Sso7-Polymerase Conjugates with Decreased Non-Specific Activity

CROSS-REFERENCE TO RELATED PATENT APPLICATIONS

[0001] The present application claims benefit of priority to US Provisional Patent Application No. 61/473,682, filed April 8, 2011, and to US Provisional Patent Application No. 61/499,873, filed June 22, 2011, each of which are incorporated by reference.

BACKGROUND OF THE INVENTION

[0002] Nucleic acid amplification reactions, such as the polymerase chain reaction (PCR), are generally template-dependent reactions in which a desired nucleic acid sequence is amplified by treating separate complementary strands of a target nucleic acid with an excess of two oligonucleotide primers. The primers are extended to form complementary primer extension products which act as templates for synthesizing the desired nucleic acid sequence. In such processes, the nucleic acid sequence between the primers on the respective DNA strands is selectively amplified.

[0003] The activity of a polymerase can be improved by joining a sequence-non-specific double-stranded nucleic acid binding domain to the enzyme, or its catalytic domain (*see, e.g.*,WO0192501). Such modified polymerases exhibit increased processivity in comparison to the unmodified enzymes.

BRIEF SUMMARY OF THE INVENTION

[0004] The present invention provides for an Sso7 polymerase conjugate protein comprising an Sso7 domain linked to a polymerase; wherein:

an amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is not lysine (K); and/or

an amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is not arginine (R),

wherein the conjugate protein has a decreased non-specific amplification activity compared to an otherwise identical control conjugate protein in which the amino acid of the Sso7

domain corresponding to K28 of SEQ ID NO:2 is lysine (K) and the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is arginine (R).

[0005] In some embodiments, the non-specific amplification activity of the Sso7 polymerase conjugate protein is reduced by at least 10% compared to non-specific amplification activity of the control conjugate protein.

[0006] In some embodiments, the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is not lysine (K). In some embodiments, the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is selected from the group consisting of serine (S), threonine (T), cytosine (C), proline (P), aspartic acid (D), glutamic acid (E), asparagine (N), glutamine (Q), alanine (A), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), leucine (L), methionine (M), arginine (R), valine (V), tryptophan (W), and tyrosine (Y).

[0007] In some embodiments, the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is not arginine (R). In some embodiments, the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is selected from the group consisting of alanine (A), cytosine (C), aspartic acid (D), glutamic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L), methionine (M), asparagine (N), glutamine (Q), serine (S), threonine (T), valine (V), tryptophan (W), tyrosine (Y), and proline (P).

[0008] In some embodiments, the polymerase substantially lacks a 3'-5' exonuclease activity.

[0009] In some embodiments, the Sso7 domain is substantially (e.g., at least 60, 75, 80, 85, 90, or 95%) identical to the first 58 or 60 amino acids of SEQ ID NO:2 or to the full-length of SEQ ID NO:2. In some embodiments, the Sso7 domain is at least 90% identical to SEQ ID NO:2.

[0010] In some embodiments, the polymerase is substantially (e.g., at least 60, 75, 80, 85, 90, or 95%) identical to SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:61. In some embodiments, the polymerase comprises SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:61.

[0011] In some embodiments, the polymerase domain has thermally stable polymerase activity. In some embodiments, the polymerase domain is a family A polymerase domain. In some embodiments, the polymerase domain is a ΔTaq polymerase domain.

[0012] In some embodiments, the polymerase domain is a family B polymerase domain. In some embodiments, the polymerase domain is from Pyrococcus.

[0013] The present invention also provides for reaction mixtures comprising a Sso7 polymerase conjugate proteins as described above or elsewhere herein. In some embodiments, the kit further comprises at least one of or more oligonucleotide primers, one or more detectably-labeled oligonucleotide probe, a buffer, nucleoside triphosphates (dNTPs), a salt, a DNA binding dye, or a stabilizer.

[0014] The present invention also provides for kits that comprise Sso7 polymerase conjugate proteins as described above or elsewhere herein. In some embodiments, the kit further comprises, in the same or a different container that contains the conjugate protein, at least one of one or more oligonucleotide primers, one or more detectably-labeled oligonucleotide probe, a buffer, nucleoside triphosphates (dNTPs), a salt, or a stabilizer.

[0015] In some embodiments, the composition comprises a buffer (when measured at a concentration of 0.1 M) that has a change of no more than 0.027 pH units per degree C when between 20° and 37° C. In some embodiments, the buffer is selected from the group consisting of HEPES, ACES, PIPES, MOPSO, BES, MOPS, TES, TAPSO, POPSO, BICINE, TAPS, and AMPSO.

[0016] The present invention also provides for methods of amplifying a target nucleic acid in a sample. In some embodiments the method comprises incubating the target nucleic acid in a reaction mixture comprising:

a. at least one primer; and

b. a Sso7 polymerase conjugate proteins as described above or elsewhere herein,

under conditions to allow for amplification of the target nucleic acid with the primer, thereby amplifying the target nucleic acid.

[0017] In some embodiments, the method comprises a polymerase chain reaction (PCR).

[0018] The present invention also provides for nucleic acids comprising a polynucleotide encoding a Sso7 polymerase conjugate proteins as described above or elsewhere herein. In some embodiments, the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is not lysine (K). In some embodiments, the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is selected from the group consisting of serine (S), threonine (T), cytosine (C), proline (P), aspartic acid (D), glutamic acid (E), asparagine (N), glutamine (Q), alanine (A), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), leucine (L), methionine (M), arginine (R), valine (V), tryptophan (W), and tyrosine (Y).

[0019] In some embodiments, the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is not arginine (R). In some embodiments, the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is selected from the group consisting of alanine (A), cytosine (C), aspartic acid (D), glutamic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L), methionine (M), asparagine (N), glutamine (Q), serine (S), threonine (T), valine (V), tryptophan (W), tyrosine (Y), and proline (P).

[0020] In some embodiments, the polymerase substantially lacks a 3'-5' exonuclease activity.

[0021] In some embodiments, the Sso7 domain is substantially (e.g., at least 60, 75, 80, 85, 90, or 95%) identical to the first 58 or 60 amino acids of SEQ ID NO:2 or to the full-length of SEQ ID NO:2. In some embodiments, the Sso7 domain is at least 90% identical to SEQ ID NO:2.

[0022] In some embodiments, the polymerase is substantially (e.g., at least 60, 75, 80, 85, 90, or 95%) identical to SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:61. In some embodiments, the polymerase comprises SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:61.

[0023] The present invention also provides for a cell comprising a recombinant expression cassette, the expression cassette comprising a promoter operably linked to a polynucleotide encoding a Sso7 polymerase conjugate proteins as described above or elsewhere herein. In some embodiments, the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is not lysine (K). In some embodiments, the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is selected from the group consisting of serine (S), threonine (T), cytosine (C), proline (P), aspartic acid (D), glutamic acid (E), asparagine (N), glutamine (Q), alanine (A), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), leucine (L), methionine (M), arginine (R), valine (V), tryptophan (W), and tyrosine (Y).

[0024] In some embodiments, the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is not arginine (R). In some embodiments, the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is selected from the group consisting of alanine (A), cytosine (C), aspartic acid (D), glutamic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L), methionine (M), asparagine (N), glutamine (Q), serine (S), threonine (T), valine (V), tryptophan (W), tyrosine (Y), and proline (P).

[0025] In some embodiments, the polymerase substantially lacks a 3'-5' exonuclease activity.

[0026] In some embodiments, the Sso7 domain is substantially (e.g., at least 60, 75, 80, 85, 90, or 95%) identical to the first 58 or 60 amino acids of SEQ ID NO:2 or to the full-length of SEQ ID NO:2. In some embodiments, the Sso7 domain is at least 90% identical to SEQ ID NO:2.

[0027] In some embodiments, the polymerase is substantially (e.g., at least 60, 75, 80, 85, 90, or 95%) identical to SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:61. In some embodiments, the polymerase comprises SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:61.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] **Figure 1.** Alignment of various Sso7 amino acid sequences (SEQ ID NOS:2, 3, 6, 4 and 5, respectively).

DEFINITIONS

[0029] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Generally, the nomenclature used herein and the laboratory procedures in cell culture, molecular genetics, organic chemistry, and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. Standard techniques are used for nucleic acid and peptide synthesis. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see generally, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated herein by reference), which are provided throughout this document. The nomenclature used herein and the laboratory procedures in analytical chemistry, and organic synthetic described below are those well known and commonly employed in the art.

[0030] The term "Sso7" or "Sso7 DNA binding domain" or "Sso7-like DNA binding domain" or "Sso7 domain" refers to nucleic acid and polypeptide polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have an amino acid sequence that has

greater than about 60% amino acid sequence identity, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or greater amino acid sequence identity to SEQ ID NO:2; (2) specifically hybridize under stringent hybridization conditions to a Sso7d nucleic acid sequence of SEQ ID NO:1 and conservatively modified variants thereof; or (3) have a nucleic acid sequence that has greater than about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher nucleotide sequence identity to SEQ ID NO:1. The term includes both full-length Sso7d polypeptides and fragments of the polypeptides that have sequence non-specific double-stranded binding activity. Sso7-like proteins include, but are not limited to, Sso7d, Sac7d and Sac7e.

