USES OF TYPE II AND TYPE I INORGANIC PYROPHOSPHATASES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/634,887, filed April 16, 2024, the entire contents of which is incorporated by reference herein.

TECHNICAL FIELD

[0002] The present disclosure provides uses of type II and type I inorganic pyrophosphatases and variant type I and type II inorganic pyrophosphatases under nonnatural conditions, particularly conditions present in industrial applications. In another aspect, the present disclosure describes compositions and methods of using inorganic pyrophosphatases with nucleoside triphosphates and in coupled reactions with a second enzyme.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0003] The official copy of the Sequence Listing is submitted concurrently with the specification as an XML file, with a file name of "CX10-270WO2_ST26.xml", a creation date of March 19, 2025, and a size of 178,983 bytes. The Sequence Listing filed is part of the specification and is incorporated in its entirety by reference herein.

BACKGROUND

[0004] A large number of physiologically relevant reactions in the cell produce pyrophosphate. Often, though not always, pyrophosphate is released from reactions coupled with nucleoside triphosphates (NTP). These reactions include cofactor (e.g. nicotinamide adenine dinucleotide (NAD)) biosynthesis, nucleotide biosynthesis, S-adenosyl-L-methionine biosynthesis, sugar metabolism, tRNA charging, fatty acid/lipid metabolism, amino acid synthesis, DNA and RNA replication, DNA ligation, and repair, among others. In this context, pyrophosphatase activity (conversion of inorganic pyrophosphate (PPi) to two orthophosphate ions (Pi)) is critical to cellular function by increasing favorable reaction equilibrium and driving reaction direction/irreversibility of coupled reactions with a second enzyme.

[0005] Many cellular enzymes are capable of catalyzing pyrophosphate conversion; however, the activity of most of these enzymes is not specific to a pyrophosphate substrate. For example, alkaline phosphatase (phoA in *E. coli*) will fully degrade pyrophosphate but will also hydrolyze other phosphates including the 5'-phosphates of cellular nucleotides. Accordingly, alkaline phosphatases may be incompatible with the cytoplasm and are generally secreted to the periplasm or extracellular space. For similar reasons, enzymes like alkaline phosphatase may not be compatible with some *in vitro* or industrial processes.

[0006] However, inorganic pyrophosphatases (IPP enzymes or IPPases) are ubiquitous in nature and are known to have both high activity in the conversion of pyrophosphate to orthophosphate and relatively low activity on other phosphorylated substrates (e.g. NTP, sugar phosphates, etc.). The most well characterized examples of inorganic pyrophosphatases (type I IPP enzymes) have been known since at least the 1960's and have been found in almost all cells from humans to bacteria. Enzymes from this class generally form homo-hexameric structures and are usually activated with magnesium. The type I IPP enzyme in *E. coli* (ppa) is also known to be essential to cell growth. Type I IPP enzymes have been used commercially for decades in coupled reactions to facilitate DNA/RNA polymerase and other reactions that produce pyrophosphate (McMillan et al. Appl Environ Microbiol. 2015, 82(2):538-548).

[0007] More recently, other pyrophosphate selective IPP enzymes have been identified. One example is the class of transmembrane proton-pumping pyrophosphatases (H(+)-PPases). Unlike type I IPP enzymes, which drive the hydrolysis of pyrophosphate to release heat, the H(+)-PPases couple the energy of inorganic pyrophosphate hydrolysis to proton movement across biological membranes. However, these enzymes are not readily conducive to *in vitro* and industrial processes. Additionally, two other classes of IPP enzymes (type II and type III) have also been characterized in recent years.

[0008] Type II IPP enzymes are much less common than type I IPP enzymes and have generally been characterized from bacterial and archaeal sources (see for example, Zyryanov et al. Biochemistry. 2004;43(4):1065-1074, Gajadeera et al. J Struct Biol. 2015;189(2):81-86, Parfenyev et al. J Biol Chem. 2001;276(27):24511-24518). There is some indication in the literature that they may prefer manganese and/or cobalt in their active sites (Zyryanov et al. Biochem J., 2002, 367(Pt 3):901-906), may have higher specific activity than type I enzymes, and may be more selective toward pyrophosphate and have lower activity toward other substrates, such as NTPs, compared to type I IPP enzymes (Baykov et al. FEBS Lett., 2017, 591(20):3225-3234). Finally, type III enzymes are derived from the haloalkanoate dehalogenase superfamily (HADSF) protein fold (see for example, Huang et al. Biochemistry, 2011, 50(41):8937-8949). The HADSF family is extremely diverse and includes enzymes that act as dehalogenases, phosphomutases, and phosphatases, among others. A very small number of type III IPP enzymes have been reported to act as pyrophosphate selective phosphatases, though potentially with unique kinetics compared to other known IPP enzymes.

[0009] Type I IPP enzymes have been used in a coupled reaction with a terminal nucleotidyl transferase (TnT) enzyme (PCT/US2023/076667, published as WO2024081770). However, limitations of type I IPP enzymes were observed. For example, type I IPP enzymes are known to have low level off-target hydrolysis activity toward nucleotides (e.g. releasing the gamma phosphate from NTP).

[0010] Therefore, IPP enzymes with higher specific activity, reduced off-target hydrolysis and other side reactions, increased thermostability, and increased solubility compared to type I IPP enzymes are

necessary to enable a variety of commercial, research, and industrial processes. In particular, IPP enzymes with improved properties would benefit coupled or paired reactions using nonnatural industrial process conditions.

SUMMARY

[0011] The present disclosure provides compositions and methods of use for type II IPP enzymes in a variety of commercial, industrial, and research processes.

[0012] In particular, the disclosure provides compositions and methods of use for IPP enzymes with improved properties under industrial process conditions and in a coupled or paired reaction with a second enzyme. Additionally, the present disclosure provides inorganic pyrophosphatases identified with improved solubility, improved thermostability, improved pyrophosphatase activity, decreased off-target hydrolysis, decreased pyrophosphorolysis by a second enzyme, decreased byproducts in a reaction with a second enzyme, and/or increased ratio of products to byproducts in a reaction with a second enzyme.

[0013] In some embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes or variants of type I IPP enzymes or variants of type II IPP enzymes and a phosphorylated compound. In some embodiments, the phosphorylated compound is natural or modified nucleoside triphosphate or a natural or modified nucleoside triphosphate analog.

[0014] In some embodiments, the method relates to use of an inorganic pyrophosphatase for increasing forward reaction in a reaction generating inorganic pyrophosphate, the method comprising providing an inorganic pyrophosphatase for converting inorganic pyrophosphate to phosphate in a reaction generating inorganic pyrophosphate as a product, wherein the inorganic pyrophosphatase comprises a type II inorganic pyrophosphatase, a type I inorganic pyrophosphatase, or a variant of type II or type I inorganic pyrophosphatase.

[0015] In some embodiments, the type II inorganic pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or to a reference sequence corresponding to SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

[0016] In some embodiments, the inorganic pyrophosphatase comprises an amino acid sequence comprising residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or an amino acid sequence comprising SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

[0017] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%,

83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0018] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0019] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0020] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 14, 45, 81, 85, 86, 89, 110, 160, 169, 177, 187, 188, 231, 269, or 308, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0021] In some embodiments, the amino acid sequence of the inorganic pyroposphatase comprises residues 12 to carboxy terminal of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or comprises SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0022] In some embodiments, the type I pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to

SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16, without the carboxy terminal histidine tag, or to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16.

[0023] In some embodiments, the inorganic pyrophosphatase comprises a type I pyrophosphatase comprising the sequence comprising SEQ ID NO: 4, 6, 8, 10, 12, or 16, or comprising the sequence comprising SEQ ID NO: 4, 6, 8, 10, 12, or 16 without the carboxy-terminal histidine tag.

[0024] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0025] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10, without the carboxy terminal histidine tag residues.

[0026] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues.

[0027] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 41, 65, 95, or 147, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0028] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, or comprises SEQ ID NO: 18, 20, 22, 24, or 26.

[0029] In some other embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes or variants of type I IPP enzymes or variants of type II IPP enzymes and a second enzyme. In some embodiments, the second enzyme produces a substrate for the IPP enzyme. In some embodiments, the second enzyme uses a natural or modified nucleoside triphosphate or a natural or modified nucleoside triphosphate analog as a substrate. In some embodiments, the second enzyme catalyzes a substrate to produce inorganic pyrophosphate. In some embodiments, the second enzyme comprises a nucleotidyl transferase, a polynucleotide ligase, a polynucleotide polymerase, or a tRNA synthetase.

[0030] In some embodiments, the second enzyme comprises a terminal nucleotidyl transferase, poly(U)polymerase or poly(A)polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, or a Polθ polymerase.

[0031] In some embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes or variants of type I IPP enzymes or variants of type II IPP enzymes in a paired extension reaction with a template independent terminal nucleotidyl transferase (or polymerase).

[0032] In any of the embodiments described herein, the composition or method may comprise a nonnatural or industrial process condition.

[0033] In another embodiment, the present disclosure provides an engineered polynucleotide encoding at least one IPP enzyme polypeptide described in the above paragraphs.

[0034] The present disclosure further provides vectors comprising at least one engineered polynucleotide, vectors, host cells, and methods of producing an IPP enzyme polypeptide.

DESCRIPTION OF THE INVENTION

[0035] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Generally, the nomenclature used herein and the laboratory procedures of cell culture, molecular genetics, microbiology, organic chemistry, analytical chemistry and nucleic acid chemistry described below are those well-known and commonly employed in the art. Such techniques are well-known and described in numerous texts and reference works well known to those of skill in the art. Standard techniques, or modifications thereof, are used for chemical syntheses and chemical analyses. All patents, patent applications, articles and publications mentioned herein, both *supra* and *infra*, are hereby expressly incorporated herein by reference.

[0036] Although any suitable methods and materials similar or equivalent to those described herein find use in the practice of the present invention, some methods and materials are described herein. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art. Accordingly, the terms defined immediately below are more fully described by reference to the invention as a whole.

[0037] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the present invention. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described. Numeric ranges are inclusive of the numbers defining the range. Thus, every numerical range disclosed herein is intended to encompass every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. It is also intended that every maximum (or minimum) numerical limitation disclosed herein includes every lower (or higher) numerical limitation, as if such lower (or higher) numerical limitations were expressly written herein.

[0038] As used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to "a polypeptide" includes more than one polypeptide. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0039] It is to be understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of." It is to be further understood that where descriptions of various embodiments use the term "optional" or "optionally" the subsequently described event or circumstance may or may not occur, and that the description includes instances where the event or circumstance occurs and instances in which it does not. It is to be understood that both the foregoing general description, and the following detailed description are exemplary and explanatory only and are not restrictive of this disclosure. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described.

Abbreviations

[0040] The abbreviations used for the genetically encoded amino acids are conventional and are as follows:

Amino Acid	Three-Letter Abbreviation	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R

Amino Acid	Three-Letter Abbreviation	One-Letter Abbreviation
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

[0041] When the three-letter abbreviations are used, unless specifically preceded by an "L" or a "D" or clear from the context in which the abbreviation is used, the amino acid may be in either the L- or D-configuration about α-carbon (C_α). For example, whereas "Ala" designates alanine without specifying the configuration about the α-carbon, "D-Ala" and "L-Ala" designate D-alanine and L-alanine, respectively.

[0042] When the one-letter abbreviations are used, upper case letters designate amino acids in the L-configuration about the α-carbon and lower-case letters designate amino acids in the D-configuration about the α-carbon. For example, "A" designates L-alanine and "a" designates D-alanine. When polypeptide sequences are presented as a string of one-letter or three-letter abbreviations (or mixtures thereof), the sequences are presented in the amino (N) to carboxy (C) direction in accordance with common convention.

[0043] The abbreviations used for the genetically encoding nucleosides are conventional and are as follows: adenosine (A); guanosine (G); cytidine (C); thymidine (T); and uridine (U). These abbreviations are also used interchangeably for nucleosides and nucleotides (nucleosides with one or more phosphate groups). Unless specifically delineated, the abbreviated nucleosides or nucleotides may be either ribonucleosides (or ribonucleotides) or 2'-deoxyribonucleosides (or 2'-deoxyribonucleotides). The nucleosides or nucleotides may also be modified at the 3' position. The

nucleosides or nucleotides may be specified as being either ribonucleosides (or ribonucleotides) or 2'-deoxyribonucleosides (or 2'-deoxyribonucleotides) on an individual basis or on an aggregate basis. The abbreviation NTP may be used to specify both ribonucleosides (or ribonucleotides) and 2'-deoxyribonucleosides (or 2'-deoxyribonucleotides) or only ribonucleosides (or ribonucleotides), depending on the context. Similarly, the abbreviations rNTP for ribonucleosides (or ribonucleotides) or 2'-deoxyribonucleosides (or 2'-deoxyribonucleotides) dNTP may also be used. When nucleic acid sequences are presented as a string of one-letter abbreviations, the sequences are presented in the 5' to 3' direction in accordance with common convention, and the phosphates are not indicated.

Definitions

[0044] In reference to the present invention, the technical and scientific terms used in the descriptions herein will have the meanings commonly understood by one of ordinary skill in the art, unless specifically defined otherwise. Accordingly, the following terms are intended to have the following meanings.

[0045] "EC" number refers to the Enzyme Nomenclature of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). The IUBMB biochemical classification is a numerical classification system for enzymes based on the chemical reactions they catalyze.

[0046] "ATCC" refers to the American Type Culture Collection whose biorepository collection includes genes and strains.

[0047] "NCBI" refers to National Center for Biological Information and the sequence databases provided therein.

[0048] "Protein," "polypeptide," and "peptide" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids, as well as polymers comprising D- and L-amino acids, and mixtures of D- and L-amino acids.

[0049] "Amino acids" are referred to herein by either their commonly known three-letter symbols or by the one-letter symbols recommended by IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single letter codes.

[0050] As used herein, "polynucleotide," "oligonucleotide," and "nucleic acid'' are used interchangeably herein and refer to two or more nucleosides or nucleotides that are covalently linked together. The polynucleotide may be wholly comprised of ribonucleotides (i.e., RNA), wholly comprised of 2' deoxyribonucleotides (i.e., DNA), wholly comprised of other synthetic nucleotides or comprised of mixtures of synthetic, ribo- and/or 2' deoxyribonucleotides. The polynucleotides may

also include modified nucleotides with substitutions, including 2' substitutions (e.g., 2'-fluoro, 2'-O-methyl, 2'-O-methoxyethyl, locked or constrained ethyl modifications, and others known to those skilled in the art). Nucleosides will be linked together via standard phosphodiester linkages or via one or more non-standard linkages, including but not limited to phosphorothioate linkages. The polynucleotide may be single-stranded or double-stranded or may include both single-stranded regions and double-stranded regions. Moreover, while a polynucleotide will typically be composed of the naturally occurring encoding nucleobases (i.e., adenine, guanine, uracil, thymine and cytosine), it may include one or more modified and/or synthetic nucleobases, such as, for example, inosine, xanthine, hypoxanthine, etc. In some embodiments, such modified or synthetic nucleobases are nucleobases encoding amino-acid sequences. Nucleobases that are modified or synthetic may comprise any known or hypothetical or future discovered modification or structure that would be recognized by one of skill in the art as a modified or synthetic nucleobase. Similarly, the terms "polynucleotide," "oligonucleotide," and "nucleic acid'' are intended to comprise any modified or synthetic structure that is now known or discovered in the future that would be recognized by one of skill in the art as being or having the function of a "polynucleotide," "oligonucleotide," or "nucleic acid.'' An example of a modified or synthetic structure having the function of a "polynucleotide," "oligonucleotide," or "nucleic acid'' is PNA or peptide nucleic acid.

[0051] As used herein, "oligo acceptor substrate" and "acceptor substrate" and "growing oligo acceptor substrate strand" and "growing oligonucleotide chain" and "growing polynucleotide strand" are used interchangeably herein and refer to any oligo or nucleotide chain or similar moiety with an exposed 3'-OH or equivalent thereof that may be recognized by a wild-type TnT or polymerase or an engineered TnT or template independent polymerase as a substrate for nucleoside addition or synthesis. In some embodiments, the acceptor substrate may be single stranded. In yet other embodiments, the acceptor substrate may be double stranded or partially doubled stranded. In some embodiments, the acceptor substrate may comprise a nucleotide chain consisting of 1-10 nucleotides, 5-20 nucleotides, 15-50 nucleotides, 30-100 nucleotides, or greater than 100 nucleotides. In some embodiments, the acceptor substrate may comprise a chemical moiety that is not a nucleotide chain but contains a free -OH capable of being recognized as a substrate by a wild-type or engineered TnT, referred to herein as a "3'-OH equivalent".

[0052] As used herein, "nucleoside triphosphate-3'-O-removable blocking group" and "nucleotide triphosphate-3'-O-removable blocking group" and "reversible terminator" and "NTP-3'-O-RBG" are used interchangeably herein and refer to a ribonucleoside triphosphate or a deoxyribonucleoside triphosphate or a synthetic or nucleoside triphosphate composed of an alternate or modified sugar with a removable blocking group attached at the 3' position of the sugar moiety. An NTP-3'-O-RBG may also include other modifications as described herein, including but not limited to modifications at the 2' position, modifications to the nucleobase, and modifications to the phosphates. A nucleotide

may also have a 3'-O-RBG, as is expected after reaction of an NTP-3'-O-RBG with an engineered TnT of the present disclosure and an oligo acceptor substrate.

[0053] As used herein, "oligo acceptor product" and "growing oligonucleotide chain" and "oligo acceptor extension product" are used interchangeably herein and refer to the product of a NTP-3'-O-RBG or other natural or modified NTP substrate and an oligo acceptor substrate, wherein a TnT or related polymerase has catalyzed the extension or addition of a nucleotide-3'-O-RBG or other natural or modified nucleotide substrate to an oligo acceptor substrate via reaction with one or more NTP-3'-O-RBGs or other natural or modified NTP substrates.

[0054] As used herein, "removable blocking group" and "blocking group" and "terminator group" and "reversible terminating group" and "inhibitor group" and related variations of these terms are used interchangeably herein and refer to a chemical group that would hinder addition of a second NTP-3'-O-RBG or other natural or modified NTP substrate to the 3' end of the growing oligo acceptor substrate strand prior to removal of the removable blocking from the first round of addition. In some embodiments, the NTP-3'-O-RBG or other natural or modified NTP substrate may comprise a removable blocking group selected from the group consisting of NTP-3'-O-NH_2, or NTP-3'-O-PO_3. In some embodiments, the NTP- 3'-O-RBG or other natural or modified NTP substrate may have a natural purine or pyrimidine base, such as adenine, guanine, cytosine, thymine, or uridine. In some embodiments, NTP- 3'-O-RBG or other natural or modified NTP substrates may have an unnatural base analog such as inosine, xanthine, hypoxanthine or another base analog, as is known in the art. In some embodiments the blocking group may comprise or may additionally comprise a modification at the 2' position.

[0055] As used herein, "template independent synthesis" refers to synthesis of an oligonucleotide or a polynucleotide without the use of template strand as a guide for synthesis of a complementary oligo or polynucleotide strand. Thus, template independent synthesis refers to an iterative process, whereby, successive nucleotides are added to a growing oligo or nucleotide chain or acceptor substrate. Template independent synthesis may be in a sequence defined manner or may be random, as is the case with the wild-type TdT in creating antigen receptor diversity.

[0056] "Coding sequence" refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0057] "Naturally-occurring" or "wild-type" refers to the form found in nature. For example, a naturally occurring or wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

[0058] As used herein, "recombinant," "engineered," "variant," and "nonnaturally occurring" when used with reference to a cell, nucleic acid, or polypeptide, refer to a material, or a material

corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature. In some embodiments, the cell, nucleic acid or polypeptide is identical to a naturally occurring cell, nucleic acid or polypeptide, but is produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or expressed native genes that are otherwise expressed at a different level.

[0059] "Percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides or polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence for optimal alignment of the two sequences. The percentage may be calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Alternatively, the percentage may be calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Those of skill in the art appreciate that there are many established algorithms available to align two sequences. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman (Smith and Waterman, Adv. Appl. Math., 1981, 2:482), by the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, J. Mol. Biol., 1970, 48:443), by the search for similarity method of Pearson and Lipman (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 1988, 85:2444), by computerized implementations of these algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package), or by visual inspection, as known in the art. Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity include, but are not limited to the BLAST and BLAST 2.0 algorithms, which are described by Altschul et al. (See, Altschul et al., J. Mol. Biol., 1990, 215: 403-410; and Altschul et al., Nucl. Acids Res., 1977, 3389-3402, respectively). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as, the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word

hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSum62 scoring matrix (See, Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA, 1989, 89:10915). Exemplary determination of sequence alignment and % sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using default parameters provided.

[0060] "Reference sequence" refers to a defined sequence used as a basis for a sequence comparison. Similarly, a "reference polypeptide" or "reference inorganic pyrophosphatase" is understood to be a polypeptide encoded by a reference sequence. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, or the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a sequence (*i.e.*, a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptide are typically performed by comparing sequences of the two polynucleotides or polypeptides over a "comparison window" to identify and compare local regions of sequence similarity. In some embodiments, a "reference sequence" can be based on a primary amino acid sequence, where the reference sequence is a sequence that can have one or more changes in the primary sequence. By way of example and not limitation, a "reference sequence based on SEQ ID NO: 46 having at the residue corresponding to X14 a leucine" or X14L refers to a reference sequence in which the corresponding residue at X14 in SEQ ID NO: 46, which is a tyrosine, has been changed to leucine.

[0061] "Comparison window" refers to a conceptual segment of contiguous nucleotide positions or amino acids residues wherein a sequence may be compared to a reference sequence. In some embodiments, the comparison window is at least 15 to 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (i.e.,

gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. In some embodiments, the comparison window can be longer than 15-20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

[0062] As used herein, "substantial identity" refers to a polynucleotide or polypeptide sequence that has at least 80 percent sequence identity, at least 85 percent identity, at least between 89 to 95 percent sequence identity, or more usually, at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 residue positions, frequently over a window of at least 30-50 residues, wherein the percentage of sequence identity is calculated by comparing the reference sequence to a sequence that includes deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. In some specific embodiments applied to polypeptides, the term "substantial identity" means that two polypeptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 89 percent sequence identity, at least 95 percent sequence identity or more (e.g., 99 percent sequence identity). In some embodiments, residue positions that are not identical in sequences being compared differ by conservative amino acid substitutions.

[0063] "Corresponding to," "reference to," and "relative to" when used in the context of the numbering of a given amino acid or polynucleotide sequence refer to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered IPP enzyme, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned.

[0064] "Mutation" refers to the alteration of a nucleic acid sequence. In some embodiments, mutations result in changes to the encoded polypeptide sequence (i.e., as compared to the original sequence without the mutation). In some embodiments, the mutation comprises a substitution, such that a different amino acid is produced. In some alternative embodiments, the mutation comprises an addition, such that an amino acid is added (e.g., insertion) to the original polypeptide sequence. In some further embodiments, the mutation comprises a deletion, such that an amino acid is deleted from the original polypeptide sequence. Any number of mutations may be present in a given sequence.

[0065] "Amino acid difference" or "residue difference" refers to a change in the amino acid residue at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in a reference sequence. The positions of amino acid differences generally are referred to herein as "Xn," where n refers to the corresponding position in the reference sequence upon which the residue

difference is based. For example, a "residue difference at position X14 as compared to SEQ ID NO: 46" refers to a change of the amino acid residue at the polypeptide position corresponding to position 14 of SEQ ID NO: 46. Thus, if the reference polypeptide of SEQ ID NO: 46 has a tyrosine at position 14, then a "residue difference at position X14 as compared to SEQ ID NO: 46" refers to an amino acid substitution of any residue other than tyrosine at the position of the polypeptide corresponding to position 14 of SEQ ID NO: 46. In most instances herein, the specific amino acid residue difference at a position is indicated as "XnY" where "Xn" specified the corresponding position as described above, and "Y" is the single letter identifier of the amino acid found in the engineered polypeptide (i.e., the different residue than in the reference polypeptide). In some embodiments, more than one amino acid can appear in a specified residue position (i.e., the alternative amino acids can be listed in the form XnY/Z, where Y and Z represent alternate amino acid residues). In some instances (e.g., in **Tables 5**, **9**, **15**, **21**, and **23**) the present invention also provides specific amino acid differences denoted by the conventional notation "AnB", where A is the single letter identifier of the residue in the reference sequence, "n" is the number of the residue position in the reference sequence, and B is the single letter identifier of the residue substitution in the sequence of the engineered polypeptide. Furthermore, in some instances, a polypeptide of the present invention can include one or more amino acid residue differences relative to a reference sequence, which is indicated by a list of the specified positions where changes are made relative to the reference sequence. In some additional embodiments, the present invention provides engineered polypeptide sequences comprising both conservative and non-conservative amino acid substitutions.

[0066] "Amino acid substitution set" and "substitution set" refers to a group of amino acid differences, i.e., substitutions, within a polypeptide sequence. In some embodiments, substitution sets comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more amino acid substitutions. In some embodiments, a substitution set refers to the set of amino acid substitutions that is present in any of the inorganic pyrophosphatase polypeptides listed in any of the Tables in the Examples. In some embodiments, the amino acid sequence comprises at least each of the amino acid substitutions in the referenced substitution set. In the substitution sets, the individual substitutions are separated by a semicolon (";"; e.g., K81R;D86S) or slash ("/"; e.g., K81R/D86S or 81R/86S).

[0067] As used herein, "conservative amino acid substitution" refers to a substitution of a residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar defined class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain is substituted with another aliphatic amino acid (e.g., alanine, valine, leucine, and isoleucine); an amino acid with an hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain (e.g., serine and threonine); an amino acid having an aromatic side chain is substituted with another amino acid having an aromatic side chain (e.g., phenylalanine, tyrosine, tryptophan, and histidine); an amino acid with a basic side chain is substituted with another amino acid with a basic side chain (e.g., lysine and

arginine); an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain (e.g., aspartic acid or glutamic acid); and/or a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively. Exemplary conservative substitutions are provided in **Table 1.1** below.

Table 1.1. Conservative Amino Acid Substitution Examples

Residue	Possible Conservative Substitutions
A, L, V, I	Other aliphatic (A, L, V, I) Other non-polar (A, L, V, I, G, M)
G, M	Other non-polar (A, L, V, I, G, M)
D, E	Other acidic (D, E)
K, R	Other basic (K, R)
N, Q, S, T	Other polar
H, Y, W, F	Other aromatic (H, Y, W, F)
C, P	None

[0068] "Non-conservative substitution" refers to substitution of an amino acid in the polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (*e.g.*, proline for glycine), (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

[0069] "Deletion" refers to modification to the polypeptide by removal of one or more amino acids from the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the reference enzyme while retaining enzymatic activity and/or retaining the improved properties of an engineered IPP enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide. In various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

[0070] "Insertion" refers to modification to the polypeptide by addition of one or more amino acids from the reference polypeptide. In some embodiments, the improved engineered IPP enzymes comprise insertions of one or more amino acids to the naturally occurring polypeptide as well as insertions of one or more amino acids to other improved IPP polypeptides. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

[0071] "Fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can be at least 14 amino acids long, at least 20 amino acids long, at least 50 amino acids long or longer, and up to 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% of the full-length IPP polypeptide, for example the polypeptide of SEQ ID NO: 2 or an IPP enzyme provided in the even-numbered sequences of SEQ ID NO: 2-90.

[0072] "Isolated polypeptide" refers to a polypeptide which is substantially separated from other contaminants that naturally accompany it, *e.g.*, protein, lipids, and polynucleotides. The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (*e.g.*, host cell or *in vitro* synthesis). The engineered IPP enzymes may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the engineered IPP enzyme can be an isolated polypeptide.

[0073] "Substantially pure polypeptide" refers to a composition in which the polypeptide species is the predominant species present (*i.e.*, on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure IPP composition will comprise about 60 % or more, about 70% or more, about 80% or more, about 90% or more, about 95% or more, and about 98% or more of all macromolecular species by mole or % weight present in the composition. In some embodiments, the object species is purified to essential homogeneity (*i.e.*, contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated engineered IPP polypeptide is a substantially pure polypeptide composition.

[0074] As used herein, "improved enzyme property" refers to at least one improved property of an enzyme. In some embodiments, the present invention provides engineered IPP enzyme polypeptides that exhibit an improvement in any enzyme property as compared to a reference IPP enzyme polypeptide and/or a wild-type IPP enzyme polypeptide, and/or another engineered IPP enzyme polypeptide. For the engineered IPP enzyme polypeptides described herein, the comparison is generally made to the wild-type enzyme from which the IPP enzyme is derived, although in some embodiments, the reference enzyme can be another improved engineered IPP enzyme. Thus, the level of "improvement" can be determined and compared between various IPP enzyme polypeptides, including wild-type, as well as engineered IPP enzymes. Improved properties include, but are not

limited, to such properties as enzymatic activity, thermostability, solvent stability, pH activity profile, cofactor requirements, refractoriness to inhibitors (*e.g.*, substrate or product inhibition), activity at elevated temperatures, increased soluble expression, decreased by-product formation, increased specific activity on PPi substrates, and/or increased activity in a paired or coupled reaction with a second enzyme.

[0075] "Increased enzymatic activity" refers to an improved property of the IPP enzyme polypeptides, which can be represented by an increase in specific activity (*e.g.*, product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (*e.g.*, percent conversion of starting amount of substrate to product in a specified time period using a specified amount of IPP enzyme) as compared to the reference IPP enzyme. Exemplary methods to determine enzyme activity are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m, V_{max} or k_{cat}, changes of which can lead to increased enzymatic activity. Improvements in enzyme activity can be from about 1.2 times the enzymatic activity of the corresponding wild-type enzyme, to as much as 2 times, 5 times, 10 times, 20 times, 25 times, 50 times or more enzymatic activity than the naturally occurring or another engineered IPP enzyme from which the IPP enzyme polypeptides were derived. IPP enzyme activity can be measured by any one of standard assays, such as by monitoring changes in properties of substrates, cofactors, or products. In some embodiments, the amount of products generated can be measured by Liquid Chromatography-Mass Spectrometry (LC-MS), HPLC, or other methods, as known in the art. Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0076] "Conversion" refers to the enzymatic conversion of the substrate(s) to the corresponding product(s). "Percent conversion" refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, the "enzymatic activity" or "activity" of a IPP enzyme polypeptide can be expressed as "percent conversion" of the substrate to the product.

[0077] "Thermostable" refers to a polypeptide that maintains similar activity (more than 60% to 80% for example) after exposure to elevated temperatures (*e.g.*, 40-80 °C) for a period of time (*e.g.*, 0.5-24 hrs) compared to the wild-type enzyme exposed to the same elevated temperature.

[0078] "Solvent stable" refers to a polypeptide that maintains similar activity (more than *e.g.*, 60% to 80%) after exposure to varying concentrations (*e.g.*, 5-99%) of solvent (ethanol, isopropyl alcohol, dimethylsulfoxide (DMSO), tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butyl acetate, methyl tert-butyl ether, etc.) for a period of time (*e.g.*, 0.5-24 hrs) compared to the wild-type enzyme exposed to the same concentration of the same solvent.

[0079] "Thermo- and solvent stable" refers to a polypeptide that is both thermostable and solvent stable.

[0080] The term "stringent hybridization conditions" is used herein to refer to conditions under which nucleic acid hybrids are stable. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrids. In general, the stability of a hybrid is a function of ion strength, temperature, G/C content, and the presence of chaotropic agents. The T_m values for polynucleotides can be calculated using known methods for predicting melting temperatures (See e.g., Baldino et al., Meth. Enzymol., 1989, 168:761-777; Bolton et al., Proc. Natl. Acad. Sci. USA, 1962, 48:1390; Bresslauer et al., Proc. Natl. Acad. Sci. USA, 1986, 83:8893-8897; Freier et al., Proc. Natl. Acad. Sci. USA, 1986, 83:9373-9377; Kierzek et al., Biochem., 1986, 25:7840-7846; Rychlik et al., 1990, Nucl. Acids Res., 1990, 18:6409-6412 (erratum, Nucl. Acids Res., 1981, 19:698); Sambrook et al., *supra*); Suggs et al., 1981, *in* Developmental Biology Using Purified Genes, Brown et al., eds., pp. 683-693, Academic Press, Cambridge, MA (1981); and Wetmur, Crit. Rev. Biochem. Mol. Biol., 1991, 26:227-259). In some embodiments, the polynucleotide encodes the polypeptide disclosed herein and hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to the complement of a sequence encoding an engineered IPP enzyme of the present invention.

[0081] "Hybridization stringency" relates to hybridization conditions, such as washing conditions, in the hybridization of nucleic acids. Generally, hybridization reactions are performed under conditions of lower stringency, followed by washes of varying but higher stringency. The term "moderately stringent hybridization" refers to conditions that permit target-DNA to bind a complementary nucleic acid that has about 60% identity, preferably about 75% identity, about 85% identity to the target DNA, with greater than about 90% identity to target-polynucleotide. Exemplary moderately stringent conditions are conditions equivalent to hybridization in 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42 °C, followed by washing in 0.2×SSPE, 0.2% SDS, at 42 °C. "High stringency hybridization" refers generally to conditions that are about 10 °C or less from the thermal melting temperature T_m as determined under the solution condition for a defined polynucleotide sequence. In some embodiments, a high stringency condition refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65 °C (*i.e.*, if a hybrid is not stable in 0.018M NaCl at 65 °C, it will not be stable under high stringency conditions, as contemplated herein). High stringency conditions can be provided, for example, by hybridization in conditions equivalent to 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42 °C, followed by washing in 0.1×SSPE, and 0.1% SDS at 65 °C. Another high stringency condition is hybridizing in conditions equivalent to hybridizing in 5X SSC containing 0.1% (w:v) SDS at 65 °C and washing in 0.1x SSC containing 0.1% SDS at 65 °C. Other high stringency hybridization conditions, as well as moderately stringent conditions, are described in the references cited above.

[0082] "Heterologous" polynucleotide refers to any polynucleotide that is introduced into a host cell by laboratory techniques and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0083] "Codon optimized" refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is efficiently expressed in the organism of interest. Although the genetic code is degenerate in that most amino acids are represented by several codons, called "synonyms" or "synonymous" codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the IPP enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0084] As used herein, "preferred, optimal, high codon usage bias codons" refers interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (See e.g., GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, Peden, University of Nottingham; McInerney, Bioinform., 1998, 14:372-73; Stenico et al., Nucl. Acids Res., 1994, 222437-46; Wright, Gene, 1990, 87:23-29). Codon usage tables are available for many different organisms (See e.g., Wada et al., Nucl. Acids Res., 1992, 20:2111-2118; Nakamura et al., Nucl. Acids Res., 2000, 28:292; Duret, et al., supra; Henaut and Danchin, *in* *Escherichia coli* and *Salmonella*, Neidhardt, et al., eds., ASM Press, Washington D.C., p. 2047-2066 (1996)). The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTS), or predicted coding regions of genomic sequences (See e.g., Mount, Bioinformatics: Sequence and Genome Analysis, Chapter 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2001); Uberbacher, Meth. Enzymol., 1996, 266:259-281; and Tiwari et al., Comput. Appl. Biosci., 1997, 13:263-270).

[0085] "Control sequence" is defined herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

[0086] "Operably linked" is defined herein as a configuration in which a control sequence is appropriately placed (*i.e.*, in a functional relationship) at a position relative to a polynucleotide of interest such that the control sequence directs or regulates the expression of the polynucleotide and/or polypeptide of interest.

