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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
07 April 2022 (07.04.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/072490 A1

(51) International Patent Classification:
C12N 9/12 (2006.01) *C12P 13/02* (2006.01)
C12N 9/14 (2006.01) *C12P 13/04* (2006.01)
C12N 9/10 (2006.01)

(21) International Application Number:
PCT/US2021/052644

(22) International Filing Date:
29 September 2021 (29.09.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/086,688 02 October 2020 (02.10.2020) US

(71) Applicant: **CODEXIS, INC.** [US/US]; 200 Penobscot Drive, Redwood City, California 94063 (US).

(72) Inventors: **NAZOR, Jovana**; 200 Penobscot Drive, Redwood City, California 94063 (US). **KRAWCZYK, Mikayla, Jianghongxia**; 200 Penobscot Drive, Redwood City, California 94063 (US). **SEIBEL, Zara, Maxine**; 200 Penobscot Drive, Redwood City, California 94063 (US). **SUBRAMANIAN, Nandhitha**; 23 Lichfield House, Rustat Avenue, Cambridge CB1 3RD (GB). **KOLEV, Joshua**; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).

(74) Agent: **NEELY WILLIS, Melanie** et al.; Codexis, Inc., 200 Penobscot Drive, Redwood City, California 94063 (US).

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

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

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** ENGINEERED PANTOTHENATE KINASE VARIANT ENZYMES

(57) **Abstract:** The present invention provides engineered pantothenate kinase (PanK) enzymes, polypeptides having PanK activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. Methods for producing PanK enzymes are also provided. The present invention further provides compositions comprising the PanK enzymes and methods of using the engineered PanK enzymes. The present invention finds particular use in the production of pharmaceutical compounds.

WO 2022/072490 A1

ENGINEERED PANTOTHENATE KINASE VARIANT ENZYMES

[0001] The present application claims priority to US Prov. Pat. Appln. Ser. No. 63/086,688, filed October 2, 2020, which is incorporated by reference in its entirety, for all purposes.

FIELD OF THE INVENTION

[0002] The present invention provides engineered pantothenate kinase (PanK) enzymes, polypeptides having PanK activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. Methods for producing PanK enzymes are also provided. The present invention further provides compositions comprising the PanK enzymes and methods of using the engineered PanK enzymes. The present invention finds particular use in the production of pharmaceutical compounds.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0003] The official copy of the Sequence Listing is submitted concurrently with the specification as an ASCII formatted text file via EFS-Web, with a file name of "CX2-207WO1_ST25.txt", a creation date of September 29, 2021 and a size of 1.26 megabytes. The Sequence Listing filed via EFS-Web is part of the specification and incorporated in its entirety by reference herein.

BACKGROUND OF THE INVENTION

[0004] The retrovirus designated as human immunodeficiency virus (HIV) is the etiological agent of acquired immune deficiency syndrome (AIDS), a complex disease that involves progressive destruction of affected individuals' immune systems and degeneration of the central and peripheral nervous systems. A common feature of retrovirus replication is reverse transcription of the viral RNA genome by a virally-encoded reverse transcriptase to generate DNA copies of HIV sequences, required for viral replication. Some compounds, such as MK-8591, are known reverse transcriptase inhibitors and have found use in the treatment of AIDS and similar diseases. While there are some compounds known to inhibit HIV reverse transcriptase, there remains a need in the art for additional compounds that are more effective in inhibiting this enzyme and thereby ameliorating the effects of AIDS.

[0005] Nucleoside analogues such as MK-8591 (Merck) are effective inhibitors of HIV's reverse transcriptase due their similarity to natural nucleosides used in the synthesis of DNA. The binding of these analogues by the reverse transcriptase stalls the synthesis of DNA by inhibiting the progressive nature of the reverse transcriptase. The stalling of the enzyme results in the premature termination of the DNA molecule making it ineffective. However, production of nucleoside analogues by standard chemical synthetic techniques can pose a challenge due to their chemical complexity.

SUMMARY OF THE INVENTION

[0006] The present invention provides engineered pantothenate kinase (PanK) enzymes, polypeptides having PanK activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. Methods for producing PanK enzymes are also provided. The present invention further provides compositions comprising the PanK enzymes and methods of using the engineered PanK enzymes. The present invention finds particular use in the production of pharmaceutical compounds.

[0007] In some embodiments, the present invention provides engineered pantothenate kinases comprising polypeptide sequences having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 4, SEQ ID NO: 44, SEQ ID NO: 320, and/or SEQ ID NO: 526, or a functional fragment thereof, wherein the engineered pantothenate kinases comprise at least one substitution or substitution set in the polypeptide sequences, and wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 4, SEQ ID NO: 44, SEQ ID NO: 320, and/or SEQ ID NO: 526.

[0008] In some embodiments, the present invention provides engineered pantothenate kinases comprising polypeptide sequences having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 4, or a functional fragment thereof, wherein the engineered pantothenate kinases comprise at least one substitution or substitution set in the polypeptide sequences, and wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 4. In some embodiments, the engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 4, or a functional fragment thereof, and wherein the engineered pantothenate kinase comprises at least one substitution or substitution set at one or more positions selected from 17, 19, 26, 31, 33, 34, 35, 37, 44, 46, 63, 73, 74, 76, 77, 79, 80, 82, 85, 91, 94, 95, 96, 103, 103/135, 103/135/141, 103/135/216, 103/135/291, 103/141, 103/141/193/216/238, 103/141/216, 103/141/238, 103/216/238, 120, 124, 126, 128, 129, 130, 135, 135/141/238, 135/141/298/323, 135/193/238/289/291, 135/219, 135/238/291, 135/248/251/291/297, 135/251/323, 135/323, 141, 141/193/251/323, 141/251, 141/251/323, 141/297, 141/323, 145, 154, 164, 193, 193/216, 194, 195, 197, 198, 216, 218, 221, 222, 225, 227, 233, 238, 247, 248, 250, 251, 274, 286, 287, 289, 290, 291, 293, 297, 308, 315, and 323, and wherein the amino acid positions of the polypeptide sequence are numbered with reference to SEQ ID NO: 4. In some additional embodiments, the engineered pantothenate kinase comprises at least one substitution or substitution set selected from 17V, 19L, 26A, 26T, 31R, 33I, 34A, 35R, 35S, 37F, 44W, 46K, 63L, 73G, 74L, 76V, 77L, 79A, 79S, 79Y, 80E, 82T, 85S, 91C, 94S, 95G, 96P, 103V, 103V/135S, 103V/135S/141L, 103V/135S/216G, 103V/135S/291K, 103V/141L, 103V/141L/193V/216G/238F, 103V/141L/216G, 103V/141L/238F, 103V/216G/238F, 120L, 120T, 124G, 126Q, 126S, 126V, 128S, 129G, 129S, 130Y, 135S, 135S/141L/238F, 135S/141L/298A/323C, 135S/193V/238F/289A/291K,

135S/219G, 135S/238F/291K, 135S/248A/251L/291K/297R, 135S/251L/323C, 135S/323C, 141L, 141L/193V/251L/323C, 141L/251L, 141L/251L/323C, 141L/297V, 141L/323C, 145I, 154H, 154R, 164P, 193V, 193V/216G, 194H, 195A, 195G, 197G, 197V, 198R, 216G, 218C, 221E, 221S, 222G, 225Y, 227T, 233C, 238F, 247T, 248A, 248S, 250F, 251L, 274A, 286D, 287L, 289A, 289S, 290G, 291G, 293L, 297R, 297V, 308A, 315L, and 323C, wherein the amino acid positions of the polypeptide sequence are numbered with reference to SEQ ID NO: 4. In some additional embodiments, the engineered pantothenate kinase comprises at least one substitution or substitution set selected from T17V, M19L, D26A, D26T, A31R, V33I, R34A, D35R, D35S, N37F, D44W, I46K, E63L, N73G, F74L, I76V, S77L, N79A, N79S, N79Y, L80E, R82T, Q85S, G91C, G94S, Q95G, R96P, I103V, I103V/T135S, I103V/T135S/H141L, I103V/T135S/S216G, I103V/T135S/M291K, I103V/H141L, I103V/H141L/I193V/S216G/Y238F, I103V/H141L/S216G, I103V/H141L/Y238F, I103V/S216G/Y238F, A120L, A120T, R124G, P126Q, P126S, P126V, H128S, R129G, R129S, R130Y, T135S, T135S/H141L/Y238F, T135S/H141L/Q298A/R323C, T135S/I193V/Y238F/K289A/M291K, T135S/D219G, T135S/Y238F/M291K, T135S/T248A/I251L/M291K/K297R, T135S/I251L/R323C, T135S/R323C, H141L, H141L/I193V/I251L/R323C, H141L/I251L, H141L/I251L/R323C, H141L/K297V, H141L/R323C, V145I, K154H, K154R, H164P, I193V, I193V/S216G, P194H, D195A, D195G, D197G, D197V, K198R, S216G, M218C, P221E, P221S, H222G, H225Y, V227T, V233C, Y238F, Q247T, T248A, T248S, Y250F, I251L, K274A, T286D, I287L, K289A, K289S, E290G, M291G, S293L, K297R, K297V, S308A, A315L, and R323C, wherein the amino acid positions of the polypeptide sequence are numbered with reference to SEQ ID NO: 4. In some embodiments, the engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered pantothenate kinase variant set forth in Table 5-1.

[0009] In some further embodiments, the engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 4. In some additional embodiments, the engineered pantothenate kinase is a variant engineered polypeptide set forth in SEQ ID NO: 4.

