

RECOMBINASE MUTANTS

BACKGROUND

5 Recombinase enzymes are useful in recombinase-mediated amplification of nucleic acids. For example, recombinase enzymes can facilitate targeting of oligonucleotides to DNA targets allow replication of DNA by a polymerase. There remains a need for modified recombinases with improved properties.

SEQUENCE LISTING

10 The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is provided as a file entitled IP1264.TXT, created September 26, 2014, which is 98Kb in size. The information in the electronic format of the Sequence Listing is incorporated herein by reference in its entirety.

BRIEF SUMMARY

15 Presented herein are recombinases for improved recombinase-mediated amplification of nucleic acids. The present inventors have surprisingly identified certain altered recombinases which have substantially improved characteristics in the seeding nucleic acids onto a patterned flow cell surface. In certain embodiments, the
20 altered recombinases of improve seeding a PCR-free library, such as a PCR-library having single-stranded adapter regions, on a patterned flow cell surface for improved cluster amplification.

In certain embodiments, the recombinase is a recombinant UvsX and comprises an amino acid substitution mutation at the position functionally equivalent to Pro256 in
25 the RB49 UvsX amino acid sequence. The wild type RB49 UvsX amino acid sequence is set forth in SEQ ID NO: 1. In certain embodiments, the recombinant UvsX comprises an amino acid sequence which comprises an amino acid that is at least 60%, 70%, 80%, 90%, 95%, 99% identical to SEQ ID NO: 1, and comprises an amino acid substitution mutation at the position functionally equivalent to Pro256 in the RB49
30 UvsX amino acid sequence. In certain embodiments, the substitution mutation

comprises a mutation to a charged residue. In certain embodiments, the substitution mutation comprises a mutation to a basic residue. In certain embodiments, the substitution mutation comprises a mutation homologous to Pro256Lys in the RB49 UvsX amino acid sequence.

5 In some embodiments, in addition to the above mutations, the recombinant UvsX can further comprise substitution mutations at positions functionally equivalent to His63 in the RB49 UvsX amino acid sequence. For example, in certain embodiments, the recombinant UvsX comprises a substitution mutation homologous to His63Ser in the RB49 UvsX amino acid sequence.

10 In some embodiments, in addition to any of the above mutations, the recombinant UvsX can further comprise a mutation selected from the group consisting of: the addition of one or more glutamic acid residues at the C-terminus; the addition of one or more aspartic acid residues at the C-terminus; and a combination thereof.

In some embodiments, the recombinant UvsX is derived from a myoviridae
15 phage selected from the group consisting of: T4, T6, Rb69, Aeh1, KVP40, Acinetobacter phage 133, Aeromonas phage 65, cyanophage P-SSM2, cyanophage PSSM4, cyanophage S-PM2, Rb32, Vibrio phage nt-1, Rb16, Rb43, and Rb49.

In some embodiments, the recombinant UvsX is derived from a myoviridae phage selected from the group consisting of: T2, Rb14, Aeromonas phage 25, phi-1,
20 Phage 31, phage 44RR2.8t, phage Rb3, and phage LZ2.

Also presented herein is a recombinant UvsX comprising the amino acid sequence of any one of SEQ ID NOs: 2 and 22-35. In certain embodiments, the recombinant UvsX comprises an amino acid sequence which comprises an amino acid that is at least 60%, 70%, 80%, 90%, 95%, 99% identical to any one of SEQ ID NOs: 2
25 and 22-35 and which comprises an amino acid substitution mutation at the position functionally equivalent to Pro256 in the RB49 UvsX amino acid sequence.

Also presented herein is a recombinant UvsX comprising a substitution mutation to the semi-conserved domain comprising the amino acid sequence of any of SEQ ID NOs: 3-5 wherein the substitution mutation comprises a mutation selected from a
30 substitution at position 7 to any residue other than Phe, Pro, Asp, Glu or Asn. In certain

embodiments, the recombinant UvsX comprises an amino acid that is at least 60%, 70%, 80%, 90%, 95%, 99% identical to a recombinase that comprises the semi-conserved domain comprising the amino acid sequence of any of SEQ ID NOs: 3-5, and wherein the recombinant UvsX comprises a substitution mutation selected from a substitution at position 7 to any residue other than Phe, Pro, Asp, Glu or Asn. In certain 5 embodiments, the mutation comprises a mutation to a charged residue. In certain embodiments, the mutation comprises a mutation to a basic residue. In certain embodiments, the mutation comprises a substitution at position 7 to Lys.

Also presented herein is a recombinant UvsX comprising a substitution mutation 10 to the semi-conserved domain comprising the amino acid sequence of any of SEQ ID NOs: 6-7 wherein the substitution mutation comprises a mutation selected from a substitution at position 12 to any residue other than Phe, Pro, Asp, Glu or Asn. In certain embodiments, the recombinant UvsX comprises an amino acid that is at least 60%, 70%, 80%, 90%, 95%, 99% identical to a recombinase that comprises the semi- 15 conserved domain comprising the amino acid sequence of any of SEQ ID NOs: 6-7, and wherein the recombinant UvsX comprises a substitution mutation selected from a substitution at position 12 to any residue other than Phe, Pro, Asp, Glu or Asn. In certain embodiments, the mutation comprises a mutation to a charged residue. In certain embodiments, the mutation comprises a mutation to a basic residue. In certain 20 embodiments, the mutation comprises a substitution at position 12 to Lys.

In some embodiments, in addition to the above mutations, the recombinant UvsX can further comprise substitution mutations at positions functionally equivalent to His63 in the RB49 UvsX amino acid sequence. For example, in certain embodiments, the recombinant UvsX comprises a substitution mutation homologous to His63Ser in 25 the RB49 UvsX amino acid sequence.

In some embodiments, in addition to any of the above mutations, the recombinant UvsX can further comprise a mutation selected from the group consisting of: the addition of one or more glutamic acid residues at the C-terminus; the addition of one or more aspartic acid residues at the C-terminus; and a combination thereof.

In some embodiments, the recombinant UvsX is derived from a myoviridae phage selected from the group consisting of: T4, T6, Rb69, Aeh1, KVP40, Acinetobacter phage 133, Aeromonas phage 65, cyanophage P-SSM2, cyanophage PSSM4, cyanophage S-PM2, Rb32, Vibrio phage nt-1, Rb16, Rb43, and Rb49.

5 In some embodiments, the recombinant UvsX is derived from a myoviridae phage selected from the group consisting of: T2, Rb14, Aeromonas phage 25, phi-1, Phage 31, phage 44RR2.8t, phage Rb3, and phage LZ2.

Also presented herein is a nucleic acid molecule encoding a recombinant UvsX as defined in any the above embodiments. Also presented herein is an expression
10 vector comprising the nucleic acid molecule described above. Also presented herein is a host cell comprising the vector described above.

