Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2831180 A1 2012/10/18

(21) 2 831 180

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2012/04/10
(87) Date publication PCT/PCT Publication Date: 2012/10/18
(85) Entrée phase nationale/National Entry: 2013/09/24
(86) N° demande PCT/PCT Application No.: EP 2012/001552
(87) N° publication PCT/PCT Publication No.: 2012/139748
(30) Priorité/Priority: 2011/04/11 (US61/474,160)

(51) Cl.Int./Int.Cl. *C12N 9/12* (2006.01)

(71) Demandeur/Applicant:
ROCHE DIAGNOSTICS GMBH, DE

(72) Inventeurs/Inventors:
BAUER, KEITH, US;
SAN FILIPPO, JOSEPH, US;
MYERS, THOMAS W., US;
REICHERT, FRED, US;
SHAHINIAN, RACHEL, US;
SUKO, SHAWN, US

(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : ADN POLYMERASES AVEC UNE ACTIVITE AMELIOREE
(54) Title: DNA POLYMERASES WITH IMPROVED ACTIVITY

***FIGURE 1***

```
          *
Z05       AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:12)
Taq       AWIEKTLEEGRRRGYVETLFGRRRYVPDLEA (SEQ ID NO:13)
Tfi       AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA (SEQ ID NO:14)
Tfl       AWIEGTLEEGRRRGYVETLFGRRRYVPDLNA (SEQ ID NO:15)
Sps17     AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA (SEQ ID NO:16)
Tth       AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:17)
Tca       AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:18)
Tma       DYIQRVVSEAKEKGYVRTLFGRKRDIPQLMA (SEQ ID NO:19)
Tne       SYIQQVVAEAKEKGYVRTLFGRKRDIPQLMA (SEQ ID NO:20)
Taf       EYLKRMKDEARKKGYVTTLFGRRRYIPQLRS (SEQ ID NO:21)
Dra       RYINHTLDFGRTHGYVETLYGRRRYVPGLSS (SEQ ID NO:23)
Bst       QYMDNIVQEAKQKGYVTTLLHRRRYLPDITS (SEQ ID NO:24)
Bca       RYMENIVQEAKQKGYVTTLLHRRRYLPDITS (SEQ ID NO:25)

X1X2X3X4X5X6X7X8X9X10X11X12X13GYVX14TL---------- (SEQ ID NO:26)
```

(57) Abrégé/Abstract:
Disclosed are DNA polymerases having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors relative to a corresponding, unmodified polymerase. The polymerases are useful in a variety of disclosed primer extension methods. Also disclosed are related compositions, including recombinant nucleic acids, vectors, and host cells, which are useful, e.g., for production of the DNA polymerases.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2012 (18.10.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/139748 A1

(51) International Patent Classification:
*C12N 9/12* (2006.01)

(21) International Application Number:
PCT/EP2012/001552

(22) International Filing Date:
10 April 2012 (10.04.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/474,160 11 April 2011 (11.04.2011) US

(71) Applicant *(for DE only)*: **ROCHE DIAGNOSTICS GMBH** [DE/DE]; Sandhofer Strasse 116, 68305 Mannheim (DE).

(71) Applicant *(for all designated States except DE, US)*: **F. HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BAUER, Keith**; 62 W. Crescent Drive, San Rafael, CA 94901 (US). **SAN FILIPPO, Joseph**; 4213 Garibaldi Place, Pleasanton, CA 94566 (US). **MYERS, Thomas, W.**; 600 Loyola Drive, Los Altos, CA 94204 (US). **REICHERT, Fred**; 675 Haas Ave, San Leandro, CA 94577-2929 (US). **SHAHINIAN, Rache**; 925 Broadway, Alameda, CA 94501 (US). **SUKO, Shawn**; 4026 La Colina Rd, El Sobrante, CA 94803 (US).

(74) Common Representative: **ROCHE DIAGNOSTICS GMBH**; Patent Department, 68305 Mannheim (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17**:

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

**Published**:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

**(54) Title**: DNA POLYMERASES WITH IMPROVED ACTIVITY

***FIGURE 1***

```
          *
Z05   AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:12)
Taq   AWIEKTLEEGRRRGYVETLFGRRRYVPDLEA (SEQ ID NO:13)
Tfi   AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA (SEQ ID NO:14)
Tfl   AWIEGTLEEGRRRGYVETLFGRRRYVPDLNA (SEQ ID NO:15)
Sps17 AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA (SEQ ID NO:16)
Tth   AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:17)
Tca   AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:18)
Tma   DYIQRVVSEAKEKGYVRTLFGRKRDIPQLMA (SEQ ID NO:19)
Tne   SYIQQVVAEAKEKGYVRTLFGRKRDIPQLMA (SEQ ID NO:20)
Taf   EYLKRMKDEARKKGYVTTLFGRRRYIPQLRS (SEQ ID NO:21)
Dra   RYINHTLDFGRTHGYVETLYGRRRYVPGLSS (SEQ ID NO:23)
Bst   QYMDNIVQEAKQKGYVTTLLHRRRYLPDITS (SEQ ID NO:24)
Bca   RYMENIVQEAKQKGYVTTLLHRRRYLPDITS (SEQ ID NO:25)
   X1X2X3X4X5X6X7X8X9X10X11X12X13GYVX14TL---------- (SEQ ID NO:26)
```

**(57) Abstract**: Disclosed are DNA polymerases having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors relative to a corresponding, unmodified polymerase. The polymerases are useful in a variety of disclosed primer extension methods. Also disclosed are related compositions, including recombinant nucleic acids, vectors, and host cells, which are useful, e.g., for production of the DNA polymerases.

WO 2012/139748 A1

# DNA POLYMERASES WITH IMPROVED ACTIVITY

## FIELD OF THE INVENTION

The present invention provides DNA polymerases with improved activities, including increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of reverse transcriptase (RT) and polymerase inhibitors, as well as use of such polymerases in various applications, including nucleic acid polynucleotide extension and amplification.

## BACKGROUND OF THE INVENTION

DNA polymerases are responsible for the replication and maintenance of the genome, a role that is central to accurately transmitting genetic information from generation to generation. DNA polymerases function in cells as the enzymes responsible for the synthesis of DNA. They polymerize deoxyribonucleoside triphosphates in the presence of a metal activator, such as $Mg^{2+}$, in an order dictated by the DNA template or polynucleotide template that is copied. *In vivo*, DNA polymerases participate in a spectrum of DNA synthetic processes including DNA replication, DNA repair, recombination, and gene amplification. During each DNA synthetic process, the DNA template is copied once or at most a few times to produce identical replicas. In contrast, *in vitro*, DNA replication can be repeated many times such as, for example, during polymerase chain reaction (*see, e.g.,* U.S. Patent No. 4,683,202).

In the initial studies with polymerase chain reaction (PCR), the DNA polymerase was added at the start of each round of DNA replication (*see* U.S. Patent No. 4,683,202, *supra*). Subsequently, it was determined that thermostable DNA polymerases could be obtained from bacteria that grow at elevated temperatures, and that these enzymes need to be added only once (*see* U.S. Patent No. 4,889,818 and U.S. Patent No. 4,965,188). At the elevated temperatures used during PCR, these enzymes are not irreversibly inactivated. As a result, one can carry out repetitive cycles of polymerase chain reactions without adding fresh enzymes at the start of each synthetic addition process. DNA polymerases, particularly thermostable polymerases, are the key to a large number of techniques in recombinant DNA studies and in medical diagnosis

of disease. For diagnostic applications in particular, a target nucleic acid sequence may be only a small portion of the DNA or RNA in question, so it may be difficult to detect the presence of a target nucleic acid sequence without amplification.

The overall folding pattern of DNA polymerases resembles the human right hand and contains three distinct subdomains of palm, fingers, and thumb. (*See* Beese *et al.*, *Science* 260:352-355, 1993); Patel *et al.*, *Biochemistry* 34:5351-5363, 1995). While the structure of the fingers and thumb subdomains vary greatly between polymerases that differ in size and in cellular functions, the catalytic palm subdomains are all superimposable. For example, motif A, which interacts with the incoming dNTP and stabilizes the transition state during chemical catalysis, is superimposable with a mean deviation of about one Å amongst mammalian pol α and prokaryotic pol I family DNA polymerases (Wang *et al.*, *Cell* 89:1087-1099, 1997). Motif A begins structurally at an antiparallel β-strand containing predominantly hydrophobic residues and continues to an α-helix. The primary amino acid sequence of DNA polymerase active sites is exceptionally conserved. In the case of motif A, for example, the sequence DYSQIELR (SEQ ID NO:22) is retained in polymerases from organisms separated by many millions years of evolution, including, *e.g.*, *Thermus aquaticus*, *Chlamydia trachomatis*, and *Escherichia coli*.

In addition to being well-conserved, the active site of DNA polymerases has also been shown to be relatively mutable, capable of accommodating certain amino acid substitutions without reducing DNA polymerase activity significantly. (*See, e.g.*, U.S. Patent No. 6,602,695). Such mutant DNA polymerases can offer various selective advantages in, *e.g.*, diagnostic and research applications comprising nucleic acid synthesis reactions.

There are at least two steps in the enzymatic process of DNA polymerization; 1) the incorporation of the incoming nucleotide and 2) the extension of the newly incorporated nucleotide. The overall faithfulness or "fidelity" of the DNA polymerase is generally thought of as a conglomerate of these two enzymatic activities, but the steps are distinct. A DNA polymerase may misincorporate the incoming nucleotide, but if it is not efficiently extended the extension rate will be severely decreased and overall product formation would be minimal. Alternatively, it is possible to have a DNA polymerase misincorporate the incoming nucleotide and readily misextend the newly formed mismatch. In this case, the overall extension rate would be high, but the overall fidelity would be low. An example of this type of enzyme would be ES112 DNA polymerase (E683R Z05 DNA polymerase; see US 7,179,590) when using

$Mn^{2+}$ as the divalent metal ion activator. The enzyme has a very high efficiency because unlike typical DNA polymerases that tend to hesitate/stall when a mismatch is encountered, the ES112 DNA polymerase readily extends the mismatch. The phenotype displayed in ES112 is more pronounced during the reverse transcription (RT) step, presumably because of structural effects of the RNA/DNA heteroduplex vs. the DNA/DNA homoduplex. A second example would be if the DNA polymerase does not readily misincorporate (may be even less likely to misincorporate), but does have increased capacity to misextend a mismatch. In this case, the fidelity is not significantly altered for the overall product. In general, this type of enzyme is more favorable for extension reactions than the characteristics of ES112 in $Mn^{2+}$ because the fidelity of the product is improved. However, this attribute can be utilized to allow the misextension of a mismatched oligonucleotide primer such as when an oligonucleotide primer of a single sequence is hybridized to a target that has sequence heterogeneity (e.g., viral targets), but the normal or lower misincorporation rate allows for completion of DNA synthesis beyond the original oligonucleotide primer. An example of this type of DNA polymerase is Z05 D580G DNA polymerase (see U.S. Patent Publication No. 2009/0148891). This type of activity is referred to as "mismatch tolerant" because it is more tolerant to mismatches in the oligonucleotide primer. While the examples above have discussed primer extension type reactions, the activity can be more significant in reactions such as RT-PCR and PCR where primer extension is reoccurring frequently. Data suggests that while enzymes such as Z05 D580G are more "tolerant" to mismatches, they also have enhanced ability to extend oligonucleotide primers containing modified bases (e.g., t-butyl benzyl modified bases) or in the presence of DNA binding dyes such as SYBR Green I (see U.S. Patent Publication No. 2009/028053).

Reverse transcription polymerase chain reaction (RT-PCR) is a technique used in many applications to detect/and or quantify RNA targets by amplification. In order to amplify RNA targets by PCR, it is necessary to first reverse transcribe the RNA template into cDNA. Typically, RT-PCR assays rely on a non-thermostable reverse transcriptase (RNA dependent DNA polymerase), derived from a mesophilic organism, for the initial cDNA synthesis step (RT). An additional thermostable DNA polymerase is required for amplification of cDNA to tolerate elevated temperatures required for nucleic acid denaturation in PCR. There are several potential benefits of using thermoactive or thermostable DNA polymerases engineered to

perform more efficient reverse transcription for RT-PCR assays. Increased reverse transcriptase activity coupled with the ability to use higher reverse transcription incubation temperatures, which allow for relaxing of RNA template secondary structure, can result in overall higher cDNA synthesis efficiency and assay sensitivity. Higher temperature incubation could also increase specificity by reducing false priming in the reverse transcription step. Enzymes with improved reverse transcription efficiency can simplify assay design by allowing for reduced RT incubation times and/or enzyme concentration. When using dUTP and UNG, nonspecific extension products containing dUMP that are formed during nonstringent set-up conditions are degraded by UNG and cannot be utilized either as primers or as templates. When using a non-thermostable reverse transcriptase (RNA dependent DNA polymerase) derived from a mesophilic organism, it is not possible to utilize the dUTP and UNG methodologies. (Myers, T.W. et al., Amplification of RNA: High Temperature Reverse Transcription and DNA Amplification with *Thermus thermophilus* DNA Polymerase, in *PCR Strategies*, Innis, M.A., Gelfand, D.H., and Sninsky, J.J., Eds., Academic Press, San Diego, CA, 58-68, (1995)). However, the use of a thermoactive or thermostable DNA polymerase of the invention for the reverse transcription step enables the reaction to be completely compatible with the utilization of the dUTP/uracil N-glycosylase (UNG) carry-over prevention system (Longo et al., Use of Uracil DNA Glycosylase to Control Carry-over Contamination in Polymerase Chain Reactions. *Gene* 93:125-128, (1990)). In addition to providing carry-over contamination control, the use of dUTP and UNG provides a "hot-start" to reduce nonspecific amplification (Innis and Gelfand (1999) *supra*).

## BRIEF SUMMARY OF THE INVENTION

Provided herein are DNA polymerases having improved activities, including increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors, relative to a corresponding, unmodified control polymerase, and methods of making and using such DNA polymerases. In some embodiments, the amino acid of the DNA polymerase corresponding to position 709 of SEQ ID NO:1 is any amino acid other than I, L, or M, and the control DNA polymerase has the same amino acid sequence as the DNA polymerase except that the amino acid of the control DNA polymerase corresponding to position 709 of SEQ ID NO:1 is I, L, or M. For example, in some embodiments, the amino acid

at the position corresponding to position 709 of SEQ ID NO:1 of the improved polymerase is selected from G, A, V, R, F, W, P, S, T, C, Y, N, Q, D, E, K, or H.

In some embodiments, the DNA polymerase having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors comprises a motif in the polymerase domain comprising

$X_1$-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-$X_7$-$X_8$-$X_9$-$X_{10}$-$X_{11}$-$X_{12}$-$X_{13}$-G-Y-V-$X_{14}$-T-L, wherein:

$X_1$ is A, D, S, E, R or Q;

$X_2$ is W or Y;

$X_3$ is any amino acid other than I, L or M;

$X_4$ is E, A, Q, K, N or D;

$X_5$ is K, G, R, Q, H or N;

$X_6$ is T, V, M or I;

$X_7$ is L, V or K;

$X_8$ is E, S, A, D or Q;

$X_9$ is E or F;

$X_{10}$ is G or A;

$X_{11}$ is R or K;

$X_{12}$ is K, R, E, T or Q;

$X_{13}$ is R, K or H; and

$X_{14}$ is E, R or T (SEQ ID NO:8).

In some embodiments $X_3$ is selected from G, A, W, P, S, T, F, Y, C, N, Q, D, E, K, V, R or H.

In some embodiments, the DNA polymerase having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors comprises a motif in the polymerase domain comprising

$X_1$-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-$X_7$-$X_8$-E-$X_{10}$-$X_{11}$-$X_{12}$-$X_{13}$-G-Y-V-$X_{14}$-T-L, wherein:

$X_1$ is A, D, or S;

$X_2$ is W or Y;

$X_3$ is any amino acid other than I;

$X_4$ is E, A, or Q;

$X_5$ is K, G, R or Q;

$X_6$ is T or V;

$X_7$ is L or V;

$X_8$ is E, S or A;

$X_{10}$ is G or A;

$X_{11}$ is R or K;

$X_{12}$ is K, R or E;

$X_{13}$ is R or K; and

$X_{14}$ is E or R (SEQ ID NO:9).

In some embodiments, the DNA polymerase having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors comprises a motif in the polymerase domain comprising

A-W-$X_3$-$X_4$-$X_5$-T-L-E-E-G-R-$X_{12}$-$X_{13}$-G-Y-V-E-T-L, wherein:

$X_3$ is any amino acid other than I;

$X_4$ is E or A;

$X_5$ is K or G;

$X_{12}$ is K or R; and

$X_{13}$ is R or K (SEQ ID NO:10).

In some embodiments, the DNA polymerase having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors comprises a motif in the polymerase domain comprising

A-W-$X_3$-$X_4$-$X_5$-T-L-E-E-G-R-$X_{12}$-$X_{13}$-G-Y-V-E-T-L, wherein:

$X_3$ is K, R, S, G, or A;

$X_4$ is E or A;

$X_5$ is K or G;

$X_{12}$ is K or R; and

$X_{13}$ is R or K (SEQ ID NO:11).

In some embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is any amino acid other than D or E. In some embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is any amino acid other than

D. In some embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is selected from the group consisting of L, G, T, Q, A, S, N, R, and K.

In some embodiments, the DNA polymerase further comprises a mutation at one or more amino acids corresponding to a position selected from 580 and 588 of SEQ ID NO:1. In some embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is any amino acid other than D or E. In some embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is selected from the group consisting of L, G, T, Q, A, S, N, R, and K. In some embodiments, the amino acid of the DNA polymerase corresponding to position 588 of SEQ ID NO:1 is any amino acid other than I. In some embodiments, the amino acid of the DNA polymerase corresponding to position 588 of SEQ ID NO:1 is selected from L, V, G, A, S, M, F, W, P, R, K, T, C, Y, N, Q, D, E or H. In some embodiments, the amino acid of the DNA polymerase corresponding to position 588 of SEQ ID NO:1 is T.

Various DNA polymerases are amenable to mutation according to the present invention. Particularly suitable are thermostable polymerases, including wild-type or naturally occurring thermostable polymerases from various species of thermophilic bacteria, as well as synthetic thermostable polymerases derived from such wild-type or naturally occurring enzymes by amino acid substitution, insertion, or deletion, or other modification. Exemplary unmodified forms of polymerase include, *e.g.*, CS5, CS6 or Z05 DNA polymerase, or a functional DNA polymerase having at least 80%, preferably at least 90%,more preferably at least 95% amino acid sequence identity thereto. Other unmodified polymerases include, *e.g.*, DNA polymerases from any of the following species of thermophilic bacteria (or a functional DNA polymerase having at least 80%, preferably at least 90%,more preferably at least 95% amino acid sequence identity to such a polymerase): *Thermotoga maritima* (SEQ ID NO:34); *Thermus aquaticus* (SEQ ID NO:2) ; *Thermus thermophilus* (SEQ ID NO:6); *Thermus flavus* (SEQ ID NO:4); *Thermus filiformis* (SEQ ID NO:3); *Thermus* sp. sps17 (SEQ ID NO:5); *Thermus* sp. Z05 (SEQ ID NO:1); *Thermotoga neopolitana* (SEQ ID NO:35); *Thermosipho africanus* (SEQ ID NO:33); *Thermus caldophilus* (SEQ ID NO:7), *Deinococcus radiodurans* (SEQ ID NO:32), *Bacillus stearothermophilus* (SEQ ID NO:36) or *Bacillus caldotenax* (SEQ ID NO:37). Suitable polymerases also include those having reverse transcriptase (RT) activity and/or the ability to

incorporate unconventional nucleotides, such as ribonucleotides or other 2'-modified nucleotides.

While thermostable DNA polymerases possessing efficient reverse transcription activity are particularly suited for performing RT-PCR, especially single enzyme RT-PCR, thermoactive, but not thermostable DNA polymerases possessing efficient reverse transcription activity also are amenable to mutation according to the present invention. For example, the attributes of increased reverse transcriptase efficiency, mismatch tolerance, extension rate, and/or tolerance of RT inhibitors are important for the RT step in an RT-PCR and this step does not need to be performed at temperatures that would inactivate a thermoactive but not thermostable DNA polymerase. Following the RT step, a thermostable DNA polymerase could either be added or it could already be included in the reaction mixture to perform the PCR amplification step. This second methodology would especially benefit by using a chemically modified thermostable DNA polymerase (or other HotStart technology to inactivate the thermostable DNA polymerase) so that it would not be fully active during the RT step. An example of a thermoactive but not thermostable DNA polymerase possessing efficient reverse transcription activity is the DNA polymerase from *Carboxydothermus hydrogenoformans* (Chy; SEQ ID NO:48). *See*, *e.g.*, US Patent Nos. 6,468,775 and 6,399,320.

In some embodiments, the DNA polymerase has at least 80%, preferably at least 90%,more preferably at least 95% amino acid sequence identity to a polymerase selected from the group consisting of:

(a) a *Thermus sp.* Z05 DNA polymerase (Z05) (SEQ ID NO:1);
(b) a *Thermus aquaticus* DNA polymerase (Taq) (SEQ ID NO:2);
(c) a *Thermus filiformis* DNA polymerase (Tfi) (SEQ ID NO:3);
(d) a *Thermus flavus* DNA polymerase (Tfl) (SEQ ID NO:4);
(e) a *Thermus sp. sps17* DNA polymerase (Sps17) (SEQ ID NO:5);
(f) a *Thermus thermophilus* DNA polymerase (Tth) (SEQ ID NO:6);
(g) a *Thermus caldophilus* DNA polymerase (Tca) (SEQ ID NO:7); and
(h) *Carboxydothermus hydrogenoformans* DNA polymerase (Chy) (SEQ ID NO:48).

In some embodiments, the DNA polymerase is a *Thermotoga* DNA polymerase. For example, in some embodiments, the DNA polymerase has at least 80%, preferably at least 90%,more

preferably at least 95% amino acid sequence identity to a polymerase selected from the group consisting of:

(a) a *Thermotoga maritima* DNA polymerase (Tma) (SEQ ID NO:34);

(b) a *Thermotoga neopolitana* DNA polymerase (Tne) (SEQ ID NO:35).

In some embodiments, the DNA polymerase has at least 80%, preferably at least 90%,more preferably at least 95% amino acid sequence identity to SEQ ID NO:1. In some embodiments, the DNA polymerase is a *Thermus sp.* Z05 DNA polymerase (Z05) DNA polymerase, and the amino acid at position 709 is any amino acid other than I. In some embodiments, the DNA polymerase is a Z05 DNA polymerase (*i.e.*, SEQ ID NO:1), and the amino acid at position 709 is any amino acid other than I, L, or M. For example, in some embodiments, the amino acid at position 709 is selected from G, A, V, R, F, W, P, S, T, C, Y, N, Q, D, E, K, or H. In some embodiments, the DNA polymerase is a Z05 DNA polymerase, and the amino acid at position 709 is K, R, S, G, or A. In some embodiments, the DNA polymerase is a Z05 DNA polymerase further comprising a substitution at position 580, and the amino acid at position 580 is any amino acid other than D or E. In some embodiments, the DNA polymerase is a Z05 DNA polymerase, and the amino acid at position 580 is any amino acid other than D. In some embodiments, the DNA polymerase is a Z05 DNA polymerase, and the amino acid at position 580 is selected from the group consisting of L, G, T, Q, A, S, N, R, and K.

In some embodiments, the mutant polymerase has increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors compared with a control DNA polymerase, wherein the amino acid of the thermostable DNA polymerase corresponding to position 588 of SEQ ID NO:1 is any amino acid other than I or V, and wherein the control DNA polymerase has the same amino acid sequence as the thermostable DNA polymerase except that the amino acid of the control DNA polymerase corresponding to position 588 of SEQ ID NO:1 is I or V. In some embodiments, the amino acid of the thermostable DNA polymerase corresponding to position 588 of SEQ ID NO:1 is selected from G, A, W, P, S, T, F, Y, C, N, Q, D, E, K, R, L, M, or H. In some embodiments, the polymerase comprises a motif in the polymerase domain comprising

Pro-Asn-Leu-Gln-Asn-$X_1$-Pro-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-Gly, wherein

$X_1$ is Ile (I), or Leu (L);

$X_2$ is any amino acid other than Ile (I) or Val (V);

$X_3$ is Arg (R) or Lys (K);

$X_4$ is Thr (T), Ser (S) or Leu (L);

$X_5$ is Pro (P) or Glu (E); and

$X_6$ is Leu (L) or Glu (E) (SEQ ID NO:29).