[0087] "Promoter sequence" refers to a nucleic acid sequence that is recognized by a host cell for expression of a polynucleotide of interest, such as a coding sequence. The promoter sequence contains transcriptional control sequences, which mediate the expression of a polynucleotide of interest. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[0088] "Suitable reaction conditions" refer to those conditions in the biocatalytic reaction solution (*e.g.*, ranges of enzyme loading, substrate loading, cofactor loading, temperature, pH, buffers, co-solvents, etc.) under which a IPP enzyme polypeptide of the present invention is capable of converting one or more substrate compounds to a product compound (*e.g.*, conversion of PPi to Pi, optionally, in a paired reaction with a second enzyme). Exemplary "suitable reaction conditions" are provided in the present invention and illustrated by the Examples.

[0089] "Composition" refers to a mixture or combination of one or more substances, wherein each substance or component of the composition retains its individual properties. As used herein, a biocatalytic composition refers to a combination of one or more substances useful for biocatalysis.

[0090] "Loading", such as in "compound loading" or "enzyme loading" or "cofactor loading" refers to the concentration or amount of a component in a reaction mixture at the start of the reaction.

[0091] "Substrate" in the context of a biocatalyst mediated process refers to the compound or molecule acted on by the biocatalyst. For example, a IPP enzyme biocatalyst used in the synthesis processes disclosed herein acts on a phosphorylated substrate, such as inorganic pyrophosphate.

[0092] "Product" in the context of a biocatalyst mediated process refers to the compound or molecule resulting from the action of the biocatalyst. For example, an exemplary product for a IPP enzyme biocatalyst used in a process disclosed herein is orthophosphate, as depicted in Scheme 1.

[0093] "Alkyl" refers to saturated hydrocarbon groups of from 1 to 18 carbon atoms inclusively, either straight chained or branched, more preferably from 1 to 8 carbon atoms inclusively, and most preferably 1 to 6 carbon atoms inclusively. An alkyl with a specified number of carbon atoms is denoted in parenthesis (e.g., $(C_1\text{-}C_6)$alkyl refers to an alkyl of 1 to 6 carbon atoms).

[0094] "Alkenyl" refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one double bond but optionally containing more than one double bond.

[0095] "Alkynyl" refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one triple bond but optionally containing more than one triple bond, and additionally optionally containing one or more double bonded moieties.

[0096] "Heteroalkyl, "heteroalkenyl," and heteroalkynyl," refer respectively, to alkyl, alkenyl and alkynyl as defined herein in which one or more of the carbon atoms are each independently replaced with the same or different heteroatoms or heteroatomic groups. Heteroatoms and/or heteroatomic groups which can replace the carbon atoms include, but are not limited to -O-, -S-, -S-O-, $-NR^g$-, -PH-, -S(O)-, -S(O)2-, -S(O) NR^g-, $-S(O)_2NR^g$, and the like, including combinations thereof, where each R^g is independently selected from hydrogen, alkyl, cycloalkyl, heterocycloalkyl, aryl, and heteroaryl.

[0097] "Amino" refers to the group $-NH_2$. Substituted amino refers to the group $-NHR^h$, NR^hR^h, and $NR^hR^hR^h$, where each R^h is independently selected from substituted or unsubstituted alkyl, cycloalkyl, cycloheteroalkyl, alkoxy, aryl, heteroaryl, heteroarylalkyl, acyl, alkoxycarbonyl, sulfanyl, sulfinyl, sulfonyl, and the like. Typical amino groups include, but are limited to, dimethylamino, diethylamino, trimethylammonium, triethylammonium, methylysulfonylamino, furanyl-oxy-sulfamino, and the like.

[0098] "Aminoalkyl" refers to an alkyl group in which one or more of the hydrogen atoms are replaced with one or more amino groups, including substituted amino groups.

[0099] "Aminocarbonyl" refers to $-C(O)NH_2$. Substituted aminocarbonyl refers to $-C(O)NR^hR^h$, where the amino group NR^hR^h is as defined herein.

[0100] "Oxy" refers to a divalent group -O-, which may have various substituents to form different oxy groups, including ethers and esters.

[0101] "Alkoxy" or "alkyloxy" are used interchangeably herein to refer to the group $-OR^z$, wherein R^z is an alkyl group, including optionally substituted alkyl groups.

[0102] "Carboxy" refers to -COOH.

[0103] "Carbonyl" refers to -C(O)-, which may have a variety of substituents to form different carbonyl groups including acids, acid halides, aldehydes, amides, esters, and ketones.

[0104] "Carboxyalkyl" refers to an alkyl in which one or more of the hydrogen atoms are replaced with one or more carboxy groups.

[0105] "Aminocarbonylalkyl" refers to an alkyl substituted with an aminocarbonyl group, as defined herein.

[0106] "Halogen" or "halo" refers to fluoro, chloro, bromo and iodo.

[0107] "Haloalkyl" refers to an alkyl group in which one or more of the hydrogen atoms are replaced with a halogen. Thus, the term "haloalkyl" is meant to include monohaloalkyls, dihaloalkyls, trihaloalkyls, etc. up to perhaloalkyls. For example, the expression "(C_1 - C_2) haloalkyl" includes 1-fluoromethyl, difluoromethyl, trifluoromethyl, 1-fluoroethyl, 1,1-difluoroethyl, 1,2-difluoroethyl, 1,1,1 trifluoroethyl, perfluoroethyl, etc.

[0108] "Hydroxy" refers to -OH.

[0109] "Hydroxyalkyl" refers to an alkyl group in which in which one or more of the hydrogen atoms are replaced with one or more hydroxy groups.

[0110] "Thiol" or "sulfanyl" refers to –SH. Substituted thiol or sulfanyl refers to –S-R^h, where R^h is an alkyl, aryl or other suitable substituent.

[0111] "Sulfonyl" refers to –SO_2-. Substituted sulfonyl refers to –SO_2-R^h, where R^h is an alkyl, aryl or other suitable substituent.

[0112] "Alkylsulfonyl" refers to –SO_2-R^z, where R^z is an alkyl, which can be optionally substituted. Typical alkylsulfonyl groups include, but are not limited to, methylsulfonyl, ethylsulfonyl, n-propylsulfonyl, and the like.

[0113] "Phosphate" as used herein refers to a functional group comprised of an orthophosphate ion (phosphorous atom covalently linked to four oxygen atoms). The orthophosphate ion is commonly found with one or more hydrogen atoms or organic groups.

[0114] "Phosphorylated" as used herein refers to the addition or presence of one of more phosphoryl groups (phosphorous atom covalently linked to the three oxygen atoms).

[0115] "Optionally substituted" as used herein with respect to the foregoing chemical groups means that positions of the chemical group occupied by hydrogen can be substituted with another atom (unless otherwise specified) exemplified by, but not limited to carbon, oxygen, nitrogen, or sulfur, or a chemical group, exemplified by, but not limited to, hydroxy, oxo, nitro, methoxy, ethoxy, alkoxy, substituted alkoxy, trifluoromethoxy, haloalkoxy, fluoro, chloro, bromo, iodo, halo, methyl, ethyl, propyl, butyl, alkyl, alkenyl, alkynyl, substituted alkyl, trifluoromethyl, haloalkyl, hydroxyalkyl, alkoxyalkyl, thio, alkylthio, acyl, carboxy, alkoxycarbonyl, carboxamido, substituted carboxamido,

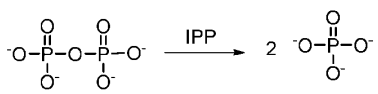
alkylsulfonyl, alkylsulfinyl, alkylsulfonylamino, sulfonamido, substituted sulfonamido, cyano, amino, substituted amino, alkylamino, dialkylamino, aminoalkyl, acylamino, amidino, amidoximo, hydroxamoyl, phenyl, aryl, substituted aryl, aryloxy, arylalkyl, arylalkenyl, arylalkynyl, pyridyl, imidazolyl, heteroaryl, substituted heteroaryl, heteroaryloxy, heteroarylalkyl, heteroarylalkenyl, heteroarylalkynyl, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloalkyl, cycloalkenyl, cycloalkylalkyl, substituted cycloalkyl, cycloalkyloxy, pyrrolidinyl, piperidinyl, morpholino, heterocycle, (heterocycle)oxy, and (heterocycle)alkyl; where preferred heteroatoms are oxygen, nitrogen, and sulfur. Additionally, where open valences exist on these substitute chemical groups they can be further substituted with alkyl, cycloalkyl, aryl, heteroaryl, and/or heterocycle groups, that where these open valences exist on carbon they can be further substituted by halogen and by oxygen-, nitrogen-, or sulfur-bonded substituents, and where multiple such open valences exist, these groups can be joined to form a ring, either by direct formation of a bond or by formation of bonds to a new heteroatom, preferably oxygen, nitrogen, or sulfur. It is further contemplated that the above substitutions can be made provided that replacing the hydrogen with the substituent does not introduce unacceptable instability to the molecules of the present invention and is otherwise chemically reasonable. One of ordinary skill in the art would understand that with respect to any chemical group described as optionally substituted, only sterically practical and/or synthetically feasible chemical groups are meant to be included. "Optionally substituted" as used herein refers to all subsequent modifiers in a term or series of chemical groups. For example, in the term "optionally substituted arylalkyl," the "alkyl" portion and the "aryl" portion of the molecule may or may not be substituted, and for the series "optionally substituted alkyl, cycloalkyl, aryl and heteroaryl," the alkyl, cycloalkyl, aryl, and heteroaryl groups, independently of the others, may or may not be substituted.

[0116] "Reaction" as used herein refers to a process in which one or more substances or compounds or substrates is converted into one or more different substances, compounds, or processes.

Use of Type II and Type I IPP Enzymes

[0117] In recent years, progress in a variety of fields has increased the demand for highly efficient inorganic pyrophosphatase enzymes with improved specific activity and other desirable qualities and reduced off-target hydrolysis and other undesirable qualities. Recently, novel type II IPP enzymes have been discovered that may possess different properties under nonnatural industrial process conditions as compared to widely used type I IPP enzymes.

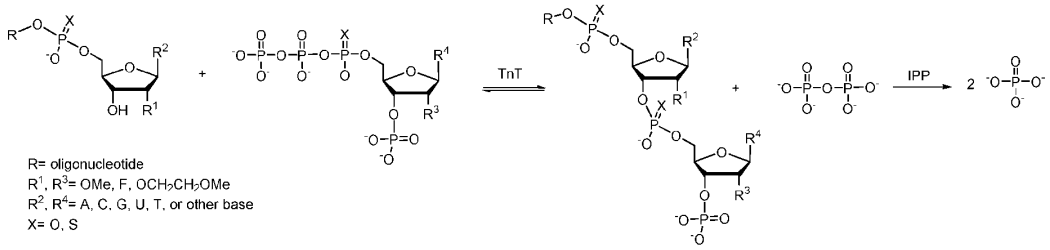
[0118] Inorganic pyrophosphatases or inorganic diphosphatases (EC 3.6.1.1) are ubiquitous and catalyze the conversion of inorganic pyrophosphate (PPi) to two orthophosphate molecules (Pi), as depicted in **Scheme I**.



Scheme 1.

[0119] This exergonic reaction may reduce inhibition caused by accumulation of inorganic phosphate and is often paired or coupled in nature with a less energetically favorable reaction, wherein the IPP enzyme reaction drives the equilibrium of the paired reaction. Examples of this include cofactor (e.g. NAD) biosynthesis, nucleotide biosynthesis, S-adenosyl-L-methionine biosynthesis, sugar metabolism, tRNA charging, fatty acid/lipid metabolism, amino acid synthesis, and DNA and RNA replication. Without being bound by any theory of operation, conversion of the inorganic pyrophosphate to phosphate by the inorganic pyrophosphatase in a paired reaction is thought to render the reaction essentially irreversible (See, e.g., Wimmer et al., Front Microbiol., 2021, 12:759359).

[0120] These paired reactions often involve a NTP cofactor or another substrate that releases inorganic pyrophosphate. An example of a paired reaction is depicted in **Scheme 2**, where the addition of a natural or modified NTP to a growing oligonucleotide chain by a TnT produces inorganic pyrophosphate, that is then degraded to orthophosphate by an IPP enzyme, driving the coupled TnT reaction in the forward direction.



Scheme 2

[0121] A variety of commercially available IPP enzymes and certain type I IPP enzymes are known to be useful in *ex vivo* (e.g., *in vitro*) or industrial applications. These include several IPP enzymes available from New England Biolabs (Ispwich, Massachuesetts): yeast IPP enzyme (catalog # M2403L), *E. coli* IPP enzyme(catalog # M0361L), and thermostable IPP enzyme (catalog # M0296L).

[0122] Recently, International patent application PCT/US2023/076667 (WO 2024/081770) reported uses of type I IPP enzymes in a coupled reaction with a terminal nucleotidyl transferase (TnT) enzyme . As described therein, the TnT catalyzes the addition of a defined modified NTP substrate to the 3'-OH end of an oligo acceptor substrate to sequentially and efficiently create a defined polynucleotide sequence without a complimentary strand or templating primer sequence. Occasionally, undesired byproducts are created by the TnT during the addition step. This includes incorporation of NTPs that have lost their blocking group, addition of more than one NTP, or the excision of a nucleotide or pyrophosphorolysis of the TnT on the growing polynucleotide chain

(where pyrophosphorolysis is defined as the reverse of the nucleotide extension reaction, resulting in reduction of the nucleotide chain by a nucleotide with the concomitant production of an NTP).

[0123] In the reported work, the type I IPP enzyme was used to degrade inorganic pyrophosphate produced by the TnT and drive the reaction toward the nucleotide extension product. The type I IPP enzyme also suppressed pyrophosphorolysis and reduced the synthesis of undesired byproducts of the coupled TnT reaction. In addition, several type I IPP enzymes were demonstrated to have high soluble expression in *E. coli*, good thermostability, pyrophosphate specific activity, and low activity toward other phosphorylated compounds.

[0124] However, limitations of type I IPP enzymes were observed. For example, in some reactions, the type I IPP enzymes were unable to fully suppress pyrophosphorolysis. Even when the type I IPP enzyme was able to suppress this unwanted side reaction, it was sometimes necessary to add the IPP enzyme at stoichiometric ratios of ~1:1 with the TnT catalyst. In addition to increased cost and other logistical complications, less common side reactions of IPP enzymes become more prominent at high enzyme concentration. For example, type I IPP enzymes are known to have low level off-target hydrolysis activity toward nucleotides (e.g. releasing the gamma phosphate from NTP) that is amplified at higher enzyme concentration. Off-target hydrolysis may be further exacerbated by the reaction conditions and buffer in a coupled reaction system (e.g. TnT reactions typically favor cobalt while type I IPP enzymes typically bind magnesium).

[0125] Therefore, IPP enzymes with improved properties as compared to type I IPP enzymes are desirable to increase the efficiency and/or increase yield of desired product in a variety of commercial, research, and industrial processes that generate pyrophosphate. In particular, IPP enzymes with improved properties would benefit coupled or paired reactions using nonnatural industrial process conditions.

[0126] In some embodiments, the present disclosure provides compositions and methods of use for type II enzymes in a variety commercial, industrial, and research processes. Use of these IPP enzymes may overcome known limitations of currently available IPP enzymes. These limitations include reduced or lower specific activity in degrading inorganic phosphate; reduced or lower activity in driving a paired reaction in the forward direction; industrial reaction condition limitations, including a preference or requirement for magnesium or high IPP enzyme concentration; limited ability to suppress undesired side reactions, including off-target hydrolysis; the limited ability to suppress pyrophosphorolysis activity in natural and unnatural polymerization or oligonucleotide synthesis reactions; and lower specific activity in commercial, industrial, and research process conditions.

[0127] In the present disclosure, type II IPP enzymes displayed higher specific activity and reduced side activity under industrial process conditions as compared to commercially available and previously utilized type I IPP enzymes. In particular, the type II IPP enzymes had improved activity at industrial process conditions and using reaction buffer for a coupled TnT nucleotide extension

reaction. Additionally, type II IPP enzymes were identified with very good soluble expression in production and high thermostability under process conditions. These improvements may translate to more efficient production of modified RNA molecules and other polynucleotides with higher purity in a coupled nucleotide extension reaction. The advantages of these type II enzymes may also translate to other nonnatural industrial processes that produce pyrophosphate including polymerization and ligation reactions. The present disclosure also provides improved variants type II IPP enzymes, as well as compositions and methods of use for commercial, industrial, and research processes.

[0128] In addition to type II iPP enzymes and variants type II IPP enzymes, the present disclosure provides type I iPPase enzymes and variants of type I iPPase enzymes identified for improved properties compared to a reference iPPase enzyme, and/or paired reactions with other second enzymes.

[0129] Accordingly, in some embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes, one or more type I IPP enzymes, variants of type II IPP enzymes, or variants of type I IPP enzymes, and a phosphorylated compound. In some embodiments, the phosphorylated compound is a natural or modified nucleoside triphosphate (NTP) or a natural or modified NTP analog.

[0130] In some other embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes, one or more type I IPP enzymes, variants of type II IPP enzymes, or variants of type I IPP enzymes, and a second enzyme. In some embodiments, the second enzyme produces a substrate for the IPP enzyme. In some embodiments, the second enzyme is a polymerase. In some embodiments, the second enzyme is a nucleic acid ligase. In some embodiments, the second enzyme uses a natural or modified NTP or a natural or modified NTP analog as a substrate. In some embodiments, the second enzyme catalyzes a substrate to produce inorganic pyrophosphate. In some embodiments, the second enzyme catalyzes conversion of a natural or modified NTP or a natural or modified NTP analog to produce inorganic pyrophosphate.

[0131] In some embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes, variants of type II IPP enzymes, or variants of type II IPP enzymes in a paired extension reaction with a template independent terminal nucleotidyl transferase (TnT) (or polymerase). In some embodiments of the present disclosure, the IPP enzymes convert PPi to Pi in a paired reaction to drive the reaction equilibrium of a paired TnT reaction, whereby a modified or natural NTP or nucleotide analog is added to an acceptor substrate or growing oligonucleotide chain. In some embodiments, the IPP enzymes convert PPi to Pi in a series of two or more paired reactions to drive the reaction equilibrium of two or more paired TnT reactions, whereby two or more modified or natural NTPs or nucleotide analogs are sequentially added to an acceptor substrate or growing oligonucleotide chain.

[0132] In some embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes, variants of type II IPP enzymes, or variants of type I IPP enzymes in a paired ligation reaction with a single strand RNA ligase (ssRNA ligase). In some embodiments of the present disclosure, the IPP enzymes convert PPi to Pi in a paired reaction to drive the reaction equilibrium of a paired ligation reaction, whereby a modified or natural NTP or nucleotide analog or chain of two or more modified or natural NTPs or nucleotide analogs are ligated or added to another modified or natural NTP or nucleotide analog or chain of two or more modified or natural NTPs or nucleotide analogs.

[0133] In some embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes, variants of type II IPP enzymes, or variants of type I IPP enzymes in a paired ligation reaction with a double stranded RNA ligase (dsRNA ligase). In some embodiments of the present disclosure, the IPP enzymes convert PPi to Pi in a paired reaction to drive the reaction equilibrium of a paired dsRNA ligation reaction, whereby a polynucleotide is ligated to another polynucleotide of a double stranded nucleic acid substrate.

[0134] In some embodiments, the present disclosure provides a method or composition comprising one or more type II IPP enzymes, one or more type I IPP enzymes, or variants of type I IPP enzymes or variants of type II IPP enzymes in a paired extension reaction with an RNA polymerase. In some embodiments of the present disclosure, the IPP enzymes convert PPi to Pi in a paired reaction to drive the reaction equilibrium of a paired RNA polymerase reaction, whereby a modified or natural NTP or nucleotide analog is added to an acceptor substrate or growing oligonucleotide chain. In some embodiments, the IPP enzymes convert PPi to Pi in a series of two or more paired reactions to drive the reaction equilibrium of two or more paired RNA polymerase reactions, whereby two or more modified or natural NTPs or nucleotide analogs are sequentially added to an acceptor substrate or growing oligonucleotide chain.

[0135] In any of the embodiments described herein, the composition or method may comprise a nonnatural or industrial process condition.

Compositions of Inorganic Pyrophosphatases for Industrial Processes

[0136] In one aspect, the present disclosure provides compositions comprising an inorganic pyrophosphatase. In particular, compositions comprising a type II inorganic pyrophosphatase, a type I inorganic pyrophosphatase, or a variant of a type I or type II inorganic pyrophosphatase may find use in a variety of commercial, research, and industrial applications.

[0137] In some embodiments of the composition, the composition comprises an inorganic pyrophosphatase and a substrate for the inorganic pyrophosphatase.

[0138] In some embodiments of the composition, the composition comprises an inorganic pyrophosphatase and a phosphorylated compound.

[0139] In some embodiments of the composition, the composition comprises an inorganic pyrophosphatase and a substrate for a second enzyme.

[0140] In some embodiments of the composition, the phosphorylated compound comprises a substrate for the second enzyme.

[0141] In some embodiments of the composition, the substrate for the second enzyme comprises a phosphorylated compound.

[0142] In some embodiments of the composition, the composition comprises an inorganic pyrophosphatase and a second enzyme that produces a substrate for the inorganic pyrophosphatase.

[0143] In some embodiments of the composition, the inorganic pyrophosphatase comprises a naturally occurring or wildtype amino acid sequence from a first species.

[0144] In some embodiments of the composition, the inorganic pyrophosphatase comprises a naturally occurring or wildtype inorganic pyrophosphatase from a first species, wherein the species comprises any of *Thermocrinis ruber*, *Aquifex pyrophilus*, *Thermus oshimai*, *Sulfolobus sp. A20*, *Geobacillus zalihae*, *Bacillus thermozeamaize*, *Bacillus smithii*, *Haloferax sp.*, *Caldibacillus thermoamylovorans*, *Streptococcus thermophilus*, *Thermobacillus xylanilyticus*, *Thermodesulfovibrio yellowstonii*, *Methanococcus aeolicus*, *Methanothermococcus sp.*, *Methanothermococcus sp. SCGC AD-155-M21*, *Methanothermococcus okinawensis*, *Methanotorris igneus*, *Idotea baltica*, *Adineta ricciae*, *Bathycoccus prasinos*, or *Ostreococcus tauri*. In some embodiments, the inorganic pyrophosphatase comprises an engineered variant of the inorganic pyrophosphatase from any of these species.

[0145] In some embodiments of the composition comprising an inorganic pyrophosphatase, the inorganic pyrophosphatase further comprises a type II inorganic pyrophosphatase. In some embodiments, the type II pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or to a reference sequence corresponding to SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

[0146] In some embodiments of the composition, the type II inorganic pyrophosphatase comprises an amino acid sequence comprising residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or an amino acid sequence comprising SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

[0147] In some embodiments, the inorganic pyrophosphatase comprises a variant with a non-naturally occurring amino acid sequence. In some embodiments, the non-naturally occurring amino acid sequence comprises an engineered or recombinant amino acid sequence with at least one amino acid sequence modification. In some embodiments, the at least one amino acid sequence modification

comprises one or more amino acid substitutions, additions, or deletions as compared to the wild type or naturally occurring sequence. In some embodiments, the non-naturally occurring amino acid sequence comprising an engineered or recombinant amino acid sequence further comprises a tag, such as a histidine tag. In some embodiments, the inorganic pyrophosphatase is used without the tag. In some embodiments, an inorganic pyrophosphatase "without the carboxy terminal histidine tag" or "without the carboxy terminal tag" in context of specific sequences refers to the amino acid sequence without the appended histidine tag. In the specified sequence, this corresponds to the 10 amino acid residues in the carboxy terminal region.

[0148] In some embodiments of the composition, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0149] In some embodiments of the composition, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0150] In some embodiments of the composition, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0151] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference at amino acid position 14, 45, 81, 85, 86, 89, 110, 160, 169, 177, 187, 188, 231, 269, or 308, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0152] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference 14L, 45E, 81R, 85V, 86S, 89D, 110L, 160L, 169E, 177E, 187L, 188E, 231T, 269N, or 308I, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0153] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference Y14L, D45E, K81R, L85V, D86S, E89D, I110L, V160L, D169E, K177E, I187L, N188E, V231T, K269N, or V308I, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0154] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference at amino acid position(s) 14/188/269, 45, 81/86/188, 85, 86/89/160/188/308, 89, 110/169/188/231, 169/188/231, 177, 187, or 188, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0155] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference(s) 14L/188E/269N, 45E, 81R/86S/188E, 85V, 86S/89D/160L/188E/308I, 89D, 110L/169E/188E/231T, 169E/188E/231T, 177E, 187L, and 188E, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0156] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference(s) Y14L/N188E/K269N, D45E, K81R/D86S/N188E, L85V, D86S/E89D/V160L/N188E/V308I, E89D, I110L/D169E/N188E/V231T, D169E/N188E/V231T, K177E, I187L, or N188E, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0157] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference at amino acid position(s) 14/269, 81/86, 86/89/160/308, 110/169/231, or 169/231, as compared to the reference sequence corresponding

to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

[0158] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference(s) 14L/269N, 81R/86S, 86S/89D/160L/308I, 110L/169E/231T, or 169E/231T, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56, or to the reference sequence corresponding to SEQ ID NO: 56.

[0159] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference(s) Y14L/K269N, K81R/D86S, D86S/E89D/V160L/V308I, I110L/D169E/V231T, or D169E/V231T, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56, or to the reference sequence corresponding to SEQ ID NO: 56.

[0160] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising residues 12 to carboxy terminal of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or comprising SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0161] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising SEQ ID NO: 76, 78, 80, 82, 84, 86, 88, or 90.

[0162] In some embodiments of the composition, the inorganic pyrophosphatase further comprises an amino acid sequence comprising residues 12 to carboxy terminal of SEQ ID NO: 18, 20, 22, 24, 26, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or comprises an amino acid sequence comprising SEQ ID NO: 18, 20, 22, 24, 26, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0163] In some embodiments of the composition, the inorganic pyrophosphatase comprises a type I inorganic pyrophosphatase. Preferably, in some embodiments, the type I inorganic pyrophosphatase is present in a composition comprising a second enzyme, wherein the second enzyme is selected from a nucleic acid ligase (e.g., ssRNA ligase, dsRNA ligase, DNA ligase, etc.), a template-dependent RNA polymerase, or a poly(X)polymerase.

[0164] In some embodiments of the compositions, the inorganic pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16 without the carboxy terminal histidine-tag, or to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16.

[0165] In some embodiments of the compositions, the inorganic pyrophosphatase comprises an amino acid sequence comprising SEQ ID NO: 4, 6, 8, 10, 12, or 16, or the sequence comprising SEQ ID NO: 4, 6, 8, 10, 12, or 16 without the carboxy terminal histidine tag.

[0166] In some embodiments of the composition, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0167] In some embodiments of the compositions, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0168] In some embodiments of the compositions, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues.

[0169] In some embodiments of the compositions, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference at amino acid position 41, 65, 95, or 147, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0170] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference 41F, 65K/R, 95S, or 147A, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0171] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising at least an amino acid residue difference V41F, P65K, P65R, E95S, or V147A, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to the reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0172] In some embodiments of the composition, the amino acid sequence of the engineered inorganic pyrophosphatase comprises SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, or comprises SEQ ID NO: 18, 20, 22, 24, or 26. In some further embodiments, the inorganic pyrophosphatase comprises an evolved or engineered variant of a type I inorganic pyrophosphatase comprising any of the amino acid sequences of SEQ ID NOs: 18, 20, 22, 24, and 26.

[0173] In some embodiments of the composition, the inorganic pyrophosphatase comprises an amino acid sequence comprising SED ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, or 90.

[0174] In some embodiments of the composition comprising an inorganic pyrophosphatase, the composition further comprises a second enzyme, particularly a second enzyme that produces a substrate for the inorganic pyrophosphatase. As discussed above, in some embodiments, the second enzyme comprises nucleic acid polymerase (e.g., a DNA polymerase, an RNA polymerase, primase, etc.), a nucleic acid ligase (e.g., DNA ligase, an RNA ligase, a double-stranded nucleic acid ligase, a single-stranded nucleic acid ligase, etc.), a template independent polymerase, or a terminal nucleotidyl transferase. These examples are non-limiting and, as a person of skill in the art will recognize, hundreds of enzymes are now known that produce a substrate for an inorganic pyrophosphatase (see for example, https://www.genome.jp/entry/C00013, Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes, Nucleic Acids Res., 2000, 28:27-30), while many more such enzymes are likely to be identified or engineered in the future. The current disclosure envisages each of these embodiments, as if specifically disclosed herein.

[0175] In some embodiments, the second enzyme that produces a substrate for the inorganic pyrophosphatase comprises an evolved enzyme or an engineered enzyme that contains one or amino acid sequence substitutions, insertions, or deletions as compared to a naturally occurring or wildtype sequence or as compared to the amino acid sequence of another evolved or engineered reference enzyme (for example, the enzymes of PCT/US2023/076667, published as WO/2024/081770).

[0176] In some specific embodiments, the second enzyme is a variant or engineered terminal nucleotidyl transferase or another template independent polymerase. Various engineered or variant TnTs and template independent polymerases are known in the art and may be used with the present invention, including terminal nucleotidyl transferases (TdTs), Pol X polymerases, Poly(N) polymerases, Polμ polymerases, Polβ polymerases, Polλ polymerases, and Polθ polymerases, and

others (see for example, U.S. Pat. 10059929, U.S. Pat. 10,760,063, U.S. Pat. 10,774,316, U.S. Pat. 11,390,858, U.S. Pat. 10,745,727, WO2020161480, US20210164008A1, US20210407509, WO2016128731, WO17216472, WO18215803, WO20072715, WO20077227, WO21122539, WO2022029427, and WO21116270, from Molecular Assemblies, Nuclera, DNA Script, and others, each of which is specifically incorporated herein).

[0177] In some embodiments, the second enzyme is a wild-type or variant or engineered terminal nucleotidyl transferase or another template independent polymerase from any of the following species: *Gallus gallus, Xenopus laevis, Oncorhynchus mykiss, Monodelphis domestica, Mus musculus, Ambystoma mexicanum, Takifugu rubripes, Raja eglanteria, Ginglymostoma cirratum, Danio rerio, Canis lupus familiaris, Lemur catta, Microcebus murinus, Rattus norvegicus, Equus caballus, Xenopus (Silurana) tropicalis, Oryctolagus cuniculus, Ailuropoda melanoleuca, Sus scrofa, Anolis carolinensis, Loxodonta Africana, Ornithorhynchus anatinus, Cavia porcellus, Heterocephalus glaber, Macaca mulatto, Sarcophilus harrisii, Sarcophilus harrisii, Otolemur garnettii, Pan paniscus, Saimiri boliviensis, Felis catus, Tupaia chinensis, Pleurodeles waltl, Orcinus orca, Trichechus manatus latirostris, Dasypus novemcinctus, Maylandia zebra, Ochotona princeps, Sorex araneus, Octodon degus, Echinops telfairi, Condylura cristata, Mustela putorius furo, Heterocephalus glaber, Mesocricetus auratus, Melopsittacus undulatus, Falco peregrinus, Chrysemys picta bellii , Microtus ochrogaster, Ictidomys tridecemlineatus, Chinchilla lanigera, Cricetulus griseus, Geospiza fortis, Pseudopodoces humilis, Columba livia, Macaca fascicularis, Pundamilia nyererei, Xiphophorus maculatus, Myotis brandtii, Pantholops hodgsonii, Latimeria chalumnae, Alligator sinensis, Pelodiscus sinensis, Myotis lucifugus, Camelus ferus, Alligator mississippiensis, Lepisosteus oculatus*, and *Homo sapiens*.

[0178] In some embodiments of the composition, the second enzyme comprises a nucleic acid or polynucleotide ligase. In some embodiments, the nucleic acid ligase comprises a DNA ligase, particularly a dsDNA ligase using co-factor ATP. In some embodiments, the DNA ligase using ATP as a co-factor comprises a wild-type or engineered DNA ligase. Exemplary wild-type and engineered DNA ligases include, among others, T4 DNA ligase, and engineered DNA ligases disclosed in U.S. Patent Nos. 10626390 and 10,837,009, and International patent publications WO2018208665 and WO2024158764, all of which are incorporated by reference herein..

[0179] In some embodiments, the DNA In some embodiments, the nucleic acid ligase comprises an RNA ligase. In some embodiments, the RNA ligase comprises a dsRNA ligase or an RNA ligase 2. In some embodiments, the RNA ligase comprises a ssRNA ligase or an RNA ligase 1. In some embodiments, the dsRNA ligase or ssRNA ligase comprises a wild-type or engineered dsRNA ligase or ssRNA ligase. Exemplary dsRNA ligase include, among others, T4 RNA ligase 2 and engineered RNA ligase variants disclosed in U.S. patent publication US2021301280, and International patent publiations WO2008094599 and WO2024138200.

[0180] In some embodiments, the composition comprises an inorganic pyrophosphatase and a substrate for the second enzyme. In some embodiments, the substrate in the composition is selected in context of the second enzyme.

[0181] In some embodiments, the substrate in the composition comprises a phosphorylated compound. In some embodiments, the phosphorylated compound comprises a triphosphate moiety.

[0182] In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises a nucleoside triphosphate or a nonnatural or modified nucleoside triphosphate analog. In some embodiments, the nucleoside triphosphate or nonnatural or modified nucleoside triphosphate analog is a deoxynucleoside triphosphate (dNTP) or a natural or modified dNTP analog. In some embodiments, the nucleoside triphosphate or a natural or modified nucleoside triphosphate analog is a ribonucleoside triphosphate (rNTP) or a nonnatural or modified ribonucleoside triphosphate analog. In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises another non naturally occurring substrate or compound. Various modifications of NTPs and dNTPs are known and have been described (WO/2024/081770, incorporated herein by reference). The descriptions and embodiments of various modifications of NTPs and dNTPs is non-limiting.

[0183] In some embodiments, where the nucleoside triphosphate is a naturally occurring nucleotide triphosphate, the amount or concentration in the composition is an unnatural amount or concentration of the nucleoside triphosphate. In some embodiments, where the second enzyme uses ATP, the ATP is present at a concentration of at least 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11 mM,, 12 mM, 13 mM, 14 mM, 15 mM, or 20 mM, or greater. In some embodiments, where the reaction uses GTP, CTP, UTP, and/or TTP, the GTP, CTP, UTP, and/or TTP is present at a concentration of at least 1.5 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11 mM,, 12 mM, 13 mM, 14 mM, 15 mM, or 20 mM, or greater. In some embodiments, wherein the second enzyme uses dATP, dGTP, dCTP, dUTP, and/or dTTP, the deoxynucleotide triphosphate is present at a concentration of at least 0.05 mM, 0.1 mM, 0.2 mM, 0.5 mM, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 10 mM, 15 mM, or 20 mM, or greater.

[0184] In some embodiments, the rNTP or dNTP comprises a modified sugar moiety. In some embodiments, the rNTP or dNTP is modified at the 3' position and/or 2' position of the sugar moiety. In some embodiments, the modification at the 3' position of the sugar moiety comprises $-NH_2$, $-NO_2$, $-(CH_2)_2-CN$, or $-PO_3$. In some embodiments, the modification at the 3' position of the sugar moiety comprises carbonitriles, phosphates, carbonates, carbamates, esters, ethers, borates, nitrates, sugars, phosphoramidates, phenylsulfenates, and sulfates. In some embodiments, the modification at the 2' position of the sugar moiety comprises a 2'-fluoro, 2'-O-methyl, 2'-O-methoxyethyl, 2'-O-alkyl, locked, or constrained ethyl modifications.