[0031] "Domain" refers to a unit of a protein or protein complex, comprising a polypeptide subsequence, a complete polypeptide sequence, or a plurality of polypeptide sequences where that unit has a defined function. The function is understood to be broadly defined and can be ligand binding, catalytic activity or can have a stabilizing effect on the structure of the protein.

[0032] "Heterologous", when used with reference to portions of a protein, indicates that the protein comprises two or more domains that are not found in the same relationship to each other in nature. Such a protein, e.g., a fusion protein, contains two or more domains from unrelated proteins arranged to make a new functional protein.

[0033] "Join" refers to any method known in the art for functionally connecting protein domains, including without limitation recombinant fusion with or without intervening domains, intein-mediated fusion, non-covalent association, and covalent bonding, including disulfide bonding; hydrogen bonding; electrostatic bonding; and conformational bonding, e.g., antibody-antigen, and biotin-avidin associations.

[0034] "Thermally stable polymerase" as used herein refers to any enzyme that catalyzes polynucleotide synthesis by addition of nucleotide units to a nucleotide chain using DNA or RNA as a template and has an optimal activity at a temperature above 45°C.

[0035] "Thermus polymerase" refers to a family A DNA polymerase isolated from any Thermus species, including without limitation *Thermus aquaticus*, *Thermus brockianus*, and *Thermus thermophilus*; any recombinant polymerases deriving from Thermus species, and any functional derivatives thereof, whether derived by genetic modification or chemical modification or other methods known in the art.

[0036] The term "amplification reaction" refers to any in vitro means for multiplying the copies of a target sequence of nucleic acid. Such methods include but are not limited to polymerase chain reaction (PCR), DNA ligase chain reaction (see U.S. Pat. Nos. 4,683,195 and 4,683,202; PCR Protocols: A Guide to Methods and Applications (Innis et al., eds, 1990)), (LCR), QBeta RNA replicase, and RNA transcription-based (such as TAS and 3SR) amplification reactions as well as others known to those of skill in the art.

[0037] "Amplifying" refers to a step of submitting a solution to conditions sufficient to allow for amplification of a polynucleotide if all of the components of the reaction are intact. Components of an amplification reaction include, e.g., primers, a polynucleotide template, polymerase, nucleotides, and the like. The term "amplifying" typically refers to an "exponential" increase in target nucleic acid. However, "amplifying" as used herein can also refer to linear increases in the numbers of a select target sequence of nucleic acid, such as is obtained with cycle sequencing.

[0038] The term "amplification reaction mixture" refers to an aqueous solution comprising the various reagents used to amplify a target nucleic acid. These include enzymes, aqueous buffers, salts, amplification primers, target nucleic acid, and nucleoside triphosphates. As discussed further herein, amplification reaction mixtures may also further include stabilizers and other additives to optimize efficiency and specificity. Depending upon the context, the mixture can be either a complete or incomplete amplification reaction mixture

[0039] "Polymerase chain reaction" or "PCR" refers to a method whereby a specific segment or subsequence of a target double-stranded DNA, is amplified in a geometric progression. PCR is well known to those of skill in the art; see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202; and PCR Protocols: A Guide to Methods and Applications, Innis et al., eds, 1990. Exemplary PCR reaction conditions typically comprise either two or three step cycles. Two step cycles have a denaturation step followed by a hybridization/elongation step. Three step cycles comprise a denaturation step followed by a hybridization step followed by a separate elongation step.

[0040] The term "specificity," as used with respect to nucleic acid amplification, refers to the likelihood of a nucleic acid amplification reaction producing specific amplification products as compared to non-specific amplification products. A "specific amplification product" refers to the polynucleotide produced by amplification with correctly matched primers and template (*i.e.*, the true target sequence). A "non-specific amplification product"

refers to the polynucleotide produced by amplification with mismatched primers and template.

[0041] The phrase "improve specificity" or "improving specificity," as used with respect to nucleic acid amplification, refers to a detectable increase in the amount of specific amplification products produced in a nucleic acid amplification as compared to the amount of non-specific amplification products produced. Specificity of an amplification reaction can be measured according to any method, including but not limited to melt-curve analysis or gel analysis. In some embodiments, the specificity of a nucleic acid amplification reaction is determined by comparing the relative yield of two products, one of which is the specific product with the expected T_m and the other the non-specific product with a lower or higher T_m than that of the specific product. A reaction mixture comprising an Sso7-polymerase conjugate having an Sso7 domain with a K28 and/or R43 mutation as described herein will have a higher relative yield in an amplification reaction of the specific product than the non-specific product (*i.e.*, the ratio of the yield as measured by the height of the melting peak of the specific product over the height of the melting peak of the non-specific product) than the relative yield of the specific product over the non-specific product in a reaction mixture having a control polymerase with the wildtype Sso7 domain. In some embodiments, an Sso7-polymerase conjugate having an Sso7 domain with a K28 and/or R43 mutation as described herein will have improved amplification specificity of at least 10%, 15%, 20%, 25%, 30%, 40%, 50%, 100%, 2-fold (200%), 2.5-fold (250%), 3-fold (300%) or greater increase in the ratio relative to reactions in which a control polymerase having the wildtype Sso7 domain.

[0042] A "primer" refers to a polynucleotide sequence that hybridizes to a sequence on a target nucleic acid and serves as a point of initiation of nucleic acid synthesis. Primers can be of a variety of lengths and are often less than 50 nucleotides in length, for example 12-30 nucleotides, in length. The length and sequences of primers for use in PCR can be designed based on principles known to those of skill in the art, see, e.g., Innis et al., supra.

[0043] A "template" refers to a polynucleotide sequence that comprises the polynucleotide to be amplified, flanked by primer hybridization sites. Thus, a "target template" comprises the target polynucleotide sequence flanked by hybridization sites for a 5' primer and a 3' primer.

[0044] As used herein, "nucleic acid" means DNA, RNA, single-stranded, double-stranded, or more highly aggregated hybridization motifs, and any chemical modifications thereof. Modifications include, but are not limited to, those providing chemical groups that

incorporate additional charge, polarizability, hydrogen bonding, electrostatic interaction, points of attachment and functionality to the nucleic acid ligand bases or to the nucleic acid ligand as a whole. Such modifications include, but are not limited to, peptide nucleic acids (PNAs), phosphodiester group modifications (e.g., phosphorothioates, methylphosphonates), 2'-position sugar modifications, 5-position pyrimidine modifications, 8-position purine modifications, modifications at exocyclic amines, substitution of 4-thiouridine, substitution of 5-bromo or 5-iodo-uracil; backbone modifications, methylations, unusual base-pairing combinations such as the isobases, isocytidine and isoguanidine and the like. Nucleic acids can also include non-natural bases, such as, for example, nitroindole. Modifications can also include 3' and 5' modifications such as capping with a fluorophore (e.g., quantum dot) or another moiety.

[0045] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers.

[0046] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, .gamma.-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e.,a carbon atom that is bound to a hydrogen atom, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0047] Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0048] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

[0049] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

[0050] The term "encoding" refers to a polynucleotide sequence encoding one or more amino acids. The term does not require a start or stop codon. An amino acid sequence can be encoded in any one of six different reading frames provided by a polynucleotide sequence.

[0051] The term "promoter" refers to regions or sequence located upstream and/or downstream from the start of transcription and which are involved in recognition and binding of RNA polymerase and other proteins to initiate transcription.

[0052] A "vector" refers to a polynucleotide, which when independent of the host chromosome, is capable replication in a host organism. Preferred vectors include plasmids and typically have an origin of replication. Vectors can comprise, e.g., transcription and

translation terminators, transcription and translation initiation sequences, and promoters useful for regulation of the expression of the particular nucleic acid.

[0053] "Recombinant" refers to a human manipulated polynucleotide or a copy or complement of a human manipulated polynucleotide. For instance, a recombinant expression cassette comprising a promoter operably linked to a second polynucleotide may include a promoter that is heterologous to the second polynucleotide as the result of human manipulation (e.g., by methods described in Sambrook et al., Molecular Cloning—A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1989) or Current Protocols in Molecular Biology Volumes 1-3, John Wiley & Sons, Inc. (1994-1998)) of an isolated nucleic acid comprising the expression cassette. In another example, a recombinant expression cassette may comprise polynucleotides combined in such a way that the polynucleotides are extremely unlikely to be found in nature. For instance, human manipulated restriction sites or plasmid vector sequences may flank or separate the promoter from the second polynucleotide. One of skill will recognize that polynucleotides can be manipulated in many ways and are not limited to the examples above.