Also presented herein is a recombinase polymerase amplification process of amplification of a target nucleic acid molecule, comprising the steps of: (a) contacting the recombinant UvsX of any of the above embodiments with a first and second nucleic
15 acid primer to form a first and second nucleoprotein primer, wherein said nucleic acid primer comprises a single stranded region at its 3' end; (b) contacting the first and the second nucleoprotein primer to said target nucleic acid molecule thereby forming a first double-stranded structure at a first portion of said first strand and forming a second double stranded structure at a second portion of said second strand such that the 3' ends
20 of said first nucleic acid primer and said second nucleic acid primer are oriented toward one another on the same double-stranded template nucleic acid molecule; (c) extending the 3' end of said first and second nucleic acid primer with one or more polymerases and dNTPs to generate a first and second double-stranded nucleic acid and a first and second displaced strands of nucleic acid; and (d) continuing the reaction through
25 repetition of (b) and (c) until a desired degree of amplification is reached.

In certain embodiments of the process, the target nucleic acid molecule comprises double stranded nucleic acid. In certain embodiments, the target nucleic acid molecule comprises single stranded nucleic acid. For example, in some embodiments, the target nucleic acid comprises a single stranded adaptor region. In certain
30 embodiments, the process is performed in the presence of a recombinase loading

protein. For example, the recombinase loading protein can be selected from the group consisting of T4 UvsY, *E. coli* recO, *E. coli* recR, and a combination thereof. In certain embodiments, the process is performed in the presence of a single strand stabilizing agent selected from the group consisting of gp32, *E. coli* SSB protein, T4 gp32 protein, and derivatives thereof. In certain embodiments, the process is performed in the presence of a crowding agent selected from the group comprising polyethylene glycol, polyethylene oxide, polystyrene, Ficoll, dextran, PVP, and albumin such that the crowding agent stimulates amplification.

In certain embodiments, the process is performed on an array of amplification sites. In certain embodiments, each amplification site comprises a plurality of amplification primers for amplification of the target nucleic acid. In certain embodiments, the array of amplification sites comprises an array of features on a surface. For example, the features can be non-contiguous and can be separated by interstitial regions of the surface that lack the amplification primers. In certain embodiments, the array of amplification sites comprises beads in solution or beads on a surface. In certain embodiments, the array of amplification sites comprises an emulsion. In certain embodiments, the process occurs isothermally.

Also presented herein is a kit for performing a recombinase polymerase reaction. In certain embodiments, the kit can comprise a recombinant UvsX as defined in any the above embodiments, and one or more of the following: a single stranded DNA binding protein; a DNA polymerase; dNTPs or a mixture of dNTPs and ddNTPs; a crowding agent; a buffer; a reducing agent; ATP or ATP analog; a recombinase loading protein; a first primer and optionally a second primer.

The details of one or more embodiments are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic showing alignment of UvsX amino acid sequences from *Enterobacteria phage* T4 (T4) (SEQ ID NO: 8), *Enterobacteria phage* T6 (T6) (SEQ ID

NO: 9), *Acinetobacter phage* 133 (Phage133) (SEQ ID NO: 10), *Enterobacteria phage* RB69 (Rb69) (SEQ ID NO: 11), *Aeromonas phage* Aeh1 (Aeh1) (SEQ ID NO: 12), *Aeromonas phage* 65 (Ae65) (SEQ ID NO: 13), *Vibrio phage* KVP40 (Kvp40) (SEQ ID NO: 14), *Enterobacteria phage* RB43 (Rb43) (SEQ ID NO: 15), *Prochlorococcus* 5 *phage* P-SSM2 (PSSM2) (SEQ ID NO: 16), and *Prochlorococcus phage* P-SSM4 (PSSM4) (SEQ ID NO: 17), as also set forth in the incorporated materials of US 2009/0029421. Residues that are positionally and/or functionally equivalent to Pro256 in the RB49 UvsX amino acid sequence are highlighted and indicated by a triangle symbol.

10 Figure 2 is a schematic showing alignment of UvsX amino acid sequences from *Enterobacteria phage* RB49 (RB49) (SEQ ID NO: 1) and *Enterobacteria phage* T4 (T4) (SEQ ID NO: 8). Residues that are positionally and/or functionally equivalent to Pro256 in the RB49 UvsX amino acid sequence are highlighted and indicated by a triangle symbol.

15 Figure 3A shows a screenshot of a cluster image of a PCR-free library seeded onto a patterned flow cell using a T4 UvsX formulation.

 Figure 3B shows a screenshot of a cluster image of a PCR-free library seeded onto a patterned flow cell using a liquid formulation that includes RB49 P256K recombinase.

20 Figure 4A shows a screenshot of a cluster image of a single stranded (ssDNA) PCR-free library seeded onto a patterned flow cell using a T4 UvsX formulation.

 Figure 4B shows a screenshot of a cluster image of a single stranded PCR-free library seeded onto a patterned flow cell using a liquid formulation that includes RB49 P256K recombinase.

25

DETAILED DESCRIPTION

 Presented herein are recombinases for improved recombinase-mediated amplification of nucleic acids. The present inventors have surprisingly identified certain altered recombinases which have substantially improved characteristics in the 30 seeding nucleic acids onto a patterned flow cell surface.

As described in greater detail hereinbelow, the inventors have surprisingly found that one or more mutations to one or more residues in the recombinase result in profound improvements in seeding a DNA library, such as, for example, a PCR-library having single-stranded adapter regions, on a patterned flow cell surface, giving
5 improved cluster amplification.

In certain embodiments, the substitution mutation comprises a mutation to a residue having a charged side chain. For example, in some embodiments, the charged amino acid is a positively charged amino acid residue. The term “positively charged amino acid” refers to a hydrophilic amino acid with a side chain pKa value of greater
10 than 7, namely a basic amino acid. Basic amino acids typically have positively charged side chains at physiological pH due to association with a hydronium ion. Naturally occurring (genetically encoded) basic amino acids include lysine (Lys, K), arginine (Arg, R) and histidine (His, H), while non-natural (non-genetically encoded, or non-standard) basic amino acids include, for example, ornithine, 2,3,-diaminopropionic acid,
15 2,4-diaminobutyric acid, 2,5,6-triaminohexanoic acid, 2-amino-4-guanidinobutanoic acid, and homoarginine. The term “negatively charged amino acid” refers to a natural or non-natural amino acid, regardless of chirality, containing, in addition to the C-terminal carboxyl group, at least one additional negatively charged group such as carboxyl, phosphate, phosphonate, sulfonate, or the like.

Also presented herein is a recombinant UvsX comprising a substitution mutation to a semi-conserved domain of the recombinant UvsX. As used herein, the term “semi-conserved domain” refers to a portion of the recombinant UvsX that is fully conserved, or at least partially conserved among various species. It has been surprisingly discovered that mutation of one or more residues in the semi-conserved domain affects
25 the recombinase activity especially in the presence of single-strand template nucleic acid, resulting in enhancement of seeding and/or amplification in recombinase-mediated amplification reactions. These mutated recombinases have improved performance in seeding of PCR-free libraries, such as a PCR-library having single-stranded adapter regions, on a patterned flow cell surface, resulting in improved cluster amplification, as
30 described in the Example section below.