[0185] In some embodiments, the rNTP or dNTP is modified at the gamma, beta, or alpha position of the triphosphate. In some embodiments, one or more phosphates of the rNTP or dNTP comprises a

substitution of an oxygen for a sulfur atom for the creation of phosphorothioate or phosphorodithioate linkages. In some embodiments, the

[0186] In some embodiments, the rNTP or dNTP comprises a natural purine or pyrimidine nucleobase, such as adenine, guanine, cytosine, thymine, or uridine. The rNTP or dNTP comprises an unnatural or modified nucleobase or nucleobase analog. Various modified nucleobases or base analogs are known to those skilled in the art, including but not limited to the following: 5-methylcytosine, pyridin-4-one, pyridin-2-one, phenyl, pseudouracil, 3-methyl uracil, dihydrouridine, naphthyl, aminophenyl, 5- alkylcytidines, 5-alkyluridines, 5-halouridines, 6-azapyrimidines, 6-alkylpyrimidines, propyne, quesosine, 2-thiouridine, 4-thiouridine, 4-acetyltidine, 5-(carboxyhydroxymethyl)uridine, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluridine, -D-galactosylqueosine, 1-methyladenosine, 1- methylinosine, 2,2-dimethylguanosine, 3-methylcytidine, 2-methyladenosine, 2- methylguanosine, N6-methyladenosine, 7-methylguanosine, 5-methoxyaminomethy1-2- thiouridine, 5-methylaminomethy luridine, 5-methylcarbon ylmethyluridine, 5- methyloxyuridine, 5-methyl-2-thiouridine, 2-methylthio-N6-isopentenyladenosine, -D mannosylqueosine, uridine-5-oxyacetic acid, 2-thiocytidine, N1-methyl-adenine, N6-methyl adenine, 8'-azido-adenine, N,N-dimethy1-adenosine, aminoally1-adenosine, 5'-methy1-urdine, pseudouridine, N1- methy1-pseudouridine, 5'-hydroxy-methy1-uridine, 2'-thio-uridine, 4'-thio uridine, hypoxanthine, xanthine, 5'-methyl-cytidine, 5'-hydroxy-methyl-cytidine, 6'-thio guanine, and N7-methyl-guanine. In some embodiments, the nucleobase modification is a removable tag, a cleavable linker, or a radio, photo, or chemical sensor. In some embodiments, the nucleobase modification is a functional element that may be used for isolation, purification, detection, protection, prevention of hydrolysis or degradation, chemical transformation, or to enable further or sequential modifications.

[0187] In some embodiments, the rNTP or dNTP comprises 5'-6-FAM-T (5'-6-carboxyfluorescein-thymidine), ddG (2',3'-dideoxyguanosine), fA (2'-deoxyfluoroadenosine), fC (2'-deoxyfluorocytidine), fG (2'-deoxyfluoroguanosine), fU (2'-deoxyfluorouridine, fA-3'P (2'-deoxyfluoroadenosine with a 3'-PO4 blocking group), fC-3'P (2'-deoxyfluorocytidine with a 3'-PO4 blocking group), fG-3'P (2'-deoxyfluoroguanosine with a 3'-PO4 blocking group), fU-3'P (2'-O-methyluridine with a 3'-PO4 blocking group), mA (2'-O-methyladenosine), mC (2'-O-methylcytidine), mG (2'-O-methylguanosine), mU (2'-O-methyluridine), mA-3'P (2'-O-methyladenosine with a 3'-PO4 blocking group), mC-3'P (2'-O-methylcytidine with a 3'-PO4 blocking group), mG-3'P (2'-O-methylguanosine with a 3'-PO4 blocking group), mU-3'P (2'-O-methyluridine with a 3'-PO4 blocking group), rA (adenosine ribonucleotide), rG (guanosine ribonucleotide), rC (cytidine ribonucleotide), rU (uridine ribonucleotide), (MOE)A (2'-O-(2-methoxyethyl)adenosine), (MOE)C (2'-O-(2-methoxyethyl)cytidine), (MOE)G' (2'-O-(2-methoxyethyl)guanosine), and (MOE)U (2'-O-(2-methoxyethyl)uridine). These examples are non-limiting.

[0188] In some specific embodiments, the substrate for the second enzyme or the phosphorylated compound comprises a modified nucleoside triphosphate or nucleoside triphosphate analog with a phosphate at the 3' position of the sugar moiety and, optionally, one or more additional modifications.

[0189] In some specific embodiments, the substrate for the second enzyme or the phosphorylated compound comprises ATP. In some specific embodiments, the substrate for the second enzyme or the phosphorylated compound comprises 3' phosphate-mATP or 3' phosphate-mUTP.

[0190] In some embodiments, the composition comprises more than one substrate or a mixture of substrates for the second enzyme. In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises a non-naturally occurring concentration. In some embodiments, the substrate for the second enzyme or the phosphorylated compound is a natural or modified NTP or NTP analog, such as ATP, 3' phosphate-mATP or 3' phosphate-mUTP, that comprises a concentration of 0.5 to 3 mM.

[0191] In some embodiments of the composition comprising an inorganic pyrophosphatase and a second enzyme that produces a substrate for the inorganic pyrophosphatase, particularly inorganic pyrophosphate.

[0192] In some embodiments, the composition comprises at least one nonnatural or industrial process condition. A nonnatural or industrial process condition may comprise any condition that differs from a condition found in an organism or living cell, including conditions related to reaction vessels, cofactors, buffers, concentrations of substrates and/or products, temperature, pH, pressure, and the like. A nonnatural or industrial process condition may comprise a condition that facilitates an industrial, commercial, or research process, as known to those of skill in the art.

[0193] In some embodiments, the inorganic pyrophosphatase and the/or second enzyme are *ex vivo* or outside of an organism or living cell. In some embodiments, the inorganic pyrophosphatase and/or the second enzyme are in or on a reaction vessel or reaction surface. In some embodiments, the composition further comprises a reaction vessel or surface. Suitable reaction vessels are well-known to those of skill in the art and include traditional glass, metal, and plastic labware (e.g. flasks, test tubes or plates), disposable, reusable, or recyclable labware, as well as slides, chips, and microfluid or drop-based formats or media. Any suitable reaction vessel or surface may be used.

[0194] In some embodiments, the composition further comprises a buffer. Suitable reaction buffers are well known in the art and include but are not limited to the following: borate, phosphate, 2-(N-morpholino)ethanesulfonic acid (MES), 3-(N-morpholino)propanesulfonic acid (MOPS), acetate, triethanolamine (TEoA), and 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), and the like. Any suitable buffer may be used. In some embodiments, the buffer is present at a concentration of 10-100 mM, 50-500 mM, or 300 mM-1 M.

[0195] In some embodiments, the composition further comprises a cofactor. In some embodiments, the cofactor is a divalent metal ion, such as Ni^{2+}, Mg^{2+}, Mn^{2+}, or Co^{2+}. In some embodiments, the cofactor is present in the form of a salt, such $MgCl_2$, $MnCl_2$, or $CoCl_2$. In some embodiments, the cofactor is a nonnatural or non-preferred cofactor for the inorganic pyrophosphatase and/or the second enzyme. In some embodiments, the cofactor is a natural or preferred cofactor for the inorganic pyrophosphatase and/or the second enzyme. In some embodiments, the cofactor is chosen based on industrial process conditions. In some embodiments, the cofactor is a preferred or natural cofactor for a second enzyme and a nonpreferred or nonnatural cofactor for the IPP enzyme. In some embodiments, the preferred or natural cofactor for a second enzyme and a nonpreferred or nonnatural cofactor for the IPP enzyme is $CoCl_2$. In some embodiments, the cofactor is present at a high concentration compared to an *in vivo* concentration. In some embodiments, the cofactor is present at a concentration of about 0.5 mM to about 5 mM, about 1 mM to about 50 mM, or about 5 mM to about 100 mM.

[0196] In some embodiments, the composition comprises an inorganic pyrophosphatase with improved properties in an industrial process, as compared to another reference inorganic pyrophosphatase. In some embodiments, the improved property as compared to a reference inorganic pyrophosphatase comprises improved solubility, improved thermostability, improved pyrophosphatase activity, decreased off-target hydrolysis, decreased pyrophosphorolysis by a second enzyme, decreased byproducts in a reaction with a second enzyme, and increased ratio of products to byproducts in a reaction with a second enzyme. In some embodiments, the reference inorganic pyrophosphatase is a type I pyrophosphatase or a commercially available pyrophosphatase. Commercially available pyrophosphatases include but are not limited to yeast inorganic pyrophosphatase (New England Biolabs #M2403L), *E.coli* inorganic pyrophosphatase (New England Biolabs # M0361L), and thermostable inorganic pyrophosphatase (New England Biolabs #M0296L). In some embodiments, the reference inorganic pyrophosphatase is a type II inorganic pyrophosphatase or an engineered inorganic pyrophosphatase.

[0197] In some embodiments, the improved property as compared to a reference inorganic pyrophosphatase comprises improved solubility as compared to a reference inorganic pyrophosphatase. In some embodiments, the inorganic pyrophosphatase with improved solubility comprises any of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, and 90.

[0198] In some embodiments, the inorganic pyrophosphatase comprises improved pyrophosphatase activity as compared to a reference inorganic pyrophosphatase. The improved pyrophosphatase activity may comprise improved activity in the conversion of inorganic pyrophosphate to orthophosphate. The improved activity may be measured as described in the Examples or through other methods, as known to those of skill in the art. In some embodiments of the composition, the

increased activity may be measured as reduced phosphorolysis of the second enzyme and/or as an increase in the ratio of products to byproducts. In some embodiments, the improved pyrophosphatase activity comprises at least 1.17-fold, 1.2-fold, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 16-fold, 17-fold, 18-fold, 19-fold, 20-fold, 21-fold, 22-fold, 23-fold, 24-fold, 25-fold, 26-fold, 27-fold, 28-fold, 29-fold, 30-fold, or more the pyrophosphatase activity of a reference pyrophosphatase. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 46, 54, 56, 58, 60, 62, 64, 66, 68, or 70 comprises at least 22-fold, 23-fold, 24-fold, 25-fold, or 26-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase of SEQ ID NO: 10. In some embodiments, the inorganic pyrophosphatase is an engineered variant of a type I pyrophosphatase of SEQ ID NO: 20, 24, and 26 and comprises at least 2-fold increased pyrophosphatase activity as compared to a type I reference pyrophosphatase of SEQ ID NO: 10. In some embodiments, the inorganic pyrophosphatase is an engineered variant of a type II pyrophosphatase of SEQ ID NO: 58, 60, and 62 and comprises at least 1.17-fold increased pyrophosphatase activity as compared to a type II reference pyrophosphatase of SEQ ID NO: 46.

[0199] In some embodiments, the inorganic pyrophosphatase comprises decreased off-target hydrolysis as compared to a reference inorganic pyrophosphatase. The decreased off-target hydrolysis may comprise decreased off-target hydrolysis of a phosphorylated compound or a substrate for a second enzyme. The decreased off-target hydrolysis may comprise decreased off-target hydrolysis of a natural or nonnatural modified NTP or NTP analog. The decreased off-target hydrolysis may be measured as described in the Examples or through other methods, as known to those of skill in the art. In some embodiments, the decreased off-target hydrolysis activity comprises at least 2-fold, 3-fold, 5-fold, 10-fold, 15-fold, 20-fold, 25-fold, 30-fold, 35-fold, 40-fold, 45-fold, 49-fold, 50-fold, 55-fold, or lower off-target hydrolysis as compared to a reference pyrophosphatase. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 46, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 comprises at least 3-fold lower off-target hydrolysis of ATP as compared to a reference inorganic pyrophosphatase of SEQ ID NO: 10. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 10, 12, 46, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 comprises at least 49-fold lower off-target hydrolysis of 3' phosphate-mATP as compared to a commercially available yeast reference pyrophosphatase (New England Biolabs #M2403L).

[0200] In some embodiments, the inorganic pyrophosphatase comprises improved pyrophosphatase activity as compared to a reference inorganic pyrophosphatase. In some embodiments of the composition and second enzyme, the increased activity may be measured as reduced phosphorolysis of the second enzyme and/or as an increase in the ratio of products to byproducts. In some embodiments of the composition, the increased activity may be measured as reduced phosphorolysis of a TnT enzyme and/or as an increase in the ratio of products to byproducts of the TnT enzyme. In some embodiments, the improved pyrophosphatase activity comprises a ratio of products to byproducts of at least 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22,

23, 24, 25, 26, 27, 28, 29, 30, 31, 32, or more. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 34, 38, 40, 42, 44, 46, 54, 56, 58, or 60 comprises an increased ratio of products to byproducts of at least 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, or 32 as compared to a reference pyrophosphatase of SEQ ID NO: 46. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 66, 68, 70, 72, or 74 comprises an increased ratio of products to byproducts of at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 as compared to a reference pyrophosphatase of SEQ ID NO: 56. In some embodiments, the inorganic pyrophosphatase is an engineered variant of a type I pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, or 26 and comprises at least a 2-fold increased pyrophosphatase activity as compared to a type I reference pyrophosphatase of SEQ ID NO: 10. In some embodiments, the inorganic pyrophosphatase is an engineered variant of a type II pyrophosphatase of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and comprises at least a 1.2-fold increased pyrophosphatase activity as compared to a type II reference pyrophosphatase of SEQ ID NO: 46.

[0201] In some embodiments, the improved property as compared to a reference inorganic pyrophosphatase comprises improved thermostability. Improved thermostability may be measured as retention of activity after a heat challenge as described in Example 19 or may be measured by other methods known to those skilled in the art. In some embodiments, the inorganic pyrophosphatase with improved thermostability comprises any of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, and 90.

[0202] In some embodiments, the inorganic pyrophosphatase polypeptides may be modified by the addition, deletion, or insertion of one or more amino acid residues at the N or C-terminal tags of the inorganic pyrophosphatase to facilitate immobilization, or increase protein yield, as described in Examples 25 and 26 and as provided in SEQ ID NOs: 76, 78, 80, 82, 84, 86, 88, and 90. In some embodiments, one or more lysine residues may be added to the N or C terminus of any of the inorganic pyrophosphatase polypeptides of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74. In some embodiments, the deletion is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 amino acid residues from the N-terminus of the polypeptide, particularly where a fusion polypeptide (e.g., histidine tag, polylysine, or combination of histidine and lysine residues, etc.) is present at the N-terminal region. In some embodiments, the deletion is 1, 2, 3, 4, or up to 5 amino acids of from the N-terminus of the polypeptide. In some embodiments, the deletion is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 amino acid residues from the C-terminus of the polypeptide, particularly where a fusion polypeptide (e.g., histidine tag, polylysine, or combination of histidine and lysine residues, etc.) is present at the C-terminal region. In some embodiments, the polypeptide comprises deletion at the N-terminal region and an insertion, e.g., fusion polypeptide, at the C-terminal region. In some embodiments, the polypeptide comprises an insertion, e.g., fusion polypeptide, at the N-terminal region and a deletion at

the C-terminal region. Exemplary inorganic pyrophosphatases with insertions and/or deletions are provided in Examples 25 and 26, and include the inorganic pyrophosphate with an amino acid sequence comprising SEQ ID NO: 76, 78, 80, 82, 84, 86, 88, or 90.

[0203] In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, and a nonnatural or modified nucleoside triphosphate or analog. In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and a second enzyme that produces a product for the inorganic pyrophosphatase and a nonnatural or modified nucleoside triphosphate or analog. In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and a second enzyme that produces a product for the inorganic pyrophosphatase and a nonnatural or industrial process condition. In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and a second enzyme that produces a product for the inorganic pyrophosphatase and a buffer. In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and a second enzyme that produces a product for the inorganic pyrophosphatase and a $CoCl_2$ at a concentration of .5 mM to 5 mM. In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and a nonnatural or modified nucleoside triphosphate or analog and a triethanolamine buffer. In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and a nonnatural or modified nucleoside triphosphate or analog and a terminal nucleotidyl transferase. In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and an engineered or variant terminal nucleotidyl transferase. In a specific embodiment, the composition comprises an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 and an engineered or variant terminal nucleotidyl transferase of SEQ ID NO: 92 or 94.

[0204] In a specific embodiment, the composition comprises an inorganic pyrophosphatase with 85%, 90%, 95%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74; a template independent polymerase, such as a terminal nucleotidyl transferase; and a substrate for the template independent polymerase. In a specific embodiment, the composition comprises an inorganic

pyrophosphatase with 85%, 90%, 95%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74 and a nonnatural or modified nucleoside triphosphate or analog. In a specific embodiment, the composition comprises an inorganic pyrophosphatase with 85%, 90%, 95%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74; a template independent polymerase, such as a terminal nucleotidyl transferase; and a nonnatural or modified nucleoside triphosphate or analog.

[0205] In a specific embodiment, the composition comprises a type II inorganic pyrophosphatase, and a template independent polymerase, such as a terminal nucleotidyl transferase, a terminal nucleotidyl transferases, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase.

[0206] In a specific embodiment, the composition comprises: an inorganic pyrophosphatase with 85%, 90%, 95%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, and a template independent polymerase, such as a terminal nucleotidyl transferase, a terminal nucleotidyl transferases, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase.

[0207] In a specific embodiment, the composition comprises an inorganic pyrophosphatase with 85%, 90%, 95%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, and a template independent polymerase, such as a terminal nucleotidyl transferase, a terminal nucleotidyl transferases, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase from a species selected from: *Gallus gallus, Xenopus laevis, Oncorhynchus mykiss, Monodelphis domestica, Mus musculus, Ambystoma mexicanum, Takifugu rubripes, Raja eglanteria, Ginglymostoma cirratum, Danio rerio, Canis lupus familiaris, Lemur catta, Microcebus murinus, Rattus norvegicus, Equus caballus, Xenopus (Silurana) tropicalis, Oryctolagus cuniculus, Ailuropoda melanoleuca, Sus scrofa, Anolis carolinensis, Loxodonta Africana, Ornithorhynchus anatinus, Cavia porcellus, Heterocephalus glaber, Macaca mulatto, Sarcophilus harrisii, Sarcophilus harrisii, Otolemur garnettii, Pan paniscus, Saimiri boliviensis, Felis catus, Tupaia chinensis, Pleurodeles waltl, Orcinus orca, Trichechus manatus latirostris, Dasypus novemcinctus, Maylandia zebra, Ochotona princeps, Sorex araneus, Octodon degus, Echinops telfairi, Condylura cristata, Mustela putorius furo, Heterocephalus glaber, Mesocricetus auratus, Melopsittacus undulatus, Falco peregrinus, Chrysemys picta bellii , Microtus ochrogaster, Ictidomys tridecemlineatus, Chinchilla lanigera, Cricetulus griseus, Geospiza fortis, Pseudopodoces humilis, Columba livia, Macaca*

fascicularis, *Pundamilia nyererei*, *Xiphophorus maculatus*, *Myotis brandtii*, *Pantholops hodgsonii*, *Latimeria chalumnae*, *Alligator sinensis*, *Pelodiscus sinensis*, *Myotis lucifugus*, *Camelus ferus*, *Alligator mississippiensis*, *Lepisosteus oculatus*, and *Homo sapiens*.

[0208] In a specific embodiment, the composition comprises a type II inorganic pyrophosphatase, and a template independent polymerase, such as a terminal nucleotidyl transferase, a terminal nucleotidyl transferases, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase from a species selected from: *Gallus gallus*, *Xenopus laevis*, *Oncorhynchus mykiss*, *Monodelphis domestica*, *Mus musculus*, *Ambystoma mexicanum*, *Takifugu rubripes*, *Raja eglanteria*, *Ginglymostoma cirratum*, *Danio rerio*, *Canis lupus familiaris*, *Lemur catta*, *Microcebus murinus*, *Rattus norvegicus*, *Equus caballus*, *Xenopus (Silurana) tropicalis*, *Oryctolagus cuniculus*, *Ailuropoda melanoleuca*, *Sus scrofa*, *Anolis carolinensis*, *Loxodonta Africana*, *Ornithorhynchus anatinus*, *Cavia porcellus*, *Heterocephalus glaber*, *Macaca mulatto*. *Sarcophilus harrisii*, *Sarcophilus harrisii*, *Otolemur garnettii*, *Pan paniscus*, *Saimiri boliviensis*, *Felis catus*, *Tupaia chinensis*, *Pleurodeles waltl*, *Orcinus orca*, *Trichechus manatus latirostris*, *Dasypus novemcinctus*, *Maylandia zebra*, *Ochotona princeps*, *Sorex araneus*, *Octodon degus*, *Echinops telfairi*, *Condylura cristata*, *Mustela putorius furo*, *Heterocephalus glaber*, *Mesocricetus auratus*, *Melopsittacus undulatus*, *Falco peregrinus*, *Chrysemys picta bellii* , *Microtus ochrogaster*, *Ictidomys tridecemlineatus*, *Chinchilla lanigera*, *Cricetulus griseus*, *Geospiza fortis*, *Pseudopodoces humilis*, *Columba livia*, *Macaca fascicularis*, *Pundamilia nyererei*, *Xiphophorus maculatus*, *Myotis brandtii*, *Pantholops hodgsonii*, *Latimeria chalumnae*, *Alligator sinensis*, *Pelodiscus sinensis*, *Myotis lucifugus*, *Camelus ferus*, *Alligator mississippiensis*, *Lepisosteus oculatus*, and *Homo sapiens*.

[0209] In any of the above embodiments, the template independent polymerase and/or the inorganic pyrophosphatase may optionally comprise an amino acid sequence with one or more amino acid substitutions, deletions, or insertions as compared to a wild-type amino acid sequence.

Methods of Using Inorganic Pyrophosphatases

[0210] In another aspect, the present disclosure provides methods of using an inorganic pyrophosphatase. In particular, methods comprising a type II inorganic pyrophosphatase or a variant of a type I or type II inorganic pyrophosphatase may find use in a variety of commercial, research, and industrial applications.

[0211] In some embodiments, the present disclosure provides a method of increasing forward reaction in a reaction generating as a by-product inorganic pyrophosphate, comprising providing an inorganic pyrophosphatase for converting inorganic pyrophosphate to phosphate (orthophosphate) in a reaction generating inorganic pyrophosphate as a by-product, wherein the inorganic pyrophosphatase comprises a type II inorganic pyrophosphatase, a type I inorganic pyrophosphatase, or a variant of a type II or type I inorganic pyrophosphatase.

[0212] A "forward" reaction as used herein refers to conversion of a substrate (or reactant) or substrates to a product or products. The conversion of the by-product inorganic pyrophosphate to phosphate renders the reaction irreversible (e.g., product(s) to substrate(s). In some embodiments, the reaction is an in vitro reaction.

[0213] In some embodiments, increasing the forward reaction by providing an inorganic pyrophosphatase increases product yield in the reaction. In some embodiments, the increase in product yield is at least 1.1, 1.2, 1.3, 1.4, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10 fold or greater compared to the reaction without the inorganic pyrophosphatase.

[0214] In some embodiments, a method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase; and

b) providing a substrate for the inorganic pyrophosphatase.

[0215] In some embodiments, the present disclosure provides a method of using an inorganic pyrophosphatase comprising at least:

a) providing an inorganic pyrophosphatase;

b) providing a substrate for the inorganic pyrophosphatase; and

c) contacting the inorganic pyrophosphatase with the substrate such that orthophosphate is produced.

[0216] In some embodiments, the present disclosure provides a method of using an inorganic pyrophosphatase comprising at least:

a) providing an inorganic pyrophosphatase; and

b) providing a phosphorylated compound.

[0217] In some embodiments, the reaction is an enzymatic reaction, as further discussed below. In some embodiments, a method of using an inorganic pyrophosphatase comprising at least:

a) providing an inorganic pyrophosphatase; and

b) providing a substrate for a second enzyme.

[0218] In some embodiments of the method, the substrate for the second enzyme is a phosphorylated compound. In some embodiments of the method, the *phosphorylated compound is a substrate* for the inorganic pyrophosphatase.

[0219] In some embodiments, the method further comprises providing a second enzyme that produces a substrate for the inorganic pyrophosphatase. Thus, in some embodiments, the method of using an inorganic pyrophosphatase comprises:

a) providing an inorganic pyrophosphatase;

b) providing a second enzyme that produces a substrate for the inorganic pyrophosphatase; and

c) contacting the inorganic pyrophosphatase with the substrate such that orthophosphate is produced.

[0220] In some embodiments, the method further comprises providing a substrate for the second enzyme and contacting the substrate with the second enzyme. Thus, in some embodiments, the method of using an inorganic pyrophosphatase comprises:

a) providing an inorganic pyrophosphatase;

b) providing a second enzyme that produces a substrate for the inorganic pyrophosphatase;

c) providing a substrate for the second enzyme;

d) contacting the substrate for the second enzyme with the second enzyme to produce a substrate for the inorganic pyrophosphatase; and

e) contacting the inorganic pyrophosphatase with the substrate such that orthophosphate is produced.

[0221] In some embodiments of the method, the inorganic pyrophosphatase comprises a naturally occurring or wildtype amino acid sequence from a first species.

[0222] In some embodiments of the method, the inorganic pyrophosphatase comprises a naturally occurring or wildtype amino acid sequence from a first species, wherein the species comprises any of *Thermocrinis ruber*, *Aquifex pyrophilus*, *Thermus oshimai*, *Sulfolobus sp. A20*, *Geobacillus zalihae*, *Bacillus thermozeamaize*, *Bacillus smithii*, *Haloferax sp.*, *Caldibacillus thermoamylovorans*, *Streptococcus thermophilus*, *Thermobacillus xylanilyticus*, *Thermodesulfovibrio yellowstonii*, *Methanococcus aeolicus*, *Methanothermococcus sp.*, *Methanothermococcus sp. SCGC AD-155-M21*, *Methanothermococcus okinawensis*, *Methanotorris igneus*, *Idotea baltica*, *Adineta ricciae*, *Bathycoccus prasinos*, or *Ostreococcus tauri*. In some embodiments, the inorganic pyrophosphatase comprises an engineered variant of the inorganic pyrophosphatase amino acid sequence from any of these species.

[0223] In some embodiments of the method, the inorganic pyrophosphatase comprises a type II inorganic pyrophosphatase, or a variant of a type II inorganic pyrophosphatase.

[0224] In some embodiments of the method, the type II inorganic pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or to a reference sequence corresponding to SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

[0225] In some embodiments of the method, the type II inorganic pyrophosphatase comprises an amino acid sequence comprising residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36,

38, 40, 42, 44, 46, 48, 50, or 52, or an amino acid sequence comprising SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52..

[0226] In some embodiments, the inorganic pyrophosphatase comprises a variant with a non-naturally occurring amino acid sequence. In some embodiments, the non-naturally occurring amino acid sequence comprises an engineered or recombinant amino acid sequence with at least one amino acid sequence modification. In some embodiments, the at least one amino acid sequence modification comprises one or more amino acid substitutions, additions, or deletions as compared to the wild type or naturally occurring sequence. In some embodiments, the non-naturally occurring amino acid sequence comprising an engineered or recombinant amino acid sequence further comprises a tag, such as a histidine tag.

[0227] In some embodiments, the variants or engineered variants are variants of a type II or a type I inorganic pyrophosphatase. In some embodiments of the method, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0228] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0229] In some embodiments of the method, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference

sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0230] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 14, 45, 81, 85, 86, 89, 110, 160, 169, 177, 187, 188, 231, 269, or 308, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0231] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference 14L, 45E, 81R, 85V, 86S, 89D, 110L, 160L, 169E, 177E, 187L, 188E, 231T, 269N, or 308I, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0232] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference Y14L, D45E, K81R, L85V, D86S, E89D, I110L, V160L, D169E, K177E, I187L, N188E, V231T, K269N, or V308I, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0233] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position(s) 14/188/269, 45, 81/86/188, 85, 86/89/160/188/308, 89, 110/169/188/231, 169/188/231, 177, 187, or 188, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0234] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) 14L/188E/269N, 45E, 81R/86S/188E, 85V, 86S/89D/160L/188E/308I, 89D, 110L/169E/188E/231T, 169E/188E/231T, 177E, 187L, or 188E, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46., or to the reference sequence corresponding to SEQ ID NO: 46.

[0235] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) Y14L/N188E/K269N, D45E, K81R/D86S/N188E, L85V, D86S/E89D/V160L/N188E/V308I, E89D, I110L/D169E/N188E/V231T, D169E/N188E/V231T, K177E, I187L, or N188E, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0236] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position(s) 14/269, 81/86, 86/89/160/308, 110/169/231, or 169/231, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

[0237] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) 14L/269N, 81R/86S, 86S/89D/160L/308I, 110L/169E/231T, or 169E/231T, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

[0238] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) Y14L/K269N, K81R/D86S, D86S/E89D/V160L/V308I, I110L/D169E/V231T, or D169E/V231T, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

[0239] In some embodiments, the amino acid sequence of the inorganic pyrophosphatase comprises residues 12 to carboxy terminal of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or comprises SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74. In some embodiments of the method, the inorganic pyrophosphatase further comprises any of the amino acid sequences of SEQ ID NOs: 18, 20, 22, 24, 26, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74.

[0240] In some embodiments of the method, the inorganic pyrophosphatase has reduced off target hydrolysis as compared to a reference inorganic pyrophosphatase.

[0241] In some embodiments, the inorganic pyrophosphatase has reduced off target hydrolysis of a natural or modified nucleoside triphosphate as compared to a reference pyrophosphatase. In some embodiments, the inorganic pyrophosphatase has at least 3-fold lower off target hydrolysis of a natural nucleoside triphosphate as compared to a reference inorganic pyrophosphatase. In some embodiments, the nucleoside triphosphate for off-target hydrolysis is ATP. In some embodiments, the modified nucleoside triphosphate for off-target hydrolysis is 3'-phosphate-2'-O-methyl-ATP. In some embodiments, the reference inorganic pyrophosphatase has a sequence corresponding to SEQ ID NO: 10.

[0242] In some embodiments, the inorganic pyrophosphatase has at least a 49-fold lower off target hydrolysis of a non-natural or modified nucleoside triphosphate as compared to a reference inorganic pyrophosphatase. In some embodiments, the non-natural or modified nucleoside triphosphate is 3' phosphate-2'-O-methyl-ATP. In some embodiments, the reference inorganic pyrophosphatase is a yeast inorganic pyrophosphatase (inorganic pyrophosphatase of New England Biolabs M2403L, US)

[0243] In some embodiments, the inorganic pyrophosphatase exhibits increased pyrophosphatase activity as compared to a reference pyrophosphatase.

[0244] In some embodiments, the inorganic pyrophosphatase comprises at least a 1.17-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase. In some embodiments, the inorganic pyrophosphatase comprises at least a 22-fold, 23-fold, 24-fold, 25-fold, or 26-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase. In some embodiments, the reference inorganic pyrophosphatase for changes in activity has a sequence corresponding to SEQ ID NO: 10.

[0245] In some embodiments of the method, the inorganic pyrophosphatase comprises a type I inorganic pyrophosphatase, or a variant of a type I inorganic pyrophosphatase.

[0246] In some embodiments, the type I pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16, without the carboxy terminal histidine tag, or to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16. In the embodiments herein, the carboxy terminal histidine tag is the ten amino acid residues at the carboxy terminal region in the referenced sequences.

[0247] In some embodiments, the inorganic pyrophosphatase comprises a type I pyrophosphatase comprising the amino acid sequence of SEQ ID NO: 4, 6, 8, 10, 12, or 16, without the carboxy terminal histidine tag or comprises the amino acid sequence of SEQ ID NO: 4, 6, 8, 10, 12, or 16.

[0248] In some embodiments, the inorganic pyrophosphatase comprises an evolved or engineered variant of a type I inorganic pyrophosphatase.

[0249] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0250] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy

terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0251] In some embodiments, the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues.

[0252] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 41, 65, 95, or 147, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0253] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference 41F, 65K/R, 95S, or 147A, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0254] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference V41F, P65K, P65R, E95S, or V147A, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0255] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues, or comprises the amino acid sequence comprising SEQ ID NO: 18, 20, 22, 24, or 26.

[0256] In some embodiments of the method, the inorganic pyrophosphatase comprises any of the amino acid sequences of SED ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, and 90.

[0257] In some embodiments of the method, the inorganic pyrophosphatase comprises increased pyrophosphatase activity as compared to a reference pyrophosphatase.

[0258] In some embodiments, the inorganic pyrophosphatase comprises at least a 2-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase.

[0259] In some embodiments, the inorganic pyrophosphatase has increased solubility as compared to a reference inorganic pyrophosphatase.

[0260] In some embodiments, the inorganic pyrophosphatase comprises an increase in thermal stability.

[0261] In some embodiments, the reaction is carried out with a second enzyme that produces a substrate for the inorganic pyrophosphatase, e.g., inorganic pyrophosphate. In some embodiments, the second enzyme comprises a heterologous enzyme (i.e., heterologous to the inorganic pyrophosphatase). In some embodiments, the reaction with the second enzyme comprises a nucleotidyl transferase reaction, a polynucleotide ligase reaction, a polynucleotide polymerase reaction, or a tRNA synthase reaction.

[0262] In some embodiments, the method comprises providing an inorganic pyrophosphatase and providing a second enzyme that produces a substrate for the inorganic pyrophosphatase. In some embodiments, the second enzyme comprises a DNA polymerase, an RNA polymerase, a nucleotide polymerase, a DNA ligase, an RNA ligase, a nucleotide ligase, a double-stranded nucleotide ligase, a single-stranded nucleotide ligase, a template independent polymerase, or a terminal nucleotidyl transferase. These examples are non-limiting and, as a person of skill in the art will recognize, hundreds of enzymes are now known that produce a substrate for an inorganic pyrophosphatase (see for example, www.genome.jp/entry/C00013, Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000)). The current disclosure envisages each of these embodiments, as if specifically disclosed herein.

[0263] In some embodiments, the second enzyme that produces a substrate for the inorganic pyrophosphatase comprises an evolved enzyme or an engineered enzyme that contains one or amino acid sequence substitutions, insertions, or deletions as compared to a naturally occurring or wildtype sequence or as compared to the amino acid sequence of another evolved or engineered reference enzyme (for example, the TnT enzymes of PCT/US2023/076667).

[0264] In some specific embodiments, the second enzyme is a variant or engineered terminal nucleotidyl transferase or another template independent polymerase. Various engineered or variant TnTs and template independent polymerases are known in the art and may be used with the present invention, including terminal nucleotidyl transferases (TdTs), Pol X polymerases, Poly(N) polymerases, Polμ polymerases, Polβ polymerases, Polλ polymerases, and Polθ polymerases, and others (see for example, U.S. Pat. 10059929, U.S. Pat. 10,760,063, U.S. Pat. 10,774,316, U.S. Pat.

11,390,858, U.S. Pat. 10,745,727, PCT/GB2020/050247, US20210164008A1, US20210407509, WO2016GB50301, WO17216472, WO18215803, WO20072715, WO20077227, WO21122539, WO2022029427, and WO21116270, from Molecular Assemblies, Nuclera, DNA Script, and others, each of which is specifically incorporated herein).