[0010] In some embodiments, the present invention provides engineered pantothenate kinases comprising polypeptide sequences having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 44, or a functional fragment thereof, wherein the engineered pantothenate kinases comprise at least one substitution or substitution set in the polypeptide sequences, and wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 44. In some embodiments, the engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 44, or a functional fragment thereof, and wherein the engineered pantothenate kinase comprises at

least one substitution or substitution set at one or more positions selected from 16/93/120/221/308, 16/141/179/282, 17, 17/34, 17/34/46, 17/34/96/128/129/164, 17/46, 17/46/63, 17/46/63/96, 17/46/63/164, 17/46/126/128/129/164, 17/46/128/129/130/164/193, 17/46/164, 17/63, 17/63/79, 17/63/79/128/129/130/164, 17/63/96, 17/63/96/164, 17/63/129/130, 17/63/164, 17/63/193, 17/79/128/129/130, 17/126, 17/126/129/130/164, 17/126/129/164, 19, 19/35/85/274, 19/35/197/218/275, 19/74/77/141/218, 19/74/145, 19/77/141/218/225, 19/85/145/197/218/308, 19/141, 19/141/225, 20/59/93, 20/59/93/141/321, 20/59/120/164/263/282, 20/164, 20/164/179/308/320, 20/282, 20/308, 26/31/33/35/76/79/94/221/293, 26/31/33/35/120/141/293, 26/31/33/76/79, 26/31/33/76/79/94, 26/31/120, 26/33/35/126/222/291, 26/33/80/126/291/293, 26/33/94/293, 26/35/79/120/126/194/198, 26/35/120/222/227/293, 26/35/126, 31/33/35/76/79/91/291, 31/33/35/76/141/293, 31/33/35/120/126/227, 31/33/35/126, 31/33/76/79/80/94/291, 31/33/76/79/95, 31/33/120, 31/35/80/94/126/141/293, 31/76/79/120/126/293, 31/76/79/293, 32/46, 33/35, 33/35/76/79/120/141/198/221, 33/35/76/80/126/221/222/227, 33/35/76/80/293, 33/35/94/126/141, 33/35/126/198/293, 33/35/126/222/227, 33/35/126/293, 33/35/221/222, 33/76/79, 33/76/79/126, 33/79/126/221/227/293, 34, 34/46, 34/46/48, 34/46/63, 34/46/63/164, 34/46/79/96/126, 34/46/79/164, 34/46/164, 34/63/126/129/193, 34/79/126/128/130, 34/164, 35, 35/74/77/197/225, 35/74/218, 35/76, 35/76/79/80/126, 35/76/80/126, 35/77/218, 35/85, 35/85/141/218, 35/85/218/225, 35/85/218/274, 35/94/120/126/293, 35/141/195/218/274, 35/145/286, 35/197/274, 35/218/308, 35/274, 46, 46/63, 46/79, 46/79/126/193, 46/96/164, 46/128/129/164, 46/129, 46/164, 46/164/193, 46/193, 59, 59/120, 63, 63/164, 63/193, 74/85/145, 74/141, 76/79/80/120, 77/85/145, 77/218, 79/96/126/129, 79/126/164, 80/126/221/227, 94/95/120/291, 94/126/293, 94/293, 120, 120/164/282/321, 126/128/129/193, 126/129/164, 128/129/130, 141/195/197/218, 164, 164/221/308, 218, 225, 263, 274, and 308, wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 44. In some additional embodiments, the engineered pantothenate kinase comprises at least one substitution or substitution set selected from 16R/93G/120V/221C/308M, 16R/141H/179V/282A, 17V, 17V/34A, 17V/34A/46K, 17V/34A/96P/128S/129G/164P, 17V/46K, 17V/46K/63L, 17V/46K/63L/96P, 17V/46K/63L/164P, 17V/46K/126V/128S/129S/164P, 17V/46K/128S/129G/130Y/164P/193V, 17V/46K/164P, 17V/63L, 17V/63L/79S, 17V/63L/79S/128S/129G/130Y/164P, 17V/63L/96P, 17V/63L/96P/164P, 17V/63L/129S/130Y, 17V/63L/164P, 17V/63L/193V, 17V/79S/128S/129S/130Y, 17V/126Q/129G/164P, 17V/126V, 17V/126V/129S/130Y/164P, 19L, 19L/35R/85S/274A, 19L/35R/197G/218C/275M, 19L/74L/77L/141H/218C, 19L/74L/145I, 19L/77L/141H/218C/225Y, 19L/85S/145I/197G/218C/308A, 19L/141H, 19L/141H/225Y, 20G/59A/93G/141H/321L, 20G/59R/93G, 20G/59R/120V/164S/263M/282A, 20G/164S, 20G/164S/179V/308M/320K, 20G/282A, 20G/308M, 26T/31R/33I/35S/76V/79A/94S/221E/293L, 26T/31R/33I/35S/120T/141H/293L, 26T/31R/33I/76V/79A, 26T/31R/33I/76V/79A/94S, 26T/31R/120T, 26T/33I/35S/126S/222G/291G, 26T/33I/80E/126S/291G/293L, 26T/33I/94S/293L,

26T/35S/79A/120T/126S/194H/198R, 26T/35S/120T/222G/227T/293L, 26T/35S/126S, 31R/33I/35S/76V/79A/91C/291G, 31R/33I/35S/76V/141H/293L, 31R/33I/35S/120T/126S/227T, 31R/33I/35S/126S, 31R/33I/76V/79A/80E/94S/291G, 31R/33I/76V/79A/95G, 31R/33I/120T, 31R/35S/80E/94S/126S/141H/293L, 31R/76V/79A/120T/126S/293L, 31R/76V/79A/293L, 32S/46K, 33F/35S/126S/293P, 33I/35S, 33I/35S/76V/79A/120T/141H/198R/221E, 33I/35S/76V/80E/126S/221E/222G/227T, 33I/35S/76V/80E/293L, 33I/35S/94S/126S/141H, 33I/35S/126S/198R/293L, 33I/35S/126S/222G/227T, 33I/35S/221E/222G, 33I/76V/79A, 33I/76V/79A/126S, 33I/79A/126S/221E/227T/293L, 34A, 34A/46K, 34A/46K/48C, 34A/46K/63L, 34A/46K/63L/164P, 34A/46K/79S/96P/126V, 34A/46K/79S/164P, 34A/46K/164P, 34A/63L/126V/129S/193V, 34A/79S/126Q/128S/130Y, 34A/164P, 35R, 35R/74L/77L/197V/225Y, 35R/74L/218C, 35R/77L/218C, 35R/85S, 35R/85S/141H/218C, 35R/85S/218C/225Y, 35R/85S/218C/274A, 35R/141H/195G/218C/274A, 35R/145I/286D, 35R/197V/274A, 35R/218C/308A, 35R/274A, 35S/76V, 35S/76V/79A/80E/126S, 35S/76V/80E/126S, 35S/94S/120T/126S/293L, 46K, 46K/63L, 46K/79S, 46K/79S/126V/193V, 46K/96P/164P, 46K/128S/129G/164P, 46K/129G, 46K/164P, 46K/164P/193V, 46K/193V, 59R, 59R/120V, 63L, 63L/164P, 63L/193V, 74L/85S/145I, 74L/141H, 76V/79A/80E/120T, 77L/85S/145I, 77L/218C, 79S/96P/126Q/129S, 79S/126V/164P, 80E/126S/221E/227T, 94S/95G/120T/291G, 94S/126S/293L, 94S/293L, 120V, 120V/164S/282A/321L, 126V/128S/129G/193V, 126V/129G/164P, 128S/129G/130Y, 141H/195G/197G/218C, 164S, 164S/221C/308M, 218C, 225Y, 263M, 274A, 308A, and 308M,wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 44. In some additional embodiments, the engineered pantothenate kinase comprises at least one substitution or substitution set selected from Q16R/N93G/A120V/P221C/S308M, Q16R/L141H/N179V/K282A, T17V, T17V/R34A, T17V/R34A/I46K, T17V/R34A/R96P/H128S/R129G/H164P, T17V/I46K, T17V/I46K/E63L, T17V/I46K/E63L/R96P, T17V/I46K/E63L/H164P, T17V/I46K/P126V/H128S/R129S/H164P, T17V/I46K/H128S/R129G/R130Y/H164P/I193V, T17V/I46K/H164P, T17V/E63L, T17V/E63L/N79S, T17V/E63L/N79S/H128S/R129G/R130Y/H164P, T17V/E63L/R96P, T17V/E63L/R96P/H164P, T17V/E63L/R129S/R130Y, T17V/E63L/H164P, T17V/E63L/I193V, T17V/N79S/H128S/R129S/R130Y, T17V/P126Q/R129G/H164P, T17V/P126V, T17V/P126V/R129S/R130Y/H164P, M19L, M19L/D35R/Q85S/K274A, M19L/D35R/D197G/M218C/L275M, M19L/F74L/S77L/L141H/M218C, M19L/F74L/V145I, M19L/S77L/L141H/M218C/H225Y, M19L/Q85S/V145I/D197G/M218C/S308A, M19L/L141H, M19L/L141H/H225Y, T20G/E59A/N93G/L141H/E321L, T20G/E59R/N93G, T20G/E59R/A120V/H164S/T263M/K282A, T20G/H164S, T20G/H164S/N179V/S308M/E320K, T20G/K282A, T20G/S308M, D26T/A31R/V33I/D35S/I76V/N79A/G94S/P221E/S293L, D26T/A31R/V33I/D35S/A120T/L141H/S293L, D26T/A31R/V33I/I76V/N79A, D26T/A31R/V33I/I76V/N79A/G94S, D26T/A31R/A120T, D26T/V33I/D35S/P126S/H222G/M291G,

D26T/V33I/L80E/P126S/M291G/S293L, D26T/V33I/G94S/S293L, D26T/D35S/N79A/A120T/P126S/P194H/K198R, D26T/D35S/A120T/H222G/V227T/S293L, D26T/D35S/P126S, A31R/V33I/D35S/I76V/N79A/G91C/M291G, A31R/V33I/D35S/I76V/L141H/S293L, A31R/V33I/D35S/A120T/P126S/V227T, A31R/V33I/D35S/P126S, A31R/V33I/I76V/N79A/L80E/G94S/M291G, A31R/V33I/I76V/N79A/Q95G, A31R/V33I/A120T, A31R/D35S/L80E/G94S/P126S/L141H/S293L, A31R/I76V/N79A/A120T/P126S/S293L, A31R/I76V/N79A/S293L, A32S/I46K, V33F/D35S/P126S/S293P, V33I/D35S, V33I/D35S/I76V/N79A/A120T/L141H/K198R/P221E, V33I/D35S/I76V/L80E/P126S/P221E/H222G/V227T, V33I/D35S/I76V/L80E/S293L, V33I/D35S/G94S/P126S/L141H, V33I/D35S/P126S/K198R/S293L, V33I/D35S/P126S/H222G/V227T, V33I/D35S/P221E/H222G, V33I/I76V/N79A, V33I/I76V/N79A/P126S, V33I/N79A/P126S/P221E/V227T/S293L, R34A, R34A/I46K, R34A/I46K/R48C, R34A/I46K/E63L, R34A/I46K/E63L/H164P, R34A/I46K/N79S/R96P/P126V, R34A/I46K/N79S/H164P, R34A/I46K/H164P, R34A/E63L/P126V/R129S/I193V, R34A/N79S/P126Q/H128S/R130Y, R34A/H164P, D35R, D35R/F74L/S77L/D197V/H225Y, D35R/F74L/M218C, D35R/S77L/M218C, D35R/Q85S, D35R/Q85S/L141H/M218C, D35R/Q85S/M218C/H225Y, D35R/Q85S/M218C/K274A, D35R/L141H/D195G/M218C/K274A, D35R/V145I/T286D, D35R/D197V/K274A, D35R/M218C/S308A, D35R/K274A, D35S/I76V, D35S/I76V/N79A/L80E/P126S, D35S/I76V/L80E/P126S, D35S/G94S/A120T/P126S/S293L, I46K, I46K/E63L, I46K/N79S, I46K/N79S/P126V/I193V, I46K/R96P/H164P, I46K/H128S/R129G/H164P, I46K/R129G, I46K/H164P, I46K/H164P/I193V, I46K/I193V, E59R, E59R/A120V, E63L, E63L/H164P, E63L/I193V, F74L/Q85S/V145I, F74L/L141H, I76V/N79A/L80E/A120T, S77L/Q85S/V145I, S77L/M218C, N79S/R96P/P126Q/R129S, N79S/P126V/H164P, L80E/P126S/P221E/V227T, G94S/Q95G/A120T/M291G, G94S/P126S/S293L, G94S/S293L, A120V, A120V/H164S/K282A/E321L, P126V/H128S/R129G/I193V, P126V/R129G/H164P, H128S/R129G/R130Y, L141H/D195G/D197G/M218C, H164S, H164S/P221C/S308M, M218C, H225Y, T263M, K274A, S308A, and S308M, wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 44. In some embodiments, the engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered pantothenate kinase variant set forth in Table 6-1. In some further embodiments, the engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 44. In some additional embodiments, the engineered pantothenate kinase is a variant engineered polypeptide set forth in SEQ ID NO: 44.