[0054] A "polymerase" refers to an enzyme that performs template-directed synthesis of polynucleotides, e.g., DNA and/or RNA. The term encompasses both the full length polypeptide and a domain that has polymerase activity. DNA polymerases are well-known to those skilled in the art, including but not limited to DNA polymerases isolated or derived from *Pyrococcus furiosus*, *Thermococcus litoralis*, and *Thermotoga maritime*, or modified versions there of. They include both DNA-dependent polymerases and RNA-dependent polymerases such as reverse transcriptase. At least five families of DNA-dependent DNA polymerases are known, although most fall into families A, B and C. There is little or no sequence similarity among the various families. Most family A polymerases are single chain proteins that can contain multiple enzymatic functions including polymerase, 3' to 5' exonuclease activity and 5' to 3' exonuclease activity. Family B polymerases typically have a single catalytic domain with polymerase and 3' to 5' exonuclease activity, as well as accessory factors. Family C polymerases are typically multi-subunit proteins with polymerizing and 3' to 5' exonuclease activity. In E. coli, three types of DNA polymerases have been found, DNA polymerases I (family A), II (family B), and III (family C). In eukaryotic cells, three different family B polymerases, DNA polymerases α, δ, and ε, are implicated in nuclear replication, and a family A polymerase, polymerase γ, is used for mitochondrial DNA replication. Other types of DNA polymerases include phage polymerases. Similarly, RNA polymerases typically include eukaryotic RNA polymerases I,

II, and III, and bacterial RNA polymerases as well as phage and viral polymerases. RNA polymerases can be DNA-dependent and RNA-dependent.

[0055] Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids 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 over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, Computer Applic. Biol. Sci. 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif., USA).

[0056] Sequences are "substantially identical" to each other if they have a specified percentage of nucleotides or amino acid residues that are the same (*e.g.*, at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection.

[0057] For sequence comparison, 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 entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program

parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0058] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. 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, Accelrys), or by manual alignment and visual inspection.

[0059] Algorithms suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (*Nuc. Acids Res.* 25:3389-402, 1977), and Altschul *et al.* (*J. Mol. Biol.* 215:403-10, 1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). 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 extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted 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 BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

[0060] The BLAST algorithm also 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 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 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0061] The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30°C. below the T_m. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Hybridization conditions are typically those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C. for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C. for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization.

[0062] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl,

1% SDS at 37°C., and a wash in 1X SSC at 45°C. Exemplary "stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C., and at least one wash in 0.2X SSC at a temperature of at least about 50°C., usually about 55°C. to about 60°C., for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[0063] The current invention provides variant Sso7 polymerase conjugates that exhibit reduced non-specific amplification activity compared to Sso7 polymerase conjugates comprising wildtype Sso7 domains. The variant Sso7 polymerase conjugates comprise at least a polymerase domain joined to a variant Sso7 binding domain. The variant Sso7 DNA-binding domains comprise an amino acid change at one or both of two positions (corresponding to K28 and/or R43 in SEQ ID NO:2) as compared to the wildtype amino acid at that position. As demonstrated in the Examples, a number of different amino acid substitutions from wildtype at these positions have been shown to reduce non-specific polymerase activity. Non-specific polymerase activity includes amplification of a nucleic acid that occur other than due to template-specific extension of a hybridized primer.

II. Sso7 Domains of the Invention

[0064] The polymerase conjugates of the invention comprise an Sso7 polypeptide sequence that has amino acid substitutions compared to the native (wildtype) Sso7d sequence. Sso7d is a small (63 amino acids, about 7 kd MW), basic chromosomal protein from the hyperthermophilic archaeabacteria *Sulfolobus solfataricus*. The protein is lysine-rich and has high thermal, acid and chemical stability. It binds to DNA in a sequence-independent manner and when bound, increase the Tm of DNA by up to 40° C under some conditions (McAfee et al., Biochemistry 34:10063-10077, 1995). Sso7d and its homologs are typically believed to be involved in packaging genomic DNA and stabilizing genomic DNA at elevated temperatures. The Sso7d protein sequence is set forth in SEQ ID NO:2.

[0065] There are several known Sso7d-like proteins (also referred to as "Sso7 proteins") including, but not limited to, Sac7a, Sac7b, Sac7d (SEQ ID NO:4), Sac7e (SEQ ID NO:5), from the hyperthermophilic archaeabacteria *Sulfolobus acidocaldarius*; Sto7e (SEQ ID NO:6) from *Sulfolobus tokodaii*, and Ssh7a and Ssh7b (SEQ ID NO:3) from *Sulfolobus shibatae* (*see, e.g.,* UniProt database accession numbers: P39476 (Sso7d); O59632 (Ssh7b); P13123 (Sac7d); P13125 (Sac7e); and Q96X56 (Sto7e)). These proteins have an identity with Sso7d that ranges from 78% to 98%. Other Sso7 domains for use in the invention may also be identified as set forth below.

[0066] As noted herein, the invention provides for amino acid changes from the native amino acid at the positions corresponding to K28 and/or R43 of SEQ ID NO:2. It should be understood that such position designations do not indicate the number of amino acids in the claimed molecule per se, but indicate where in the claimed molecule the residue occurs when the claimed molecule sequence is maximally aligned with SEQ ID NO:2. In the context of variant Sso7 domains, "correspondence" to a Sso7-like protein sequence is based on the convention of numbering according to amino acid position number of a particular sequence (i.e., SEQ ID NO:2) and then aligning the Sso7-like protein sequence in a manner that maximizes the percentage of sequence identity to SEQ ID NO:2. Alignment can be performed either manually or using a sequence comparison algorithm (e.g., using the NCBI BLAST program with default parameters (*see, e.g.,* Altschul *et al., Nucl. Acids Res.* 25:3389-3402, 1997). An example of an alignment of several Sso7-like proteins with SEQ ID NO:2 is shown in Figure 1. The corresponding sequences can be summarized as follows:

	Actual position of amino acid corresponding to K28 of SEQ ID NO:2	**Actual position of amino acid corresponding to R43 of SEQ ID NO:2**
Sso7 (SEQ ID NO:2)	28	43
Ssh7b (SEQ ID NO:3)	28	43
Sto7e (SEQ ID NO:4)	28	42
Sac7d (SEQ ID NO:5)	28	42
Sac7e (SEQ ID NO:6)	28	42

[0067] Any Sso7 DNA binding protein domain can be substituted at the K28 and/or R43 position corresponding to SEQ ID NO:2. Thus, for example, in some embodiments, the variant Sso7 polypeptide sequence is substantially (e.g., at least 60, 70, 80, 85, 90, or 95%) identical to any of, e.g., SEQ ID NOS:2, 3, 4, 5, or 6, and comprises an amino acid other than K at the amino acid position corresponding to K28. In some embodiments, the amino acid position corresponding to K28 is serine (S), threonine (T), cytosine (C), proline (P), aspartic acid (D), glutamic acid (E), asparagine (N), glutamine (Q), alanine (A), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), leucine (L), methionine (M), arginine (R), valine (V), tryptophan (W), or tyrosine (Y).

[0068] In some embodiments, the variant Sso7 polypeptide sequence is substantially (e.g., at least 60, 70, 80, 85, 90, or 95%) identical to any of, e.g., SEQ ID NOS:2, 3, 4, 5, or 6, and comprises an amino acid other than R at the amino acid position corresponding to R43. In some embodiments, the amino acid position corresponding to R43 is alanine (A), cytosine (C), aspartic acid (D), glutamic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L), methionine (M), asparagine (N), glutamine (Q), serine (S), threonine (T), valine (V), tryptophan (W), tyrosine (Y), or proline (P).

[0069] In some embodiments, the variant Sso7 polypeptide sequence is substantially (e.g., at least 60, 70, 80, 85, 90, or 95%) identical to any of, e.g., SEQ ID NOS:2, 3, 4, 5, or 6, and comprises an amino acid other than K at the amino acid position corresponding to K28 and an amino acid other than R at the amino acid position corresponding to R43. For example, in some embodiments, the amino acid at position K28 is selected from: serine (S), threonine (T), cytosine (C), proline (P), aspartic acid (D), glutamic acid (E), asparagine (N), glutamine (Q), alanine (A), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), leucine (L), valine (V), tryptophan (W), or tyrosine (Y) and the amino acid at position R43 is selected from: alanine (A), cytosine (C), aspartic acid (D), glutamic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L), methionine (M), asparagine (N), glutamine (Q), serine (S), threonine (T), valine (V), tryptophan (W), tyrosine (Y), or proline (P).

[0070] There are many ways of generating these alterations or variants of a given nucleic acid sequence. Well-known methods include, e.g., site-directed mutagenesis, PCR amplification using degenerate oligonucleotides, chemical synthesis of a desired oligonucleotide (e.g., in conjunction with ligation and/or cloning to generate large nucleic

acids) and other well-known techniques. See, Giliman & Smith, Gene 8:81-97 (1979), Roberts, et al., Nature 328:731-734 (1987) and Sambrook, Innis, and Ausubel (all supra).

