

X family DNA polymerases and Uses Thereof

Technical field

The present invention relates generally to the field of nucleic acid synthesis. In particular, the specification discloses tardigrade PolX family enzymes with terminal deoxynucleotidyl transferase activity and their use for nucleic acid synthesis.

Background

Conventional phosphoramidite-based DNA synthesis has been the dominant method for producing synthetic DNA, but this method suffers from several inherent limitations. For example, there is a practical constraint on the length of DNA that can be reliably produced using phosphoramidite chemistry, as overall yield and purity tend to decrease exponentially with increasing length of target sequences. The use of harsh reagents and conditions during synthesis can also introduce unwanted modifications, such as depurination or deamination, further compromising the fidelity of the final DNA construct.

In recent years, there have been considerable efforts to develop enzymatic template-independent DNA synthesis to overcome the efficiency and environmental issues associated with phosphoramidite synthesis, with terminal deoxynucleotidyltransferase (TdT) being the predominantly used enzyme. However, TdT suffers from a lack of step-wise control over nucleotide addition, and multiple nucleotides may be added in a random manner, thus necessitating the use of modified nucleotides to increase synthesis accuracy. There remains a need for improved solutions for the *de novo* synthesis of high-quality, long DNA sequences.

Accordingly, it is generally desirable to overcome or ameliorate one or more of the above-mentioned difficulties.

Summary

Disclosed herein is a polypeptide that is distinguished from a wild-type *Ramazzottius varieornatus* X family DNA polymerase (RvPolX) by at least one amino acid substitution in the wild-type RvPolX amino acid sequence, wherein the at least one amino acid substitution is at a position corresponding to positions 281, 513 and/or 522 as set forth in SEQ ID NO: 1.

Disclosed herein is a polynucleotide comprising a nucleic acid sequence encoding a polypeptide as defined herein.

Disclosed herein is an expression construct comprising the polynucleotide as defined herein.

Disclosed herein is a host cell comprising the polynucleotide as defined herein or the expression construct as defined herein.

Disclosed herein is a method of producing a polypeptide as defined herein, the method comprising a) culturing a host cell as defined herein under suitable conditions to allow expression of the polypeptide, and b) isolating the polypeptide.

Disclosed herein is a method of synthesizing a nucleic acid molecule comprising the step of contacting a nucleic acid primer with at least one nucleotide and a polypeptide as defined herein, under conditions for the addition of the at least one nucleotide to the nucleic acid primer by the polypeptide.

Disclosed herein is the use of a polypeptide as defined herein for synthesizing a nucleic acid molecule with a nucleic acid primer and at least one nucleotide.

Disclosed herein is a kit for synthesizing a nucleic acid molecule comprising a polypeptide as defined herein and at least one nucleotide.

Brief description of the drawings

Embodiments of the present invention are hereafter described, by way of non-limiting example only, with reference to the accompanying drawings in which:

Figure 1. Domain structure and template-independent DNA polymerase activity of RvPolX. (A) Domain structure of RvPolX, with the BRCA1 C-terminal (BRCT) domain in blue, and the low-complexity (LC) region in red. The catalytic domain consists of the 8 kDa domain (yellow), fingers domain (orange), palm domain (pink), and thumb domain (green). The loop1 region within the palm domain is colored in black. Denaturing urea-PAGE assay of RvPolX-catalyzed template-independent extension of a 14-mer single-stranded DNA (ssDNA14) under varying conditions, including various (B) pH buffers, (C) salt concentrations, (D) divalent metal ions, and (E) ssDNA substrates (the substrates ssDNA14A, ssDNA14T, ssDNA14G, and ssDNA14C consist of ssDNA extended by A, T, G or C, respectively).

Figure 2. Domain structure and 3D structures of PolX family members. (A) BRCA1 C-terminal (BRCT) domain in blue, and the low-complexity (LC) region in red. The catalytic domain consists of the 8 kDa domain (yellow), fingers domain (orange), palm domain (pink), and thumb domain (green). The loop1 region within the palm domain is colored in black. (B) A red arrow is used to indicate the loop1 region on the 3D structures. Alphafold structure of Pol mu is used in this figure as the loop1 region in its crystal structure is not resolved.

Figure 3. Denaturing urea-PAGE analysis of the RvPolX template-independent DNA polymerase assays with four different dNTPs and different divalent metal ions.

Figure 4. Contribution of the RvPolX N-terminal region to its activity and stability. (A) Denaturing urea-PAGE analysis of the DNA polymerase assays of WT and Δ 1-258aa performed at 25 °C, 40 °C, and 45 °C, in the presence or absence of 500 mM NaCl. Each assay was conducted in triplicate, and the band intensities of the extension products were quantified and plotted on the graph. Error bars reflect the standard deviations, and significance was determined by one-way analysis of variance (ANOVA) with Dunnett's posttest comparing mutants to WT (vehicle control). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ns, not significant. The far-UV CD spectra analysis of (B) WT and (C) Δ 1-258aa, was conducted at a range of temperatures (25°C to 50°C), in the presence of 500 mM NaCl.

Figure 5. Protein engineering of RvPolX. (A) Left: Twelve amino acid residues around the dNTP binding pocket were selected for saturation mutagenesis. Right: The binding mode between RvPolX containing both G513A and R522I mutations and dATP nucleotide. Hydrophobic residues at these two mutation sites (G513 and R522) generate a continuous hydrophobic patch (W514, A513, V521, I522, and M518) at the nucleotide binding pocket. The base of dATP is well accommodated in the hydrophobic patch with the hydrophilic ssDNA (5'-AGCCTG-3'). Denaturing urea-PAGE analysis shows the effect of RvPolX mutations on template-independent DNA polymerase activity at (B) 500 mM NaCl and (C) 1000 mM NaCl, with four different dNTPs. Each assay was conducted in triplicate, and the band intensities of the extension products were quantified and plotted on the graph. Error bars reflect the standard deviations, and the significance was determined by one-way analysis of variance (ANOVA) with Dunnett's posttest comparing mutants to WT (vehicle control). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ns, not significant.

Figure 6. Protein engineering of RvPolX. Denaturing urea-PAGE analysis of the effect on DNA polymerization of (A) selected single mutation mutants and (B) the combination of G513A and R522I/M/V mutants.

Figure 7. Kinetic analysis and microscale thermophoresis (MST) studies of WT and G513A+R522I. Denaturing urea-PAGE analysis shows time-dependent DNA polymerization catalyzed by (A) WT and (B) G513A+R522I with four different dNTPs. The labels on the products bands (+1, +2, +3, etc.) indicate ssDNA14 extended by a corresponding number of nucleotides. (C) Time-dependent changes in intensity for selected product bands. Each assay was conducted in triplicate, and the data are presented as mean values, with error bars reflecting the standard deviations, and the solid lines representing the exponential fits. (D) MST titration curves for the interactions between WT and G513A+R522I with four different dNTPs, used to determine the dissociation constants (K_d) for each dNTP. Each assay was conducted in triplicate, and the data are presented as mean values, with error bars reflecting the standard errors.

Figure 8. Denaturing urea-PAGE analysis of the G513A+R522I template-independent DNA polymerase assays with different divalent metal ions. Various concentrations of (A) single and (B) mixture divalent metal ions were tested.

Figure 9. Denaturing urea-PAGE analysis of the G513A+R522I and HsTdT template-independent DNA polymerase assays. Denaturing urea-PAGE analysis shows the effect of G513A+R522I on template-independent DNA polymerase activity at (A) 500 mM KCl and (B) 1000 mM KCl, with four different dNTPs. Each assay was conducted in triplicate, and the band intensities of the extension products were quantified and plotted on the graph. Error bars reflect the standard deviations. Statistical differences were based on Student's t test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ns, not significant.

Figure 10. Denaturing urea-PAGE analysis of template-independent DNA polymerase assays using 3'-ONH₂ modified nucleotides with G513A+R522I and commercial calf thymus TdT. (A) Denaturing urea-PAGE analysis demonstrates the effect of 3'-ONH₂ modified nucleotides on template-independent DNA polymerase activity between G513A+R522I and commercial calf thymus TdT. The assay was conducted in triplicate, and percentage of band intensity corresponding to the one-base extended product (+1) was quantified and plotted on the graph. (B) Denaturing urea-PAGE analysis of the investigation into the exonuclease activity of 1 μ M G513A+R522I and 1 μ M commercial calf thymus TdT against 50 nM ssDNA14 after 40 minutes of incubation in the absence of nucleotides at 25°C. The assay was conducted in triplicate, and the percentage of band intensity of the degraded bands was quantified and plotted on the graph. Error bars represent standard deviations. Statistical differences were determined using Student's t test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ns, not significant.

Figure 11. WebLogo presentation of the GW motif and the R522 residue within the thumb domain. Six distinct groups were generated from the 230 PolX family proteins, including Pol β , Pol β -like proteins, Pol λ , Pol λ -like proteins, TdT/Pol μ , and TdT/Pol μ -like proteins. Multiple alignments of each group were analyzed using WebLogo (<http://weblogo.berkeley.edu/logo.cgi>). Both the GW motif and the R522 residue within the thumb domain are boxed in black and red, respectively.

Detailed description

The present specification teaches a polypeptide comprising a wild-type *Ramazzottius varieornatus* X family DNA polymerase (RvPolX) enzyme, and variants thereof. The polypeptide may comprise an amino acid sequence having at least 70% sequence identity to an amino acid sequence in SEQ ID NO: 1 or 3, or a catalytic fragment thereof.

Disclosed herein is a polypeptide that is distinguished from a wild-type RvPolX by at least one amino acid substitution in the wild-type RvPolX amino acid sequence, wherein the at least one amino acid substitution is at a position corresponding to positions 281, 513 and/or 522 as set forth in SEQ ID NO: 1.

Without being bound by theory, the inventors have discovered and characterized a template-independent DNA polymerase enzyme, named RvPolX herein, from the extremotolerant tardigrade species *R. varieornatus*. It is found that RvPolX exhibits a distinctive ability for template-independent DNA polymerization, retaining its efficacy even in the presence of elevated salt concentrations, surpassing that of the extensively researched TdT enzyme. The inventors have further identified conditions under which RvPolX is able to limit the number of nucleotide extension per cycle to one, which can potentially eliminate the need for iterative protection and deprotection steps during DNA synthesis. Furthermore, the inventors have identified mutations that improve the activity and the nucleotide promiscuity of the enzyme. It was found that RvPolX can accept 3'-modified reversible terminator nucleotides, which are common modified nucleotides used in enzymatic DNA synthesis. Finally, combinatorial mutants were engineered which exhibit improved DNA synthesis activity using either natural nucleotides or 3'-modified reversible terminator nucleotides as substrates.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues is a synthetic non-naturally-occurring amino acid, such as a chemical analogue

of a corresponding naturally-occurring amino acid, as well as to naturally-occurring amino acid polymers. These terms do not exclude modifications, for example, glycosylations, acetylations, phosphorylations and the like. Soluble forms of the subject proteinaceous molecules are particularly useful. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid including, for example, unnatural amino acids or polypeptides with substituted linkages.