[0011] In some embodiments, the present invention provides engineered pantothenate kinases comprising polypeptide sequences having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%,

94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 320, or a functional fragment thereof, wherein the engineered pantothenate kinases comprise at least one substitution or substitution set in the polypeptide sequences, and wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 320. In some embodiments, the engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 320, or a functional fragment thereof, and wherein the engineered pantothenate kinase comprises at least one substitution or substitution set at one or more positions selected from 20/31/35/48/63/74/94/120/164, 20/31/35/74/79/164/218/225/282, 20/35/120/197, 20/48/63/94/120/145/197/218/227, 20/48/218/227, 20/94/120, 20/120/145/197/227, 20/197/225/227, 24, 33, 46, 48, 48/63, 48/63/74/79/120/129/227/282, 48/63/94/275, 48/63/145, 63, 63/120/197/218/275, 63/129/145, 63/218/275/282, 74/79/94/275, 94/120, 95, 130, 149, 164, 178, 196, 201, 221, 271, 277, 282, 289, 293, 298, 314, and 321, wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 320. In some additional embodiments, the engineered pantothenate kinase comprises at least one substitution or substitution set selected from 20G/31R/35D/48C/63L/74L/94S/120T/164P, 20G/31R/35D/74L/79A/164S/218C/225Y/282A, 20G/35D/120T/197G, 20G/48C/63L/94S/120T/145I/197G/218C/227V, 20G/48C/218C/227V, 20G/94S/120T, 20G/120T/145I/197G/227V, 20G/197G/225Y/227V, 24H, 24V, 33V, 46A, 48C/63L, 48C/63L/74L/79S/120T/129G/227V/282A, 48C/63L/94S/275M, 48C/63L/145I, 48V, 63L, 63L/120T/197G/218C/275M, 63L/129S/145I, 63L/218C/275M/282A, 74L/79A/94S/275M, 94S/120T, 95D, 95E, 130L, 149A, 164G, 178D, 178H, 196I, 201G, 221G, 271C, 271M, 277R, 282D, 289C, 293Y, 298D, 298E, 314G, and 321L,wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 320. In some additional embodiments, the engineered pantothenate kinase comprises at least one substitution or substitution set selected from T20G/A31R/S35D/R48C/E63L/F74L/G94S/A120T/H164P, T20G/A31R/S35D/F74L/N79A/H164S/M218C/H225Y/K282A, T20G/S35D/A120T/D197G, T20G/R48C/E63L/G94S/A120T/V145I/D197G/M218C/T227V, T20G/R48C/M218C/T227V, T20G/G94S/A120T, T20G/A120T/V145I/D197G/T227V, T20G/D197G/H225Y/T227V, Q24H, Q24V, I33V, I46A, R48C/E63L, R48C/E63L/F74L/N79S/A120T/R129G/T227V/K282A, R48C/E63L/G94S/L275M, R48C/E63L/V145I, R48V, E63L, E63L/A120T/D197G/M218C/L275M, E63L/R129S/V145I, E63L/M218C/L275M/K282A, F74L/N79A/G94S/L275M, G94S/A120T, Q95D, Q95E, R130L, R149A, H164G, P178D, P178H, G196I, V201G, P221G, G271C, G271M, K277R, K282D, K289C, S293Y, Q298D, Q298E, S314G, and E321L, wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 320. In some embodiments, the engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence

of at least one engineered pantothenate kinase variant set forth in Table 7-1. In some further embodiments, the engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 320. In some additional embodiments, the engineered pantothenate kinase is a variant engineered polypeptide set forth in SEQ ID NO: 320.

[0012] In some embodiments, the present invention provides engineered pantothenate kinases comprising polypeptide sequences having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 526, or a functional fragment thereof, wherein the engineered pantothenate kinases comprise at least one substitution or substitution set in the polypeptide sequences, and wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 526. In some embodiments, the engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 526, or a functional fragment thereof, and wherein the engineered pantothenate kinase comprises at least one substitution or substitution set at one or more positions selected from 24, 24/33, 24/46/221, 24/95/164/178/201/282, 24/149/178/282/293/298, 24/149/201/282, 24/149/282/314, 24/178/201/271/282/293/298/314, 24/201/271/282/293/298, 24/201/282, 24/227, 33/95/104, 46/95/282, 48, 95/201/282/293/314, 164/201/271/314, 178/293/298, 201, 201/282, 221/289, 282/293/298/314/321, and 298, wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 526. In some additional embodiments, the engineered pantothenate kinase comprises at least one substitution or substitution set selected from 24H/33V, 24H/46V/221G, 24H/227V, 24V, 24V/95D/164G/178D/201G/282D, 24V/149A/178D/282D/293Y/298E, 24V/149A/201G/282D, 24V/149A/282D/314G, 24V/178D/201G/271C/282D/293Y/298E/314G, 24V/201G/271C/282D/293Y/298E, 24V/201G/282D, 33V/95N/104T, 46V/95E/282A, 48V, 95D/201G/282D/293Y/314G, 164G/201G/271C/314G, 178D/293Y/298D, 201G, 201G/282D, 221G/289C, 282D/293Y/298E/314G/321L, and 298D,wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 526. In some additional embodiments, the engineered pantothenate kinase comprises at least one substitution or substitution set selected from Q24H/I33V, Q24H/I46V/P221G, Q24H/T227V, Q24V, Q24V/Q95D/H164G/P178D/V201G/K282D, Q24V/R149A/P178D/K282D/S293Y/Q298E, Q24V/R149A/V201G/K282D, Q24V/R149A/K282D/S314G, Q24V/P178D/V201G/G271C/K282D/S293Y/Q298E/S314G, Q24V/V201G/G271C/K282D/S293Y/Q298E, Q24V/V201G/K282D, I33V/Q95N/A104T, I46V/Q95E/K282A, C48V, Q95D/V201G/K282D/S293Y/S314G, H164G/V201G/G271C/S314G, P178D/S293Y/Q298D, V201G, V201G/K282D, P221G/K289C, K282D/S293Y/Q298E/S314G/E321L, and Q298D, wherein the amino acid positions of the polypeptide sequences are numbered with reference to SEQ ID NO: 526. In some embodiments, the

engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered pantothenate kinase variant set forth in Table 8-1. In some further embodiments, the engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 526. In some additional embodiments, the engineered pantothenate kinase is a variant engineered polypeptide set forth in SEQ ID NO: 526.

[0013] In some further embodiments, the engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered pantothenate kinase variant set forth in the even numbered sequences of SEQ ID NOS: 4-630. In some additional embodiments, the engineered pantothenate kinase comprises a polypeptide sequence forth in the even numbered sequences of SEQ ID NOS: 4-630. In some further embodiments, the engineered pantothenate kinase comprises at least one improved property compared to wild-type *E. coli* pantothenate kinase. In still some additional embodiments, the improved property comprises improved activity on a substrate, as compared to a wild-type pantothenate kinase. In some further embodiments, the substrate comprises an alcohol. In some further embodiments, the substrate comprises an aldehyde. In some additional embodiments, the improved property comprises improved production of phospho-ethynyl glycerol, as compared to a wild-type pantothenate kinase. In some further embodiments, the engineered pantothenate kinase is purified. The present invention also provides compositions comprising at least one engineered pantothenate kinase provided herein. In some embodiments, the present invention provides compositions comprising one engineered pantothenate kinase provided herein.

[0014] The present invention also provides polynucleotide sequences encoding at least one engineered pantothenate kinase provided herein. In some embodiments, the polynucleotide sequence encoding at least one engineered pantothenate kinase comprises at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 3, SEQ ID NO: 43, SEQ ID NO: 319, and/or SEQ ID NO: 525. In some embodiments, the polynucleotide sequence of the engineered pantothenate kinase comprises at least one substitution at one or more positions. In some further embodiments, the polynucleotide sequence encoding at least one engineered pantothenate kinase comprises at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 3, or a functional fragment thereof. In some further embodiments, the polynucleotide sequence encoding at least one engineered pantothenate kinase comprises at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 43, or a functional fragment thereof. In some further embodiments, the polynucleotide sequence encoding at least one engineered pantothenate kinase comprises at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 319, or a functional

fragment thereof. In some further embodiments, the polynucleotide sequence encoding at least one engineered pantothenate kinase comprises at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 525, or a functional fragment thereof. In some additional embodiments, the polynucleotide sequence is operably linked to a control sequence. In some embodiments, the polynucleotide sequence is codon optimized. In some further embodiments, the polynucleotide comprises an odd-numbered sequence of SEQ ID NOS: 1-629. The present invention also provides expression vectors comprising at least one polynucleotide sequence encoding a pantothenate kinase provided herein. In some embodiments, the expression vectors comprise one polynucleotide sequence encoding a pantothenate kinase provided herein. The present invention also provides host cells comprising at least one expression vector provided herein. The present invention also provides host cells comprising at least one polynucleotide sequence encoding a pantothenate kinase provided herein.

[0015] The present invention also provides methods of producing an engineered pantothenate kinase in a host cell, comprising culturing the host cell provided herein, under suitable conditions, such that at least one engineered pantothenate kinase is produced. In some embodiments, the methods further comprise recovering at least one engineered pantothenate kinase from the culture and/or host cell. In some additional embodiments, the methods further comprise the step of purifying the at least one engineered pantothenate kinase.