In some embodiments, the semi-conserved domain comprises amino acids having the sequence set forth in any of SEQ ID NOs: 3-7. SEQ ID NOs: 3-7 correspond to residues in the semi-conserved domain among various species. SEQ ID NO: 3 corresponds to residues 251-258 of the T4 UvsX amino acid sequence, which is set forth herein as SEQ ID NO: 8. An alignment showing the conservation among various species in the semi-conserved domain is set forth in Figures 1 and 2. The UvsX sequences shown in Figure 1 were obtained from Genbank database accession numbers NP_049656 (T4), YP_004300647 (Phage 133); NP_861734 (RB69); NP_943894.1 (Aeh1); YP_004300858 (Ae65); NP_899256 (KVP40); YP_239013 (RB43); YP_214417 (P-SSM2); YP_214708 (P-SSM4); and from US Publication No. 2009/0029421 (T6). Figure 2 is a schematic showing alignment of UvsX amino acid sequences from *Enterobacteria phage* RB49 (RB49) (SEQ ID NO: 1) and *Enterobacteria phage* T4 (T4) (SEQ ID NO: 8). Residues that are positionally and/or functionally equivalent to Pro256 in the RB49 UvsX amino acid sequence are highlighted and indicated by a triangle. The UvsX sequences shown in Figure 2 were obtained from Genbank database accession numbers NP_891595 (RB49) and NP_049656 (T4).

Mutations to one or more residues in the semi-conserved domain have been surprisingly found to increase the recombinase activity especially in the presence of single-strand template nucleic acid, resulting in enhancement of seeding and/or amplification in recombinase-mediated amplification reactions. These mutated recombinases have improved performance in seeding of PCR-free libraries, such as a PCR-library having single-stranded adapter regions, on a patterned flow cell surface, resulting in improved cluster amplification, as described in the Example section below. For example, in some embodiments of the recombinant UvsX presented herein, the substitution mutation comprises a mutation at position 7 of any of SEQ ID NOs: 3-5 to any residue other than Phe, Pro, Asp, Glu or Asn. In certain embodiments, the recombinant UvsX comprises a mutation to Lys at position 7 of any of SEQ ID NOs: 3-5. In some embodiments of the recombinant UvsX presented herein, the substitution mutation comprises a mutation at position 12 of any of SEQ ID NOs: 6-7 to

any residue other than Phe, Pro, Asp, Glu or Asn. In certain embodiments, the recombinant UvsX comprises a mutation to Lys at position 12 of any of SEQ ID NOs: 6-7.

In some embodiments, the recombinase is a UvsX protein. Any phage recombinase can be used in the embodiments presented herein, including, for example phage recombinases such as UvsX or UvsX-like recombinase derived from a myoviridae phage such as, for example, T4, T6, Rb69, Aeh1, KVP40, Acinetobacter phage 133, Aeromonas phage 65, cyanophage P-SSM2, cyanophage PSSM4, cyanophage S-PM2, Rb32, Vibrio phage nt-1, Rb16, Rb43, and Rb49. In certain embodiments, the recombinase is a UvsX or UvsX-like recombinase derived from a myoviridae phage such as, for example, T2, Rb14, Aeromonas phage 25, phi-1, Phage 31, phage 44RR2.8t, phage Rb3, and phage LZ2. It will be readily apparent to one of skill in the art that other recombinase proteins can be used in the embodiments presented herein. Suitable recombinase proteins can be identified by homology to UvsX using any number of a number of methods known in the art, such as, for example, BLAST alignment, as described in greater detail below.

By "functionally equivalent" it is meant that the control recombinase, in the case of studies using a different recombinase entirely, will contain the amino acid substitution that is considered to occur at the amino acid position in the other recombinase that has the same functional role in the enzyme. As an example, the mutation at position 257 from Phenylalanine to Lysine (F257K) in the T4 UvsX would be functionally equivalent to a substitution at position 256 from Proline to Lysine (P256K) in RB49 UvsX.

Generally functionally equivalent substitution mutations in two or more different recombinases occur at homologous amino acid positions in the amino acid sequences of the recombinases. Hence, use herein of the term "functionally equivalent" also encompasses mutations that are "positionally equivalent" or "homologous" to a given mutation, regardless of whether or not the particular function of the mutated amino acid is known. It is possible to identify positionally equivalent or homologous amino acid residues in the amino acid sequences of two or more different recombinases

on the basis of sequence alignment and/or molecular modelling. An example of sequence alignment to identify positionally equivalent and/or functionally equivalent residues is set forth in Figure 1, which sets forth an alignment of UvsX amino acid sequences from *Enterobacteria phage* T4 (T4) (SEQ ID NO: 8), *Enterobacteria phage* T6 (T6) (SEQ ID NO: 9), *Acinetobacter phage* 133 (Phage133) (SEQ ID NO: 10), *Enterobacteria phage* RB69 (Rb69) (SEQ ID NO: 11), *Aeromonas phage* Aeh1 (Aeh1) (SEQ ID NO: 12), *Aeromonas phage* 65 (Ae65) (SEQ ID NO: 13), *Vibrio phage* KVP40 (Kvp40) (SEQ ID NO: 14), *Enterobacteria phage* RB43 (Rb43) (SEQ ID NO: 15), *Prochlorococcus phage* P-SSM2 (PSSM2) (SEQ ID NO: 16), and *Prochlorococcus phage* P-SSM4 (PSSM4) (SEQ ID NO: 17), as also set forth in the incorporated materials of US 2009/0029421. The UvsX sequences shown in Figure 1 were obtained from Genbank database accession numbers NP_049656 (T4), YP_004300647 (Phage 133); NP_861734 (RB69); NP_943894.1 (Aeh1); YP_004300858 (Ae65); NP_899256 (KVP40); YP_239013 (RB43); YP_214417 (P-SSM2); YP_214708 (P-SSM4); and from US Publication No. 2009/0029421 (T6).

Figure 2 is a schematic showing alignment of UvsX amino acid sequences from *Enterobacteria phage* RB49 (RB49) (SEQ ID NO: 1) and *Enterobacteria phage* T4 (T4) (SEQ ID NO: 8). Residues that are positionally and/or functionally equivalent to Pro256 in the RB49 UvsX amino acid sequence are highlighted and indicated by a triangle. The UvsX sequences shown in Figure 2 were obtained from Genbank database accession numbers NP_891595 (RB49) and NP_049656 (T4).

A positionally equivalent and/or functionally equivalent residue can be determined for one or more of any number of other UvsX sequences by aligning those sequences with that of a reference sequence such as T4 and RB49. As a non-limiting example, UvsX sequences from *Synechococcus phage* S-PM2, *Enterobacteria phage* RB32, *Vibrio phage* nt-1, *Enterobacteria phage* RB16 are set forth as SEQ ID NOs: 18-21, and obtained from Genbank database accession numbers YP_195169.1; YP_802982.1; YP_008125207.1; YP_003858336.1, can be aligned with a reference UvsX sequence such as, for example T4 UvsX (SEQ ID NO:8) and RB49 UvsX (SEQ ID NO: 1) and positionally equivalent and/or functionally equivalent residues are

identified. By way of example, the residues shown in the table below are identified as positionally equivalent and/or functionally equivalent to Pro256 in the RB49 UvsX amino acid sequence. It will be readily appreciated by one of skill in the art that positionally equivalent and/or functionally equivalent positions for other UvsX proteins

5 can be ascertained by following a similar approach.