[0265] In some embodiments, the second enzyme is a wild-type or variant or engineered terminal nucleotidyl transferase or another template independent polymerase from any of the following species: *Gallus gallus, Xenopus laevis, Oncorhynchus mykiss, Monodelphis domestica, Mus musculus, Ambystoma mexicanum, Takifugu rubripes, Raja eglanteria, Ginglymostoma cirratum, Danio rerio, Canis lupus familiaris, Lemur catta, Microcebus murinus, Rattus norvegicus, Equus caballus, Xenopus (Silurana) tropicalis, Oryctolagus cuniculus, Ailuropoda melanoleuca, Sus scrofa, Anolis carolinensis, Loxodonta Africana, Ornithorhynchus anatinus, Cavia porcellus, Heterocephalus glaber, Macaca mulatto, Sarcophilus harrisii, Sarcophilus harrisii, Otolemur garnettii, Pan paniscus, Saimiri boliviensis, Felis catus, Tupaia chinensis, Pleurodeles waltl, Orcinus orca, Trichechus manatus latirostris, Dasypus novemcinctus, Maylandia zebra, Ochotona princeps, Sorex araneus, Octodon degus, Echinops telfairi, Condylura cristata, Mustela putorius furo, Heterocephalus glaber, Mesocricetus auratus, Melopsittacus undulatus, Falco peregrinus, Chrysemys picta bellii , Microtus ochrogaster, Ictidomys tridecemlineatus, Chinchilla lanigera, Cricetulus griseus, Geospiza fortis, Pseudopodoces humilis, Columba livia, Macaca fascicularis, Pundamilia nyererei, Xiphophorus maculatus, Myotis brandtii, Pantholops hodgsonii, Latimeria chalumnae, Alligator sinensis, Pelodiscus sinensis, Myotis lucifugus, Camelus ferus, Alligator mississippiensis, Lepisosteus oculatus*, and *Homo sapiens*.

[0266] In some embodiments, the substrate for the second enzyme comprises a triphosphate moiety. In some embodiments, the phosphorylated compound comprises a triphosphate moiety.

[0267] In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises a nucleoside triphosphate or a nonnatural or modified nucleoside triphosphate analog. In some embodiments, the nucleoside triphosphate or nonnatural or modified nucleoside triphosphate analog is a deoxynucleoside triphosphate (dNTP) or a natural or modified dNTP analog. In some embodiments, the nucleoside triphosphate or a natural or modified nucleoside triphosphate analog is a ribonucleoside triphosphate (rNTP) or a nonnatural or modified ribonucleoside triphosphate analog. In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises another non naturally occurring substrate or compound. Various modifications of NTPs and dNTPs are known and have been described (PCT/US2023/076667, incorporated herein by reference). The descriptions and embodiments of various modifications of NTPs and dNTPs is non-limiting.

[0268] In some embodiments, the rNTP or dNTP comprises a modified sugar moiety. In some embodiments, the rNTP or dNTP is modified at the 3' position and/or 2' position of the sugar moiety.

In some embodiments, the modification at the 3' position of the sugar moiety comprises $-NH_2$, $-NO_2$, $-(CH_2)_2-CN$, or $-PO_3$. In some embodiments, the modification at the 3' position of the sugar moiety comprises carbonitriles, phosphates, carbonates, carbamates, esters, ethers, borates, nitrates, sugars, phosphoramidates, phenylsulfenates, and sulfates. In some embodiments, the modification at the 2' position of the sugar moiety comprises a 2'-flouro, 2'-O-methyl, 2'-O-methoxyethyl, 2'-O-alkyl, locked or constrained ethyl modifications.

[0269] In some embodiments, the rNTP or dNTP is modified at the gamma, beta, or alpha position of the triphosphate. In some embodiments, one or more phosphates of the rNTP or dNTP comprises a substitution of an oxygen for a sulfur atom for the creation of phosphorothioate or phosphorodithioate linkages.

[0270] In some embodiments, the rNTP or dNTP comprises a natural purine or pyrimidine nucleobase, such as adenine, guanine, cytosine, thymine, or uridine. The rNTP or dNTP comprises an unnatural or modified nucleobase or nucleobase analog. Various modified nucleobases or base analogs are known to those skilled in the art, including but not limited to the following: 5-methylcytosine, pyridin-4-one, pyridin-2-one, phenyl, pseudouracil, 3-methyl uracil, dihydrouridine, naphthyl, aminophenyl, 5- alkylcytidines, 5-alkyluridines, 5-halouridines, 6-azapyrimidines, 6-alkylpyrimidines, propyne, quesosine, 2-thiouridine, 4-thiouridine, 4-acetyltidine, 5-(carboxyhydroxymethyl)uridine, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluridine, -D-galactosylqueosine, 1-methyladenosine, 1- methylinosine, 2,2-dimethylguanosine, 3-methylcytidine, 2-methyladenosine, 2- methylguanosine, N6-methyladenosine, 7-methylguanosine, 5-methoxyaminomethy1-2- thiouridine, 5-methylaminomethy luridine, 5-methylcarbon ylmethyluridine, 5- methyloxyuridine, 5-methyl-2-thiouridine, 2-methylthio-N6-isopentenyladenosine, -D mannosylqueosine, uridine-5-oxyacetic acid, 2-thiocytidine, N1-methyl-adenine, N6-methyl adenine, 8'-azido-adenine, N,N-dimethy1-adenosine, aminoally1-adenosine, 5'-methy1-urdine, pseudouridine, N1- methy1-pseudouridine, 5'-hydroxy-methy1-uridine, 2'-thio-uridine, 4'-thio uridine, hypoxanthine, xanthine, 5'-methyl-cytidine, 5'-hydroxy-methyl-cytidine, 6'-thio guanine, and N7-methyl-guanine. In some embodiments, the nucleobase modification is a removable tag, a cleavable linker, or a radio, photo, or chemical sensor. In some embodiments, the nucleobase modification is a functional element that may be used for isolation, purification, detection, protection, prevention of hydrolysis or degradation, chemical transformation, or to enable further or sequential modifications.

[0271] In some embodiments, the rNTP or dNTP comprises 5'-6-FAM-T (5'-6-carboxyfluorescein-thymidine), ddG (2',3'-dideoxyguanosine), fA (2'-deoxyfluoroadenosine), fC (2'-deoxyfluorocytidine), fG (2'-deoxyfluoroguanosine), fU (2'-deoxyfluorouridine, fA-3'P (2'-deoxyfluoroadenosine with a 3'-PO4 blocking group), fC-3'P (2'-deoxyfluorocytidine with a 3'-PO4 blocking group), fG-3'P (2'-deoxyfluoroguanosine with a 3'-PO4 blocking group), fU-3'P (2'-O-methyluridine with a 3'-PO4

blocking group), mA (2'-O-methyladenosine), mC (2'-O-methylcytidine), mG (2'-O-methylguanosine), mU (2'-O-methyluridine), mA-3'P (2'-O-methyladenosine with a 3'-PO4 blocking group), mC-3'P (2'-O-methylcytidine with a 3'-PO4 blocking group), mG-3'P (2'-O-methylguanosine with a 3'-PO4 blocking group), mU-3'P (2'-O-methyluridine with a 3'-PO4 blocking group), rA (adenosine ribonucleotide), rG (guanosine ribonucleotide), rC (cytidine ribonucleotide), rU (uridine ribonucleotide), (MOE)A (2'-O-(2-methoxyethyl)adenosine), (MOE)C (2'-O-(2-methoxyethyl)cytidine), (MOE)G' (2'-O-(2-methoxyethyl)guanosine), and (MOE)U (2'-O-(2-methoxyethyl)uridine). These examples are non-limiting.

[0272] In some specific embodiments, the substrate for the second enzyme or the phosphorylated compound comprises a modified nucleoside triphosphate or nucleoside triphosphate analog with a phosphate at the 3' position of the sugar moiety and, optionally, one or more additional modifications.

[0273] In some specific embodiments, the substrate for the second enzyme or the phosphorylated compound comprises ATP, GTP, CTP, UTP, or TTP. In some embodiments, as provided above, the substrate for the second enzyme or the phosphorylated compound comprises a 2'-modified NTP. In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises a 2'-O'-methyl-ATP, 2'-O'-methyl-GTP, 2'-O'-methyl-CTP, or 2'-O-methyl-UTP. In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises 2'-fluoro-ATP, 2'-fluoro-GTP, 2'-fluoro-CTP, or 2'-fluoro-UTP.

[0274] In some specific embodiments, the substrate for the second enzyme or the phosphorylated compound comprises a 3'-phosphate NTP. In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises 3'-phosphate-2'-O'-methyl-ATP, 3'-phosphate-2'-O'-methyl-GTP, 3'-phosphate-2'-O'-methyl-CTP, or 3'-phosphate-2'-O-methyl-UTP. In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises 3'-phosphate-2'-fluoro-ATP, 3'-phosphate-2'-fluoro-GTP, 3'-phosphate-2'-fluoro-CTP, or 3'-phosphate-2'-fluoro-UTP.

[0275] In some embodiments, the method comprises more than one substrate or a mixture of substrates for the second enzyme. In some embodiments, the substrate for the second enzyme or the phosphorylated compound comprises a non-naturally occurring concentration. In some embodiments, where the reaction (e.g., second enzyme) uses ATP, the ATP is present at a concentration of at least 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11 mM,, 12 mM, 13 mM, 14 mM, 15 mM, or 20 mM, or greater. In some embodiments, where the reaction uses GTP, CTP, UTP, and/or TTP, the GTP, CTP, UTP, and/or TTP is present at a concentration of at least 1.5 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11 mM,, 12 mM, 13 mM, 14 mM, 15 mM, or 20 mM, or greater. In some embodiments, wherein the reaction (e.g., second enzyme) uses dATP, dGTP, dCTP, dUTP, and/or dTTP, the deoxynucleotide triphosphate is present at a concentration of at least 0.05 mM, 0.1 mM, 0.2 mM, 0.5 mM, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 10 mM, 15 mM, or 20 mM, or greater.

[0276] In some embodiments, the substrate for the second enzyme or the phosphorylated compound is a natural or modified NTP or NTP analog, such as ATP, 3' phosphate-2'-O-methyl-ATP or 3' phosphate-2'-O-methyl-UTP, that comprises a concentration of 0.5 to 3 mM, or 1 to 10 mM, or 2 to 20 mM.

[0277] In some embodiments of the method comprising providing an inorganic pyrophosphatase and providing a second enzyme that produces a substrate for the inorganic pyrophosphatase, e.g., the second enzyme produces inorganic pyrophosphate.

[0278] In some embodiments, the method comprises at least one nonnatural or industrial process condition. A nonnatural or industrial process condition may comprise any condition that differs from a condition found in an organism or living cell, including conditions related to reaction vessels, cofactors, buffers, concentrations of substrates and/or products, temperature, pH, pressure, and the like. A nonnatural or industrial process condition may comprise a condition that facilitates an industrial, commercial, or research process, as known to those of skill in the art.

[0279] In some embodiments, the inorganic pyrophosphatase and the/or second enzyme are *ex vivo* or outside of an organism or living cell. In some embodiments, the inorganic pyrophosphatase and the/or second enzyme are present *in vitro*. In some embodiments, the inorganic pyrophosphatase and/or the second enzyme are in or on a reaction vessel or reaction surface. In some embodiments, the method further comprises a reaction vessel or surface. Suitable reaction vessels are well-known to those of skill in the art and include traditional glass, metal, and plastic labware (e.g. flasks, test tubes or plates), disposable, reusable, or recyclable labware, as well as slides, chips, and microfluid or drop-based formats or media. Any suitable reaction vessel or surface may be used.

[0280] In some embodiments, the method further comprises a buffer. Suitable reaction buffers are well known in the art and include but are not limited to the following: borate, phosphate, 2-(N-morpholino)ethanesulfonic acid (MES), 3-(N-morpholino)propanesulfonic acid (MOPS), acetate, triethanolamine (TEoA), and 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), and the like. Any suitable buffer may be used. In some embodiments, the buffer is present at a concentration of 10-100 mM, 50-500 mM, or 300 mM-1 M.

[0281] In some embodiments, the method further comprises a cofactor, for example where the reaction uses a cofactor. In some embodiments, the cofactor is a divalent metal ion, such as Ni^{2+}, Mg^{2+}, Mn^{2+}, or Co^{2+}. In some embodiments, the cofactor is present in the form of a salt, such $MgCl_2$, $MnCl_2$, or $CoCl_2$. In some embodiments, the cofactor is a nonnatural or non-preferred cofactor for the inorganic pyrophosphatase and/or the second enzyme. In some embodiments, the cofactor is a natural or preferred cofactor for the inorganic pyrophosphatase and/or the second enzyme. In some embodiments, the cofactor is chosen based on industrial process conditions. In some embodiments, the cofactor is a preferred or natural cofactor for a second enzyme and a nonpreferred or nonnatural cofactor for the IPP enzyme. In some embodiments, the preferred or natural cofactor for a second

enzyme and a nonpreferred or nonnatural cofactor for the IPP enzyme is $CoCl_2$. In some embodiments, the cofactor is present at a high concentration compared to an *in vivo* concentration. In some embodiments, the cofactor is present at a concentration of 0.5 mM to 5 mM.

[0282] In some embodiments, the method comprises an inorganic pyrophosphatase with improved properties in an industrial process, as compared to another reference inorganic pyrophosphatase. In some embodiments, the improved property as compared to a reference inorganic pyrophosphatase comprises improved solubility, improved thermostability, improved pyrophosphatase activity, decreased off-target hydrolysis, decreased pyrophosphorolysis by a second enzyme, decreased byproducts in a reaction with a second enzyme, and increased ratio of products to byproducts in a reaction with a second enzyme. In some embodiments, the reference inorganic pyrophosphatase is a type I pyrophosphatase or a commercially available pyrophosphatase. Commercially available pyrophosphatases include but are not limited to yeast inorganic pyrophosphatase (New England Biolabs #M2403L), *E.coli* inorganic pyrophosphatase (New England Biolabs # M0361L), and thermostable inorganic pyrophosphatase (New England Biolabs #M0296L). In some embodiments, the reference inorganic pyrophosphatase is a type II inorganic pyrophosphatase or an engineered inorganic pyrophosphatase.

[0283] In some embodiments, the improved property as compared to a reference inorganic pyrophosphatase comprises improved solubility as compared to a reference inorganic pyrophosphatase. In some embodiments, the inorganic pyrophosphatase with improved solubility comprises any of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, and 90.

[0284] In some embodiments, the inorganic pyrophosphatase comprises improved pyrophosphatase activity as compared to a reference inorganic pyrophosphatase. The improved pyrophosphatase activity may comprise improved activity in the conversion of inorganic pyrophosphate to orthophosphate. The improved activity may be measured as described in the Examples or through other methods, as known to those of skill in the art. In some embodiments of the method, the increased activity may be measured as reduced phosphorolysis of the second enzyme and/or as an increase in the ratio of products to byproducts. In some embodiments, the improved pyrophosphatase activity comprises at least 1.17-fold, 1.2-fold, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 16-fold, 17-fold, 18-fold, 19-fold, 20-fold, 21-fold, 22-fold, 23-fold, 24-fold, 25-fold, 26-fold, 27-fold, 28-fold, 29-fold, 30-fold, or more the pyrophosphatase activity of a reference pyrophosphatase. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 46, 54, 56, 58, 60, 62, 64, 66, 68, or 70 comprises at least 22-fold, 23-fold, 24-fold, 25-fold, or 26-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase of SEQ ID NO: 10. In some embodiments, the inorganic pyrophosphatase is an engineered variant of a type I pyrophosphatase of SEQ ID NO: 20, 24, and 26 and comprises at least

2-fold increased pyrophosphatase activity as compared to a type I reference pyrophosphatase of SEQ ID NO: 10. In some embodiments, the inorganic pyrophosphatase is an engineered variant of a type II pyrophosphatase of SEQ ID NO: 58, 60, and 62 and comprises at least 1.17-fold increased pyrophosphatase activity as compared to a type II reference pyrophosphatase of SEQ ID NO: 46.

[0285] In some embodiments, the inorganic pyrophosphatase comprises decreased off-target hydrolysis as compared to a reference inorganic pyrophosphatase. The decreased off-target hydrolysis may comprise decreased off-target hydrolysis of a phosphorylated compound or a substrate for a second enzyme. The decreased off-target hydrolysis may comprise decreased off-target hydrolysis of a natural or nonnatural modified NTP or NTP analog. The decreased off-target hydrolysis may be measured as described in the Examples or through other methods, as known to those of skill in the art. In some embodiments, the decreased off-target hydrolysis activity comprises at least 2-fold, 3-fold, 5-fold, 10-fold, 15-fold, 20-fold, 25-fold, 30-fold, 35-fold, 40-fold, 45-fold, 49-fold, 50-fold, 55-fold, or lower off-target hydrolysis as compared to a reference pyrophosphatase. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 46, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 comprises at least 3-fold lower off-target hydrolysis of ATP as compared to a reference inorganic pyrophosphatase of SEQ ID NO: 10. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 10, 12, 46, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74 comprises at least 49-fold lower off-target hydrolysis of 3' phosphate-mATP as compared to a commercially available yeast reference pyrophosphatase (New England Biolabs #M2403L).

[0286] In some embodiments of the method, the inorganic pyrophosphatase comprises improved pyrophosphatase activity as compared to a reference inorganic pyrophosphatase. In some embodiments of the method, the increased activity may be measured as reduced phosphorolysis of the second enzyme and/or as an increase in the ratio of products to byproducts. In some embodiments, the increased activity may be measured as reduced phosphorolysis of a TnT enzyme and/or as an increase in the ratio of products to byproducts of the TnT enzyme (see, e.g., Anderson et al., Nucleic Acids Research, 1999, 27(15):3190–3196). In some embodiments, the improved pyrophosphatase activity comprises a ratio of products to byproducts of at least 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, or more. In some embodiments, the inorganic pyrophosphatase of SEQ ID NO: 34, 38, 40, 42, 44, 46, 56, 58, or 60 comprises an increased ratio of products to byproducts of at least 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, or 32 as compared to a reference pyrophosphatase of SEQ ID NO: 46. In some embodiments, the variant type II inorganic pyrophosphatase of SEQ ID NO: 66, 68, 70, 72 or 74 comprises an increased ratio of products to byproducts of at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 as compared to a reference pyrophosphatase of SEQ ID NO: 56. In some embodiments, the inorganic pyrophosphatase is an engineered variant of a type I pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, or 26 and comprises at least a 2-fold increased pyrophosphatase activity as compared to a type I reference pyrophosphatase of SEQ ID NO: 10. In

some embodiments, the inorganic pyrophosphatase is an engineered variant of a type II pyrophosphatase of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72 or 74 and comprises at least a 1.2-fold increased pyrophosphatase activity as compared to a type II reference pyrophosphatase of SEQ ID NO: 46.

[0287] In some embodiments, the improved property as compared to a reference inorganic pyrophosphatase comprises improved thermostability. Improved thermostability may be measured as retention of activity after a heat challenge as described in Example 19 or may be measured by other methods known to those skilled in the art. In some embodiments, the inorganic pyrophosphatase with improved thermostability comprises any of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74.

[0288] In some embodiments, the inorganic pyrophosphatase polypeptides may be modified by the addition, deletion, or insertion of one or more amino acid residues at the N or C-terminal tags of the inorganic pyrophosphatase to facilitate immobilization or increase protein yield, as described in Examples 25 and 26 and as provided in SEQ ID NOs: 76, 78, 80, 82, 84, 86, 88, and 90. In some embodiments, one or more lysine residues may be added to the N or C terminus of any of the inorganic pyrophosphatase polypeptides of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74.

[0289] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74; and

b) providing a nonnatural or modified nucleoside triphosphate or analog.

[0290] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74;

b) providing a second enzyme that produces a product for the inorganic pyrophosphatase; and

c) providing a nonnatural or modified nucleoside triphosphate or analog.

[0291] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74;

b) providing a second enzyme that produces a product for the inorganic pyrophosphatase; and

c) a nonnatural or industrial process condition.

[0292] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74;

b) providing a second enzyme that produces a product for the inorganic pyrophosphatase; and

c) a triethanolamine buffer.

[0293] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74;

b) providing a second enzyme that produces a product for the inorganic pyrophosphatase; and

c) $CoCl_2$ at a concentration of 0.5 mM to 5 mM.

[0294] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74;

b) providing a terminal nucleotidyl transferase; and

a) a nonnatural or modified nucleoside triphosphate or analog.

[0295] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74; and

b) providing an engineered or variant terminal nucleotidyl transferase.

[0296] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises at least:

a) providing an inorganic pyrophosphatase of SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74; and

b) providing an engineered or variant terminal nucleotidyl transferase of SEQ ID NO: 92 or 94.

[0297] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises:

a) providing an inorganic pyrophosphatase with 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more amino acid sequence identity to any of SEQ ID NO: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74;

b) providing a template independent polymerase, such as a terminal nucleotidyl transferase;

c) providing a substrate for the template independent polymerase;

d) contacting the substrate with the template independent polymerase such that organic pyrophosphate is produced; and

e) contacting the inorganic pyrophosphatase with the organic pyrophosphate such that orthophosphate is produced.

[0298] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises:

a) providing an inorganic pyrophosphatase with 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74; and

b) providing a nonnatural or modified nucleoside triphosphate or analog.

[0299] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises:

a) providing an inorganic pyrophosphatase with 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74;

b) providing a template independent polymerase, such as a terminal nucleotidyl transferase;

c) providing a nonnatural or modified nucleoside triphosphate or analog;

d) contacting the nonnatural or modified nucleoside triphosphate or analog with the template independent polymerase such that organic pyrophosphate is produced; and

e) contacting the inorganic pyrophosphatase with the organic pyrophosphate such that orthophosphate is produced.

[0300] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises:

a) providing a type II inorganic pyrophosphatase; and

b) providing a template independent polymerase, such as a terminal nucleotidyl transferase, a terminal nucleotidyl transferases, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase.

[0301] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises:

a) providing an inorganic pyrophosphatase with 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74; and

b) providing a template independent polymerase, such as a terminal nucleotidyl transferase, a terminal nucleotidyl transferases, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase.

[0302] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises:

a) providing an inorganic pyrophosphatase with 70%, 75%, 80%, 81%, 82%, 83%, 84%,

85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more amino acid sequence identity to any of SEQ ID NOs: 10, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, and

b) providing a template independent polymerase, such as a terminal nucleotidyl transferase, a terminal nucleotidyl transferases, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase from a species selected from: *Gallus gallus, Xenopus laevis, Oncorhynchus mykiss, Monodelphis domestica, Mus musculus, Ambystoma mexicanum, Takifugu rubripes, Raja eglanteria, Ginglymostoma cirratum, Danio rerio, Canis lupus familiaris, Lemur catta, Microcebus murinus, Rattus norvegicus, Equus caballus, Xenopus (Silurana) tropicalis, Oryctolagus cuniculus, Ailuropoda melanoleuca, Sus scrofa, Anolis carolinensis, Loxodonta Africana, Ornithorhynchus anatinus, Cavia porcellus, Heterocephalus glaber, Macaca mulatto, Sarcophilus harrisii, Sarcophilus harrisii, Otolemur garnettii, Pan paniscus, Saimiri boliviensis, Felis catus, Tupaia chinensis, Pleurodeles waltl, Orcinus orca, Trichechus manatus latirostris, Dasypus novemcinctus, Maylandia zebra, Ochotona princeps, Sorex araneus, Octodon degus, Echinops telfairi, Condylura cristata, Mustela putorius furo, Heterocephalus glaber, Mesocricetus auratus, Melopsittacus undulatus, Falco peregrinus, Chrysemys picta bellii , Microtus ochrogaster, Ictidomys tridecemlineatus, Chinchilla lanigera, Cricetulus griseus, Geospiza fortis, Pseudopodoces humilis, Columba livia, Macaca fascicularis, Pundamilia nyererei, Xiphophorus maculatus, Myotis brandtii, Pantholops hodgsonii, Latimeria chalumnae, Alligator sinensis, Pelodiscus sinensis, Myotis lucifugus, Camelus ferus, Alligator mississippiensis, Lepisosteus oculatus*, and *Homo sapiens*.

[0303] In a specific embodiment, the method of using an inorganic pyrophosphatase comprises:

a) providing a type II inorganic pyrophosphatase; and

b) providing a template independent polymerase, such as a terminal nucleotidyl transferase, a terminal nucleotidyl transferases, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, or a Polθ polymerase from a species selected from: *Gallus gallus, Xenopus laevis, Oncorhynchus mykiss, Monodelphis domestica, Mus musculus, Ambystoma mexicanum, Takifugu rubripes, Raja eglanteria, Ginglymostoma cirratum, Danio rerio, Canis lupus familiaris, Lemur catta, Microcebus murinus, Rattus norvegicus, Equus caballus, Xenopus (Silurana) tropicalis, Oryctolagus cuniculus, Ailuropoda melanoleuca, Sus scrofa, Anolis carolinensis, Loxodonta Africana, Ornithorhynchus anatinus, Cavia porcellus, Heterocephalus glaber, Macaca mulatto, Sarcophilus harrisii, Sarcophilus harrisii, Otolemur garnettii, Pan paniscus, Saimiri boliviensis, Felis catus, Tupaia chinensis, Pleurodeles waltl, Orcinus orca, Trichechus manatus latirostris, Dasypus novemcinctus, Maylandia zebra, Ochotona princeps, Sorex araneus, Octodon degus, Echinops telfairi, Condylura cristata, Mustela putorius furo, Heterocephalus glaber, Mesocricetus auratus, Melopsittacus undulatus, Falco peregrinus, Chrysemys picta bellii , Microtus ochrogaster, Ictidomys tridecemlineatus, Chinchilla*

lanigera, *Cricetulus griseus*, *Geospiza fortis*, *Pseudopodoces humilis*, *Columba livia*, *Macaca fascicularis*, *Pundamilia nyererei*, *Xiphophorus maculatus*, *Myotis brandtii*, *Pantholops hodgsonii*, *Latimeria chalumnae*, *Alligator sinensis*, *Pelodiscus sinensis*, *Myotis lucifugus*, *Camelus ferus*, *Alligator mississippiensis*, *Lepisosteus oculatus*, and *Homo sapiens*.

[0304] In any of the above embodiments, the template independent polymerase and/or the inorganic pyrophosphatase may optionally comprise an amino acid sequence with one or more amino acid substitutions, deletions, or insertions as compared to a wild-type amino acid sequence. Exemplary template independent polymerase variants are described in, for example, WO2020077227 and WO2017216472.

Type II and Type I, and Variant Type II and Variant Type I Inorganic Pyrophosphatases

[0305] In another aspect, the present disclosure provides type II inorganic pyrophosphatase polypeptides, type I inorganic pyrophosphatase polypeptides, and evolved or engineered variants of type I and type II inorganic pyrophosphatase polypeptides useful in methods and compositions comprising nonnatural conditions present in industrial applications. In some embodiments, the polypeptides described herein have improved solubility, improved thermostability, improved pyrophosphatase activity, decreased off-target hydrolysis, decreased pyrophosphorolysis by a second enzyme, decreased byproducts in a reaction with a second enzyme, and/or increased ratio of products to byproducts in a reaction with a second enzyme.

[0306] The present invention provides inorganic pyrophosphatase polypeptides, polynucleotides encoding the polypeptides, methods of preparing the polypeptides, and methods for using the polypeptides. Where the description relates to polypeptides, it is to be understood that it can describe the polynucleotides encoding the polypeptides.

[0307] Suitable reaction conditions under which the above-described improved properties of the engineered polypeptides carry out the desired reaction can be determined with respect to concentrations or amounts of polypeptide, substrate, co-substrate, buffer, solvent, pH, conditions including temperature and reaction time, and/or conditions with the polypeptide immobilized on a solid support, as further described below and in the Examples.

[0308] In some embodiments, the exemplary engineered inorganic pyrophosphatases comprise an amino acid sequence that has one or more residue differences as compared to SEQ ID NO: 10 at the residue positions indicated in **Tables 3**, **7**, and **13**. In some embodiments, the exemplary engineered inorganic pyrophosphatases comprise an amino acid sequence that has one or more residue differences as compared to SEQ ID NO: 46 at the residue positions indicated in **Tables 5**, **9**, **15**, **21**, **22** and **26.1**. In some embodiments, the exemplary engineered inorganic pyrophosphatases comprise an amino acid sequence that has one or more residue differences as compared to SEQ ID NO: 56 at the residue positions indicated in **Tables 23**, **24.1**, and **24.2**.

[0309] The structure and function information for the exemplary engineered polypeptides of the present invention are based on the improved solubility as compared to a reference polypeptide, increased pyrophosphatase activity, decreased off-target hydrolysis, and increased ratio of products to byproducts in a paired or coupled reaction, as further described in the Examples. The odd-numbered sequence identifiers (i.e., SEQ ID NOs) in these Tables refer to the nucleotide sequence encoding the amino acid sequence provided by the even-numbered SEQ ID NOs in these Tables. Exemplary sequences are provided in the electronic sequence listing file accompanying this invention, which is hereby incorporated by reference herein. The amino acid residue differences are based on comparison to the reference sequence of SEQ ID NOs: 10, 46, and 56, as indicated.

[0310] The wild-type type I inorganic pyrophosphatase from *Sulfolobus sp. A20* (SEQ ID NO: 10) and the wild-type type II inorganic pyrophosphatase from *Methanotorris igneus* (SEQ ID NO: 46) were selected for evolution. The polypeptides of the present disclosure are engineered variants of SEQ ID NOs: 10 and 46 with a N-terminal 6-histidine tag.

[0311] The polypeptides of the present disclosure have residue differences that result in improved properties for an inorganic pyrophosphatase in an industrial method or composition. Various residue differences, at both conserved and non-conserved positions, have been discovered to be related to improvements in various enzymes properties, including improved solubility, improved thermostability, improved pyrophosphatase activity, decreased off-target hydrolysis, decreased pyrophosphorolysis by a second enzyme, decreased byproducts in a reaction with a second enzyme, and/or increased ratio of products to byproducts in a reaction with a second enzyme.

[0312] The activity of each engineered inorganic pyrophosphatase relative to the reference polypeptide of SEQ ID NOs: 10, 46, and/or 56 was determined as solubility and/or activity as described in the Examples herein. In some embodiments, a shake flask purified enzyme (SFP) is used to assess the properties of the engineered inorganic pyrophosphatase, the results of which are provided in the Examples.

[0313] In some embodiments, the specific enzyme properties are associated with the residues differences as compared to SEQ ID NOs: 10, 46, and/or 56 at the residue positions indicated herein. In some embodiments, residue differences affecting polypeptide expression can be used to increase expression of the engineered inorganic pyrophosphatases.

[0314] In light of the guidance provided herein, it is further contemplated that any of the exemplary engineered polypeptides comprising the sequences of SEQ ID NOs: 10, 46, and/or 56 find use as the starting amino acid sequence for synthesizing other inorganic pyrophosphatase polypeptides, for example by subsequent rounds of evolution that incorporate new combinations of various amino acid differences from other polypeptides in **Tables 3**, **5**, **7**, **9**, **13**, **15**, **21**, **22**, **23**, **24.1**, **24.2**, **246.1** and **26.2**, and other residue positions described herein. Further improvements may be generated by including

amino acid differences at residue positions that had been maintained as unchanged throughout earlier rounds of evolution.

[0315] In some embodiments, the engineered inorganic pyrophosphatase polypeptide is a variant of a type I inorganic pyrophosphatase and has increased soluble protein expression, increased pyrophosphatase activity, decreased off-target hydrolysis, and/or improved thermal stability.

[0316] In some embodiments, an engineered inorganic pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0317] In some embodiments, an engineered inorganic pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0318] In some embodiments, an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26.

[0319] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 41, 65, 95, or 147, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0320] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference 41F, 65K/R, 95S, or 147A, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0321] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference V41F, P65K, P65R, E95S, or V147A, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0322] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, or comprises SEQ ID NO: 18, 20, 22, 24, or 26.

[0323] In some embodiments, the engineered inorganic pyrophosphatase polypeptide is a variant of a type II inorganic pyrophosphatase and has increased soluble protein expression, increased pyrophosphatase activity, decreased off-target hydrolysis, increased ratio of products to byproducts in a reaction with a second enzyme, and/or improved thermal stability

[0324] In some embodiments, an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0325] In some embodiments, an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0326] In some embodiments, an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence

corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0327] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 14, 45, 81, 85, 86, 89, 110, 160, 169, 177, 187, 188, 231, 269, or 308, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0328] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference 14L, 45E, 81R, 85V, 86S, 89D, 110L, 160L, 169E, 177E, 187L, 188E, 231T, 269N, or 308I, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0329] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference Y14L, D45E, K81R, L85V, D86S, E89D, I110L, V160L, D169E, K177E, I187L, N188E, V231T, K269N, or V308I, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0330] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position(s) 14/188/269, 45, 81/86/188, 85, 86/89/160/188/308, 89, 110/169/188/231, 169/188/231, 177, 187, or 188, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0331] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) 14L/188E/269N, 45E, 81R/86S/188E, 85V, 86S/89D/160L/188E/308I, 89D, 110L/169E/188E/231T, 169E/188E/231T, 177E, 187L, and 188E, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0332] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) Y14L/N188E/K269N, D45E, K81R/D86S/N188E, L85V, D86S/E89D/V160L/N188E/V308I, E89D, I110L/D169E/N188E/V231T, D169E/N188E/V231T, K177E, I187L, or N188E, as compared to the reference sequence

corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0333] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position(s) 14/269, 81/86, 86/89/160/308, 110/169/231, or 169/231, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

[0334] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) 14L/269N, 81R/86S, 86S/89D/160L/308I, 110L/169E/231T, or 169E/231T, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

[0335] In some embodiments, the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) Y14L/K269N, K81R/D86S, D86S/E89D/V160L/V308I, I110L/D169E/V231T, or D169E/V231T, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

[0336] In some embodiments, the amino acid sequence of the inorganic pyrophosphatase comprises residues 12 to carboxy terminal of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, or 90, or comprises SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, or 90.

[0337] As will be appreciated by the skilled artisan, in some embodiments, one or a combination of residue differences above that is selected can be kept constant (i.e., maintained) in the engineered inorganic pyrophosphatase as a core feature, and additional residue differences at other residue positions incorporated into the sequence to generate additional engineered inorganic pyrophosphatase polypeptides with improved properties. Accordingly, it is to be understood for any engineered inorganic pyrophosphatase containing one or a subset of the residue differences above, the present invention contemplates other engineered inorganic pyrophosphatases that comprise the one or subset of the residue differences, and additionally one or more residue differences at the other residue positions disclosed herein.

[0338] As noted above, the engineered inorganic pyrophosphatase polypeptides are also capable of converting substrates (e.g., inorganic pyrophosphate) to products (e.g., orthophosphate). In some embodiments, the engineered inorganic pyrophosphatase polypeptide is capable of converting the substrate compounds to the product compound with at least 1.17-fold, 1.2-fold, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold,

16-fold, 17-fold, 18-fold, 19-fold, 20-fold, 21-fold, 22-fold, 23-fold, 24-fold, 25-fold, 26-fold, 27-fold, 28-fold, 29-fold, 30-fold, or more activity relative to the activity of the reference polypeptide of SEQ ID NOs: 10, 46, and/or 56.