The terms “wild-type” or “parent” are used interchangeably herein and refer to the non-mutated version of a polypeptide as it appears naturally. The terms “mutant”, “variant”, “engineered polypeptide” and “engineered protein” are used interchangeably herein to refer to a polypeptide derived from a wild-type protein and comprising one or more amino acid modifications, e.g., an amino acid substitution, insertion and/or deletion. The variants may be obtained by various techniques well known in the art, e.g., site-directed mutagenesis, random mutagenesis or directed evolution.

The term “substitution” as used herein refers to the presence of an amino acid residue at a certain position of the derivative sequence which is different from the amino acid residue which is present or absent at the corresponding position in the reference sequence. Preferably, the term “substitution” refers to the replacement of an amino acid residue by another selected from the naturally-occurring standard 20 amino acid residues (G, P, A, V, L, I, M, C, F, Y, W, H, K, R, Q, N, E, D, S and T). The sign “+” herein indicates a combination of substitutions. The following terminology is used herein to designate a substitution: G513A denotes that the amino acid residue at position 513 (glycine, G) of the wild-type sequence is changed to an alanine (A). R522I/M/V denotes that the amino acid residue at position 522 (arginine, R) of the wild-type sequence is substituted with either an isoleucine (I), methionine (M) or valine (V).

A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art, which can be generally sub-classified as follows:

Table A. Amino acid sub-classification

Sub-classes	Amino acids
Acidic	Aspartic acid, Glutamic acid
Basic	Noncyclic: Arginine, Lysine; Cyclic: Histidine
Charged	Aspartic acid, Glutamic acid, Arginine, Lysine, Histidine
Small	Glycine, Serine, Alanine, Threonine, Proline
Polar/neutral	Asparagine, Histidine, Glutamine, Cysteine, Serine, Threonine
Polar/large	Asparagine, Glutamine
Hydrophobic	Tyrosine, Valine, Isoleucine, Leucine, Methionine, Phenylalanine, Tryptophan
Aromatic	Tryptophan, Tyrosine, Phenylalanine
Residues that influence chain orientation	Glycine and Proline

Conservative amino acid substitution also includes groupings based on side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. For example, it is reasonable to expect that replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid will not have a major effect on the properties of the resulting variant polypeptide. Whether an amino acid change results in a functional polypeptide can readily be determined by assaying its activity. Conservative substitutions are shown in the table below under the heading of exemplary substitutions. Amino acid substitutions falling within the scope of the invention, are, in general, accomplished by selecting substitutions that do not differ significantly in their effect on maintaining (a) the structure of the peptide backbone in the area of the substitution, (b) the charge or

hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. After the substitutions are introduced, the variants are screened for biological activity.

Table B. Amino acid substitutions

Original Residue	Exemplary Substitutions
Ala	Val, Leu, Ile
Arg	Lys, Gln, Asn
Asn	Gln, His, Lys, Arg
Asp	Glu
Cys	Ser
Gln	Asn, His, Lys,
Glu	Asp, Lys
Gly	Pro
His	Asn, Gln, Lys, Arg
Ile	Lcu, Val, Met, Ala, Phe
Lcu	Ile, Val, Met, Ala, Phe
Lys	Arg, Gln, Asn
Met	Lcu, Ile, Phe
Phe	Lcu, Val, Ile, Ala
Pro	Gly
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe, Thr, Ser
Val	Ile, Leu, Met, Phe, Ala

As used herein, a “position corresponding to” or recitation that amino acid positions “correspond to” amino acid positions in a reference sequence, such as set forth in the sequence listing, refers to amino acid positions identified upon alignment with the disclosed sequence to maximise identity using a standard alignment algorithm or software (such as the BLAST, ClustalW, ClustalOmega, MUSCLE, TCOffee or ProbCons software). By aligning the sequences, one skilled in the art can identify corresponding residues.

The term “sequence identity” as used herein refers to the extent that sequences are identical on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over

a window of comparison. Thus, a “percentage of sequence identity” is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G and I) or the identical amino acid residue (e.g., Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) 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 (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity.

The term “at least 70%” as used herein includes at least 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more.

The term “polynucleotide” or “nucleic acid” are used interchangeably herein to refer to a polymer of nucleotides, which can be mRNA, RNA, cRNA, cDNA or DNA. The term typically refers to polymeric form of nucleotides of at least 10 bases in length, either ribonucleotides or deoxynucleotides or a modified form of either type of nucleotide. The term includes single and double stranded forms of DNA.

The term “encoding” or “encodes” refers to the inherent property of specific sequences of nucleotides in a polynucleotide, such as a gene, a cDNA, or an mRNA, to serve as templates for synthesis of other polymers and macromolecules in biological processes having either a defined sequence of nucleotides (i.e. rRNA, tRNA and mRNA) or a defined sequence of amino acids and the biological properties resulting therefrom. Thus, a gene encodes a protein if transcription of a gene and translation of mRNA corresponding to that gene produces the protein in a cell or other biological system. Both the coding strand and the non-coding strand, used as the template for transcription of a gene or cDNA, can be referred to as encoding the protein or other product of that gene or cDNA.

An “expression construct” generally includes at least a control sequence operably linked to a nucleotide sequence of interest. In this manner, for example, promoters in operable connection with the nucleotide sequences to be expressed are provided in expression

constructs for expression in an organism or part thereof including a host cell. For the practice of the present invention, conventional compositions and methods for preparing and using constructs and host cells are well known to one skilled in the art, see for example, *Molecular Cloning: A Laboratory Manual*, 3rd edition Volumes 1, 2, and 3. J. F. Sambrook, D. W. Russell, and N. Irwin, Cold Spring Harbor Laboratory Press, 2000.

By “expression vector” or “vector” is meant a nucleic acid molecule, preferably a DNA molecule derived, for example, from a plasmid, bacteriophage, or plant virus, into which a nucleic acid sequence may be inserted or cloned. A vector preferably contains one or more unique restriction sites and may be capable of autonomous replication in a defined host cell including a target cell or tissue or a progenitor cell or tissue thereof, or be integrable with the genome of the defined host such that the cloned sequence is reproducible. Accordingly, the vector may be an autonomously replicating vector, i.e., a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a linear or closed circular plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. A vector system may comprise a single vector or plasmid, two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vector may also include a selection marker such as an antibiotic resistance gene that can be used for selection of suitable transformants. Examples of such resistance genes are well known to those of skill in the art.

A “host cell” herein refers to any cell useful for the production of a polypeptide or polypeptide variant disclosed herein, and may be a prokaryotic or eukaryotic cell. The prokaryotic host cell may be any Gram-positive or Gram-negative bacterium. The host cell may also be a eukaryotic cell, such as a yeast, fungal, mammalian, insect or plant cell.

RvPolX and engineered variants

The present disclosure provides a *Ramazzottius varieornatus* X family DNA polymerase (RvPolX) and RvPolX variants which can be used to synthesise polynucleotides of predetermined sequences, such as DNA or RNA, without the use of a template strand. The polymerases are surprisingly stable at salt concentrations beyond the tolerance of many commercially available DNA polymerases, and may be useful for various applications in which high salt concentration is an advantage. For example, the polymerases are useful for nanopore sequencing where a high salt concentration may be used to boost the signal-to-noise ratio for ionic current-based measurements. The polymerases of this disclosure also allow modified nucleotides, in particular 3'-O-modified nucleotides, to be used in enzyme-mediated nucleic acid synthesis.

In one embodiment, provided herein is a polypeptide comprising an amino acid sequence having at least 70% sequence identity (such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity) to an amino acid sequence of SEQ ID NO: 1, or a fragment thereof.

In one embodiment, provided herein is a polypeptide that is distinguished from a wild-type RvPolX by at least one amino acid substitution in the wild-type RvPolX amino acid sequence, wherein the at least one amino acid substitution is at a position corresponding to positions 281, 513 and/or 522 as set forth in SEQ ID NO: 1, wherein the polypeptide comprises or consists of an amino acid sequence having at least 70% sequence identity (such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity) to an amino acid sequence as set forth in SEQ ID NO: 1, or a fragment thereof.

The fragment may be a catalytic fragment that is capable of template-independent DNA polymerisation. Thus, in some embodiments, the polypeptide as defined herein comprises or consists of an amino acid sequence having at least 70% sequence identity

(such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity) to the catalytic domain of wild-type RvPolX enzyme (i.e., SEQ ID NO: 3).

In one embodiment, provided herein is a polypeptide that is distinguished from a wild-type RvPolX by at least one amino acid substitution in the wild-type RvPolX amino acid sequence, wherein the at least one amino acid substitution is at a position corresponding to positions 281, 513 and/or 522 as set forth in SEQ ID NO: 1 (i.e. at a position corresponding to positions 23, 255 and/or 264 as set forth in SEQ ID NO: 3), wherein the polypeptide comprises or consists of an amino acid sequence having at least 70% (such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity) sequence identity to SEQ ID NO: 3 or a fragment thereof.

In some embodiments, the polypeptide comprises at least two amino acid substitutions at positions corresponding to positions 281, 513 and/or 522 as set forth in SEQ ID NO: 1.

In some embodiments, the polypeptide comprises three amino acid substitutions at positions corresponding to positions 281, 513 and 522 as set forth in SEQ ID NO: 1.

In one embodiment, the amino acid substitution at the position corresponding to position 281, 513 and/or 522 is a substitution to a hydrophobic amino acid residue, i.e., a substitution to A, C, F, I, L, M, V, W or Y. The inventors have found that, advantageously, a hydrophobic amino acid at one or more of these positions can improve the catalytic activity of the polypeptide.

The amino acid substitution at the position corresponding to position 281 may be a substitution to A, F, I, L, V or Y. In one embodiment, the amino acid substitution corresponding to position 281 is a substitution to L.

The amino acid substitution at the position corresponding to position 513 may be a substitution to A or C.

The amino acid substitution at the position corresponding to position 522 may be a substitution to I, M or V.

In some embodiments, the polypeptide comprises at least two of the following:

- a) a substitution to A, F, I, L, V or Y at a position corresponding to position 281 of SEQ ID NO: 1;
- b) a substitution to A or C at a position corresponding to position 513 of SEQ ID NO: 1; and/or
- c) a substitution to I, M or V at a position corresponding to position 522 of SEQ ID NO: 1.

In some embodiments, the polypeptide comprises:

- a) a substitution to A, F, I, L, V or Y at a position corresponding to position 281 of SEQ ID NO: 1;
- b) a substitution to A or C at a position corresponding to position 513 of SEQ ID NO: 1; and
- c) a substitution to I, M or V at a position corresponding to position 522 of SEQ ID NO: 1.

In one embodiment, the polypeptide is distinguished from the wild-type RvPolX by at least two amino acid substitutions in the wild-type RvPolX amino acid sequence. The polypeptide may comprise an amino acid substitution of i) G513A or G513C, and ii) R522I, R522M or R522V.