Phage Species	SEQ ID NO:	Positionally/Functionally Equivalent Position
T4	8	Phe257
T6	9	Phe259
Acinetobacter phage 133	10	Pro257
Rb69	11	Pro258
Aeh1	12	Pro269
Aeromonas phage 65	13	Asp266
KVP40	14	Pro267
Rb43	15	Pro259
cyanophage P-SSM2	16	Gln261
cyanophage PSSM4	17	Glu264
cyanophage S-PM2	18	Glu264
Rb32	19	Phe259
Vibrio phage nt-1	20	Pro267
Rb16	21	Pro259

The recombinant UvsX proteins described hereinabove can comprise additional substitution mutations that are known to enhance one or more aspects of recombinase

10 activity, stability or any other desirable property. For example, in some embodiments, in addition to any of the above mutations, the recombinant UvsX can further comprise substitution mutations at positions functionally equivalent His63 in the RB49 UvsX amino acid sequence as is known in the art and exemplified by the disclosure of US

2009/0029421, which is incorporated by reference in its entirety. For example, in certain embodiments, the recombinant UvsX comprises a substitution mutation homologous to His63Ser in the RB49 UvsX amino acid sequence.

5 In some embodiments, in addition to any of the above mutations, the recombinant UvsX can comprise additional substitution, deletion and/or addition mutations as compared to a wild type recombinase. Any of a variety of substitution mutations at one or more of positions can be made, as is known in the art and exemplified by the incorporated materials of 2009/0029421. For example, in some
10 embodiments, in addition to the above mutations, the recombinant UvsX can further comprise a mutation selected from the group consisting of: the addition of one or more glutamic acid residues at the C-terminus; the addition of one or more aspartic acid residues at the C-terminus; and a combination thereof.

Mutating Recombinases

15 Various types of mutagenesis are optionally used in the present disclosure, e.g., to modify recombinases to produce variants, e.g., in accordance with recombinase models and model predictions, or using random or semi-random mutational approaches. In general, any available mutagenesis procedure can be used for making recombinase mutants. Such mutagenesis procedures optionally include selection of mutant nucleic
20 acids and polypeptides for one or more activity of interest (e.g., enhanced seeding and/or amplification on a solid support). Procedures that can be used include, but are not limited to: site-directed point mutagenesis, random point mutagenesis, in vitro or in vivo homologous recombination (DNA shuffling and combinatorial overlap PCR), mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis,
25 phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA, point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, degenerate PCR, double-strand break repair, and many others known to persons of skill. The starting recombinase for mutation can be any of those noted herein,

including available recombinases mutants such as those identified e.g., in US 2009/0029421, which is incorporated by reference in its entirety.

Optionally, mutagenesis can be guided by known information from a naturally occurring recombinase molecule, or of a known altered or mutated recombinase (e.g.,
5 using an existing mutant recombinase as noted in the preceding references), e.g., sequence, sequence comparisons, physical properties, crystal structure and/or the like as discussed above. However, in another class of embodiments, modification can be essentially random (e.g., as in classical or "family" DNA shuffling, see, e.g., Cramer et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed
10 evolution" *Nature* 391:288-291).

Additional information on mutation formats is found in: Sambrook et al., *Molecular Cloning--A Laboratory Manual* (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 2000 ("Sambrook"); *Current Protocols in Molecular Biology*, F. M. Ausubel et al., eds., Current Protocols, a joint venture
15 between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2011) ("Ausubel") and *PCR Protocols A Guide to Methods and Applications* (Innis et al. eds) Academic Press Inc. San Diego, Calif. (1990) ("Innis"). The following publications and references cited within provide additional detail on mutation formats: Arnold, Protein engineering for unusual environments, *Current*
20 *Opinion in Biotechnology* 4:450-455 (1993); Bass et al., Mutant Trp repressors with new DNA-binding specificities, *Science* 242:240-245 (1988); Bordo and Argos (1991) Suggestions for "Safe" Residue Substitutions in Site-directed Mutagenesis 217:721-729; Botstein & Shortle, Strategies and applications of in vitro mutagenesis, *Science* 229:1193-1201 (1985); Carter et al., Improved oligonucleotide site-directed
25 mutagenesis using M13 vectors, *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, Site-directed mutagenesis, *Biochem. J.* 237:1-7 (1986); Carter, Improved oligonucleotide-directed mutagenesis using M13 vectors, *Methods in Enzymol.* 154: 382-403 (1987); Dale et al., Oligonucleotide-directed random mutagenesis using the phosphorothioate method, *Methods Mol. Biol.* 57:369-374 (1996); Eghtedarzadeh & Henikoff, Use of
30 oligonucleotides to generate large deletions, *Nucl. Acids Res.* 14: 5115 (1986); Fritz et

al., Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, *Nucl. Acids Res.* 16: 6987-6999 (1988); Grundstrom et al., Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis, *Nucl. Acids Res.* 13: 3305-3316 (1985); Hayes (2002) Combining

5 Computational and Experimental Screening for rapid Optimization of Protein Properties PNAS 99(25) 15926-15931; Kunkel, The efficiency of oligonucleotide directed mutagenesis, in *Nucleic Acids & Molecular Biology* (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin)) (1987); Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection, *Proc. Natl. Acad. Sci. USA* 82:488-492

10 (1985); Kunkel et al., Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods in Enzymol.* 154, 367-382 (1987); Kramer et al., The gapped duplex DNA approach to oligonucleotide-directed mutation construction, *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer & Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, *Methods in Enzymol.* 154:350-367 (1987); Kramer et al.,

15 Point Mismatch Repair, *Cell* 38:879-887 (1984); Kramer et al., Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, *Nucl. Acids Res.* 16: 7207 (1988); Ling et al., Approaches to DNA mutagenesis: an overview, *Anal Biochem.* 254(2): 157-178 (1997); Lorimer and Pastan *Nucleic Acids Res.* 23, 3067-8 (1995); Mandecki, Oligonucleotide-directed

20 double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis, *Proc. Natl. Acad. Sci. USA*, 83:7177-7181(1986); Nakamaye & Eckstein, Inhibition of restriction endonuclease *Nci* I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 14: 9679-9698 (1986); Nambiar et al., Total synthesis and cloning of a gene coding for the

25 ribonuclease S protein, *Science* 223: 1299-1301(1984); Sakamar and Khorana, Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin), *Nucl. Acids Res.* 14: 6361-6372 (1988); Sayers et al., Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 16:791-802 (1988); Sayers et al., Strand specific