The mutant or improved polymerases can include other, non-substitutional modifications. One such modification is a thermally reversible covalent modification that inactivates the enzyme, but which is reversed to activate the enzyme upon incubation at an elevated temperature, such as a temperature typically used for polynucleotide extension. Exemplary reagents for such thermally reversible modifications are described in U.S. Patent Nos. 5,773, 258 and 5,677,152.

In some embodiments, the reverse transcriptase activity is determined by performing real-time RT-PCR amplification and detection of a Hepatitis C Virus (HCV) transcript generated from the first 800 bases of HCV genotype Ib 5'NTR in pSP64 poly(A) (Promega). Two or more reaction mixtures can have titrated numbers of copies of the Hepatitis C Virus (HCV) transcript (*e.g.*, 1:5 titrations, 1:10 titrations, *e.g.*, 10,000 copies, 1000 copies, 100 copies, 10 copies, 1 copy, 0 copies in several reaction mixtures). The reverse transcriptase ability of a polymerase of the invention can be compared to the reverse transcriptase ability of a reference polymerase (*e.g.*, a naturally occurring or unmodified polymerase), over a preselected unit of time, as described herein. Polymerases with improved reverse transcriptase ability will amplify the transcript with greater efficiency, or will require a lower number of PCR cycles to amplify the transcript (*i.e.*, exhibit a lower Cp value, as calculated herein), in comparison to a naturally occurring or unmodified polymerase. Moreover, in some embodiments, polymerases with improved RT function also have improved replication of long RNA (e.g., at least 500 or 1000 or 2000 or 5000 or more nucleotides long) templates.

In various other aspects, the present invention provides a recombinant nucleic acid encoding a mutant or improved DNA polymerase as described herein, a vector comprising the recombinant nucleic acid, and a host cell transformed with the vector. In certain embodiments, the vector is an expression vector. Host cells comprising such expression vectors are useful in methods of the invention for producing the mutant or improved polymerase by culturing the host cells under conditions suitable for expression of the recombinant nucleic acid. The polymerases of the invention may be contained in reaction mixtures and/or kits. The embodiments of the

recombinant nucleic acids, host cells, vectors, expression vectors, reaction mixtures and kits are as described above and herein.

In yet another aspect, a method for conducting polynucleotide extension is provided. The method generally includes contacting a DNA polymerase having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors as described herein with a primer, a polynucleotide template, and nucleoside triphosphates under conditions suitable for extension of the primer, thereby producing an extended primer. The polynucleotide template can be, for example, an RNA or DNA template. In certain embodiments the polynucleotide template is RNA.The nucleotide triphosphates can include unconventional nucleotides such as, *e.g.*, ribonucleotides and/or labeled nucleotides. Further, the primer and/or template can include one or more nucleotide analogs. In some variations, the polynucleotide extension method is a method for polynucleotide amplification that includes contacting the mutant or improved DNA polymerase with a primer pair, the polynucleotide template, and the nucleoside triphosphates under conditions suitable for amplification of the polynucleotide. The polynucleotide extension reaction can be, *e.g.*, PCR, isothermal extension, or sequencing (*e.g.*, 454 sequencing reaction). In certain embodiments the primer extension method comprises a polymerase chain reaction (PCR). The polynucleotide template can be from any type of biological sample.

Optionally, the primer extension reaction comprises an actual or potential inhibitor of a reference or unmodified polymerase. The inhibitor can inhibit the nucleic acid extension rate and/or the reverse transcription efficiency of a reference or unmodified (control) polymerase. In some embodiments, the inhibitor is hemoglobin, or a degradation product thereof. For example, in some embodiments, the hemoglobin degradation product is a heme breakdown product, such as hemin, hematoporphyrin, or bilirubin. In some embodiments, the inhibitor is an iron-chelator or a purple pigment. In other embodiments, the inhibitor is heparin or melanin. In certain embodiments, the inhibitor is an intercalating dye. In some embodiments, the intercalating dye is [2-[N-bis-(3-dimethylaminopropyl)-amino]-4-[2,3-dihydro-3-methyl-(benzo-1,3-thiazol-2-yl)-methylidene]-1-phenyl-quinolinium]+. In some embodiments, the intercalating dye is [2-[N-(3-dimethylaminopropyl)-N-propylamino]-4-[2,3-dihydro-3-methyl-(benzo-1,3-thiazol-2-yl)-methylidene]-1-phenyl-quinolinium]+. In some embodiments, the intercalating dye is not [2-[N-(3-dimethylaminopropyl)-N-propylamino]-4-[2,3-dihydro-3-methyl-(benzo-1,3-thiazol-2-yl)-

methylidene]-1-phenyl-quinolinium]+. In some embodiments, the conditions suitable for extension comprise $Mg^{++}$. In some embodiments, the conditions suitable for extension comprise $Mn^{++}$.

The present invention also provides a kit useful in such a polynucleotide extension method. Generally, the kit includes at least one container providing a mutant or improved DNA polymerase as described herein. In certain embodiments, the kit further includes one or more additional containers providing one or more additional reagents. For example, in specific variations, the one or more additional containers provide nucleoside triphosphates; a buffer suitable for polynucleotide extension; and/or one or more primer or probe polynucleotides, hybridizable, under polynucleotide extension conditions, to a predetermined polynucleotide template. The polynucleotide template can be from any type of biological sample.

Further provided are reaction mixtures comprising the polymerases of the invention. The reaction mixtures can also contain a template nucleic acid (DNA and/or RNA), one or more primer or probe polynucleotides, nucleoside triphosphates (including, *e.g.*, deoxyribonucleoside triphosphates, ribonucleoside triphosphates, labeled nucleoside triphosphates, unconventional nucleoside triphosphates), buffers, salts, labels (*e.g.*, fluorophores). In some embodiments the polynucleotide template is RNA. In some embodiments, the reaction mixtures comprise an iron chelator or a purple dye. In certain embodiments, the reaction mixtures comprise hemoglobin, or a degradation product of hemoglobin. For example, in certain embodiments, the degradation products of hemoglobin include heme breakdown products such as hemin, hematin, hematophoryn, and bilirubin. In other embodiments, the reaction mixtures comprise heparin or a salt thereof. Optionally, the reaction mixture comprises an intercalating dye (including but not limited to those described above or elsewhere herein). In certain embodiments, the reaction mixture contains a template nucleic acid that is isolated from blood. In other embodiments, the template nucleic acid is RNA and the reaction mixture comprises heparin or a salt thereof.

Further embodiments of the invention are described herein.

## DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although essentially any methods and materials similar to those described herein can be used in

the practice or testing of the present invention, only exemplary methods and materials are described. For purposes of the present invention, the following terms are defined below.

The terms "a," "an," and "the" include plural referents, unless the context clearly indicates otherwise.

An "amino acid" refers to any monomer unit that can be incorporated into a peptide, polypeptide, or protein. As used herein, the term "amino acid" includes the following twenty natural or genetically encoded alpha-amino acids: alanine (Ala or A), arginine (Arg or R), asparagine (Asn or N), aspartic acid (Asp or D), cysteine (Cys or C), glutamine (Gln or Q), glutamic acid (Glu or E), glycine (Gly or G), histidine (His or H), isoleucine (Ile or I), leucine (Leu or L), lysine (Lys or K), methionine (Met or M), phenylalanine (Phe or F), proline (Pro or P), serine (Ser or S), threonine (Thr or T), tryptophan (Trp or W), tyrosine (Tyr or Y), and valine (Val or V). In cases where "X" residues are undefined, these should be defined as "any amino acid." The structures of these twenty natural amino acids are shown in, e.g., Stryer et al., Biochemistry, 5th ed., Freeman and Company (2002). Additional amino acids, such as selenocysteine and pyrrolysine, can also be genetically coded for (Stadtman (1996) "Selenocysteine," Annu Rev Biochem. 65:83-100 and Ibba et al. (2002) "Genetic code: introducing pyrrolysine," Curr Biol. 12(13):R464-R466). The term "amino acid" also includes unnatural amino acids, modified amino acids (e.g., having modified side chains and/or backbones), and amino acid analogs. See, e.g., Zhang et al. (2004) "Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells," Proc. Natl. Acad. Sci. U.S.A. 101(24):8882-8887, Anderson et al. (2004) "An expanded genetic code with a functional quadruplet codon" Proc. Natl. Acad. Sci. U.S.A. 101(20):7566-7571, Ikeda et al. (2003) "Synthesis of a novel histidine analogue and its efficient incorporation into a protein in vivo," Protein Eng. Des. Sel. 16(9):699-706, Chin et al. (2003) "An Expanded Eukaryotic Genetic Code," Science 301(5635):964-967, James et al. (2001) "Kinetic characterization of ribonuclease S mutants containing photoisomerizable phenylazophenylalanine residues," Protein Eng. Des. Sel. 14(12):983-991, Kohrer et al. (2001) "Import of amber and ochre suppressor tRNAs into mammalian cells: A general approach to site-specific insertion of amino acid analogues into proteins," Proc. Natl. Acad. Sci. U.S.A. 98(25):14310-14315, Bacher et al. (2001) "Selection and Characterization of Escherichia coli Variants Capable of Growth on an Otherwise Toxic Tryptophan Analogue," J. Bacteriol. 183(18):5414-5425, Hamano-Takaku et

al. (2000) "A Mutant Escherichia coli Tyrosyl-tRNA Synthetase Utilizes the Unnatural Amino Acid Azatyrosine More Efficiently than Tyrosine," J. Biol. Chem. 275(51):40324-40328, and Budisa et al. (2001) "Proteins with {beta}-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids," Protein Sci. 10(7):1281-1292.

To further illustrate, an amino acid is typically an organic acid that includes a substituted or unsubstituted amino group, a substituted or unsubstituted carboxy group, and one or more side chains or groups, or analogs of any of these groups. Exemplary side chains include, e.g., thiol, seleno, sulfonyl, alkyl, aryl, acyl, keto, azido, hydroxyl, hydrazine, cyano, halo, hydrazide, alkenyl, alkynl, ether, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, ester, thioacid, hydroxylamine, or any combination of these groups. Other representative amino acids include, but are not limited to, amino acids comprising photoactivatable cross-linkers, metal binding amino acids, spin-labeled amino acids, fluorescent amino acids, metal-containing amino acids, amino acids with novel functional groups, amino acids that covalently or noncovalently interact with other molecules, photocaged and/or photoisomerizable amino acids, radioactive amino acids, amino acids comprising biotin or a biotin analog, glycosylated amino acids, other carbohydrate modified amino acids, amino acids comprising polyethylene glycol or polyether, heavy atom substituted amino acids, chemically cleavable and/or photocleavable amino acids, carbon-linked sugar-containing amino acids, redox-active amino acids, amino thioacid containing amino acids, and amino acids comprising one or more toxic moieties.

The term "biological sample" encompasses a variety of sample types obtained from an organism and can be used in a diagnostic or monitoring assay. The term encompasses urine, urine sediment, blood, saliva, and other liquid samples of biological origin, solid tissue samples, such as a biopsy specimen or tissue cultures or cells derived therefrom and the progeny thereof. The term encompasses samples that have been manipulated in any way after their procurement, such as by treatment with reagents, solubilization, sedimentation, or enrichment for certain components. The term encompasses a clinical sample, and also includes cells in cell culture, cell supernatants, cell lysates, serum, plasma, biological fluids, and tissue samples.

The term "mutant," in the context of DNA polymerases of the present invention, means a polypeptide, typically recombinant, that comprises one or more amino acid substitutions relative to a corresponding, functional DNA polymerase.

The term "unmodified form," in the context of a mutant polymerase, is a term used herein for purposes of defining a mutant DNA polymerase of the present invention: the term "unmodified form" refers to a functional DNA polymerase that has the amino acid sequence of the mutant polymerase except at one or more amino acid position(s) specified as characterizing the mutant polymerase. Thus, reference to a mutant DNA polymerase in terms of (a) its unmodified form and (b) one or more specified amino acid substitutions means that, with the exception of the specified amino acid substitution(s), the mutant polymerase otherwise has an amino acid sequence identical to the unmodified form in the specified motif. The "unmodified polymerase" (and therefore also the modified form having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors) may contain additional mutations to provide desired functionality, e.g., improved incorporation of dideoxyribonucleotides, ribonucleotides, ribonucleotide analogs, dye-labeled nucleotides, modulating 5'-nuclease activity, modulating 3'-nuclease (or proofreading) activity, or the like. Accordingly, in carrying out the present invention as described herein, the unmodified form of a DNA polymerase is predetermined. The unmodified form of a DNA polymerase can be, for example, a wild-type and/or a naturally occurring DNA polymerase, or a DNA polymerase that has already been intentionally modified. An unmodified form of the polymerase is preferably a thermostable DNA polymerase, such as DNA polymerases from various thermophilic bacteria, as well as functional variants thereof having substantial sequence identity to a wild-type or naturally occurring thermostable polymerase. Such variants can include, for example, chimeric DNA polymerases such as, for example, the chimeric DNA polymerases described in U.S. Patent Nos. 6,228,628 and 7,148,049. In certain embodiments, the unmodified form of a polymerase has reverse transcriptase (RT) activity.

The term "thermostable polymerase," refers to an enzyme that is stable to heat, is heat resistant, and retains sufficient activity to effect subsequent polynucleotide extension reactions and does not become irreversibly denatured (inactivated) when subjected to the elevated temperatures for the time necessary to effect denaturation of double-stranded nucleic acids. The heating conditions necessary for nucleic acid denaturation are well known in the art and are exemplified in, *e.g.*, U.S. Patent Nos. 4,683,202, 4,683,195, and 4,965,188. As used herein, a thermostable polymerase is suitable for use in a temperature cycling reaction such as the polymerase chain reaction ("PCR"). Irreversible denaturation for purposes herein refers to permanent and

complete loss of enzymatic activity. For a thermostable polymerase, enzymatic activity refers to the catalysis of the combination of the nucleotides in the proper manner to form polynucleotide extension products that are complementary to a template nucleic acid strand. Thermostable DNA polymerases from thermophilic bacteria include, *e.g.*, DNA polymerases from *Thermotoga maritima*, *Thermus aquaticus*, *Thermus thermophilus*, *Thermus flavus*, *Thermus filiformis*, *Thermus* species sps17, *Thermus* species Z05, *Thermus caldophilus*, *Bacillus caldotenax*, *Thermotoga neopolitana*, and *Thermosipho africanus*.

The term "thermoactive" refers to an enzyme that maintains catalytic properties at temperatures commonly used for reverse transcription or anneal/extension steps in RT-PCR and/or PCR reactions (i.e., 45-80 °C). Thermostable enzymes are those which are not irreversibly inactivated or denatured when subjected to elevated temperatures necessary for nucleic acid denaturation. Thermoactive enzymes may or may not be thermostable. Thermoactive DNA polymerases can be DNA or RNA dependent from thermophilic species or from mesophilic species including, but not limited to, *Escherichia coli, Moloney murine leukemia viruses, and Avian myoblastosis virus.*

As used herein, a "chimeric" protein refers to a protein whose amino acid sequence represents a fusion product of subsequences of the amino acid sequences from at least two distinct proteins. A chimeric protein typically is not produced by direct manipulation of amino acid sequences, but, rather, is expressed from a "chimeric" gene that encodes the chimeric amino acid sequence. In certain embodiments, for example, an unmodified form of a mutant DNA polymerase of the present invention is a chimeric protein that consists of an amino-terminal (N-terminal) region derived from a Thermus species DNA polymerase and a carboxy-terminal (C-terminal) region derived from Tma DNA polymerase. The N-terminal region refers to a region extending from the N-terminus (amino acid position 1) to an internal amino acid. Similarly, the C-terminal region refers to a region extending from an internal amino acid to the C-terminus.

The term "aptamer" refers to a single-stranded DNA that recognizes and binds to DNA polymerase, and efficiently inhibits the polymerase activity as described in U.S. Pat. No. 5,693,502. Use of aptamer and dUTP/UNG in RT-PCR is also discussed, for example, in Smith, E.S. et al, (Amplification of RNA: High-temperature Reverse Transcription and DNA Amplification with a Magnesium-activated Thermostable DNA Polymerase, in PCR Primer: A

Laboratory Manual, 2nd Edition, Dieffenbach, C.W. and Dveksler, G.S., Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 211-219, (2003)).

In the context of mutant DNA polymerases, "correspondence" to another sequence (*e.g.*, regions, fragments, nucleotide or amino acid positions, or the like) is based on the convention of numbering according to nucleotide or amino acid position number and then aligning the sequences in a manner that maximizes the percentage of sequence identity. An amino acid "corresponding to position [X] of [specific sequence]" refers to an amino acid in a polypeptide of interest that aligns with the equivalent amino acid of a specified sequence. Generally, as described herein, the amino acid corresponding to a position of a polymerase can be determined using an alignment algorithm such as BLAST as described below. Because not all positions within a given "corresponding region" need be identical, non-matching positions within a corresponding region may be regarded as "corresponding positions." Accordingly, as used herein, referral to an "amino acid position corresponding to amino acid position [X]" of a specified DNA polymerase refers to equivalent positions, based on alignment, in other DNA polymerases and structural homologues and families. In some embodiments of the present invention, "correspondence" of amino acid positions are determined with respect to a region of the polymerase comprising one or more motifs of SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 32, 33, 34, 35, 36, 37, or 48. When a polymerase polypeptide sequence differs from SEQ ID NOS:1, 2, 3, 4, 5, 6, 7, 32, 33, 34, 35, 36, 37, or 48 (*e.g.*, by changes in amino acids or addition or deletion of amino acids), it may be that a particular mutation associated with improved activity as discussed herein will not be in the same position number as it is in SEQ ID NOS:1, 2, 3, 4, 5, 6, 7, 32, 33, 34, 35, 36, 37, or 48. This is illustrated, for example, in Table 1.

“Recombinant” as used herein, refers to an amino acid sequence or a nucleotide sequence that has been intentionally modified by recombinant methods. By the term "recombinant nucleic acid" herein is meant a nucleic acid, originally formed *in vitro*, in general, by the manipulation of a nucleic acid by restriction endonucleases, in a form not normally found in nature. Thus an isolated, mutant DNA polymerase nucleic acid, in a linear form, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell, it will replicate non-recombinantly, *i.e.*, using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such

nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention. A "recombinant protein" is a protein made using recombinant techniques, *i.e.*, through the expression of a recombinant nucleic acid as depicted above.

A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation.

The term "host cell" refers to both single-cellular prokaryote and eukaryote organisms (*e.g.*, bacteria, yeast, and actinomycetes) and single cells from higher order plants or animals when being grown in cell culture.

The term "vector" refers to a piece of DNA, typically double-stranded, which may have inserted into it a piece of foreign DNA. The vector or may be, for example, of plasmid origin. Vectors contain "replicon" polynucleotide sequences that facilitate the autonomous replication of the vector in a host cell. Foreign DNA is defined as heterologous DNA, which is DNA not naturally found in the host cell, which, for example, replicates the vector molecule, encodes a selectable or screenable marker, or encodes a transgene. The vector is used to transport the foreign or heterologous DNA into a suitable host cell. Once in the host cell, the vector can replicate independently of or coincidental with the host chromosomal DNA, and several copies of the vector and its inserted DNA can be generated. In addition, the vector can also contain the necessary elements that permit transcription of the inserted DNA into an mRNA molecule or otherwise cause replication of the inserted DNA into multiple copies of RNA. Some expression vectors additionally contain sequence elements adjacent to the inserted DNA that increase the half-life of the expressed mRNA and/or allow translation of the mRNA into a protein molecule. Many molecules of mRNA and polypeptide encoded by the inserted DNA can thus be rapidly synthesized.

The term "nucleotide," in addition to referring to the naturally occurring ribonucleotide or deoxyribonucleotide monomers, shall herein be understood to refer to related structural variants thereof, including derivatives and analogs, that are functionally equivalent with respect to the particular context in which the nucleotide is being used (*e.g.*, hybridization to a complementary base), unless the context clearly indicates otherwise.

The term "nucleic acid" or "polynucleotide" refers to a polymer that can be corresponded to a ribose nucleic acid (RNA) or deoxyribose nucleic acid (DNA) polymer, or an analog thereof. This includes polymers of nucleotides such as RNA and DNA, as well as synthetic forms, modified (*e.g.*, chemically or biochemically modified) forms thereof, and mixed polymers (*e.g.*, including both RNA and DNA subunits). Exemplary modifications include methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, and the like), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, and the like), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids and the like). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Typically, the nucleotide monomers are linked via phosphodiester bonds, although synthetic forms of nucleic acids can comprise other linkages (*e.g.*, peptide nucleic acids as described in Nielsen *et al.* (*Science* 254:1497-1500, 1991). A nucleic acid can be or can include, e.g., a chromosome or chromosomal segment, a vector (*e.g.*, an expression vector), an expression cassette, a naked DNA or RNA polymer, the product of a polymerase chain reaction (PCR), an oligonucleotide, a probe, and a primer. A nucleic acid can be, *e.g.*, single-stranded, double-stranded, or triple-stranded and is not limited to any particular length. Unless otherwise indicated, a particular nucleic acid sequence optionally comprises or encodes complementary sequences, in addition to any sequence explicitly indicated.

The term "oligonucleotide" refers to a nucleic acid that includes at least two nucleic acid monomer units (e.g., nucleotides). An oligonucleotide typically includes from about six to about 175 nucleic acid monomer units, more typically from about eight to about 100 nucleic acid monomer units, and still more typically from about 10 to about 50 nucleic acid monomer units (e.g., about 15, about 20, about 25, about 30, about 35, or more nucleic acid monomer units). The exact size of an oligonucleotide will depend on many factors, including the ultimate function or use of the oligonucleotide. Oligonucleotides are optionally prepared by any suitable method, including, but not limited to, isolation of an existing or natural sequence, DNA replication or amplification, reverse transcription, cloning and restriction digestion of appropriate sequences, or direct chemical synthesis by a method such as the phosphotriester

method of Narang *et al.* (*Meth. Enzymol.* 68:90-99, 1979); the phosphodiester method of Brown *et al.* (*Meth. Enzymol.* 68:109-151, 1979); the diethylphosphoramidite method of Beaucage *et al.* (*Tetrahedron Lett.* 22:1859-1862, 1981); the triester method of Matteucci *et al.* (*J. Am. Chem. Soc.* 103:3185-3191, 1981); automated synthesis methods; or the solid support method of U.S. Pat. No. 4,458,066 or other methods known to those skilled in the art.

The term "primer" as used herein refers to a polynucleotide capable of acting as a point of initiation of template-directed nucleic acid synthesis when placed under conditions in which polynucleotide extension is initiated (*e.g.*, under conditions comprising the presence of requisite nucleoside triphosphates (as dictated by the template that is copied) and a polymerase in an appropriate buffer and at a suitable temperature or cycle(s) of temperatures (e.g., as in a polymerase chain reaction)). To further illustrate, primers can also be used in a variety of other oligonuceotide-mediated synthesis processes, including as initiators of *de novo* RNA synthesis and *in vitro* transcription-related processes (e.g., nucleic acid sequence-based amplification (NASBA), transcription mediated amplification (TMA), etc.). A primer is typically a single-stranded oligonucleotide (*e.g.*, oligodeoxyribonucleotide). The appropriate length of a primer depends on the intended use of the primer but typically ranges from 6 to 40 nucleotides, more typically from 15 to 35 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template for primer elongation to occur. In certain embodiments, the term "primer pair" means a set of primers including a 5' sense primer (sometimes called “forward”) that hybridizes with the complement of the 5' end of the nucleic acid sequence to be amplified and a 3' antisense primer (sometimes called “reverse”) that hybridizes with the 3' end of the sequence to be amplified (*e.g.,* if the target sequence is expressed as RNA or is an RNA). A primer can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include $^{32}P$, fluorescent dyes, electron-dense reagents, enzymes (as commonly used in ELISA assays), biotin, or haptens and proteins for which antisera or monoclonal antibodies are available.