In one embodiment, the polypeptide is distinguished from the wild-type RvPolX by three amino acid substitutions at positions 281, 513 and 522. The polypeptide may comprise an amino acid substitution of i) E281L, ii) G513A or G513C, and iii) R522I, R522M or R522V.

In some embodiments, the polypeptide comprises an amino acid sequence having at least 70% sequence identity (such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity) to an amino acid sequence in SEQ ID NO: 3–26. In some embodiments, the polypeptide comprises or consists of an amino acid sequence in SEQ ID NO: 3–26.

In some embodiments, the polypeptide comprises an amino acid sequence having at least 70% sequence identity (such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity) to an amino acid sequence in SEQ ID NO: 15–20. In some embodiments, the polypeptide comprises or consists of an amino acid sequence in SEQ ID NO: 15–20.

Amino acid sequence of wild-type RvPolX (Uniprot ID A0A1D1UV65) (positions 281, 513 and 522 underlined)

MYTQRKLRTSKRIAPAEPPKKKEQAEDTDEDATLNHSPKKKPREVPSKKTFDKS
QPDDKDIFNGKTVYFVPVEIPQARRNLLARRVEEFGGKVADELHLGGKTPLD
YIVVDDRVALDTICRAMDNNFHDNFEEKMLQLLDKASVVRAQWLSQCLLERKI
VDDSEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDESYHSDDDRK
SYVSNFSDLGDMEDLHDVPGSQGDVDAAKLQKSNFFLSPAGGSRSGTNPNNAH
IIEQLEEMMKTYRNMGDEFRGYAYARASAIRKFPDPILGMEQIKQLKNVGPRI
ASHIVEIVETGHMEKKDGAESNEATKAMNLFVNMHGVGKTAQKWVQMGY
RTLDDLKEKVKLTPEQAVGLKYYDEFQEKMTREVEEIVRIVEDTVDSIRSGFI
FKVCGSYLRGREMCGDVDLLMTHSDGHSHTILKPLVGKLHEMGFLTDDLW
LTEGRHKYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWSGSMYFVRSIR
DYAHKKGMSLTDYSLNINVVRVNGEKNVNEGIPLETPTEESIFEALGLPYLRPEE
RDH (SEQ ID NO: 1)

Nucleic acid sequence of wild-type RvPolX

ATGTATACACAAAGAAAACAAAGGACTTCAAAGCGCATCGCCCCCGCTGA
GCCGAAAAAGAAAGAGCAGGCGGAAGACACGGACGAGGATGCTACCCTG
AACCCTCGCCGAAGAAGAAGCCTCGTGAAGTGCCAAGCAAGAAGACCTT

TGATAAGAGCCAACCGGATGACAAGGACATTTTTAACGGTAAGACCGTTT
ATTTTGTCCCGGTGGAAATCCCGCAAGCGCGTCGTAATCTGCTGGCACGCC
GTGTAGAGGAGTTCGGTGGCAAAGTTGCGGATGAACTGCACCTGGGCGGC
AAAACGCCGTTGGACTATATCGTTGTAGATGACCGTGTGCGCTGGACACC
ATATGTCGTGCCATGGATAATAACTTCCACGACAACCTCGAGAAAATGCTG
CAGTTGCTAGACAAGGCTTCGGTTGTGCGTGCGCAATGGCTGTCTCAATGT
CTGTTAGAGCGTAAAATCGTGGACGACTCAGAAGCGAACATCAACGAGTT
CTTGAAGAATCACATTTCGTGCTATGGCAAATAAAGCTGATCAGCCGAGCC
GTTCCCCGCTTCCGTTTTCCCCGCTGCGTGACTCCAGCCCCGAGCCGTGATG
AGTCGTATCATTCCGATGACGATCGTAAGAGCTACGTTTCCAACTTTAGCG
ACCTTGCGGACATGGAAGACCTGCACGATGTGCCGGGTCTCAAGGTGAC
GTGGATGCCGCGAAACTGCAAAGTCCAACCTCTTCCTGAGCCCAGCGGG
TGGCTCACGTAGCGGCACCAATCCGAATGCGCACATCATTGAACAGCTTG
AGGAGATGATGAAAACCTATCGCAACATGGGTGACGAATTCCGTGGTTAC
GCGTATGCGCGAGCGATTAGCGCAATTTCGCAAATTTCCGGATCCGATTCTG
GGCATGGAACAAATTAAGCAGTTGAAAAACGTCGGTCCGCGTATTGCGAG
CCACATCGTCGAGATTGTGGAGACAGGTCACATGGAAAAAAGGATGGTG
CAGAGTCTAACGAAGCCACGAAGGCAATGAATCTGTTCTGTAACATGCAT
GGTGTGGCACGAAAACCGCTCAGAAATGGGTTTCAGATGGGCTACCGCAC
CCTGGATGACTTGAAAGAAAAGGTGAAATTGACCCCGAACAGGCGGTTG
GTCTGAAATACTATGATGAGTTTCAGGAGAAGATGACCCGTAGGGAGGTG
GAGGAGATCGTGCGCATCGTCGAGGACACCGTTGATAGCATCCGCAGCGG
CTTTATCTTCAAGGTGTGCGGTAGCTACCTCCGCGGTCTGTGAGATGTGCGG
TGATGTTGACCTGCTGATGACCCATTCCGACGGCCACAGCCATAATACCAT
CCTGAAACCGTTGGTTGGAAAGTTGCACGAAATGGGGTTTCTGACCGACG
ACCTGTGGCTGACTGAGGGTCGTCATAAAAAAGTACTTCGGCGTGTGCAAAT
TGTTTAAAGAAGGCAGCAAACATCGTCGGCTGGACATCATTATCCCACCGT
ATAACGAATGGGCAACGTGCGTTCTGGGCTGGAGTGGCTCTATGTACTTCG
TGCGCTCCATCCGTGATTACGCGCACAAGAAGGGTATGAGCCTCACTGACT
ACAGCCTGAATATCAACGTAGTTTCGCGTCAACGGTGAAAAAGTTAACGAA
GGTATTCCGTTAGAAACCCCGACTGAGGAAAGCATTTCGAGGCGCTGGG
TCTGCCGTACTTACGTCCGGAAGAGCGCGATCATTA

(SEQ ID NO: 2)

Amino acid sequence of wild-type RvPolX catalytic domain

NPNAHIIIEQLEEMMKTYRNMGDEFGRGYAYARAISAIRKFDPILGMEQIKQLK
NVGPRIASHIVEIVETGHMEKKDGAESNEATKAMNLFVNMHGVGKTAKQW
VQMGYRTLDDLKEKVKLTPQAVGLKYYDEFQEKMTREVVEIVRIVEDTV
SIRSGFIFKVCGSYLRGREMCGDVDLLMTHSDGHSHTILKPLVGKLHEMGFL
TDDLWLTEGRHKKYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWGS
SMYFVRSIRDYAHKKGMSLTDYSLNINVVRVNGEKNVNEGIPTPTESIFEALGLP
YLRPEERDH (SEQ ID NO: 3)

Table 1. Exemplary RvPolX variants

SEQ ID NO	Mutations	Amino acid sequence (mutation positions underlined, mutated residues in bold)
4	E281A	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPN <u>AHII</u> EQLEEMMKTYRNM <u>GDA</u> FRGYAYARAISAIRKFDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGKTAKQK WVQMGYRTLDDLKEKVKLTPQAVGLKYYDEFQEKMTR REVVEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWGS SMYFV <u>R</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNVNEGIPL EPTPTESIFEALGLPYLRPEERDH
5	E281F	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPN <u>AHII</u> EQLEEMMKTYRNM <u>GDF</u>

		FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWWSGSMY FVRSIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
6	E281I	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGDIF RGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWWSGSMY FVRSIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
7	E281L	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWWSGSMY

		FVRSIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
8	E281V	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIIEQLEEMMKTYRNMGDY FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWSGSMY FVRSIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
9	E281Y	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIIEQLEEMMKTYRNMGDY FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWSGSMY FVRSIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
10	G513A	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD

		SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGDE FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTCTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLAWSGSMY FVRSIRDYAHKKGMSLTDYSLNINVVRVNGEKNVNEGIPLE TPTEESIFEALGLPYLRPEERDH
11	G513C	MYTQRKLRTSKRIAPAEPPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNF EKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGDE FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTCTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLCWSGSMY FVRSIRDYAHKKGMSLTDYSLNINVVRVNGEKNVNEGIPLE TPTEESIFEALGLPYLRPEERDH
12	R522I	MYTQRKLRTSKRIAPAEPPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNF EKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGDE FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTCTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR

		REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWGS MYFV <u>I</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKN NEGIPLETPTTEESIFEALGLPYLRPEERDH
13	R522M	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>E</u> FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWGS SMYFV <u>M</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKN NEGIPLETPTTEESIFEALGLPYLRPEERDH
14	R522V	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>E</u> FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVLGWGS SMYFV <u>S</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKN NEGIPLETPTTEESIFEALGLPYLRPEERDH

15	G513A, R522I	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNF EKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>E</u> FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>A</u> WSGSMY FV <u>I</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNVEGIPIET PTEESIFEALGLPYLRPEERDH
16	G513C, R522I	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNF EKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>E</u> FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>C</u> WSGSMY FV <u>I</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNVEGIPIET PTEESIFEALGLPYLRPEERDH
17	G513A, R522M	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNF EKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL

		QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGDE FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>A</u> WSGSMY FV <u>M</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKVNEGIPLE TPTEESIFEALGLPYLRPEERDH
18	G513C, R522M	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNF EKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGDE FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>C</u> WSGSMY FV <u>M</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKVNEGIPLE TPTEESIFEALGLPYLRPEERDH
19	G513A, R522V	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNF EKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGDE FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK

		KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>A</u> WSGSMY FV <u>V</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
20	G513C, R522V	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>E</u> FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>C</u> WSGSMY FV <u>V</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
21	E281L, G513A, R522I	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>L</u> FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQAVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>A</u> WSGSMY FV <u>I</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLET PTEESIFEALGLPYLRPEERDH
22	E281L, G513C, R522I	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSQPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA

		MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>L</u> FRGYAYARASAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>C</u> WSGSMY FV <u>I</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLET PTEESIFEALGLPYLRPEERDH
23	E281L, G513A, R522M	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKV ADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>L</u> FRGYAYARASAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGREMCGDVDL LMTHSDGHSHTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>A</u> WSGSMY FV <u>M</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
24	E281L, G513C, R522M	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKV ADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>L</u> FRGYAYARASAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK

		WVQMGYRTLDDLKEKVKLTPAQVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>C</u> WSGSMY FV <u>M</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
25	E281L, G513A, R522V	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>L</u> FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>A</u> WSGSMY FV <u>V</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH
26	E281L, G513C, R522V	MYTQRKLRTSKRIAPAEPKKKEQAEDTDEDATLNHSPKKK PREVPSKKTFDKSPDDKDIFNGKTVYFVPVEIPQARRNLL ARRVEEFGGKVADELHLGGKTPLDYIVVDDRVALDTICRA MDNNFHDNFEKMLQLLDKASVVRAQWLSQCLLERKIVDD SEANINEFLKNHIRAMANKADQPSRSPLPFSPLRDSSPSRDE SYHSDDDRKSYVSNFSDLGDMEDLHDVPGSQGDVDAAKL QKSNFFLSPAGGSRSGTNPNAHIEQLEEMMKTYRNMGD <u>L</u> FRGYAYARAISAIRKFPDPILGMEQIKQLKNVGPRIASHIVEI VETGHMEKKDGAESNEATKAMNLFVNMHGVGTKTAQK WVQMGYRTLDDLKEKVKLTPAQVGLKYYDEFQEKMTR REVEEIVRIVEDTVDSIRSGFIFKVCGSYLRGEMCGDVDL LMTHSDGHSNTILKPLVGKLHEMGFLTDDLWLTEGRHK KYFGVCKLFKEGSKHRRLDIIIPPYNEWATCVL <u>C</u> WSGSMY FV <u>V</u> SIRDYAHKKGMSLTDYSLNINVVRVNGEKNNEGIPLE TPTEESIFEALGLPYLRPEERDH

Polypeptides herein may further comprise a fusion partner (such as a fusion peptide or domain) at its N-terminus, C-terminus or both. In preferred embodiments, the fusion partner is recombinantly attached to the polypeptide and is expressed together with the polypeptide. The fusion partner may be used for purification, identification, increasing expression, secretability or increasing catalytic activity. For instance, a hexahistidine tag may be added to enable affinity chromatography during protein purification, or a solubility tag may be added to enhance protein solubility during expression. Other such fusion partners are extensively described in the literature and thus all fusion partners known to a skilled person are contemplated in the present disclosure.

The RvPolX variant polypeptides of this disclosure are generally capable of both template-dependent and template-independent nucleic acid synthesis. Template-dependent nucleic acid synthesis refers to the process in which the sequence of a newly synthesized nucleic acid molecule is determined by the sequence of an existing nucleic acid template. In template-independent nucleic acid synthesis, nucleic acids are synthesized without relying on a template strand to dictate the sequence.

Template-dependent and template-independent polymerisation activity may be measured using methods well known in the art. Exemplary methods are provided in the examples of this disclosure.

In embodiments, RvPolX variant polypeptides herein exhibit measurable polymerase activity at salt concentrations of 100 mM and above, such as at a salt concentration of about 100 mM, about 200 mM, about 300 mM, about 400 mM, about 500 mM, about 600 mM, about 700 mM, about 800 mM, about 900 mM, about 1000 mM, about 1100 mM, about 1200 mM, about 1300 mM, about 1400 mM, or about 1500 mM. In some embodiments, RvPolX variant polypeptides herein exhibit measurable polymerase activity at least in a range of salt concentrations from about 400 mM to about 1500 mM. The salt may be any salt, such as NaCl or KCl.

In embodiments, RvPolX variant polypeptides herein exhibit measurable polymerase activity at least in a range of pH from about pH 6 to about pH 10, such as at a pH of

about 6, about 6.5, about 7, about 7.5, about 8, about 8.5, about 9, about 9.5, or about 10.

In embodiments, RvPolX variant polypeptides herein exhibit measurable polymerase activity at least in a range of temperatures from about 20°C to about 40°C, such as at a temperature of about 20°C, about 25°C, about 30°C, about 35°C, or about 40°C.

In embodiments, RvPolX variant polypeptides herein are capable of incorporating nucleotides (including ribonucleotides and deoxyribonucleotides) and modified nucleotides, including 3'-O-modified nucleotides such as 3'-O-NH₂ nucleotides.

In the context of the present disclosure, a "3'-O-modified nucleotide" refers to a ribonucleotide (rNTP) or deoxyribonucleotide (dNTP) containing a blocking moiety at the 3' position of the sugar that prevents the formation of a phosphodiester bond. The blocking moiety may be a chemical group which can be removed through a specific cleaving reaction to allow further nucleotide addition, and such a nucleotide is also referred to as a "reversible terminator nucleotide". Reversible blocking moieties include but are not limited to acetyl, azido, azidomethyl, amino, aminoxy, allyl, cyanoethyl, dimethoxytrityl, 2-nitrobenzyl, N-oxime, propargyl, tert-butoxy ethoxy, tert-butyl dimethyl silyl, or trimethyl(silyl)ethoxymethyl.

The RvPolX polypeptides can be provided in various forms, such as an isolated preparation, as a substantially purified enzyme, or as cell extracts or lysates of such cells. The enzymes can be lyophilized, spray-dried, or be in the form of a concentrate or crude paste.

Nucleic acids, vectors, host cells

Disclosed herein is a polynucleotide comprising a nucleic acid sequence encoding a polypeptide as defined herein.

Disclosed herein is an expression construct comprising the polynucleotide as defined herein. The gene encoding the polypeptide in the expression construct may be operably

linked to a promoter (such as a constitutive or inducible promoter) for polypeptide expression. Suitable promoters may be selected based on the expression host.

Codons in the polynucleotide or expression construct may be selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria may be used to express the gene in bacteria; preferred codons used in yeast may be used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. Not all codons need to be replaced to optimize the codon usage in the gene since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon-optimized polynucleotides encoding the RvPolX polypeptides may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full-length coding region.

Disclosed herein is a vector comprising the polynucleotide or expression construct as defined herein. The vector may be any vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and can result in the expression of the RvPolX polynucleotide sequence in a suitable host cell. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

Disclosed herein is a host cell comprising the polynucleotide, expression construct or vector as defined herein.

The host cell may be transformed, transfected or transduced with the polynucleotide, expression construct or vector in a transient or stable manner. The expression construct or vector disclosed herein may be maintained in the host cell as a chromosomal integrant or as a self-replicating extra-chromosomal vector. The term "host cell" also encompasses any progeny of a parent host cell that is not identical to the parent host cell due to mutations that occur during replication. The host cell may be any cell useful in the production of a polypeptide or polypeptide variant disclosed herein, e.g., a prokaryote or a eukaryote. The prokaryotic host cell may be any Gram-positive or

Gram-negative bacterium. The host cell may also be a eukaryotic cell, such as a yeast, fungal, mammalian, insect or plant cell.

In some embodiments the host cell is selected from the group consisting of Gram-positive bacteria, Gram-negative bacteria, filamentous fungi, yeast and algae. Non-limiting examples of host cells include *Escherichia coli*, *Lactococcus lactis*, *Bacillus sp.*, and *Pichia sp.*

Protein production

Disclosed herein is a method of producing a polypeptide as defined herein, the method comprising a) culturing a host cell as defined herein under suitable conditions to allow expression of the polypeptide, and b) isolating the polypeptide.

Those skilled in the field of molecular biology will understand that any of a wide variety of expression systems may be used to provide the polypeptides of this disclosure. An isolated polypeptide of the invention may be produced in a prokaryotic host (e.g., *E. coli*) or in a eukaryotic host (e.g., *Saccharomyces cerevisiae*, insect cells, e.g., Sf21 cells, or mammalian cells, e.g., NIH 3T3, HeLa, COS cells). Such cells are available from a wide range of sources (e.g., the American Type Culture Collection, Rockland, MD). Non-limiting examples of insect cells are, *Spodoptera frugiperda* (Sf) cells, e.g., Sf9, Sf21, *Trichoplusia ni* cells, e.g., High Five cells, and *Drosophila* S2 cells. Examples of fungi (including yeast) host cells are *S. cerevisiae*, *Kluyveromyces lactis* (*K. lactis*), species of *Candida* including *C. albicans* and *C. glabrata*, *Aspergillus nidulans*, *Schizosaccharomyces pombe* (*S. pombe*), *Komagataella pastoris* (previously known as *Pichia pastoris*), and *Yarrowia lipolytica*. Examples of mammalian cells are COS cells, baby hamster kidney cells, mouse L cells, LNCaP cells, Chinese hamster ovary (CHO) cells, human embryonic kidney (HEK) cells, African green monkey cells, CV1 cells, HeLa cells, MDCK cells, Vero and Hep-2 cells. *Xenopus laevis* oocytes, or other cells of amphibian origin, may also be used. Prokaryotic host cells include bacterial cells, for example, *E. coli*, *B. subtilis*, and mycobacteria.

Depending on the expression system and host cell selected, the polypeptides are produced by growing host cells transformed by a polynucleotide or an expression vector under conditions whereby the recombinant polypeptides are expressed. The selection of appropriate culture media and growth conditions is within the capacity of one of ordinary skill in the art. Polypeptides herein can be produced in the cells and/or culture supernatant, and both may be used to recover the polypeptide.

Methods to grow cells that produce the polypeptides of the invention include, but are not limited to, batch, batch-fed, continuous and perfusion cell culture techniques. Typically, cell culture is performed under sterile, controlled temperature and atmospheric conditions. A bioreactor is a chamber used to culture cells in which environmental conditions such as temperature, atmosphere, agitation and/or pH can be monitored. The bioreactor can be a stainless steel chamber or a pre-sterilised plastic bag (e.g., Cellbag.RTM., Wave Biotech, Bridgewater, N.J.). The bioreactor may be dimensioned for cultures of about 20 L to about 50,000 L.

In some embodiments, the method further comprises recovering the polypeptide from the cell culture. Polypeptides may be isolated from the cell and/or media fraction using methods that preserve their integrity, such as by gradient centrifugation, e.g., caesium chloride, sucrose and iodixanol, as well as standard purification techniques including, e.g., ion exchange, gel filtration or affinity chromatography. Conditions for purifying a particular enzyme will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, presence of a tag, etc., and will be apparent to those having skill in the art.

Kits

In some embodiments, the RvPolX polypeptides described herein are provided in the form of kits.

Disclosed herein is a kit for synthesizing a nucleic acid molecule, comprising a RvPolX polypeptide and at least one nucleotide, wherein the RvPolX polypeptide i) comprises

an amino acid sequence having at least 70% sequence identity to an amino acid sequence of SEQ ID NO: 1, or ii) is a RvPolX variant polypeptide as defined herein.

Disclosed herein is a kit for synthesizing a nucleic acid molecule, comprising a RvPolX variant polypeptide as defined herein and at least one nucleotide.

The nucleotide may be a ribonucleotide (rNTP), deoxyribonucleotide (dNTP) or a 3'-O-modified rNTP or dNTP. The 3' modification may be a 3'-O-acetyl, 3'-O-amino, 3'-aminoxy, 3'-O-methyl, 3'-O-azido, 3'-O-azidomethyl, 3'-O-cyanoethyl, 3'-O-allyl, 3'-O-(N-oxime), 3'-O-acetyl, 3'-O-propargyl, 3'-O-2-nitrobenzyl, 3'-O-dimethoxytrityl, 3'-O-tert-butyl dimethyl silyl, or 3'-O-trimethyl(silyl)ethoxymethyl. In one embodiment, the nucleotide is a deoxyribonucleotide (dNTP). In one embodiment, the nucleotide is a 3'-O-modified dNTP.