30 cleavage of phosphorothioate-containing DNA by reaction with restriction

endonucleases in the presence of ethidium bromide, (1988) Nucl. Acids Res. 16: 803-814; Sieber, et al., Nature Biotechnology, 19:456-460 (2001); Smith, In vitro mutagenesis, Ann. Rev. Genet. 19:423-462 (1985); Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Stemmer, Nature 370, 389-91(1994); Taylor et al., The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, Nucl. Acids Res. 13: 8749-8764 (1985); Taylor et al., The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, Nucl. Acids Res. 13: 8765-8787 (1985); Wells et al., Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, Phil. Trans. R. Soc. Lond. A 317: 415-423 (1986); Wells et al., Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, Gene 34:315-323 (1985); Zoller & Smith, Oligonucleotide-directed mutagenesis using M 13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment, Nucleic Acids Res. 10:6487-6500 (1982); Zoller & Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, Methods in Enzymol. 100:468-500 (1983); Zoller & Smith, Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template, Methods in Enzymol. 154:329-350 (1987); Clackson et al. (1991) "Making antibody fragments using phage display libraries" Nature 352:624-628; Gibbs et al. (2001) "Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling" Gene 271:13-20; and Hiraga and Arnold (2003) "General method for sequence-independent site-directed chimeragenesis: J. Mol. Biol. 330:287-296. Additional details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

Making and Isolating Recombinant Recombinase

Generally, nucleic acids encoding a recombinase as presented herein can be made by cloning, recombination, in vitro synthesis, in vitro amplification and/or other

available methods. A variety of recombinant methods can be used for expressing an expression vector that encodes a recombinase as presented herein. Methods for making recombinant nucleic acids, expression and isolation of expressed products are well known and described in the art. A number of exemplary mutations and combinations of mutations, as well as strategies for design of desirable mutations, are described herein.

Additional useful references for mutation, recombinant and in vitro nucleic acid manipulation methods (including cloning, expression, PCR, and the like) include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, Calif. (Berger); Kaufman et al. (2003) Handbook of Molecular and Cellular Methods in Biology and Medicine Second Edition Ceske (ed) CRC Press (Kaufman); and The Nucleic Acid Protocols Handbook Ralph Rapley (ed) (2000) Cold Spring Harbor, Humana Press Inc (Rapley); Chen et al. (ed) PCR Cloning Protocols, Second Edition (Methods in Molecular Biology, volume 192) Humana Press; and in Viljoen et al. (2005) Molecular Diagnostic PCR Handbook Springer, ISBN 1402034032.

In addition, a plethora of kits are commercially available for the purification of plasmids or other relevant nucleic acids from cells, (see, e.g., EasyPrep.TM., FlexiPrep.TM. both from Pharmacia Biotech; StrataClean.TM., from Stratagene; and, QIAprep.TM. from Qiagen). Any isolated and/or purified nucleic acid can be further manipulated to produce other nucleic acids, used to transfect cells, incorporated into related vectors to infect organisms for expression, and/or the like. Typical cloning vectors contain transcription and translation terminators, transcription and translation initiation sequences, and promoters useful for regulation of the expression of the particular target nucleic acid. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in eukaryotes, or prokaryotes, or both, (e.g., shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and integration in prokaryotes, eukaryotes, or both.

Other useful references, e.g. for cell isolation and culture (e.g., for subsequent nucleic acid isolation) include Freshney (1994) Culture of Animal Cells, a Manual of

Basic Technique, third edition, Wiley-Liss, New York and the references cited therein; Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, N.Y.; Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, Fla.

Nucleic acids encoding the recombinant recombinases of disclosed herein are also a feature of embodiments presented herein. A particular amino acid can be encoded by multiple codons, and certain translation systems (e.g., prokaryotic or eukaryotic cells) often exhibit codon bias, e.g., different organisms often prefer one of the several synonymous codons that encode the same amino acid. As such, nucleic acids presented herein are optionally "codon optimized," meaning that the nucleic acids are synthesized to include codons that are preferred by the particular translation system being employed to express the recombinase. For example, when it is desirable to express the recombinase in a bacterial cell (or even a particular strain of bacteria), the nucleic acid can be synthesized to include codons most frequently found in the genome of that bacterial cell, for efficient expression of the recombinase. A similar strategy can be employed when it is desirable to express the recombinase in a eukaryotic cell, e.g., the nucleic acid can include codons preferred by that eukaryotic cell.

A variety of protein isolation and detection methods are known and can be used to isolate recombinases, e.g., from recombinant cultures of cells expressing the recombinant recombinases presented herein. A variety of protein isolation and detection methods are well known in the art, including, e.g., those set forth in R. Scopes, Protein Purification, Springer-Verlag, N.Y. (1982); Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N.Y. (1990); Sandana (1997) Bioseparation of Proteins, Academic Press, Inc.; Bollag et al. (1996) Protein Methods, 2nd Edition Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ, Harris and Angal (1990) Protein Purification Applications: A Practical Approach IRL Press at Oxford, Oxford, England; Harris and Angal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, England;

Scopes (1993) Protein Purification: Principles and Practice 3rd Edition Springer Verlag, NY; Janson and Ryden (1998) Protein Purification: Principles, High Resolution Methods and Applications, Second Edition Wiley-VCH, NY; and Walker (1998) Protein Protocols on CD-ROM Humana Press, NJ; and the references cited therein.

5 Additional details regarding protein purification and detection methods can be found in Satinder Ahuja ed., Handbook of Bioseparations, Academic Press (2000).

Methods of use

The altered recombinases presented herein can be used in a recombinase-mediated amplification procedure, such as a recombinase polymerase amplification (RPA) technique. Briefly, RPA can be initiated by contacting a target nucleic acid with a recombinase and a single stranded nucleic acid primer specific for the target nucleic acid molecule. The hybridized primer can then be extended by a polymerase, such as a polymerase capable of strand displacement in the presence of dNTPs to generate a double stranded target nucleic acid molecule and a displaced strand of nucleic acid molecule. Further amplification can take place by recombinase-mediated targeting of primers to the displaced strand of nucleic acid molecule and extension of the primer to generate a double stranded nucleic acid molecule. The RPA process can be modulated by combination of the above-described components with, for example, recombinase-loading factors, specific strand-displacing polymerases and a robust energy regeneration system. Exemplary RPA procedures, systems and components that can be readily adapted for use with the recombinant UvsX proteins of the present disclosure are described, for example, in US Pat. Nos. 8,071,308; 7,399,590, 7,485,428, 7,270,981, 8,030,000, 7,666,598, 7,763,427, 8,017,399, 8,062,850, and 7,435,561, each of which is

10
15
20
25

In some embodiments, isothermal amplification can be performed using kinetic exclusion amplification (KEA), also referred to as exclusion amplification (ExAmp). A nucleic acid library of the present disclosure can be made using a method that exploits kinetic exclusion. Kinetic exclusion can occur when a process occurs at a sufficiently rapid rate to effectively exclude another event or process from occurring. Take for

30

example the making of a nucleic acid array where sites of the array are randomly seeded with target nucleic acids from a solution and copies of the target nucleic acid are generated in an amplification process to fill each of the seeded sites to capacity. In accordance with the kinetic exclusion methods of the present disclosure, the seeding and amplification processes can proceed simultaneously under conditions where the amplification rate exceeds the seeding rate. As such, the relatively rapid rate at which copies are made at a site that has been seeded by a first target nucleic acid will effectively exclude a second nucleic acid from seeding the site for amplification. Kinetic exclusion amplification methods can be performed as described in detail in the disclosure of U.S. Application Pub. No. 2013/0338042, which is incorporated herein by reference in its entirety.