[0071] In some embodiments, the variant Sso7 domain or polypeptide sequence is identical or substantially identical (*e.g.*, has at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity) to an amino acid sequence of any of SEQ ID NOS:7-32. In some embodiments, the Sso7 domain or protein has the amino acid sequence of any of SEQ ID NOS:7-32.

[0072] In one example of generating an Sso7 sequence of the invention, site directed mutagenesis is used to substitute an amino acid residue. The nucleic acid sequence is substituting by synthesizing an oligonucleotide primer that contains the mutation. In some embodiments, the primer is hybridized to an Sso7 nucleic acid, e.g., SEQ ID NO:1, and a new sequence amplified. The amplification product with the mutation may then ligated into an expression vector.

[0073] In some embodiments, polypeptide sequences are altered as above, i.e., by changing the corresponding nucleic acid sequence and expressing the polypeptide. However, polypeptide sequences can also be generated synthetically using commercially available peptide synthesizers to produce a desired polypeptide (see, Merrifield, and Stewart & Young, supra).

[0074] Finally, the substituted Sso7 sequences are evaluated by using techniques such as those described herein to identify the fusion polymerases that exhibit reduced non-specific polymerase activity.

III. Polymerases

[0075] In general it is believed that essentially any polymerase can be linked to a variant Sso7 protein sequence of the invention to form the Sso7-polymerase conjugate protein of the invention, thereby improving various activities of the polymerase. In one exemplary embodiment, the Sso7 polymerase conjugate protein comprises a polymerase domain derived from two parental polymerases, Pfu and DeepVent. Such polymerases are described for example in U.S. Application Publication Nos. 20040219558; 20040214194; 20040191825; 20030162173, each of which is hereby incorporated by reference in its entirety for all purposes and in particular for all teachings related to hybrid polymerases. In some

embodiments, the polymerase is substantially (e.g., at least 60, 70, 80, 85, 90, or 95%) identical to SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:61.

[0076] A variety of polymerases can be used as at least a portion of the polymerase domain of polymerase. At least five families of DNA-dependent DNA polymerases are known, although most fall into families A, B and C. There is little or no structural or sequence similarity among the various families. Most family A polymerases are single chain proteins that can contain multiple enzymatic functions including polymerase, 3' to 5' exonuclease activity and 5' to 3' exonuclease activity. Family B polymerases typically have a single catalytic domain with polymerase and 3' to 5' exonuclease activity, as well as accessory factors. Family C polymerases are typically multi-subunit proteins with polymerizing and 3' to 5' exonuclease activity. In *E. coli*, three types of DNA polymerases have been found, DNA polymerases I (family A), II (family B), and III (family C). In eukaryotic cells, three different family B polymerases, DNA polymerases α, δ, and ε are implicated in nuclear replication, and a family A polymerase, polymerase γ, is used for mitochondrial DNA replication. Other types of DNA polymerases include phage polymerases. Any of these polymerases, combinations of all or portions of these polymerases, as well as chimeras or hybrids between two or more of such polymerases or their equivalents can be used to form a portion or all of the polymerase domain of Sso7 polymerase conjugate proteins of the invention.

[0077] In one exemplary embodiment, the polymerase conjugates of the invention have a polymerase domain derived from two parental polymerases, Pfu and DeepVent. Such polymerases are described for example in U.S. Application Publication Nos. 20040219558; 20040214194; 20040191825; 20030162173, each of which is hereby incorporated by reference in its entirety for all purposes and in particular for all teachings related to hybrid polymerases. In some embodiments, the polymerase is substantially (e.g., at least 60, 70, 80, 85, 90, or 95%) identical to SEQ ID NO:34 (optionally including a linker such as SEQ ID NO:62) or SEQ ID NO:33 (optionally including a linker such as SEQ ID NO:62) or SEQ ID NO:61.

[0078] Further, in some embodiments, non-thermostable polymerases may also be used in accordance with the invention. For example, the large fragment of E. coli DNA Polymerase I (Klenow) (the Klenow Fragment) with mutation (D355A, E357A) abolishes the 3'→5' exonuclease activity. This enzyme or equivalent enzymes can be used, for example, in embodiments where the amplification reaction is not performed at high temperatures.

[0079] In some embodiments, the hybrid polymerases of the invention include a polymerase domain comprising mutations that reduce or abolish exonuclease activity of any hybrid polymerase comprising such a polymerase domain in comparison to a Sso7 polymerase conjugate protein comprising a polymerase domain that does not have such mutations. A variety of mutations can be introduced into a native polymerase domain to reduce or eliminate 3'-5' exonuclease activity. For example, U.S. Pat. Nos. 6,015,668; 5,939,301 and 5,948,614 describe mutations of a metal-binding aspartate to an alanine residue in the 3'-5' exonuclease domain of the Tma and Tne DNA polymerases. These mutations reduce the 3'-5' exonuclease activities of these enzymes to below detectable levels. Similarly, U.S. Pat. No. 5,882,904 describes an analogous aspartate-to-alanine mutation in *Thermococcus barossi*, and U.S. Pat. No. 5,489,523 teaches the double-mutant D141A E143A of the *Pyrococcus wosei* DNA polymerases. Both of these mutant polymerases have virtually no detectable 3'-5' exonuclease activity. Methods of assaying 3'-5' exonuclease activity are well-known in the art. *See, e.g.*, Freemont *et al.*, *Proteins* 1:66 (1986); Derbyshire et al., *EMBO J.* 16:17 (1991) and Derbyshire *et al.*, *Methods in Enzymology* 262:363 85 (1995). It will be understood that while the above-described mutations were originally identified in one polymerase, one can generally introduce such mutations into other polymerases to reduce or eliminate exonuclease activity. In a specific embodiment, **a** polymerases of the invention comprise the double point mutation D141A/E143A in the polymerase domain. The phrase "corresponding to a position," in reference to polymerase amino acids, refers to an amino acid that aligns with the same amino acid (*e.g.*, D141 or E143) in a reference polymerase amino acid sequence (e.g., SEQ ID NO:2, optionally including a linker such as SEQ ID NO:62). Sequence comparisons can be performed using any BLAST including BLAST 2.2 algorithm with default parameters, described in Altschul *et al.*, *Nuc. Acids Res.* 25:3389 3402 (1977) and Altschul *et al.*, *J. Mol. Biol.* 215:403 410 (1990), respectively. Thus, in some embodiments, Sso7 polymerase conjugate proteins comprise mutations in the exonuclease domain. For example, in some embodiments, the Sso7d polymerase conjugate is substantially (e.g., at least 80%, 85%, 90%, 95%, 98% or 100%) identical to SEQ ID NO:61, comprises one or more mutation reducing exonuclease activity while substantially retaining polymerase activity, and is conjugated to an Sso7d DNA binding domain as described herein. In still further embodiments, such Sso7 polymerase conjugate proteins comprise an amino acid sequence according to SEQ ID NO:33 (optionally including a linker such as SEQ ID NO:62).

IV. Conjugates

[0080] The Sso7 DNA binding domain and the polymerase can be joined to form a Sso7-polymerase conjugate by methods well known to those of skill in the art. These methods include both chemical and recombinant means.

[0081] Chemical joining the Sso7 protein to the polymerase can be performed, for example as described in Bioconjugate Techniques, Hermanson, Ed., Academic Press (1996). Joining can include, for example, derivitization for the purpose of linking the two proteins to each other, either directly or through a linking compound, by methods that are well known in the art of protein chemistry. For example, in one chemical conjugation embodiment, the means of linking the catalytic domain and the nucleic acid binding domain comprises a heterobifunctional-coupling reagent which ultimately contributes to formation of an intermolecular disulfide bond between the two moieties. Other types of coupling reagents that are useful in this capacity for the present invention are described, for example, in U.S. Patent 4,545,985. Alternatively, an intermolecular disulfide may conveniently be formed between cysteines in each moiety, which occur naturally or are inserted by genetic engineering. The means of linking moieties may also use thioether linkages between heterobifunctional crosslinking reagents or specific low pH cleavable crosslinkers or specific protease cleavable linkers or other cleavable or noncleavable chemical linkages.

[0082] The means of linking the Sso7 and polymerase domains of the conjugate protein may also comprise a peptidyl bond formed between moieties that are separately synthesized by standard peptide synthesis chemistry or recombinant means. The conjugate protein itself can also be produced using chemical methods to synthesize an amino acid sequence in whole or in part. For example, peptides can be synthesized by solid phase techniques, such as, e.g., the Merrifield solid phase synthesis method, in which amino acids are sequentially added to a growing chain of amino acids (see, Merrifield (1963) J. Am. Chem. Soc., 85:2149-2146). Equipment for automated synthesis of polypeptides is commercially available from suppliers such as PE Corp. (Foster City, CA), and may generally be operated according to the manufacturer's instructions. The synthesized peptides can then be cleaved from the resin, and purified, e.g., by preparative high performance liquid chromatography (see Creighton, Proteins Structures and Molecular Principles, 50-60 (1983)). The composition of the synthetic polypeptides or of subfragments of the polypeptide, may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; see Creighton, Proteins, Structures and Molecular Principles, pp. 34-49 (1983)).