The kit may further comprise a nucleic acid primer, such as a single- or double-stranded DNA or RNA oligonucleotide. The nucleic acid primer may comprise a tag at one or both of the 3' and 5' ends for detection or further processing (e.g., purification), such as but not limited to a biotin moiety.

The kit may further comprise one or more divalent metal cations, such as Mg^{2+} , Mn^{2+} and/or Co^{2+} .

One or more components of the kit may be provided in lyophilised or dry form for reconstitution in a suitable solvent. The kit may include solvents and buffers for nucleic acid synthesis by the polymerase.

The kit can further include substrates for assessing the activity of the RvPolX enzymes, as well as reagents for detecting the products. The kit can also include reagent dispensers and instructions for use of the kit.

Methods of nucleic acid synthesis

Disclosed herein is the use of a RvPolX polypeptide for nucleic acid synthesis, wherein the RvPolX polypeptide i) comprises an amino acid sequence having at least 70% sequence identity to an amino acid sequence of SEQ ID NO: 1, or ii) is a RvPolX variant polypeptide as defined herein.

Disclosed herein is the use of a polypeptide as defined herein for synthesizing a nucleic acid molecule with a nucleic acid primer and at least one nucleotide.

Disclosed herein is a method of synthesizing a nucleic acid molecule, the method comprising contacting a nucleic acid primer with at least one nucleotide and a polypeptide as defined herein, under conditions for the addition of the at least one nucleotide to the nucleic acid primer by the polypeptide.

In some embodiments, the nucleic acid primer is a single-stranded or double-stranded DNA molecule.

In some embodiments, the primer comprises a tag at one or both of the 3' and 5' ends. In one embodiment, the tag comprises a biotin moiety.

The primer and/or polypeptide may be provided on a solid support, such as a membrane, resin, or other solid phase carrier including but not limited to beads, particles, granules, a gel, or a surface. Alternatively, the primer and/or polypeptide may be provided in solution.

In some embodiments, the nucleotide is a ribonucleotide (rNTP), deoxyribonucleotide (dNTP) or a 3'-O-modified rNTP or dNTP. The 3' modification may be a 3'-O-acetyl, 3'-O-amino, 3'-aminooxy, 3'-O-methyl, 3'-O-azido, 3'-O-azidomethyl, 3'-O-cyanoethyl, 3'-O-allyl, 3'-O-(N-oxime), 3'-O-acetyl, 3'-O-propargyl, 3'-O-2-nitrobenzyl, 3'-O-dimethoxytrityl, 3'-O-tert-butyl dimethyl silyl, or 3'-O-trimethyl(silyl)ethoxymethyl. In one embodiment, the nucleotide is a deoxyribonucleotide. In one embodiment, the nucleotide is a 3'-O-NH₂ dNTP.

In some embodiments of the method, the contacting is performed in the presence of Mn^{2+} , Mg^{2+} and/or Co^{2+} . In one embodiment, the contacting is performed in the presence of Mn^{2+} . In one embodiment, the contacting is performed in the presence of Mn^{2+} and Mg^{2+} .

In some embodiments, the Mn^{2+} and/or Mg^{2+} is each present at a concentration of about 0.5 mM to about 20 mM, such as a concentration of about 0.5 mM, about 1 mM, about 2 mM, about 3 mM, about 4 mM, about 5 mM, about 6 mM, about 7 mM, about 8 mM, about 9 mM, about 10 mM, about 11 mM, about 12 mM, about 13 mM, about 14 mM, about 15 mM, about 16 mM, about 17 mM, about 18 mM, about 19 mM, or about 20 mM.

In some embodiments, the Co^{2+} is present at a concentration of about 0.5 mM to about 3 mM, such as a concentration of about 0.5 mM, about 1 mM, about 1.5 mM, about 2 mM, about 2.5 mM, or about 3 mM.

In some embodiments, the contacting is performed at a temperature of about 20°C to about 40°C, such as about 20°C, about 25°C, about 30°C, about 31°C, about 32°C, about 33°C, about 34°C, about 35°C, about 36°C, about 37°C, about 38°C, about 39°C, or about 40°C.

In some embodiments, the contacting is performed at a pH of about 6 to about 10, such as a pH of about 6, about 6.5, about 7, about 7.5, about 8, about 8.5, about 9, about 9.5, or about 10.

In some embodiments, the contacting is performed at a salt concentration of about 100 mM or more. The salt may be a monovalent salt, such as NaCl or KCl.

In some embodiments, the contacting is performed at a high salt concentration, such as a salt concentration of at least 400 mM. In some embodiments, the contacting is performed at a salt concentration of about 400 mM to about 1500 mM, such as a salt concentration of about 400 mM, about 500 mM, about 600 mM, about 700 mM, about

800 mM, about 900 mM, about 1000 mM, about 1100 mM, about 1200 mM, about 1300 mM, about 1400 mM, or about 1500 mM.

Where 3'-O-modified nucleotides are used, the method may comprise a step of deprotecting the primer being extended at the protected 3'-O-position. Methods of deprotection are well known in the art and the skilled person may select appropriate deprotection reagents and conditions according to the 3'-O-protecting group used.

In embodiments where the primer is immobilised to a solid support, the method may also comprise cleaving or releasing the extended primer from the solid support. In some embodiments, the method comprises isolating the extended primer, for example, using the tag on the primer.

As used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative (or).

As used in this application, the singular form "a," "an," and "the" include plural references unless the context clearly dictates otherwise. For example, the term "an agent" includes a plurality of agents, including mixtures thereof.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications which fall within the spirit and scope. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

Unless otherwise defined, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

Certain embodiments of the invention will now be described with reference to the following examples which are intended for the purpose of illustration only and are not intended to limit the scope of the generality hereinbefore described.

EXAMPLES

Materials and Methods

Expression Constructs

Synthesis of the codon-optimized, full-length gene sequence of RvPolX (UniProt accession number: A0A1D1UV65), in a pET-28a (+) vector (Novagen) with a cleavable N-terminal His-SUMO tag, was carried out by Twist Bioscience (US). To prepare mutant for CD spectroscopy studies, mutant with $\Delta 1-258$ was generated. For protein engineering studies, the selected residues (R276, E281, C425, S427, R432, G513, W514, M518, Y519, R525, R522, and D537) located at the incoming nucleotide binding pocket were subjected to mutations. Twelve single residues were chosen, and each was mutated to 19 different amino acids using the NEBuilder HiFi DNA Assembly master kit (New England Biolabs), following the manufacturer's instructions.

Recombinant Protein Production

The constructs were transformed into *Escherichia coli* Rosetta 2 (DE3). The bacterial culture carrying the plasmid of interest was grown in LB medium supplemented with antibiotics at 37°C with shaking at 250 rpm. Upon reaching OD₆₀₀ reach 0.8, 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) was introduced, and the culture was then incubated at 16°C with shaking at 200 rpm overnight. The bacterial cultures were subsequently harvested, lysed, and the supernatant proteins were purified using TALON® Metal Affinity Resin (Takara). To remove the N-terminal His-SUMO tag, the purified proteins were incubated with SUMO protease overnight at 4°C with an approximate molar ratio of 1:25. The purified proteins were further subjected to purification via HiTrap™ Heparin HP affinity columns (GE Healthcare Life Sciences). Finally, 4-20% mini-PROTEAN® TGX™ precast protein gel (Bio-Rad) was used to analyze the purified recombinant proteins.

DNA Polymerase Activity Assays

A 5'-Alexa Fluor 488-labeled 14-mer single stranded DNA (Alexa-ssDNA14) oligonucleotide (5'-Alexa488-TACGCATTAGCCTG-3') was used as the DNA primer. Assay mixtures were prepared on ice. The reaction was initiated by adding 0.3 mM of the respective dNTP, incubated at 25°C for 5 to 30 minutes, and inactivated at 95°C for 2 minutes using a thermocycler. The products were analyzed by gel electrophoresis on an 18% acrylamide gel with 8 M urea (urea-PAGE). The gel bands were visualised through the fluorescence of the Alexa488 fluorophore using ChemiDoc MP Imaging System (Bio-Rad), and the band intensities were quantified using Image Lab software (Bio-Rad).

To assay primer elongation by RvPolX under different pH conditions, a reaction mixture containing 4 μM RvPolX, 50 nM Alexa-ssDNA14, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 1% glycerol, and 1 mM MnCl₂ was prepared in various pH buffers, including 50 mM MES pH 6.0, 50 mM Tris-HCl pH 6.8, 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.4, 50 mM Tris-HCl pH 8.0, or 50 mM Tris-HCl pH 8.8. The reaction was initiated by adding nucleotide.

To assay primer elongation by RvPolX in the presence of different divalent ions, a reaction mixture containing 50 mM Tris-HCl pH 8.8, 4 μ M RvPolX, 50 nM Alexa-ssDNA14, 1 mM TCEP, and 1% glycerol was prepared. Various divalent ions, including 5 mM CaCl_2 , 5 mM LiCl_2 , 5 mM MgCl_2 , 1 mM MnCl_2 , or 1 mM ZnCl_2 , were added individually. To assay the primer elongation by RvPolX in the presence of a mixture of two different divalent ions, 1 mM MnCl_2 and 5 mM MgCl_2 were added. The reaction was initiated by adding nucleotide, and terminated after 5 minutes.

To assay primer elongation by RvPolX at different salt concentrations, various concentrations of NaCl (ranging from 0 to 700 mM) were added to the reaction mixture containing 50 mM Tris-HCl pH 8.8, 4 μ M RvPolX, 50 nM Alexa-ssDNA14, 1 mM TCEP, 1 mM MnCl_2 , 5 mM MgCl_2 , and 1% glycerol. The reaction was initiated by adding nucleotide solution containing various concentrations of NaCl (ranging from 0 to 700 mM), followed by rapid stirring. The reaction was terminated after 30 minutes.

To assay primer elongation by RvPolX at different temperatures, a reaction mixture containing 50 mM Tris-HCl pH 8.8, 4 μ M RvPolX, 50 nM Alexa-ssDNA14, 1 mM TCEP, 1 mM MnCl_2 , 5 mM MgCl_2 , and 1% glycerol was prepared. The reaction mixture and nucleotide solutions were pre-incubated separately at different temperatures for 5 minutes. The reaction was then initiated by adding nucleotide solution, followed by rapid stirring. The reaction was terminated after 5 minutes.

To compare the activity of RvPolX variants and HsTdT under high salt conditions, a lower concentration of enzyme and dNTPs were used. Assay mixtures were prepared on ice with 0.5 μ M G513A+R522I or HsTdT, 50 nM Alexa-ssDNA14, 1 mM MnCl_2 , 5 mM MgCl_2 , 1 mM TCEP, 1% glycerol, and 50 mM Tris-HCl pH 8.8 in the presence of 500 mM or 1000 mM KCl. The reaction was initiated by adding 50 μ M of the respective dNTP, followed by incubation at 25°C for 10 to 30 minutes, and inactivated at 95°C for 2 minutes using a thermocycler. The products were analyzed by gel electrophoresis on an 18% acrylamide gel with 8 M urea (urea-PAGE) and visualised using ChemiDoc MP Imaging System (Bio-Rad).