In some embodiments, the target nucleic acid that is amplified is fully double stranded. In some embodiments, the target nucleic acid that is amplified comprises a region of double stranded nucleic acid, and also comprises a region having single stranded nucleic acid. In certain embodiments, the target nucleic acid comprises one or more forked adapters with a region of about 5, 10, 15, 20, 25, 30, 35, 40 or more than about 40 bases of single stranded sequence at each end of the library fragments. Design and use of forked adapters is described in greater detail in the disclosures of U.S. Pat. Nos. 7,742,463 and 8,563,748, each of which is incorporated herein by reference in its entirety.

Kinetic exclusion can exploit a relatively slow rate for making a first copy of a target nucleic acid vs. a relatively rapid rate for making subsequent copies of the target nucleic acid or of the first copy. In the example of the previous paragraph, kinetic exclusion occurs due to the relatively slow rate of target nucleic acid seeding (e.g. relatively slow diffusion or transport) vs. the relatively rapid rate at which amplification occurs to fill the site with copies of the nucleic acid seed. In another exemplary embodiment, kinetic exclusion can occur due to a delay in the formation of a first copy of a target nucleic acid that has seeded a site (e.g. delayed or slow activation) vs. the relatively rapid rate at which subsequent copies are made to fill the site. In this example, an individual site may have been seeded with several different target nucleic

acids (e.g. several target nucleic acids can be present at each site prior to amplification). However, first copy formation for any given target nucleic acid can be activated randomly such that the average rate of first copy formation is relatively slow compared to the rate at which subsequent copies are generated. In this case, although an individual site may have been seeded with several different target nucleic acids, kinetic exclusion will allow only one of those target nucleic acids to be amplified. More specifically, once a first target nucleic acid has been activated for amplification, the site will rapidly fill to capacity with its copies, thereby preventing copies of a second target nucleic acid from being made at the site.

10 An amplification reagent can include further components that facilitate amplicon formation and in some cases increase the rate of amplicon formation. Recombinase, such as for example UvsX, can facilitate amplicon formation by allowing repeated invasion/extension. More specifically, recombinase can facilitate invasion of a target nucleic acid by the polymerase and extension of a primer by the polymerase using the target nucleic acid as a template for amplicon formation. This process can be repeated as a chain reaction where amplicons produced from each round of invasion/extension serve as templates in a subsequent round. The process can occur more rapidly than standard PCR since a denaturation cycle (e.g. via heating or chemical denaturation) is not required. As such, recombinase-facilitated amplification can be carried out isothermally. It is generally desirable to include ATP, or other nucleotides (or in some cases non-hydrolyzable analogs thereof) in a recombinase-facilitated amplification reagent to facilitate amplification. A mixture of recombinase and single stranded binding (SSB) protein is particularly useful as SSB can further facilitate amplification. Exemplary formulations for recombinase-facilitated amplification include those sold commercially as TwistAmp kits by TwistDx (Cambridge, UK). Useful components of recombinase-facilitated amplification reagent and reaction conditions are set forth in US 5,223,414 and US 7,399,590, each of which is incorporated herein by reference.

Sequence Comparison, Identity, and Homology

The terms "identical" or "percent identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using
5 one of the sequence comparison algorithms described below (or other algorithms available to persons of skill) or by visual inspection.

The phrase "substantially identical," in the context of two nucleic acids or polypeptides (e.g., DNAs encoding a recombinase, or the amino acid sequence of a recombinase) refers to two or more sequences or subsequences that have at least about
10 60%, about 80%, about 90-95%, about 98%, about 99% or more nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm or by visual inspection. Such "substantially identical" sequences are typically considered to be "homologous," without reference to actual ancestry. Preferably, the "substantial identity" exists over a
15 region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably, the sequences are substantially identical over at least about 150 residues, or over the full length of the two sequences to be compared.

Proteins and/or protein sequences are "homologous" when they are derived,
20 naturally or artificially, from a common ancestral protein or protein sequence. Similarly, nucleic acids and/or nucleic acid sequences are homologous when they are derived, naturally or artificially, from a common ancestral nucleic acid or nucleic acid sequence. Homology is generally inferred from sequence similarity between two or more nucleic acids or proteins (or sequences thereof). The precise percentage of similarity between
25 sequences that is useful in establishing homology varies with the nucleic acid and protein at issue, but as little as 25% sequence similarity over 50, 100, 150 or more residues is routinely used to establish homology. Higher levels of sequence similarity, e.g., 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% or more, can also be used to establish homology. Methods for determining sequence similarity percentages (e.g.,

BLASTP and BLASTN using default parameters) are described herein and are generally available.

For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally *Current Protocols in Molecular Biology*, Ausubel et al., eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., supplemented through 2004).

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased.

Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted
5 when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as
10 defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

In addition to calculating percent sequence identity, the BLAST algorithm also
15 performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid
20 is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Nucleic acids encoding altered recombinases

25 Further presented herein are nucleic acid molecules encoding the altered recombinase enzymes presented herein. For any given altered recombinase which is a mutant version of a recombinase for which the amino acid sequence and preferably also the wild type nucleotide sequence encoding the recombinase is known, it is possible to obtain a nucleotide sequence encoding the mutant according to the basic principles of
30 molecular biology. For example, given that the wild type nucleotide sequence encoding

RB49 UvsX recombinase is known, it is possible to deduce a nucleotide sequence encoding any given mutant version of RB49 UvsX having one or more amino acid substitutions using the standard genetic code. Similarly, nucleotide sequences can readily be derived for mutant versions other recombinases such as, for example, T4, T6,
5 Rb69, Aeh1, KVP40, Acinetobacter phage 133, Aeromonas phage 65, cyanophage P-SSM2, cyanophage PSSM4, cyanophage S-PM2, Rb32, Vibrio phage nt-1, Rb16, Rb43, T2, Rb14, Aeromonas phage 25, phi-1, Phage 31, phage 44RR2.8t, phage Rb3, and phage LZ2, etc. Nucleic acid molecules having the required nucleotide sequence may then be constructed using standard molecular biology techniques known in the art.

10 In accordance with the embodiments presented herein, a defined nucleic acid includes not only the identical nucleic acid but also any minor base variations including, in particular, substitutions in cases which result in a synonymous codon (a different codon specifying the same amino acid residue) due to the degenerate code in conservative amino acid substitutions. The term “nucleic acid sequence” also includes
15 the complementary sequence to any single stranded sequence given regarding base variations.

The nucleic acid molecules described herein may also, advantageously, be included in a suitable expression vector to express the recombinase proteins encoded therefrom in a suitable host. Incorporation of cloned DNA into a suitable expression
20 vector for subsequent transformation of said cell and subsequent selection of the transformed cells is well known to those skilled in the art as provided in Sambrook et al. (1989), Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, which is incorporated by reference in its entirety.

Such an expression vector includes a vector having a nucleic acid according to
25 the embodiments presented herein operably linked to regulatory sequences, such as promoter regions, that are capable of effecting expression of said DNA fragments. The term “operably linked” refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. Such vectors may be transformed into a suitable host cell to provide for the expression of a protein
30 according to the embodiments presented herein.