The term "conventional" or "natural" when referring to nucleic acid bases, nucleoside triphosphates, or nucleotides refers to those which occur naturally in the polynucleotide being described (*i.e.*, for DNA these are dATP, dGTP, dCTP and dTTP). Additionally, dITP, and 7-

deaza-dGTP are frequently utilized in place of dGTP and 7-deaza-dATP can be utilized in place of dATP in *in vitro* DNA synthesis reactions, such as sequencing. Collectively, these may be referred to as dNTPs.

The term "unconventional" or "modified" when referring to a nucleic acid base, nucleoside, or nucleotide includes modification, derivations, or analogues of conventional bases, nucleosides, or nucleotides that naturally occur in a particular polynucleotide. Certain unconventional nucleotides are modified at the 2' position of the ribose sugar in comparison to conventional dNTPs. Thus, although for RNA the naturally occurring nucleotides are ribonucleotides (*i.e.*, ATP, GTP, CTP, UTP, collectively rNTPs), because these nucleotides have a hydroxyl group at the 2' position of the sugar, which, by comparison is absent in dNTPs, as used herein, ribonucleotides are unconventional nucleotides as substrates for DNA polymerases. As used herein, unconventional nucleotides include, but are not limited to, compounds used as terminators for nucleic acid sequencing. Exemplary terminator compounds include but are not limited to those compounds that have a 2',3' dideoxy structure and are referred to as dideoxynucleoside triphosphates. The dideoxynucleoside triphosphates ddATP, ddTTP, ddCTP and ddGTP are referred to collectively as ddNTPs. Additional examples of terminator compounds include 2'-$PO_4$ analogs of ribonucleotides (*see, e.g.*, U.S. Application Publication Nos. 2005/0037991 and 2005/0037398). Other unconventional nucleotides include phosphorothioate dNTPs ([α-S]dNTPs), 5'-[α-borano]-dNTPs, [α]-methyl-phosphonate dNTPs, and ribonucleoside triphosphates (rNTPs). Unconventional bases may be labeled with radioactive isotopes such as $^{32}P$, $^{33}P$, or $^{35}S$; fluorescent labels; chemiluminescent labels; bioluminescent labels; hapten labels such as biotin; or enzyme labels such as streptavidin or avidin. Fluorescent labels may include dyes that are negatively charged, such as dyes of the fluorescein family, or dyes that are neutral in charge, such as dyes of the rhodamine family, or dyes that are positively charged, such as dyes of the cyanine family. Dyes of the fluorescein family include, *e.g.*, FAM, HEX, TET, JOE, NAN and ZOE. Dyes of the rhodamine family include Texas Red, ROX, R110, R6G, and TAMRA. Various dyes or nucleotides labeled with FAM, HEX, TET, JOE, NAN, ZOE, ROX, R110, R6G, Texas Red and TAMRA are marketed by Perkin-Elmer (Boston, MA), Applied Biosystems (Foster City, CA), or Invitrogen/Molecular Probes (Eugene, OR). Dyes of the cyanine family include Cy2, Cy3, Cy5, and Cy7 and are

marketed by GE Healthcare UK Limited (Amersham Place, Little Chalfont, Buckinghamshire, England).

As used herein, "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the sequence in the comparison window can comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

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. 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 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% 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. Sequences are "substantially identical" to each other if they are at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, or at least 55% identical. These definitions also refer to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 nucleotides in length, or more typically over a region that is 100 to 500 or 1000 or more nucleotides in length.

The terms "similarity" or "percent similarity," in the context of two or more polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of amino acid residues that are either the same or similar as defined by a conservative amino acid substitutions (*e.g.*, 60% similarity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% similar 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. Sequences are "substantially similar" to each other if they are at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, or at least 55% similar to each other. Optionally, this similarly exists over a region that is at least about 50 amino acids in length, or more typically over a region that is at least about 100 to 500 or 1000 or more amino acids in length.

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

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, for example, by the local homology algorithm of Smith and Waterman (*Adv. Appl. Math.* 2:482, 1970), by the homology alignment algorithm of Needleman and Wunsch (*J. Mol. Biol.* 48:443, 1970), by the search for similarity method of Pearson and Lipman (*Proc. Natl. Acad. Sci. USA* 85:2444, 1988), by computerized implementations of these algorithms (*e.g.*, GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (*see*, *e.g.*, Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995 supplement)).

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

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see*, *e.g.*, Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-87, 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, typically less than about 0.01, and more typically less than about 0.001.

The term "reverse transcription efficiency" refers to the fraction of RNA molecules that are reverse transcribed as cDNA in a given reverse transcription reaction. In certain embodiments, the mutant DNA polymerases of the invention have improved reverse transcription efficiencies relative to unmodified forms of these DNA polymerases. That is, these mutant DNA polymerases reverse transcribe a higher fraction of RNA templates than their unmodified forms

under a particular set of reaction conditions. Reverse transcription efficiency can be measured, for example, by measuring the crossing point (Cp) of a PCR reaction using a RNA template, and comparing the Cp value to a Cp value of a control reaction in which a DNA template of the same sequence (except U's are replaced with T's) is amplified, wherein the RNA and DNA amplifications use a common primer set and the same polymerase, e.g., as described in the examples. A test polymerase has improved RT efficiency when the test polymerase has a decreased Cp value compared to a control polymerase when RNA is used as a template, but has a substantially unchanged Cp value relative to the control polymerase when DNA is used as a template. In some embodiments a polymerase of the invention has an improved RT efficiency such that the Cp is at least one, two, three, four, or five units less than the corresponding control polymerase on the RNA template.

The term "mismatch tolerance" refers to the ability of a polymerase to tolerate a mismatch-containing sequence when extending a nucleic acid (e.g., a primer or other oligonucleotide) in a template-dependent manner by attaching (e.g., covalently) one or more nucleotides to the nucleic acid. The term "3' mismatch tolerance" refers to the ability of a polymerase to tolerate a mismatch-containing (nearly complementary) sequence where the nucleic acid to be extended (e.g., a primer or other oligonucleotide) has a mismatch with its template at the 3' terminal nucleotide of the primer. Mismatches to the template may also be located at the 3' penultimate nucleotide of the primer, or at another position within the sequence of the primer.

The term "mismatch discrimination" refers to the ability of a polymerase to distinguish a fully complementary sequence from a mismatch-containing sequence when extending a nucleic acid (e.g., a primer or other oligonucleotide) in a template-dependent manner by attaching (e.g., covalently) one or more nucleotides to the nucleic acid. The term "3'-mismatch discrimination" refers to the ability of a polymerase to distinguish a fully complementary sequence from a mismatch-containing (nearly complementary) sequence where the nucleic acid to be extended (e.g., a primer or other oligonucleotide) has a mismatch at the nucleic acid's 3' terminus compared to the template to which the nucleic acid hybridizes. The term "mismatch" refers to the existence of one or more base mispairings (or "noncomplementary base oppositions") within a stretch of otherwise complementary duplex-forming (or potentially duplex-forming) sequences.

The term "Cp value" or "crossing point" value refers to a value that allows quantification of input target nucleic acids. The Cp value can be determined according to the second-derivative maximum method (Van Luu-The, et al., "Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction," BioTechniques, Vol. 38, No. 2, February 2005, pp. 287–293). In the second derivative method, a Cp corresponds to the first peak of a second derivative curve. This peak corresponds to the beginning of a log-linear phase. The second derivative method calculates a second derivative value of the real-time fluorescence intensity curve, and only one value is obtained. The original Cp method is based on a locally defined, differentiable approximation of the intensity values, *e.g.*, by a polynomial function. Then the third derivative is computed. The Cp value is the smallest root of the third derivative. The Cp can also be determined using the fit point method, in which the Cp is determined by the intersection of a parallel to the threshold line in the log-linear region (Van Luu-The, et al., BioTechniques, Vol. 38, No. 2, February 2005, pp. 287–293). The Cp value provided by the LightCycler instrument offered by Roche by calculation according to the second-derivative maximum method.

The term "PCR efficiency" refers to an indication of cycle to cycle amplification efficiency. PCR efficiency is calculated for each condition using the equation: % PCR efficiency = $(10^{(-\text{slope})}-1) \times 100$, wherein the slope was calculated by linear regression with the log copy number plotted on the y-axis and Cp plotted on the x-axis. PCR efficiency can be measured using a perfectly matched or mismatched primer template.

The term "nucleic acid extension rate" refers the rate at which a biocatalyst (*e.g.*, an enzyme, such as a polymerase, ligase, or the like) extends a nucleic acid (*e.g.*, a primer or other oligonucleotide) in a template-dependent or template-independent manner by attaching (*e.g.*, covalently) one or more nucleotides to the nucleic acid. To illustrate, certain mutant DNA polymerases described herein have improved nucleic acid extension rates relative to unmodified forms of these DNA polymerases, such that they can extend primers at higher rates than these unmodified forms under a given set of reaction conditions.

The term "tolerance of RT and polymerase inhibitors" refers to the ability of a polymerase to maintain activity (polymerase or reverse transcription activity) in the presence of an amount of an inhibitor that would inhibit the polymerase activity or reverse transcription activity of a control polymerase. In some embodiments, the improved polymerase is capable of polymerase

or reverse transcription activity in the presence of an amount of the inhibitor that would essentially eliminate the control polymerase activity. A "control polymerase" refers to a polymerase that comprises an isoleucine (I) corresponding to position 709 of SEQ ID NO:1 but is otherwise identical to the improved polymerase.

The term "5'-nuclease probe" refers to an oligonucleotide that comprises at least one light emitting labeling moiety and that is used in a 5'-nuclease reaction to effect target nucleic acid detection. In some embodiments, for example, a 5'-nuclease probe includes only a single light emitting moiety (e.g., a fluorescent dye, etc.). In certain embodiments, 5'-nuclease probes include regions of self-complementarity such that the probes are capable of forming hairpin structures under selected conditions. To further illustrate, in some embodiments a 5'-nuclease probe comprises at least two labeling moieties and emits radiation of increased intensity after one of the two labels is cleaved or otherwise separated from the oligonucleotide. In certain embodiments, a 5'-nuclease probe is labeled with two different fluorescent dyes, e.g., a 5' terminus reporter dye and the 3' terminus quencher dye or moiety. In some embodiments, 5'-nuclease probes are labeled at one or more positions other than, or in addition to, terminal positions. When the probe is intact, energy transfer typically occurs between the two fluorophores such that fluorescent emission from the reporter dye is quenched at least in part. During an extension step of a polymerase chain reaction, for example, a 5'-nuclease probe bound to a template nucleic acid is cleaved by the 5' to 3' nuclease activity of, e.g., a *Taq* polymerase or another polymerase having this activity such that the fluorescent emission of the reporter dye is no longer quenched. Exemplary 5'-nuclease probes are also described in, e.g., U.S. Pat. No. 5,210,015, U.S. Pat. No. 5,994,056, and U.S. Pat. No. 6,171,785. In other embodiments, a 5' nuclease probe may be labeled with two or more different reporter dyes and a 3' terminus quencher dye or moiety.

The term "FRET" or "fluorescent resonance energy transfer" or "Foerster resonance energy transfer" refers to a transfer of energy between at least two chromophores, a donor chromophore and an acceptor chromophore (referred to as a quencher). The donor typically transfers the energy to the acceptor when the donor is excited by light radiation with a suitable wavelength. The acceptor typically re-emits the transferred energy in the form of light radiation with a different wavelength. When the acceptor is a "dark" quencher, it dissipates the transferred energy in a form other than light. Whether a particular fluorophore acts as a donor or an

acceptor depends on the properties of the other member of the FRET pair. Commonly used donor-acceptor pairs include the FAM-TAMRA pair. Commonly used quenchers are DABCYL and TAMRA. Commonly used dark quenchers include BlackHole Quenchers™ (BHQ), (Biosearch Technologies, Inc., Novato, Cal.), Iowa Black™ (Integrated DNA Tech., Inc., Coralville, Iowa), and BlackBerry™ Quencher 650 (BBQ-650) (Berry & Assoc., Dexter, Mich.).

## BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1** depicts an amino acid sequence alignment of a region from the polymerase domain of exemplary DNA polymerases from various species of bacteria: *Thermus* species Z05 (Z05) (SEQ ID NO:12), *Thermus aquaticus* (Taq) (SEQ ID NO:13), *Thermus filiformus* (Tfi) (SEQ ID NO:14), *Thermus flavus* (Tfl) (SEQ ID NO:15), *Thermus* species sps17 (Sps17) (SEQ ID NO:16), *Thermus thermophilus* (Tth) (SEQ ID NO:17), *Thermus caldophilus* (Tca) (SEQ ID NO:18), *Thermotoga maritima* (Tma) (SEQ ID NO:19), *Thermotoga neopolitana* (Tne) (SEQ ID NO:20), *Thermosipho africanus* (Taf) (SEQ ID NO:21), *Deinococcus radiodurans* (Dra) (SEQ ID NO:23), *Bacillus stearothermophilus* (Bst) (SEQ ID NO:24), and *Bacillus caldotenax* (Bca) (SEQ ID NO:25). In addition, the polypeptide regions shown comprise the amino acid motif $X_1$-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-$X_7$-$X_8$-$X_9$-$X_{10}$-$X_{11}$-$X_{12}$-$X_{13}$-G-Y-V-$X_{14}$-T-L (SEQ ID NO:26), the variable positions of which are further defined herein. This motif is highlighted in bold type for each polymerase sequence. Amino acid positions amenable to mutation in accordance with the present invention are indicated with an asterisk (*). Gaps in the alignments are indicated with a dot (.).

**Figure 2** provides sequence identities among the following DNA Polymerase I enzymes: *Thermus sp.* Z05 DNA polymerase (Z05); *Thermus aquaticus* DNA polymerase (Taq); *Thermus filiformis* DNA polymerase (Tfi); *Thermus flavus* DNA polymerase (Tfl); *Thermus sp. sps17* DNA polymerase (Sps17); *Thermus thermophilus* DNA polymerase (Tth); *Thermus caldophilus* DNA polymerase (Tca); *Deinococcus radiodurans* DNA polymerase (Dra); *Thermotoga maritima* DNA polymerase (Tma); *Thermotoga neopolitana* DNA polymerase (Tne); *Thermosipho africanus* DNA polymerase (Taf); *Bacillus stearothermophilus* DNA polymerase (Bst); and *Bacillus caldotenax* DNA polymerase (Bca). **(A)** sequence identities over the entire

polymerase I enzyme (corresponding to amino acids 1-834 of Z05); and **(B)** sequence identities over the polymerase sub domain corresponding to amino acids 420-834 of Z05.

**Figure 3** provides sequence identities among various *Thermus* sp DNA Polymerase I enzymes: *Thermus sp.* Z05 DNA polymerase (Z05); *Thermus aquaticus* DNA polymerase (Taq); *Thermus filiformis* DNA polymerase (Tfi); *Thermus flavus* DNA polymerase (Tfl); *Thermus sp. sps17* DNA polymerase (Sps17); *Thermus thermophilus* DNA polymerase (Tth); and *Thermus caldophilus* DNA polymerase (Tca). **(A)** sequence identities over the entire polymerase I enzyme (corresponding to amino acids 1-834 of Z05); and **(B)** sequence identities over the polymerase sub domain corresponding to amino acids 420-834 of Z05.

## DETAILED DESCRIPTION

The present invention provides improved DNA polymerases in which one or more amino acids in the polymerase domain have been mutated relative to a functional DNA polymerase. The DNA polymerases of the invention are active enzymes having increased reverse transcriptase efficiency (e.g., in the presence of $Mn^{2+}$ and $Mg^{2+}$ divalent cations) relative to the unmodified form of the polymerase and/or increased mismatch tolerance, extension rate and tolerance of RT and polymerase inhibitors. In certain embodiments, the mutant DNA polymerases may be used at lower concentrations for superior or equivalent performance as the parent enzymes.

DNA polymerases that more efficiently perform reverse transcription are helpful, for example, in a variety of applications involving assays that employ RT-PCR to detect and/or quantify RNA targets. The DNA polymerases are therefore useful in a variety of applications involving polynucleotide extension as well as reverse transcription or amplification of polynucleotide templates, including, for example, applications in recombinant DNA studies and medical diagnosis of disease. The mutant DNA polymerases are also particularly useful, because of their tolerance for mis-matches, for detecting targets that possibly have variable sequences (e.g., viral targets, or cancer and other disease genetic markers).

DNA polymerases of the invention can be characterized by having the following motif:

$X_1$-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-$X_7$-$X_8$-$X_9$-$X_{10}$-$X_{11}$-$X_{12}$-$X_{13}$-Gly-Tyr-Val-$X_{14}$-Thr-Leu

(also referred to herein in the one-letter code as $X_1$-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-$X_7$-$X_8$-$X_9$-$X_{10}$-$X_{11}$-$X_{12}$-$X_{13}$-G-Y-V-$X_{14}$-T-L) (SEQ ID NO:8); wherein

$X_1$ is Ala (A), Asp (D), Ser (S), Glu (E), Arg (R) or Gln (Q);

$X_2$ is Trp (W) or Tyr (Y);

$X_3$ is any amino acid other than Ile (I), Leu (L) or Met (M);

$X_4$ is Glu (E), Ala (A), Gln (Q), Lys (K), Asn (N) or Asp (D);

$X_5$ is Lys (K), Gly (G), Arg (R), Gln (Q), His (H) or Asn (N);

$X_6$ is Thr (T), Val (V), Met (M) or Ile (I);

$X_7$ is Leu (L), Val (V) or Lys (K);

$X_8$ is Glu (E), Ser (S), Ala (A), Asp (D) or Gln (Q);

$X_9$ is Glu (E) or Phe (F);

$X_{10}$ is Gly (G) or Ala (A);

$X_{11}$ is Arg (R) or Lys (K);

$X_{12}$ is Lys (K), Arg (R), Glu (E), Thr (T) or Gln (Q);

$X_{13}$ is Arg (R), Lys (K) or His (H); and

$X_{14}$ is Glu (E), Arg (R) or Thr (T).

In some embodiments, $X_3$ is is selected from G, A, W, P, S, T, F, Y, C, N, Q, D, E, K, V, R or H.

In some embodiments, DNA polymerases of the invention can be characterized by having the following motif:

$X_1$-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-$X_7$-$X_8$-Glu-$X_{10}$-$X_{11}$-$X_{12}$-$X_{13}$-Gly-Tyr-Val-$X_{a14}$-Thr-Leu (also referred to herein in the one-letter code as $X_1$-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-$X_7$-$X_8$-E-$X_{10}$-$X_{11}$-$X_{12}$-$X_{13}$-G-Y-V-$X_{14}$-T-L) (SEQ ID NO:9); wherein

$X_1$ is Ala (A), Asp (D) or Ser (S);

$X_2$ is Trp (W) or Tyr (Y);

$X_3$ is any amino acid other than Ile (I);

$X_4$ is Glu (E), Ala (A) or Gln (Q);

$X_5$ is Lys (K), Gly (G), Arg (R) or Gln (Q);

$X_6$ is Thr (T) or Val (V);

$X_7$ is Leu (L) or Val (V);

$X_8$ is Glu (E), Ser (S) or Ala (A);

$X_{10}$ is Gly (G) or Ala (A);

$X_{11}$ is Arg (R) or Lys (K);

$X_{12}$ is Lys (K), Arg (R) or Glu (E);

$X_{13}$ is Arg (R) or Lys (K); and

$X_{14}$ is Glu (E) or Arg (R)

In some embodiments, DNA polymerases of the invention can be characterized by having the following motif:

Ala-Trp-$X_3$-$X_4$-$X_5$-Thr-Leu-Glu-Glu-Gly-Arg-$X_{12}$-$X_{13}$-Gly-Tyr-Val-Glu-Thr-Leu (also referred to herein in the one-letter code as A-W-$X_3$-$X_4$-$X_5$-T-L-E-E-G-R-$X_{12}$-$X_{13}$-G-Y-V-E-T-L) (SEQ ID NO:10); wherein

$X_3$ is any amino acid other than Ile (I);

$X_4$ is Glu (E) or Ala (A);

$X_5$ is Lys (K) or Gly (G);

$X_{12}$ is Lys (K) or Arg (R); and

$X_{13}$ is Arg (R) or Lys (K).

In some embodiments, the DNA polymerase comprising the motif of SEQ ID NO:9 or SEQ ID NO:10 is not SEQ ID NO:2. In some embodiments, the amino acid corresponding to position $X_3$ of SEQ ID NO:9 or SEQ ID NO:10 is any amino acid other than Leu (L).

In some embodiments, DNA polymerases of the invention can be characterized by having the following motif:

Ala-Trp-$X_3$-$X_4$-$X_5$-Thr-Leu-Glu-Glu-Gly-Arg-$X_{12}$-$X_{13}$-Gly-Tyr-Val-Glu-Thr-Leu (also referred to herein in the one-letter code as A-W-$X_3$-$X_4$-$X_5$-T-L-E-E-G-R- $X_{12}$-$X_{13}$-G-Y-V-E-T-L) (SEQ ID NO:11); wherein

$X_3$ is Lys (K), Arg (R), Ser (S), Gly (G) or Ala (A);

$X_4$ is Glu (E) or Ala (A);

$X_5$ is Lys (K) or Gly (G);

$X_{12}$ is Lys (K) or Arg (R); and

$X_{13}$ is Arg (R) or Lys (K).

This motif is present within the "fingers" domain of many Family A type DNA-dependent DNA polymerases, particularly thermostable DNA polymerases from thermophilic bacteria (Li et al., *EMBO J.* 17:7514-7525, 1998). For example, Figure 1 shows an amino acid sequence alignment of a region from the "fingers" domain of DNA polymerases from several species of bacteria:

*Bacillus caldotenax*, *Bacillus stearothermophilus*, *Deinococcus radiodurans*, *Thermosipho africanus*, *Thermotoga maritima*, *Thermotoga neopolitana*, *Thermus aquaticus*, *Thermus caldophilus*, *Thermus filiformus*, *Thermus flavus*, *Thermus* sp. sps17, *Thermus* sp. Z05, and *Thermus thermophilus*. As shown, the native sequence corresponding to the motif above is present in each of these polymerases, indicating a conserved function for this region of the polymerase. Figure 2 provides sequence identities among these DNA polymerases.

Accordingly, in some embodiments, the invention provides for a polymerase comprising SEQ ID NO:8, 9, 10, or 11, having the improved activity and/or characteristics described herein, and wherein the DNA polymerase is otherwise a wild-type or a naturally occurring DNA polymerase, such as, for example, a polymerase from any of the species of thermophilic bacteria listed above, or is substantially identical to such a wild-type or a naturally occurring DNA polymerase. For example, in some embodiments, the polymerase of the invention comprises SEQ ID NO:8, 9, 10, or 11 and is at least 80%, 85%, 90%, or 95% identical to SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 32, 33, 34, 35, 36, 37, or 48. In one variation, the unmodified form of the polymerase is from a species of the genus *Thermus*. In other embodiments of the invention, the unmodified polymerase is from a thermophilic species other than *Thermus*, e.g., *Thermotoga*. The full nucleic acid and amino acid sequence for numerous thermostable DNA polymerases are available. The sequences each of *Thermus aquaticus* (Taq) (SEQ ID NO:2), *Thermus thermophilus* (Tth) (SEQ ID NO:6), *Thermus* species Z05 (SEQ ID NO:1), *Thermus* species sps17 (SEQ ID NO:5), *Thermotoga maritima* (Tma) (SEQ ID NO:34), and *Thermosipho africanus* (Taf) (SEQ ID NO:33) polymerase have been published in PCT International Patent Publication No. WO 92/06200. The sequence for the DNA polymerase from *Thermus flavus* (SEQ ID NO:4) has been published in Akhmetzjanov and Vakhitov (*Nucleic Acids Research* 20:5839, 1992). The sequence of the thermostable DNA polymerase from *Thermus caldophilus* (SEQ ID NO:7) is found in EMBL/GenBank Accession No. U62584. The sequence of the thermostable DNA polymerase from *Thermus filiformis* (SEQ ID NO:3) can be recovered from ATCC Deposit No. 42380 using, e.g., the methods provided in U.S. Pat. No. 4,889,818, as well as the sequence information provided in Table 1. The sequence of the *Thermotoga neapolitana* DNA polymerase (SEQ ID NO:35) is from GeneSeq Patent Data Base Accession No. R98144 and PCT WO 97/09451. The sequence of the thermostable DNA polymerase from *Bacillus caldotenax* (SEQ ID NO:37) is described in, e.g., Uemori et al. (*J Biochem (Tokyo)* 113(3):401-

410, 1993; *see also*, Swiss-Prot database Accession No. Q04957 and GenBank Accession Nos. D12982 and BAA02361). Examples of unmodified forms of DNA polymerases that can be modified as described herein are also described in, e.g., U.S. Pat. Nos. 6,228,628; 6,346,379; 7,030,220; 6,881,559; 6,794,177; 6,468,775, and U.S. Pat. Nos. 7,148,049; 7,179,590; 7,410,782; 7,378,262. Representative full length polymerase sequences are also provided in the sequence listing.