DESCRIPTION OF THE INVENTION

[0016] The present invention provides engineered pantothenate kinase (PanK) enzymes, polypeptides having PanK activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. Methods for producing PanK enzymes are also provided. The present invention further provides compositions comprising the PanK enzymes and methods of using the engineered PanK enzymes. The present invention finds particular use in the production of pharmaceutical compounds.

[0017] 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.

[0018] 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.

[0019] 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.

Abbreviations

[0020] The abbreviations used for the genetically encoded amino acids are conventional and are as follows: alanine (Ala or A), arginine (Arg or R), asparagine (Asn or N), aspartate (Asp or D), cysteine (Cys or C), glutamate (Glu or E), glutamine (Gln or Q), histidine (His or H), isoleucine (Ile or I), leucine (Leu or L), lysine (Lys or K), methionine (Met or M), phenylalanine (Phe or F), proline (Pro or P), serine (Ser or S), threonine (Thr or T), tryptophan (Trp or W), tyrosine (Tyr or Y), and valine (Val or V).

[0021] 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. 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.

[0022] 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). Unless specifically delineated, the abbreviated nucleosides may be either ribonucleosides or 2'-deoxyribonucleosides. The nucleosides may be specified as being either ribonucleosides or 2'-deoxyribonucleosides on an individual basis or on an aggregate basis. 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

[0023] 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.

[0024] As used herein, 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.

[0025] Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting. Thus, as used herein, the term "comprising" and its cognates are used in their inclusive sense (*i.e.*, equivalent to the term "including" and its corresponding cognates).

[0026] It is to be further 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."

[0027] As used herein, the term "about" means an acceptable error for a particular value. In some instances, "about" means within 0.05%, 0.5%, 1.0%, or 2.0%, of a given value range. In some instances, "about" means within 1, 2, 3, or 4 standard deviations of a given value.

[0028] As used herein, "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.

[0029] As used herein, "ATCC" refers to the American Type Culture Collection whose biorepository collection includes genes and strains.

[0030] As used herein, "NCBI" refers to National Center for Biological Information and the sequence databases provided therein.

[0031] As used herein, "pantothenate kinase," "PanK," refer to enzymes (EC 2.7.1.33), that phosphorylate pantothenate to form 4'-phosphopantothenate or variant enzymes derived from such PanK enzymes, whether or not such variant enzymes retain the same functionality as the source (i.e., "parent") enzyme. This includes both naturally-occurring or wild-type enzymes and enzymes generated by human engineering.

[0032] "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 or phosphorylation). 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.

[0033] "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.

[0034] As used herein, "hydrophilic amino acid or residue" refers to an amino acid or residue having a side chain exhibiting a hydrophobicity of less than zero according to the normalized consensus hydrophobicity scale of Eisenberg et al., (Eisenberg et al., J. Mol. Biol., 179:125-142 [1984]). Genetically encoded hydrophilic amino acids include L-Thr (T), L-Ser (S), L-His (H), L-Glu (E), L-Asn (N), L-Gln (Q), L-Asp (D), L-Lys (K) and L-Arg (R).

[0035] As used herein, "acidic amino acid or residue" refers to a hydrophilic amino acid or residue having a side chain exhibiting a pKa value of less than about 6 when the amino acid is included in a peptide or polypeptide. Acidic amino acids typically have negatively charged side chains at physiological pH due to loss of a hydrogen ion. Genetically encoded acidic amino acids include L-Glu (E) and L-Asp (D).

[0036] As used herein, "basic amino acid or residue" refers to a hydrophilic amino acid or residue having a side chain exhibiting a pKa value of greater than about 6 when the amino acid is included in a peptide or polypeptide. Basic amino acids typically have positively charged side chains at physiological pH due to association with hydronium ion. Genetically encoded basic amino acids include L-Arg (R) and L-Lys (K).

[0037] As used herein, "polar amino acid or residue" refers to a hydrophilic amino acid or residue having a side chain that is uncharged at physiological pH, but which has at least one bond in which the pair of electrons shared in common by two atoms is held more closely by one of the atoms. Genetically encoded polar amino acids include L-Asn (N), L-Gln (Q), L-Ser (S) and L-Thr (T).

[0038] As used herein, "hydrophobic amino acid or residue" refers to an amino acid or residue having a side chain exhibiting a hydrophobicity of greater than zero according to the normalized consensus hydrophobicity scale of Eisenberg et al., (Eisenberg et al., J. Mol. Biol., 179:125-142 [1984]). Genetically encoded hydrophobic amino acids include L-Pro (P), L-Ile (I), L-Phe (F), L-Val (V), L-Leu (L), L-Trp (W), L-Met (M), L-Ala (A) and L-Tyr (Y).

[0039] As used herein, "aromatic amino acid or residue" refers to a hydrophilic or hydrophobic amino acid or residue having a side chain that includes at least one aromatic or heteroaromatic ring. Genetically encoded aromatic amino acids include L-Phe (F), L-Tyr (Y) and L-Trp (W). Although owing to the pKa of its heteroaromatic nitrogen atom L-His (H) it is sometimes classified as a basic residue, or as an aromatic residue as its side chain includes a heteroaromatic ring, herein histidine is classified as a hydrophilic residue or as a "constrained residue" (see below).

[0040] As used herein, "constrained amino acid or residue" refers to an amino acid or residue that has a constrained geometry. Herein, constrained residues include L-Pro (P) and L-His (H). Histidine has a constrained geometry because it has a relatively small imidazole ring. Proline has a constrained geometry because it also has a five membered ring.

[0041] As used herein, "non-polar amino acid or residue" refers to a hydrophobic amino acid or residue having a side chain that is uncharged at physiological pH and which has bonds in which the pair of electrons shared in common by two atoms is generally held equally by each of the two atoms (i.e., the side chain is not polar). Genetically encoded non-polar amino acids include L-Gly (G), L-Leu (L), L-Val (V), L-Ile (I), L-Met (M) and L-Ala (A).

[0042] As used herein, "aliphatic amino acid or residue" refers to a hydrophobic amino acid or residue having an aliphatic hydrocarbon side chain. Genetically encoded aliphatic amino acids include L-Ala (A), L-Val (V), L-Leu (L) and L-Ile (I). It is noted that cysteine (or "L-Cys" or "[C]") is unusual in that it can form disulfide bridges with other L-Cys (C) amino acids or other sulfanyl- or sulfhydryl-containing amino acids. The "cysteine-like residues" include cysteine and other amino acids that contain sulfhydryl moieties that are available for formation of disulfide bridges. The ability of L-Cys (C) (and other amino acids with -SH containing side chains) to exist in a peptide in either the reduced free -SH or oxidized disulfide-bridged form affects whether L-Cys (C) contributes net hydrophobic or hydrophilic character to a peptide. While L-Cys (C) exhibits a hydrophobicity of 0.29 according to the normalized consensus scale of Eisenberg (Eisenberg et al., 1984, *supra*), it is to be understood that for purposes of the present disclosure, L-Cys (C) is categorized into its own unique group.

[0043] As used herein, "small amino acid or residue" refers to an amino acid or residue having a side chain that is composed of a total three or fewer carbon and/or heteroatoms (excluding the α-carbon and hydrogens). The small amino acids or residues may be further categorized as aliphatic, non-polar, polar or acidic small amino acids or residues, in accordance with the above definitions. Genetically-encoded small amino acids include L-Ala (A), L-Val (V), L-Cys (C), L-Asn (N), L-Ser (S), L-Thr (T) and L-Asp (D).

[0044] As used herein, "hydroxyl-containing amino acid or residue" refers to an amino acid containing a hydroxyl (-OH) moiety. Genetically-encoded hydroxyl-containing amino acids include L-Ser (S) L-Thr (T) and L-Tyr (Y).

[0045] As used herein, "polynucleotide" and "nucleic acid' refer to two or more 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), or comprised of mixtures of ribo- and 2' deoxyribonucleotides. While the nucleosides will typically be linked together via standard phosphodiester linkages, the polynucleotides may include one or more non-standard 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.

[0046] As used herein, "nucleoside" refers to glycosylamines comprising a nucleobase (i.e., a nitrogenous base), and a 5-carbon sugar (e.g., ribose or deoxyribose). Non-limiting examples of nucleosides include cytidine, uridine, adenosine, guanosine, thymidine, and inosine. In contrast, the term "nucleotide" refers to the glycosylamines comprising a nucleobase, a 5-carbon sugar, and one or more phosphate groups. In some embodiments, nucleosides can be phosphorylated by kinases to produce nucleotides.

[0047] As used herein, "nucleoside diphosphate" refers to glycosylamines comprising a nucleobase (i.e., a nitrogenous base), a 5-carbon sugar (e.g., ribose or deoxyribose), and a diphosphate (i.e., pyrophosphate) moiety. In some embodiments herein, "nucleoside diphosphate" is abbreviated as "NDP." Non-limiting examples of nucleoside diphosphates include cytidine diphosphate (CDP), uridine diphosphate (UDP), adenosine diphosphate (ADP), guanosine diphosphate (GDP), thymidine diphosphate (TDP), and inosine diphosphate (IDP). The terms "nucleoside" and "nucleotide" may be used interchangeably in some contexts.

[0048] As used herein, "coding sequence" refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0049] As used herein, the terms "biocatalysis," "biocatalytic," "biotransformation," and "biosynthesis" refer to the use of enzymes to perform chemical reactions on organic compounds.

[0050] As used herein, "wild-type" and "naturally-occurring" refer to the form found in nature. For example, a 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.

[0051] As used herein, "recombinant," "engineered," "variant," and "non-naturally occurring" when used with reference to a cell, nucleic acid, or polypeptide, refers 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 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 express native genes that are otherwise expressed at a different level.

[0052] The term "percent (%) sequence identity" is used 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 by any suitable method, including, but not limited to the local homology algorithm of Smith and Waterman (Smith and Waterman, Adv. Appl. Math., 2:482 [1981]), by the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, J. Mol. Biol., 48:443 [1970]), by the search for similarity method of Pearson and Lipman (Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85:2444 [1988]), by computerized implementations of these algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the 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., 215: 403-410 [1990]; and Altschul et al., Nucl. Acids Res., 3389-3402 [1977], 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 (See, 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 word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (See, Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 [1989]). 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.

[0053] As used herein, "reference sequence" refers to a defined sequence used as a basis for a sequence and/or activity comparison. 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, at least 100 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 polypeptides 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.

[0054] As used herein, "comparison window" refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acid residues wherein a sequence may be compared to a reference sequence of at least 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. The comparison window can be longer than 20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

[0055] As used herein, "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 pantothenate kinase, 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.