[0339] As noted above, the engineered inorganic pyrophosphatase polypeptides are also capable of reducing the off-target hydrolysis of a phosphorylated compound (e.g., ATP, 3' phos-mATP, or another natural or nonnatural NTP or analog). In some embodiments, the engineered inorganic pyrophosphatase polypeptide is capable of decreasing the off-target hydrolysis of a phosphorylated compound by at least 2-fold, 3-fold, 5-fold, 10-fold, 15-fold, 20-fold, 25-fold, 30-fold, 35-fold, 40-fold, 45-fold, 49-fold, 50-fold, 55-fold, or more relative to the activity of the reference polypeptide of SEQ ID NOs: 10, 46, and/or 56 or another reference inorganic pyrophosphatase.

[0340] In some embodiments, the inorganic pyrophosphatase capable of converting the substrate compounds to the product compounds with at least 1.17-fold, 2-fold, or 22-fold the activity relative to SEQ ID NOs: 10, 46, and/or 56, comprises an amino acid sequence selected from the even-numbered sequences in SEQ ID NOs: 18-90.

[0341] In some embodiments, the inorganic pyrophosphatase capable of reducing the off-target hydrolysis of a phosphorylated compound by at least 2-fold, 3-fold, or 49-fold the activity relative to SEQ ID NOs: 10, 46, and/or 56, comprises an amino acid sequence selected from the even-numbered sequences in SEQ ID NOs: 18-90.

[0342] In some embodiments, the inorganic pyrophosphatase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 10, 46, and/or 56, that increases soluble expression or isolated protein yield of the inorganic pyrophosphatase in a bacterial host cell, particularly in *E. coli*, as compared to a wild-type or engineered reference inorganic pyrophosphatase, comprises an amino acid sequence selected from the even-numbered sequences in SEQ ID NOs: 18-90.

[0343] In some embodiments, the engineered inorganic pyrophosphatase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 10, 46, and/or 56, that increases thermostability of the engineered inorganic pyrophosphatase, as compared to a wild-type or engineered reference inorganic pyrophosphatase, comprises an amino acid sequence selected from the even-numbered sequences of SEQ ID NOs: 18-90.

[0344] In some embodiments, the engineered inorganic pyrophosphatase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 10, 46, and/or 56, that increases the ratio of product to byproduct in a paired or coupled reaction with a second enzyme, as compared to a wild-type or engineered reference inorganic pyrophosphatase, comprises an amino acid sequence selected from the even-numbered sequences of SEQ ID NOs: 18-90.

[0345] In some embodiments, the engineered inorganic pyrophosphatase with improved properties has an amino acid sequence comprising a sequence selected from the even-numbered sequences of SEQ ID NOs: 18-90.

[0346] In some embodiments, the engineered inorganic pyrophosphatase, comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identity to one of the sequences of SEQ ID NOs: 10, 46, and/or 56 , as provided in the Examples.

[0347] In addition to the residue positions specified above, any of the engineered inorganic pyrophosphatase polypeptides disclosed herein can further comprise other residue differences relative to SEQ ID NOs: 10, 46, and/or 56, at other residue positions (i.e., residue positions other than those included herein). Residue differences at these other residue positions can provide for additional variations in the amino acid sequence without adversely affecting the ability of the polypeptide to carry out the conversion of substrate to product. Accordingly, in some embodiments, in addition to the amino acid residue differences present in any one of the engineered inorganic pyrophosphatase polypeptides selected from the even-numbered sequences in the range of SEQ ID NOs: 18-90, the sequence can further comprise 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-26, 1-30, 1-35, 1-40, 1-45, 1-50, 1-100, or 1-150 residue differences at other amino acid residue positions as compared to the SEQ ID NOs: 10, 46, and/or 56. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 100, or 150 residue positions. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 residue positions. The residue differences at these other positions can be conservative changes or non-conservative changes. In some embodiments, the residue differences can comprise conservative substitutions and non-conservative substitutions as compared to the inorganic pyrophosphatase polypeptide of SEQ ID NOs: 10, 46, and/or 56.

[0348] In some embodiments, the present invention also provides engineered polypeptides that comprise a fragment of any of the engineered inorganic pyrophosphatase polypeptides described herein that retains the functional activity and/or improved property of that engineered inorganic pyrophosphatase. Accordingly, in some embodiments, the present invention provides a polypeptide fragment capable of converting substrate to product under suitable reaction conditions, wherein the fragment comprises at least about 90%, 95%, 96%, 97%, 98%, or 99% of a full-length or truncated amino acid sequence of an engineered inorganic pyrophosphatase of the present invention, such as an exemplary inorganic pyrophosphatase polypeptide selected from the even-numbered sequences in the range of SEQ ID NOs: 18-90. In some embodiments, the engineered inorganic pyrophosphatase can have an amino acid sequence comprising a deletion in any one of the inorganic pyrophosphatase

polypeptide sequences described herein, such as the exemplary engineered polypeptides of the even-numbered sequences in the range of SEQ ID NOs: 18-90.

[0349] Thus, for each and every embodiment of the engineered inorganic pyrophosphatase polypeptides of the invention, the amino acid sequence can comprise deletions of one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, up to 20% of the total number of amino acids, or up to 30% of the total number of amino acids of the inorganic pyrophosphatase polypeptides, where the associated functional activity and/or improved properties of the engineered inorganic pyrophosphatase described herein are maintained. In some embodiments, the deletions can comprise 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 amino acid residues. In some embodiments, the number of deletions can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, or 50 amino acid residues. In some embodiments, the deletions can comprise deletions of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 amino acid residues.

[0350] In some embodiments, the engineered inorganic pyrophosphatase polypeptide herein can have an amino acid sequence comprising an insertion as compared to any one of the engineered inorganic pyrophosphatase polypeptides described herein, such as the exemplary engineered polypeptides of the even-numbered sequences in the range of SEQ ID NOs: 18-90. Thus, for each and every embodiment of the inorganic pyrophosphatase polypeptides of the invention, the insertions can comprise one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, 30 or more amino acids, 40 or more amino acids, or 50 or more amino acids, where the associated functional activity and/or improved properties of the engineered inorganic pyrophosphatase described herein is maintained. The insertions can be to amino or carboxy terminus, or internal portions of the inorganic pyrophosphatase polypeptide.

[0351] In some embodiments, the engineered inorganic pyrophosphatase described herein can have an amino acid sequence comprising a sequence selected from the even-numbered sequences in the range of SEQ ID NOs: 18-90, and optionally one or several (e.g., up to 3, 4, 5, or up to 10) amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, 1-50, 1-75, 1-100, or 1-150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally around 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

[0352] In some embodiments, the polypeptide comprises a deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 amino acid residues from the N-terminus of the polypeptide, particularly where a fusion polypeptide (e.g., histidine tag, polylysine, or combination of histidine and lysine residues, etc.) is present at the N-terminal region. In some embodiments, the deletion is 1, 2, 3, 4, or up to 5 amino acids of from the N-terminus of the polypeptide. In some embodiments, the deletion is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 amino acid residues from the C-terminus of the polypeptide, particularly where a fusion polypeptide (e.g., histidine tag, polylysine, or combination of histidine and lysine residues, etc.) is present at the C-terminal region. In some embodiments, the polypeptide comprises deletion at the N-terminal region and an insertion, e.g., fusion polypeptide, at the C-terminal region. In some embodiments, the polypeptide comprises an insertion, e.g., fusion polypeptide, at the N-terminal region and a deletion at the C-terminal region. Exemplary inorganic pyrophosphatases with insertions and/or deletions are provided in Examples 25 and 26, and include the inorganic pyrophosphatase with an amino acid sequence comprising SEQ ID NO: 76, 78, 80, 82, 84, 86, 88, or 90.

[0353] In the above embodiments, the suitable reaction conditions for the engineered polypeptides are provided in the Examples herein.

[0354] In some embodiments, the inorganic pyrophosphatase is expressed as a fusion protein. In some embodiments, the inorganic pyrophosphatase described herein can be fused to a variety of polypeptide sequences, such as, by way of example and not limitation, polypeptide tags that can be used for detection and/or purification, cell localization signals (e.g., secretion signals), polypeptides with enzymatic activity, and/or for providing a reactive amino acid for immobilization on a support medium. In some embodiments, the fusion polypeptide of the inorganic pyrophosphatase comprises a glycine-histidine or histidine-tag (His-tag). In some embodiments, the fusion polypeptide of the inorganic pyrophosphatase comprises one or more lysine residues, for example lysine residues of 1-12 amino acids or longer in length, for example to enhance expression and/or provide for immobilization on a support medium. In some embodiments, the fusion protein of the inorganic pyrophosphatase comprises an epitope tag, such as c-myc, FLAG, V5, or hemagglutinin (HA). In some embodiments, the fusion is to the amino (N-) terminus of inorganic pyrophosphatase polypeptide. In some embodiments, the fusion polypeptide comprises a mixture of fusion polypeptides, for example, a mixture of histidine and lysine residues, such as provided in the Examples. In some embodiments, the fusion is to the carboxy (C-) terminus of the inorganic pyrophosphatase polypeptide. In some embodiments, the polypeptides described herein can be used with or without fusions to other polypeptides.

[0355] In some embodiments, the engineered inorganic pyrophosphatase polypeptides of the invention can be fused to another second polypeptide, such as a polypeptide with a different enzymatic activity. In some embodiments, the present provides a fusion polypeptide comprising an engineered inorganic pyrophosphatase polypeptide fused to a second polypeptide with terminal

nucleotidyl transferase activity. For example, synthetic genes encoding an *N*-terminal and *C*-terminal hexahistidine tagged version of an inorganic pyrophosphatase polypeptide (e.g., polypeptide of SEQ ID NOs: 1-90) can be fused to gene encoding a TnT variant polypeptide. Typically, the polypeptides (e.g., IPP and TnT) are fused via a polypeptide linker (e.g., a GSGGTG linker) introduced in the construct between the genes encoding the polypeptides. Such fusion proteins can be constructed using well-established techniques (e.g., Gibson assembly cloning) and expressed in *E. coli* (e.g., a strain derived from W3110).

[0356] It is to be understood that the polypeptides described herein are not restricted to the genetically encoded amino acids. In addition to the genetically encoded amino acids, the polypeptides described herein may be comprised, either in whole or in part, of naturally occurring and/or synthetic non-encoded amino acids. Certain commonly encountered non-encoded amino acids of which the polypeptides described herein may be comprised include, but are not limited to: the D-stereoisomers of the genetically-encoded amino acids; 2,3-diaminopropionic acid (Dpr); α-aminoisobutyric acid (Aib); ε-aminohexanoic acid (Aha); δ-aminovaleric acid (Ava); N-methylglycine or sarcosine (MeGly or Sar); ornithine (Orn); citrulline (Cit); t-butylalanine (Bua); t-butylglycine (Bug); N-methylisoleucine (MeIle); phenylglycine (Phg); cyclohexylalanine (Cha); norleucine (Nle); naphthylalanine (Nal); 2-chlorophenylalanine (Ocf); 3-chlorophenylalanine (Mcf); 4-chlorophenylalanine (Pcf); 2-fluorophenylalanine (Off); 3-fluorophenylalanine (Mff); 4-fluorophenylalanine (Pff); 2-bromophenylalanine (Obf); 3-bromophenylalanine (Mbf); 4-bromophenylalanine (Pbf); 2-methylphenylalanine (Omf); 3-methylphenylalanine (Mmf); 4-methylphenylalanine (Pmf); 2-nitrophenylalanine (Onf); 3-nitrophenylalanine (Mnf); 4-nitrophenylalanine (Pnf); 2-cyanophenylalanine (Ocf); 3-cyanophenylalanine (Mcf); 4-cyanophenylalanine (Pcf); 2-trifluoromethylphenylalanine (Otf); 3-trifluoromethylphenylalanine (Mtf); 4-trifluoromethylphenylalanine (Ptf); 4-aminophenylalanine (Paf); 4-iodophenylalanine (Pif); 4-aminomethylphenylalanine (Pamf); 2,4-dichlorophenylalanine (Opef); 3,4-dichlorophenylalanine (Mpcf); 2,4-difluorophenylalanine (Opff); 3,4-difluorophenylalanine (Mpff); pyrid-2-ylalanine (2pAla); pyrid-3-ylalanine (3pAla); pyrid-4-ylalanine (4pAla); naphth-1-ylalanine (1nAla); naphth-2-ylalanine (2nAla); thiazolylalanine (taAla); benzothienylalanine (bAla); thienylalanine (tAla); furylalanine (fAla); homophenylalanine (hPhe); homotyrosine (hTyr); homotryptophan (hTrp); pentafluorophenylalanine (5ff); styrylkalanine (sAla); authrylalanine (aAla); 3,3-diphenylalanine (Dfa); 3-amino-5-phenypentanoic acid (Afp); penicillamine (Pen); 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (Tic); β-2-thienylalanine (Thi); methionine sulfoxide (Mso); N(w)-nitroarginine (nArg); homolysine (hLys); phosphonomethylphenylalanine (pmPhe); phosphoserine (pSer); phosphothreonine (pThr); homoaspartic acid (hAsp); homoglutanic acid (hGlu); 1-aminocyclopent-(2 or 3)-ene-4 carboxylic acid; pipecolic acid (PA), azetidine-3-carboxylic acid (ACA); 1-aminocyclopentane-3-carboxylic acid; allylglycine (aGly); propargylglycine (pgGly); homoalanine (hAla); norvaline (nVal); homoleucine (hLeu), homovaline (hVal); homoisoleucine (hIle);

homoarginine (hArg); N-acetyl lysine (AcLys); 2,4-diaminobutyric acid (Dbu); 2,3-diaminobutyric acid (Dab); N-methylvaline (MeVal); homocysteine (hCys); homoserine (hSer); hydroxyproline (Hyp) and homoproline (hPro). Additional non-encoded amino acids of which the polypeptides described herein may be comprised will be apparent to those of skill in the art (See e.g., the various amino acids provided in Fasman, CRC Practical Handbook of Biochemistry and Molecular Biology, CRC Press, Boca Raton, FL, pp. 3-70 (1989), and the references cited therein, all of which are incorporated by reference). These amino acids may be in either the L- or D-configuration.

[0357] Those of skill in the art will recognize that amino acids or residues bearing side chain protecting groups may also comprise the polypeptides described herein. Non-limiting examples of such protected amino acids, which in this case belong to the aromatic category, include (protecting groups listed in parentheses), but are not limited to: Arg(tos), Cys(methylbenzyl), Cys (nitropyridinesulfenyl), Glu(δ-benzylester), Gln(xanthyl), Asn(N-δ-xanthyl), His(bom), His(benzyl), His(tos), Lys(fmoc), Lys(tos), Ser(O-benzyl), Thr (O-benzyl) and Tyr(O-benzyl).

[0358] Non-encoding amino acids that are conformationally constrained of which the polypeptides described herein may be composed include, but are not limited to, N-methyl amino acids (L-configuration); 1-aminocyclopent-(2 or 3)-ene-4-carboxylic acid; pipecolic acid; azetidine-3-carboxylic acid; homoproline (hPro); and 1-aminocyclopentane-3-carboxylic acid.

[0359] In some embodiments, the engineered polypeptides can be in various forms, for example, such as an isolated preparation, as a substantially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, and/or as cell extracts and/or lysates of such cells. The enzymes can be lyophilized, spray-dried, precipitated or be in the form of a crude paste, as further discussed below.

[0360] In some embodiments, the polypeptides described herein is provided in solution, as a lyophilizate, or is immobilized on a substrate or support medium. In some embodiments, the substrate or support medium is a solid substrate, porous substrate, membrane, or particles. The enzyme can be entrapped in matrixes or membranes. In some embodiments, matrices include polymeric materials such as calcium-alginate, agar, k-carrageenin, polyacrylamide, agarose or derivatives thereof (e.g., cross-linked agarose), and collagen, or solid matrices, such as activated carbon, porous ceramic, and diatomaceous earth. In some embodiments, the matrix is a particle, a membrane, or a fiber. Types of membranes include, among others, nylon, cellulose, polysulfone, or polyacrylate.

[0361] In some embodiments, the polypeptide is immobilized on the surface of a support material. In some embodiments, the enzyme is adsorbed on the support material. In some embodiments, the enzyme is immobilized on the support material by covalent attachment. Support materials include, among others, inorganic materials, such as alumina, silica, porous glass, ceramics, diatomaceous earth, clay, and bentonite, or organic materials, such as cellulose (CMC, DEAE-cellulose), starch, activated carbon, polyacrylate, polyacrylamide, polystyrene, and ion-exchange resins, such as

Amberlite, Sephadex, and Dowex. Exemplary support mediums and immobilizations are described in the Examples.

[0362] In some embodiments, the polypeptides described herein are provided in the form of kits. The enzymes in the kits may be present individually or as a plurality of enzymes. The kits can further include reagents for carrying out the enzymatic reactions, substrates for assessing the activity of enzymes, as well as reagents for detecting the products. The kits can also include reagent dispensers and instructions for use of the kits.

[0363] In some embodiments, the kits of the present invention include arrays comprising a plurality of different inorganic pyrophosphatase polypeptides at different addressable position, wherein the different polypeptides are different variants of a reference sequence each having at least one different improved enzyme property. In some embodiments, a plurality of polypeptides immobilized on solid supports are configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments. The array can be used to test a variety of substrate compounds for conversion by the polypeptides. Such arrays comprising a plurality of engineered polypeptides and methods of their use are known in the art (See e.g., WO2009/008908A2).

Polynucleotides, Expression Vectors, and Host Cells

[0364] In another aspect, the present disclosure provides polynucleotides encoding any of the inorganic pyrophosphatase polypeptides or engineered variants thereof, as described herein. The polynucleotides may be operatively linked to one or more heterologous regulatory sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. Expression constructs containing a heterologous polynucleotide encoding the engineered inorganic pyrophosphatase are introduced into appropriate host cells to express the corresponding inorganic pyrophosphatase polypeptide.

[0365] As will be apparent to the skilled artisan, availability of a protein sequence and the knowledge of the codons corresponding to the various amino acids provide a description of all the polynucleotides capable of encoding the subject polypeptides. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons, allows an extremely large number of nucleic acids to be made, all of which encode the improved inorganic pyrophosphatase enzymes. Thus, having knowledge of a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the protein. In this regard, the present invention specifically contemplates each and every possible variation of polynucleotides that could be made encoding the polypeptides described herein by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide described herein, including the amino acid sequences presented in **Tables 3**, **5**, **7**, **9**, **13**,

15, **21**, **22**, **23**, **24.1**, **24.2**, **26.1**, and **26.2**, and disclosed in the Sequence Listing incorporated by reference herein as the sequences of even-numbered SEQ ID NO. in the range of SEQ ID NOs: 2-90.

[0366] In various embodiments, the codons are preferably selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria are used to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. In some embodiments, all codons need not be replaced to optimize the codon usage of the inorganic pyrophosphatase 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 inorganic pyrophosphatase enzymes may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full-length coding region.

[0367] In some embodiments, the polynucleotide comprises a codon optimized nucleotide sequence encoding the inorganic pyrophosphatase polypeptide amino acid sequence, as represented by SEQ ID NOs: 10, 46, and/or 56. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity to the codon optimized nucleic acid sequences encoding the even-numbered sequences in the range of SEQ ID NOs: 2-90. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences in the odd-numbered sequences in the range of SEQ ID NOs: 1-89, with or without the nucleotide sequences encoding the corresponding histidine-tag fusion polypeptide, as described herein. In some embodiments, the codon optimized sequences of the odd-numbered sequences in the range of SEQ ID NOs: 1-89, enhance expression of the encoded inorganic pyrophosphatase, providing preparations of enzyme capable of converting substrate to product.

[0368] In some embodiments, the polynucleotide comprises a polynucleotide sequence comprising at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 9, 45, and/or 55 and/or or a functional fragment thereof, wherein said polynucleotide sequence encodes an engineered polypeptide comprising at least one substitution at one or more amino acid positions

[0369] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to residues 12 to carboxy

terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or to a reference sequence corresponding to SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

[0370] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an inorganic pyrophosphatase comprising an amino acid sequence comprising residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or an amino acid sequence comprising SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

[0371] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding at least one engineered terminal deoxynucleotidyl transferase comprising a sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 10, 46, and/or 56, with or without the corresponding histidine-tag residues.

[0372] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0373] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

[0374] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54,

56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0375] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence comprising at least an amino acid residue difference at amino acid position 14, 45, 81, 85, 86, 89, 110, 160, 169, 177, 187, 188, 231, 269, or 308, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

[0376] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence comprising at least an amino acid residue difference at amino acid position(s) 14/188/269, 45, 81/86/188, 85, 86/89/160/188/308, 89, 110/169/188/231, 169/188/231, 177, 187, or 188, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

[0377] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence comprising at least an amino acid residue difference at amino acid position(s) 14/269, 81/86, 86/89/160/308, 110/169/231, or 169/231, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

[0378] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence comprising residues 12 to carboxy terminal of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or comprising SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

[0379] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16, without the carboxy terminal histidine tag, or to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16.

[0380] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding the inorganic pyrophosphatase of SEQ ID NO: 4, 6, 8, 10, 12, or 16.

[0381] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 10,

or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0382] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0383] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues.

[0384] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising an amino acid sequence comprising at least an amino acid residue difference at amino acid position 41, 65, 95, or 147, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

[0385] In some embodiments, the polynucleotide comprises a polynucleotide sequence encoding an engineered inorganic pyrophosphatase comprising the amino acid sequence comprising SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, or comprising SEQ ID NO: 18, 20, 22, 24, or 26.

[0386] In some embodiments, as described above, the polynucleotide encodes an engineered inorganic pyrophosphatase polypeptide with improved properties as compared to SEQ ID NOs: 10, 46, and/or 56, wherein the polypeptide comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity to a reference sequence selected from SEQ ID NOs: 10, 46, and/56,

wherein the sequence is selected from the even-numbered sequences in the range of SEQ ID NOs: 2-90. In some embodiments, the reference amino acid sequence is selected from the even-numbered sequences in the range of SEQ ID NOs: 2-90. In some embodiments, the reference amino acid sequence is SEQ ID NO: 10, while in some other embodiments, the reference sequence is SEQ ID NO: 46, while in some other embodiments, the reference sequence is SEQ ID NO: 56.

[0387] In some embodiments, the polynucleotide encodes an inorganic pyrophosphatase polypeptide capable of converting one or more substrates to product with improved properties as compared to SEQ ID NOs: 10, 46, and/or 56, wherein the polypeptide comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to reference sequence SEQ ID NOs: 10, 46, and/or 56, with or without the nucleotide sequence encoding the histidine-tag.

[0388] In some embodiments, the polynucleotide encoding the engineered inorganic pyrophosphatase comprises a polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 1-89, with or without the nucleotide sequence encoding the histidine-tag, as described herein. In some embodiments, the polynucleotide sequence comprises SEQ ID NOs: 9. 45, and/or 55, with or without the nucleotide sequence encoding the histidine-tag.

[0389] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference sequence selected from the odd-numbered sequences in SEQ ID NOs: 1-89, or a complement thereof, and encode an inorganic pyrophosphatase.

[0390] \In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the odd-numbered sequences in the range of SEQ ID NOs: 1-89 or a complement thereof and encode an inorganic pyrophosphatase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity to SEQ ID NOs: 10, 46, and/or 56, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 10, 46, and/or 56, as described above and in the Examples, below.

[0391] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered inorganic pyrophosphatase polypeptide with improved properties comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity to SEQ ID NOs: 10, 46, and/or 56. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity at the

nucleotide level to a reference polynucleotide encoding the engineered inorganic pyrophosphatase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 1-89.

[0392] In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered inorganic pyrophosphatase polypeptide with improved properties comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity to SEQ ID NOs: 10, 46, and/or 56. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered inorganic pyrophosphatase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 1-89.

[0393] In some embodiments, an isolated polynucleotide encoding any of the inorganic pyrophosphatase polypeptides, including the engineered inorganic pyrophosphatase polypeptides provided herein is manipulated in a variety of ways to provide for expression of the polypeptide. In some embodiments, the polynucleotides encoding the polypeptides are provided as expression vectors where one or more control sequences is present to regulate the expression of the polynucleotides and/or polypeptides. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

[0394] In some embodiments, the control sequences include among other sequences, promoters, leader sequences, polyadenylation sequences, propeptide sequences, signal peptide sequences, and transcription terminators. As known in the art, suitable promoters can be selected based on the host cells used. For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present application, include, but are not limited to the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (dagA), *Bacillus subtilis* levansucrase gene (sacB), *Bacillus licheniformis* alpha-amylase gene (amyL), *Bacillus stearothermophilus* maltogenic amylase gene (amyM), *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus licheniformis* penicillinase gene (penP), *Bacillus subtilis* xylA and xylB genes, and prokaryotic beta-lactamase gene (See e.g., Villa-Kamaroff et al., Proc. Natl Acad. Sci. USA 75: 3727-3731 [1978]), as well as the *tac* promoter (See e.g., DeBoer et al., Proc. Natl Acad. Sci. USA 80: 21-25 [1983]). Exemplary promoters for filamentous fungal host cells, include promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum*

trypsin-like protease (See e.g., WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof. Exemplary yeast cell promoters can be from the genes can be from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are known in the art (See e.g., Romanos et al., Yeast, 1992, 8:423-488). Exemplary promoters for use in insect cells include, but are not limited to, polyhedrin, p10, ELT, OpIE2, and hr5/ie1 promoters. Exemplary promoters for use in mammalian cells include, but are not limited to, those from cytomegalovirus (CMV), chicken β-actin promoter fused with the CMV enhancer, Simian virus 40 (SV40), from Homo sapiens phosphoglycerate kinase, beta actin, elongation factor-1a or glyceraldehyde-3-phosphate dehydrogenase, and from Gallus β-actin.

[0395] In some embodiments, the control sequence is a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice finds use in the present invention. For example, exemplary transcription terminators for filamentous fungal host cells can be obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease. Exemplary terminators for yeast host cells can be obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are known in the art (See e.g., Romanos et al., *supra*). Exemplary terminators for insect cells and mammalian cells include, but are not limited to, those from cytomegalovirus (CMV), Simian virus 40 (SV40), from Homo sapiens growth hormone hGH, from bovine growth hormone BGH, and from human or rabbit beta globulin.

[0396] In some embodiments, the control sequence is a suitable leader sequence, a non-translated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used. Exemplary leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells include but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP). The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of

the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells include but are not limited to those from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are also known in the art (See e.g., Guo and Sherman, Mol. Cell. Bio., 1995, 15:5983-5990).

[0397] In some embodiments, the control sequence is a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. Any signal peptide coding region that directs the expressed polypeptide into the secretory pathway of a host cell of choice finds use for expression of the engineered inorganic pyrophosphatase polypeptides provided herein. Effective signal peptide coding regions for bacterial host cells include but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NClB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. Further signal peptides are known in the art (See e.g., Simonen and Palva, Microbiol. Rev., 1993, 57:109-137). Effective signal peptide coding regions for filamentous fungal host cells include but are not limited to the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase. Useful signal peptides for yeast host cells include but are not limited to those from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Useful signal peptides for insect and mammalian host cells include but are not limited to, those from the genes for immunoglobulin gamma (IgG) and the signal peptide in a human secreted protein, such as human beta-galactosidase polypeptide.

[0398] In some embodiments, the control sequence is a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is referred to as a "proenzyme," "propolypeptide," or "zymogen," in some cases). A propolypeptide can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region includes but is not limited to the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila*

lactase (See e.g., WO 95/33836). Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

[0399] In some embodiments, regulatory sequences are also utilized. These sequences facilitate the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those that cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include, but are not limited to the lac, tac, and trp operator systems. In yeast host cells, suitable regulatory systems include, but are not limited to the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include, but are not limited to the TAKA alpha-amylase promoter, Aspergillus niger glucoamylase promoter, and Aspergillus oryzae glucoamylase promoter. Exemplary inducible promoters regulated by exogenous agents include the zinc-inducible sheep metallothionine (MT) promoter, dexamethasone (Dex)-inducible promoter, mouse mammary tumor virus (MMTV) promoter; ecdysone insect promoter, tetracycline-inducible promoter system, RU486-inducible promoter system, and the rapamycin-inducible promoter system.

[0400] The present invention also provides recombinant expression vectors comprising a polynucleotide encoding an engineered inorganic pyrophosphatase polypeptide, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. In some embodiments, the various nucleic acid and control sequences described above are combined together to produce a recombinant expression vector which includes one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the variant inorganic pyrophosphatase polypeptide at such sites. Alternatively, the polynucleotide sequence(s) of the present invention are expressed by inserting the polynucleotide sequence or a nucleic acid construct comprising the polynucleotide sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[0401] The recombinant expression 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 variant inorganic pyrophosphatase polynucleotide sequence. 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.

[0402] In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extra-chromosomal element, a minichromosome, or

an artificial chromosome). The vector may contain any means for assuring self-replication. In some alternative embodiments, 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. Furthermore, a single vector or plasmid or 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 may be used.

[0403] In some embodiment, the recombinant polynucleotides may be provided on a non-replicating expression vector or plasmid. In some embodiments, the non-replicating expression vector or plasmid can be based on viral vectors defective in replication (see, e.g., Travieso et al., npj Vaccines, 2022, Vol. 7, Article 75).

[0404] In some embodiments, the expression vector preferably contains one or more selectable markers, which permit easy selection of transformed cells. A "selectable marker" is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophy, and the like. Examples of bacterial selectable markers include but are not limited to the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferases), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof.

[0405] In another aspect, the present invention provides a host cell comprising a polynucleotide encoding at least one engineered inorganic pyrophosphatase polypeptide of the present invention, the polynucleotide being operatively linked to one or more control sequences for expression of the engineered inorganic pyrophosphatase enzyme(s) in the host cell. Host cells for use in expressing the polypeptides encoded by the expression vectors of the present invention are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, *Vibrio fluvialis*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* and *Pichia pastoris* [ATCC Accession No. 201178]); insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, BHK, 293, and Bowes melanoma cells; and plant cells. Exemplary host cells are *Escherichia coli* strains (e.g., W3110 (ΔfhuA) and BL21).

[0406] In some embodiments, the host cell strain comprises a knockout of one or more genes, in particular phosphatase genes. In some embodiments, the host cell comprises a knockout or single gene deletion of *E. coli genes aphA*, *surE*, *phoA*, and/or *cpdB*, as described below in the Examples. In some embodiments, the host cell comprising a knockout of one or more phosphatase genes has increased production of the product and/or decreased de-phosphorylation of the product or substrate.

[0407] Accordingly, in another aspect, the present invention provides methods for producing the engineered inorganic pyrophosphatase polypeptides, where the methods comprise culturing a host cell capable of expressing a polynucleotide encoding the engineered inorganic pyrophosphatase polypeptide under conditions suitable for expression of the polypeptide. In some embodiments, the methods further comprise the steps of isolating and/or purifying the inorganic pyrophosphatase polypeptides, as described herein.

[0408] In some embodiments, the inorganic pyrophosphatase polypeptide expressed in a host cell is recovered from the cells and/or the culture medium using any one or more of the known techniques for protein purification, including, among others, lysozyme or detergent treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography, such as described herein. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available (e.g., CelLytic B™, Sigma-Aldrich, St. Louis MO).

[0409] Chromatographic techniques for isolation/purification of the inorganic pyrophosphatase polypeptides include, among others, reverse phase chromatography, high-performance liquid chromatography, ion-exchange chromatography, hydrophobic-interaction chromatography, size-exclusion chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying the inorganic pyrophosphatase depends, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., and will be apparent to those having skill in the art. In some embodiments, affinity techniques may be used to isolate the inorganic pyrophosphatase. For affinity chromatography purification, an antibody that specifically binds inorganic pyrophosphatase polypeptide may be used. In some embodiments, an affinity tag, e.g., His-tag, can be introduced into the inorganic pyrophosphatase polypeptide for purposes of isolation/purification. In some embodiments, the affinity purification can use a specific ligand bound by the inorganic pyrophosphatase or dye affinity column (See e.g., EP0641862; Stellwagen, "Dye Affinity Chromatography," *In* Current Protocols in Protein Science, Unit 9.2-9.2.16, (2001)).

[0410] Appropriate culture media and growth conditions for the above-described host cells are well known in the art. Polynucleotides for expression of the inorganic pyrophosphatase polypeptides may be introduced into cells by various methods known in the art. Techniques include, among others, electroporation, biolistic particle bombardment, liposome mediated transfection, calcium chloride transfection, and protoplast fusion.

[0411] The engineered inorganic pyrophosphatases with the properties disclosed herein can be obtained by subjecting the polynucleotide encoding the naturally occurring or engineered inorganic pyrophosphatase polypeptide to mutagenesis and/or directed evolution methods known in the art, and as described herein. An exemplary directed evolution technique is mutagenesis and/or DNA shuffling (See e.g., Stemmer, Proc. Natl. Acad. Sci. USA, 1994, 91:10747-10751; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and U.S. Pat. 6,537,746). Other directed

evolution procedures that can be used include, among others, staggered extension process (StEP), in vitro recombination (See e.g., Zhao et al., Nat. Biotechnol., 1998, 16:258–261), mutagenic PCR (See e.g., Caldwell et al., PCR Methods Appl., 1994, 3:S136-S140), and cassette mutagenesis (See e.g., Black et al., Proc. Natl. Acad. Sci. USA, 1996, 93:3525-3529).

[0412] For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed. Mutagenesis and directed evolution methods are known in the art (See e.g., US Patent Nos. 5,605,793, 5,811,238, 5,830,721, 5,834,252, 5,837,458, 5,928,905, 6,096,548, 6,117,679, 6,132,970, 6,165,793, 6,180,406, 6,251,674, 6,265,201, 6,277,638, 6,287,861, 6,287,862, 6,291,242, 6,297,053, 6,303,344, 6,309,883, 6,319,713, 6,319,714, 6,323,030, 6,326,204, 6,335,160, 6,335,198, 6,344,356, 6,352,859, 6,355,484, 6,358,740, 6,358,742, 6,365,377, 6,365,408, 6,368,861, 6,372,497, 6,337,186, 6,376,246, 6,379,964, 6,387,702, 6,391,552, 6,391,640, 6,395,547, 6,406,855, 6,406,910, 6,413,745, 6,413,774, 6,420,175, 6,423,542, 6,426,224, 6,436,675, 6,444,468, 6,455,253, 6,479,652, 6,482,647, 6,483,011, 6,484,105, 6,489,146, 6,500,617, 6,500,639, 6,506,602, 6,506,603, 6,518,065, 6,519,065, 6,521,453, 6,528,311, 6,537,746, 6,573,098, 6,576,467, 6,579,678, 6,586,182, 6,602,986, 6,605,430, 6,613,514, 6,653,072, 6,686,515, 6,703,240, 6,716,631, 6,825,001, 6,902,922, 6,917,882, 6,946,296, 6,961,664, 6,995,017, 7,024,312, 7,058,515, 7,105,297, 7,148,054, 7,220,566, 7,288,375, 7,384,387, 7,421,347, 7,430,477, 7,462,469, 7,534,564, 7,620,500, 7,620,502, 7,629,170, 7,702,464, 7,747,391, 7,747,393, 7,751,986, 7,776,598, 7,783,428, 7,795,030, 7,853,410, 7,868,138, 7,783,428, 7,873,477, 7,873,499, 7,904,249, 7,957,912, 7,981,614, 8,014,961, 8,029,988, 8,048,674, 8,058,001, 8,076,138, 8,108,150, 8,170,806, 8,224,580, 8,377,681, 8,383,346, 8,457,903, 8,504,498, 8,589,085, 8,762,066, 8,768,871, 9,593,326, and all related US, as well as PCT and non-US counterparts; Ling et al., Anal. Biochem., 1997, 254(2):157-78; Dale et al., Meth. Mol. Biol., 1996, 57:369-74; Smith, Ann. Rev. Genet., 1985, 19:423-462; Botstein et al., Science, 1985, 229:1193-1201; Carter, Biochem. J., 1986, 237:1-7; Kramer et al., Cell, 1984, 38:879-887; Wells et al., Gene, 1985, 34:315-323; Minshull et al., Curr. Op. Chem. Biol., 1999, 3:284-290; Christians et al., Nat. Biotechnol., 1999, 17:259-264; Crameri et al., Nature, 1998, 391:288-291; Crameri, et al., Nat. Biotechnol., 1997, 15:436-438; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 1997, 94:4504-4509; Crameri et al., Nat. Biotechnol., 1996, 14:315-319; Stemmer, Nature, 1994, 370:389-391; Stemmer, Proc. Nat. Acad. Sci. USA, 1994, 91:10747-10751; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; and WO 2009/152336, all of which are incorporated herein by reference).