Also amenable to the mutations described herein are functional DNA polymerases that have been previously modified (*e.g.*, by amino acid substitution, addition, or deletion). In some embodiments, such functional modified polymerases retain the amino acid motif of SEQ ID NO:8 (or a motif of SEQ ID NO:9, 10 or 11), and optionally the amino acid motif of SEQ ID NO:38. Thus, suitable unmodified DNA polymerases also include functional variants of wild-type or naturally occurring polymerases. Such variants typically will have substantial sequence identity or similarity to the wild-type or naturally occurring polymerase, typically at least 80% sequence identity and more typically at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity.

In some embodiments, the polymerase of the invention, as well as having a polymerase domain comprising SEQ ID NOS:8, 9, 10, or 11 also comprises a nuclease domain (e.g., corresponding to positions 1 to 291 of Z05).

In some embodiments, a polymerase of the invention is a chimeric polymerase, i.e., comprising polypeptide regions from two or more enzymes. Examples of such chimeric DNA polymerases are described in, *e.g.*, U.S. Patent No. 6,228,628. Particularly suitable are chimeric CS-family DNA polymerases, which include the CS5 (SEQ ID NO:27) and CS6 (SEQ ID NO:28) polymerases and variants thereof having substantial amino acid sequence identity or similarity to SEQ ID NO:27 or SEQ ID NO:28 (typically at least 80% amino acid sequence identity and more typically at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% amino acid sequence identity) and can thus be modified to contain SEQ ID NO:8. The CS5 and CS6 DNA polymerases are chimeric enzymes derived from *Thermus* sp. Z05 and *Thermotoga maritima* (*Tma*) DNA polymerases. They comprise the N-terminal 5'-nuclease domain of the *Thermus* enzyme and the C-terminal 3'-5' exonuclease and the polymerase domains of the *Tma* enzyme. These enzymes have efficient reverse transcriptase activity, can extend nucleotide analog-containing primers, and can incorporate alpha-phosphorothioate dNTPs, dUTP, dITP, and also

fluorescein- and cyanine-dye family labeled dNTPs. The CS5 and CS6 polymerases are also efficient $Mg^{2+}$-activated PCR enzymes. The CS5 and CS6 chimeric polymerases are further described in, e.g., U.S. Pat. No. 7,148,049.

In some embodiments, the amino acid substitutions are single amino acid substitutions. The DNA polymerases provided herein can comprise one or more amino acid substitutions in the active site relative to the unmodified polymerase. In some embodiments, the amino acid substitution(s) comprise at least position $X_3$ of the motif set forth in SEQ ID NO:8 (or a motif of SEQ ID NO:9, 10 or 11). Amino acid substitution at this position confers increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors, yielding a mutant DNA polymerase with an increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors relative to the unmodified polymerase. Typically, the amino acid at position $X_3$ is substituted with an amino acid that does not correspond to the native sequence within the motif set forth in SEQ ID NO:8 (or a motif of SEQ ID NO:9, 10 or 11). Thus, typically, the amino acid at position $X_3$, if substituted, is not Ile (I), Leu (L) or Met (M) as these positions occur in naturally-occurring polymerases. *See*, *e.g.*, Figure 1. In certain embodiments, amino acid substitutions include G, A, W, P, S, T, F, Y, C, N, Q, D, E, K, V, R or H at position $X_3$. In certain embodiments, amino acid substitutions include Lysine (K), Arginine (R), Serine (S), Glycine (G) or Alanine (A) at position $X_3$. Other suitable amino acid substitution(s) at one or more of the identified sites can be determined using, *e.g.*, known methods of site-directed mutagenesis and determination of polynucleotide extension performance in assays described further herein or otherwise known to persons of skill in the art.

In some embodiments, the polymerase of the invention comprises SEQ ID NO:8, 9, 10, or 11 and further comprises one or more additional amino acid changes (*e.g.*, by amino acid substitution, addition, or deletion) compared to a native polymerase. In some embodiments, such polymerases retain the amino acid motif of SEQ ID NO:8 (or a motif of SEQ ID NO:9, 10 or 11), and further comprise the amino acid motif of SEQ ID NO:38 (corresponding to the D580X mutation of Z05 (SEQ ID NO:1)) as follows:

Thr-Gly-Arg-Leu-Ser-Ser-$X_7$-$X_8$-Pro-Asn-Leu-Gln-Asn
(also referred to herein in the one-letter code as
T-G-R-L-S-S-$X_7$-$X_8$-P-N-L-Q-N) (SEQ ID NO:38); wherein

$X_7$ is Ser (S) or Thr (T); and

$X_8$ is any amino acid other than Asp (D) or Glu (E)

The mutation characterized by SEQ ID NO:38 is discussed in more detail in, e.g., US Patent Publication No. 2009/0148891. Such functional variant polymerases typically will have substantial sequence identity or similarity to the wild-type or naturally occurring polymerase (e.g., SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 32, 33, 34, 35, 36, 37, or 48), typically at least 80% amino acid sequence identity and more typically at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% amino acid sequence identity.

In some embodiments, the polymerase of the invention comprises SEQ ID NO:8, 9, 10, or 11 and further comprises one or more additional amino acid changes (*e.g.*, by amino acid substitution, addition, or deletion) compared to a native polymerase. In some embodiments, such polymerases retain the amino acid motif of SEQ ID NO:8 (or a motif of SEQ ID NO:9, 10 or 11), and further comprise the amino acid motif of SEQ ID NO:29 (corresponding to the I588X mutation of Z05 (SEQ ID NO:1)) as follows:

Pro-Asn-Leu-Gln-Asn-$X_1$-Pro-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-Gly (also referred to herein in the one-letter code as P-N-L-Q-N-$X_1$-P-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-G) (SEQ ID NO:29); wherein

$X_1$ is Ile (I), or Leu (L);

$X_2$ is any amino acid other than Ile (I) or Val (V);

$X_3$ is Arg (R) or Lys (K);

$X_4$ is Thr (T), Ser (S) or Leu (L);

$X_5$ is Pro (P) or Glu (E); and

$X_6$ is Leu (L) or Glu (E).

In some embodiments, such functional variant polymerases typically will have substantial sequence identity or similarity to the wild-type or naturally occurring polymerase (e.g., SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 32, 33, 34, 35, 36, 37, or 48), typically at least 80% amino acid sequence identity and more typically at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% amino acid sequence identity.

In some embodiments, the DNA polymerase of the invention comprises amino acid substitution at position $X_3$ (e.g., as in a motif selected from SEQ ID NO:8, 9, 10 or 11) and comprises an amino acid substitution corresponding to SEQ ID NO:38 and SEQ ID NO:29.

Other suitable amino acid substitution(s) at one or more of the identified sites can be determined using, *e.g.*, known methods of site-directed mutagenesis and determination of polynucleotide extension performance in assays described further herein or otherwise known to persons of skill in the art, e.g., amino acid substitutions described in U.S. Pat. Application Publication Nos. 2009/0148891 and 2009/0280539.

Because the precise length of DNA polymerases vary, the precise amino acid positions corresponding to each of $X_3$ (e.g., of SEQ ID NOs: 8, 9, 10, and 11), $X_8$ (of SEQ ID NO:38), and $X_2$ (of SEQ ID NO:29) can vary depending on the particular mutant polymerase used. Amino acid and nucleic acid sequence alignment programs are readily available (*see, e.g.,* those referred to *supra*) and, given the particular motifs identified herein, serve to assist in the identification of the exact amino acids (and corresponding codons) for modification in accordance with the present invention. The positions corresponding to each of $X_3$, $X_8$ and $X_2$ are shown in Table 1 for representative chimeric thermostable DNA polymerases and thermostable DNA polymerases from exemplary thermophilic species.

**Table 1.** Amino Acid Positions Corresponding to Motif Positions $X_3$ (e.g., of SEQ ID NOs: 8, 9, 10, and 11), $X_8$ (of SEQ ID NO:38), and $X_2$ (of SEQ ID NO:29) in Exemplary Polymerases.

| **Organism or Chimeric Sequence** | **Amino Acid Position** | | |
|---|---|---|---|
| Consensus (SEQ ID NO:) | **$X_3$** | **$X_8$** (of SEQ ID NO:38) | **$X_2$** (of SEQ ID NO:29) |
| *T. thermophilus* (6) | 709 | 580 | 588 |
| *T. caldophilus* (7) | 709 | 580 | 588 |
| *T.* sp. Z05 (1) | 709 | 580 | 588 |
| *T. aquaticus* (2) | 707 | 578 | 586 |
| *T. flavus* (4) | 706 | 577 | 585 |
| *T. filiformis* (3) | 705 | 576 | 584 |
| *T.* sp. sps17 (5) | 705 | 576 | 584 |

| Organism or Chimeric Sequence | Amino Acid Position | | |
|---|---|---|---|
| Consensus (SEQ ID NO:) | $X_3$ | $X_8$ (of SEQ ID NO:38) | $X_2$ (of SEQ ID NO:29) |
| *T. maritima* (34) | 770 | 640 | 648 |
| *T. neapolitana* (35) | 770 | 640 | 648 |
| *T. africanus* (33) | 769 | 639 | 647 |
| *B. caldotenax* (37) | 751 | 621 | 629 |
| *B. stearothermophilus* (36) | 750 | 620 | 628 |
| CS5 (27) | 770 | 640 | 648 |
| CS6 (28) | 770 | 640 | 648 |

In some embodiments, the DNA polymerase of the present invention is derived from *Thermus sp.* Z05 DNA polymerase (SEQ ID NO:1) or a variant thereof (*e.g.*, carrying the D580G mutation or the like). As referred to above, in *Thermus* sp. Z05 DNA polymerase, position $X_3$ corresponds to Isoleucine (I) at position 709; position $X_8$ corresponds to Aspartate (D) at position 580. Thus, in certain variations of the invention, the mutant polymerase comprises at least one amino acid substitution, relative to a *Thermus sp.* Z05 DNA polymerase, at I709 and/or D580. Thus, typically, the amino acid at position 709 is not I. In some embodiments, the amino acid at position 709 is selected from G, A, V, L, R, M, F, W, P, S, T, C, Y, N, Q, D, E, K, or H. In certain embodiments, amino acid residue at position 709 is K, R, S, G or A. In certain embodiments, amino acid residues at position D580 can be selected from Leucine (L), Glycine (G), Threonine (T), Glutamine (Q), Alanine (A), Serine (S), Asparagine (N), Arginine (R), and Lysine (K). Further, in certain embodiments, the amino acid at position 588 of SEQ ID NO: 1 is any amino acid other than I. In some embodiments, the amino acid at position 588 of SEQ ID NO: 1 is selected from L, V, G, A, S, M, F, W, P, R, K, T, C, Y, N, Q, D, E or H. In some embodiments, the amino acid at position 588 of SEQ ID NO: 1 is T.

Exemplary *Thermus sp.* Z05 DNA polymerase mutants include those comprising the amino acid substitution(s) I709K (or I709R, I709S, I709G, I709A), and/or I588T, and/or D580G. In some embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, *e.g.*, amino acid residue substitutions I709K (or I709R, I709S, I709G, I709A), I588T, and D580G. In some

embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, *e.g.*, amino acid residue substitutions I709K and D580G, or I709R and D580G, I709S and D580G, I709G and D580G, or I709A and D580G. In some embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, *e.g.*, amino acid residue substitutions I709K and I588T, or I709R and I588T, I709S and I588T, I709G and I588T, or I709A and I588T. In certain embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, *e.g.*, amino acid residue substitutions independently selected from I709K, I588T, and/or D580G. In certain embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, *e.g.*, amino acid residue substitutions independently selected from I709R, I588T, and/or D580G. In certain embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, *e.g.*, amino acid residue substitutions independently selected from I709S, I588T, and/or D580G. In certain embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, *e.g.*, amino acid residue substitutions independently selected from I709G, I588T, and/or D580G. In certain embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, *e.g.*, amino acid residue substitutions independently selected from I709A, I588T, and/or D580G.

In addition to mutation of the motifs of SEQ ID NOS:8, 9, 10, 11, 29 and 38 as described herein, the DNA polymerases of the present invention can also include other, non-substitutional modification(s). Such modifications can include, for example, covalent modifications known in the art to confer an additional advantage in applications comprising polynucleotide extension. For example, one such modification is a thermally reversible covalent modification that inactivates the enzyme, but which is reversed to activate the enzyme upon incubation at an elevated temperature, such as a temperature typically used for polynucleotide extension. Exemplary reagents for such thermally reversible modifications are described in U.S. Patent Nos. 5,773, 258 and 5,677,152.

The DNA polymerases of the present invention can be constructed by mutating the DNA sequences that encode the corresponding unmodified polymerase (*e.g.*, a wild-type polymerase or a corresponding variant from which the polymerase of the invention is derived), such as by using techniques commonly referred to as site-directed mutagenesis. Nucleic acid molecules encoding the unmodified form of the polymerase can be mutated by a variety of polymerase chain reaction (PCR) techniques well-known to one of ordinary skill in the art. (*See, e.g.*, *PCR Strategies* (M. A. Innis, D. H. Gelfand, and J. J. Sninsky eds., 1995, Academic Press, San

Diego, CA) at Chapter 14; *PCR Protocols : A Guide to Methods and Applications* (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White eds., Academic Press, NY, 1990).

By way of non-limiting example, the two primer system, utilized in the Transformer Site-Directed Mutagenesis kit from Clontech, may be employed for introducing site-directed mutants into a polynucleotide encoding an unmodified form of the polymerase. Following denaturation of the target plasmid in this system, two primers are simultaneously annealed to the plasmid; one of these primers contains the desired site-directed mutation, the other contains a mutation at another point in the plasmid resulting in elimination of a restriction site. Second strand synthesis is then carried out, tightly linking these two mutations, and the resulting plasmids are transformed into a mutS strain of *E. coli*. Plasmid DNA is isolated from the transformed bacteria, restricted with the relevant restriction enzyme (thereby linearizing the unmutated plasmids), and then retransformed into *E. coli*. This system allows for generation of mutations directly in an expression plasmid, without the necessity of subcloning or generation of single-stranded phagemids. The tight linkage of the two mutations and the subsequent linearization of unmutated plasmids result in high mutation efficiency and allow minimal screening. Following synthesis of the initial restriction site primer, this method requires the use of only one new primer type per mutation site. Rather than prepare each positional mutant separately, a set of "designed degenerate" oligonucleotide primers can be synthesized in order to introduce all of the desired mutations at a given site simultaneously. Transformants can be screened by sequencing the plasmid DNA through the mutagenized region to identify and sort mutant clones. Each mutant DNA can then be restricted and analyzed by electrophoresis, such as for example, on a Mutation Detection Enhancement gel (Mallinckrodt Baker, Inc., Phillipsburg, NJ) to confirm that no other alterations in the sequence have occurred (by band shift comparison to the unmutagenized control). Alternatively, the entire DNA region can be sequenced to confirm that no additional mutational events have occurred outside of the targeted region.

DNA polymerases with more than one amino acid substituted can be generated in various ways. In the case of amino acids located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If however, the amino acids are located some distance from each other (separated by more than ten amino acids, for example) it is more difficult to generate a single

oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed. In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions. An alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: DNA encoding the unmodified polymerase is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third round of mutagenesis, and so on. Alternatively, the multi-site mutagenesis method of Seyfang & Jin (*Anal. Biochem.* 324:285-291. 2004) may be utilized.

Accordingly, also provided are recombinant nucleic acids encoding any of the DNA polymerases of the present invention. Using a nucleic acid of the present invention, encoding a DNA polymerase, a variety of vectors can be made. Any vector containing replicon and control sequences that are derived from a species compatible with the host cell can be used in the practice of the invention. Generally, expression vectors include transcriptional and translational regulatory nucleic acid regions operably linked to the nucleic acid encoding the DNA polymerase. The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. In addition, the vector may contain a Positive Retroregulatory Element (PRE) to enhance the half-life of the transcribed mRNA (*see* Gelfand *et al.* U.S. Patent No. 4,666,848). The transcriptional and translational regulatory nucleic acid regions will generally be appropriate to the host cell used to express the polymerase. Numerous types of appropriate expression vectors, and suitable regulatory sequences are known in the art for a variety of host cells. In general, the transcriptional and translational regulatory sequences

may include, *e.g.*, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In typical embodiments, the regulatory sequences include a promoter and transcriptional start and stop sequences. Vectors also typically include a polylinker region containing several restriction sites for insertion of foreign DNA. In certain embodiments, "fusion flags" are used to facilitate purification and, if desired, subsequent removal of tag/flag sequence, e.g., "His-Tag". However, these are generally unnecessary when purifying a thermoactive and/or thermostable protein from a mesophilic host (e.g., *E. coli*) where a "heat-step" may be employed. The construction of suitable vectors containing DNA encoding replication sequences, regulatory sequences, phenotypic selection genes, and the polymerase of interest are prepared using standard recombinant DNA procedures. Isolated plasmids, viral vectors, and DNA fragments are cleaved, tailored, and ligated together in a specific order to generate the desired vectors, as is well-known in the art (*see, e.g.*, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, NY, 2nd ed. 1989)).

In certain embodiments, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selection genes are well known in the art and will vary with the host cell used. Suitable selection genes can include, for example, genes coding for ampicillin and/or tetracycline resistance, which enables cells transformed with these vectors to grow in the presence of these antibiotics.

In one aspect of the present invention, a nucleic acid encoding a DNA polymerase is introduced into a cell, either alone or in combination with a vector. By "introduced into" or grammatical equivalents herein is meant that the nucleic acids enter the cells in a manner suitable for subsequent integration, amplification, and/or expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type. Exemplary methods include $CaPO_4$ precipitation, liposome fusion, LIPOFECTIN®, electroporation, viral infection, and the like.

In some embodiments, prokaryotes are typically used as host cells for the initial cloning steps of the present invention. They are particularly useful for rapid production of large amounts of DNA, for production of single-stranded DNA templates used for site-directed mutagenesis, for screening many mutants simultaneously, and for DNA sequencing of the mutants generated. Suitable prokaryotic host cells include *E. coli* K12 strain 94 (ATCC No. 31,446), *E. coli* strain W3110 (ATCC No. 27,325), *E. coli* K12 strain DG116 (ATCC No. 53,606), *E. coli* X1776

(ATCC No. 31,537), and *E. coli* B; however many other strains of *E. coli*, such as HB101, JM101, NM522, NM538, NM539, and many other species and genera of prokaryotes including bacilli such as *Bacillus subtilis*, other enterobacteriaceae such as *Salmonella typhimurium* or *Serratia marcesans*, and various *Pseudomonas* species can all be used as hosts. Prokaryotic host cells or other host cells with rigid cell walls are typically transformed using the calcium chloride method as described in section 1.82 of Sambrook *et al.*, *supra*. Alternatively, electroporation can be used for transformation of these cells. Prokaryote transformation techniques are set forth in, for example Dower, in *Genetic Engineering, Principles and Methods* 12:275-296 (Plenum Publishing Corp., 1990); Hanahan *et al.*, *Meth. Enzymol.*, 204:63, 1991. Plasmids typically used for transformation of *E. coli* include pBR322, pUCI8, pUCI9, pUCIl8, pUC119, and Bluescript M13, all of which are described in sections 1.12-1.20 of Sambrook *et al.*, *supra*. However, many other suitable vectors are available as well.

The DNA polymerases of the present invention are typically produced by culturing a host cell transformed with an expression vector containing a nucleic acid encoding the DNA polymerase, under the appropriate conditions to induce or cause expression of the DNA polymerase. Methods of culturing transformed host cells under conditions suitable for protein expression are well-known in the art (*see, e.g.*, Sambrook *et al.*, *supra*). Suitable host cells for production of the polymerases from lambda pL promotor-containing plasmid vectors include *E. coli* strain DG116 (ATCC No. 53606) (*see* US Pat. No. 5,079,352 and Lawyer, F.C. *et al.*, *PCR Methods and Applications* 2:275-87, 1993). Following expression, the polymerase can be harvested and isolated. Methods for purifying the thermostable DNA polymerase are described in, for example, Lawyer *et al.*, *supra*. Once purified, the ability of the DNA polymerases to have improved RT efficiency, increased mis-match tolerance, extension rate and/or tolerance of RT and polymerase inhibitors can be tested (e.g., as described in the examples).

The improved DNA polymerases of the present invention may be used for any purpose in which such enzyme activity is necessary or desired. Accordingly, in another aspect of the invention, methods of polynucleotide extension (e.g., PCR) using the polymerases are provided. Conditions suitable for polynucleotide extension are known in the art. (*See, e.g.*, Sambrook *et al.*, *supra*. *See also* Ausubel *et al.*, *Short Protocols in Molecular Biology* (4th ed., John Wiley & Sons 1999). Generally, a primer is annealed, *i.e.*, hybridized, to a target nucleic acid to form a primer-template complex. The primer-template complex is contacted with the DNA polymerase

and nucleoside triphosphates in a suitable environment to permit the addition of one or more nucleotides to the 3' end of the primer, thereby producing an extended primer complementary to the target nucleic acid. The primer can include, *e.g.*, one or more nucleotide analog(s). In addition, the nucleoside triphosphates can be conventional nucleotides, unconventional nucleotides (*e.g.*, ribonucleotides or labeled nucleotides), or a mixture thereof. In some variations, the polynucleotide extension reaction comprises amplification of a target nucleic acid. Conditions suitable for nucleic acid amplification using a DNA polymerase and a primer pair are also known in the art (*e.g.*, PCR amplification methods). (*See, e.g.*, Sambrook *et al.*, *supra*; Ausubel *et al.*, *supra*; *PCR Applications: Protocols for Functional Genomics* (Innis *et al.* eds., Academic Press 1999). In other, non-mutually exclusive embodiments, the polynucleotide extension reaction comprises reverse transcription of an RNA template (*e.g.*, RT-PCR). In some embodiments, the improved polymerases find use in 454 sequencing (Margulies, M et al. 2005, Nature, 437, 376-380).

Optionally, the primer extension reaction comprises an actual or potential inhibitor of a reference or unmodified polymerase. The inhibitor can inhibit, for example, the nucleic acid extension rate and/or the reverse transcription efficiency of a reference or unmodified (control) polymerase. In some embodiments, the inhibitor is hemoglobin, or a degradation product thereof. For example, in some embodiments, the hemoglobin degradation product is a heme breakdown product, such as hemin, hematoporphyrin, or bilirubin. In some embodiments, the inhibitor is an iron-chelator or a purple pigment. In other embodiments, the inhibitor is heparin. In certain embodiments, the inhibitor is an intercalating dye. In certain embodiments, the inhibitor is melanin, which has been described as a polymerase inhibitor. *See*, *e.g*, Ekhardt, *et al.*, *Biochem Biophys Res Commun*. 271(3):726-30 (2000).