[0056] 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.

[0057] As used herein, "amino acid difference" and "residue difference" refer to a difference 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. In some cases, the reference sequence has a histidine tag, but the numbering is maintained relative to the equivalent reference sequence without the histidine tag. 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 X93 as compared to SEQ ID NO: 4" refers to a difference of the amino acid residue at the polypeptide position corresponding to position 93 of SEQ ID NO: 4. Thus, if the reference polypeptide of SEQ ID NO: 4 has a serine at position 93, then a "residue difference at position X93 as compared to SEQ ID NO: 4" an amino acid substitution of any residue other than serine at the position of the polypeptide corresponding to position 93 of SEQ ID NO: 4. 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 instances (e.g., in the Table presented in the Examples), 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. 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 residue differences are present relative to the reference sequence. In some embodiments, where more than one amino acid can be used in a specific residue position of a polypeptide, the various amino acid residues that can be used are separated by a "/" (e.g., X307H/X307P or X307H/P). The slash may also be used to indicate multiple substitutions within a given variant (i.e., there is more than one substitution present in a given sequence, such as in a combinatorial variant). In some embodiments, the present invention

includes engineered polypeptide sequences comprising one or more amino acid differences comprising conservative or non-conservative amino acid substitutions. In some additional embodiments, the present invention provides engineered polypeptide sequences comprising both conservative and non-conservative amino acid substitutions.

[0058] 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, in some embodiments, 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 an hydroxyl side chain (e.g., serine and threonine); an amino acids having aromatic side chains 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 basis 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.

[0059] As used herein, "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.

[0060] As used herein, "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 pantothenate kinase 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. Deletions are typically indicated by "-" in amino acid sequences.

[0061] As used herein, "insertion" refers to modification to the polypeptide by addition of one or more amino acids from the reference polypeptide. 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.

[0062] The term "amino acid substitution set" or "substitution set" refers to a group of amino acid substitutions in a polypeptide sequence, as compared to a reference sequence. A substitution set can have 1, 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 variant pantothenate kinases listed in the Tables provided in the Examples

[0063] A "functional fragment" and "biologically active fragment" are used interchangeably herein to refer to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion(s) and/or internal deletions, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence to which it is being compared (e.g., a full-length engineered pantothenate kinase of the present invention) and that retains substantially all of the activity of the full-length polypeptide.

[0064] As used herein, "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., within a host cell or via in vitro synthesis). The recombinant pantothenate kinase polypeptides 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 recombinant pantothenate kinase polypeptides can be an isolated polypeptide.

[0065] As used herein, "substantially pure polypeptide" or "purified protein" 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. However, in some embodiments, the composition comprising pantothenate kinase comprises pantothenate kinase that is less than 50% pure (e.g., about 10%, about 20%, about 30%, about 40%, or about 50%) Generally, a substantially pure pantothenate kinase composition comprises 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 recombinant pantothenate kinase polypeptides are substantially pure polypeptide compositions.

[0066] As used herein, "improved enzyme property" refers to at least one improved property of an enzyme. In some embodiments, the present invention provides engineered pantothenate kinase

polypeptides that exhibit an improvement in any enzyme property as compared to a reference pantothenate kinase polypeptide and/or a wild-type pantothenate kinase polypeptide, and/or another engineered pantothenate kinase polypeptide. Thus, the level of "improvement" can be determined and compared between various pantothenate kinase polypeptides, including wild-type, as well as engineered pantothenate kinases. Improved properties include, but are not limited, to such properties as increased protein expression, increased thermoactivity, increased thermostability, increased pH activity, increased stability, increased enzymatic activity, increased substrate specificity or affinity, increased specific activity, increased resistance to substrate or end-product inhibition, increased chemical stability, improved chemoselectivity, improved solvent stability, increased tolerance to acidic pH, increased tolerance to proteolytic activity (i.e., reduced sensitivity to proteolysis), reduced aggregation, increased solubility, and altered temperature profile. In additional embodiments, the term is used in reference to the at least one improved property of pantothenate kinase enzymes. In some embodiments, the present invention provides engineered pantothenate kinase polypeptides that exhibit an improvement in any enzyme property as compared to a reference pantothenate kinase polypeptide and/or a wild-type pantothenate kinase polypeptide, and/or another engineered pantothenate kinase polypeptide. Thus, the level of "improvement" can be determined and compared between various pantothenate kinase polypeptides, including wild-type, as well as engineered pantothenate kinases.

[0067] As used herein, "increased enzymatic activity" and "enhanced catalytic activity" refer to an improved property of the engineered 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 enzyme) as compared to the reference enzyme. In some embodiments, the terms refer to an improved property of engineered pantothenate kinase polypeptides provided herein, 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 pantothenate kinase) as compared to the reference pantothenate kinase enzyme. In some embodiments, the terms are used in reference to improved pantothenate kinase enzymes provided herein. Exemplary methods to determine enzyme activity of the engineered pantothenate kinases of the present invention are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of Km, Vmax or kcat, changes of which can lead to increased enzymatic activity. For example, improvements in enzyme activity can be from about 1.1 fold the enzymatic activity of the corresponding wild-type enzyme, to as much as 2-fold, 5-fold, 10-fold, 20-fold, 25-fold, 50-fold, 75-fold, 100-fold, 150-fold, 200-fold or more enzymatic activity than the naturally occurring pantothenate kinase or another engineered pantothenate kinase from which the pantothenate kinase polypeptides were derived.

[0068] As used herein, "conversion" refers to the enzymatic conversion (or biotransformation) of a 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 pantothenate kinase polypeptide can be expressed as "percent conversion" of the substrate to the product in a specific period of time.

[0069] Enzymes with "generalist properties" (or "generalist enzymes") refer to enzymes that exhibit improved activity for a wide range of substrates, as compared to a parental sequence. Generalist enzymes do not necessarily demonstrate improved activity for every possible substrate. In some embodiments, the present invention provides pantothenate kinase variants with generalist properties, in that they demonstrate similar or improved activity relative to the parental gene for a wide range of sterically and electronically diverse substrates. In addition, the generalist enzymes provided herein were engineered to be improved across a wide range of diverse molecules to increase the production of metabolites/products.

[0070] 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 (Tm) 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 Tm values for polynucleotides can be calculated using known methods for predicting melting temperatures (See e.g., Baldino et al., Meth. Enzymol., 168:761-777 [1989]; Bolton et al., Proc. Natl. Acad. Sci. USA 48:1390 [1962]; Bresslauer et al., Proc. Natl. Acad. Sci. USA 83:8893-8897 [1986]; Freier et al., Proc. Natl. Acad. Sci. USA 83:9373-9377 [1986]; Kierzek et al., Biochem., 25:7840-7846 [1986]; Rychlik et al., Nucl. Acids Res., 18:6409-6412 [1990] (erratum, Nucl. Acids Res., 19:698 [1991]); 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. 26:227-259 [1991]). 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 pantothenate kinase enzyme of the present invention.

[0071] As used herein, "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 Tm 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.

[0072] As used herein, "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 pantothenate kinase enzymes may be codon optimized for optimal production in the host organism selected for expression.

[0073] As used herein, "preferred," "optimal," and "high codon usage bias" codons when used alone or in combination refer(s) 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., 14:372-73 [1998]; Stenico et al., Nucl. Acids Res., 222437-46 [1994]; and Wright, Gene 87:23-29 [1990]). Codon usage tables are available for many different organisms (See e.g., Wada et al., Nucl. Acids Res., 20:2111-2118 [1992]; Nakamura et al., Nucl. Acids Res., 28:292 [2000]; 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., 266:259-281 [1996]; and Tiwari et al., Comput. Appl. Biosci., 13:263-270 [1997]).

[0074] As used herein, "control sequence" includes 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 sequence, signal peptide sequence, initiation 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.

[0075] "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.

[0076] "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.

[0077] The phrase "suitable reaction conditions" refers to those conditions in the enzymatic conversion reaction solution (e.g., ranges of enzyme loading, substrate loading, temperature, pH, buffers, co-solvents, etc.) under which a pantothenate kinase polypeptide of the present invention is capable of converting a substrate to the desired product compound. Some exemplary "suitable reaction conditions" are provided herein.

[0078] As used herein, "loading," such as in "compound loading" or "enzyme loading" refers to the concentration or amount of a component in a reaction mixture at the start of the reaction.

[0079] As used herein, "substrate" in the context of an enzymatic conversion reaction process refers to the compound or molecule acted on by the engineered enzymes provided herein (e.g., engineered pantothenate kinase polypeptides).

[0080] As used herein, "increasing" yield of a product (e.g., a deoxyribose phosphate analog) from a reaction occurs when a particular component present during the reaction (e.g., a pantothenate kinase enzyme) causes more product to be produced, compared with a reaction conducted under the same conditions with the same substrate and other substituents, but in the absence of the component of interest.

[0081] A reaction is said to be "substantially free" of a particular enzyme if the amount of that enzyme compared with other enzymes that participate in catalyzing the reaction is less than about 2%, about 1%, or about 0.1% (wt/wt).

[0082] As used herein, "fractionating" a liquid (e.g., a culture broth) means applying a separation process (e.g., salt precipitation, column chromatography, size exclusion, and filtration) or a combination of such processes to provide a solution in which a desired protein comprises a greater percentage of total protein in the solution than in the initial liquid product.

[0083] As used herein, "starting composition" refers to any composition that comprises at least one substrate. In some embodiments, the starting composition comprises any suitable substrate.

[0084] As used herein, "product" in the context of an enzymatic conversion process refers to the compound or molecule resulting from the action of an enzymatic polypeptide on a substrate.

[0085] As used herein, "equilibration" as used herein refers to the process resulting in a steady state concentration of chemical species in a chemical or enzymatic reaction (e.g., interconversion of two species A and B), including interconversion of stereoisomers, as determined by the forward rate constant and the reverse rate constant of the chemical or enzymatic reaction.

[0086] As used herein, "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., (C1-C4)alkyl refers to an alkyl of 1 to 4 carbon atoms).

[0087] As used herein, "alkenyl" refers to 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.

[0088] As used herein, "alkynyl" refers to 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.

[0089] As used herein, "heteroalkyl, "heteroalkenyl," and heteroalkynyl," refer 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-, -NRα-, -PH-, -S(O)-, $-S(O)_2$, -S(O) NRα-, $-S(O)_2$NRα-, and the like, including combinations thereof, where each Rα is independently selected from hydrogen, alkyl, heteroalkyl, cycloalkyl, heterocycloalkyl, aryl, and heteroaryl.