[0083] In addition, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the sequence. Non-classical amino acids include, but are not limited to, the D-isomers of the common amino acids, α-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, γ-Abu, ε-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxy-proline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, fluoro-amino acids, designer amino acids such as β-methyl amino acids, Cα-methyl amino acids, Nα-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

[0084] In another embodiment, the Sso7 and polymerase domains are joined via a linking group. The linking group can be a chemical crosslinking agent, including, for example, succinimidyl-(N-maleimidomethyl)-cyclohexane-1-carboxylate (SMCC). The linking group can also be an additional amino acid sequence(s), including, for example, a polyalanine, polyglycine or similarly, linking group.

[0085] Alternatively, in some embodiments, the coding sequences of each polypeptide in the fusion protein are directly joined at their amino- or carboxy-terminus via a peptide bond in any order. Alternatively, an amino acid linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such an amino acid linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Typical peptide linker sequences contain Gly, Ser, Val and Thr residues. Other near neutral amino acids, such as Ala can also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al. (1985) Gene 40:39-46; Murphy et al. (1986) Proc. Natl. Acad. Sci. USA 83:8258-8262; U.S. Patent Nos. 4,935,233 and 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length, e.g., 3, 4, 6, or 10 amino acids in length, but can be 100 or 200 amino acids in length. Linker sequences may not be required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference. An exemplary peptide linker is shown in SEQ ID NO:62.

[0086] Exemplary conjugates of the invention include those that comprise a polypeptide identical, or substantially (e.g., at least 60, 70, 80, 85, 90, or 95%) identical to any of SEQ ID NOS:35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60 so long as the conjugates include the non-native K28 or R43 amino acid in the sequence.

[0087] Other chemical linkers include carbohydrate linkers, lipid linkers, fatty acid linkers, polyether linkers, e.g., PEG, etc. For example, poly(ethylene glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterobifunctional linkages.

[0088] Other methods of joining the Sso7 and polymerase domains include ionic binding by expressing negative and positive tails and indirect binding through antibodies and streptavidin-biotin interactions. (See, e.g., Bioconjugate Techniques, supra). The domains may also be joined together through an intermediate interacting sequence. For example, an Sso7d-interacting sequence, i.e., a sequence that binds to Sso7d, can be joined to a polymerase. The resulting fusion protein can then be allowed to associate non-covalently with the Sso7d to generate an Sso7d-polymerase conjugate.

[0089] In some embodiments, a conjugate Sso7-polymerase protein of the invention is produced by recombinant expression of a nucleic acid encoding the protein. Such a fusion product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the product by methods known in the art.

[0090] Nucleic acids encoding the domains to be incorporated into the fusion proteins of the invention can be obtained using routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook and Russell, Molecular Cloning, A Laboratory Manual (3rd ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 1994-1999).

[0091] Nucleic acid sequences that encode the Sso7 and polymerase polypeptides can be obtained using any of a variety of methods. In some embodiments, the nucleic acid sequences encoding the polypeptides are cloned from cDNA and genomic DNA libraries by hybridization with probes, or isolated using amplification techniques with oligonucleotide primers. More commonly, amplification techniques are used to amplify and isolate the Sso7 and polymerase sequences using a DNA or RNA template (see, e.g., Dieffenfach & Dveksler,

PCR Primers: A Laboratory Manual (1995)). Alternatively, overlapping oligonucleotides can be produced synthetically and joined to produce one or more of the domains. Nucleic acids encoding catalytic or double-stranded nucleic acid binding domains can also be isolated from expression libraries using antibodies as probes.

[0092] In an example of obtaining a nucleic acid encoding an Sso7 or polymerase domain using PCR, the nucleic acid sequence or subsequence is PCR amplified, using a sense primer containing one restriction site and an antisense primer containing another restriction site. This will produce a nucleic acid encoding the desired domain sequence or subsequence and having terminal restriction sites. This nucleic acid can then be ligated into a vector containing a nucleic acid encoding the second domain and having the appropriate corresponding restriction sites. The domains can be directly joined or may be separated by a linker, or other, protein sequence. Suitable PCR primers can be determined by one of skill in the art using the sequence information provided in GenBank or other sources. Appropriate restriction sites can also be added to the nucleic acid encoding the protein or protein subsequence by site-directed mutagenesis. The plasmid containing the domain-encoding nucleotide sequence or subsequence is cleaved with the appropriate restriction endonuclease and then ligated into an appropriate vector for amplification and/or expression according to standard methods.

[0093] One of skill will also recognize that modifications can additionally be made to the Sso7 and polymerase domains without diminishing their biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of a domain into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, the addition of codons at either terminus of the polynucleotide that encodes the binding domain to provide, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (e.g., poly His) placed on either terminus to create conveniently located restriction sites or termination codons or purification sequences.

[0094] One or more of the domains may also be modified to facilitate the linkage of the two domains to obtain the polynucleotides that encode the fusion polypeptides of the invention. Thus, Sso7 and polymerase domains that are modified by such methods are also part of the invention. For example, a codon for a cysteine residue can be placed at either end of a domain so that the domain can be linked by, for example, a sulfide linkage. The modification can be performed using either recombinant or chemical methods (see, e.g., Pierce Chemical Co. catalog, Rockford IL).

[0095] The Sso7 and polymerase domains of the recombinant fusion protein can be joined by linker domains, usually polypeptide sequences including Gly, Ser, Ala, and Val such as those described above. In some embodiments, proline residues are incorporated into the linker to prevent the formation of significant secondary structural elements by the linker.

[0096] In some embodiments, the recombinant nucleic acids the recombinant nucleic acids encoding the proteins of the invention are modified to provide preferred codons which enhance translation of the nucleic acid in a selected organism (e.g., yeast preferred codons are substituted into a coding nucleic acid for expression in yeast).

[0097] A variety of expression systems known to those of ordinary skill in the art can be used to express the protein conjugates of the invention. In some embodiments, the polynucleotide that encodes the fusion polypeptide is placed under the control of a promoter that is functional in the desired host cell. An extremely wide variety of promoters are available, and can be used in the expression vectors of the invention, depending on the particular application. Ordinarily, the promoter selected depends upon the cell in which the promoter is to be active. Other expression control sequences such as ribosome binding sites, transcription termination sites and the like are also optionally included. Constructs that include one or more of these control sequences are termed "expression cassettes." Accordingly, the nucleic acids that encode the joined polypeptides are incorporated for high level expression in a desired host cell.

[0098] Expression control sequences that are suitable for use in a particular host cell are often obtained by cloning a gene that is expressed in that cell. Commonly used prokaryotic control sequences, which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (lac) promoter systems (Change et al., Nature (1977) 198: 1056), the tryptophan (trp) promoter system (Goeddel et al., Nucleic Acids Res. (1980) 8: 4057), the tac promoter (DeBoer, et al., Proc. Natl. Acad. Sci. U.S.A. (1983) 80:21 25); and the lambda-derived PL promoter and N-gene ribosome binding site (Shimatake et al., Nature (1981) 292: 128). The particular promoter system is not critical to the invention, any available promoter that functions in prokaryotes can be used. Standard bacterial expression vectors include plasmids such as pBR322-based plasmids, e.g., pBLUESCRIPT, pSKF, pET23D, and fusion expression systems such as GST and LacZ. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation.

[0099] For expression of fusion polypeptides in prokaryotic cells other than E. coli, a promoter that functions in the particular prokaryotic species is required. Such promoters can be obtained from genes that have been cloned from the species, or heterologous promoters can be used. For example, the hybrid trp-lac promoter functions in Bacillus in addition to E. coli. These and other suitable bacterial promoters are well known in the art and are described, e.g., in Sambrook et al. and Ausubel et al. Bacterial expression systems for expressing the proteins of the invention are available in, e.g., *E. coli*, *Bacillus* sp., and *Salmonella* (Palva et al., Gene 22:229-235 (1983); Mosbach et al., Nature 302:543-545 (1983). Kits for such expression systems are commercially available.

[0100] Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available. In yeast, vectors include Yeast Integrating plasmids (e.g., YIp5) and Yeast Replicating plasmids (the YRp series plasmids) and pGPD-2. Expression vectors containing regulatory elements from eukaryotic viruses are typically used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the CMV promoter, SV40 early promoter, SV40 later promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

[0101] Either constitutive or regulated promoters can be used in the present invention. Regulated promoters can be advantageous because the host cells can be grown to high densities before expression of the fusion polypeptides is induced. High level expression of heterologous proteins slows cell growth in some situations. An inducible promoter is a promoter that directs expression of a gene where the level of expression is alterable by environmental or developmental factors such as, for example, temperature, pH, anaerobic or aerobic conditions, light, transcription factors and chemicals.