The DNA polymerases of the present invention can be used to extend templates in the presence of polynucleotide templates isolated from samples comprising polymerase inhibitors, e.g., such as blood. For example, the DNA polymerases of the present invention can be used to extend templates in the presence of hemoglobin, a major component of blood, or in the presence of a hemoglobin degradation product. Hemoglobin can be degraded to various heme breakdown products, such as hemin, hematin, hematoporphyrin, and bilirubin. Thus, in certain embodiments, the DNA polymerases of the present invention can be used to extend templates in the presence of hemoglobin degradation products, including but not limited to, hemin, hematin,

hematoporphyrin, and bilirubin. In certain embodiments, the hemoglobin degradation product is hemin. In some embodiments, the DNA polymerases of the present invention can be used to extend templates in the presence of about 0.5 to 20.0 μM, about 0.5 to 10.0 μM, about 0.5 to 5.0 μM, about 1.0 to 10.0 μM, about 1.0 to 5.0 μM, about 2.0 to 5.0 μM, or about 2.0 to 3.0 μM hemin. In other embodiments, the DNA polymerases of the present invention can be used to extend templates in the presence of at least about 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 10.0, 20.0, or greater than 20 μM hemin. The breakdown products of hemoglobin include iron-chelators and purple pigments. Thus, in some embodiments, the DNA polymerases of the present invention can be used to extend templates in the presence of iron-chelators and/or purple pigments. In other embodiments, the DNA polymerases of the present invention can be used to extend templates in the presence of amounts of hemoglobin degradation products that would inhibit extension of the same template by a reference or control DNA polymerase.

The DNA polymerases of the present invention can be used to extend templates in the presence of heparin. Heparin is commonly present as an anticoagulant in samples isolated from blood. In some embodiments, the DNA polymerases of the present invention can be used to extend templates in the presence of about 1.0 to 400 ng/μl, 1.0 to 300 ng/μl, 1.0 to 200 ng/μl, 5.0 to 400 ng/μl, 5.0 to 300 ng/μl, 5.0 to 200 ng/μl, 10.0 to 400 ng/μl, 10.0 to 300 ng/μl, or 10.0 to 200 ng/μl heparin. In some embodiments, the DNA polymerases of the present invention can be used to extend templates in the presence of at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400 ng/μl, or greater than 400 ng/μl of heparin. In other embodiments, the DNA polymerases of the present invention can be used to extend templates in the presence of amounts of heparin that would inhibit extension of the same template by a reference or control DNA polymerase.

In some embodiments, an improved polymerase of the invention is used in a reverse transcription reaction. In some embodiments, the reverse transcription reaction is carried out in a mixture containing the RNA template, one or more primer(s), and a thermostable DNA polymerase of the invention. The reaction mixture typically contains all four standard deoxyribonucleoside triphosphates (dNTPs) and a buffer containing a divalent cation and a monovalent cation. Exemplary cations include, e.g., $Mg^{2+}$, although other cations, such as $Mn^{2+}$ or $Co^{2+}$ can activate DNA polymerases. In other embodiments, the reverse transcription reaction is carried out with a thermo-active DNA polymerase of the invention. In particular

embodiments, the improved polymerase of the invention allows for more efficient amplification of RNA templates without compromising the efficient amplification of a DNA template in the presence of $Mn^{2+}$ or $Mg^{2+}$, as described in the examples.

The most efficient RT activity in thermostable DNA polymerases has been achieved using $Mn^{2+}$ as the divalent metal ion activator. However, it is well known that when $Mn^{2+}$ is present in reactions the fidelity of DNA polymerases is lower. Unless one is trying to generate mutations, it is generally favored to maintain a higher fidelity. Fortunately, most conventional sequencing, PCR and RT-PCR applications do not require high fidelity conditions because the detection systems generally are looking at a population of products. With the advent of next generation sequencing, digital PCR, etc., the fidelity of the product is more important and methods that allow for higher fidelity DNA synthesis are critical. Achieving efficient RT activity using $Mg^{2+}$ as the divalent metal ion activator is an excellent way to substantially increase the fidelity of the DNA polymerase and allow for more reliable copying of the nucleic acid target.

Because the polymerases described herein can also have increased mismatch tolerance, the polymerases find use in methods where variation of the target template is likely and yet the template is nevertheless desired to be amplified regardless of the variation at the target template. An example of such templates can include, for example, viral, bacterial, or other pathogen sequences. In many embodiments, it is desirable to determine simply whether an individual (human or non-human animal) has a viral or other infection, regardless of the precise viral variant that has infected the individual. As an example, one can use a primer pair to amplify HCV using a polymerase of the invention and detect the presence of the HCV even if the particular virus infecting the individual has a mutation resulting in a mismatch at the primer hybridization site.

Target nucleic acids can come from a biological or synthetic source. The target can be, for example, DNA or RNA. Generally, where amplicons are generated, the amplicons will be composed of DNA, though ribonucleotides or synthetic nucleotides can also be incorporated into the amplicon. Where one wishes to detect an RNA, the amplification process will typically involve the use of reverse transcription, including for example, reverse transcription PCR (RT-PCR).

Specific target sequences can include, e.g., viral nucleic acids (e.g., human immunodeficiency virus (HIV), hepatitis virus B (HBV), (cytomegalovirus (CMV), parvo B19 virus, Epstein-Barr

virus, hepatitis virus C (HCV), human papilloma virus (HPV), Japanese encephalitis virus (JEV), West Nile virus (WNV), St. Louis encephalitis virus (SLEV), Murray Valley encephalitis virus, and Kunjin virus), bacterial nucleic acids (e.g., *S. aureus*, *Neisseria meningitidis*, *Plasmodium falciparum*, *Chlamydia muridarum*, *Chlamydia trachomatis*), mycobacteria, fungal nucleic acids, or nucleic acids from animals or plants. In some embodiments, the target nucleic acids are animal (e.g., human) nucleic acids or are derived from an animal (e.g., human) sample (i.e., viral or other pathogenic organism nucleic acids may be present in a sample from an animal biopsy, blood sample, urine sample, fecal sample, saliva, etc.). In some embodiments, the target nucleic acids are, for example, human genetic regions that may include variants associated with disease (e.g., cancer, diabetes, etc.). Because in some embodiments the polymerases of the invention have mismatch tolerance, such enzymes are particularly useful, for example, where a diversity of related sequences could be in a target sequence. As an example, the invention can be used to detect viral pathogens, where the viral pathogens have sufficient variation in their genomes to make it difficult or impossible to design a single or small set of primers that will amplify most or all possible viral genomes or in cancer or other disease genetic markers where variation in sequence is known or likely to occur.

Other methods for detecting extension products or amplification products using the improved polymerases described herein include the use of fluorescent double-stranded nucleotide binding dyes or fluorescent double-stranded nucleotide intercalating dyes. Examples of fluorescent double-stranded DNA binding dyes include SYBR-green (Molecular Probes). The double stranded DNA binding dyes can be used in conjunction with melting curve analysis to measure primer extension products and/or amplification products. The melting curve analysis can be performed on a real-time PCR instrument, such as the ABI 5700/7000 (96 well format) or ABI 7900 (384 well format) instrument with onboard software (SDS 2.1). Alternatively, the melting curve analysis can be performed as an end point analysis. Exemplary methods of melting point analysis are described in U.S. Patent Publication No. 2006/0172324.

In another aspect of the present invention, kits are provided for use in primer extension methods described herein. In some embodiments, the kit is compartmentalized for ease of use and contains at least one container providing an improved DNA polymerase in accordance with the present invention. One or more additional containers providing additional reagent(s) can also be included. In some embodiments, the kit can also include a blood collection tube, container, or

unit that comprises heparin or a salt thereof, or releases heparin into solution. The blood collection unit can be a heparinized tube. Such additional containers can include any reagents or other elements recognized by the skilled artisan for use in primer extension procedures in accordance with the methods described above, including reagents for use in, *e.g.*, nucleic acid amplification procedures (*e.g.*, PCR, RT-PCR), DNA sequencing procedures, or DNA labeling procedures. For example, in certain embodiments, the kit further includes a container providing a 5' sense primer hybridizable, under primer extension conditions, to a predetermined polynucleotide template, or a primer pair comprising the 5' sense primer and a corresponding 3' antisense primer. In other, non-mutually exclusive variations, the kit includes one or more containers providing nucleoside triphosphates (conventional and/or unconventional). In specific embodiments, the kit includes alpha-phosphorothioate dNTPs, dUTP, dITP, and/or labeled dNTPs such as, *e.g.*, fluorescein- or cyanin-dye family dNTPs. In still other, non-mutually exclusive embodiments, the kit includes one or more containers providing a buffer suitable for a primer extension reaction.

In another aspect of the present invention, reaction mixtures are provided comprising the polymerases with increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors as described herein. The reaction mixtures can further comprise reagents for use in, *e.g.*, nucleic acid amplification procedures (*e.g.*, PCR, RT-PCR), DNA sequencing procedures, or DNA labeling procedures. For example, in certain embodiments, the reaction mixtures comprise a buffer suitable for a primer extension reaction. The reaction mixtures can also contain a template nucleic acid (DNA and/or RNA), one or more primer or probe polynucleotides, nucleoside triphosphates (including, *e.g.*, deoxyribonucleotides, ribonucleotides, labeled nucleotides, unconventional nucleotides), salts (*e.g.*, $Mn^{2+}$, $Mg^{2+}$), labels (*e.g.*, fluorophores). In some embodiments, the reaction mixtures contain a 5'-sense primer hybridizable, under primer extension conditions, to a predetermined polynucleotide template, or a primer pair comprising the 5'-sense primer and a corresponding 3' antisense primer. In some embodiments, the reaction mixtures contain alpha-phosphorothioate dNTPs, dUTP, dITP, and/or labeled dNTPs such as, *e.g.*, fluorescein- or cyanin-dye family dNTPs. In some embodiments, the reaction mixtures comprise an iron chelator or a purple dye. In certain embodiments, the reaction mixtures comprise hemoglobin, or a degradation product of hemoglobin. For example, in certain embodiments, the degradation products of hemoglobin

include heme breakdown products such as hemin, hematin, hematophoryn, and bilirubin. In other embodiments, the reaction mixtures comprise heparin or a salt thereof. In certain embodiments, the reaction mixture contains a template nucleic acid that is isolated from blood. In other embodiments, the template nucleic acid is RNA and the reaction mixture comprises heparin or a salt thereof.

## EXAMPLES

The following examples are offered to illustrate, but not to limit the claimed invention.

***Example 1:*** **Library Generation**

In brief, the steps in this screening process included library generation, expression and partial purification of the mutant enzymes, screening of the enzymes for the desired properties, DNA sequencing, clonal purification, and further characterization of selected candidate mutants. Each of these steps is described further below.

***Clonal Library generation:*** A nucleic acid encoding the polymerase domain of Z05 D580G DNA polymerase was subjected to error-prone (mutagenic) PCR between Blp I and Bgl II restriction sites of a plasmid including this nucleic acid sequence. The amplified sequence is provided as SEQ ID NO:39. The primers used for this are given below:

Forward Primer: 5'- CTACCTCCTGGACCCCTCCAA-3' (SEQ ID NO:30); and,

Reverse Primer: 5'- ATAACCAACTGGTAGTGGCGTGTAA-3' (SEQ ID NO:31)

PCR was performed using a range of $Mg^{2+}$ concentrations from 1.8-3.6 mM, in order to generate libraries with a range of mutation rates. Buffer conditions were 50 mM Bicine pH 8.2, 115 mM KOAc, 8% w/v glycerol, and 0.2 mM each dNTPs. A GeneAmp® AccuRT Hot Start PCR enzyme was used at 0.15 U/μL. Starting with $5x10^5$ copies of linearized Z05 D580G plasmid DNA per reaction volume of 50 μL, reactions were denatured using a temperature of 94 °C for 60 seconds, then 30 cycles of amplification were performed, using a denaturation temperature of 94 °C for 15 seconds, an annealing temperature of 60 °C for 15 seconds, an extension temperature of 72 °C for 120 seconds, and followed by a final extension at a temperature of 72 °C for 5 minutes.

The resulting amplicon was purified with a QIAquick PCR Purification Kit (Qiagen, Inc., Valencia, CA, USA) and cut with Blp I and Bgl II, and then re-purified with a QIAquick PCR Purification Kit. A Z05 D580G vector plasmid was prepared by cutting with the same two restriction enzymes and treating with alkaline phosphatase, recombinant (RAS, cat# 03359123001) and purified with a QIAquick PCR Purification Kit. The cut vector and the mutated insert were mixed at a 1:3 ratio and treated with T4 DNA ligase for 5 minutes at room temperature (NEB Quick Ligation™ Kit). The ligations were purified with a QIAquick PCR Purification Kit and transformed into an *E.coli* host strain by electroporation.

Aliquots of the expressed cultures were plated on ampicillin-selective medium in order to determine the number of unique transformants in each transformation. Transformations were stored at -70 °C to -80 °C in the presence of glycerol as a cryo-protectant.

Each library was then spread on large format ampicillin-selective agar plates. Individual colonies were transferred to 384-well plates containing 2X Luria broth with ampicillin and 10% w/v glycerol using an automated colony picker (QPix2, Genetix Ltd). These plates were incubated overnight at 30 °C to allow the cultures to grow and then stored at -70 °C to -80 °C. The glycerol added to the 2X Luria broth was low enough to permit culture growth and yet high enough to provide cryo-protection. Several thousand colonies at several mutagenesis ($Mg^{2+}$) levels were prepared in this way for later use.

***Extract library preparation Part 1—Fermentation:*** From the clonal libraries described above, a corresponding library of partially purified extracts suitable for screening purposes was prepared. The first step of this process was to make small-scale expression cultures of each clone. These cultures were grown in 96-well format; therefore there were 4 expression culture plates for each 384-well library plate. 0.5 µL was transferred from each well of the clonal library plate to a well of a 96 well seed plate, containing 150 µL of Medium A (see Table 3 below). This seed plate was shaken overnight at 1150 rpm at 30 °C, in an iEMS plate incubator/shaker (ThermoElectron). These seed cultures were then used to inoculate the same medium, this time inoculating 20 µL into 250 µL Medium A in large format 96 well plates (Nunc # 267334). These plates were incubated overnight at 37 °C with shaking. The expression plasmid contained transcriptional control elements, which allow for expression at 37 °C but not at 30 °C. After overnight incubation, the cultures expressed the clone protein at typically 1-10%

of total cell protein. The cells from these cultures were harvested by centrifugation. These cells were either frozen (-20 ºC) or processed immediately, as described below.

**Table 2.** Medium A (Filter-sterilized prior to use)

| Component | Concentration |
|---|---|
| $MgSO_4.7H_2O$ | 0.2 g/L |
| Citric acid.$H_2O$ | 2 g/L |
| $K_2HPO_4$ | 10 g/L |
| $NaNH_4PO_4.4H_2O$ | 3.5 g/L |
| $MgSO_4$ | 2 mM |
| Casamino acids | 2.5 g/L |
| Glucose | 2 g/L |
| Thiamine.HCl | 10 mg/L |
| Ampicillin | 100 mg/L |

***Extract library preparation Part 2—Extraction:*** Cell pellets from the fermentation step were resuspended in 25 µL Lysis buffer (Table 3 below) and transferred to 384-well thermocycler plates and sealed. Note that the buffer contained lysozyme to assist in cell lysis, and DNase to remove DNA from the extract. To lyse the cells the plates were incubated at 37 ºC for 15 minutes, frozen overnight at -20 °C, and incubated again at 37 ºC for 15 minutes. Ammonium sulfate was added (1.5 µL of a 2M solution) and the plates incubated at 75 ºC for 15 minutes in order to precipitate and inactivate contaminating proteins, including the exogenously added nucleases. The plates were centrifuged at 3000 x g for 15 minutes at 4 ºC and the supernatants transferred to a fresh 384-well thermocycler plate. These extract plates were frozen at -20 ºC for later use in screens. Each well contained about 0.5-3 µM of the mutant library polymerase enzyme.

**Table 3.** Lysis Buffer

| Component | Concentration or Percentage |
|---|---|
| Tris pH 7.5 | 50 mM |
| EDTA | 1 mM |
| $MgCl_2$ | 6 mM |
| Tween 20 | 0.5% v/v |
| Lysozyme (from powder) | 1 mg/mL |
| DNase I | 0.05 Units/µL |

***Example 2:*** **Identification of mutant DNA polymerases with improved reverse transcription efficiency**

**Screening extract libraries for improved reverse transcription efficiency:** The extract library was screened by comparing Cp (Crossing Point) values from growth curves generated by fluorescent 5' nuclease (TaqMan) activity of crude enzyme extracts in a RT-PCR system from amplification of a 240 base pair amplicon from Hepatitis C Virus (HCV) transcript JP2-5, containing the first 800 bases of HCV genotype Ib 5'NTR in pSP64 poly(A) (Promega).

Reactions were carried out on the Roche LC 480 kinetic thermocycler in 384 well format with each well containing 1.5 µL of an individual enzyme extract diluted 5-fold with buffer containing 20 mM Tris-HCl, pH 8, 100 mM KCl, 0.1 mM EDTA, and 0.1% Tween-20 added to 18.5 µL of RT-PCR master mix described in Table 4. The thermocycling conditions were: 1 minute at 65 °C ("RT" step); 5 cycles of 94 °C for 15 seconds followed by 60 °C for 30 seconds; and 45 cycles of 91 °C for 15 seconds followed by 60 °C for 30 seconds.

**Table 4. RT-PCR Master Mix**

| **Component** | **Concentration** |
|---|---|
| Tricine pH 8.3 | 50 mM |
| KOAc | 100 mM |
| Glycerol | 5% (v/v) |
| DMSO | 2 % (v/v) |

| | |
|---|---|
| Primer 1 | 200 nM |
| Primer 2 | 200 nM |
| TaqMan Probe | 75 nM |
| Aptamer | 200 nM |
| dATP | 200 µM |
| dCTP | 200 µM |
| dGTP | 200 µM |
| dUTP | 400 µM |
| UNG | .04 Units/µL |
| RNA Target | 5000 copies/µL |
| $Mn(OAc)_2$ | 2.1 mM |

Approximately 5000 clones were screened using the above protocol. Twenty one clones were chosen from the original pool for rescreening based on earliest Crossing Point (Cp) values and fluorescent plateau values above an arbitrary cut off as calculated by the Abs Quant/2nd derivative max method. Culture wells corresponding to the top extracts were sampled to fresh growth medium and re-grown to produce new culture plates containing the best mutants, as well as a number of parental Z05 D580X (X= G, K, and R) cultures to be used for comparisons. These culture plates were then used to make fresh crude extracts which were quantified and rescreened at 20 nM concentrations with the same master mix conditions as described in Table 1. Table 5 shows the Cp values obtained from the FAM signal increase due to cleavage of the TaqMan probe. Results show that the polymerase expressed by clone 0813-L15 amplifies the RNA target with higher efficiency than the Z05 D580G parental.

**Table 5. Cp values obtained with mutant polymerases amplifying an RNA template.**

| Clone | Average Cp |
|---|---|
| 0813-L15 | 18.5 |
| Z05 D580R | 24.0 |
| Z05 D580K | 24.5 |
| Z05 D580G | 27.5 |

The DNA sequence of the mutated region of the polymerase gene was sequenced to determine the mutation(s) that were present in any single clone. Clone 0815-L15 was chosen for further testing, so mutant polymerase protein was expressed in flask culture, purified to homogeneity, and quantified.

**Use of Z05 D580G mutant in $Mn^{2+}$-based RT-PCR:** Sequencing results revealed that the polymerase expressed by clone 0813-L15 carries mutations I709K and A803S in addition to the parental D580G mutation. Purified mutant Z05 D580G_I709K_A803S (0813-L15) was compared to parental Z05 D580G in TaqMan $Mn^{2+}$ -based RT-PCR. Reverse transcription and PCR efficiencies were measured by comparing Cp values from amplifications of JP2-5 RNA transcript and pJP2-5 DNA linear plasmid digested with the restriction endonuclease *Eco*RI. Oligonucleotides and Master Mix conditions (Table 1) were the same as used in the original screen. Each reaction had either 100,000 copies of JP2-5 transcript, 100,000 copies of pJP2-5 linear plasmid DNA, or 1000 copies of pJP2-5 linear plasmid DNA. All targets were amplified with Primer 1 and Primer 2, as described above, in duplicate reactions to generate a 240 base pair amplicon .All reactions were performed on the Roche Light Cycler 480 thermal cycler with a reaction volume of 15 µL. Crossing Point (Cps) were calculated by the Abs Quant/$2^{nd}$ derivative max method and averaged. Amplifications were carried out using a range of DNA Polymerase concentrations from 2.5 nM- 30 nM. Thermocycling conditions were: 1 minute at 65 ºC ("RT" step); 5 cycles of 94 ºC for 15 seconds followed by 60 ºC for 30 seconds; and 45 cycles of 91 ºC for 15 seconds followed by 60 ºC for 30 seconds. Table 6 shows Cp values obtained from fluorescent signal increase due to cleavage of the TaqMan probe at 20 nM enzyme condition.

**Table 6. Cp values obtained with mutant polymerases when amplifying RNA and DNA templates in the presence of $Mn^{2+}$.**

| Enzyme | RNA $10^5$ copies Cp | DNA $10^5$ copies Cp | DNA $10^3$ copies Cp |
|---|---|---|---|
| Z05 D580G | 31.6 | 19.7 | 27.5 |
| Z05 D580G_I709K_A803S | 20.3 | 18.9 | 26.6 |

The results indicate that mutant Z05 D580G_I709K_A803S allows for more efficient amplification of an RNA target without compromise of PCR efficiency on a DNA target, as compared to the parental enzyme.

**Use of Z05 D580G mutant in $Mg^{2+}$ -based RT-PCR:** The purified mutant Z05 D580G_I709K_A803S was also compared to parental Z05 D580G for the ability to perform TaqMan RT-PCR in the presence of $Mg^{2+}$ . The master mix conditions used were identical to those described in Table 1, except that the KOAc concentration was varied from 20 mM- 160 mM and $Mn(OAc)_2$ was replaced with 2.1 mM $Mg(OAc)_2$. Each reaction had 30 nM enzyme and either 100,000 copies of JP2-5 transcript, 100,000 copies of pJP2-5 linear plasmid DNA, or 1000 copies of pJP2-5 linear plasmid DNA. All targets were amplified with the same primer set in duplicate reactions to generate a 240 base pair amplicon. PCR and RT-PCR efficiencies were determined by comparing Cp values between DNA and RNA .All reactions were performed on the Roche Light Cycler 480 thermal cycler with a reaction volume of 15 µL. Crossing Point (Cps) were calculated by the Abs Quant/$2^{nd}$ derivative max method and Cps were averaged. Thermocycling conditions were: 65 °C- 5 minutes, 70 °C- 5 minutes, and 75 °C- 5 minutes (three temperature"RT"step); 5 cycles of 94 °C for 15 seconds followed by 62 °C for 30 seconds; and 45 cycles of 91 °C for 15 seconds followed by 62 °C for 30 seconds. Table 7 shows Cp values obtained from fluorescent signal increase due to cleavage of the TaqMan probe at the 40 nM KOAc condition.

**Table7. Cp values obtained with mutant polymerases when amplifying RNA and DNA templates in the presence of $Mg^{2+}$.**

| Enzyme | RNA $10^5$ copies Cp | DNA $10^5$ copies Cp | DNA $10^3$ copies Cp |
|---|---|---|---|
| Z05 D580G | 28.4 | 18.5 | 24.7 |
| Z05 D580G I709K_A803S | 20.6 | 17.8 | 23.8 |

The results indicate that mutant Z05 D580G_I709K_A803S performs $Mg^{2+}$-based RT PCR with significantly greater efficiency than Z05 D580G under these conditions.

**Determination of phenotype-conferring mutation(s):** The polymerase expressed by the 0815-L15 clone displays the most improvement in RNA amplification over parental Z05 D580G in the RT-PCR screen. The 0815-L15 clone expresses a double mutant polymerase carrying mutations I709K and A803 in addition to the parental D580G mutation. Based on the nature of the amino acid change and the proximity of the A803S to the C- terminus of the protein, we predicted that the I709K mutation is responsible for the observed phenotype. A Z05 D580G_I709K mutant was constructed by PCR -based site-directed mutagenesis, purified, quantified, and compared to 0815-L15 (Z05 D580G_I709K_A803S) in $Mg^{2+}$ activated TaqMan RT-PCR with varying KOAc concentration from 20 mM- 160 mM and 30 nM enzyme. Master Mix conditions were the same as those described previously in Table 1 except $Mn(OAc)_2$ was replaced with 2.1 mM $Mg(OAc)_2$. Each reaction had either 100,000 copies of JP2-5 transcript, 100,000 copies of pJP2-5 linear plasmid DNA, or 1000 copies of pJP2-5 linear plasmid DNA. All targets were amplified with the same primer set in duplicate reactions to generate a 240 base pair amplicon. The PCR and RT-PCR efficiencies were determined by comparing Cp values between DNA and RNA .All reactions were performed on the Roche Light Cycler 480 thermal cycler with a reaction volume of 15 µL. Crossing Point (Cps) were calculated by the Abs Quant/2nd derivative max method and Cps were averaged. The thermocycling conditions were: 2 minutes at 50 ºC ("UNG" step); 65 ºC- 5 minutes, 68 ºC- 5 minutes, and 72 ºC- 5 minutes (three temperature "RT"step); 5 cycles of 94 ºC for 15 seconds followed by 62 ºC for 30 seconds; and 45 cycles of 91 ºC for 15 seconds followed by 62 ºC for 30 seconds. Table 8 shows the Cp values obtained from fluorescent signal increase due to cleavage of the TaqMan probe at the 60 mM KOAc condition.