[0413] In some embodiments, the enzyme clones obtained following mutagenesis treatment are screened by subjecting the enzymes to a defined temperature (or other assay conditions, such as testing the enzyme's activity over a broad range of substrates) and measuring the amount of enzyme activity remaining after heat treatments or other assay conditions. Clones containing a polynucleotide encoding an inorganic pyrophosphatase polypeptide are then sequenced to identify the nucleotide sequence changes (if any) and used to express the enzyme in a host cell. Measuring enzyme activity

from the expression libraries can be performed using any suitable method known in the art (e.g., standard biochemistry techniques, such as HPLC analysis).

[0414] In some embodiments, the clones obtained following mutagenesis treatment can be screened for engineered inorganic pyrophosphatases having one or more desired improved enzyme properties (e.g., improved activity, reduced off target activity, increased stability, etc.). Measuring enzyme activity from the expression libraries can be performed using the standard biochemistry techniques, such as HPLC analysis, LC-MS analysis, RapidFire-MS analysis, and/or capillary electrophoresis analysis.

[0415] When the sequence of the engineered polypeptide is known, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (*e.g.*, by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides encoding portions of the inorganic pyrophosphatase can be prepared by chemical synthesis as known in the art (e.g., the classical phosphoramidite method of Beaucage et al., Tet. Lett., 1981, 22:1859-69, or the method described by Matthes et al., EMBO J., 1984, 3:801-05) as typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized (*e.g.*, in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In addition, essentially any nucleic acid can be obtained from any of a variety of commercial sources. In some embodiments, additional variations can be created by synthesizing oligonucleotides containing deletions, insertions, and/or substitutions, and combining the oligonucleotides in various permutations to create engineered inorganic pyrophosphatases with improved properties.

[0416] Accordingly, in some embodiments, a method for preparing the engineered inorganic pyrophosphatase polypeptide comprises: (a) synthesizing a polynucleotide encoding a polypeptide comprising an amino acid sequence having at least about 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to an amino acid sequence selected from SEQ ID NOs: 10, 46, and/or 56, and having one or more residue differences as compared to SEQ ID NOs: 10, 46, and/or 56; and (b) expressing the engineered inorganic pyrophosphatase polypeptide encoded by the polynucleotide.

[0417] In some embodiments of the method, the polynucleotide encodes an engineered inorganic pyrophosphatase that has optionally one or several (e.g., up to 3, 4, 5, or up to 10) amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, 1-50, 1-75, 1-100, or 1-150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally around 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 55, 60, 65, 70,

80, 90, 100, 110, 120, 130, 140, or 150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

Methods of Using the Inorganic Pyrophosphatase Enzymes

[0418] In some embodiments, the inorganic pyrophosphatase enzymes described herein find use in a composition or method under industrial process conditions, as described herein.

[0419] In the embodiments provided herein and illustrated in the Examples, various ranges of suitable reaction conditions that can be used in the processes, include but are not limited to, substrate loading, co-substrate loading, pH, temperature, buffer, solvent system, cofactor, polypeptide loading, and reaction time. Further suitable reaction conditions for carrying out the process for biocatalytic conversion of substrate compounds to product compounds using an engineered inorganic pyrophosphatase described herein can be determined in view of the guidance provided herein by experimentation that includes, but is not limited to, contacting the engineered inorganic pyrophosphatase polypeptide and one or more substrate compounds under experimental reaction conditions of concentration, pH, temperature, and solvent conditions, and detecting the product compound.

[0420] In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each substrate of at least about 25 μM to 50 μM, 50 μM to 100 μM, 100 μM to 200 μM, 200 μM to 300 μM, 300 μM to 500 μM, 500 μM to 1 mM, 1 mM to 1.5 mM, 1.5 mM to 3 mM, 3 mM to 5 mM, or 5 mM to 10 mM. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each substrate of at least about 0.5 g/L, at least about 1 g/L, at least about 5 g/L, at least about 10 g/L, at least about 15 g/L, at least about 20 g/L, or at least about 30 g/L, or even greater, as permissible for the substrate compound.

[0421] In some embodiments, where the reaction uses a second enzyme and a substrate for the second enzyme, the suitable reaction conditions comprise a substrate compound loading for each substrate for the second enzyme of at least about 25 μM to 50 μM, 50 μM to 100 μM, 100 μM to 200 μM, 200 μM to 300 μM, 300 μM to 500 μM, 500 μM to 1 mM, 1 mM to 1.5 mM, 1.5 mM to 3 mM, 3 mM to 5 mM, 5 mM to 10 mM, or 10 mM to 20 mM. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each substrate of the second enzyme of at least about 0.5 g/L, at least about 1 g/L, at least about 5 g/L, at least about 10 g/L, at least about 15 g/L, at least about 20 g/L, or at least about 30 g/L, or even greater, as permissible for the substrate compound for the second enzyme.

[0422] In carrying out the inorganic pyrophosphatase -mediated synthesis processes described herein, the engineered polypeptide, e.g., inorganic pyrophosphatase, may be added to the reaction mixture in the form of a purified enzyme, partially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, as cell extracts and/or lysates of such cells, and/or as an enzyme immobilized

on a solid support. Whole cells transformed with gene(s) encoding the engineered inorganic pyrophosphatase enzyme or cell extracts, lysates thereof, and isolated enzymes may be employed in a variety of different forms, including solid (e.g., lyophilized, spray-dried, and the like) or semisolid (e.g., a crude paste). The cell extracts or cell lysates may be partially purified by precipitation (ammonium sulfate, polyethyleneimine, heat treatment or the like, followed by a desalting procedure prior to lyophilization (e.g., ultrafiltration, dialysis, etc.). Any of the enzyme preparations (including whole cell preparations) may be stabilized by crosslinking using known crosslinking agents, such as, for example, glutaraldehyde or immobilization to a solid phase (e.g., Eupergit C, and the like). In some embodiments, where the reaction involves a second enzyme, the second enzyme may be provided in various forms, as described in the foregoing for the inorganic pyrophosphatase.

[0423] The gene(s) encoding the engineered inorganic pyrophosphatase polypeptides can be transformed into host cell separately or together into the same host cell. For example, in some embodiments one set of host cells can be transformed with gene(s) encoding one engineered inorganic pyrophosphatase polypeptide, and another set can be transformed with gene(s) encoding another inorganic pyrophosphatase. Both sets of transformed cells can be utilized together in the reaction mixture in the form of whole cells, or in the form of lysates or extracts derived therefrom. In other embodiments, a host cell can be transformed with gene(s) encoding multiple engineered inorganic pyrophosphatase polypeptides. In some embodiments the engineered polypeptides can be expressed in the form of secreted polypeptides, and the culture medium containing the secreted polypeptides can be used for the inorganic pyrophosphatase reaction.

[0424] In some embodiments, an improved activity of the engineered inorganic pyrophosphatase polypeptides disclosed herein provides for processes wherein higher percentage conversion can be achieved with lower concentrations of the engineered polypeptide. In some embodiments of the process, the suitable reaction conditions comprise an engineered polypeptide amount of about 1% (w/w), 2% (w/w), 5% (w/w), 10% (w/w), 20% (w/w), 30% (w/w), 40% (w/w), 50% (w/w), 75% (w/w), 100% (w/w) or more of substrate compound loading.

[0425] In some embodiments, the engineered polypeptide is present at a molar ratio of engineered polypeptide to substrate of about 50 to 1, 25 to 1, 10 to 1, 5 to 1, 1 to 1, 1 to 5, 1 to 10, 1 to 25, 1 to 50, 1 to 100, 1 to 250, or 1 to 500. In some embodiments, the engineered polypeptide is present at a molar ratio of engineered polypeptide to substrate from a range of about 50 to 1 to a range of about 1 to 500.

[0426] In some embodiments, the engineered polypeptide is present at about 1pg/L to about 1000 pg/L, about 1 ng/L to about 1000 ng/L; about 1 ug/L to about 1000 ug/L; about 1 mg/L to about 100 mg/L, ,0.01 g/L to about 50 g/L; about 0.01 to about 0.1 g/L; about 0.05 g/L to about 50 g/L; about 0.1 g/L to about 40 g/L; about 1 g/L to about 40 g/L; about 2 g/L to about 40 g/L; about 5 g/L to about 40 g/L; about 5 g/L to about 30 g/L; about 0.1 g/L to about 10 g/L; about 0.5 g/L to about 10 g/L;

about 1 g/L to about 10 g/L; about 0.1 g/L to about 5 g/L; about 0.5 g/L to about 5 g/L; or about 0.1 g/L to about 2 g/L. In some embodiments, the inorganic pyrophosphatase polypeptide is present at about 0.01 g/L, 0.05 g/L, 0.1 g/L, 0.2 g/L, 0.5 g/L, 1, 2 g/L, 5 g/L, 10 g/L, 15 g/L, 20 g/L, 25 g/L, 30 g/L, 35 g/L, 40 g/L, or 50 g/L.

[0427] In some embodiments, the suitable reaction conditions comprise a divalent metal cofactor. In some embodiments, the divalent metal cofactor is Mg^{+2}, Zn^{+2}, Co^{+2}, or Mn^{+2}, or any combination thereof. The divalent metal is present at concentrations of about 150 μM; about 200 μM; about 250 μM, about 500 μM; or about 1000 μM. In some embodiments, the divalent metal is present at concentrations of about 1 to 1000 μM; about 50 to 400 μM; about 100 to 300 μM; or about 200 to 600 μM; about 500 to 1000 μM.

[0428] In some embodiments, the divalent metal cofactor is cobalt. In some embodiments, the cobalt (II) chloride is present at concentrations of about 150 μM; about 200 μM; about 250 μM, about 500 μM; or about 1000 μM. In some embodiments, the cobalt (II) chloride is present at concentrations of about 1 to 1000 μM; about 50 to 400 μM; about 100 to 300 μM; or about 200 to 600 μM; about 500 to 1000 μM. In some embodiments, the cobalt (II) chloride is present at concentrations of about 150 μM; about 200 μM; about 250 μM, about 500 μM; or about 1000 μM.

[0429] During the course of the reaction, the pH of the reaction mixture may change. The pH of the reaction mixture may be maintained at a desired pH or within a desired pH range. This may be done by the addition of an acid or a base, before and/or during the course of the reaction. Alternatively, the pH may be controlled by using a buffer. Accordingly, in some embodiments, the reaction condition comprises a buffer. Suitable buffers to maintain desired pH ranges are known in the art and include, by way of example and not limitation, borate, phosphate, 2-(N-morpholino)ethanesulfonic acid (MES), 3-(N-morpholino)propanesulfonic acid (MOPS), acetate, triethanolamine (TEoA), and 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), and the like. In some embodiments, the reaction conditions comprise water as a suitable solvent with no buffer present.

[0430] In the embodiments of the process, the reaction conditions comprise a suitable pH. The desired pH or desired pH range can be maintained by use of an acid or base, an appropriate buffer, or a combination of buffering and acid or base addition. The pH of the reaction mixture can be controlled before and/or during the course of the reaction. In some embodiments, the suitable reaction conditions comprise a solution pH from about 4 to about 10, pH from about 5 to about 10, pH from about 5 to about 9, pH from about 6 to about 9, pH from about 6 to about 8. In some embodiments, the reaction conditions comprise a solution pH of about 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, or 10.

[0431] In the embodiments of the processes herein, a suitable temperature is used for the reaction conditions, for example, taking into consideration the increase in reaction rate at higher temperatures, and the activity of the enzyme during the reaction time period. Accordingly, in some embodiments,

the suitable reaction conditions comprise a temperature of about 10 °C to about 95 °C, about 10 °C to about 75 °C, about 15 °C to about 95 °C, about 20 °C to about 95 °C, about 20 °C to about 65 °C, about 25 °C to about 70 °C, or about 50 °C to about 70 °C. In some embodiments, the suitable reaction conditions comprise a temperature of about 10 °C, 15 °C, 20 °C, 25 °C, 30 °C, 35 °C, 40 °C, 45 °C, 50 °C, 55 °C, 60 °C, 65 °C, 70 °C, 75 °C, 80 °C, 85 °C, 90 °C or 95 °C. In some embodiments, the temperature during the enzymatic reaction can be maintained at a specific temperature throughout the course of the reaction. In some embodiments, the temperature during the enzymatic reaction can be adjusted over a temperature profile during the course of the reaction.

[0432] In some embodiments, the processes of the invention are carried out in a solvent. Suitable solvents include water, aqueous buffer solutions, organic solvents, polymeric solvents, and/or co-solvent systems, which generally comprise aqueous solvents, organic solvents and/or polymeric solvents. The aqueous solvent (water or aqueous co-solvent system) may be pH-buffered or unbuffered. In some embodiments, the processes using the engineered inorganic pyrophosphatase polypeptides can be carried out in an aqueous co-solvent system comprising an organic solvent (e.g., ethanol, isopropanol (IPA), dimethyl sulfoxide (DMSO), dimethylformamide (DMF) ethyl acetate, butyl acetate, 1-octanol, heptane, octane, methyl t butyl ether (MTBE), toluene, and the like), ionic or polar solvents (e.g., 1-ethyl 4 methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate, 1-butyl 3 methylimidazolium hexafluorophosphate, glycerol, polyethylene glycols, and the like). In some embodiments, the co-solvent can be a polar solvent, such as a polyol, dimethylsulfoxide (DMSO), or lower alcohol. The non-aqueous co- solvent component of an aqueous co-solvent system may be miscible with the aqueous component, providing a single liquid phase, or may be partly miscible or immiscible with the aqueous component, providing two liquid phases. Exemplary aqueous co-solvent systems can comprise water and one or more co-solvents selected from an organic solvent, polar solvent, and polyol solvent. In general, the co-solvent component of an aqueous co-solvent system is chosen such that it does not adversely inactivate the inorganic pyrophosphatase enzyme under the reaction conditions. Appropriate co-solvent systems can be readily identified by measuring the enzymatic activity of the specified engineered inorganic pyrophosphatase enzyme with a defined substrate of interest in the candidate solvent system, utilizing an enzyme activity assay, such as those described herein.

[0433] In some embodiments of the process, the suitable reaction conditions comprise an aqueous co-solvent, where the co-solvent comprises DMSO at about 1% to about 50% (v/v), about 1 to about 40% (v/v), about 2% to about 40% (v/v), about 5% to about 30% (v/v), about 10% to about 30% (v/v), or about 10% to about 20% (v/v). In some embodiments of the process, the suitable reaction conditions can comprise an aqueous co-solvent comprising ethanol at about 1% (v/v), about 5% (v/v), about 10% (v/v), about 15% (v/v), about 20% (v/v), about 25% (v/v), about 30% (v/v), about 35% (v/v), about 40% (v/v), about 45% (v/v), or about 50% (v/v).

[0434] In some embodiments, the reaction conditions comprise a surfactant for stabilizing or enhancing the reaction. Surfactants can comprise non-ionic, cationic, anionic and/or amphiphilic surfactants. Exemplary surfactants, include by way of example and not limitation, nonyl phenoxypolyethoxylethanol (NP40), TRITON™ X-100 polyethylene glycol *tert*-octylphenyl ether, polyoxyethylene-stearylamine, cetyltrimethylammonium bromide, sodium oleylamidosulfate, polyoxyethylene-sorbitanmonostearate, hexadecyldimethylamine, etc. Any surfactant that may stabilize or enhance the reaction may be employed. The concentration of the surfactant to be employed in the reaction may be generally from 0.1 to 50 mg/mL, particularly from 1 to 20 mg/mL.

[0435] In some embodiments, the reaction conditions include an antifoam agent, which aids in reducing or preventing formation of foam in the reaction solution, such as when the reaction solutions are mixed or sparged. Anti-foam agents include non-polar oils (e.g., minerals, silicones, etc.), polar oils (e.g., fatty acids, alkyl amines, alkyl amides, alkyl sulfates, etc.), and hydrophobic (e.g., treated silica, polypropylene, etc.), some of which also function as surfactants. Exemplary anti-foam agents include Y-30® (Dow Corning), poly-glycol copolymers, oxy/ethoxylated alcohols, and polydimethylsiloxanes. In some embodiments, the anti-foam can be present at about 0.001% (v/v) to about 5% (v/v), about 0.01% (v/v) to about 5% (v/v), about 0.1% (v/v) to about 5% (v/v), or about 0.1% (v/v) to about 2% (v/v). In some embodiments, the anti-foam agent can be present at about 0.001% (v/v), about 0.01% (v/v), about 0.1% (v/v), about 0.5% (v/v), about 1% (v/v), about 2% (v/v), about 3% (v/v), about 4% (v/v), or about 5% (v/v) or more as desirable to promote the reaction.

[0436] The quantities of reactants used in the inorganic pyrophosphatase reaction will generally vary depending on the quantities of product desired, and concomitantly the amount of substrate employed. Those having ordinary skill in the art will understand how to vary these quantities to tailor them to the desired level of productivity and scale of production in view of the guidance provided in the present disclosure.

[0437] In some embodiments, the order of addition of reactants is not critical. The reactants may be added together at the same time to a solvent (e.g., monophasic solvent, biphasic aqueous co-solvent system, and the like), or alternatively, some of the reactants may be added separately, and some together at different time points. For example, the cofactor, co-substrate and substrate may be added first to the solvent.

[0438] The solid reactants (e.g., enzyme, salts, etc.) may be provided to the reaction in a variety of different forms, including powder (e.g., lyophilized, spray dried, and the like), solution, emulsion, suspension, and the like. The reactants can be readily lyophilized or spray dried using methods and equipment that are known to those having ordinary skill in the art. For example, the protein solution can be frozen at -80 °C in small aliquots, then added to a pre-chilled lyophilization chamber, followed by the application of a vacuum.

[0439] For improved mixing efficiency when an aqueous co-solvent system is used, the inorganic pyrophosphatase, and co-substrate may be added and mixed into the aqueous phase first. The substrate may be added and mixed in, followed by the organic phase or the substrate may be dissolved in the organic phase and mixed in. Alternatively, the substrate may be premixed in the organic phase, prior to addition to the aqueous phase.

[0440] The processes of the present invention are generally allowed to proceed until further conversion of substrate to product does not change significantly with reaction time (e.g., less than 10% of substrate being converted, or less than 5% of substrate being converted). In some embodiments, the reaction is allowed to proceed until there is complete or near complete conversion of substrate to product. Transformation of substrate to product can be monitored using known methods by detecting substrate and/or product, with or without derivatization. Suitable analytical methods include gas chromatography, HPLC, MS, and the like. In some embodiments, after suitable conversion to product, the reactants are separated from the product and additional reactants are added to the product.

[0441] Any of the processes disclosed herein using the engineered polypeptides for the preparation of products can be carried out under a range of suitable reaction conditions, including but not limited to ranges of substrates, temperature, pH, solvent system, substrate loading, polypeptide loading, cofactor loading, and reaction time. In some embodiments, additional reaction components or additional techniques carried out to supplement the reaction conditions. These can include taking measures to stabilize or prevent inactivation of the enzyme, reduce product inhibition, shift reaction equilibrium to formation of the desired product.

[0442] In some embodiments, the engineered inorganic pyrophosphatase polypeptides can be provided on a solid support, such as a membrane, resin, solid carrier, or other solid phase material. A solid support can be composed of organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of a solid support can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar. Solid supports can be porous or non-porous and can have swelling or non-swelling characteristics. A solid support can be configured in the form of a well, depression, or other container, vessel, feature, or location.

[0443] In some embodiments, the engineered inorganic pyrophosphatase polypeptides of the present invention can be immobilized on a solid support such that they retain their improved activity, and/or other improved properties relative to the reference polypeptide of SEQ ID NOs: 10, 46, and/or 56. In such embodiments, the immobilized polypeptides can facilitate the biocatalytic conversion of the substrate compounds or other suitable substrates to the product and after the reaction is complete are

easily retained (e.g., by retaining beads on which polypeptide is immobilized) and then reused or recycled in subsequent reactions. Such immobilized enzyme processes allow for further efficiency and cost reduction. Accordingly, it is further contemplated that any of the compositions or methods of using the inorganic pyrophosphatase polypeptides of the present invention can be carried out using the inorganic pyrophosphatase polypeptides bound or immobilized on a solid support.

[0444] Methods of enzyme immobilization are known in the art. The engineered polypeptides can be bound non-covalently or covalently. Various methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are known in the art (See e.g., Yi et al., Proc. Biochem., 2007, 42(5):895-898; Martin et al., Appl. Microbiol. Biotechnol.,2007, 76(4): 843-851; Koszelewski et al., J. Mol. Cat. B: Enzymatic, 2010, 63:39-44; Truppo et al., Org. Proc. Res. Dev., published online: dx.doi.org/10.1021/op200157c; Hermanson, Bioconjugate Techniques, 2nd ed., Academic Press, Cambridge, MA (2008); Mateo et al., Biotechnol. Prog., 2002, 18(3):629-34; and "Bioconjugation Protocols: Strategies and Methods," *In* Methods in Molecular Biology, Niemeyer (ed.), Humana Press, New York, NY (2004); the disclosures of each which are incorporated by reference herein). Solid supports useful for immobilizing the engineered inorganic pyrophosphatase of the present invention include but are not limited to beads or resins comprising polymethacrylate with epoxide functional groups, polymethacrylate with amino epoxide functional groups, styrene/DVB copolymer or polymethacrylate with octadecyl functional groups. Exemplary solid supports useful for immobilizing the engineered inorganic pyrophosphatase polypeptides of the present invention include, but are not limited to, EnginZyme (including, EziG-1, EziG-1, and EziG-3), chitosan beads, Eupergit C, and SEPABEADs (Mitsubishi) (including EC-EP, EC-HFA/S, EXA252, EXE119 and EXE120), and are also provided in the Examples. In some embodiments, the solid support comprises IB-COV7 resin, an polyacrylate resin with epoxide groups for covalent immobilization.

[0445] In some embodiments, the inorganic pyrophosphatase is co-immobilized of a support medium with a second enzyme used in a reaction. In some embodiments, the inorganic pyrophosphatase is co-immobilized with a DNA polymerase, an RNA polymerase, a nucleotide polymerase, a DNA ligase, an RNA ligase, a nucleotide ligase, a double-stranded nucleotide ligase, a single-stranded nucleotide ligase, a template independent polymerase, or a terminal nucleotidyl transferase. Other second enzyme that can be used co-immobilized with an inorganic pyrophosphatase is described herein.

[0446] In some embodiments, as discussed herein, the inorganic pyrophosphatase polypeptides may be modified by the addition, deletion, or insertion of one or more amino acid residues to facilitate immobilization. In some embodiments, introduction of 1, 3 or 5 lysine residues at the N or C-terminal tags of the inorganic pyrophosphatase may facilitate immobilization and increase soluble protein yield, as described in Examples 25 and 26. In some embodiments, one or more lysine residues may

be added to the N or C terminus of any of the inorganic pyrophosphatase polypeptides of SEQ ID NOs: 2-90.

[0447] In further embodiments, any of the above-described processes for the conversion of one or more substrate compounds to product compound can further comprise one or more steps selected from: extraction; isolation; purification; and crystallization of product compound. As is known to those skilled in the art, acidic compounds such as NTPs may exist in various salt forms that can be used interchangeably in the methods described herein. All such forms are specifically envisaged for use in the methods described herein. Methods, techniques, and protocols for extracting, isolating, purifying, and/or crystallizing the product from biocatalytic reaction mixtures produced by the above disclosed processes are known to the ordinary artisan and/or accessed through routine experimentation. Additionally, illustrative methods are provided in the Examples below.

[0448] Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXAMPLES

[0449] The following Examples, including experiments and results achieved, are provided for illustrative purposes only and are not to be construed as limiting the present invention. Indeed, there are various suitable sources for many of the reagents and equipment described below. It is not intended that the present invention be limited to any particular source for any reagent or equipment item.

[0450] In the experimental disclosure below, the following abbreviations apply: M (molar); mM (millimolar), μM and uM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg (micrograms); L and l (liter); ml and mL (milliliter); ul, μl, uL, and μL (microliter); cm (centimeters); mm (millimeters); μM and μιη (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); psi and PSI (pounds per square inch); °C (degrees Celsius); RT and rt (room temperature); CV (coefficient of variability); CAM and cam (chloramphenicol); PMBS (polymyxin B sulfate); IPTG (isopropyl β-D-l-thiogalactopyranoside); LB (lysogeny broth); TB (terrific broth); SFP (shake flask powder); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); nt (nucleotide; polynucleotide); aa (amino acid; polypeptide); E. coli W3110 (commonly used laboratory E. coli strain, available from the Coli Genetic Stock Center [CGSC], New Haven, CT); HTP (high throughput); HPLC (high pressure liquid chromatography); HPLC-UV (HPLC-Ultraviolet Visible Detector); 1H NMR (proton nuclear magnetic resonance spectroscopy); FIOPC (fold improvements over positive control); Sigma and Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO; Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Microfluidics (Microfluidics, Westwood, MA); Life Technologies (Life Technologies, a part of Fisher Scientific, Waltham, MA); Amresco (Amresco, LLC, Solon, OH); Carbosynth (Carbosynth, Ltd., Berkshire, UK); Varian (Varian Medical

Systems, Palo Alto, CA); Agilent (Agilent Technologies, Inc., Santa Clara, CA); Infors (Infors USA Inc., Annapolis Junction, MD); and Thermotron (Thermotron, Inc., Holland, MI).

Example 1

Inorganic pyrophosphatase (IPP) gene Acquisition and Construction of Expression Vectors

[0451] Family I inorganic pyrophosphatase (IPP) enzymes from bacteria were acquired as synthetic gene constructs SEQ ID NOs: 2, 4-16 encoding a *C*-terminal 6-histidine tagged version of the WT IPP. Family II IPP enzymes from bacteria, archaea, and eukaryotes were acquired as synthetic gene constructs SEQ ID NOs: 28-52 encoding a *N*-terminal 6-histidine tagged version of the WT IPP. Genes were designed with codon optimization for *E. coli* expression, synthesized, and subcloned into the *E. coli* expression vector pCK100900i (See e.g., US Pat. No. 7,629,157 and US Pat. Appln. Publn. 2016/0244787, both of which are hereby incorporated by reference). Each plasmid construct was transformed into an *E. coli* strain derived from W3110. Directed evolution techniques were used to generate libraries of gene variants from these plasmids (See e.g., US Pat. No. 8,383,346 and WO 2010/144103, both of which are hereby incorporated by reference). The substitutions in the enzyme variants described herein are indicated with reference to the 6-histidine tagged version of the WT IPP enzymes or variants thereof, as indicated.

Example 2

Shake Flask Expression and Purification of IPP Family I Enzymes

Shake Flask Expression

[0452] The enzymes described above were streaked out onto agar plates containing 30 μg/mL chloramphenicol and 0.01% glucose (v/v), for single colonies. Then, a single colony was used to inoculate a 5-mL tube containing Terrific Broth (TB) media (Teknova) and 30 μg/mL chloramphenicol, which was then shaking for 20 hours at 30 °C, 250 rpm. A glycerol stock was then prepared by mixing the overnight culture with 50% glycerol at a 2:1 ratio. These cultures were thawed and 30 μL from each glycerol stock was used to inoculate 1 L shake flasks containing: 160 mL Terrific Broth (TB) media (Teknova), 30 μg/mL chloramphenicol, 0.03% (v/v) lactose and 0.075% (v/v) glucose. The shake flasks were grown at 32 °C for 18 hours at 250 rpm. Following this incubation period, the cultures were centrifuged at 4,000 rpm for 10 min at 4 °C. The culture supernatant was discarded, and the pellets were resuspended in 30 mL of 50 mM Tris-HCl, pH 7.5. The cell suspension was chilled in an ice bath and lysed using a Microfluidizer cell disruptor (Microfluidics M-110L). The crude lysate was pelleted by centrifugation (10,000 rpm for 90 min at 4 °C) in an Avanti J-series centrifuge and JLA-16.250 fixed rotor (Beckman Coulter). The supernatant was then filtered through a 0.2 μm PES membrane to further clarify the lysate. The clarified IPP lysates were then supplemented with 20 mM imidazole and 500 mM NaCl.

Purification of IPP from Shake Flask Lysates

[0453] The clarified IPP lysates were then purified using an AKTA Start purification system and a 5 mL HisTrap FF column (GE Healthcare) using the AC Step HiF setting (the run parameters are provided below in **Table 1.2**). The SF wash buffer comprised 50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, 0.02% v/v Triton X-100 reagent.

Table 1.2 Purification Parameters

Parameter	Volume
Column volume	5 mL
Flow rate	8 mL/min
Pressure limit	0.3 MPa
Sample volume	35 mL
Equilibration volume	5 column volumes (CV) = 25 mL
Wash Unbound volume	20 CV = 100 mLs
Elution	Isocratic (step)
Elution volume	5 CV = 25 mLs
Fraction volume	1.5 mLs
RE-equilibration volume	5 CV = 25 mLs

[0454] Elution fractions containing protein were identified by UV absorption (A280) and pooled, then dialyzed overnight in dialysis buffer (20 mM Tris-HCl, pH 7.4, 100 mM KCl, 0.1 mM EDTA, and 50% glycerol) in a 3.5K Slide-A-Lyzer™ dialysis cassette (Thermo Fisher) for buffer exchange. IPP concentrations from the preparations were measured by absorption at 280 nm.

[0455] Elution fractions containing protein were identified by UV absorption (A280) and pooled, then dialyzed overnight in dialysis buffer (20 mM Tris-HCl, pH 7.4, 100 mM KCl, 0.1 mM EDTA, and 50% glycerol) in a 3.5K Slide-A-Lyzer™ dialysis cassette (Thermo Fisher) for buffer exchange. IPP concentrations in the preparations were measured by absorption at 280 nm. Soluble protein concentrations are summarized in **Table 2** below, showing a fold improvement in soluble protein production following shake-flask purification relative to the inorganic pyrophosphatase from *Haloferax sp.*, SEQ ID NO: 2.

[0456] Protein solubility is defined as the measured protein conc. (mg/mL) at A280 divided by the product of the extinction coefficient and the molecular weight of each enzyme.

Table 2. Soluble Enzyme Production of Family I Enzymes Relative to SEQ ID NO: 2

SEQ ID NO: (nt/aa)	FIOP 1	Source organism
3/4	++	*Thermocrinis ruber*
5/6	++	*Aquifex pyrophilus*
7/8	++	*Thermus oshimai*
9/10	+++	*Sulfolobus sp. A20*

Table 2. Soluble Enzyme Production of Family I Enzymes Relative to SEQ ID NO: 2		
SEQ ID NO: (nt/aa)	**FIOP 1**	**Source organism**
11/12	+++	*Geobacillus zalihae*
13/14	+++	*Bacillus thermozeamaize*
15/16	++	*Bacillus smithii*
1/2		*Haloferax sp.*
Levels of increased activity were determined for FIOP 1 relative to SEQ ID NO: 2 and defined as follows: "+" > 1.0, "++" > 1.58, "+++" > 5.51.		

Example 3

Shake Flask Expression and Purification of IPP Family I Evolved Variants

[0457] IPP enzymes of SEQ ID NOs: 10 and 18-26 were produced in shake flask and purified as described in Example 2. Soluble protein concentrations are summarized in **Table 3** below, showing a fold improvement in soluble protein production following shake-flask purification relative to the inorganic pyrophosphatase from *Sulfolobus sp. A20.* SEQ ID NO: 10.

Table 3. Soluble Enzyme Production Relative to SEQ ID NO: 10		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 10)**	**FIOP 1**
17/18	E95S	+++
19/20	P65R	+++
21/22	P65K	++
23/24	V147A	+
25/26	V41F	++
Levels of increased activity were determined FIOP activity 1 relative to SEQ ID NO: 10 and are defined as follows: "+" > 0.05, "++" > 0.38, "+++" > 0.65.		

Example 4

Shake Flask Expression and Purification of IPP Family II Enzymes

[0458] IPP enzymes of SEQ ID NOs: 28-52 were produced in shake flask and purified as described in Example 2. Soluble protein concentrations are summarized in **Table 4** below, showing a fold improvement in soluble protein production following shake-flask purification relative to the inorganic pyrophosphatase from *Ostreococcus tauri*, SEQ ID NO: 28.

Table 4: Soluble Enzyme Production Relative to SEQ ID NO: 28		
SEQ ID NO: (nt/aa)	**FIOP 1**	**Source organism**
29/30	+++	*Caldibacillus thermoamylovorans*

Table 4: Soluble Enzyme Production Relative to SEQ ID NO: 28		
31/32	+++	*Streptococcus thermophilus*
33/34	++++	*Thermobacillus xylanilyticus*
35/36	++++	*Thermodesulfovibrio yellowstonii*
37/38	++	*Methanococcus aeolicus*
39/40	++++	*Methanothermococcus sp.*
41/42	++	*Methanothermococcus sp. SCGC AD-155-M21*
43/44	+++	*Methanothermococcus okinawensis*
45/46	+++	*Methanotorris igneus*
47/48	+	*Idotea baltica*
49/50	+++	*Adineta ricciae*
51/52	++	*Bathycoccus prasinos*
27/28		*Ostreococcus tauri*
Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 28 and defined as follows: "+" > 1.0, "++" > 1.36, "+++" > 4.51, "++++" > 7.0.		

Example 5

Shake Flask Expression and Purification of IPP Family II Evolved Variants

[0459] IPP enzymes of SEQ ID NOs: 46 and 54-74, were produced in shake flask and purified as described in Example 2. Soluble protein concentrations are summarized in **Table 5** below, showing a fold improvement in soluble protein production following shake-flask purification relative to the inorganic pyrophosphatase from *Methanotorris igneus*, SEQ ID NO: 46.

Table 5. Soluble Enzyme Production Relative to SEQ ID NO: 46		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 46)**	**FIOP 1**
53/54	D45E	+
55/56	N188E	+++
57/58	E89D	+
59/60	L85V	++++
61/62	I187L	+++
63/64	K177E	++
65/66	K81R/D86S/N188E	++++
67/68	D86S/E89D/V160L/N188E/V308I	+
69/70	D169E/N188E/V231T	++
71/72	I110L/D169E/N188E/V231T	++++
73/74	Y14L/N188E/K269N	+
Levels of increased activity were determined FIOP activity 1 relative to SEQ ID NO: 46 and are defined as follows: "+" > 0.63, "++" > 1.0, "+++" > 1.08, "++++" > 1.23.		

Example 6

Measurement of Pyrophosphatase Activity by Absorbance Detection Assay of Family I Enzymes

[0460] IPP enzymes of SEQ ID NOs: 2-16 were produced in shake flask and purified as described in Example 2.