To assay primer elongation by G513A+R522I in the presence of various concentrations of single and mixed divalent metal ions, a reaction mixture containing 50 mM Tris-HCl pH 8.8, 0.5 μ M RvPolX, 50 nM Alexa-ssDNA14, 1 mM TCEP, 500 mM NaCl, and 1% glycerol was prepared. For single metal ion assays, CoCl₂, MgCl₂, and MnCl₂ were individually tested at 0.5, 1, 2, 3, 5, and 10 mM concentrations. For mixtures of two divalent metal ions, combinations of 0, 1, 5, and 10 mM MnCl₂ and MgCl₂ were evaluated. The reaction was initiated by adding 50 μ M of the respective dNTP, followed by incubation at 25°C and terminated after 10 minutes using a thermocycler. The products were analyzed by gel electrophoresis on an 18% acrylamide gel with 8 M urea (urea-PAGE) and visualised using ChemiDoc MP Imaging System (Bio-Rad).

To assess and compare primer elongation between G513A+R522I and commercial calf thymus TdT using modified nucleotides, a reaction mixture for G513A+R522I was prepared with 1 μ M G513A+R522I, 50 nM Alexa-ssDNA14, 1 mM MnCl₂, 5 mM MgCl₂, 1 mM TCEP, 1% glycerol, and 50 mM Tris-HCl pH 8.8. For commercial calf thymus TdT, a reaction mixture was prepared with 1 μ M TdT (New England Biolabs), 1X Terminal Transferase Reaction Buffer (New England Biolabs), 0.25 mM CoCl₂ (New England Biolabs), and 50 nM Alexa-ssDNA14. The reaction was initiated by adding 0.25 mM of the respective 3'-ONH₂ modified nucleotides (Firebird Biomolecular Sciences), followed by incubation at 25°C for 40 minutes. Subsequently, the reaction was inactivated at 95°C for 2 minutes using a thermocycler. The products were then analyzed by gel electrophoresis on an 18% acrylamide gel with 8 M urea (urea-PAGE) and visualised using ChemiDoc MP Imaging System (Bio-Rad).

Circular Dichroism Spectroscopy

The protein solutions were diluted to a final concentration of 3 μ M in a buffer containing 25 mM HEPES at pH 7.4, 10% glycerol, and 500 mM NaCl. Prior to the measurements, all samples were incubated at room temperature for 5 minutes. A quartz cuvette with a path length of 0.1 cm, filled with 300 μ L of the sample, was used for the measurements. Circular dichroism (CD) measurements were conducted using a Jasco J-815 Spectrometer (Jasco), at a wavelength range of 200 to 250 nm, with the incubation temperature gradually increased from 25 to 50°C at intervals of 0.1 nm, and 10

accumulations recorded for each reading. The resulting data was analyzed using the web-based server BeStSel (Beta Structure Selection).

Computational Modelling

The protein sequence of RvPolX was retrieved from the UniProt database (accession number: A0A1D1UV65). The whole modelling protocol was given as follows: (1) three initial models of RvPolX were firstly predicted by AlphaFold v2.0 program; (2) the best model with the top ranked score was selected and its N-terminus (1-234aa) containing a long-disordered loop was truncated; (3) this pruned model was subjected to the sidechain optimization in Sybyl-X package (v2.1), which afforded diverse models; (4) the model with the lowest energy was selected for the subsequent modelling; (5) two coordinated metal ions (e.g., Mn^{2+}) in the catalytic site were modelled with the following procedure: one ion was positioned in the center of two carboxyl groups of D437 and D439, while the other was placed in the center of three carboxyl groups of D437, D439, and D498; subsequently, two metal ions and the sidechains of D437, D439, and D498 were optimized in Sybyl-X package to relax their conformations, while the remaining part of enzyme was fixed. (6) a single-strand DNA (ssDNA, 5'-AGCCTG-3') was placed in the binding pocket. More specifically, the coordinates of ssDNA backbone were initialized from the crystal structure (PDB ID: 4I27) of its close homolog TdT with ssDNA by aligning the structure of RvPolX to that of TdT; and then the missing bases were added by tleap program in the AmberTool package; only the ssDNA was optimized in the binding pocket by the Sybyl package, while the whole enzyme and two metal ions were restrained during the optimization; both ssDNA and its neighbouring sidechains on the RvPolX were fully relaxed with the restraint of the remaining system; (7) the sidechains of RvPolX and bases of ssDNA were optimized with the restraint of backbone; (8) the whole system including RvPolX, ssDNA, and two metal ions is fully optimized for subsequent molecular dynamics simulation. The complex model from the previous protocol was further optimized by molecular dynamics simulation in the explicit water environment with the following steps. (1) the complex model was solvated by TIP3P water model and neutralized by counter ions with the NaCl concentration of 150 mM; (2) the whole system was minimized with position restraint of the heavy atoms by exerting a harmonic force constant of $5.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, and

subsequently fully relaxed without any restraint; (3) the whole system was gradually heated to 298.15K with the position restraint of $5.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for 25ps; (4) the whole system was equilibrated in the NVT ensemble for 25ps; (5) the whole system was subjected to three rounds of 500ps equilibration in NPT ensemble with the successive decrease of harmonic force constants (5.0, 1.0, and $0.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, respectively); (6) the whole system was subjected to the production simulation without any restraint at the pressure of 1 atm and temperature of 298.15K; Monte-Carlo barostat and Langevin thermostat were employed to control the pressure and temperature, respectively; ff14SB and OC15 force field parameters were applied for RvPolX and ssDNA, respectively; the production simulation time was set to 50ns; Smooth Particle Mesh Ewald (SPME) method was employed to calculate the long-range electrostatic interaction with the cutoff of 9 Å; the whole simulation was conducted for three times in the AMBER22 program.

Five typical models were retrieved from the previous molecular dynamics simulations by K-Mean cluster analysis, minimized, and processed in the Sybyl package to remove water molecules and counter ions. Only two models with an open binding site exposure to the incoming nucleotide were kept for subsequent molecular docking simulations. The incoming nucleotide dGTP were modelled and optimized in Sybyl package (v2.1) and subsequently docked into the binding site of RvPolX/ssDNA complex model by the GOLD program v2018. The docking parameters were provided as follows: scoring function is ChemScore; population size is 400; selection pressure is 1.1; number of operations is 500,000; number of islands is 5; niche size is 2; crossover frequency is 95%; mutation frequency is 95%; migration frequency is 10%. After docking simulations, all the sampled conformations for each incoming nucleotide were further filtered by the following criteria: (1) the base of the incoming nucleotide can form favourable π - π stacking with the ss DNA (5'-AGCCTG-3'); (2) the phosphate group(s) of the incoming nucleotide can form the salt bridge with the nearby residues such as lysine or arginine; (3) the phosphate group(s) of the incoming nucleotide can form the coordination interaction with a metal ion in the catalytic pocket. According to this filtering, a reasonable conformation (**Figure 5A**) with the incoming nucleotide dGTP was achieved, minimized in the Sybyl package, and finally rescored in the GOLD program.

Protein Engineering and Screening of RvPolX Mutants

Plasmids encoding mutations to the RvPolX dNTP binding pocket residues were transformed into Rosetta™ 2(DE3) *E. coli* cells. For protein expression, a single colony of each mutant was selected and grown in 4 mL LB cultures. The cells were harvested by centrifugation at 3,300 *g* for 1 minute and then lysed by resuspending them in 500 µL of B-PER™ Bacterial Protein Extraction Reagent (Thermo Scientific), followed by incubation at room temperature for 30 minutes. After centrifugation at 16,000 *g* for 20 minutes at 4 °C, the supernatant was incubated with 20 µL of pre-washed Dynabeads™ (Thermo Scientific, US) in cold room for 1 hour. After that, the beads were separated and washed 3 times with wash buffer (25 mM HEPES pH 7.4, 500 mM NaCl, 0.01% Tween-20, 10 mM imidazole). The immobilized protein on the Dynabeads™ was incubated with a reaction buffer containing 25 mM HEPES pH 7.4, 1 µM ssDNA14, 1 mM dCTP, 1 mM TCEP, 1 mM MnCl₂, 5 mM MgCl₂, and 1% glycerol at 25°C and 1000 rpm for 60 minutes. After centrifugation, 30 µL of supernatant was analyzed using a UHPLC system (Shimadzu, Japan). Each analysis involved injecting 10 µL of the sample onto a Clarity 2.6 µm Oligo-XT 100 Å, LC Column 100 x 4.6 mm (Phenomenex). Mobile phase A (15 mM trimethylamine, 400 mM 1,1,1,3,3,3-hexafluoro-2-propanol, 10% methanol, pH 8.0) and mobile phase B (100% methanol) were prepared. A 15-minute gradient program ranging from 16% to 36% mobile phase B was employed to analyze the extended products.

Microscale Thermophoresis

Microscale thermophoresis (MST) study was conducted to investigate the interaction between RvPolX and a mutant with four dNTPs. This study utilized the Monolith™ NT.labelFree system by NanoTemper Technologies. A series of samples containing nucleotides (at concentrations ranging from 0.01 to 0.15 µM) and enzyme (0.8 µM) in a reaction buffer consisting of 50 mM HEPES at pH 7.4, 5 mM MgCl₂, 1 mM MnCl₂, and 2 mM TCEP were incubated at room temperature for 10 minutes before being loaded individually into Monolith™ NT.labelFree capillaries. The measurements were conducted at room temperature with 20% LED power and 20% MST power, and the

resulting data were analyzed using MO Affinity Analysis to determine the dissociation constants.

Kinetic Analysis

To monitor the time-dependent primer elongation, a reaction mixture containing 4 μM RvPolX, 50 mM Tris-HCl pH 8.8, 50 nM Alexa-ssDNA14, 1 mM TCEP, 1 mM MnCl_2 , 5 mM MgCl_2 , and 1% glycerol was prepared. After a 5 minutes pre-incubation at 25 $^\circ\text{C}$, the reaction was initiated by adding nucleotide. The reaction was halted at specific time intervals and analyzed using 18% denaturing urea-PAGE, visualized using Chemidoc (BioRad), and the bands quantified using Image Lab (Bio-Rad). Curve fitting was performed using the Graphpad Prism 9. The oligonucleotide chain growth polymerization kinetics was modelled as a Poisson distribution. For each of the assays, a specific oligonucleotide product band (k) that allowed us to observe the time-dependent accumulation and subsequent depletion of the oligonucleotide product, was selected for analysis. The apparent rate of chain elongation (V_{obs}) was estimated by fitting the changes in band intensity (I) over time (t) to a Poisson probability distribution function, with a normalization constant (A).

$$I = At^k e^{-V_{\text{obs}}t}$$

Example 1: RvPolX displays template-independent DNA polymerization activity and high salt tolerance

The tardigrade *R. varieornatus* contains three PolXs, of which two belong to the Pol λ -like group (UniProt IDs: A0A1D1UV65 and A0A1D1UJX5, 34 and 38% sequence identities to human Pol λ), and one belongs to the Pol β -like group (A0A1D1W154, 32% sequence identity to human Pol β). One of the *R. varieornatus* Pol λ -like enzymes (UniProt ID: A0A1D1UV65, designated as RvPolX) was recombinantly expressed in and purified from *E. coli* cultures for further biochemical analysis.