**Table 8. Cp values obtained using mutant polymerases to amplify RNA and DNA templates.**

| Enzyme | RNA $10^5$ copies Cp | DNA $10^5$ copies Cp | DNA $10^3$ copies Cp |
|---|---|---|---|
| Z05 D580G | 29.2 | 17.0 | 23.0 |
| Z05 D580G_I709K_A803S | 19.3 | 16.6 | 22.6 |
| Z05 D580G_I709K | 19.0 | 16.7 | 22.5 |

Z05 D580G_I709K and Z05 D580G_I709K_A803S have similar Cp values on both RNA and DNA targets, demonstrating that the I709K mutation confers the observed improvement in RT-PCR performance.

**Various Amino Acid Substitutions at the I709 position:** The effect of various substitutions at the I709 position on $Mg^{2+}$ - based TaqMan RT-PCR efficiency of Z05 D580G DNA polymerase was examined. First, the mutations were created in Z05 D580G DNA polymerase, utilizing a PCR -based site-directed mutagenesis technique, and the mutant enzymes purified and quantified. Z05D580G_I709 mutants K (Lysine), A (Alanine), G (Glycine), S (Serine), R (Arginine), L (Leucine), and D (Apartic Acid) were compared to parental Z05 D580G in $Mg^{2+}$ activated TaqMan RT-PCR with varying KOAc concentration from 20 mM- 160 mM and 10 nM enzyme. Master Mix conditions were the same as those described previously in Table 1 except 2.0 mM $Mg(OAc)_2$ was used. Each reaction had either 100,000 copies of JP2-5 transcript, 100,000 copies of pJP2-5 linear plasmid DNA, or 1000 copies of pJP2-5 linear plasmid DNA. All targets were amplified with the same primer set in duplicate reactions to generate a 240 base pair amplicon. All reactions were performed on the Roche Light Cycler 480 thermal cycler with a reaction volume of 15 µL. Crossing Point (Cps) were calculated by the Abs Quant/$2^{nd}$ derivative max method and Cps were averaged. The thermocycling conditions were: 3 minutes at 50 ºC ("UNG" step); 65 ºC- 5 minutes, 68 ºC- 5 minutes, and 72 ºC- 5 minutes (three temperature "RT"step); 5 cycles of 95 ºC for 15 seconds followed by 62 ºC for 30 seconds; and 45 cycles of 91 ºC for 15 seconds followed by 62 ºC for 30 seconds. Table 9 shows the Cp values obtained from fluorescent signal increase due to cleavage of the TaqMan probe at the 80 mM KOAc condition.

**Table 9. Cp values obtained using polymerases having various substitutions at the I709 position to amplify RNA and DNA templates.**

| Enzyme | RNA $10^5$ copies Cp | DNA $10^5$ copies Cp | DNA $10^3$ copies Cp |
|---|---|---|---|
| Z05 D580G | 35.0 | 17.7 | 23.8 |
| Z05 D580G_I709K | 20.2 | 17.6 | 23.6 |
| Z05 D580G_I709R | 21.3 | 17.4 | 23.2 |
| Z05 D580G_I709S | 27.6 | 16.8 | 22.8 |

| | | | |
|---|---|---|---|
| Z05 D580G_I709G | 19.2 | 16.4 | 22.5 |
| Z05 D580G_I709L | 34.2 | 17.2 | 23.3 |
| Z05 D580G_I709D | NS | NS | NS |
| Z05 D580G_I709A | 28.4 | 17.1 | 23.1 |

NS= No TaqMan generated growth curves

This example shows that several amino acid substitutions at position 709 of Z05 D580G DNA polymerase result in more efficient amplification of RNA targets.

***Example 3:*** **Screening extract libraries for improved 3' primer mismatch tolerance.**

The extract library of Example 1 was screened for improved 3' primer mismatch tolerance by comparing the final fluorescence after extension by an enzyme of a primer (DG48; SEQ ID NO:40, Table 10) perfectly matched to the sequence of primer M13mp18 vs. the final fluorescence of a primer (FR744; SEQ ID NO:42; Table 10) with a 3' A:A mismatch.

DG48 Perfect Match:

5'-GGGAAGGGCGATCGGTGCGGGCCTCTTCGC-3' (SEQ ID NO:40)

FR744 A:A Mismatch:

5'-GGGAAGGGCGATCGGTGCGGGCCTCTTCGCA-3' (SEQ ID NO:42)

The enzyme extracts above were diluted 10-fold for primer extension reactions by combining 2.5 µL extract with 22.5 µL of a buffer containing 20 mM Tris-HCl, pH 8, 100 mM KCl, 0.1 mM EDTA, and 0.2% Tween-20 in a 384-well thermocycler plate, covering and heating for 10 minutes at 90 °C. Control reactions with perfect match primer combined 0.5 µL of the diluted extract with 15 µL master mix in 384-well PCR plates. Extension of the primed template was monitored every 15 seconds in a modified kinetic thermal cycler using a CCD camera (see, Watson, *supra*). Master mix contained 1 nM primed primer template, 25 mM Tricine, pH 8.3, 100 mM KOAc, 0.6X SYBR Green I, 200 µM each dNTP, 100 nM Aptamer, and 2.5 mM $Mg(OAc)_2$. In order to distinguish extension-derived fluorescence from background fluorescence, parallel wells were included in the experiment in which primer strand extension was prevented by leaving out the nucleotides from the reaction master mix. Reactions with the 3'-mismatched primer (FR744, SEQ ID NO:42) were performed as above except 1.0 µL the diluted extract was added to each reaction.

Approximately 5700 mutant extracts were screened using the above protocol. Clones were selected based on maximum fluorescence relative to a starting baseline after 1 minute of extension at 40 °C followed by 8.5 minutes of extension at 64 °C. Based on this criteria a relatively small number of extracts were chosen for purification and further testing. They were first streaked on selective agar plates to ensure clonal purity, and the DNA sequence of the mutated region of the polymerase gene was sequenced to determine the mutation(s) that were present in any single clone. In parallel with this work, mutant polymerase protein was expressed in flask culture, purified to homogeneity, and quantified.

***Example 4:*** **Primer extension of a variety of 3'-mismatches to an M13 template.**

This example demonstrates that substitutions at positions 588 and 709 results in a polymerase having improved efficiency extending a template using 3' mismatched primers.

Purified Z05 D580G I588T I709K was compared to the parental enzyme Z05 D580G in primer extension of a variety of 3'-primer mismatches to an M13mp18 template. Templates and primers are listed below in Table 10:

**Table 10. Primers used to extend an M13mp18 template.**

| Name | Description | Sequence (5'- 3') | SEQ ID NO: |
|---|---|---|---|
| M13mp18 | Template | | |
| DG48 | Perfect Match | GGGAAGGGCGATCGGTGCGGGCCTCTTCGC | 40 |
| FR743 | T:G Mismatch | GGGAAGGGCGATCGGTGCGGGCCTCTTCGT | 41 |
| FR744 | A:A Mismatch | GGGAAGGGCGATCGGTGCGGGCCTCTTCGCA | 42 |
| FR745 | A:C Mismatch | GGGAAGGGCGATCGGTGCGGGCCTCTTCA | 43 |
| FR750 | T:T Mismatch | GGGAAGGGCGATCGGTGCGGGCCTCTTCGCTT | 44 |
| FR751 | C:T Mismatch | GGGAAGGGCGATCGGTGCGGGCCTCTTCGCTC | 45 |
| FR752 | C:C Mismatch | GGGAAGGGCGATCGGTGCGGGCCTCTTCC | 46 |
| FR753 | T:C Mismatch | GGGAAGGGCGATCGGTGCGGGCCTCTTCT | 47 |

Primers were pre-annealed to M13mp18 template at a 10:1 primer:template ratio and added to extension reactions at 1 nM final concentration with 5 nM enzyme and 25 mM Tricine, pH 8.3,

100 mM KOAc, 0.6X SYBR Green I, 200 μM each dNTP, 100 nM Aptamer, and 2.5 mM $Mg(OAc)_2$. Reactions were performed in triplicate with extension of the primed template monitored every 15 seconds in a modified kinetic thermal cycler using a CCD camera (see, Watson, *supra*). The replicates were averaged and maximum slope for each condition was calculated as the change in fluorescence over time. Results are shown in Table 11 below.

**Table 11. Mutant polymerase extension rates of mismatched primers.**

| Enzyme | Matched Primer | Mismatched Primers | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | DG48 | FR743 | FR744 | FR745 | FR750 | FR751 | FR752 | FR753 |
| | C:G | T:G | A:A | A:C | T:T | C:T | C:C | T:C |
| Z05 D580G | 6.1 | 4.8 | 0.3 | 1.5 | 0.3 | 1.2 | 0.3 | 0.5 |
| Z05-D I588T I709K | 13.3 | 14.0 | 0.5 | 9.4 | 3.7 | 15.6 | 0.7 | 7.2 |

This example shows that Z05 D580G I588T I709K is about two-fold faster at extending a perfect matched primer template than the parental enzyme Z05 D580G and about two to greater than 10-fold faster at extending the 3'-mismatched primer templates depending on the terminal 3'-mismatch.

***Example 5:*** **Amplification of mutant BRAF plasmid template compared to wild-type BRAF human genomic template using mutant polymerases.**

This example demonstrates that the 588 and 709 mutations result in a polymerase having improved mismatch tolerance compared to the parental enzyme.

Purified Z05 D580G I588T I709K was compared to the parental enzyme Z05 D580G in TaqMan PCR for improved mismatch tolerance of a mutant BRAF V600E target in a background of wild-type Human Genomic DNA.

The forward primer is perfectly matched to the mutant sequence and has a single 3' A:A mismatch to the wild type sequence. Reactions have either 10,000 copies (33 ng) of wild-type human genomic cell line DNA or have 100 or 10,000 copies of a linearized plasmid containing the BRAF V600E mutant sequence in a final volume of 16 μL. To allow for the different salt optima of the enzymes, amplifications were performed using a range of KCl concentrations from 40 to 145 mM. Buffer conditions were 50 mM Tris-HCl pH 8.0, 2.5 mM $MgCl_2$, 0.2 mM

each dNTP, 0.02 U/μL UNG, and 200 nM Aptamer. Forward and Reverse primers were at 100 nM and the TaqMan probe was at 25 nM. All DNA polymerases were assayed at 20 nM and add 2 % (v/v) enzyme storage buffer (50% v/v glycerol, 100 mM KCl, 20 mM Tris pH 8.0, 0.1 mM EDTA, 1mM DTT, 0.5% Tween 20). The reactions were performed in a Roche LightCycler 480 thermal cycler and denatured using a temperature of 95 °C for 60 seconds, then 99 cycles of amplification were performed, using a denaturation temperature of 92 °C for 10 seconds and an annealing temperature of 62 °C for 30 seconds.

Reactions were run in duplicate, crossing points ("Cps") were calculated by the Abs Quant/2nd derivative Max method and the Cps were averaged. The averaged Cp values are shown in Table 12 as well as calculated PCR efficiency at the KCl concentration for each enzyme which resulted in the earliest high copy mutant Cp. High Copy delta Cp is equal to the difference between the average Cp values of the reactions with 10,000 copy of 3'-mismatched wild-type genomic target and the average Cp values of the reactions with 10,000 copy of perfect match plasmid target.

**Table 12. Cp values of mutant polymerases using a 3' mismatched primer.**

| Enzyme | Optimum KCL (mM) | 10,000 copy Wild-Type Cp | 100 copy Mutant Plasmid Cp | 10,000 copy Mutant Plasmid Cp | % PCR Efficiency | High copy ΔCp |
|---|---|---|---|---|---|---|
| Z05 D580G | 100 | 32.5 | 33.1 | 26.9 | 109 | 5.7 |
| Z05 D580G I588T I709K | 100 | 30.0 | 33.2 | 26.6 | 100 | 3.4 |

This example demonstrates that Z05 D580G I588T I709K results a 2.3 cycle improvement in the high copy ΔCp , demonstrating improved tolerance of a 3'-terminal A:A mismatch in this PCR system.

***Example 6:*** **Mutant polymerases have improved activity in the presence of inhibitors.**

This example demonstrates that the I709K mutation results in improved RT-PCR efficiency in the presence of known inhibitors of DNA polymerases.

**Hemin**

Hemoglobin, a critical component in blood, can be degraded to various heme breakdown products, such as hemin, hematin, hematoporphyrin, and eventually bilirubin. Since these molecules are both iron-chelators and purple pigments, they might utilize several mechanisms to inhibit polymerase and/or reverse transcriptase activity.

A model system using an HCV RNA transcript was used to determine the inhibitory effects of hemin in RT-PCR using Z05, Z05 D580G, or Z05 D580G I709K polymerases. 45 U DNA pol Z05, Z05 D580G, or Z05 D580G I709K were tested inRT-PCR conditions (120 mM KOAc, 3.3 mM $Mn^{2+}$, 60 mM Tricine; 50 uL total) amplifying 1,000 copies of an HCV RNA transcript with and without the addition of 2.5 uM hemin (40 fold molar excess to DNA pol). These reactions were run in a Roche LightCycler 480 Real Time PCR Instrument with a 12 minute RT step followed by 50 cycles of denaturation and extension. Real time fluorescence was detected in the JA270 and CY5.5 channels during the last 50 cycles. The Cp (crossing point) values from growth curves generated by fluorescent 5' nuclease (TaqMan) activity for each reaction was determined using the instrument's "2nd derivative Max analysis" method. The Cps of all normal reactions were compared to those with hemin, as shown in Table 13. In the presence of 2.5 uM hemin, no amplification of HCV RNA was observed by Z05, whereas the variant Z05 D580G detected HCV with a Cp delay of 3.4 cycles (*vs.* no hemin control)and Z05 D580G I709K detected HCV with a 2.5 cycle earlier Cp (*vs.* no hemin control).

**Table 13. Cp values of mutant polymerases in the presence and absence of hemin.**

| Enzyme | (-) HEMIN Cp | (+) HEMIN Cp |
|---|---|---|
| Z05 | 31.9 | No signal |
| Z05 D580G | 29.2 | 32.6 |
| Z05 D580G I709K | 28.2 | 25.7 |

Agarose gel electrophoresis confirmed that these effects were due to reduced amplification, not quenching by the porphyrin hemin molecule. Similar results were obtained with HCV DNA templates, suggesting that hemin acts as general PCR inhibitor.

**Heparin**

Heparin is a highly sulfated glycosaminoglycan and contains one of the highest negative charge densities of any known biological molecule. As such, it can mimic nucleic acid substrates and is often used as a non-specific competitor in protein-DNA/RNA binding assays. Whereas hemin acts a general polymerase and PCR inhibitor, heparin preferentially inhibits reverse transcription by, for example, Z05-based DNA polymerases.

Using the HCV RNA RT-PCR amplification model system described above, the presence of 100, 200, 400 or 1000 ng/uL of heparin was tested to determine inhibitory effects using Z05, Z05 D580G, or Z05 D580G I709K polymerases. The Cps of all normal reactions were compared to those with heparin (Table 14). Whereas the wild-type Z05 enzyme was unable to amplify HCV RNA in the presence of 12.5 ng/uL heparin, the Z05 D580G and Z05 D580G I709K mutants were able to tolerate up to 200 or 1000 ng/ul heparin with minimal Cp delays, suggesting that these variants are tolerant of at least 15 – 80 fold more heparin, respectively.

A direct comparison between RNA and DNA substrates revealed that amplification of DNA by Z05 D580G and Z05 D580G I709K is completely unaffected by the presence of high levels of heparin. Overall, these data support the notion that heparin is an inhibitor that more specifically inhibits reverse transcription. The resistance of a DNA pol to heparin is directly correlated with the intrinsic RT activity for each particular enzyme.

**Table 14. Cp values of mutant polymerases in the presence of increasing amounts of heparin.**

| Enzyme | (-) HEPARIN Cp | (+) 100 ng/ul HEPARIN Cp | (+) 200 ng/ul HEPARIN Cp | (+) 400 ng/ul HEPARIN Cp | (+) 1000 ng/ul HEPARIN Cp |
|---|---|---|---|---|---|
| Z05 | 33.2 | No signal | No signal | No signal | No signal |
| Z05 D580G | 29.4 | 32.2 | 38.7 | No signal | No signal |
| Z05 D580G I709K | 28.3 | 28.4 | 28.5 | 29.2 | 33.4 |

This example shows that the I709K mutation results in improved RT-PCR efficiency in the presence of the inhibitors Hemin and Heparin.

***Example 7*: Mutant polymerases have improved primer mismatch tolerance when extending an RNA template.**

This example demonstrates that the D580G and I709K mutations result in polymerases having improved tolerance for primer mismatches to an RNA template.

**Mismatch Tolerance**

HCV RNA transcripts were mutated in regions under primer 3'-ends such that terminal, N-1, and N-2 mismatches could be systematically evaluated in RT-PCR conditions (120 mM KOAc, 3.3 mM $Mn^{2+}$, 60 mM Tricine; 50 uL total reaction volume) with polymerases Z05, Z05 D580G and Z05 D580G I709K. These reactions were run in a Roche LightCycler 480 Real Time PCR Instrument with a 12 minute RT step followed by 50 cycles of denaturation and extension. Real time fluorescence was detected in the JA270 and CY5.5 channels during the last 50 cycles. The Cp (crossing point) values from growth curves generated by fluorescent 5' nuclease (TaqMan) activity for each reaction was determined using the instrument's "2nd derivative Max analysis" method. The primer mismatch tolerance for various DNA polymerases were determined by comparing the Cp values. As shown in Table 15, Z05 D580G consistently had much earlier Cp values than Z05 when mismatched primers were used (N refers to position on the primer 3'-terminus with the primer:template mismatch indicated below). Importantly, Z05 D580G was able to detect several mismatches that the parental Z05 enzyme could not.

**TABLE 15. Cp values of Z05 and Z05 D580G polymerases using mismatched primers.**

| Enzyme | No mismatch | N A:A | N A:G | N-1 T:C | N-1 T:T | N-2 T:C | N-2 T:T |
|---|---|---|---|---|---|---|---|
| Z05 | 25.2 | No signal | 49.4 | 48.9 | 33.2 | 47.2 | No signal |
| Z05 D580G | 24.7 | 33.1 | 31.2 | 32.7 | 30.1 | 30.9 | 34.3 |

In Table 16, delta Cp values were determined by comparing the Cp values of Z05 D580G to Z05 D580G I709K for each mismatch as indicated. Thus, positive delta Cp values indicate how many cycles earlier TaqMan® signal was detected by Z05 D580G I709K. Overall, the mutant Z05 D580G I709K displays the greatest primer mismatch tolerance, providing a 4 cycle

improvement on average over the parental enzyme Z05 D580G for the mismatches shown, reflecting a 16 fold improvement in PCR performance.

**TABLE 16. delta Cp of Z05 D580G vs. Z05 D580G I709K polymerases using mismatched primers.**

| | No mismatch | N C:T | N C:A | N-1 A:A | N-1 A:C | N-2 C:C | N-2 Insertion |
|---|---|---|---|---|---|---|---|
| delta Cp | 1.3 | 4.9 | 2.6 | 5.2 | 2.4 | 4.0 | 4.8 |

The above example shows that substitutions as positions 580 and 709 of the Z05 polymerase result in improved RT-PCR efficiency when using mismatched primers.

***Example 8*: The 709 mutation improves RT-PCR efficiency.**

This example demonstrates that the I709K single mutant in the Z05 DNA polymerase results in a polymerase having improved RT-PCR efficiency without reduced efficiency when amplifying a DNA template.

The I709K mutation was subcloned into Z05 DNA polymerase backbone as a single mutant.After expression and purification, RT-PCR efficiencies of mutant Z05 I709K were compared with DNA polymerases Z05, Z05 D580G, and Z05 D580G I709K in $Mn^{2+}$- based TaqMan® RT-PCR. Master Mix conditions were the same as those described previously in Table 4, except $Mn(OAc)_2$ concentration was 1.5 mM, UNG concentration was 0.2U/μL, and probe concentration was 100 nM. Each DNA polymerase was diluted in buffer containing 20 mMTris-HCl, pH 8, 100 mMKCl, 0.1 mM EDTA, and 0.1% Tween-20 to make individual 5X enzyme stocks.Then 3μL of 5X enzyme stock was added to the appropriatereaction well for a final enzyme concentration of 20 nMin a total reaction volume of 15 μL. Each reaction had either 100,000 copies of JP2-5 transcript, 100,000 copies of pJP2-5 linear plasmid DNA, or 1000 copies of pJP2-5 linear plasmid DNA. All targets were amplified with the same primer set in replicates of four reactions to generate a 240 base pair amplicon. All reactions were performed on the Roche Light Cycler 480 thermal cycler. Crossing Point (Cps) values were calculated by the Abs Quant/$2^{nd}$ derivative max method and averaged. The thermocycling conditions were: 2 minutes at 50 ºC ("UNG" step); 55 ºC- 30seconds, 60 ºC- 1 minute, and 65 ºC- 1.5 minutes (three temperature "RT"step); 5 cycles of 94 ºC for 15 seconds followed by 62

°C for 30 seconds; and 45 cycles of 91 °C for 15 seconds followed by 62 °C for 30 seconds. Table 17 shows the Cp values obtained from fluorescent signal increase due to cleavage of the TaqMan® probe.

**Table 17. Cp values of mutant polymerases in RT-PCR.**

| Enzyme | RNA $10^5$ copies Cp | DNA $10^5$ copies Cp | DNA $10^3$ copies Cp |
|---|---|---|---|
| Z05 | 34.3 | 17.8 | 25.1 |
| Z05 D580G | 22.0 | 17.7 | 24.8 |
| Z05 I709K | 20.8 | 17.6 | 24.7 |
| Z05 D580G I709K | 18.6 | 17.5 | 24.5 |

This example shows that the I709K mutation results in increased reverse transcription and amplification efficiency using an RNA template when compared to the Z05 parental enzyme without a decrease in amplification efficiency using a DNA template.

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.