The nucleic acid molecule may encode a mature protein or a protein having a prosequence, including that encoding a leader sequence on the preprotein which is then cleaved by the host cell to form a mature protein. The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, and optionally a
5 promoter for the expression of said nucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable markers, such as, for example, an antibiotic resistance gene.

Regulatory elements required for expression include promoter sequences to bind RNA polymerase and to direct an appropriate level of transcription initiation and also
10 translation initiation sequences for ribosome binding. For example, a bacterial expression vector may include a promoter such as the lac promoter and for translation initiation the Shine-Dalgarno sequence and the start codon AUG. Similarly, a eukaryotic expression vector may include a heterologous or homologous promoter for RNA polymerase II, a downstream polyadenylation signal, the start codon AUG, and a
15 termination codon for detachment of the ribosome. Such vectors may be obtained commercially or be assembled from the sequences described by methods well known in the art.

Transcription of DNA encoding the recombinase by higher eukaryotes may be optimised by including an enhancer sequence in the vector. Enhancers are cis-acting
20 elements of DNA that act on a promoter to increase the level of transcription. Vectors will also generally include origins of replication in addition to the selectable markers.

EXAMPLE 1

This example provides methods of seeding a PCR-free library on a patterned
25 flow cell surface for improved cluster amplification. In one embodiment, the method of the invention uses a seeding formulation that includes a UvsX comprising mutations set forth hereinabove, for example, a RB49 UvsX mutant comprising Pro256Lys (as set forth herein as SEQ ID NO: 2, referred to herein as “RB49 P256K”). It was surprisingly found that recombinase-mediated amplification using that substantially
30 improves the seeding of PCR-libraries with single-stranded adapter regions onto a

patterned flow cell surface. In another embodiment, the method of the invention uses a seeding formulation that includes a relatively high concentration of DNA polymerase (e.g., eBsu polymerase) in combination with RB49 P256K recombinase.

To evaluate the efficacy of the RB49 P256K formulation in seeding a PCR-free library onto a patterned flow cell surface, a PCR-free library was generated using a TruSeq® DNA PCR-free sample preparation kit (Illumina, Inc.). PCR-free libraries generated using the TruSeq® library preparation kit have forked adapters with a region of about 40 bases of single stranded sequence at each end of the library fragments.

Figure 3A shows a screenshot 100 of a cluster image of a PCR-free library seeded onto a patterned flow cell using a standard formulation comprising T4 UvsX recombinase (as set forth herein as SEQ ID NO: 8, referred to herein as “T4 UvsX”). Figure 3B shows a screenshot 150 of a cluster image of a PCR-free library seeded onto a patterned flow cell using a liquid formulation that includes RB49 P256K recombinase. In this example, the library was mixed with the T4 UvsX formulation or the RB49 P256K formulation to 100pM final concentration, flushed onto a flow cell, and incubated on a cBot at 38 °C. After a 1 hour incubation period, the temperature was lowered to 20 °C and the flow cell was washed with HT2 wash buffer (Illumina). Clusters were stained with a 1:5,000 dilution of SYBR® Green (Life Technologies) in 0.1 M Tris/0.1 M sodium ascorbate and imaged on a fluorescence microscope. Referring to Figure 3A, cluster density generated by seeding a PCR-free library onto a patterned flow cell with a standard formulation (e.g., T4 UvsX) is relatively sparse. Referring to Figure 3B, the density of clusters generated by seeding a PCR-free library onto a patterned flow cell using a formulation that includes RB49 P256K recombinase is substantially improved.

Figure 4A shows a screenshot 200 of a cluster image of a single stranded (ssDNA) PCR-free library seeded onto a patterned flow cell using a standard T4 UvsX formulation. Figure 4B shows a screenshot 250 of a cluster image of a single stranded PCR-free library seeded onto a patterned flow cell using a liquid formulation that includes RB49 P256K recombinase. In this example, a double-stranded PCR-free library was denatured using NaOH and subsequently seeded onto the patterned flow cell

at a concentration of 50 pM. Referring to Figure 4A, cluster density generated by seeding a ssDNA, PCR-free DNA library onto a patterned flow cell with a standard formulation (e.g., T4 UvsX) is relatively sparse. Referring to Figure 4B, the density of clusters generated by seeding a ssDNA PCR-free library onto a patterned flow cell
5 using a formulation that includes RB49 P256K recombinase is substantially improved.

EXAMPLE 2

Improved amplification using RB49 P256K mutants

This example describes a comparison of amplification performance between
10 recombinases with and without the P256K mutation described herein. For the purposes of this example, “control” RB49 UvsX (set forth in SEQ ID NO: 1) further comprises a H63S mutation. A P256K mutant is generated by further mutating the control to bear a Lys residue at position 256, as set forth herein by SEQ ID NO: 2).

Clustering of a PCR-free library on a patterned flow cell is performed on a cBot
15 as described above in Example 1, using either control or P256K mutant. Sequencing is then performed on a HiSeq instrument (Illumina, Inc.) and the sequencing results are analyzed to determine callability of a variety of regions which are typically poorly represented in previous sequencing data.

Callability is a measure of the fraction of sites at which a single nucleotide
20 polymorphism (SNP) is called correctly. Ideally, this value is 1 (for 100%) meaning that at 100% of the sites within a particular type of region (i.e., high GC, etc) the SNPs are called correctly. Coverage is a measure of the fraction of sites which have a coverage $> n$, where n is typically 30x (i.e., the standard coverage for a human genome). The fosmid promoters are a set of 100 gene promoters which were identified as poorly
25 represented in previous sequencing data. The promoters were cloned into fosmid vectors. A High GC region may be defined as a region with at least 100 bp where GC content is equal to or over 75% ($N50 (G + C \geq 0.75) 100 N50$). A Huge GC region may be defined as a region with at least 100 bp where GC content is equal to or over 85% ($N50 (G + C \geq 0.85) 100 N50$). A Low GC region may be defined as a region with at
30 least 100 bp where GC content is equal to or less than 40% ($N50 (G + C \geq 0.40) 100$

N50). A High AT region may be defined as a region with at least 100 bp where AT content is equal to or over 75% ($N50 (A + T \geq 0.75) 100 N50$), downsampled to ~ 50k regions. A Huge AT region may be defined as a region with at least 100 bp where AT content is equal to or over 85% ($N50 (A + T \geq 0.85) 100 N50$), downsampled to ~ 50K regions. An AT dinucleotide repeat region may be defined as a region that includes long stretches of ATAT repeats.

A comparison of callability data for control vs. P256K mutants demonstrates that the P256K mutant shows unexpected and significant improvements in callability of one or more of fosmid promoter regions, High GC, Huge GC, Low GC, High AT, Huge AT, and AT dinucleotide repeat regions, compared to that of the control.

EXAMPLE 3

Improved amplification using mutants having mutations homologous to P256K

The performance comparison described above in Example 2 is repeated for other recombinases. In this example, “control” recombinases are generated by modifying the wild type recombinase to comprise a mutation homologous to the H63S in RB49, as set forth in the “control” column in the table below. The “P256K homolog” mutants are generated by further modifying the controls to bear a mutation homologous to the P256K in RB49 as set forth in the “P256K homolog” column in the table below.