[0461] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 100 µM pyrophosphate, 1.25 nM IPP, 100 mM triethanolamine (pH 7.8), and 1 mM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for IPP, were pre-mixed in a single solution, and was aliquoted into each well of the 96-well plates (ii) IPP solution was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at 50 °C for 60 minutes, then held at 4 °C until the reaction was quenched. The reaction was quenched with 20 uL of 75% MeOH and the plate was vortexed and centrifuged for 5 minutes at 4000 rpm. The reactions were further diluted with 30 uL milli-Q water. After the reaction quench and dilution, the phosphate detection assay was setup using 96-well Greiner clear bottom plates and a commercially available phosphate detection kit (Sigma Aldrich, *MAK308*). 75 uL of the Malachite Green Reagent was mixed with 25 uL of the processed reaction, which was then left to incubate at room temperature for 30 minutes. After the incubation period, the plate was assayed by capturing the endpoint absorbance at 620 nm. For analysis, the blank well values were subtracted from sample well values for an accurate capture of absorbance. Absorbance FIOP was calculated as the absorbance of the enzyme divided by the absorbance relative to the inorganic pyrophosphatase from *Haloferax sp* SEQ ID NO: 2. The results are shown in **Table 6**.

Table 6: IPP Family I Inorganic Pyrophosphatase Activity Detection by Absorbance at 620 nm relative to SEQ ID NO: 2

SEQ ID NO: (nt/aa)	FIOP 1	Source organism
3/4	+++	*Thermocrinis ruber*
5/6	++	*Aquifex pyrophilus*
7/8	+++	*Thermus oshimai*
9/10	++	*Sulfolobus sp. A20*
11/12	++++	*Geobacillus zalihae*
13/14	++++	*Bacillus thermozeamaize*
15/16	++++	*Bacillus smithii*
Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 2 and defined as follows: "+" > 1.0, "++" > 9.7, "+++" > 38.3, "++++" > 120.8.		

Example 7

Measurement of Pyrophosphatase Activity of Family I Evolved Variants

[0462] IPP enzymes of SEQ ID NOs: 10 and 18-26 were produced in shake flask and purified as described in Example 2.

[0463] Reactions were performed and processed as described in example 6. Absorbance FIOP was calculated as the absorbance of the enzyme at 260 nm divided by the absorbance relative to the inorganic pyrophosphatase from *Sulfolobus sp. A20.* SEQ ID NO: 10. The results are shown in **Table 7.**

Table 7: Inorganic Pyrophosphatase Activity Detection by Absorbance at 620 nm Relative to SEQ ID NO: 10		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 10)**	**FIOP 1**
17/18	E95S	+
19/20	P65R	+++
21/22	P65K	+
23/24	V147A	+++
25/26	V41F	+++
Levels of increased activity were determined FIOP activity 1 relative to SEQ ID NO: 10 and are defined as follows: "+" > 0.82, "++" > 1.0, "+++" > 2.38.		

Example 8

Measurement of Pyrophosphatase Activity of Family II Enzymes

[0464] IPP enzymes of SEQ ID NOs: 28-52 were produced in shake flask and purified as described in Example 2.

[0465] Reactions were performed and processed as described in example 6. Absorbance FIOP was calculated as the absorbance of the enzyme divided by the absorbance relative to the inorganic pyrophosphatase from *Ostreococcus tauri,* SEQ ID NO: 28. The results are shown in **Table 8.**

Table 8: IPP Family II Inorganic Pyrophosphatase Activity Detection by Absorbance at 620 nm Relative to SEQ ID NO: 28		
SEQ ID NO: (nt/aa)	**FIOP 1**	**Source organism**
29/30	++	*Caldibacillus thermoamylovorans*
31/32	++	*Streptococcus thermophilus*
33/34	+++	*Thermobacillus xylanilyticus*
35/36	+++	*Thermodesulfovibrio yellowstonii*
37/38	++++	*Methanococcus aeolicus*
39/40	++++	*Methanothermococcus sp.*
41/42	++++	*Methanothermococcus sp. SCGC AD-155-M21*
43/44	++++	*Methanothermococcus okinawensis*
45/46	++++	*Methanotorris igneus*

Table 8: IPP Family II Inorganic Pyrophosphatase Activity Detection by Absorbance at 620 nm Relative to SEQ ID NO: 28		
SEQ ID NO: (nt/aa)	**FIOP 1**	**Source organism**
47/48	+++	*Idotea baltica*
49/50	+++	*Adineta ricciae*
51/52	++	*Bathycoccus prasinos*
Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 28 and defined as follows: "+" > 1.0, "++" > 9.7, "+++" > 38.3, "++++" > 120.8.		

Example 9

Measurement of Pyrophosphatase Activity by of Family II Evolved Variants

[0466] IPP enzymes of SEQ ID NOs: 46 and 54-74, were produced in shake flask and purified as described in Example 2.

[0467] Reactions were performed and processed as described in example 6. Absorbance FIOP was calculated as the absorbance of the enzyme divided by the absorbance relative to the inorganic pyrophosphatase from *Methanotorris igneus*, SEQ ID NO: 46. The results are shown in **Table 9**.

Table 9: Inorganic Pyrophosphatase Activity Detection by Absorbance at 620 nm Relative to SEQ ID NO: 46		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 46)**	**FIOP 1**
53/54	D45E	+++
55/56	N188E	+++
57/58	E89D	++++
59/60	L85V	++++
61/62	I187L	++++
63/64	K177E	+++
65/66	K81R/D86S/N188E	+++
67/68	D86S/E89D/V160L/N188E/V308I	+++
69/70	D169E/N188E/V231T	++
71/72	I110L/D169E/N188E/V231T	+
73/74	Y14L/N188E/K269N	+
Levels of increased activity were determined FIOP activity 1 relative to SEQ ID NO: 46 and are defined as follows: "+" > 0.89, "++" > 1.0, "+++" > 1.07, "++++" > 1.17.		

Example 10

Measurement of Pyrophosphatase Activity of Family I, Family II and Commercial Enzymes

[0468] IPP enzymes of SEQ ID NOs: 10, 12, 46, and 54-74 were produced in shake flask and purified as described in Example 2.

[0469] Reactions were performed and processed as described in example 6. Absorbance FIOP was calculated as the absorbance of the enzyme divided by the absorbance relative to the inorganic pyrophosphatase from *Sulfolobus sp. A20* (SEQ ID NO: 10). The results are shown in **Table 10**.

Table 10: IPP Family I, Family II and Commercial Enzymes Absorbance at 620 nm Relative to SEQ ID NO: 10

SEQ ID NO: (nt/aa)	FIOP 1
11/12	+
45/46	++
53/54	++
55/56	++
57/58	+++
59/60	+++
61/62	+++
63/64	++
65/66	+++
67/68	++
69/70	++
71/72	+
73/74	+
Yeast IPP (M2403L)	
Thermal Stable IPP (M0296L)	+
E.coli IPP (M0361L)	+
Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 10 and defined as follows: "+" > 15.2, "++" > 22.0, "+++" > 24.9.	

Example 11

Off-target Phosphate Hydrolysis of IPP Family I Enzymes with Nucleotide Triphosphates

[0470] IPP enzymes of SEQ ID NOs: 2-16 were produced in shake flask and purified as described in Example 2. Off target hydrolytic activity of IPPs against 3'-phos-mATP and ATP was examined by incubating IPP with the nucleotide with species decomposition analyzed by HPLC (see Example 12).

[0471] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 1 mM nucleotide triphosphate, 20 µM IPP, 100 mM triethanolamine (pH 7.8), and 1 mM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for IPP, were pre-mixed in a single solution, and was aliquoted into each well of the 96-well plates (ii) IPP solution was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable

aluminum seal and incubated in a thermocycler at 50 °C for 4 hours then held at 4 °C until the reaction was quenched. The reactions were quenched by the addition of 75 uL of 50 % MeOH, vortexed and centrifuged for 10 minutes at 4,000 rpm. The reactions were further diluted 1:1 with milli-Q water, to a nucleotide triphosphate concentration of 50 uM. The samples were then analyzed on HPLC, as described in example 12. FIOP 1 was calculated as the percent of the ATP remaining of the enzyme divided by the percent of ATP remaining relative to the inorganic pyrophosphatase from *Thermus Oshimai* SEQ ID NO: 8. FIOP 2 was calculated as the percent of the 3'-phosphate mATP remaining of the enzyme divided by the percent of 3'-phos-mATP remaining relative to the inorganic pyrophosphatase from *Thermus Oshimai* SEQ ID NO: 8. The results are shown in **Table 11**.

Table 11. IPP Decomposition of ATP and 3'-Phos-mATP Relative to SEQ ID NO: 8

SEQ ID NO: (nt/aa)	FIOP Activity 1 Relative to SEQ ID NO: 8	FIOP Activity 2 Relative to SEQ ID NO: 8
3/4	+++	+
5/6	+++	++
7/8	+	+
9/10	+++	+
11/12		+
13/14		+
15/16		+
1/2	++++	++

Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 8 and for activity 2 relative to SEQ ID NO: 8 and defined as follows: "+" > 0.48, "++" > 1.0, "+++" > 13.5, "++++" > 135.0. "+" > 0.48, "++" > 1.0, "+++" > 13.5, "++++" > 135.0.

Example 12

High performance liquid chromatography (HPLC) analysis of nucleotides

[0472] Samples reactions are quenched and diluted as described above in example 11. 100 µL is transferred to a 96 well round bottom plate and then 5 µL is injected onto an Ultimate 3000 HPLC system using a PAL autosampler according to the method outlined in **Table 12**.

Table 12. Achiral HPLC Parameters (Ion Pairing Gradient)

Instrument	Ultimate 3000 HPLC System with a PAL autosampler
Column	Agilent Zorbax RR StableBond Aq, 150 x 3.0 mm x 3.5 µm
Mobile Phases	A: 50 mM sodium phosphate, 2 mM tetrabutylammonium bisulfate, pH 7.0

Table 12. Achiral HPLC Parameters (Ion Pairing Gradient)	
	B: Acetonitrile C: Water
LC Gradient	- 0 – 4.0 min: 4.0 % B, 48.5 – 16.2 % C - 4.0 – 4.0 min: 4.0 % B, 16.2 – 48.5 % C - 4.0 – 5.5 min: 4.0 % B, 48.5 % C *Note: Depending on column age, adjust % B (acetonitrile, isocratic) up or down in order to have NQPs retain and elute within 5.5 min window.
Flow rate	1.2 mL/min
Run time	5.5 min
Column temperature	30 °C
Injection volume	5 µL
UV Detector	260 nm

Example 13

Off-target Phosphate Hydrolysis of IPP Family I Evolved Variants with Nucleotide Triphosphates

[0473] IPP variants of SEQ ID NOs: 18-26 were produced in shake flask and purified as described in Example 2. Off target hydrolytic activity of IPPs against 3'-phos-mATP and ATP was examined by incubating IPP with the nucleotide with species decomposition analyzed by HPLC (see Example 12).

[0474] Reactions were performed and processed as described in example 11. The samples were then analyzed on HPLC, as described in example 12. FIOP 1 was calculated as the percent of the ATP remaining of the variant divided by the percent of ATP remaining relative to the inorganic pyrophosphatase from *Sulfolobus sp. A20* SEQ ID NO: 10. FIOP 2 was calculated as the percent of the 3'-phosphate mATP remaining of the enzyme divided by the percent of 3'-phos-mATP remaining relative to the inorganic pyrophosphatase from *Sulfolobus sp. A20* SEQ ID NO: 10. The results are shown in **Table 13**.

Table 13. Decomposition of ATP and 3'-Phos-mATP Relative to SEQ ID NO: 10			
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 10)**	**FIOP Activity 1 Relative to SEQ ID NO: 10**	**FIOP Activity 2 Relative to SEQ ID NO: 10**
17/18	E95S		+
19/20	P65R	++++	+++
21/22	P65K	++++	+++
23/24	V147A	+	+
25/26	V41F	+++	++
Levels of increased activity were determined for FIOP activity 1 relative to SEQ ID NO: 10 and are defined as follows: "+" > 0.3, "++" > 1.0, "+++" > 1.04, "++++" > 12.0. Levels of increased activity			

Table 13. Decomposition of ATP and 3'-Phos-mATP Relative to SEQ ID NO: 10			
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 10)**	**FIOP Activity 1 Relative to SEQ ID NO: 10**	**FIOP Activity 2 Relative to SEQ ID NO: 10**
were determined for FIOP activity 2 relative to SEQ ID NO: 10 and are defined as follows: "+" > 0.3, "++" > 1.0, "+++" > 1.04, "++++" > 12.0.			

Example 14

Off-target Phosphate Hydrolysis of IPP Family II Enzymes with Nucleotide Triphosphates

[0475] IPP enzymes of SEQ ID NOs: 28-52 were produced in shake flask and purified as described in Example 2. Off target hydrolytic activity of IPPs against 3'-phos-mATP and ATP was examined by incubating IPP with the nucleotide with species decomposition analyzed by HPLC (see Example 12).

[0476] Reactions were performed and processed as described in Example 11. The samples were then analyzed on HPLC, as described in Example 12. FIOP 1 was calculated as the percent of the ATP remaining of the enzyme divided by the percent of ATP remaining relative to the inorganic pyrophosphatase from *Streptococcus thermophilus* SEQ ID NO: 32. FIOP 2 was calculated as the percent of the 3'-phosphate mATP remaining of the enzyme divided by the percent of 3'-phos-mATP remaining relative to the inorganic pyrophosphatase from *Streptococcus thermophilus* SEQ ID NO: 32. The results are shown in **Table 14**.

Table 14. Decomposition of ATP and 3'-Phos-mATP Relative to SEQ ID NO: 32		
SEQ ID NO: (nt/aa)	**FIOP Activity 1 Relative to SEQ ID NO: 32**	**FIOP Activity 2 Relative to SEQ ID NO: 32**
29/30	+++	++
33/34	++++	+++
35/36	++++	++
37/38	++	+++
39/40	++++	+++
41/42	+++	++
43/44	+++	++
45/46	++++	+++
47/48	+++	+
49/50	+++	+
51/52	+++	+
27/28	+++	+
Levels of increased activity were determined for FIOP 1 relative to SEQ ID NO: 32 for FIOP 2 relative to SEQ ID NO: 32 and defined as follows: "+" > 1.0, "++" > 1.26, "+++" > 3.55, "++++" > 4.79. "+" > 0.59, "++" > 0.95, "+++" > 1.0.		

Example 15

Off-target Phosphate Hydrolysis of IPP Family II Evolved Variants with Nucleotide Triphosphates

[0477] IPP enzymes of SEQ ID NOs: 46 and 54-74 were produced in shake flask and purified as described in Example 2. Off target hydrolytic activity of IPPs against 3'-phos-mATP and ATP was examined by incubating IPP with the nucleotide with species decomposition analyzed by HPLC (see Example 12).

[0478] Reactions were performed and processed as described in example 11. The samples were then analyzed on HPLC, as described in example 12. FIOP 1 was calculated as the percent of the ATP remaining of the enzyme divided by the percent of ATP remaining relative to the inorganic pyrophosphatase from *Methanotorris igneus* SEQ ID NO: 46. FIOP 2 was calculated as the percent of the 3'-phosphate mATP remaining of the enzyme divided by the percent of 3'-phos-mATP remaining relative to the inorganic pyrophosphatase from *Methanotorris igneus* SEQ ID NO: 46. The results are shown in **Table 15**.

Table 15. IPP Family II Evolved Variants Decomposition of ATP and 3'-Phos-mATP Relative to SEQ ID NO: 46

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 46)	FIOP Activity 1 Relative to SEQ ID NO: 46	FIOP Activity 2 Relative to SEQ ID NO: 46
53/54	D45E	+	+++
55/56	N188E	+	++
57/58	E89D	+	+
59/60	L85V	++	+++
61/62	I187L	+++	++
63/64	K177E	++	++
65/66	K81R/D86S/N188E	+	++
67/68	D86S/E89D/V160L/N188E/V308I	+	+
69/70	D169E/N188E/V231T	+	++
71/72	I110L/D169E/N188E/V231T	+	+
73/74	Y14L/N188E/K269N	+	++

Levels of increased activity were determined for FIOP activity 1 relative to SEQ ID NO: 46 and are defined as follows: "+" > 0.92, "++" > 1.0, "+++" > 1.01. Levels of increased activity were determined for FIOP activity 2 relative to SEQ ID NO: 46 and are defined as follows: "+" > 0.92, "++" > 1.0, "+++" > 1.01.

Example 16

Off-target Phosphate Hydrolysis of IPP Family I, Family II, and Commercial Enzymes with Nucleotide Triphosphates

[0479] IPP enzymes of SEQ ID NOs: 10, 12, 46, and 54-74, were produced in shake flask and purified as described in Example 2. Off target hydrolytic activity of IPPs against 3'-phos-mATP and ATP was examined by incubating IPP with the nucleotide with species decomposition analyzed by HPLC (see Example 12).

[0480] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 1 mM nucleotide triphosphate, 5 µM IPP, 100 mM triethanolamine (pH 7.8), and 1 mM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for IPP, were pre-mixed in a single solution, and was aliquoted into each well of the 96-well plates (ii) IPP solution was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at 50 °C for 16 hours then held at 4 °C until the reaction was quenched. The reactions were quenched by the addition of 75 uL of 50 % MeOH, vortexed and centrifuged for 10 minutes at 4000 rpm. The reactions were further diluted 1:1 with milli-Q water, to a nucleotide triphosphate concentration of 50 uM. The samples were then analyzed on HPLC, as described in example 12. FIOP 1 was calculated as the percent of the ATP remaining of the enzyme divided by the percent of ATP remaining relative to the inorganic pyrophosphatase from *Sulfolobus sp. A20* SEQ ID NO: 10. FIOP 2 was calculated as the percent of the 3'-phosphate mATP remaining of the enzyme divided by the percent of 3'-phos-mATP remaining relative to the inorganic pyrophosphatase from *Sulfolobus sp. A20* SEQ ID NO: 10. The results are shown in **Table 16**.

Table 16. IPP Family I, Family II and Commercial Enzymes Decomposition of ATP and 3'-Phos-mATP Relative to SEQ ID NO: 10

SEQ ID NO: (nt/aa)	FIOP Activity 1 Relative to SEQ ID NO: 10	FIOP Activity 2 Relative to SEQ ID NO: 10
11/12		+
45/46	+++	++
53/54	+++	++
55/56	+++	++
57/58	+++	++
59/60	+++	++
61/62	+++	++
63/64	+++	++
65/66	+++	++
67/68	+++	++
69/70	+++	++
71/72	+++	++
73/74	+++	++
Yeast IPP (M2403L)		
Thermal Stable IPP (M0296L)		+
E.coli IPP (M0361L)		

Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 10 and for activity 2 relative to SEQ ID NO: 10 and defined as follows: "+" > 0.9, "++" > 1.02, "+++" > 3.0. "+" > 0.9, "++" > 1.02, "+++" > 3.0.

Example 17

Activity of shake-flask purified IPP enzymes with TnT (SEQ ID NO: 92), 3'-phosphate-blocked nucleotides and modified RNA oligonucleotide substrates

[0481] IPP enzymes of SEQ ID NOs: 10, 34, and 38-46 were produced in shake flask and purified as described in Example 2.

[0482] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 2.4 mM unlabeled oligonucleotide, 100 µM FAM-labeled oligonucleotide, 3000 µM nucleotide triphosphate mUTP-3'P, 10 µM TnT (SEQ ID NO: 92) solution, 0.0625 µM IPP, 100 mM triethanolamine (pH 7.8), and 1 mM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for TnT, were pre-mixed in a single solution, and was aliquoted into each well of the 96-well plates (ii) TnT solution was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at 50 °C for 60 minutes, then held at 4 °C until the reaction was quenched. Reactions were quenched and processed for CE analysis as described in Example 18.

Example 18

Capillary Electrophoresis (CE) Analysis of Oligonucleotides

Sample preparation for reaction analysis using CE

[0483] For analysis of the reaction samples, capillary electrophoresis was performed using an ABI 3500xl Genetic Analyzer (ThermoFisher). Reactions (1 µL) were quenched by the addition of 99 µL of 1 mM aqueous EDTA. Quenched reactions were diluted in water to 37.5 nM oligonucleotide, and a 2-µL aliquot of this solution was transferred to a new 96-well MicroAmp Optical PCR plate or 384-well MicroAmp Optical PCR plate containing 18 µL Hi-Di™ Formamide (ThermoFisher) containing an appropriate size standard (LIZ or Alexa633). The ABI3500xl was configured with POP6 polymer, 50 cm capillaries, and a 55 °C oven temperature. Pre-run settings were 18KV for 50 sec. Injection was 10KV for 2 sec, and the run settings were 19KV for 850 seconds. FAM-labeled oligo substrates and products were identified by their sizes relative to the sizing ladder.

[0484] The FIOPs, which are calculated as the ratio of percent product conversion to percent byproduct conversion compared to the reference sequence value, are summarized in **Table 18** below, showing differences in enzyme activity following the reaction.

[0485] Percent product conversion was calculated as the percent product observed of the enzyme, defined as the sum of the area of products divided by the sum of the total peak area. The products formed are shown in **Table 18**.

[0486] Percent byproduct conversion was calculated as the percent byproduct of the enzyme, defined as the sum of the area of byproducts divided by the sum of the total peak area.

Table 18. Suppression of Byproduct in an IPP- TnT-Coupled Reaction

SEQ ID NO: (nt/aa)	FIOP (Ratio of % Product: % Byproduct) Relative to SEQ ID NO: 10
33/34	+
37/38	++
39/40	++
41/42	+
43/44	
45/46	++
Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 10 and defined as follows: "+" > 15.3, "++" > 18.3.	
FAM-labeled oligo: SEQ ID NO: 95 Unlabeled oligo: TTTTTTTTmA/i2FG/mU/32FG/ FAM-labeled Product (observed): SEQ ID NO: 96 Unlabeled product: TTTTTTTTmA/i2FG/mU/i2FG/mU/3'Phos/	

Example 19

Thermal stability of shake-flask purified IPP enzymes with TnT (SEQ ID NO: 92), 3'-phosphate-blocked nucleotides and modified RNA oligonucleotide substrates

[0487] IPP enzymes of SEQ ID NOs: 10, 34, and 38-46 were produced in shake flask and purified as described in Example 2.

[0488] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 1.9 mM unlabeled oligonucleotide, 100 µM FAM-labeled oligonucleotide, 3000 µM nucleotide triphosphate, 10 µM TnT (SEQ ID NO: 92) solution, 0.1 µM IPP, 100 mM triethanolamine (pH 7.8), and 1 mM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for TnT, were pre-mixed in a single solution, and was aliquoted into each well of the 96-well plates (ii) the reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at 85.5 °C for 15 minutes to thermally challenge the IPP (iii) after the incubation period, TnT (SEQ ID NO: 92) solution was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at 50 °C for 30 minutes, then held at 4 °C until the reaction was quenched. Reactions were quenched and processed for CE analysis as described in Example 18.

[0489] The FIOPs, which are calculated as the ratio of percent product conversion to percent byproduct conversion compared to the reference sequence value, are summarized in **Table 19** below, showing differences in enzyme activity following the reaction.

[0490] Percent product conversion was calculated as the percent observed product observed of the enzyme, defined as the sum of the area of products divided by the sum of the total peak area. The products formed are shown in **Table 19**.

[0491] Percent byproduct conversion was calculated as the percent byproduct of the enzyme, defined as the sum of the area of byproducts divided by the sum of the total peak area.

Table 19. Suppression of Byproduct in an IPP- TnT-Coupled Reaction -T8mC*mA*mGmA/ 5'-6-FAM-T16mC*mA*mGmA/ mATP-3'P

SEQ ID NO: (nt/aa)	FIOP (Ratio of % Product: % Byproduct)
33/34	+
37/38	++
39/40	++
41/42	+++
43/44	++
45/46	++++
Observed Product: T8mC*mA*mGmAmA-3'P, 5'-6-FAM-T16mC*mA*mGmAmA-3'P (observed)	
Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 10 and defined as follows: "+" > 1.0, "++" > 1.1, "+++" > 2.5, "++++" > 48.0.	
FAM-labeled oligo: SEQ ID NO: 97 Unlabeled oligo: TTT TTT TTmC* mA*mGmA FAM-labeled Product (observed): SEQ ID NO: 98 Unlabeled product: TTT TTT TTmC* mA*mGmAmA/3Phos/	

Example 20

Shake Flask Expression and Purification of IPP Variants Relative to SEQ ID NO: 46

Shake Flask Expression

[0492] Variants identified in a primary high throughput library screening as having improved properties relative to SEQ ID NO: 46 were selected for shake flask scale up and validation. The improved properties of interest include: an increase in product conversion, decrease in byproduct conversion, and protein solubility.

[0493] Six improved variants, as described above, SEQ ID NO: 54-64 and SEQ ID NO: 46 were expressed and purified as described in Example 2. IPP concentrations from the preparations were measured by absorption at 280 nm. Soluble protein concentrations are summarized in **Table 20** below,

showing a fold improvement in soluble protein production following shake-flask purification relative to the inorganic pyrophosphatase SEQ ID NO: 46.

Table 20. Soluble Enzyme Production Relative to SEQ ID NO: 46	
SEQ ID NO: (nt/aa)	**FIOP 1**
53/54	+
55/56	+++
57/58	++
59/60	+++
61/62	++
63/64	++
Levels of increased activity were determined for activity 1 relative to SEQ ID NO: 46 and defined as follows: "+" > 1.0, "++" > 1.15, "+++" > 1.4.	

Example 21

Activity improvement of shake-flask purified IPP enzymes with TnT (SEQ ID NO: 94), 3'-phosphate-blocked nucleotides and modified RNA oligonucleotide substrates

[0494] IPP enzymes of SEQ ID NOs: 46 and 54-58, were produced in shake flask and purified as described in Example 2.

[0495] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 2.4 mM unlabeled oligonucleotide, 100 µM FAM-labeled oligonucleotide, 3000 µM nucleotide triphosphate, 10 µM TnT (SEQ ID NO: 94) solution, 0.0039 µM IPP, 100 mM triethanolamine (pH 7.8), and 1 mM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for TnT, were pre-mixed in a single solution, and was aliquoted into each well of the 96-well plates (ii) TnT solution, SEQ ID NO: 94, was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at 50 °C for 60 minutes, then held at 4 °C until the reaction was quenched. Reactions were quenched and processed for CE analysis as described in Example 18.

[0496] The FIOPs, which are calculated as the ratio of percent product conversion to percent byproduct conversion compared to the reference sequence value, are summarized in **Table 21** below, showing differences in enzyme activity following the reaction.

[0497] Percent product conversion was calculated as the percent product observed of the enzyme, defined as the sum of the area of products divided by the sum of the total peak area. The products formed are shown in **Table 21**.

[0498] Percent byproduct conversion was calculated as the percent byproduct of the enzyme, defined as the sum of the area of byproducts divided by the sum of the total peak area.

Table 21. Suppression of Byproduct in an IPP- TnT-Coupled Reaction		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 46)**	**FIOP Activity 1 Relative to SEQ ID NO: 46**
53/54	D45E	++
55/56	N188E	+++
57/58	E89D	+++
59/60	L85V	++
Levels of increased activity were determined fIOP activity 1 relative to SEQ ID NO: 46 and are defined as follows: "+" > 1.0, "++" > 1.12, "+++" > 2.26.		
FAM-labeled oligo: SEQ ID NO: 97 Unlabeled oligo: TTT TTT TTmC* mA*mGmA FAM-labeled Product (observed): SEQ ID NO: 98 Unlabeled product: TTT TTT TTmC* mA*mGmAmA/3Phos/		

Example 22

Activity improvement of shake-flask purified IPP variants with TnT (SEQ ID NO: 94), 3'-phosphate-blocked nucleotides and modified RNA oligonucleotide substrates

[0499] IPP variants of SEQ ID NOs: 54-58 were produced in shake flask and purified as described in Example 2.

[0500] Reactions were performed in 96-well format 200 µL BioRad PCR plates. Reactions included 2.4 mM unlabeled oligonucleotide, 100 µM FAM-labeled oligonucleotide, 3000 µM nucleotide triphosphate, 10 µM TnT (SEQ ID NO: 94) solution, 0.00097 µM IPP, 100 mM triethanolamine (pH 7.8), and 1 mM cobalt (II) chloride. The reactions were set up as follows: (i) all reaction components, except for TnT, were pre-mixed in a single solution, and was aliquoted into each well of the 96-well plates (ii) TnT solution, SEQ ID NO: 94, was then added into the wells to initiate the reaction. The reaction plate was heat-sealed with a peelable aluminum seal and incubated in a thermocycler at 50 °C for 60 minutes, then held at 4 °C until the reaction was quenched. Reactions were quenched and processed for CE analysis as described in Example 18.

[0501] The FIOPs, which are calculated as the ratio of percent product conversion to percent byproduct conversion compared to the reference sequence value, are summarized in **Table 22** below, showing differences in enzyme activity following the reaction.

[0502] Percent product conversion was calculated as the percent product of the variant, defined as the sum of the area of products divided by the sum of the total peak area. The products formed are shown in **Table 22**.

[0503] Percent byproduct conversion was calculated as the percent byproduct of the variant, defined as the sum of the area of byproducts divided by the sum of the total peak area.

Table 22. Suppression of Byproduct in an IPP- TnT-Coupled Reaction		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 46)**	**FIOP activity 1 relative to SEQ ID NO: 46**
53/54	D45E	++
55/56	N188E	++
57/58	E89D	++
Levels of increased activity were determined fIOP activity 1 relative to SEQ ID NO: 46 and are defined as follows: "+" > 1.0, "++" > 1.33.		
FAM-labeled oligo: SEQ ID NO: 95 Unlabeled oligo: TTTTTTTTmA/i2FG/mU/32FG/ FAM-labeled Product (observed): SEQ ID NO: 96 Unlabeled product: TTTTTTTTmA/i2FG/mU/i2FG/mU/3Phos/		

Example 23

Shake Flask Expression and Purification of IPP Variants Relative to SEQ ID NO: 56

[0504] Variants identified in a primary high throughput library screening as having improved properties relative to SEQ ID NO: 56 were selected for shake flask scale up and validation. The improved properties of interest include: an increase in product conversion, decrease in byproduct conversion, and protein solubility.

[0505] IPP variants of SEQ ID NOs: 56 and 66-74, were produced in shake flask and purified as described in Example 1. Soluble protein concentrations are summarized in **Table 23**, showing a fold improvement in soluble protein production following shake flask purification, relative to SEQ ID NO: 56.

Table 23: Soluble Enzyme Production relative to SEQ ID NO: 56		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 56)**	**FIOP activity 1 Relative to SEQ ID NO: 56**
65/66	K81R/D86S	+++
67/68	D86S/E89D/V160L/V308I	+
69/70	D169E/V231T	++
71/72	I110L/D169E/V231T	+++
73/74	Y14L/K269N	+
Levels of increased activity were determined fIOP activity 1 relative to SEQ ID NO: 56 and are defined as follows: "+" > 0.58, "++" > 0.95, "+++" > 1.15.		

Example 24

Activity improvement of shake-flask purified IPP variants with TnT (SEQ ID NO: 94), 3'-phosphate-blocked nucleotides and modified RNA oligonucleotide substrates

[0506] Reactions were set up with variants described in Example 22 and were performed as described in Example 21.

[0507] The samples were prepared and analyzed as described in Example 18.

[0508] The FIOPs, which are calculated as the ratio of percent product conversion to percent byproduct conversion compared to the reference sequence value, are summarized in **Table 24.1** and **24.2** below, showing differences in enzyme activity following the reaction.

[0509] Percent product conversion was calculated as the percent product observed of the variant, defined as the sum of the area of products divided by the sum of the total peak area. The products formed are shown in **Table 24.1** and **Table 24.2**.

[0510] Percent byproduct conversion was calculated as the percent byproduct of the variant, defined as the sum of the area of byproducts divided by the sum of the total peak area.

Table 24.1. Suppression of Byproduct in an IPP- TnT-Coupled Reaction

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 56)	FIOP activity 1 Relative to SEQ ID NO: 56
65/66	K81R/D86S	+
67/68	D86S/E89D/V160L/V308I	++
69/70	D169E/V231T	++
71/72	I110L/D169E/V231T	+++
73/74	Y14L/K269N	++
Levels of increased activity were determined fIOP activity 1 relative to SEQ ID NO: 56 and are defined as follows: "+" > 0.93, "++" > 1.48, "+++" > 2.9.		
FAM-labeled oligo: SEQ ID NO: 97 Unlabeled oligo: TTT TTT TTmC* mA*mGmA FAM-labeled Product (observed): SEQ ID NO: 98 Unlabeled product: TTT TTT TTmC* mA*mGmAmA/3Phos/		

Table 24.2. Suppression of Byproduct in an IPP- TnT-Coupled Reaction

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 56)	FIOP activity 1 Relative to SEQ ID NO: 56
65/66	K81R/D86S	+
67/68	D86S/E89D/V160L/V308I	+
69/70	D169E/V231T	++
71/72	I110L/D169E/V231T	++
73/74	Y14L/K269N	+
Levels of increased activity were determined fIOP activity 1 relative to SEQ ID NO: 56 and are defined as follows: "+" > 1.17, "++" > 2.37.		

Table 24.2. Suppression of Byproduct in an IPP- TnT-Coupled Reaction
FAM-labeled oligo: SEQ ID NO: 95
Unlabeled oligo: TTTTTTTTmA/i2FG/mU/32FG/
FAM-labeled Product (observed): SEQ ID NO: 96
Unlabeled product: TTTTTTTTmA/i2FG/mU/i2FG/mU/3Phos/

Example 25

Generation of N and C-Terminal Tags Containing Lysine Residues on SEQ ID NO: 46 and 72

[0511] Eight IPP enzymes were modified by the introduction of 1, 3 or 5 lysine residues at the N or C-terminal tags. These modifications were constructed on wild type (WT) SEQ ID NO: 46 and evolved SEQ ID NO: 72, through well-established cloning techniques (PCR, restriction digestion).

Example 26

Shake Flask Expression and Purification of IPP Constructs Modified with Lysine Residues

Shake Flask Expression

[0512] Variants described above of SEQ ID NOs: 46, 72, and 76-90 were produced in shake flask and purified as described in Example 2.

[0513] Elution fractions containing protein were identified by UV absorption (A280) and pooled, then dialyzed overnight in dialysis buffer (20 mM Tris-HCl, pH 7.4, 100 mM KCl, 0.1 mM EDTA, and 50% glycerol) in a 3.5K Slide-A-Lyzer™ dialysis cassette (Thermo Fisher) for buffer exchange. IPP concentrations in the preparations were measured by absorption at 280 nm. Soluble protein concentrations are summarized in **Table 26.1** and **Table 26.2** below, showing a fold improvement in soluble protein production following shake-flask purification relative to SEQ ID NO: 46 and SEQ ID NO: 72, respectively.

Table 26.1. Soluble Enzyme Production Relative to SEQ ID NO: 46		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 46)**	**FIOP 1**
75/76	-1M/-2K/-3K/M4K	+++
77/78	-1M/-2K/-3K/-4K/-5K/M6K	++
79/80	*319G/*320G/-321S/-322G/-323K/-324*/-325*	+
81/82	*319G/*320G/-321S/-322G/-323K/-324K/-325K/-326*/-327*	+
Levels of increased activity were determined fIOP activity 1 relative to SEQ ID NO: 46 and are defined as follows: "+" > 0.5, "++" > 0.93, "+++" > 1.8.		