[0102] For *E. coli* and other bacterial host cells, inducible promoters are known to those of skill in the art. These include, for example, the lac promoter, the bacteriophage lambda PL promoter, the hybrid trp-lac promoter (Amann et al. (1983) Gene 25: 167; de Boer et al. (1983) Proc. Nat'l. Acad. Sci. USA 80: 21), and the bacteriophage T7 promoter (Studier et al. (1986) J. Mol. Biol.; Tabor et al. (1985) Proc. Nat'l. Acad. Sci. USA 82: 1074-8). These promoters and their use are discussed in Sambrook et al., supra.

[0103] Inducible promoters for other organisms are also well known to those of skill in the art. These include, for example, the metallothionein promoter, the heat shock promoter, as well as many others.

[0104] Translational coupling may be used to enhance expression. The strategy uses a short upstream open reading frame derived from a highly expressed gene native to the translational system, which is placed downstream of the promoter, and a ribosome binding site followed after a few amino acid codons by a termination codon. Just prior to the termination codon is a second ribosome binding site, and following the termination codon is a start codon for the initiation of translation. The system dissolves secondary structure in the RNA, allowing for the efficient initiation of translation. See Squires, et. al. (1988), J. Biol. Chem. 263: 16297-16302.

[0105] The construction of polynucleotide constructs generally requires the use of vectors able to replicate in bacteria. Such vectors are commonly used in the art. A plethora of kits are commercially available for the purification of plasmids from bacteria (for example, EasyPrepJ, FlexiPrepJ, from Pharmacia Biotech; StrataCleanJ, from Stratagene; and, QIAexpress Expression System, Qiagen). The isolated and purified plasmids can then be further manipulated to produce other plasmids, and used to transform cells.

[0106] The fusion polypeptides can be expressed intracellularly, or can be secreted from the cell. Intracellular expression often results in high yields. If necessary, the amount of soluble, active fusion polypeptide may be increased by performing refolding procedures (see, e.g., Sambrook et al., supra.; Marston et al., Bio/Technology (1984) 2: 800; Schoner et al., Bio/Technology (1985) 3: 151). Fusion polypeptides of the invention can be expressed in a variety of host cells, including E. coli, other bacterial hosts, yeast, and various higher eukaryotic cells such as the COS, CHO and HeLa cells lines and myeloma cell lines. The host cells can be mammalian cells, insect cells, or microorganisms, such as, for example, yeast cells, bacterial cells, or fungal cells.

[0107] Once expressed, the recombinant fusion polypeptides can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (see, generally, 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)). Substantially pure compositions of at least about 90 to 95% homogeneity are preferred, and 98 to 99% or more

homogeneity are most preferred. Once purified, partially or to homogeneity as desired, the polypeptides may then be used (e.g., as immunogens for antibody production).

[0108] To facilitate purification of the fusion polypeptides of the invention, the nucleic acids that encode the fusion polypeptides can also include a coding sequence for an epitope or "tag" for which an affinity binding reagent is available. Examples of suitable epitopes include the myc and V-5 reporter genes; expression vectors useful for recombinant production of fusion polypeptides having these epitopes are commercially available (e.g., Invitrogen (Carlsbad CA) vectors pcDNA3.1/Myc-His and pcDNA3.1/V5-His are suitable for expression in mammalian cells). Additional expression vectors suitable for attaching a tag to the fusion proteins of the invention, and corresponding detection systems are known to those of skill in the art, and several are commercially available (e.g., FLAG" (Kodak, Rochester NY). Another example of a suitable tag is a polyhistidine sequence, which is capable of binding to metal chelate affinity ligands. Typically, six adjacent histidines (SEQ ID NO:63) are used, although one can use more or less than six. Suitable metal chelate affinity ligands that can serve as the binding moiety for a polyhistidine tag include nitrilo-tri-acetic acid (NTA) (Hochuli, E. (1990) "Purification of recombinant proteins with metal chelating adsorbents" In Genetic Engineering: Principles and Methods, J.K. Setlow, Ed., Plenum Press, NY; commercially available from Qiagen (Santa Clarita, CA)).

V. Methods

[0109] As discussed herein, the present invention provides conjugate proteins for use in nucleic acid amplification reactions. Such amplification reactions include without limitation polymerase chain reaction (PCR), DNA ligase chain reaction (LCR), QBeta RNA replicase, and RNA transcription-based (such as TAS and 3SR) amplification reactions as well as others known to those of skill in the art. Polymerase chain reactions that can be conducted using the compositions described herein include without limitation reverse-transcription PCR (rt-PCR) and quantitative PCR (qPCR).

[0110] As will be appreciated, any combination of the different components described herein is encompassed by the present invention, as are amplification reactions utilizing any combination of different components of the invention. For example, amplification reactions of the invention may utilize Sso7 polymerase conjugates comprising one or mutations that remove or completely abolish exonuclease activity, particularly 3'-5' exonuclease activity. Such amplification reactions may further utilize such mutant hybrid polymerases combined

with hot-start antibodies. Some amplification reactions of the invention may utilize mutant hybrid polymerases comprising the D141A/E143A double point mutation combined with additives such as sarcosine or heparin. Some amplification reactions of the invention may also utilize mutant hybrid polymerases lacking 3'-5' exonuclease activity combined with hot-start antibodies and with additives such as sarcosine or heparin.

[0111] In some embodiments, dye-based qPCR detection methods are used to monitor amplification reactions utilizing components of the invention. Such detection methods generally rely on monitoring the increase in fluorescence signal due to the binding of DNA-binding dye to the amplified DNA. For example, SYBR Green I, a commonly used fluorescent DNA binding dye, binds all double–stranded DNA and detection is monitored by measuring the increase in fluorescence throughout the cycle.

[0112] In other embodiments, probe-based qPCR detection methods are used to monitor amplification reactions utilizing components of the invention. Such detection methods generally rely on the sequence-specific detection of a desired PCR product. Unlike dye-based qPCR methods that detect all double-stranded DNA, probe-based qPCR utilizes a fluorescent-labeled target-specific probe, which detects specific sequences in the amplified DNA.

VI. Reaction Mixtures

[0113] The present invention also provides reaction mixtures comprising the Sso7-polymerase conjugates of the invention. Optionally, an antibody is complexed with the polymerase. The reaction mixtures can optionally comprise one or more dNTPs, one or more oligonucleotides, a biological sample comprising a target nucleic acid, and/or a double stranded DNA binding dye or other reagent useful for PCR or primer extension reactions.

[0114] In certain aspects, it may be desirable to include an additional compound as an additive to improve efficiency in amplification reactions, including but not limited to qPCR. In some embodiments, inclusion of the additive is sufficient to increase efficiency of the polymerase by at least 5, 10, 15, 20, 25, 35, 40, or 50% or more compared to a control mixture lacking the additive.

[0115] In some embodiments, the additive is an osmolyte included in an amplification reaction of the invention to improve efficiency. Members of the osmolyte family have been shown to improve the thermal stability of proteins (Santoro, Biochemistry, 1992) as well as decrease DNA double helix stability (Chadalavada, FEBS Letters, 1997). In some

embodiments, osmolytes are small molecules or compounds which are produced by living organisms in response to environmental stresses such as extreme temperatures, dehydration, or salinity and which protect their cellular components and help to maintain optimal cytosolic conditions. Osmolytes of use in the present invention may include without limitation sarcosine, trimethylamine N-oxide (TMAO), dimethylsulfoniopropionate, and trimethylglycine.

[0116] In some embodiments, concentrations of about 100 to about 1000 mM of osmolytes are used in methods and kits of the present invention. In still further embodiments, concentrations of about 50 to about 700, about 100 to about 600, about 150 to about 500, about 200 to about 400 mM, and about 300 to about 350 mM osmolytes are used in methods and kits of the invention. In some embodiments, the osmolyte used in methods, reaction mixtures, and kits of the invention is sarcosine (optionally at the above-listed concentrations).

[0117] In some embodiments, particularly in the amplification of low-copy target nucleic acids, efficiency decreases due to the binding of polymerase to non-primed double-stranded nucleic acid targets. Binding of the polymerase to the double-stranded targets will prevent those targets from denaturation, hybridizing to primers, and undergoing an amplification reaction. To improve the specificity of the polymerase for primed templates, in some embodiments reaction mixtures of the invention utilize heparin. Heparin molecules, which are negatively charged, can be included in the reaction mixture to mimic the electrostatic property of double stranded nucleic acids. The addition of heparin can, without being limited to a mechanism of action, prevent excess polymerase from binding to the double-stranded template until a single-stranded primed-template becomes available. In some exemplary embodiments, heparin is used in methods and kits of the invention at concentrations of about 50 to about 750 pg/µl. In further exemplary embodiments, heparin is used in methods and kits of the invention at concentrations of about 75 to about 700, about 100 to about 600, about 125 to about 500, about 150 to about 400, about 175 to about 300, and about 200 to about 250 pg/µl. Other molecules known in the art can be used in a similar manner to prevent non-specific binding of the polymerase to non-primed double-stranded template.

