sequence of step (a) as a substrate, until at least one recombinase is obtained that is active on the asymmetric target sequence within the LTR of the retrovirus DNA of step (a);

(k) isolating the nucleic acid encoding the least one recombinase obtained in step (j) from the library; and

(l) cloning the nucleic acid obtained in step (k) into a suitable expression vector.

2. Method according to claim 1, wherein the at least one known recombinase whose target sequence is used in step (a) and upon which molecular directed evolution is applied in step (f) is part of a recombinase library.

3. Method according to any one of claims 1 or 2, wherein the expression vector in step (1) is selected from the group consisting of retroviral vectors, lentiviral vectors, spumavirus vectors and adenoviral vectors.

4. Method for preparing a tailored recombinase, wherein said method comprises the method according to any one of claims 1 to 3 and the further step of expressing the tailored recombinase from the nucleic acid encoding the recombinase inserted into the expression vector in a suitable host cell, wherein the recombinase is optionally expressed as a fusion polypeptide comprising the amino acid sequence of the tailored recombinase.

5. Method for preparing a transformed cell, wherein said method comprises the method according to any one of claims 1 to 3 and the further step of introducing the expression vector in vitro into a cell, wherein the cell preferably is an adult stem cell.

6. Method according to any one of claims 1 to 5, further comprising the step of preparing the expression vector, the recombinase or cell as a pharmaceutical composition.

7. Nucleic acid encoding a tailored recombinase, which tailored recombinase is capable of recombining asymmetric target sequences comprising SEQ ID NO:1 within the LTR of proviral DNA of a plurality of HIV-1 strains, wherein the amino acid sequence of the tailored recombinase has at least 99% sequence identity to a sequence according to SEQ ID NO: 38, wherein said tailored recombinase comprises all amino acid exchanges as compared to the Cre sequence (SEQ ID NO: 36) of P12S, P15L, M44V, K86N, G93A, A175S, and P307A, wherein the nucleic acid optionally is an expression vector.

8. Nucleic acid of claim 7, as obtainable from step (k) of the method according to any one of claims 1 to 3 or expression vector as obtainable from the method according to any one of claims 1 to 3.

9. Nucleic acid according to any one of claims 7 to 8, wherein the amino acid sequence of the tailored recombinase comprises the sequence according to SEQ ID NO: 38.

10. Nucleic acid according to claim 9, wherein the amino acid sequence of the tailored recombinase comprises the sequence according to SEQ ID NO: 39.

11. Nucleic acid according to claim 10, wherein the amino acid sequence of the tailored recombinase is selected from the group consisting of SEQ ID NO: 40-64 and a combination of sequences from the group consisting of SEQ ID NO: 40-64.

12. Tailored recombinase encoded by the nucleic acid of any one of claims 7 to 11, which is optionally expressed as a fusion protein.

13. Transformed cell comprising a nucleic acid of any one of claims 7 to 11, wherein said cell preferably is a stem cell from the hematopoietic lineage.

14. Pharmaceutical composition comprising a nucleic acid of any one of claims 7 to 11, a tailored recombinase of claim 12 and/or a transformed cell according to claim 13.

15. Pharmaceutical composition according to claim 14, wherein the pharmaceutical composition is for use in treatment or prevention of HIV-1 infection in a subject.

16. Pharmaceutical composition according to claim 15, wherein the pharmaceutical composition is formulated for administration to a subject, if proviral DNA found in a sample obtained from the subject comprises the asymmetric target sequence of SEQ ID NO:1 on which the recombinase has been selected.

17. Expression vector when prepared by the method of claim 1.

18. Tailored recombinase when prepared by the method of claim 4.

19. Transformed cell when prepared by the method of claim 5.