# INFORMAL SEQUENCE LISTING

**SEQ ID NO:1** ***Thermus sp.*** **Z05 DNA polymerase (Z05)**

MKAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAK
APSFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGFEADDVLATLAKKAEREGYEVRIL
TADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGLKPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLK
EWGSLENILKNLDRVKPESVRERIKAHLEDLKLSLELSRVRSDLPLEVDFARRREPDREGLRAFLERLEF
GSLLHEFGLLEAPAPLEEAPWPPPEGAFVGFVLSRPEPMWAELKALAACKEGRVHRAKDPLAGLKDLKEV
RGLLAKDLAVLALREGLDLAPSDDPMLLAYLLDPSNTTPEGVARRYGGEWTEDAAHRALLAERLQQNLLE
RLKGEEKLLWLYQEVEKPLSRVLAHMEATGVRLDVAYLKALSLELAEEIRRLEEEVFRLAGHPFNLNSRD
QLERVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQHRELTKLKNTYVDPLPGLVHPRTG
RLHTRFNQTATATGRLSSSDPNLQNIPIRTPLGQRIRRAFVAEAGWALVALDYSQIELRVLAHLSGDENL
IRVFQEGKDIHTQTASWMFGVSPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELAIPYEEAVAFIERYFQ
SFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKL
FPHLREMGARMLLQVHDELLLEAPQARAEEVAALAKEAMEKAYPLAVPLEVEVGIGEDWLSAKG

**SEQ ID NO:2** ***Thermus aquaticus*** **DNA polymerase (Taq)**

MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKA
PSFRHEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILT
ADKDLYQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEE
WGSLEALLKNLDRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGS
LLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARG
LLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRL
EGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQL
ERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRL
HTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIR
VFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSF
PKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFP
RLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE

**SEQ ID NO:3** ***Thermus filiformis*** **DNA polymerase (Tfi)**

MLPLLEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGEVAIVVFDAKAPSF
RHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGLVRLEVPGFEADDVLATLARKAEREGYEVRILSADR
DLYQLLSDRIHLLHPEGEVLTPGWLQERYGLSPERWVEYRALVGDPSDNLPGVPGIGEKTALKLLKEWGS
LEAILKNLDQVKPERVWEAIRNNLDKLQMSLELSRLRTDLPLEVDFAKRREPTGKGLKAFLERLEFGSLL
HEFGLLEAPKEAEEAPWPPPGGAFLGFLLSRPEPMWAELLALAGAKEGRVHRAEDPVGALKDLKEIRGLL
AKDLSVLALREGREIPPGDDPMLLAYLLDPGNTNPEGVARRYGGEWKEDAAARALLSERLWQALYPRVAE
EERLLWLYREVERPLAQVLAHMEATGVRLDVPYLEALSQEVAFELERLEAEVHRLAGHPFNLNSRDQLER
VLFDELGLPPIGKTEKTGKRSTSAAVLELLREAHPIVGRILEYRELMKLKSTYIDPLPRLVHPKTGRLHT
RFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRKAFIAEEGHLLVALDYSQIELRVLAHLSGDENLIRVF
REGKDIHTETAAWMFGVPPEGVDGAMRRAAKTVNFGVLYGMSAHRLSQELSIPYEEAAAFIERYFQSFPK
VRAWIAKTLEEGRKKGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRL
RPLGVRILLQVHDELVLEAPKARAEEAAQLAKETMEGVYPLSVPLEVEVGMGEDWLSAKE

**SEQ ID NO:4** ***Thermus flavus*** **DNA polymerase (Tfl)**

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDVVVVVFDAKAP
SFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGLVRLEVPGFEADDVLATLAKRAEKEGYEVRILTA
DRDLYQLLSERIAILHPEGYLITPAWLYEKYGLRPEQWVDYRALAGDPSDNIPGVKGIGEKTAQRLIREW
GSLENLFQHLDQVKPSLREKLQAGMEALALSRKLSQVHTDLPLEVDFGRRRTPNLEGLRAFLERLEFGSL
LHEFGLLEGPKAAEEAPWPPPEGAFLGFSFSRPEPMWAELLALAGAWEGRLHRAQDPLRGLRDLKGVRGI
LAKDLAVLALREGLDLFPEDDPMLLAYLLDPSNTTPEGVARRYGGEWTEDAGERALLAERLFQTLKERLK
GEERLLWLYEEVEKPLSRVLARMEATGVRLDVAYLQALSLEVEAEVRQLEEEVFRLAGHPFNLNSRDQLE
RVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVDRILQYRELTKLKNTYIDPLPALVHPKTGRLH

TRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFVAEEGWVLVVLDYSQIELRVLAHLSGDENLIRV
FQEGRDIHTQTASWMFGVSPEGVDPLMRRAAKTINFGVLYGMSAHRLSGELSIPYEEAVAFIERYFQSYP
KVRAWIEGTLEEGRRRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVRLFPR
LQELGARMLLQVHDELVLEAPKDRAERVAALAKEVMEGVWPLQVPLEVEVGLGEDWLSAKE

**SEQ ID NO:5** ***Thermus sp.*** **Sps17 DNA polymerase (Sps17)**

MLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGEVAIVVFDAKAPSF
RHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGLVRLEVPGFEADDVLATLAKKAEREGYEVRILSADR
DLYQLLSDRIHLLHPEGEVLTPGWLQERYGLSPERWVEYRALVGDPSDNLPGVPGIGEKTALKLLKEWGS
LEAILKNLDQVKPERVREAIRNNLDKLQMSLELSRLRTDLPLEVDFAKRREPDWEGLKAFLERLEFGSLL
HEFGLLEAPKEAEEAPWPPPGGAFLGFLLSRPEPMWAELLALAGAKEGRVHRAEDPVGALKDLKEIRGLL
AKDLSVLALREGREIPPGDDPMLLAYLLDPGNTNPEGVARRYGGEWKEDAAARALLSERLWQALYPRVAE
EERLLWLYREVERPLAQVLAHMEATGVRLDVPYLEALSQEVAFELERLEAEVHRLAGHPFNLNSRDQLER
VLFDELGLPPIGKTEKTGKRSTSAAVLELLREAHPIVGRILEYRELMKLKSTYIDPLPRLVHPKTGRLHT
RFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRKAFIAEEGHLLVALDYSQIELRVLAHLSGDENLIRVF
REGKDIHTETAAWMFGVPPEGVDGAMRRAAKTVNFGVLYGMSAHRLSQELSIPYEEAAAFIERYFQSFPK
VRAWIAKTLEEGRKKGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRL
RPLGVRILLQVHDELVLEAPKARAEEAAQLAKETMEGVYPLSVPLEVEVGMGEDWLSAKA

**SEQ ID NO:6** ***Thermus thermophilus*** **DNA polymerase (Tth)**

MEAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAK
APSFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRIL
TADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLK
EWGSLENLLKNLDRVKPENVREKIKAHLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEF
GSLLHEFGLLEAPAPLEEAPWPPPEGAFVGFVLSRPEPMWAELKALAACRDGRVHRAADPLAGLKDLKEV
RGLLAKDLAVLASREGLDLVPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEDAAHRALLSERLHRNLLK
RLEGEEKLLWLYHEVEKPLSRVLAHMEATGVRRDVAYLQALSLELAEEIRRLEEEVFRLAGHPFNLNSRD
QLERVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQHRELTKLKNTYVDPLPSLVHPRTG
RLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFVAEAGWALVALDYSQIELRVLAHLSGDENL
IRVFQEGKDIHTQTASWMFGVPPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELAIPYEEAVAFIERYFQ
SFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKL
FPRLREMGARMLLQVHDELLLEAPQARAEEVAALAKEAMEKAYPLAVPLEVEVGMGEDWLSAKG

**SEQ ID NO:7** ***Thermus caldophilus*** **DNA polymerase (Tca)**

MEAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAK
APSFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKNPEKEGYEVRIL
TADRDLDQLVSDRVAVLHPEGHLITPEWLWQKYGLKPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLK
EWGSLENLLKNLDRVKPENVREKIKAHLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEF
GSLLHEFGLLEAPAPLEEAPWPPPEGAFVGFVLSRPEPMWAELKALAACRDGRVHRAADPLAGLKDLKEV
RGLLAKDLAVLASREGLDLVPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEDAAHRALLSERLHRNLLK
RLQGEEKLLWLYHEVEKPLSRVLAHMEATGVRLDVAYLQALSLELAEEIRRLEEEVFRLAGHPFNLNSRD
QLERVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQHRELTKLKNTYVDPLPSLVHPNTG
RLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFVAEAGWALVALDYSQIELRVLAHLSGDENL
IRVFQEGKDIHTQTASWMFGVPPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELAIPYEEAVAFIERYFQ
SFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKL
FPRLREMGARMLLQVHDELLLEAPQAGAEEVAALAKEAMEKAYPLAVPLEVEVGMGEDWLSAKG

**SEQ ID NO:8**

$X_1X_2X_3X_4X_5X_6X_7X_8X_9X_{10}X_{11}X_{12}X_{13}$GYV$X_{14}$TL, wherein $X_1$ is A, D, S, E, R or Q; $X_2$ is W or Y; $X_3$ is any amino acid other than I, L or M; $X_4$ is E, A, Q, K, N or D; $X_5$ is K, G, R, Q, H or N; $X_6$ is T, V, M or I; $X_7$ is L, V or K; $X_8$ is E, S, A, D or Q; $X_9$ is E or F; $X_{10}$ is G or A; $X_{11}$ is R or K; $X_{12}$ is K, R, E, T or Q; $X_{13}$ is R, K or H; and $X_{14}$ is E, R or T.

**SEQ ID NO:9**
$X_1X_2X_3X_4X_5X_6X_7X_8EX_{10}X_{11}X_{12}X_{13}GYVX_{14}TL$, wherein $X_1$ is A, D, or S; $X_2$ is W or Y; $X_3$ is any amino acid other than I; $X_4$ is E, A, or Q; $X_5$ is K, G, R or Q; $X_6$ is T or V; $X_7$ is L or V; $X_8$ is E, S or A; $X_{10}$ is G or A; $X_{11}$ is R or K; $X_{12}$ is K, R or E; $X_{13}$ is R or K; and $X_{14}$ is E or R.

**SEQ ID NO:10**
$AWX_3X_4X_5TLEEGRX_{12}X_{13}GYVETL$, wherein $X_3$ is any amino acid other than I; $X_4$ is E or A; $X_5$ is K or G; $X_{12}$ is K or R; and $X_{13}$ is R or K.

**SEQ ID NO:11**
$AWX_3X_4X_5TLEEGRX_{12}X_{13}GYVETL$, wherein $X_3$ is K, R, S, G, or A; $X_4$ is E or A; $X_5$ is K or G; $X_{12}$ is K or R; and $X_{13}$ is R or K.

**SEQ ID NO:12 ZO5**
AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA

**SEQIDNO:13 Taq**
AWIEKTLEEGRRRGYVETLFGRRRYVPDLEA

**SEQIDNO:14 Tfi**
AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA

**SEQIDNO:15 Tfl**
AWIEGTLEEGRRRGYVETLFGRRRYVPDLNA

**SEQIDNO:16 Sps17**
AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA

**SEQIDNO:17 Tth**
AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA

**SEQIDNO:18 Tca**
AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA

**SEQIDNO:19 Tma**
DYIQRVVSEAKEKGYVRTLFGRKRDIPQLMA

**SEQIDNO:20 Tne**
SYIQQVVAEAKEKGYVRTLFGRKRDIPQLMA

**SEQIDNO:21 Taf**
EYLKRMKDEARKKGYVTTLFGRRRYIPQLRS

**SEQIDNO:22 DNA polymerase acitve site motif A**
DYSQIELR

**SEQ ID NO:23 Dra**
RYINHTLDFGRTHGYVETLYGRRRYVPGLSS

**SEQIDNO:24 Bst**
QYMDNIVQEAKQKGYVTTLLHRRRYLPDITS

**SEQIDNO:25 Bca**
RYMENIVQEAKQKGYVTTLLHRRRYLPDITS

**SEQ ID NO:26 native consensus motif**
$X_1X_2X_3X_4X_5X_6X_7X_8X_9X_{10}X_{11}X_{12}X_{13}GYVX_{14}TL$, wherein $X_1$ is A, D, S, E, R or Q; $X_2$ is W or Y; $X_3$ is I, L or M; $X_4$ is E, A, Q, K, N or D; $X_5$ is K, G, R, Q,

H or N; $X_6$ is T, V, M or I; $X_7$ is L, V or K; $X_8$ is E, S, A, D or Q; $X_9$ is E or F; $X_{10}$ is G or A; $X_{11}$ is R or K; $X_{12}$ is K, R, E, T or Q; $X_{13}$ is R, K or H; and $X_{14}$ is E, R or T.

**SEQ ID NO:27 CS5 DNA polymerase**

MKAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAK
APSFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGFEADDVLATLAKKAEREGYEVRIL
TADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGLKPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLK
EWGSLENILKNLDRVKPESVRERIKAHLEDLKLSLELSRVRSDLPLEVDFARRREPDREGLRAFLERLEF
GSLLHEFGLLEESEPVGYRIVKDLVEFEKLIEKLRESPSFAIDLETSSLDPFDCDIVGISVSFKPKEAYY
IPLHHRNAQNLDEKEVLKKLKEILEDPGAKIVGQNLKFDYKVLMVKGVEPVPPYFDTMIAAYLLEPNEKK
FNLDDLALKFLGYKMTSYQELMSFSFPLFGFSFADVPVEKAANYSCEDADITYRLYKTLSLKLHEADLEN
VFYKIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEEIYRIAGEPFNINSPKQVSRILFEKL
GIKPRGKTTKTGDYSTRIEVLEELAGEHEIIPLILEYRKIQKLKSTYIDALPKMVNPKTGRIHASFNQTG
TATGRLSSSDPNLQNLPTKSEEGKEIRKAIVPQDPNWWIVSADYSQIELRILAHLSGDENLLRAFEEGID
VHTLTASRIFNVKPEEVTEEMRRAGKMVNFSIIYGVTPYGLSVRLGVPVKEAEKMIVNYFVLYPKVRDYI
QRVVSEAKEKGYVRTLFGRKRDIPQLMARDRNTQAEGERIAINTPIQGTAADIIKLAMIEIDRELKERKM
RSKMIIQVHDELVFEVPNEEKDALVELVKDRMTNVVKLSVPLEVDVTIGKTWS

**SEQ ID NO:28 CS6 DNA polymerase**

MKAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAK
APSFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGFEADDVLATLAKKAEREGYEVRIL
TADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGLKPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLK
EWGSLENILKNLDRVKPESVRERIKAHLEDLKLSLELSRVRSDLPLEVDFARRREPDREGLRAFLERLEF
GSLLHEFGLLEESEPVGYRIVKDLVEFEKLIEKLRESPSFAIALATSSLDPFDCDIVGISVSFKPKEAYY
IPLHHRNAQNLDEKEVLKKLKEILEDPGAKIVGQNLKFDYKVLMVKGVEPVPPYFDTMIAAYLLEPNEKK
FNLDDLALKFLGYKMTSYQELMSFSFPLFGFSFADVPVEKAANYSCEDADITYRLYKTLSLKLHEADLEN
VFYKIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEEIYRIAGEPFNINSPKQVSRILFEKL
GIKPRGKTTKTGDYSTRIEVLEELAGEHEIIPLILEYRKIQKLKSTYIDALPKMVNPKTGRIHASFNQTG
TATGRLSSSDPNLQNLPTKSEEGKEIRKAIVPQDPNWWIVSADYSQIELRILAHLSGDENLLRAFEEGID
VHTLTASRIFNVKPEEVTEEMRRAGKMVNFSIIYGVTPYGLSVRLGVPVKEAEKMIVNYFVLYPKVRDYI
QRVVSEAKEKGYVRTLFGRKRDIPQLMARDRNTQAEGERIAINTPIQGTAADIIKLAMIEIDRELKERKM
RSKMIIQVHDELVFEVPNEEKDALVELVKDRMTNVVKLSVPLEVDVTIGKTWS

**SEQ ID NO:29**

PNLQN$X_1$P$X_2X_3X_4X_5X_6$G, wherein $X_1$ is I or L; $X_2$ is any amino acid other than I or V; $X_3$ is R or K; $X_4$ is T, S or L; $X_5$ is P or E; and $X_6$ is L or E.

**SEQ ID NO:30 Forward Primer**

5'- CTACCTCCTGGACCCCTCCAA-3'

**SEQ ID NO:31 Reverse Primer**

5'- ATAACCAACTGGTAGTGGCGTGTAA-3'

**SEQ ID NO:32 *Deinococcus radiodurans* DNA polymerase (Dra)**

MADASPDPSKPDALVLIDGHALAFRSYFALPPLNNSKGEMTDAIVGFMKLLLRLARQKSNQVIVVFDPPV
KTLRHEQYEGYKSGRAQTPEDLRGQINRIRALVDALGFPRLEEPGYEADDVIASLTRMAEGKGYEVRIVT
SDRDAYQLLDEHVKVIANDFSLIGPAQVEEKYGVTVRQWVDYRALTGDASDNIPGAKGIGPKTAAKLLQE
YGTLEKVYEAAHAGTLKPDGTRKKLLDSEENVKFSHDLSCMVTDLPLDIEFGVRRLPDNPLVTEDLLTEL
ELHSLRPMILGLNGPEQDGHAPDDLLEREHAQTPEEDEAAALPAFSAPELAEWQTPAEGAVWGYVLSRED
DLTAALLAAATFEDGVARPARVSEPDEWAQAEAPENLFGELLPSDKPLTKKEQKALEKAQKDAEKARAKL
REQFPATVDEAEFVGQRTVTAAAAKALAAHLSVRGTVVEPGDDPLLYAYLLDPANTNMPVVAKRYLDREW
PADAPTRAAITGHLVRELPPLLDDARRKMYDEMEKPLSGVLGRMEVRGVQVDSDFLQTLSIQAGVRLADL
ESQIHEYAGEEFHIRSPKQLETVLYDKLELASSKKTKLTGQRSTAVSALEPLRDAHPIIPLVLEFRELDK

LRGTYLDPIPNLVNPHTGRLHTTFAQTAVATGRLSSLNPNLQNIPIRSELGREIRKGFIAEDGFTLIAAD
YSQIELRLLAHIADDPLMQQAFVEGADIHRRTAAQVLGLDEATVDANQRRAAKTVNFGVLYGMSAHRLSN
DLGIPYAEAATFIEIYFATYPGIRRYINHTLDFGRTHGYVETLYGRRRYVPGLSSRNRVQREAEERLAYN
MPIQGTAADIMKLAMVQLDPQLDAIGARMLLQVHDELLIEAPLDKAEQVAALTKKVMENVVQLKVPLAVE
VGTGPNWFDTK

**SEQ ID NO:33** ***Thermosipho africanus*** **DNA polymerase (Taf)**

MGKMFLFDGTGLVYRAFYAIDQSLQTSSGLHTNAVYGLTKMLIKFLKEHISIGKDACVFVLDSKGGSKKR
KDILETYKANRPSTPDLLLEQIPYVEELVDALGIKVLKIEGFEADDIIATLSKKFESDFEKVNIITGDKD
LLQLVSDKVFVWRVERGITDLVLYDRNKVIEKYGIYPEQFKDYLSLVGDQIDNIPGVKGIGKKTAVSLLK
KYNSLENVLKNINLLTEKLRRLLEDSKEDLQKSIELVELIYDVPMDVEKDEIIYRGYNPDKLLKVLKKYE
FSSIIKELNLQEKLEKEYILVDNEDKLKKLAEEIEKYKTFSIDTETTSLDPFEAKLVGISISTMEGKAYY
IPVSHFGAKNISKSLIDKFLKQILQEKDYNIVGQNLKFDYEIFKSMGFSPNVPHFDTMIAAYLLNPDEKR
FNLEELSLKYLGYKMISFDELVNENVPLFGNDFSYVPLERAVEYSCEDADVTYRIFRKLGRKIYENEMEK
LFYEIEMPLIDVLSEMELNGVYFDEEYLKELSKKYQEKMDGIKEKVFEIAGETFNLNSSTQVAYILFEKL
NIAPYKKTATGKFSTNAEVLEELSKEHEIAKLLLEYRKYQKLKSTYIDSIPLSINRKTNRVHTTFHQTGT
STGRLSSSNPNLQNLPTRSEEGKEIRKAVRPQRQDWWILGADYSQIELRVLAHVSKDENLLKAFKEDLDI
HTITAAKIFGVSEMFVSEQMRRVGKMVNFAIIYGVSPYGLSKRIGLSVSETKKIIDNYFRYYKGVFEYLK
RMKDEARKKGYVTTLFGRRRYIPQLRSKNGNRVQEGERIAVNTPIQGTAADIIKIAMINIHNRLKKENLR
SKMILQVHDELVFEVPDNELEIVKDLVRDEMENAVKLDVPLKVDVYYGKEWE

**SEQ ID NO:34** ***Thermotoga maritima*** **DNA polymerase (Tma)**

MARLFLFDGTALAYRAYYALDRSLSTSTGIPTNATYGVARMLVRFIKDHIIVGKDYVAVAFDKKAATFRH
KLLETYKAQRPKTPDLLIQQLPYIKKLVEALGMKVLEVEGYEADDIIATLAVKGLPLFDEIFIVTGDKDM
LQLVNEKIKVWRIVKGISDLELYDAQKVKEKYGVEPQQIPDLLALTGDEIDNIPGVTGIGEKTAVQLLEK
YKDLEDILNHVRELPQKVRKALLRDRENAILSKKLAILETNVPIEINWEELRYQGYDREKLLPLLKELEF
ASIMKELQLYEESEPVGYRIVKDLVEFEKLIEKLRESPSFAIDLETSSLDPFDCDIVGISVSFKPKEAYY
IPLHHRNAQNLDEKEVLKKLKEILEDPGAKIVGQNLKFDYKVLMVKGVEPVPPYFDTMIAAYLLEPNEKK
FNLDDLALKFLGYKMTSYQELMSFSFPLFGFSFADVPVEKAANYSCEDADITYRLYKTLSLKLHEADLEN
VFYKIEMPLVNVLARMELNGVYVDTEFLKKLSEEYGKKLEELAEEIYRIAGEPFNINSPKQVSRILFEKL
GIKPRGKTTKTGDYSTRIEVLEELAGEHEIIPLILEYRKIQKLKSTYIDALPKMVNPKTGRIHASFNQTG
TATGRLSSSDPNLQNLPTKSEEGKEIRKAIVPQDPNWWIVSADYSQIELRILAHLSGDENLLRAFEEGID
VHTLTASRIFNVKPEEVTEEMRRAGKMVNFSIIYGVTPYGLSVRLGVPVKEAEKMIVNYFVLYPKVRDYI
QRVVSEAKEKGYVRTLFGRKRDIPQLMARDRNTQAEGERIAINTPIQGTAADIIKLAMIEIDRELKERKM
RSKMIIQVHDELVFEVPNEEKDALVELVKDRMTNVVKLSVPLEVDVTIGKTWS

**SEQ ID NO:35** ***Thermotoga neopolitana*** **DNA polymerase (Tne)**

MARLFLFDGTALAYRAYYALDRSLSTSTGIPTNAVYGVARMLVKFIKEHIIPEKDYAAVAFDKKAATFRH
KLLVSDKAQRPKTPALLVQQLPYIKRLIEALGFKVLELEGYEADDIIATLAVRAARFLMRFSLITGDKDM
LQLVNEKIKVWRIVKGISDLELYDSKKVKERYGVEPHQIPDLLALTGDDIDNIPGVTGIGEKTAVQLLGK
YRNLEYILEHARELPQRVRKALLRDREVAILSKKLATLVTNAPVEVDWEEMKYRGYDKRKLLPILKELEF
ASIMKELQLYEEAEPTGYEIVKDHKTFEDLIEKLKEVPSFALDLETSSLDPFNCEIVGISVSFKPKTAYY
IPLHHRNAHNLDETLVLSKLKEILEDPSSKIVGQNLKYDYKVLMVKGISPVYPHFDTMIAAYLLEPNEKK
FNLEDLSLKFLGYKMTSYQELMSFSSPLFGFSFADVPVDKAAEYSCEDADITYRLYKILSMKLHEAELEN
VFYRIEMPLVNVLARMEFNWVYVDTEFLKKLSEEYGKKLEELAEKIYQIAGEPFNINSPKQVSNILFEKL
GIKPRGKTTKTGDYSTRIEVLEEIANEHEIVPLILEFRKILKLKSTYIDTLPKLVNPKTGRFHASFHQTG
TATGRLSSSDPNLQNLPTKSEEGKEIRKAIVPQDPDWWIVSADYSQIELRILAHLSGDENLVKAFEEGID
VHTLTASRIYNVKPEEVNEEMRRVGKMVNFSIIYGVTPYGLSVRLGIPVKEAEKMIISYFTLYPKVRSYI
QQVVAEAKEKGYVRTLFGRKRDIPQLMARDKNTQSEGERIAINTPIQGTAADIIKLAMIDIDEELRKRNM
KSRMIIQVHDELVFEVPDEEKEELVDLVKNKMTNVVKLSVPLEVDISIGKSWS