[0118] In some embodiments, the reaction mixtures comprise an agent that improves the specificity of nucleic acid amplification when added to an amplification reaction mixture prior to amplification of the target nucleic acid molecule. In some embodiments, the agent is selected from arginine (e.g., L-arginine or D-arginine), spermidine, and spermine. In some embodiments, the agent is arginine.

[0119] In some embodiments, the agent that improves the specificity of nucleic acid amplification is present in the amplification reaction mixture at a concentration of about 1 mM to about 500 mM. In some embodiments, the agent is present at a concentration of about 1 mM to about 100 mM, about 1 mM to about 75 mM, about 1 mM to about 50 mM, about 1 mM to about 25 mM, or about 5 mM to about 15 mM.

[0120] The arginine, spermidine, and/or spermine agents of the present invention may be provided as salts. Examples of applicable salt forms include hydrochlorides, hydrobromides, sulfates, methanesulfonates, nitrates, maleates, acetates, citrates, fumarates, tartrates (e.g., (+)-tartrates, (-)-tartrates or mixtures thereof including racemic mixtures), succinates, and benzoates. These salts may be prepared by methods known to those skilled in art. Also included are base addition salts such as sodium, potassium, calcium, ammonium, organic amino, or magnesium salt, or a similar salt. When an agent of the present invention contains relatively basic functionalities, acid addition salts can be obtained by contacting the neutral form of such compounds with a sufficient amount of the desired acid, either neat or in a suitable inert solvent. Examples of acceptable acid addition salts include those derived from inorganic acids like hydrochloric, hydrobromic, nitric, carbonic, monohydrogencarbonic, phosphoric, monohydrogenphosphoric, dihydrogenphosphoric, sulfuric, monohydrogensulfuric, hydriodic, or phosphorous acids and the like, as well as the salts derived organic acids like acetic, propionic, isobutyric, maleic, malonic, benzoic, succinic, suberic, fumaric, lactic, mandelic, phthalic, benzenesulfonic, p-tolylsulfonic, citric, tartaric, methanesulfonic, and the like. In some embodiments, arginine, spermidine, and/or spermine salts are monohydrochloride, dihydrochloride, trihydrochloride, or tetrahydrochloride salts.

[0121] In some embodiments, an amplification reaction mixture of the present invention comprises: a polymerase (*e.g.*, a polymerase-Sso7 conjugate) as described herein at a concentration of about 1 U/ml to about 75 U/ml (*e.g.*, about 1 U/ml, 5 U/ml, 10 U/ml, 15 U/ml, 20 U/ml, 25 U/ml, 30 U/ml, 35 U/ml, 40 U/ml, 45 U/ml, 50 U/ml, 55 U/ml, 60 U/ml, 65 U/ml, 70 U/ml, or 75 U/ml); arginine, spermidine, or spermine or a salt thereof at a concentration of about 1 mM to about 100 mM (*e.g.*, about 1 mM, 5 mM, 10 mM, 15 mM, 20 mM, 25 mM, 30mM, 35 mM, 40 mM, 45 mM, 50 mM, 55 mM, 60 mM, 65 mM, 70 mM, 75 mM, 80 mM, 85 mM, 90 mM, 95 mM, or 100 mM); dNTPs at a concentration of about 0.1 mM to about 10 mM (*e.g.*, about 0.1 mM, 0.2 mM, 0.3 mM, 0.4 mM, 0.5 mM, 0.6 mM, 0.7 mM, 0.8 mM, 0.9 mM, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, or 10 mM); magnesium, *e.g.*, $MgCl_2$, at a concentration of about 1 mM to about 20 mM (*e.g.*, about 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11 mM, 12

mM, 13 mM, 14 mM, 15 mM, 16 mM, 17 mM, 18 mM, 19 mM, or 20 mM); $(NH_4)_2SO_4$ at a concentration of about 10 mM to about 100 mM (*e.g.*, about 10 mM, 20 mM, 30 mM, 40 mM, 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, or 100 mM); potassium, *e.g*, KCl, at a concentration of about 50 mM to about 200 mM (*e.g.*, about 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, or 200 mM); a buffer, *e.g.*, Tris pH 8.5-9.5 at a concentration of about 50 mM to about 200 mM (*e.g.*, about 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, or 200 mM) or a buffer (when measured at a concentration of 0.1 M) that has a change of no more than 0.027 pH units per degree C when between 20° and 37° C. at a concentration of about 5 mM to about 200 mM (*e.g.*, about 5 mM, 10 mM, 25 mM, 40 mM, 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, or 200 mM); a disaccharide, *e.g.*, trehalose, at a concentration of about 100 mM to about 500 mM (*e.g.*, about 100 mM, 125 mM, 150 mM, 175 mM, 200 mM, 225 mM, 250 mM, 275 mM, 300 mM, 325 mM, 350 mM, 375 mM, 400 mM, 425 mM, 450 mM, 475 mM, or 500 mM); one or more osmolytes, *e.g*, sarcosine, trimethylamine N-oxide (TMAO), dimethylsulfoniopropionate, and trimethylglycine, at a concentration of about 50 mM to about 200 mM (*e.g.*, about 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, or 200 mM); Tween-20 at a concentration of about 0.1% to about 0.5% (*e.g.*, about 0.1%, 0.2%, 0.3%, 0.4%, or 0.5%); glycerol at a concentration of about 1% to about 10% (*e.g.*, about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%); DMSO at a concentration of about 1% to about 10% (*e.g.*, about 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, 5.5%, 6%, 6.5%, 7%, 7.5%, 8%, 8.5%, 9%, 9.5%, or 10%); fluorescein at a concentration of about 0.001% to about 0.01% (*e.g.*, about 0.001%, 0.002%, 0.003%, 0.004%, 0.005%, 0.006%, 0.007%, 0.008%, 0.009%, or 0.01%); and DNA binding dye (e.g., cyanine dye) at a concentration of about 0.5X to about 5X (*e.g.*, about 0.5X, 0.6X, 0.7X, 0.8X, 0.9X, 1X, 1.5X, 2X, 2.5X, 3X, 3.5X, 4X, 4.5X, or 5X).

[0122] It has been discovered that inclusion of POPSO in an amplification reaction can improve amplification specificity, and based on this discovery, it is believed that any buffer (e.g. ,POPSO) having a change of no more than 0.027 pH units per degree C when between 20° and 37° C will have the same effect. Thus, the reaction mixtiures described above can include, e.g., HEPES ((4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)), ACES (N-(2-Acetamido)-2-aminoethanesulfonic acid), PIPES (piperazine-N,N′-bis(2-ethanesulfonic acid),

MOPSO (3-(N-Morpholino)-2-hydroxypropanesulfonic Acid), BES (N,N-Bis(2-hydroxyethyl)-2-aminoethanesulfonic Acid), MOPS (3-(N-morpholino)propanesulfonic acid), TES (N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid), TAPSO (3-[N-Tris(hydroxymethyl)methylamino]-2-hydroxypropanesulfonic Acid), POPSO (Piperazine-N,N'-bis(2-hydroxypropanesulfonic acid)), BICINE (N,N-bis(2-hydroxyethyl)glycine), TAPS (N-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid), or AMPSO (N-(1,1-Dimethyl-2-hydroxyethyl)-3-amino-2-hydroxypropanesulfonic acid).

VII. Kits

[0123] In one aspect, the present invention provides kits for conducting nucleic acid amplification reactions. In some embodiments, such kits include a Sso7-polymerase conjugate, and optionally dNTPs, and at least one buffer (e.g., Triss or a buffer (e.g., HEPES or POPSO) having a change of no more than 0.027 pH units per degree C when between 20° and 37° C). Such kits may also include primers, probes, stabilizers and other additives (e.g., arginine, spermidine, spermine, heparin and/or sarcosine) to increase the efficiency of the amplification reactions. Such kits may also include one or more primers as well as instructions for conducting nucleic acid amplification reactions using the components of the kits.

[0124] In a further aspect, the present invention provides kits that include components that improve the efficiency and specificity of nucleic acid amplification reactions over reactions conducted using conventional reaction conditions and reactants. Such additional components are described further herein and include without limitation hot-start antibodies, and/or additives such as sarcosine and heparin.

EXAMPLES

[0125] The following examples are intended to illustrate, but not to limit, the claimed invention.