[0090] As used herein, "alkoxy" refers to the group –ORβ wherein R β is an alkyl group is as defined above including optionally substituted alkyl groups as also defined herein.

[0091] As used herein, "aryl" refers to an unsaturated aromatic carbocyclic group of from 6 to 12 carbon atoms inclusively having a single ring (e.g., phenyl) or multiple condensed rings (e.g., naphthyl or anthryl). Exemplary aryls include phenyl, pyridyl, naphthyl and the like.

[0092] As used herein, "amino" refers to the group -NH_2. Substituted amino refers to the group –NHRδ, NRδRδ, and NRδRδRδ, where each Rδ 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.

[0093] As used herein, "oxo" refers to =O.

[0094] As used herein, "oxy" refers to a divalent group -O-, which may have various substituents to form different oxy groups, including ethers and esters.

[0095] As used herein, "carboxy" refers to -COOH.

[0096] As used herein, "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.

[0097] As used herein, "alkyloxycarbonyl" refers to -C(O)ORε, where Rε is an alkyl group as defined herein, which can be optionally substituted.

[0098] As used herein, "aminocarbonyl" refers to -C(O)NH_2. Substituted aminocarbonyl refers to –C(O)NRδRδ, where the amino group NRδRδ is as defined herein.

[0099] As used herein, "halogen" and "halo" refer to fluoro, chloro, bromo and iodo.

[0100] As used herein, "hydroxy" refers to -OH.

[0101] As used herein, "cyano" refers to -CN.

[0102] As used herein, "heteroaryl" refers to an aromatic heterocyclic group of from 1 to 10 carbon atoms inclusively and 1 to 4 heteroatoms inclusively selected from oxygen, nitrogen and sulfur within the ring. Such heteroaryl groups can have a single ring (e.g., pyridyl or furyl) or multiple condensed rings (e.g., indolizinyl or benzothienyl).

[0103] As used herein, "heteroarylalkyl" refers to an alkyl substituted with a heteroaryl (i.e., heteroaryl-alkyl- groups), preferably having from 1 to 6 carbon atoms inclusively in the alkyl moiety and from 5 to 12 ring atoms inclusively in the heteroaryl moiety. Such heteroarylalkyl groups are exemplified by pyridylmethyl and the like.

[0104] As used herein, "heteroarylalkenyl" refers to an alkenyl substituted with a heteroaryl (i.e., heteroaryl-alkenyl- groups), preferably having from 2 to 6 carbon atoms inclusively in the alkenyl moiety and from 5 to 12 ring atoms inclusively in the heteroaryl moiety.

[0105] As used herein, "heteroarylalkynyl" refers to an alkynyl substituted with a heteroaryl (i.e., heteroaryl-alkynyl- groups), preferably having from 2 to 6 carbon atoms inclusively in the alkynyl moiety and from 5 to 12 ring atoms inclusively in the heteroaryl moiety.

[0106] As used herein, "heterocycle," "heterocyclic," and interchangeably "heterocycloalkyl," refer to a saturated or unsaturated group having a single ring or multiple condensed rings, from 2 to 10 carbon ring atoms inclusively and from 1 to 4 hetero ring atoms inclusively selected from nitrogen, sulfur or oxygen within the ring. Such heterocyclic groups can have a single ring (e.g., piperidinyl or tetrahydrofuryl) or multiple condensed rings (e.g., indolinyl, dihydrobenzofuran or quinuclidinyl). Examples of heterocycles include, but are not limited to, furan, thiophene, thiazole, oxazole, pyrrole, imidazole, pyrazole, pyridine, pyrazine, pyrimidine, pyridazine, indolizine, isoindole, indole, indazole, purine, quinolizine, isoquinoline, quinoline, phthalazine, naphthylpyridine, quinoxaline, quinazoline, cinnoline, pteridine, carbazole, carboline, phenanthridine, acridine, phenanthroline, isothiazole, phenazine, isoxazole, phenoxazine, phenothiazine, imidazolidine, imidazoline, piperidine, piperazine, pyrrolidine, indoline and the like.

[0107] As used herein, "membered ring" is meant to embrace any cyclic structure. The number preceding the term "membered" denotes the number of skeletal atoms that constitute the ring. Thus, for example, cyclohexyl, pyridine, pyran and thiopyran are 6-membered rings and cyclopentyl, pyrrole, furan, and thiophene are 5-membered rings.

[0108] Unless otherwise specified, positions occupied by hydrogen in the foregoing groups can be further substituted with substituents 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; and preferred heteroatoms are oxygen, nitrogen, and sulfur. It is understood that where open valences exist on these substituents 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 understood 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.

[0109] As used herein the term "culturing" refers to the growing of a population of microbial cells under any suitable conditions (e.g., using a liquid, gel or solid medium).

[0110] Recombinant polypeptides can be produced using any suitable methods known in the art. Genes encoding the wild-type polypeptide of interest can be cloned in vectors, such as plasmids, and expressed in desired hosts, such as E. coli, etc. Variants of recombinant polypeptides can be generated by various methods known in the art. Indeed, there is a wide variety of different mutagenesis techniques well known to those skilled in the art. In addition, mutagenesis kits are also available from many commercial molecular biology suppliers. Methods are available to make specific substitutions at defined amino acids (site-directed), specific or random mutations in a localized region of the gene (regio-specific), or random mutagenesis over the entire gene (e.g., saturation mutagenesis). Numerous suitable methods are known to those in the art to generate enzyme variants, including but not limited to site-directed mutagenesis of single-stranded DNA or double-stranded DNA using PCR, cassette mutagenesis, gene synthesis, error-prone PCR, shuffling, and chemical saturation mutagenesis, or any other suitable method known in the art. Mutagenesis and directed evolution methods can be readily applied to enzyme-encoding polynucleotides to generate variant libraries that can be expressed, screened, and assayed. Any suitable mutagenesis and directed evolution methods find use in the present invention and are well 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., 254(2):157-78 [1997]; Dale et al., Meth. Mol. Biol., 57:369-74 [1996]; Smith, Ann. Rev. Genet., 19:423-462 [1985]; Botstein et al., Science, 229:1193-1201 [1985]; Carter, Biochem. J., 237:1-7 [1986]; Kramer et al., Cell, 38:879-887 [1984]; Wells et al., Gene, 34:315-323 [1985]; Minshull et al.,

Curr. Op. Chem. Biol., 3:284-290 [1999]; Christians et al., Nat. Biotechnol., 17:259-264 [1999]; Crameri et al., Nature, 391:288-291 [1998]; Crameri, et al., Nat. Biotechnol., 15:436-438 [1997]; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 94:4504-4509 [1997]; Crameri et al., Nat. Biotechnol., 14:315-319 [1996]; Stemmer, Nature, 370:389-391 [1994]; Stemmer, Proc. Nat. Acad. Sci. USA, 91:10747-10751 [1994]; 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).

[0111] In some embodiments, the enzyme clones obtained following mutagenesis treatment are screened by subjecting the enzyme preparations to a defined temperature (or other assay conditions) and measuring the amount of enzyme activity remaining after heat treatments or other suitable assay conditions. Clones containing a polynucleotide encoding a polypeptide are then isolated from the gene, 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).

[0112] After the variants are produced, they can be screened for any desired property (e.g., high or increased activity, or low or reduced activity, increased thermal activity, increased thermal stability, and/or acidic pH stability, etc.). In some embodiments, "recombinant pantothenate kinase polypeptides" (also referred to herein as "engineered pantothenate kinase polypeptides," "variant pantothenate kinase enzymes," "pantothenate kinase variants," and "pantothenate kinase combinatorial variants") find use. In some embodiments, "recombinant pantothenate kinase polypeptides" (also referred to as "engineered pantothenate kinase polypeptides," "variant pantothenate kinase enzymes," "pantothenate kinase variants," and "pantothenate kinase combinatorial variants") find use.

[0113] As used herein, a "vector" is a DNA construct for introducing a DNA sequence into a cell. In some embodiments, the vector is an expression vector that is operably linked to a suitable control sequence capable of effecting the expression in a suitable host of the polypeptide encoded in the DNA sequence. In some embodiments, an "expression vector" has a promoter sequence operably linked to the DNA sequence (e.g., transgene) to drive expression in a host cell, and in some embodiments, also comprises a transcription terminator sequence.

[0114] As used herein, the term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, and post-translational modification. In some embodiments, the term also encompasses secretion of the polypeptide from a cell.

[0115] As used herein, the term "produces" refers to the production of proteins and/or other compounds by cells. It is intended that the term encompass any step involved in the production of polypeptides including, but not limited to, transcription, post-transcriptional modification, translation, and post-translational modification. In some embodiments, the term also encompasses secretion of the polypeptide from a cell.

[0116] As used herein, an amino acid or nucleotide sequence (e.g., a promoter sequence, signal peptide, terminator sequence, etc.) is "heterologous" to another sequence with which it is operably linked if the two sequences are not associated in nature. For example a "heterologous polynucleotide" is 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.

[0117] As used herein, the terms "host cell" and "host strain" refer to suitable hosts for expression vectors comprising DNA provided herein (e.g., the polynucleotides encoding the pantothenate kinase variants). In some embodiments, the host cells are prokaryotic or eukaryotic cells that have been transformed or transfected with vectors constructed using recombinant DNA techniques as known in the art.

[0118] The term "analogue" means a polypeptide having more than 70% sequence identity but less than 100% sequence identity (e.g., more than 75%, 78%, 80%, 83%, 85%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity) with a reference polypeptide. In some embodiments, analogues means polypeptides that contain one or more non-naturally occurring amino acid residues including, but not limited, to homoarginine, ornithine and norvaline, as well as naturally occurring amino acids. In some embodiments, analogues also include one or more D-amino acid residues and non-peptide linkages between two or more amino acid residues.

[0119] The term "effective amount" means an amount sufficient to produce the desired result. One of general skill in the art may determine what the effective amount by using routine experimentation.

[0120] The terms "isolated" and "purified" are used to refer to a molecule (e.g., an isolated nucleic acid, polypeptide, etc.) or other component that is removed from at least one other component with which it is naturally associated. The term "purified" does not require absolute purity, rather it is intended as a relative definition.

[0121] As used herein, "stereoselectivity" refers to the preferential formation in a chemical or enzymatic reaction of one stereoisomer over another. Stereoselectivity can be partial, where the formation of one stereoisomer is favored over the other, or it may be complete where only one stereoisomer is formed. When the stereoisomers are enantiomers, the stereoselectivity is referred to as enantioselectivity, the fraction (typically reported as a percentage) of one enantiomer in the sum of both. It is commonly alternatively reported in the art (typically as a percentage) as the enantiomeric excess ("e.e.") calculated therefrom according to the formula [major enantiomer − minor enantiomer]/[major enantiomer + minor enantiomer]. Where the stereoisomers are diastereoisomers, the stereoselectivity is referred to as diastereoselectivity, the fraction (typically reported as a percentage) of one diastereomer in a mixture of two diastereomers, commonly alternatively reported as the diastereomeric excess ("d.e."). Enantiomeric excess and diastereomeric excess are types of stereomeric excess.