For example, for T6 UvsX, control is generated by modifying wild type T6 UvsX (SEQ ID NO: 9) to bear a H66S mutation. The P256K homolog is further modified to bear both H66S and F259K mutations.

WT backbone	WT backbone SEQ ID NO:	Control	P256K homolog
T4	8	64S	64S F257K
T6	9	H66S	H66S

			F259K
Acinetobacter phage 133	10	H64S	H64S P257K
Rb69	11	H64S	H64S P258K
Aeh1	12	H76S	H76S P269K
Aeromonas phage 65	13	H73S	H73S D266K
KVP40	14	H64S	H64S P267K
Rb43	15	H66S	H66S P259K
cyanophage P-SSM2	16	T62S	T62S Q261K
cyanophage PSSM4	17	T65S	T65S E264K
cyanophage S-PM2	18	T65S	T65S E264K
Rb32	19	H66S	H66S F259K
Vibrio phage nt-1	20	H64S	H64S P267K
Rb16	21	H66S	H66S P259K

Clustering of a PCR-free library on a patterned flow cell is performed on a cBot as described above in Example 1, using either control or P256K mutant. Sequencing is then performed on a HiSeq instrument (Illumina, Inc.) and the sequencing results are

analyzed as described above in Example 2 to determine callability of a variety of regions which are typically poorly represented in previous sequencing data.

5 A comparison of callability data for control vs. P256K homolog mutants demonstrates that the P256K homolog mutants show unexpected and significant improvements in callability of one or more of fosmid promoter regions, High GC, Huge GC, Low GC, High AT, Huge AT, and AT dinucleotide repeat regions, compared to that of the control.

10 Throughout this application various publications, patents and/or patent applications have been referenced. The disclosure of these publications in their entirety is hereby incorporated by reference in this application.

The term comprising is intended herein to be open-ended, including not only the recited elements, but further encompassing any additional elements.

15 A number of embodiments have been described. Nevertheless, it will be understood that various modifications may be made. Accordingly, other embodiments are within the scope of the following claims.

CLAIMS

1. A UvsX recombinase polypeptide comprising an amino acid sequence that is at least 90% identical to SEQ ID NO: 1, which UvsX recombinase polypeptide comprises an amino acid substitution mutation which is a mutation to a basic residue at the position functionally equivalent to position 256 in the Enterobacteria phage RB49 UvsX of SEQ ID NO: 1, the UvsX recombinase polypeptide having increased recombinase activity compared to a T4 UvsX recombinase polypeptide having SEQ ID NO: 8 or a RB49 UvsX recombinase polypeptide having all of the amino acid sequence of SEQ ID NO: 1 except for a His to Ser substitution at amino acid position 63 of SEQ ID NO: 1.
2. A UvsX recombinase polypeptide according to Claim 1, in which the amino acid substitution mutation replaces the residue at the position functionally equivalent to position 256 in the polypeptide of SEQ ID NO: 1 with a lysine (Lys).
3. A UvsX recombinase polypeptide according to Claim 1, in which the UvsX recombinase polypeptide further comprises a mutation to a serine residue at a position functionally equivalent to position 63 in the polypeptide of SEQ ID NO: 1.
4. A UvsX recombinase polypeptide according to Claim 1, further comprising the addition of one or more glutamic acid residues at the C-terminus; the addition of one or more aspartic acid residues at the C-terminus; or a combination thereof.
5. A UvsX recombinase polypeptide according to Claim 1, in which the UvsX recombinase polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 2.
6. A UvsX recombinase polypeptide comprising the amino acid sequence SEQ ID NO: 2.
7. A UvsX recombinase polypeptide comprising an amino acid sequence that is at least 90% identical to SEQ ID NO: 1 and comprising a semiconserved domain, the semiconserved domain comprising all of SEQ ID NO: 3, 4 or 5 except for a substitution at position 7 of SEQ ID NO: 3, 4, or 5 to a basic residue, the UvsX recombinase polypeptide having increased recombinase activity compared to a T4 UvsX recombinase polypeptide having SEQ ID NO: 8 or a RB49 UvsX recombinase polypeptide having all of the amino acid sequence of SEQ ID NO: 1 except for a His to Ser substitution at amino acid position 63 of SEQ ID NO: 1.

8. A UvsX recombinase polypeptide according to Claim 7, in which the substitution at position 7 comprises a mutation to Lys.
9. A UvsX recombinase polypeptide comprising an amino acid sequence that is at least 90% identical to SEQ ID NO: 1 and comprising a semiconserved domain, the semiconserved domain comprising all of SEQ ID NO: 6 or 7 except for a substitution at position 12 of SEQ ID NO: 6 or 7 to a basic residue, the UvsX recombinase polypeptide having increased recombinase activity compared to a T4 UvsX recombinase polypeptide having SEQ ID NO: 8 or a RB49 UvsX recombinase polypeptide having all of the amino acid sequence of SEQ ID NO:1 except for a His to Ser substitution at amino acid position 63 of SEQ ID NO: 1.
10. A UvsX recombinase polypeptide according to Claim 9, in which the substitution at position 12 comprises a substitution to Lys.
11. A UvsX recombinase polypeptide comprising all of the amino acid sequence of SEQ ID NO: 1, in which the proline (Pro) at position 256 is substituted for a basic residue, in which the basic residue is not lysine (Lys).
12. A UvsX recombinase polypeptide according to Claim 6, further comprising the addition of one or more glutamic acid residues at the C-terminus; the addition of one or more aspartic acid residues at the C-terminus; or a combination thereof.
13. A UvsX recombinase polypeptide according to Claim 1, in which the amino acid substitution mutation replaces the residue at the position functionally equivalent to position 256 of the polypeptide of SEQ ID NO: 1 with a lysine (Lys) and further comprising an amino acid substitution mutation replacing the residue at a position functionally equivalent to position 63 of the polypeptide of SEQ ID NO: 1 with serine (Ser).
14. A UvsX recombinase polypeptide according to Claim 1, in which recombinase activity is increased in the presence of a single stranded template nucleic acid compared to the T4 UvsX recombinase polypeptide having SEQ ID NO: 8 or the RB49 UvsX recombinase polypeptide having all of the amino acid sequence of SEQ ID NO: 1 except for a His to Ser substitution at amino acid position 63 of SEQ ID NO: 1.
15. A UvsX recombinase polypeptide according to Claim 1, the UvsX recombinase polypeptide having improved seeding and/or amplification on a solid support compared to the T4 UvsX recombinase polypeptide having SEQ ID NO: 8 or the RB49 UvsX recombinase

polypeptide having all of the amino acid sequence of SEQ ID NO: 1 except for a His to Ser substitution at amino acid position 63 of SEQ ID NO: 1.

16. A UvsX recombinase polypeptide according to Claim 1, in which the basic residue comprises lysine (Lys), arginine (Arg), or histidine (His).

5 17. A UvsX recombinase polypeptide according to Claim 1, the UvsX recombinase polypeptide comprising an amino acid sequence that is at least 95% identical to SEQ ID NO: 1.

18. A UvsX recombinase polypeptide according to Claim 1, the UvsX recombinase polypeptide comprising an amino acid sequence that is at least 99% identical to SEQ ID NO:

10 1.