Table 26.2. Soluble Enzyme Production Relative to SEQ ID NO: 72		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 72)**	**FIOP 1**
83/84	-1M/-2K/-3K/M4K	+++
85/86	-1M/-2K/-3K/-4K/-5K/M6K	+
87/88	*319G/*320G/-321S/-322G/-323K/-324*/-325*	+++
89/90	*319G/*320G/-321S/-322G/-323K/-324K/-325K/-326*/-327*	+++
Levels of increased activity were determined fIOP activity 1 relative to SEQ ID NO: 72 and are defined as follows: "+" > 0.67, "++" > 1.0, "+++" > 1.26.		

Example 27

Immobilization of Lysine Tagged IPP Constructs on Epoxy and Affinity Resin

[0514] Purified IPP enzyme constructs from example 26 were immobilized on epoxy and affinity resins, as described below.

[0515] Epoxy resin (IB-COV-7) and affinity resin (IB-HIS-2, pre-loaded with $CoCl_2$) was purchased from ChiralVision and weighed out into 10 wells of a 1.0 mL deep 96 well plate. The resin was then charged with IPP at a target of 0.7 weight percent IPP vs wet weight of resin, according to the immobilization procedure summarized in Example 28 (epoxy resin) and Example 29 (affinity resin), in 50 mM TEoAHCl (pH 7.8) for 24 h at 25 °C with 500 rpm agitation.

[0516] After immobilization, the A280 of the resulting supernatant solution was compared to that of a solution of each enzyme that had been incubated under the same conditions but without resin present. The percent immobilization was then calculated as: $\left(1 - \frac{A280}{A280_{neg.\ ctrl.}}\right) \times 100$. The results are shown in **Table 27**.

Table 27. Immobilization of IPP on Affinity and Epoxy Resin		
Resin	**SEQ ID NO: (nt/aa)**	**Percent of IPP Immobilized**
IB-HIS-2	71/72	+
IB-HIS-2	75/76	++
IB-HIS-2	77/78	++
IB-HIS-2	79/80	+
IB-HIS-2	81/82	+
IB-HIS-2	83/84	++
IB-HIS-2	85/86	++
IB-HIS-2	87/88	+
IB-HIS-2	89/90	+
IB-COV-7	71/72	+
IB-COV-7	75/76	+
IB-COV-7	77/78	+
IB-COV-7	79/80	+
IB-COV-7	81/82	+

Table 27. Immobilization of IPP on Affinity and Epoxy Resin		
Resin	**SEQ ID NO: (nt/aa)**	**Percent of IPP Immobilized**
IB-COV-7	83/84	+
IB-COV-7	85/86	+
IB-COV-7	87/88	++
IB-COV-7	89/90	++
Levels of immobilization are determined by ratio of A280 vs negative control and defined as follows: "+" 0 to 50%, "++" > 50%.		

[0517] A reaction stock was prepared, consisting of pyrophosphate (1 mM), $CoCl_2$ (0.5 mM), TEoAHCl (50 mM), pH 7.8. The reaction stock was added to resin containing wells at a volume of 10 µL per mg of resin and incubated for 10 minutes with 500 rpm agitation at a temperature 40 °C. A 10 µL aliquot was removed from each well, diluted to a theoretical max of 40 µM phosphate in RO water, and analyzed for phosphate concentration using a malachite green phosphate assay kit purchased from Sigma Aldrich (MAK307). The results are shown in **Table 28**.

Table 28. Activity of IPP Immobilized on Affinity and Epoxy Resin		
Resin	**IPP Construct**	**Phosphate Produced (mM)**
IB-HIS-2	71/72	++
IB-HIS-2	75/76	+
IB-HIS-2	77/78	++
IB-HIS-2	79/80	++
IB-HIS-2	81/82	++
IB-HIS-2	83/84	++
IB-HIS-2	85/86	++
IB-HIS-2	87/88	++
IB-HIS-2	89/90	++
IB-COV-7	71/72	++
IB-COV-7	75/76	++
IB-COV-7	77/78	+
IB-COV-7	79/80	+
IB-COV-7	81/82	+
IB-COV-7	83/84	+
IB-COV-7	85/86	+
IB-COV-7	87/88	+
IB-COV-7	89/90	+
Levels of phosphate production are determined by malachite green assay and are defined as follows: "+" 0 to 30%, "++" > 30%.		

Example 28

General Enzyme Immobilization Procedure on Epoxide Functionalized Covalent Resins

[0518] The desired protein variant was first produced in shake flask and purified as described in Example 1. The storage buffer from an aliquot of the desired protein stock was then exchanged for

10-500 mM TEAHCl or MOPS buffer, pH 7-9, by diluting10x with TEAHCl followed by concentration through a Sartirous VivaSpin 6 (30,000 MWCO) spin filter at 4000 rpm and 4 °C. The resulting small volume is then diluted another 2.5 x with 10-500 mM TEAHCl or MOPS buffer, pH 7-9, to give the original volume.

[0519] Between 2 and 50 mg epoxide functionalized resin was weighed out, either wet or dry, into a 96 well plate with a volume of 2 mL per well or into a 2 mL Eppendorf tube. For larger preparations, 0.5 to 1.0 g resin was weighed out into a 15 mL conical tube. The volume of protein stock containing the desired wt. % of protein vs resin mass was transferred to the 96 well plate or tube containing resin.

[0520] The samples were then incubated at 4-50 °C for 12 to 48 h using either 500 rpm agitation or on a tube rotator operating at 10 rpm. For smaller preparations utilizing epoxide functionalized covalent resins in 96 well plates or 2.0 mL Eppendorf tubes, the resin was washed three times with 300 µL wash volumes of TEoAHCl (50 mM, pH 7.8) containing 500 mM NaCl, allowing the mixture to agitate at room temperature at 500 rpm for 5 minutes during each wash. Between each wash the tubes were briefly centrifuged to collect resin at the bottom. The resins were then washed with three times with 300 µL wash volumes of TEoAHCl (50 mM, pH 7.8), allowing the mixture to agitate at room temperature at 500 rpm for 5 minutes during each wash. Between each wash the tubes were briefly centrifuged to collect resin at the bottom. In the case of a larger preparation in a 15 mL conical tube, the resin was collected by vacuum filtration and then washed three times with 10 mL of TEAHCl (50 mM, pH 7.8) containing 500 mM NaCl, followed by an additional three wash volumes of 10 mL TEAHCl (50 mM, pH 7.8), allowing the resin to mix well on a rotator for 10 minutes at room temperature during each wash.

Example 29

General Enzyme Immobilization Procedure on Affinity, Adsorbent, and Ion-Exchange Resins

[0521] The desired iPPase protein variant was first produced in shake flask and purified as described in Example 1. The storage buffer from an aliquot of the desired protein stock was then exchanged for TEAHCl (500 mM, pH 7.8) by diluting10x with TEAHCl followed by concentration through a Sartirous VivaSpin 6 (30,000 MWCO) spin filter at 4000 rpm and 4 °C. The resulting small volume is then diluted another 2.5 x with TEAHCl (500 mM, pH 7.8) to give the original volume.

[0522] Between 2 and 50 mg affinity, adsorbent, or ion-exchange resin was weighed out, either wet or dry, into a 96 well plate with a volume of 2 mL per well or into a 2 mL Eppendorf tube. For larger preparations, 0.5 to 1.0 g adsorbent or ion-exchange resin was weighed out into a 15 mL conical tube. The volume of protein stock containing the desired wt. % of protein vs resin mass was transferred to the 96 well plate or tube containing resin.

[0523] The samples were then incubated at 4-25 °C for 24 to 48 h using either 500 rpm agitation or on a tube rotator operating at 10 rpm. Before subjecting to activity assays, the supernatant was removed, and the resin was washed three times. For smaller preparations utilizing adsorbent or ion-

exchange resins in 96 well plates or 2.0 mL Eppendorf tubes, 300 µL wash volumes of TEAHCl (50 mM, pH 7.8) were used, allowing the mixture to agitate at room temperature at 500 rpm for 5 minutes during each wash. Between each wash the tubes were briefly centrifuged to collect resin at the bottom. In the case of a larger preparation in a 15 mL conical tube, the resin was collected by vacuum filtration and then washed three times with 10 mL of TEAHCl (50 mM, pH 7.8), allowing the resin to mix well on a rotator for 10 minutes at room temperature during each wash.

[0524] While the invention has been described with reference to the specific embodiments, various changes can be made and equivalents can be substituted to adapt to a particular situation, material, composition of matter, process, process step or steps, thereby achieving benefits of the invention without departing from the scope of what is claimed.

[0525] For all purposes, each and every publication and patent document cited in this disclosure is incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

CLAIMS

What is claimed is:

1. A method of increasing forward reaction in a reaction generating inorganic pyrophosphate, comprising providing an inorganic pyrophosphatase for converting inorganic pyrophosphate to phosphate in a reaction generating inorganic pyrophosphate as a product, wherein the inorganic pyrophosphatase comprises a type II inorganic pyrophosphatase, a type I inorganic pyrophosphatase, or a variant of type II or type I inorganic pyrophosphatase.

2. The method of Claim 1, wherein increasing the forward reaction increases product yield in the reaction.

3. The method of Claim 1 or 2, wherein the inorganic pyrophosphatase comprises a type II inorganic pyrophosphatase.

4. The method of any one of Claims 1-3, wherein the type II pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or to a reference sequence corresponding to SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

5. The method of Claim 4, wherein the inorganic pyrophosphatase comprises an amino acid sequence comprising residues 12 to carboxy terminal of SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, or an amino acid sequence comprising SEQ ID NO: 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52.

6. The method of Claim 1 or 2, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

7. The method of Claim 1 or 2, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74,

wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

8. The method of Claim 1 or 2, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

9. The method of any one of Claims 6-8, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 14, 45, 81, 85, 86, 89, 110, 160, 169, 177, 187, 188, 231, 269, or 308, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

10. The method of any one of Claims 6-8, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference 14L, 45E, 81R, 85V, 86S, 89D, 110L, 160L, 169E, 177E, 187L, 188E, 231T, 269N, or 308I, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

11. The method of Claim 6, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position(s) 14/188/269, 45, 81/86/188, 85, 86/89/160/188/308, 89, 110/169/188/231, 169/188/231, 177, 187, or 188, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

12. The method of Claim 1 or 11, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) 14L/188E/269N, 45E, 81R/86S/188E, 85V, 86S/89D/160L/188E/308I, 89D, 110L/169E/188E/231T, 169E/188E/231T, 177E, 187L, and 188E, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

13. The method of Claim 6, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position(s) 14/269, 81/86, 86/89/160/308, 110/169/231, or 169/231, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56, or to the reference sequence corresponding to SEQ ID NO: 56.

14. The method of Claim 6 or 13, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) 14L/269N, 81R/86S, 86S/89D/160L/308I, 110L/169E/231T, or 169E/231T, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56, or to the reference sequence corresponding to SEQ ID NO: 56.

15. The method Claim 6, wherein the amino acid sequence of the inorganic pyrophosphatase comprises residues 12 to carboxy terminal of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or comprises SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

16. The method of any one of Claims 6-15, wherein the inorganic pyrophosphatase has reduced off target hydrolysis as compared to a reference inorganic pyrophosphatase.

17. The method of Claim 16, wherein the inorganic pyrophosphatase has reduced off target hydrolysis of a natural or modified nucleoside triphosphate as compared to a reference pyrophosphatase.

18. The method of Claim 17, wherein the inorganic pyrophosphatase has at least 3-fold lower off target hydrolysis of a natural nucleoside triphosphate as compared to a reference inorganic pyrophosphatase.

19. The method of Claim 18, wherein the nucleoside triphosphate is ATP.

20. The method of any one of Claims 16-19, wherein the reference inorganic pyrophosphatase has a sequence corresponding to SEQ ID NO: 10.

21. The method of any one of Claims 6-15, wherein the wherein the inorganic pyrophosphatase has at least a 49-fold lower off target hydrolysis of a non-natural or modified nucleoside triphosphate as compared to a reference inorganic pyrophosphatase.

22. The method of Claim 21, wherein the non-natural or modified nucleoside triphosphate is 3' phosphate-2'-O-methyl-ATP.

23. The method of Claim 21 or 22, wherein the reference inorganic pyrophosphatase is a yeast inorganic pyrophosphatase of Saccharomyces cerevisiae ppa.

24. The method of any of Claims 6-15, wherein the inorganic pyrophosphatase exhibits increased pyrophosphatase activity as compared to a reference pyrophosphatase.

25. The method of Claim 24, wherein the inorganic pyrophosphatase comprises at least a 1.17-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase.

26. The method of Claim 24, wherein the inorganic pyrophosphatase comprises at least a 22-fold, 23-fold, 24-fold, 25-fold, or 26-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase.

27. The method of any one of Claims 24-26, wherein the reference inorganic pyrophosphatase has a sequence corresponding to SEQ ID NO: 10.

28. The method of Claim 1 or 2, wherein the inorganic pyrophosphatase comprises a type I inorganic pyrophosphatase, or a variant of a type I inorganic pyrophosphatase.

29. The method of Claim 28, wherein the type I pyrophosphatase comprises an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16, without the carboxy terminal histidine tag, or to a reference sequence corresponding to SEQ ID NO: 2, 4, 6, 8, 10, 12, or 16.

30. The method of Claim 29, wherein the inorganic pyrophosphatase comprises a type I pyrophosphatase comprising the sequence comprising SEQ ID NO: 4, 6, 8, 10, 12, or 16, or comprising the sequence comprising SEQ ID NO: 4, 6, 8, 10, 12, or 16 without the carboxy-terminal histidine tag.

31. The method of Claim 28, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

32. The method of Claim 28, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10, without the carboxy terminal histidine tag residues.

33. The method of Claim 28, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues.

34. The method of any one of Claims 31 or 32, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 41, 65, 95, or 147, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

35. The method of any one of Claims 31, 32, and 34, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference 41F, 65K/R, 95S, or 147A, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

36. The method of any one of Claims 31-35, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, or comprises SEQ ID NO: 18, 20, 22, 24, or 26.

37. The method of any of Claims 31-36, wherein the inorganic pyrophosphatase comprises increased pyrophosphatase activity as compared to a reference pyrophosphatase.

38. The method of Claim 37, wherein the inorganic pyrophosphatase comprises at least a 2-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase.

39. The method of any of Claims 4-38, wherein the inorganic pyrophosphatase has increased solubility as compared to a reference inorganic pyrophosphatase.

40. The method of any of Claims 4-39, wherein the inorganic pyrophosphatase has increased thermal stability as compared to a reference inorganic pyrophosphatase.

41. The method of any one of Claims 1-40, wherein the reaction is carried out with a second enzyme that produces inorganic pyrophosphate.

42. The method of Claim 41, wherein the second enzyme comprises a heterologous enzyme.

43. The method of Claim 41 or 42, wherein the second enzyme comprises a template independent nucleotidyl transferase, a polynucleotide ligase, a polynucleotide polymerase, or a tRNA synthetase.

44. The method of Claim 43, wherein the second enzyme comprises a nucleotidyl transferase, poly(U)polymerase or poly(A)polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase.

45. The method of Claim 43, wherein the nucleotidyl transferase comprises a terminal nucleotidyl transferase.

46. The method of Claim 41 or 42, wherein the second enzyme comprises a terminal nucleotidyl transferase, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase from a species selected from: *Gallus gallus, Xenopus laevis, Oncorhynchus mykiss, Monodelphis domestica, Mus musculus, Ambystoma mexicanum, Takifugu rubripes, Raja eglanteria, Ginglymostoma cirratum, Danio rerio, Canis lupus familiaris, Lemur catta, Microcebus murinus, Rattus norvegicus, Equus caballus, Xenopus (Silurana) tropicalis, Oryctolagus cuniculus, Ailuropoda melanoleuca, Sus scrofa, Anolis carolinensis, Loxodonta Africana, Ornithorhynchus anatinus, Cavia porcellus, Heterocephalus glaber, Macaca mulatto, Sarcophilus harrisii, Sarcophilus harrisii, Otolemur garnettii, Pan paniscus, Saimiri boliviensis, Felis catus, Tupaia chinensis, Pleurodeles waltl, Orcinus orca, Trichechus manatus latirostris, Dasypus novemcinctus, Maylandia zebra, Ochotona princeps, Sorex araneus, Octodon degus, Echinops telfairi, Condylura cristata, Mustela putorius furo, Heterocephalus glaber, Mesocricetus auratus, Melopsittacus undulatus, Falco peregrinus, Chrysemys picta bellii , Microtus ochrogaster, Ictidomys tridecemlineatus, Chinchilla lanigera, Cricetulus griseus, Geospiza fortis, Pseudopodoces humilis, Columba livia, Macaca fascicularis, Pundamilia nyererei, Xiphophorus maculatus, Myotis brandtii, Pantholops hodgsonii, Latimeria chalumnae, Alligator sinensis, Pelodiscus sinensis, Myotis lucifugus, Camelus ferus, Alligator mississippiensis, Lepisosteus oculatus,* and *Homo sapiens*.

47. The method of any one of Claims 44-46, wherein the reaction with the nucleotidyl transferase comprises a natural or unnatural or modified nucleoside triphosphate substrate.

48. The method of Claim 47, wherein the non-natural or modified nucleoside triphosphate is 3' phosphate-mATP.

49. The method of Claim 43, wherein the polynucleotide ligase comprises a single stranded DNA ligase, a double stranded DNA ligase, a single stranded RNA ligase, or a double stranded RNA ligase.

50. The method of Claim 49, wherein the reaction with the polynucleotide ligase comprises a nucleotide co-factor comprises a natural nucleotide triphosphate or a modified nucleotide triphosphate.

51. The method of Claim 50, wherein the nucleotide co-factor is ATP.

52. The method of Claim 50 or 51, wherein the polynucleotide ligase comprises a singled stranded RNA ligase, wherein the substrate for the single stranded RNA ligase comprises a polynucleotide acceptor comprising at least a modified nucleoside and/or internucleoside linkage; and a polynucleotide donor comprising at least a modified nucleoside and/or internucleoside linkage, or a nucleotide donor comprising a modified nucleoside.

53. The method of Claim 50 or 51, wherein the polynucleotide ligase comprises a double stranded RNA ligase, wherein the substrate for the double stranded RNA ligase comprises a polynucleotide acceptor comprising at least a modified nucleoside and/or internucleoside linkage; a polynucleotide donor comprising at least a modified nucleoside and/or internucleoside linkage, and a polynucleotide strand comprising a region complementary to the polynucleotide acceptor and polynucleotide donor, wherein the complementary polynucleotide strand comprises at least a modified nucleoside and/or internucleoside linkage.

54. The method of Claim 42, wherein the polynucleotide polymerase comprises a template-dependent DNA polymerase.

55. The method of Claim 53, wherein the substrate for the template-dependent DNA polymerase comprises a natural or unnatural or modified deoxynucleoside triphosphate.

56. The method of Claim 42, wherein the polynucleotide polymerase comprises an RNA polymerase.

57. The method of Claim 56, wherein the substrate for the RNA polymerase comprises a natural or unnatural or modified nucleoside triphosphate.

58. The method of any one of Claims 1-57, wherein the inorganic pyrophosphatase is immobilized on a support medium.

59. The method of any one of Claims 1-57, wherein method further comprises a reaction vessel containing the reaction.

60. The method of any of Claims 1-59, wherein the method further comprises at least one non-natural or industrial process condition.

61. The method of Claim 60, wherein the non-natural or industrial process condition comprises one or more of the following: a buffer, high salt concentration as compared to a naturally occurring concentration, high substrate concentration as compared to a naturally occurring concentration, one or more non-natural or non-preferred cofactors (such as a divalent metal cofactor),

an isolated or purified inorganic pyrophosphatase, high inorganic pyrophosphatase concentration as compared to a naturally occurring concentration, high temperature as compared to a naturally occurring temperature, an organic solvent, high pressure, or a high or low pH as compared to a naturally occurring pH.

62. An engineered inorganic pyrophosphatase comprising a polypeptide sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to residues SEQ ID NO: 10, 46, and 56, and one or more amino acid residue differences relative to the reference sequence.

63. The engineered inorganic pyrophosphatase of Claim 62, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

64. The engineered inorganic pyrophosphatase of Claim 62, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 46 or 56, or to a reference sequence corresponding to SEQ ID NO: 46 or 56.

65. The engineered inorganic pyrophosphatase of Claim 62, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to amino acid residues 12 to carboxy terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to amino acid residues 12 to carboxy

terminus of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or to a reference sequence corresponding to SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

66. The engineered inorganic pyrophosphatase of any one of Claims 62-65, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 14, 45, 81, 85, 86, 89, 110, 160, 169, 177, 187, 188, 231, 269, or 308, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

67. The engineered inorganic pyrophosphatase of any one of Claims 62-65 and 66, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference 14L, 45E, 81R, 85V, 86S, 89D, 110L, 160L, 169E, 177E, 187L, 188E, 231T, 269N, or 308I, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46 or 56, or to the reference sequence corresponding to SEQ ID NO: 46 or 56.

68. The engineered inorganic pyrophosphatase of any one of Claims 62-65, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position(s) 14/188/269, 45, 81/86/188, 85, 86/89/160/188/308, 89, 110/169/188/231, 169/188/231, 177, 187, or 188, or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46, or to the reference sequence corresponding to SEQ ID NO: 46.

69. The engineered inorganic pyrophosphatase of any one of Claims 62-65 and 68, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s) 14L/188E/269N, 45E, 81R/86S/188E, 85V, 86S/89D/160L/188E/308I, 89D, 110L/169E/188E/231T, 169E/188E/231T, 177E, 187L, and 188E, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 46., or to the reference sequence corresponding to SEQ ID NO: 46.

70. The engineered inorganic pyrophosphatase of any one of Claims 62-65, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position(s) 14/269, 81/86, 86/89/160/308, 110/169/231, or 169/231, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

71. The engineered inorganic pyrophosphatase of any one of Claims 62-65 and 70, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference(s)

14L/269N, 81R/86S, 86S/89D/160L/308I, 110L/169E/231T, or 169E/231T, and/or any combinations thereof, as compared to the reference sequence corresponding to residues 12 to carboxy terminal of SEQ ID NO: 56 or to the reference sequence corresponding to SEQ ID NO: 56.

72. The engineered inorganic pyrophosphatase of any one of Claims 62-65, wherein the amino acid sequence of the inorganic pyrophosphatase comprises residues 12 to carboxy terminal of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74, or comprises SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, or 74.

73. The engineered inorganic pyrophosphatase of any one of Claims 62-72, wherein the inorganic pyrophosphatase has reduced off target hydrolysis as compared to a reference inorganic pyrophosphatase.

74. The engineered inorganic pyrophosphatase of Claim 73, wherein the inorganic pyrophosphatase has reduced off target hydrolysis of a natural or modified nucleoside triphosphate as compared to a reference inorganic pyrophosphatase.

75. The engineered inorganic pyrophosphatase of Claim 74, wherein the inorganic pyrophosphatase has at least 3-fold lower off target hydrolysis of a natural nucleoside triphosphate as compared to a reference inorganic pyrophosphatase.

76. The engineered inorganic pyrophosphatase of Claim 74 or 75, wherein the nucleoside triphosphate is ATP.

77. The engineered inorganic pyrophosphatase of any one of Claims 73-76, wherein the reference inorganic pyrophosphatase has a sequence corresponding to SEQ ID NO: 10.

78. The engineered inorganic pyrophosphatase of any one of Claims 62-73, wherein the inorganic pyrophosphatase has at least a 49-fold lower off target hydrolysis of a non-natural or modified nucleoside triphosphate as compared to a reference inorganic pyrophosphatase.

79. The engineered inorganic pyrophosphatase of Claim 78, wherein the non-natural or modified nucleoside triphosphate is 3' phosphate-2'-O-methyl-ATP.

80. The engineered inorganic pyrophosphatase of Claim 78 or79, wherein the reference inorganic pyrophosphatase is a yeast inorganic pyrophosphatase of Saccharomyces cerevisiae ppa.

81. The engineered inorganic pyrophosphatase of any one of Claims 62-73, wherein the inorganic pyrophosphatase exhibits increased pyrophosphatase activity as compared to a reference pyrophosphatase.

82. The engineered inorganic pyrophosphatase of Claim 81, wherein the inorganic pyrophosphatase comprises at least a 1.17-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase.

83. The engineered inorganic pyrophosphatase of Claim 81, wherein the inorganic pyrophosphatase comprises at least a 22-fold, 23-fold, 24-fold, 25-fold, or 26-fold increased pyrophosphatase activity as compared to a reference pyrophosphatase.

84. The engineered inorganic pyrophosphatase of any one of Claims 81-83, wherein the reference inorganic pyrophosphatase has a sequence corresponding to SEQ ID NO: 10 or 46.

85. The engineered inorganic pyrophosphatase of Claim 62, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

86. The engineered inorganic pyrophosphatase of Claim 62, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26 without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10, without the carboxy terminal histidine tag residues.

87. The engineered inorganic pyrophosphatase of Claim 62, wherein the inorganic pyrophosphatase comprises an engineered inorganic pyrophosphatase comprising an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, wherein the amino acid sequence comprises one or more amino acid residue differences relative to the reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, or to a reference sequence corresponding to SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues.

88. The engineered inorganic pyrophosphatase of any one of Claims 85 or 86, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference at amino acid position 41, 65, 95, or 147, or any combinations thereof, relative

to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

89. The engineered inorganic pyrophosphatase of any one of Claims 85, 86, and 88, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises at least an amino acid residue difference 41F, 65K, 65R, 95S, or 147A, or any combinations thereof, relative to the reference sequence corresponding to SEQ ID NO: 10, or to a reference sequence corresponding to SEQ ID NO: 10 without the carboxy terminal histidine tag residues.

90. The engineered inorganic pyrophosphatase of any one of Claims 85-89, wherein the amino acid sequence of the engineered inorganic pyrophosphatase comprises SEQ ID NO: 18, 20, 22, 24, or 26, without the carboxy terminal histidine tag residues, or comprises SEQ ID NO: 18, 20, 22, 24, or 26.

91. The engineered inorganic pyrophosphatase of any of Claims 85-90, comprising at least one improved property, as compared to a wild-type or reference inorganic pyrophosphatase.

92. The engineered inorganic pyrophosphatase of Claim 91, wherein said improved property comprises improved solubility, improved thermostability, improved pyrophosphatase activity, decreased off target hydrolysis, decreased pyrophosphorolysis by a second enzyme, decreased byproducts in a reaction with a second enzyme, and/or increased ratio of products to byproducts in a reaction with a second enzyme.

93. The engineered inorganic pyrophosphatase of Claim 91 or 92, wherein said inorganic pyrophosphatase comprises at least 1.17-fold, 2-fold, or 22-fold improved activity as compared to a reference inorganic pyrophosphatase.

94. The engineered inorganic pyrophosphatase of Claim 91 or 92, wherein said inorganic pyrophosphatase comprises at least 2-fold, 3-fold, or 49-fold decreased off target hydrolysis as compared to a reference inorganic pyrophosphatase.

95. The engineered inorganic pyrophosphatase of any one of Claims 62-94, wherein the engineered inorganic pyrophosphatase is a purified preparation.

96. A recombinant polynucleotide comprising a polynucleotide sequence encoding at least one engineered inorganic pyrophosphatase of any one of Claims 62-94.

97. The recombinant polynucleotide of Claim 96, comprising a polynucleotide sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference polynucleotide sequence corresponding to SEQ ID NO: 9, 45, or 55, wherein the recombinant polynucleotide encodes an inorganic pyrophosphatase.

98. The recombinant polynucleotide of Claim 96, comprising a polynucleotide sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference polynucleotide sequence corresponding to SEQ ID NO: 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, or 73, wherein the recombinant polynucleotide encodes an inorganic pyrophosphatase.

99. The recombinant polynucleotide of Claim 96, comprising a polynucleotide sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a reference polynucleotide sequence corresponding to SEQ ID NO: 17, 19, 21, 23, or 25, wherein the recombinant polynucleotide encodes an inorganic pyrophosphatase.

100. The polynucleotide of any of Claims 96-99, wherein said polynucleotide sequence is codon-optimized.

101. The polynucleotide of any of Claims 96-100, wherein said polynucleotide sequence is operably linked to a control sequence.

102. An expression vector comprising at least one polynucleotide of any one of Claims 96-101.

103. A host cell comprising at least one expression vector of Claim 102.

104. A method of producing an engineered inorganic pyrophosphatase polypeptide in a host cell comprising culturing a host cell of Claim 103, under suitable culture conditions, such that at least one engineered inorganic pyrophosphatase is produced.

105. The method of Claim 104, further comprising recovering at least one inorganic pyrophosphatase from the culture and/or host cells.

106. The method of Claim 104 or 105, further comprising purifying said at least one engineered inorganic pyrophosphatase.

107. A composition comprising an inorganic pyrophosphatase and (i) a natural or modified nucleoside triphosphate or natural or modified nucleoside triphosphate analog or (ii) a substrate for a second enzyme.

108. A composition comprising an inorganic pyrophosphatase and a second enzyme that produces a substrate for the inorganic pyrophosphatase.

109. The composition of Claim 108, wherein the composition further comprises (i) a natural or modified nucleoside triphosphate or natural or modified nucleoside triphosphate analog or (ii) a substrate for a second enzyme.

110. The composition of Claim 109, wherein the natural or modified nucleoside triphosphate or natural or modified nucleoside triphosphate analog comprises a modified nucleoside

triphosphate or nucleoside triphosphate analog with a phosphate at the 3'-position of the sugar moiety and, optionally, one or more additional modifications.

111. The composition of any of Claims 107-110, wherein the inorganic pyrophosphatase comprises a type II inorganic pyrophosphatase or a variant of type I or type II inorganic pyrophosphatase.

112. The composition of any one of Claims 107-111, wherein the inorganic pyrophosphatase has reduced off target hydrolysis of a natural nucleoside triphosphate as compared to a reference inorganic pyrophosphatase.

113. The composition of Claim 112, wherein the nucleoside triphosphate is ATP.

114. The composition of any one of Claims 107-113, wherein the inorganic pyrophosphatase comprises the inorganic pyrophosphatase of any one of Claims 62-95.

115. The composition of any one of Claims 108-114, wherein the second enzyme comprises a nucleotidyl transferase, a polynucleotide ligase, a polynucleotide polymerase, or a tRNA synthetase.

116. The composition of Claim 115, wherein the second enzyme comprises a nucleotidyl transferase, poly(U)polymerase or poly(A)polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, and a Polθ polymerase.

117. The composition of Claim 116, wherein the nucleotidyl transferase comprises a terminal nucleotidyl transferase.

118. The composition of Claim 117, comprising the terminal nucleotidyl transferase and a natural or unnatural or modified nucleoside triphosphate substrate.

119. The composition of Claim 118, wherein the non-natural or modified nucleoside triphosphate is 3' phosphate-mATP.

120. The composition of Claim 115, wherein the second enzyme comprises a terminal nucleotidyl transferase, a Pol X polymerase, a Poly(N) polymerase, a Polμ polymerase, a Polβ polymerase, a Polλ polymerase, or a Polθ polymerase from a species selected from: *Gallus gallus, Xenopus laevis, Oncorhynchus mykiss, Monodelphis domestica, Mus musculus, Ambystoma mexicanum, Takifugu rubripes, Raja eglanteria, Ginglymostoma cirratum, Danio rerio, Canis lupus familiaris, Lemur catta, Microcebus murinus, Rattus norvegicus, Equus caballus, Xenopus (Silurana) tropicalis, Oryctolagus cuniculus, Ailuropoda melanoleuca, Sus scrofa, Anolis carolinensis, Loxodonta Africana, Ornithorhynchus anatinus, Cavia porcellus, Heterocephalus glaber, Macaca mulatto, Sarcophilus harrisii, Sarcophilus harrisii, Otolemur garnettii, Pan paniscus, Saimiri boliviensis, Felis catus, Tupaia chinensis, Pleurodeles waltl, Orcinus orca, Trichechus manatus latirostris, Dasypus novemcinctus, Maylandia zebra,*

Ochotona princeps, *Sorex araneus*, *Octodon degus*, *Echinops telfairi*, *Condylura cristata*, *Mustela putorius furo*, *Heterocephalus glaber*, *Mesocricetus auratus*, *Melopsittacus undulatus*, *Falco peregrinus*, *Chrysemys picta bellii* , *Microtus ochrogaster*, *Ictidomys tridecemlineatus*, *Chinchilla lanigera*, *Cricetulus griseus*, *Geospiza fortis*, *Pseudopodoces humilis*, *Columba livia*, *Macaca fascicularis*, *Pundamilia nyererei*, *Xiphophorus maculatus*, *Myotis brandtii*, *Pantholops hodgsonii*, *Latimeria chalumnae*, *Alligator sinensis*, *Pelodiscus sinensis*, *Myotis lucifugus*, *Camelus ferus*, *Alligator mississippiensis*, *Lepisosteus oculatus*, and *Homo sapiens*.

121. The composition of Claim 115, wherein the polynucleotide ligase comprises a single stranded DNA ligase, a double stranded DNA ligase, a single stranded RNA ligase, or a double stranded RNA ligase.

122. The composition of Claim 121, comprising the polynucleotide ligase and a nucleotide co-factor, wherein the nucleotide co-factor comprises a natural nucleotide triphosphate or a modified nucleotide triphosphate.

123. The composition of Claim 121 or 122, wherein the polynucleotide ligase comprises a singled stranded RNA ligase, and the composition further comprises a substrate for the single stranded RNA ligase, wherein the substrate comprises a polynucleotide acceptor comprising at least a modified nucleoside and/or internucleoside linkage; and/or a polynucleotide donor comprising at least a modified nucleoside and/or internucleoside linkage, and/or a nucleotide donor comprising a modified nucleoside.

124. The composition of Claim 121 or 122, wherein the polynucleotide ligase comprises a double stranded RNA ligase, and the composition further comprises a substrate for the double stranded RNA ligase, wherein the substrate comprises a polynucleotide acceptor comprising at least a modified nucleoside and/or internucleoside linkage; and/or a polynucleotide donor comprising at least a modified nucleoside and/or internucleoside linkage, and/or a polynucleotide strand comprising a region complementary to the polynucleotide acceptor and polynucleotide donor, wherein the complementary polynucleotide strand comprises at least a modified nucleoside and/or internucleoside linkage.

125. The composition of Claim 115, wherein the polynucleotide polymerase comprises a template-dependent DNA polymerase.

126. The composition of Claim 125, comprising the template-dependent DNA polymerase, and a substrate for the template-dependent DNA polymerase, wherein the substrate comprises a natural or unnatural or modified deoxynucleoside triphosphate.

127. The composition of Claim 115, wherein the polynucleotide polymerase comprises an RNA polymerase.

128. The composition of Claim 127, wherein the substrate for the RNA polymerase comprises a natural or unnatural or modified nucleoside triphosphate.

129. The composition of any one of Claims 107-128, wherein the inorganic pyrophosphatase is immobilized on a support medium.

130. The composition of any one of Claims 107-129, further comprising a non-natural or industrial process condition comprising one or more of the following: a buffer, high salt concentration as compared to a naturally occurring concentration, high substrate concentration as compared to a naturally occurring concentration, one or more non-natural or non-preferred cofactors (such as a divalent metal cofactor), an isolated or purified inorganic pyrophosphatase, high inorganic pyrophosphatase concentration as compared to a naturally occurring concentration, an organic solvent, or a high or low pH as compared to a naturally occurring pH.