[0122] As used herein, “regioselectivity” and “regioselective reaction” refer to a reaction in which one direction of bond making or breaking occurs preferentially over all other possible directions. Reactions can completely (100%) regioselective if the discrimination is complete, substantially regioselective (at least 75%), or partially regioselective (x%, wherein the percentage is set dependent upon the reaction of interest), if the product of reaction at one site predominates over the product of reaction at other sites.

[0123] As used herein, “chemoselectivity” refers to the preferential formation in a chemical or enzymatic reaction of one product over another.

[0124] As used herein, “pH stable” refers to a pantothenate kinase polypeptide that maintains similar activity (e.g., more than 60% to 80%) after exposure to high or low pH (e.g., 4.5-6 or 8 to 12) for a period of time (e.g., 0.5-24 hrs) compared to the untreated enzyme.

[0125] As used herein, “thermostable” refers to a pantothenate kinase 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 h) compared to the wild-type enzyme exposed to the same elevated temperature.

[0126] As used herein, “solvent stable” refers to a pantothenate kinase 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 h) compared to the wild-type enzyme exposed to the same concentration of the same solvent.

[0127] As used herein, “thermo- and solvent stable” refers to a pantothenate kinase polypeptide that is both thermostable and solvent stable.

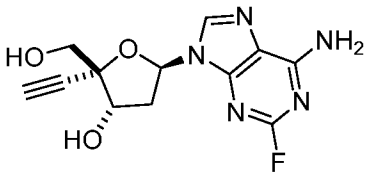
[0128] As used herein, “optional” and “optionally” mean that the subsequently described event or circumstance may or may not occur, and that the description includes instances where the event or circumstance occurs and instances in which it does not. One of ordinary skill in the art would understand that with respect to any molecule described as containing one or more optional substituents, only sterically practical and/or synthetically feasible compounds are meant to be included.

[0129] As used herein, “optionally substituted” 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.

DETAILED DESCRIPTION OF THE INVENTION

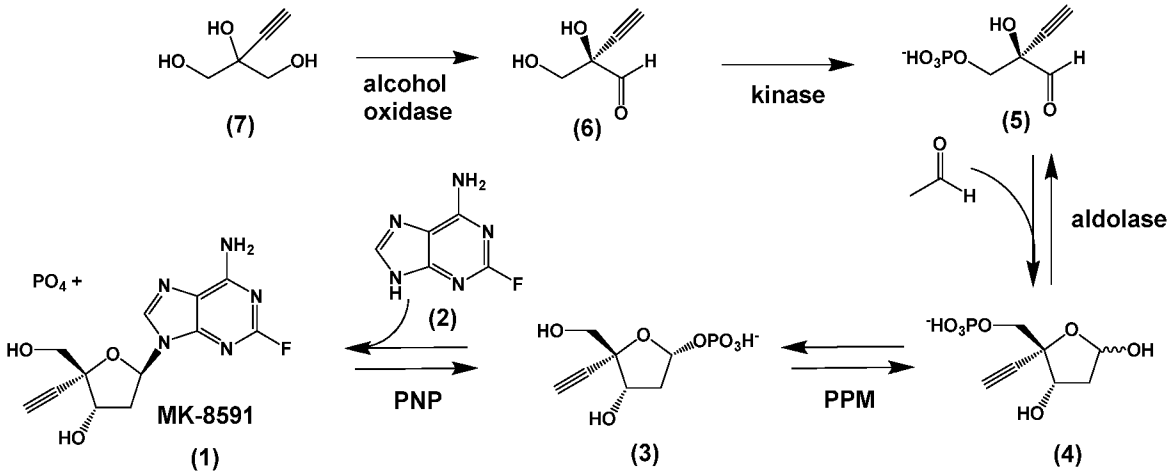
[0130] The present invention provides engineered pantothenate kinase (PanK) enzymes, polypeptides having PanK activity, and the polynucleotides encoding these enzymes, as well as the vectors and host cells comprising these polynucleotides and polypeptides. Methods for producing PanK enzymes are also provided. The present invention further provides compositions comprising the PanK enzymes and methods of using the engineered PanK enzymes. The present invention finds particular use in the production of pharmaceutical compounds.

[0131] In some embodiments, the present invention provides enzymes suitable for the production of phosphorylated glycerol derivatives and glyceraldehyde derivatives with bulky substituents on the C2 carbon of glycerol, especially phosphorylated ethynyl-glycerols and ethynyl- glyceraldehydes that are intermediates for the in vitro enzymatic synthesis of the non-natural nucleoside analog shown of compound (**1**).



Compound (1)

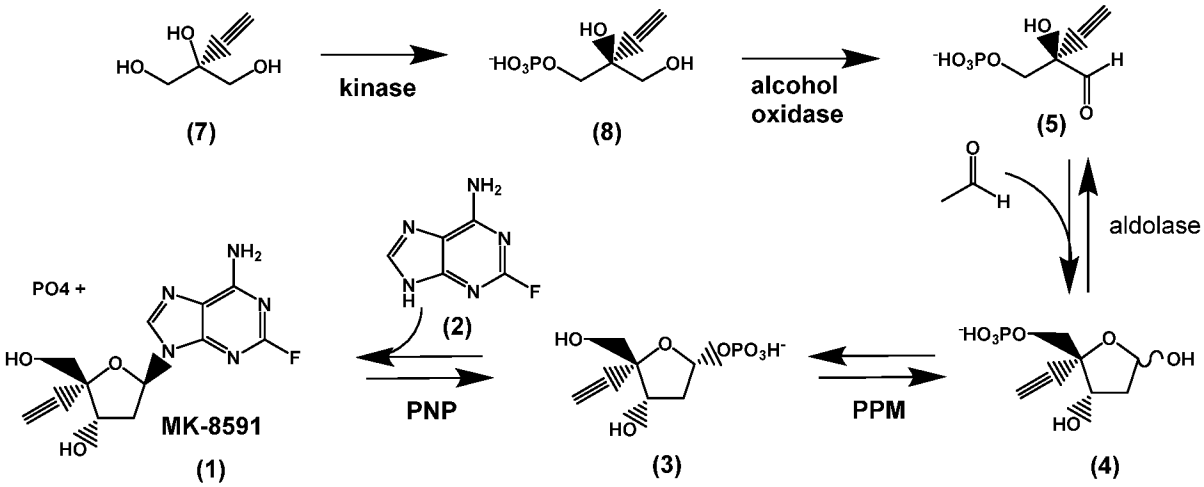
[0132] Production of phosphorylated glyceraldehyde derivatives such as compound (**5**), can be difficult. However, the corresponding non-phosphorylated glyceraldehyde derivatives (**6**) can be made by oxidizing the glycerol derivative (**7**) with an alcohol oxidase. Once the glycerol aldehyde is formed it can be phosphorylated into the desired intermediate (**5**) by PanK as shown in Scheme I.



Scheme I

[0133] Alternate methods of producing compound (**5**) may have advantages in the efficient production of compound (**1**) at industrial scale. PanK enzymes, including the improved enzymes of the present disclosure, may have improved phosphorylation activity on the triglycerol of compound

(7) as compared to the glyceraldehyde of compound (6). Thus, producing compound (5) from compound (7) by first phosphorylating compound (7) using a PanK enzyme to produce compound (8) and then oxidizing to compound (5) using an alcohol oxidase may improve production of compound (1) (see Scheme 2, below). Phosphorylation of compound (7) prior to oxidation by the alcohol oxidase may also allow coupling of the oxidase and aldose reactions may reduce product inhibition and increase enzyme efficiency.



Scheme 2

Engineered PanK Polypeptides

[0134] The present invention provides engineered PanK 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 also describes the polynucleotides encoding the polypeptides. In some embodiments, the present invention provides engineered, non-naturally occurring PanK enzymes with improved properties as compared to wild-type PanK enzymes. Any suitable reaction conditions find use in the present invention. In some embodiments, methods are used to analyze the improved properties of the engineered polypeptides to carry out the isomerization reaction. In some embodiments, the reaction conditions are modified with regard to concentrations or amounts of engineered PanK, substrate(s), buffer(s), solvent(s), co-factors, pH, conditions including temperature and reaction time, and/or conditions with the engineered PanK polypeptide immobilized on a solid support, as further described below and in the Examples.

[0135] In some embodiments, additional reaction components or additional techniques are utilized to supplement the reaction conditions. In some embodiments, these include taking measures to stabilize or prevent inactivation of the enzyme, reduce product inhibition, shift reaction equilibrium to desired product formation.

[0136] In some further embodiments, any of the above described process for the conversion of substrate compound to product compound can further comprise one or more steps selected from: extraction, isolation, purification, crystallization, filtration, and/or lyophilization of product compound(s). Methods, techniques, and protocols for extracting, isolating, purifying, and/or crystallizing the product(s) from biocatalytic reaction mixtures produced by the processes provided herein are known to the ordinary artisan and/or accessed through routine experimentation. Additionally, illustrative methods are provided in the Examples below.

Methods of Using the Engineered Pantothenate Kinase Enzymes

[0137] In some embodiments, the PanK enzymes described herein find use in processes for converting compound (**7**) to compound (**8**). Generally, the process for performing the phosphorylation reaction comprises contacting or incubating the substrate compound in presence of a co-substrate, such as adenylate triphosphate (ATP), with an acetate kinase enzyme and acetyl phosphate used to recycle the resulting adenylate triphosphate (ADP) back to ATP.

[0138] 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, reductant, divalent transition metal, pH, temperature, buffer, solvent system, 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 PanK polypeptide described herein can be readily optimized in view of the guidance provided herein by routine experimentation that includes, but is not limited to, contacting the engineered PanK polypeptide and substrate compound under experimental reaction conditions of concentration, pH, temperature, and solvent conditions, and detecting the product compound.