The domain structure of RvPolX consists of an N-terminal BRCT domain and a C-terminal catalytic domain, similar to TdT, Pol μ , and Pol λ (**Figure 1A**). A multiple

sequence alignment of the catalytic domains of RvPolX and the mammalian PolXs suggest a conserved fold, which includes the 8 kDa, fingers, palm, and thumb subdomains. The TdT crystal structure contains a specific loop (loop1) that obstructs template strand binding, and is reported to contribute to its template-independent polymerase activity. Loop1 is also present in Pol μ , and although this loop is disordered in the Pol μ crystal structures, it is nevertheless also thought to play a role during template-independent polymerization (**Figure 2**). Deletion of loop1 was reported to decrease template-independent activity and increase template-dependent activity in both TdT and Pol μ . An equivalent loop1 region is absent in Pol β , Pol λ , and is also absent in the sequence of RvPolX (**Figure 1A**).

To investigate whether RvPolX can catalyze template-independent DNA polymerization, RvPolX (4 μ M) was incubated with a 14-mer single-stranded DNA (ssDNA14, 0.05 μ M) and different dNTPs (0.3 mM). Despite lacking the loop1 element, the assays revealed that RvPolX displays modest template-independent DNA polymerization activity (**Figures 1B-E**). RvPolX activity increased with increasing pH (pH 5.0 to pH 8.8, **Figure 1B**), and decreased with increasing NaCl concentration (0-700 mM, **Figure 1D**). The isoelectric point of RvPolX was calculated to be 6.80, with a net charge of -11.1 at pH 8.8. A high negative surface charge density is found in many halophilic proteins, and could contribute to the stability of RvPolX under the reaction conditions. Activity was detectable up to 700 mM NaCl for incorporation of dGTP, dTTP, and dCTP, and up to 500 mM NaCl for incorporation of dATP, suggesting that RvPolX might be suitable for further development as a salt-tolerant polymerase. In assays with various divalent metal ions (Ca²⁺, Co²⁺, Mg²⁺, Mn²⁺, and Zn²⁺), RvPolX activity was highest with Mn²⁺ as the metal cofactor, followed by Mg²⁺ and Co²⁺ (**Figure 3**). In assays with different combinations of divalent metal ions, activity was highest when both Mg²⁺ and Mn²⁺ were present (**Figure 1C**). Subsequent assays were carried out at pH 8.8, with Mg²⁺ (5 mM) and Mn²⁺ (1 mM).

Next, the effect of the oligonucleotide substrate sequence on RvPolX activity was examined, by employing 15-mer oligonucleotide substrates (ssDNA14A, ssDNA14T, ssDNA14G, and ssDNA14C) composed of ssDNA14 with a single base (A, T, G, or C) added at the 3' end. The assays revealed that RvPolX did not exhibit a strong preference

for the specific 3' base of the oligonucleotide substrate (**Figure 1E**). Conversely, across a range of reaction conditions, RvPolX exhibits a preference for the dNTP substrate in the following order: dGTP > dTTP > dATP = dCTP (**Figures 1B-E**). Collectively, the assays demonstrate that RvPolX catalyzes template-independent DNA polymerization, and might be suitable for further development as a salt-tolerant polymerase.

The RvPolX N-terminal Region Contributes to its Stability

To investigate the possible contributions of the RvPolX N-terminal region to its activity and stability, a RvPolX N-terminal 258-amino acid deletion mutant (Δ 1-258aa) was constructed which contained only the catalytic domain, and its activity was compared to that of RvPolX wild-type (WT). For assays in the absence of NaCl, the activity of WT was constant at 25°C and 40°C, but decreased at 45°C (**Figure 4C**). Activity of Δ 1-258aa is comparable to WT at 25°C and 40°C, but significantly less than WT at 45°C. For assays in the presence of 500 mM NaCl, the activity of WT decreased between 25°C and 40°C, and was eliminated at 45°C. Activity of Δ 1-258aa is significantly less than WT at 25°C and 40°C, and was also eliminated at 45°C. The assays demonstrated that deletion of the RvPolX N-terminal region decreased its resilience to high temperatures, particularly under high salt concentrations. Thus, although the RvPolX N-terminal region is not essential for template-independent DNA polymerase activity, it plays a role in preserving activity under high temperature and salt conditions.

It was next investigated whether the RvPolX N-terminal region contributes to its structural stability under high salt concentrations. Far-UV circular dichroism (CD) spectroscopy experiments of WT and Δ 1-258aa were conducted at a range of temperatures (25°C to 50°C), in the presence of 500 mM NaCl (**Figures 4A-B**). At 25-35°C, the CD spectra of both WT and Δ 1-258aa exhibited characteristic features of α -helices. At 40°C and above, drastic changes to the CD spectra were observed for Δ 1-258aa, indicating a loss of α -helical secondary structure. A corresponding drastic change was not observed in WT, suggesting that deletion of the N-terminal amino acids (1-258aa) adversely affects the stability of the RvPolX secondary structure. Subsequent experiments were thus carried out with full-length RvPolX.

Discussion

Among the X family DNA polymerases, the mammalian Pol β , Pol λ , Pol μ , and TdT are the four representative enzymes that have been subjects of in-depth studies. In particular, TdT has garnered significant attention for *de novo* DNA synthesis applications, especially in the context of DNA-based data storage, due to its high template-independent polymerase activity. The findings on RvPolX, from the extremotolerant tardigrade *R. varieornatus*, expands the scope of biochemically characterized PolX enzymes to invertebrates. It is demonstrated that RvPolX possesses modest template-independent DNA polymerase activity, despite sharing only 21% sequence identity with TdT. Notably, while TdT becomes inactive at salt concentration above 300 mM, RvPolX exhibits significantly higher salt tolerance, retaining catalytic activity at salt concentrations even up to 1000 mM. Salt tolerant versions of the template-dependent DNA polymerase from Bacillus phage phi29 have previously been employed in nanopore devices for DNA sequencing. The development of salt-tolerant template-independent polymerases could enable applications in nanopore-based devices for *de novo* DNA synthesis, involving electrophoretic control of the DNA synthesis process, and *in situ* proofreading by nanopore sequencing.

Analysis of the phylogenetic tree of PolX in animals (Eumetazoa) reveals that numerous diverse sequences remain to be explored. In vertebrates, homologues of all four PolX enzymes are found in mammals and fish, while Pol μ is absent in birds, and TdT is absent in amphibians. Among the diverse invertebrate phyla, the presence of PolX enzymes is more sporadic and required further investigation. In mammals, the PolX enzymes are expressed in a tissue-specific manner, and serve distinct functions related to DNA repair. As the PolX enzyme with the highest template-independent DNA polymerase activity, TdT plays specific roles in vertebrate adaptive immunity. This adaptation is absent in invertebrates, suggesting that the invertebrate PolXs may instead serve as functional counterparts of Pol β and Pol λ . Many of the PolX variants, including RvPolX, contain an N-terminal BRCT domain, suggesting their involvement in repair of DNA double-stranded breaks by NHEJ. In this context, template-independent DNA polymerase activity has been proposed as one of the mechanisms that provides flexibility in the repair of various types of double-strand breaks. Thus, a systematic investigation of

diverse invertebrate PolX enzymes, particularly those containing a BRCT domain, has the potential to enrich the collection of template-independent polymerases with unique characteristics suitable for various DNA synthesis applications.

Example 2: Protein Engineering of RvPolX

Targeted mutagenesis of active site residues in RvPolX was carried out to enhance its salt-tolerant template-independent DNA polymerase activity. A structural model of the ternary complex of RvPolX and its substrates was constructed by molecular docking of the AlphaFold structure of RvPolX with ssDNA14, two Mn^{2+} ions, and each of the four dNTPs (**Figure 5A**). From the structural model, 12 amino acid residues located in and around the dNTP binding pocket (R276, E281, C425, S427, R432, G513, W514, M518, Y519, R525, R522, and D537, **Figure 5A**) were selected for saturation mutagenesis and high throughput activity screening.

A total of 240 clones from the mutant library were individually expressed and purified using Dynabeads™, and screened for template-independent DNA polymerase activity, with reaction products analyzed by high-performance liquid chromatography (HPLC). In general, substitutions to hydrophobic amino acids (e.g., A, C, F, I, L, M, V, W, and Y) at these 12 sites lead to improved catalytic activity. Four mutants G513A, R522I, R522M, and R522V displayed increased activity for dATP incorporation relative to the WT (**Figure 6A**), and were thus selected for further development.

To explore the potential synergistic effects of the G513A and R522I/M/V mutations, double mutants were generated and assayed for template-independent DNA polymerization. Among three double mutants, the combination of G513A and R522I (G513A+R522I) led to the highest activity (**Figure 6B**). The extent of conversion of the oligonucleotide substrate for WT and selected mutants, in the presence of 500 mM and 1000 mM NaCl, was estimated by quantifying the ssDNA14 and extension product bands in each gel lane. For assays in the presence of 500 mM NaCl (**Figure 5B**), G513A+R522I displayed higher conversion of the oligonucleotide substrate relative to WT, for incorporation of dATP (7.9-fold) and dGTP (1.8-fold). The enhancements were further amplified in the presence of 1000 mM NaCl (**Figure 5C**), for incorporation of

dTTP (1.7-fold), dCTP (1.5-fold), dGTP (2.3-fold), and dATP (increase in extension product from undetectable to 8%). Overall, the G513A+R522I mutant exhibits an enhanced salt-tolerant polymerase activity, and increased promiscuity towards dNTP substrates, with a particularly marked increase in proficiency for incorporation of dATP. As depicted in **Figure 5A**, the formation of a continuous hydrophobic patch in the nucleotide binding pocket is crucial for further improvement in catalytic activity. A potentially optimal combination of residues in this hydrophobic patch can be A or C at 513, W at 514, M at 518, V at 521, and I, M or V at 522.

Biochemical Investigation of RvPolX G513A+R522I

For a more quantitative comparison of the activity of the G513A+R522I with RvPolX WT, the assays were repeated for each of the four dNTPs in the absence of salt. As shown in **Figures 7A** and **7B**, the oligonucleotide chain elongation products were measured over a course of 300 or 1800 s, and the chain growth polymerization kinetics modelled as a Poisson distribution. For each of the assays, an oligonucleotide product band was selected for which the time-dependent rise and subsequent fall in band intensity could be observed during the course of the assay. The apparent rate of chain elongation (V_{obs}) was estimated by fitting the corresponding time-dependent accumulation and subsequent depletion of the product to a Poisson probability distribution function (**Figure 7C**). For dATP, it was not possible to select a product band containing the rise and fall phases for both WT and G513A+R522I. Therefore, the +3 band was monitored, which contains the rise and fall phases for G513A+R522I, and the rise and plateau phases for WT. We also monitored the +1 band containing the rise and fall phases for WT, showing that the rates calculated for the +1 and +3 bands are comparable. As shown in **Table 2**, for each of the four dNTPs, the calculated V_{obs} for G513A+R522I was significantly higher than those for WT (2-, 3-, 7-, and 36-fold higher for G, C, T, and A, respectively).