**SEQ ID NO:36** ***Bacillus stearothermophilus*** **DNA polymerase (Bst)**

MKNKLVLIDGNSVAYRAFFALPLLHNDKGIHTNAVYGFTMMLNKILAEEQPTHILVAFDAGKTTFRHETF
QDYKGGRQQTPPELSEQFPLLRELLKAYRIPAYELDHYEADDIIGTMAARAEREGFAVKVISGDRDLTQL
ASPQVTVEITKKGITDIESYTPETVVEKYGLTPEQIVDLKGLMGDKSDNIPGVPGIGEKTAVKLLKQFGT
VENVLASIDEIKGEKLKENLRQYRDLALLSKQLAAICRDAPVELTLDDIVYKGEDREKVVALFQELGFQS
FLDKMAVQTDEGEKPLAGMDFAIADSVTDEMLADKAALVVEVVGDNYHHAPIVGIALANERGRFFLRPET
ALADPKFLAWLGDETKKKTMFDSKRAAVALKWKGIELRGVVFDLLLAAYLLDPAQAAGDVAAVAKMHQYE
AVRSDEAVYGKGAKRTVPDEPTLAEHLARKAAAIWALEEPLMDELRRNEQDRLLTELEQPLAGILANMEF
TGVKVDTKRLEQMGAELTEQLQAVERRIYELAGQEFNINSPKQLGTVLFDKLQLPVLKKTKTGYSTSADV
LEKLAPHHEIVEHILHYRQLGKLQSTYIEGLLKVVHPVTGKVHTMFNQALTQTGRLSSVEPNLQNIPIRL
EEGRKIRQAFVPSEPDWLIFAADYSQIELRVLAHIAEDDNLIEAFRRGLDIHTKTAMDIFHVSEEDVTAN
MRRQAKAVNFGIVYGISDYGLAQNLNITRKEAAEFIERYFASFPGVKQYMDNIVQEAKQKGYVTTLLHRR
RYLPDITSRNFNVRSFAERTAMNTPIQGSAADIIKKAMIDLSVRLREERLQARLLLQVHDELILEAPKEE
IERLCRLVPEVMEQAVALRVPLKVDYHYGPTWYDAK

**SEQ ID NO:37** ***Bacillus caldotenax*** **DNA polymerase (Bca)**

MKKKLVLIDGSSVAYRAFFALPLLHNDKGIHTNAVYGFTMMLNKILAEEEPTHMLVAFDAGKTTFRHEAF
QEYKGGRQQTPPELSEQFPLLRELLRAYRIPAYELENYEADDIIGTLAARAEQEGFEVKVISGDRDLTQL
ASPHVTVDITKKGITDIEPYTPEAVREKYGLTPEQIVDLKGLMGDKSDNIPGVPGIGEKTAVKLLRQFGT
VENVLASIDEIKGEKLKETLRQHREMALLSKKLAAIRRDAPVELSLDDIAYQGEDREKVVALFKELGFQS
FLEKMESPSSEEEKPLAKMAFTLADRVTEEMLADKAALVVEVVEENYHDAPIVGIAVVNEHGRFFLRPET
ALADPQFVAWLGDETKKKSMFDSKRAAVALKWKGIELCGVSFDLLLAAYLLDPAQGVDDVAAAAKMKQYE
AVRPDEAVYGKGAKRAVPDEPVLAEHLVRKAAAIWALERPFLDELRRNEQDRLLVELEQPLSSILAEMEF
AGVKVDTKRLEQMGEELAEQLRTVEQRIYELAGQEFNINSPKQLGVILFEKLQLPVLKKSKTGYSTSADV
LEKLAPYHEIVENILQHYRQLGKLQSTYIEGLLKVVRPDTKKVHTIFNQALTQTGRLSSTEPNLQNIPIR
LEEGRKIRQAFVPSESDWLIFAADYSQIELRVLAHIAEDDNLMEAFRRDLDIHTKTAMDIFQVSEDEVTP
NMRRQAKAVNFGIVYGISDYGLAQNLNISRKEAAEFIERYFESFPGVKRYMENIVQEAKQKGYVTTLLHR
RRYLPDITSRNFNVRSFAERMAMNTPIQGSAADIIKKAMIDLNARLKEERLQARLLLQVHDELILEAPKE
EMERLCRLVPEVMEQAVTLRVPLKVDYHYGSTWYDAK

**SEQ ID NO:38 modified Z05 D580 motif**

T-G-R-L-S-S-$X_7$-$X_8$-P-N-L-Q-N, wherein $X_7$ is S or T; and $X_8$ is any amino acid other than D or E.

**SEQ ID NO:39 synthetic amplicon encoding Z05 D580G DNA polymerase**

ctacctcctggacccctccaacaccacccccgaggggtggcccggcgctacgggggggagtggacggag
gacgccgcccaccgggccctcctcgctgagcggctccagcaaaacctcttggaacgcctcaagggagagg
aaaagctcctttggctctaccaagaggtggaaaagcccctctcccgggtcctggcccacatggaggccac
cggggtaaggctggacgtggcctatctaaaggccctttccctggagcttgcggaggagattcgccgcctc
gaggaggaggtcttccgcctggcgggccacccccttcaacctgaactcccgtgaccagctagagcgggtgc
tctttgacgagcttaggcttcccgccctgggcaagacgcaaaagacggggaagcgctccaccagcgccgc
ggtgctggaggccctcagggaggcccaccccatcgtggagaagatcctccagcaccgggagctcaccaag
ctcaagaacacctacgtagacccccctcccgggcctcgtccacccgaggacgggccgcctccacacccgct
tcaaccagacagccacggccacgggaaggctctctagctccgggcccaacctgcagaacatccccatccg
cacccccttgggccagaggatccgccgggccttcgtggccgaggcgggatgggcgttggtggccctggac
tatagccagatagagctccgggtcctcgcccacctctccggggacgagaacctgatcagggtcttccagg
aggggaaggacatccacacccagaccgcaagctggatgttcggcgtctccccggaggccgtggacccccct
gatgcgccgggcggccaagacggtgaacttcggcgtcctctacggcatgtccgcccataggctctcccag
gagcttgccatcccctacgaggaggcggtggcctttatagagcgctacttccaaagcttccccaaggtgc
gggcctggatagaaaagaccctggaggaggggaggaagcggggctacgtggaaaccctcttcggaagaag
gcgctacgtgcccgacctcaacgcccgggtgaagagcgtcagggaggccgcggagcgcatggccttcaac
atgcccgtccagggcaccgccgccgacctcatgaagctcgccatggtgaagctcttcccccacctccggg
agatgggggcccgcatgctcctccaggtccacgacgagctcctcctggaggcccccccaagcgcgggccga
ggaggtggcggctttggccaaggaggccatggagaaggcctatcccctcgccgtgcccctggaggtggag

gtggggatcggggaggactggctttccgccaagggctgatatcagatctccctgattatgcgtcagtcta
tgaagaaaaatcgtatacagatggacgaagagagaatccttgtgaatttaacagagggtatagggattac
acgccactaccagttggttat

**SEQ ID NO:40**
gggaagggcgatcggtgcgggcctcttcgc

**SEQ ID NO:41**
gggaagggcgatcggtgcgggcctcttcgt

**SEQ ID NO:42**
gggaagggcgatcggtgcgggcctcttcgca

**SEQ ID NO:43**
gggaagggcgatcggtgcgggcctcttca

**SEQ ID NO:44**
gggaagggcgatcggtgcgggcctcttcgctt

**SEQ ID NO:45**
gggaagggcgatcggtgcgggcctcttcgctc

**SEQ ID NO:46**
gggaagggcgatcggtgcgggcctcttcc

**SEQ ID NO:47**
gggaagggcgatcggtgcgggcctcttct

**SEQ ID NO:48** ***Carboxydothermus hydrogenoformans*** **DNA polymerase (Chy)**
MGKVVLVDGNSLLHRAFFALPPLKTTKGEPTGAVYEFLTMLFRVIKDEKPEYLAVAFDISRKTFRTEQFTAYKGHRK
EAPDELVPQFALVREVLKVLNVPYIELDGYEADDIIGHLSRAFAGQGHEVVIYTADRDMLQLVDEKTVVYLTKKGIT
ELVKMDLAAILENYGLKPKQLVDVKGLMGDPSDNIPGVPGIGEKTALDLIKTYGSVEEVLARKDELKPKLREKLAEH
ENLAKISKQLATILREIPLEISLEDLKVKEPNYEEVAKLFLHLEFKSFLKEIEPKIKKEYQEGKDLVQVETVETEGQ
IAVVFSDGFYVDDGEKTKFYSLDRLNEIEEIFRNKKIITDDAKGIYHVCLEKGLTFPEVCFDARIAAYVLNPADQNP
GLKGLYLKYDLPVYEDVSLNIRGLFYLKKEMMRKIFEQEQERLFYEIELPLTPVLAQMEHTGIQVDREALKEMSLEL
GEQIEELIREIYVLAGEEFNLNSPRQLGVILFEKLGLPVIKKTKTGYSTDAEVLEELLPFHEIIGKILNYRQLMKLK
STYTDGLMPLINERTGKLHTTFNQTGTLTGRLASSEPNLQNIPIRLELGRKLRKMFIPSPGYDYIVSADYSQIELRL
LAHFSEEPKLIEAYQKGEDIHRKTASEVFGVSLEEVTPEMRAHAKSVNFGIVYGISDFGLGRDLKIPREVAGKYIKN
YFANYPKVREYLDELVRTAREKGYVTTLFGRRRYIPELSSKNRTVQGFGERTAMNTPLQGSAADIIKLAMINVEKEL
KARKLKSRLLLSVHDELVLEVPAEELEEVKALVKGVMESVVELKVPLIAEVGAGKNWYEAK

# CLAIMS

1. A DNA polymerase having increased reverse transcriptase efficiency, mismatch tolerance, extension rate and/or tolerance of RT and polymerase inhibitors compared with a control DNA polymerase, wherein the amino acid of the DNA polymerase corresponding to position 709 of SEQ ID NO:1 is any amino acid other than I, L, or M, and wherein the control DNA polymerase has the same amino acid sequence as the DNA polymerase except that the amino acid of the control DNA polymerase corresponding to position 709 of SEQ ID NO:1 is I, L, or M.

2. The DNA polymerase of claim 1, comprising a motif in the polymerase domain comprising
$X_1$-$X_2$-$X_3$-$X_4$-$X_5$-$X_6$-$X_7$-$X_8$-$X_9$-$X_{10}$-$X_{11}$-$X_{12}$-$X_{13}$-G-Y-V-$X_{14}$-T-L (SEQ ID NO:8), wherein:
$X_1$ is A, D, S, E, R or Q;
$X_2$ is W or Y;
$X_3$ is any amino acid other than I, L or M;
$X_4$ is E, A, Q, K, N or D;
$X_5$ is K, G, R, Q, H or N;
$X_6$ is T, V, M or I;
$X_7$ is L, V or K;
$X_8$ is E, S, A, D or Q;
$X_9$ is E or F;
$X_{10}$ is G or A;
$X_{11}$ is R or K;
$X_{12}$ is K, R, E, T or Q;
$X_{13}$ is R, K or H; and
$X_{14}$ is E, R or T.

3. The DNA polymerase of claim 1, comprising a motif in the polymerase domain comprising
A-W-$X_3$-$X_4$-$X_5$-T-L-E-E-G-R-$X_{12}$-$X_{13}$-G-Y-V-E-T-L (SEQ ID NO:11), wherein:
$X_3$ is K, R, S, G, or A;

$X_4$ is E or A;

$X_5$ is K or G;

$X_{12}$ is K or R; and

$X_{13}$ is R or K.

4. The DNA polymerase of any one of claims 1 to 3, wherein the amino acid corresponding to position 580 of SEQ ID NO:1 is any amino acid other than D or E.

5. The DNA polymerase of any one of claims 1 to 4, wherein the amino acid corresponding to position 580 of SEQ ID NO:1 is selected from the group consisting of L, G, T, Q, A, S, N, R, and K.

6. The DNA polymerase of any one of claims 1 to 5, wherein the DNA polymerase has at least 80%, preferably at least 90%, more preferably at least 95% amino acid sequence identity to a polymerase selected from the group consisting of:

(a) SEQ ID NO:1;

(b) SEQ ID NO:2;

(c) SEQ ID NO:3;

(d) SEQ ID NO:4;

(e) SEQ ID NO:5;

(f) SEQ ID NO:6;

(g) SEQ ID NO:7;

(h) SEQ ID NO:32;

(i) SEQ ID NO:33;

(j) SEQ ID NO:34;

(k) SEQ ID NO:35;

(l) SEQ ID NO:36; and

(m) SEQ ID NO:37.

7. The DNA polymerase of any one of claims 1 to 6, wherein the polymerase has at least 80%, preferably at least 90%, more preferably at least 95% amino acid sequence identity to SEQ ID NO:1.

8. The DNA polymerase of claim 7, wherein the amino acid at position 580 is selected from the group consisting of L, G, T, Q, A, S, N, R, and K.

9. A recombinant nucleic acid encoding the DNA polymerase according to any one of claims 1 to 8.

10. A method for conducting primer extension, comprising:

contacting a DNA polymerase according to any one of claims 1 to 8 with a primer, a polynucleotide template, and nucleoside triphosphates under conditions suitable for extension of the primer, thereby producing an extended primer.

11. The method of claim 10, wherein the method occurs in the presence of at least one inhibitor of DNA polymerase activity and/or reverse transcription activity.

12. A kit for producing an extended primer, comprising at least one container providing a DNA polymerase according to any one of claims 1 to 8.

13. The kit according to claim 12, further comprising one or more additional containers selected from the group consisting of:

(a) a container providing a primer hybridizable, under primer extension conditions, to a predetermined polynucleotide template;

(b) a container providing nucleoside triphosphates; and

(c) a container providing a buffer suitable for primer extension.

14. A reaction mixture comprising a DNA polymerase according to any one of claims 1 to 8, at least one primer, a polynucleotide template, and nucleoside triphosphates.

15. The reaction mixture of claim 14, wherein the mixture comprises at least one inhibitor of DNA polymerase activity and/or reverse transcription activity.

***FIGURE 1***

```
          *
Z05    AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:12)
Taq    AWIEKTLEEGRRRGYVETLFGRRRYVPDLEA (SEQ ID NO:13)
Tfi    AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA (SEQ ID NO:14)
Tfl    AWIEGTLEEGRRRGYVETLFGRRRYVPDLNA (SEQ ID NO:15)
Sps17  AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA (SEQ ID NO:16)
Tth    AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:17)
Tca    AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:18)
Tma    DYIQRVVSEAKEKGYVRTLFGRKRDIPQLMA (SEQ ID NO:19)
Tne    SYIQQVVAEAKEKGYVRTLFGRKRDIPQLMA (SEQ ID NO:20)
Taf    EYLKRMKDEARKKGYVTTLFGRRRYIPQLRS (SEQ ID NO:21)
Dra    RYINHTLDFGRTHGYVETLYGRRRYVPGLSS (SEQ ID NO:23)
Bst    QYMDNIVQEAKQKGYVTTLLHRRRYLPDITS (SEQ ID NO:24)
Bca    RYMENIVQEAKQKGYVTTLLHRRRYLPDITS (SEQ ID NO:25)
   X1X2X3X4X5X6X7X8X9X10X11X12X13GYVX14TL---------- (SEQ ID NO:26)
```

# *FIGURE 2*

| A. | Sequence identities over the entire polymerase I enzyme (corresponding to amino acids 1-834 of Z05) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Name** | **Z05** | **Taq** | **Tfi** | **Tfl** | **Sps17** | **Tth** | **Tca** | **Dra** | **Tma** | **Tne** | **Taf** | **Bst** | **Bca** | |
| **Z05** | | 0.864 | 0.833 | 0.859 | 0.839 | 0.962 | 0.958 | 0.459 | 0.374 | 0.368 | 0.359 | 0.407 | 0.408 | |
| **Taq** | 0.864 | | 0.831 | 0.854 | 0.836 | 0.872 | 0.864 | 0.468 | 0.382 | 0.368 | 0.351 | 0.397 | 0.397 | |
| **Tfi** | 0.833 | 0.831 | | 0.82 | 0.991 | 0.829 | 0.824 | 0.45 | 0.371 | 0.375 | 0.353 | 0.405 | 0.397 | |
| **Tfl** | 0.859 | 0.854 | 0.82 | | 0.824 | 0.853 | 0.848 | 0.462 | 0.381 | 0.374 | 0.356 | 0.397 | 0.398 | |
| **Sps17** | 0.839 | 0.836 | 0.991 | 0.824 | | 0.835 | 0.83 | 0.452 | 0.375 | 0.377 | 0.355 | 0.407 | 0.399 | |
| **Tth** | 0.962 | 0.872 | 0.829 | 0.853 | 0.835 | | 0.989 | 0.463 | 0.373 | 0.367 | 0.358 | 0.406 | 0.406 | |
| **Tca** | 0.958 | 0.864 | 0.824 | 0.848 | 0.83 | 0.989 | | 0.46 | 0.371 | 0.365 | 0.356 | 0.404 | 0.404 | |
| **Dra** | 0.459 | 0.468 | 0.45 | 0.462 | 0.452 | 0.463 | 0.46 | | 0.334 | 0.325 | 0.314 | 0.338 | 0.339 | |
| **Tma** | 0.374 | 0.382 | 0.371 | 0.381 | 0.375 | 0.373 | 0.371 | 0.334 | | 0.854 | 0.567 | 0.37 | 0.377 | |
| **Tne** | 0.368 | 0.368 | 0.375 | 0.374 | 0.377 | 0.367 | 0.365 | 0.325 | 0.854 | | 0.558 | 0.377 | 0.376 | |
| **Taf** | 0.359 | 0.351 | 0.353 | 0.356 | 0.355 | 0.358 | 0.356 | 0.314 | 0.567 | 0.558 | | 0.356 | 0.364 | |
| **Bst** | 0.407 | 0.397 | 0.405 | 0.397 | 0.407 | 0.406 | 0.404 | 0.338 | 0.37 | 0.377 | 0.356 | | 0.881 | |
| **Bca** | 0.408 | 0.397 | 0.397 | 0.398 | 0.399 | 0.406 | 0.404 | 0.339 | 0.377 | 0.376 | 0.364 | 0.881 | | |
| | | | | | | | | | | | | | | |
| **B.** | **Sequence identities over polymerase sub domain only (corresponding to amino acids 420-834 of Z05)** | | | | | | | | | | | | | |
| **Name** | **Z05** | **Taq** | **Tfi** | **Tfl** | **Sps17** | **Tth** | **Tca** | **Dra** | **Tma** | **Tne** | **Taf** | **Bst** | **Bca** | |
| **Z05** | | 0.901 | 0.845 | 0.891 | 0.845 | 0.975 | 0.973 | 0.563 | 0.483 | 0.478 | 0.44 | 0.498 | 0.49 | |
| **Taq** | 0.901 | | 0.879 | 0.901 | 0.877 | 0.906 | 0.901 | 0.561 | 0.488 | 0.473 | 0.44 | 0.503 | 0.495 | |
| **Tfi** | 0.845 | 0.879 | | 0.857 | 0.997 | 0.853 | 0.853 | 0.566 | 0.495 | 0.49 | 0.449 | 0.512 | 0.49 | |
| **Tfl** | 0.891 | 0.901 | 0.857 | | 0.855 | 0.889 | 0.889 | 0.571 | 0.492 | 0.48 | 0.444 | 0.494 | 0.485 | |
| **Sps17** | 0.845 | 0.877 | 0.997 | 0.855 | | 0.853 | 0.853 | 0.566 | 0.495 | 0.49 | 0.449 | 0.512 | 0.49 | |
| **Tth** | 0.975 | 0.906 | 0.853 | 0.889 | 0.853 | | 0.99 | 0.563 | 0.478 | 0.473 | 0.437 | 0.496 | 0.488 | |
| **Tca** | 0.973 | 0.901 | 0.853 | 0.889 | 0.853 | 0.99 | | 0.563 | 0.478 | 0.473 | 0.437 | 0.496 | 0.488 | |
| **Dra** | 0.563 | 0.561 | 0.566 | 0.571 | 0.566 | 0.563 | 0.563 | | 0.45 | 0.448 | 0.426 | 0.474 | 0.454 | |
| **Tma** | 0.483 | 0.488 | 0.495 | 0.492 | 0.495 | 0.478 | 0.478 | 0.45 | | 0.883 | 0.622 | 0.474 | 0.475 | |
| **Tne** | 0.478 | 0.473 | 0.49 | 0.48 | 0.49 | 0.473 | 0.473 | 0.448 | 0.883 | | 0.615 | 0.476 | 0.473 | |
| **Taf** | 0.44 | 0.44 | 0.449 | 0.444 | 0.449 | 0.437 | 0.437 | 0.426 | 0.622 | 0.615 | | 0.46 | 0.473 | |
| **Bst** | 0.498 | 0.503 | 0.512 | 0.494 | 0.512 | 0.496 | 0.496 | 0.474 | 0.474 | 0.476 | 0.46 | | 0.898 | |
| **Bca** | 0.49 | 0.495 | 0.49 | 0.485 | 0.49 | 0.488 | 0.488 | 0.454 | 0.475 | 0.473 | 0.473 | 0.898 | | |

# *FIGURE 3*

| **A. Sequence identities over the entire polymerase I enzyme (corresponding to amino acids 1-834 of Z05)** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Name | **Z05** | **Tth** | **Tfi** | **Tfl** | **Tca** | **Taq** | **Sps17** |
| **Z05** | | 0.962 | 0.833 | 0.859 | 0.958 | 0.864 | 0.839 |
| **Tth** | 0.962 | | 0.829 | 0.853 | 0.989 | 0.872 | 0.835 |
| **Tfi** | 0.833 | 0.829 | | 0.82 | 0.824 | 0.831 | 0.991 |
| **Tfl** | 0.859 | 0.853 | 0.82 | | 0.848 | 0.854 | 0.824 |
| **Tca** | 0.958 | 0.989 | 0.824 | 0.848 | | 0.864 | 0.83 |
| **Taq** | 0.864 | 0.872 | 0.831 | 0.854 | 0.864 | | 0.836 |
| **Sps17** | 0.839 | 0.835 | 0.991 | 0.824 | 0.83 | 0.836 | |
| | | | | | | | |
| **B. Sequence identities over polymerase sub domain only (corresponding to amino acids 420-834 of Z05)** | | | | | | | |
| Name | **Z05** | **Tth** | **Tfi** | **Tfl** | **Tca** | **Taq** | **Sps17** |
| **Z05** | | 0.975 | 0.845 | 0.891 | 0.973 | 0.901 | 0.845 |
| **Tth** | 0.975 | | 0.853 | 0.889 | 0.99 | 0.906 | 0.853 |
| **Tfi** | 0.845 | 0.853 | | 0.857 | 0.853 | 0.879 | 0.997 |
| **Tfl** | 0.891 | 0.889 | 0.857 | | 0.889 | 0.901 | 0.855 |
| **Tca** | 0.973 | 0.99 | 0.853 | 0.889 | | 0.901 | 0.853 |
| **Taq** | 0.901 | 0.906 | 0.879 | 0.901 | 0.901 | | 0.877 |
| **Sps17** | 0.845 | 0.853 | 0.997 | 0.855 | 0.853 | 0.877 | |

# *FIGURE 1*

```
          *
Z05    AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:12)
Taq    AWIEKTLEEGRRRGYVETLFGRRRYVPDLEA (SEQ ID NO:13)
Tfi    AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA (SEQ ID NO:14)
Tfl    AWIEGTLEEGRRRGYVETLFGRRRYVPDLNA (SEQ ID NO:15)
Sps17  AWIAKTLEEGRKKGYVETLFGRRRYVPDLNA (SEQ ID NO:16)
Tth    AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:17)
Tca    AWIEKTLEEGRKRGYVETLFGRRRYVPDLNA (SEQ ID NO:18)
Tma    DYIQRVVSEAKEKGYVRTLFGRKRDIPQLMA (SEQ ID NO:19)
Tne    SYIQQVVAEAKEKGYVRTLFGRKRDIPQLMA (SEQ ID NO:20)
Taf    EYLKRMKDEARKKGYVTTLFGRRRYIPQLRS (SEQ ID NO:21)
Dra    RYINHTLDFGRTHGYVETLYGRRRYVPGLSS (SEQ ID NO:23)
Bst    QYMDNIVQEAKQKGYVTTLLHRRRYLPDITS (SEQ ID NO:24)
Bca    RYMENIVQEAKQKGYVTTLLHRRRYLPDITS (SEQ ID NO:25)
   X1X2X3X4X5X6X7X8X9X10X11X12X13GYVX14TL---------- (SEQ ID NO:26)
```