Example 1

[0126] A series of Sso7d-polymerase conjugates were generated based on SEQ ID NO:2 for the polymerase domain and SEQ ID NO:3 as the Sso7d domain. The Sso7d domain was mutated at positions 28 (wild-type K) or 43 (wild-type R). The resulting conjugates were tested by amplifying a template DNA molecule and detecting the resulting product with SYBR GREEN using melting temperature analysis. A two-step PCR was performed to

amplify 18s amplicon (18s68) using 1 ng of HeLa cell derived cDNA as template. qPCR was performed on Bio-Rad CFX96 qPCR instrument using regular 2-step PCR protocol with 5s denaturation step at 95°C and followed by 30s anneal-extension step at 61°C in each amplification cycle. Forty cycles of PCR amplification were performed and followed by melt-curve analysis using 0.5°C temperature increments with 5s hold in each step. The melt-curve analysis was used to evaluate PCR specificity, which is a technique to characterize double-stranded DNA (dsDNA) based on their dissociation (melting) behavior as they transition from dsDNA to single-stranded DNA (ssDNA) with increasing temperature (Tm). In general, target sequence was amplified by PCR prior to melt-curve analysis. According to the nucleotide sequence, length, GC content, and strand complementarity, melting of PCR products will give a single peak of specific melting temperature. Therefore, single melt-peak indicates one specific PCR product. Multiple peaks indicate the presence of non-specific products in addition to the specific one.

[0127] While the wild-type conjugate generated considerable non-specific products (*i.e.*, two peaks in melt-curve analysis), a number of substitutions at the position corresponding to K28 resulted in improved activity, *i.e.*, reduced non-specific polymerase activity. Mutants in which R28 was changed to any of S, T, C, P, D, E, N, or Q had improved activity (specificity) compared to the R28 wildtype. R28M and R28R were not significantly better than wildtype. The K28 variant Sso7d-polymerase conjugates tested included those having sequences according to SEQ ID NOS:35-42.

[0128] A two-step PCR was then performed to amplify beta-Actin amplicon (ActB86), using 1 ng of HeLa cell derived cDNA as template and under the same conditions as described above to evaluate PCR specificity for R43 variant Sso7d-polymerase conjugates. A number of substitutions resulted in decreased nonspecific polymerase activity. For example, mutants in which R43 was changed to any of G, A, S, T, C, V, L, I, M, F, Y, D, E, N, Q, H, K, or W had improved activity (specificity) compared to the R43 wildtype. R43M had a slight non-specific shoulder, but was still significantly improved compared to wildtype. The R43 variant Sso7-polymerase conjugates tested included those having sequences according to SEQ ID NOS:43-60.

[0129] In addition, it was discovered that non-specific polymerase activity can also be reduced by adding reagents to the PCR reaction mixture. PCR was performed to amplify 18s amplicon (18s68) using 10 ng, 1 ng, 100 pg, and 10 pg of HeLa cell derived cDNA as template. qPCR was performed using a regular 2-step PCR protocol with 5s denaturation step at 98°C and followed by 30s anneal-extension step at 60°C in each amplification cycle.

Forty cycles of PCR amplifications were carried out and followed by melt-curve analysis as described above. The addition of L-Arginine monohydrochloride (10 mM), spermidine trihydrochloride (5 mM), or spermine tetrahydrochloride (5 or 10 mM) further reduced non-specific activity of the Sso7d-polymerase conjugate variants tested. This was further demonstrated in qPCR results using reaction mixtures containing 10 mM L-arginine monohydrochloride at different template concentrations (from 10 ng to 10 pg). Compared to the controls lacking free arginine, the corresponding reactions containing free arginine had more specific product and considerably less non-specific product at each template concentration.

[0130] Next, the effect of arginine in enhancing inhibitor tolerance of qPCR reagent mixture was tested. Amplification reaction mixtures contained an exonuclease deficient polymerase (SEQ ID NO:33) conjugated to a mutated Sso7d domain (SEQ ID NO:38) at a final concentration of about 24 U/ml and a polymerase inhibitor (one of two different chocolates (Enlveonet or Tanzanie), a common PCR inhibitor). Arginine was omitted from one set of samples and added at a concentration of 10 mM to another set of samples. The final percentage concentration of chocolate in the qPCR reaction ranged from 0-2%. PCR was performed to amplify ADAR amplicon (ADAR_162) using 1 ng of HeLa cell derived cDNA as template. qPCR was performed using regular 2-step PCR protocol with 5s denaturation step at 95°C and followed by 30s anneal-extension step at 60°C in each amplification cycle. Forty cycles of PCR amplifications were carried out and followed by melt-curve analysis as described above. Success of PCR amplification (as reflected by an amplification and also evaluated by the Ct value) in the presence of different concentrations of inhibitor indicates how well the reaction mixture tolerates the PCR inhibitor. For both chocolates, in the presence of higher concentrations of inhibitor, poor or no amplification was observed for the reagent mixture in the absence of arginine. In contrast, reagent mixture supplemented with arginine still exhibited good PCR amplification even in the presence of 2% inhibitor, suggesting higher inhibitor tolerance.

[0131] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

WHAT IS CLAIMED IS:

1. An Sso7 polymerase conjugate protein comprising an Sso7 domain linked to a polymerase; wherein:

an amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is not lysine (K); and/or

an amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is not arginine (R),

wherein the conjugate protein has a decreased non-specific amplification activity compared to an otherwise identical control conjugate protein in which the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is lysine (K) and the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is arginine (R).

2. The Sso7 polymerase conjugate protein of claim 1, wherein the non-specific amplification activity of the Sso7 polymerase conjugate protein is reduced by at least 10% compared to non-specific amplification activity of the control conjugate protein.

3. The Sso7 polymerase conjugate protein of claim 1, wherein the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is not lysine (K).

4. The Sso7 polymerase conjugate protein of claim 3, wherein the amino acid of the Sso7 domain corresponding to K28 of SEQ ID NO:2 is selected from the group consisting of serine (S), threonine (T), cytosine (C), proline (P), aspartic acid (D), glutamic acid (E), asparagine (N), glutamine (Q), alanine (A), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), leucine (L), methionine (M), arginine (R), valine (V), tryptophan (W), and tyrosine (Y).

5. The Sso7 polymerase conjugate protein of claim 1, wherein the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is not arginine (R).

6. The Sso7 polymerase conjugate protein of claim 5, wherein the amino acid of the Sso7 domain corresponding to R43 of SEQ ID NO:2 is selected from the group consisting of alanine (A), cytosine (C), aspartic acid (D), glutamic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), lysine (K), leucine (L), methionine (M), asparagine (N), glutamine (Q), serine (S), threonine (T), valine (V), tryptophan (W), tyrosine (Y), and proline (P).

7. The Sso7 polymerase conjugate protein of claim 1, wherein the polymerase substantially lacks a 3'-5' exonuclease activity.

8. The Sso7 polymerase conjugate protein of claim 1, wherein the Sso7 domain is substantially (e.g., at least 60, 75, 80, 85, 90, or 95%) identical to the first 58 or 60 amino acids of SEQ ID NO:2 or to the full-length of SEQ ID NO:2. .

9. The Sso7 polymerase conjugate protein of claim 1, wherein the Sso7 domain is at least 90% identical to SEQ ID NO:2.

10. The Sso7 polymerase conjugate protein of claim 1, wherein the polymerase is substantially (e.g., at least 60, 75, 80, 85, 90, or 95%) identical to SEQ ID NO:33 or SEQ ID NO:61.

11. The Sso7 polymerase conjugate protein of claim 10, wherein the polymerase comprises SEQ ID NO: 33 or SEQ ID NO:61.

12. The Sso7 polymerase conjugate protein of claim 1, wherein the polymerase domain has thermally stable polymerase activity.

13. The Sso7 polymerase conjugate protein of claim 12, wherein the polymerase domain is a family A polymerase domain.

14. The Sso7 polymerase conjugate protein of claim 13, wherein the polymerase domain is a ΔTaq polymerase domain.

15. The Sso7 polymerase conjugate protein of claim 12, wherein the polymerase domain is a family B polymerase domain.

16. The Sso7 polymerase conjugate protein of claim 15, wherein the polymerase domain is from *Pyrococcus*.

17. A reaction mixture comprising the Sso7 polymerase conjugate protein of any of claims 1-16.

18. The reaction mixture of claim 17, further comprising at least one of or more oligonucleotide primers, one or more detectably-labeled oligonucleotide probe, a buffer, nucleoside triphosphates, a salt, a DNA binding dye, or a stabilizer.

19. A kit comprising the Sso7 polymerase conjugate protein of any of claims 1-16.

20. The kit of claim 19, further comprising in the same or a different container that contains the conjugate protein, at least one of one or more oligonucleotide primers, one or more detectably-labeled oligonucleotide probe, a buffer, nucleoside triphosphates, a salt, a DNA binding dye, or a stabilizer.

21. A method of amplifying a target nucleic acid in a sample, the method comprising incubating the target nucleic acid in a reaction mixture comprising:

a. at least one primer; and

b. the Sso7 polymerase conjugate protein of any of claims 1-16,

under conditions to allow for amplification of the target nucleic acid with the primer, thereby amplifying the target nucleic acid.

22. The method of claim 21, wherein the method comprises a polymerase chain reaction (PCR).

23. A nucleic acid comprising a polynucleotide encoding the Sso7 polymerase conjugate protein of any of claims 1-16.

24. A cell comprising a recombinant expression cassette, the expression cassette comprising a promoter operably linked to the polynucleotide of claim 23.