Table 2. The apparent rates of chain elongation (V_{obs}) of WT and G513A+R522I RvPolX.

Band of extension product	V_{obs} of WT (s ⁻¹)	V_{obs} of G513A+R522I (s ⁻¹)
---------------------------	------------------------------------	---

dATP	3	0.0009 ± 0.0001	0.0380 ± 0.0014
dATP	1	0.0011 ± 0.0000	N.A.
dCTP	2	0.0104 ± 0.0007	0.0356 ± 0.0018
dTTP	3	0.0066 ± 0.0002	0.0439 ± 0.0020
dGTP	6	0.0744 ± 0.0010	0.1597 ± 0.0065

To examine the possible mechanistic origin for the enhancement of catalytic activity of the G513A+R522I, the dissociation constants for binding of the four dNTPs to RvPolX WT and G513A+R522I were determined using label-free microscale thermophoresis (MST, **Figure 7D**). For both WT and G513A+R522I, the lowest K_d (corresponding to the highest affinity) was observed for dATP, followed by dTTP, dCTP, and then dGTPs. The K_d for binding of dGTP to G513A+R522I was ~5-fold lower than WT, while the K_d for binding of the other dNTPs to the G513A+R522I was comparable to WT. The absence of correlation between dNTP binding affinity and enhancement of catalytic activity indicates the need for alternative explanations, particularly for the non-conservative R522I mutation. The residue corresponding to R522 is conserved in Pol λ , Pol μ , and TdT, and in the crystal structure of human Pol μ gap filling pre- and post-catalytic complexes, this residue interacts with the phosphodiester backbone of the template DNA strand. In the template-independent DNA polymerase assays, the oligonucleotide substrate ssDNA14 could bind to either the primer elongation site or the template binding site. Therefore, one conceivable hypothesis regarding the effect of the R522I mutation is its potential to destabilize non-productive binding of ssDNA14 to the template binding site.

It was found that the RvPolX variant had maximal activity in the presence of a combination of Mg^{2+} and Mn^{2+} , which aligns with other PolX enzymes (**Figure 8**).

A comparison of polymerase activity of human terminal deoxynucleotidyl transferase (HsTdT) and the engineered RvPolX variant (G513A+R522I) under high salt conditions revealed that both enzymes retained catalytic activity at 500 mM and 1000 mM KCl (**Figure 9**). TdT exhibited higher activity for incorporating dGTP and dCTP, while RvPolX showed greater activity for incorporating dATP and dTTP at 1000 mM KCl. Thus, the engineered RvPolX variant demonstrates comparable activity to HsTdT under

high salt conditions and may be used as an alternative to HsTdT for biotechnological applications.

A comparison of template-independent DNA polymerase activity of the RvPolX (G513A+R522I) variant and commercial calf thymus TdT (New England Biolabs) using 3'-ONH₂-modified nucleotides as substrate showed that the RvPolX variant has significantly higher activity than TdT in incorporating protected 3'-ONH₂ dNTPs (**Figure 10A**). The RvPolX variant also exhibited relatively low exonuclease activity compared to the TdT in the absence of nucleotides (**Figure 10B**).

Discussion

Previous structural studies showed that loop1 of TdT adopts a lariat-like conformation, impeding the binding of a continuous template DNA strand, a feature believed to be important in its template-independent DNA polymerase activity. Deletion of loop1 or substitution of amino acid residues within the loop reduces or completely abolishes template-independent DNA polymerase activity. Interestingly, loop1 is absent in both RvPolX and the well-studied Pol λ , which has also been reported to have modest template-independent DNA polymerase activity. During template-independent polymerization, Pol λ exhibits a preference for incorporating pyrimidine nucleotides, while wild-type RvPolX exhibits a preference for dGTP, with lower activities for pyrimidine nucleotides, and limited activity for dATP. Nevertheless, these experiments demonstrated that the activity of RvPolX can be enhanced by engineering of the dNTP binding pocket (**Figure 5**).

By employing a combination of computational modelling of the active site, mutagenesis of the dNTP binding pocket, and HPLC-based activity screening under stringent reaction conditions with 350 mM NaCl, two mutations were identified (G513A and R522I) which increased enzyme activity. Combining the two mutations led to a synergistic increase in activity for incorporation of all four dNTPs, particularly dATP (~35-fold), yielding a polymerase with overall higher activity and substrate promiscuity (**Table 2**). Modifications to this procedure, such as mutagenesis of the oligonucleotide-

binding pocket or screening under alternative stringent reaction conditions, can be adapted to further enhance the activity and robustness of this enzyme.

In RvPolX, G513 is part of a GW motif, which is also present in TdT and Pol μ , but is replaced with a YF motif in Pol λ and Pol β , thought to underlie the different nucleotide selectivity of these enzymes. The Y residue in Pol λ and Pol β is positioned near the dNTP 2' carbon and is proposed to act as a steric gate, preventing the binding and incorporation of rNTPs. Substitution of Y with G in TdT and Pol μ expands the nucleotide binding pocket, allowing the binding and incorporation of both dNTPs and rNTPs. Interestingly, RvPolX is the only Pol λ -like enzyme containing the GW motif, while the others contain YF, HY or other motifs (**Figure 11**). Given its proximity to the dNTP substrate, the increased activity of the G513A mutation may result from the binding of the dNTP in a more reactive conformation, or through rigidification of the dNTP binding pocket.

In conclusion, this research on RvPolX led to the development of an enzyme with salt-tolerant template-independent DNA polymerase activity, adding to the enzymatic toolkit available for *de novo* DNA synthesis, and enabling applications such as integration into nanopore-based systems. In the context of *R. varieornatus*, the properties of this enzyme may relate to its expected role in the NHEJ pathway of DNA repair in the presence of various stressors.

It will be appreciated that many further modifications and permutations of various aspects of the described embodiments are possible. Accordingly, the described aspects are intended to embrace all such alterations, modifications, and variations that fall within the spirit and scope of the appended claims.

CLAIMS

1. A polypeptide that is distinguished from a wild-type *Ramazzottius varieornatus* X family DNA polymerase (RvPolX) by at least one amino acid substitution in the wild-type RvPolX amino acid sequence, wherein the at least one amino acid substitution is at a position corresponding to positions 281, 513 and/or 522 as set forth in SEQ ID NO: 1.
2. The polypeptide of claim 1, wherein the polypeptide comprises an amino acid sequence having at least 70% sequence identity to an amino acid sequence of SEQ ID NO: 3 or a fragment thereof.
3. The polypeptide of claim 1 or 2, wherein the amino acid substitution at the position corresponding to position 281, 513 and/or 522 is a substitution to a hydrophobic amino acid residue.
4. The polypeptide of any one of claims 1 to 3, wherein the amino acid substitution at the position corresponding to position 281 is a substitution to A, F, I, L, V or Y.
5. The polypeptide of claim 4, wherein the amino acid substitution at the position corresponding to position 281 is a substitution to L.
6. The polypeptide of any one of claims 1 to 5, wherein the amino acid substitution at the position corresponding to position 513 is a substitution to A or C.
7. The polypeptide of any one of claims 1 to 6, wherein the amino acid substitution at the position corresponding to position 522 is a substitution to I, M or V.
8. The polypeptide of any one of claims 1 to 7, wherein the polypeptide is distinguished from the wild-type RvPolX by at least two amino acid substitutions in the wild-type RvPolX amino acid sequence.
9. The polypeptide of claim 8, wherein the polypeptide comprises an amino acid substitution of i) G513A or G513C, and ii) R522I, R522M or R522V.

10. The polypeptide of any one of claims 1 to 9, wherein the polypeptide is distinguished from the wild-type RvPolX by three amino acid substitutions at positions 281, 513 and 522.
11. The polypeptide of claim 10, wherein the polypeptide comprises an amino acid substitution of i) E281L, ii) G513A or G513C, and iii) R522I, R522M or R522V.
12. The polypeptide of any one of claims 4 to 11, wherein the polypeptide comprises an amino acid sequence having at least 70% sequence identity to an amino acid sequence in SEQ ID NO: 4–26.
13. The polypeptide of any one of claims 1 to 12, wherein the polypeptide is capable of template-independent nucleic acid synthesis.
14. The polypeptide of claim 13, wherein the polypeptide is capable of nucleic acid synthesis at a salt concentration of about 100 mM to about 1500 mM.
15. A polynucleotide comprising a nucleic acid sequence encoding a polypeptide of any of the preceding claims.
16. An expression construct comprising the polynucleotide of claim 15.
17. A host cell comprising the polynucleotide of claim 15 or the expression construct of claim 16.
18. A method of producing a polypeptide of any one of claims 1 to 14, the method comprising a) culturing a host cell of claim 17 under suitable conditions to allow expression of the polypeptide, and b) isolating the polypeptide.
19. A method of synthesizing a nucleic acid molecule, the method comprising contacting a nucleic acid primer with at least one nucleotide and a polypeptide of any one of claims 1 to 14 under conditions for the addition of the at least one nucleotide to the nucleic acid primer by the polypeptide.
20. The method of claim 19, wherein the nucleic acid primer is a single-stranded or double-stranded DNA molecule.

21. The method of claim 19 or 20, wherein the nucleotide is a deoxyribonucleotide (dNTP) or a 3'-O-NH₂ dNTP.
22. The method of any one of claims 19 to 21, wherein the contacting is performed in the presence of Mn²⁺, Mg²⁺ and/or Co²⁺.
23. The method of claim 22, wherein the Mn²⁺ and/or Mg²⁺ is each present at a concentration of about 0.5 mM to about 20 mM.
24. The method of claim 22, wherein the Co²⁺ is present at a concentration of about 0.5 mM to about 3 mM.
25. The method of any one of claims 19 to 24, wherein the contacting is performed at a salt concentration of at least about 100 mM.
26. The method of claim 25, wherein the contacting is performed at a salt concentration of about 400 mM to about 1500 mM.
27. The method of any one of claims 19 to 26, wherein the contacting is performed at a temperature of about 20°C to about 40°C.
28. The method of any one of claims 19 to 27, wherein the contacting is performed at a pH of about 6 to about 10.
29. Use of a polypeptide of any one of claims 1 to 14 for synthesizing a nucleic acid molecule with a nucleic acid primer and at least one nucleotide.
30. A kit for synthesizing a nucleic acid molecule, comprising a polypeptide of any one of claims 1 to 14 and at least one nucleotide.
31. The kit of claim 30, wherein the nucleotide is a deoxyribonucleotide (dNTP) or a 3'-O-modified dNTP.