[0139] Suitable reaction conditions using the engineered PanK polypeptides typically comprise a co-substrate, which is used stoichiometrically in the phosphorylation reaction. Generally, the co-substrate for PanK enzymes is ATP. Other phosphoryl group donors that are capable of serving as co-substrates for PanK enzymes can be used. Because the co-substrate is used stoichiometrically, the co-substrate is present at an equimolar or higher amount than that of the substrate compound (i.e., the molar concentration of co-substrate is equivalent to or higher than the molar concentration of substrate compound). In some embodiments, the suitable reaction conditions can comprise a co-substrate molar concentration of at least 1 fold, 1.5 fold, 2 fold, 3 fold 4 fold or 5 fold or more than the molar concentration of the substrate compound. In some embodiments, the suitable reaction conditions can comprise a co-substrate concentration, particularly ATP, of about 0.0005 M to about 2 M, 0.01 M to about 2 M, 0.1 M to about 2 M, 0.2 M to about 2 M, about 0.5 M to about 2 M, or about 1 M to about 2 M. In some embodiments, the reaction conditions comprise a co-substrate concentration of about 0.0001 M, 0.001 M, 0.01 M, 0.1 M, 0.2 M, 0.3 M, 0.4 M, 0.5 M, 0.6 M, 0.7 M, 0.8 M, 1 M, 1.5 M, or 2 M. In some embodiments, additional co-substrate can be added during the reaction. In some

additional embodiments, the co-substrate is present in lower concentrations due to the presence of a recycling system, such as acetate kinase and acetyl phosphate, that recycles ADP to ATP.

[0140] Substrate compound in the reaction mixtures can be varied, taking into consideration, for example, the desired amount of product compound, the effect of substrate concentration on enzyme activity, stability of enzyme under reaction conditions, and the percent conversion of substrate to product. In some embodiments, the suitable reaction conditions comprise a substrate compound loading of at least about 0.5 to about 200 g/L, 1 to about 200 g/L, 5 to about 150 g/L, about 10 to about 100 g/L, 20 to about 100 g/L or about 50 to about 100 g/L. In some embodiments, the suitable reaction conditions comprise a substrate compound loading 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, at least about 30 g/L, at least about 50 g/L, at least about 75 g/L, at least about 100 g/L, at least about 150 g/L or at least about 200 g/L, or even greater. The values for substrate loadings provided herein are based on the molecular weight of 2-ethynylglycerol; however, it also contemplated that the equivalent molar amounts of various alcohol or aldehyde analogues also can be used in the process.

[0141] In carrying out the PanK mediated processes described herein, the engineered polypeptide 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 PanK 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).

[0142] The gene(s) encoding the engineered PanK polypeptides can be transformed into host cells 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 PanK polypeptide and another set can be transformed with gene(s) encoding another engineered PanK polypeptide. 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 PanK 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 PanK reaction.

[0143] In some embodiments, the improved activity and/or selectivity of the engineered PanK 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 0.03% (w/w), 0.05 % (w/w), 0.1 % (w/w), 0.15 % (w/w), 0.2 % (w/w), 0.3 % (w/w), 0.4 % (w/w), 0.5 % (w/w), 1 % (w/w), 2% (w/w), 5% (w/w), 10% (w/w), 20% (w/w) or more of substrate compound loading.

[0144] In some embodiments, the engineered polypeptide is present at about 0.01 g/L to about 15 g/L; about 0.05 g/L to about 15 g/L; about 0.1 g/L to about 10 g/L; about 1 g/L to about 8 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 PanK 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 g/L, 2 g/L, 5 g/L, 10 g/L, or 15 g/L.

[0145] In some embodiments, the reaction conditions also comprise a divalent metal capable of serving as a cofactor in the reaction. Generally, the divalent metal co-factor is magnesium (i.e., Mg^{+2}). The magnesium ion may be provided in various forms, such as magnesium chloride ($MgCl_2$). While magnesium ion is the metal co-factor found in the naturally occurring PanK enzyme and functions efficiently in the engineered enzymes, it is to be understood that other divalent metals capable of acting as a co-factor can be used in the processes. In some embodiments, the reaction conditions can comprises a divalent metal cofactor, particularly Mg^{+2}, at a concentration of about 1 mM to 1 M , 1 mM to 100 mM, 1 mM to about 50 mM, 25 mM to about 35 mM, about 30 mM to about 60 mM or about 55 mM to about 65 mM. In some embodiments, the reaction conditions comprise a divalent metal co-factor concentration of about 1 mM, 10 mM, 20 mM, 30 mM, 40 mM, 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, or 100 mM.

[0146] 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, and 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), and the like. In some embodiments, the buffer is tris. In some embodiments of the process, the suitable reaction conditions comprise a buffer (e.g., tris) concentration of from about 0.01 to about 0.4 M, 0.05 to about 0.4 M, 0.1 to about 0.3 M, or about 0.1 to about 0.2 M. In some embodiments, the reaction condition comprises a buffer (e.g., tris) concentration of about 0.01, 0.02, 0.03, 0.04, 0.05, 0.07, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.3, or 0.4 M.

[0147] In some embodiments, the reaction condition comprises a wet organic solvent. Suitable wet organic solvents are known in the art and include, by way of example and not limitation, wet isopropyl alcohol, wet toluene, and wet methyl tertiary butyl ether.

[0148] In the embodiments of the process, the reaction conditions can 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.

[0149] In the embodiments of the processes herein, a suitable temperature can be 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 60°C, about 10°C to about 55°C, about 15°C to about 60°C, about 20°C to about 60°C, about 20°C to about 55°C, about 25°C to about 55°C, or about 30°C to about 50°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, or 60°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.

[0150] In some embodiments, the reaction conditions can 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, 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.

[0151] In some embodiments, the reaction conditions can 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.

[0152] The quantities of reactants used in the kinase reaction will generally vary depending on the quantities of product desired, and concomitantly the amount of PanK substrate employed. Those having ordinary skill in the art will readily understand how to vary these quantities to tailor them to the desired level of productivity and scale of production.

[0153] 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, PanK enzyme, and substrate may be added first to the solvent.

[0154] 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.

[0155] For improved mixing efficiency when an aqueous co-solvent system is used, the PanK enzyme, and cofactor may be added and mixed into the aqueous phase first. The organic phase may then be added and mixed in, followed by addition of the PanK enzyme substrate and co-substrate. Alternatively, the PanK enzyme substrate may be premixed in the organic phase, prior to addition to the aqueous phase.

[0156] The phosphorylation process is generally allowed to proceed until further conversion of substrate to phosphorylated 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.

[0157] In some embodiments of the process, the suitable reaction conditions comprise a substrate loading of at least about 5 g/L, 10 g/L, 20 g/L, 30 g/L, 40 g/L, 50 g/L, 60 g/L, 70 g/L, 100 g/L, or more, and wherein the method results in at least about 50%, 60%, 70%, 80%, 90%, 95% or greater conversion of substrate compound to product compound in about 48 h or less, in about 36 h or less, in about 24 h or less, or in about 3 h or less.

[0158] In further embodiments of the processes for converting substrate compound to product compound using the engineered PanK polypeptides, the suitable reaction conditions can comprise an initial substrate loading to the reaction solution which is then contacted by the polypeptide. This

reaction solution is then further supplemented with additional substrate compound as a continuous or batchwise addition over time at a rate of at least about 1 g/L/h, at least about 2 g/L/h, at least about 4 g/L/h, at least about 6 g/L/h, or higher. Thus, according to these suitable reaction conditions, polypeptide is added to a solution having an initial substrate loading of at least about 20 g/L, 30 g/L, or 40 g/L. This addition of polypeptide is then followed by continuous addition of further substrate to the solution at a rate of about 2 g/L/h, 4 g/L/h, or 6 g/L/h until a much higher final substrate loading of at least about 30 g/L, 40 g/L, 50 g/L, 60 g/L, 70 g/L, 100 g/L, 150 g/L, 200 g/L or more, is reached. Accordingly, in some embodiments of the process, the suitable reaction conditions comprise addition of the polypeptide to a solution having an initial substrate loading of at least about 20 g/L, 30 g/L, or 40 g/L followed by addition of further substrate to the solution at a rate of about 2 g/L/h, 4 g/L/h, or 6 g/L/h until a final substrate loading of at least about 30 g/L, 40 g/L, 50 g/L, 60 g/L, 70 g/L, 100 g/L or more, is reached. This substrate supplementation reaction condition allows for higher substrate loadings to be achieved while maintaining high rates of conversion of substrate to phosphorylated product of at least about 50%, 60%, 70%, 80%, 90% or greater conversion of substrate.

[0159] In some embodiments, acetate kinase and acetyl phosphate recycle ADP to ATP. In some embodiments, acetate kinase and acetyl phosphate recycle an ADP analogue to an ATP analogue.

[0160] In some embodiments of the processes, the reaction using an engineered PanK polypeptide can comprise the following suitable reaction conditions: (a) substrate loading at about 50 g/L; (b) about 0.5 g/L of the engineered polypeptide; (c) 1.25 eq acetyl phosphate; (d) about 0.1 mol% ATP ; (e) about 10 mM MgCl2; (f) about 0.125 mg/mL acetate kinase; (g) a pH of about 8.0; (h) temperature of about 25 °C; and (i) reaction time of about 18 hrs.

[0161] In some embodiments, additional reaction components or additional techniques are 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 product formation.

[0162] In further embodiments, any of the above described process for the conversion of substrate compound to product compound can further comprise one or more steps selected from: extraction; isolation; purification; and crystallization of product compound. 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.

[0163] Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

Engineered PanK Polynucleotides Encoding Engineered Polypeptides, Expression Vectors and Host Cells

[0164] The present invention provides polynucleotides encoding the engineered enzyme polypeptides described herein. In some embodiments, the polynucleotides are operatively linked to one or more

heterologous regulatory sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. In some embodiments, expression constructs containing at least one heterologous polynucleotide encoding the engineered enzyme polypeptide(s) are introduced into appropriate host cells to express the corresponding enzyme polypeptide(s).

[0165] 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 an engineered enzyme (e.g., PanK) polypeptide. Thus, the present invention provides methods and compositions for the production of each and every possible variation of enzyme polynucleotides that could be made that encode the enzyme 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 the Examples (e.g., in the various Tables).

[0166] In some embodiments, the codons are preferably optimized for utilization by the chosen host cell for protein production. For example, preferred codons used in bacteria are typically used for expression in bacteria. Consequently, codon optimized polynucleotides encoding the engineered enzyme polypeptides contain preferred codons at about 40%, 50%, 60%, 70%, 80%, 90%, or greater than 90% of the codon positions in the full length coding region.

[0167] In some embodiments, the enzyme polynucleotide encodes an engineered polypeptide having enzyme activity with the properties disclosed herein, wherein the polypeptide comprises an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to a reference sequence selected from the SEQ ID NOS provided herein, or the amino acid sequence of any variant (e.g., those provided in the Examples), and one or more residue differences as compared to the reference polynucleotide(s), or the amino acid sequence of any variant as disclosed in the Examples (for example 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more amino acid residue positions). In some embodiments, the reference polypeptide sequence comprises SEQ ID NO: 4, while in some other embodiments, the reference polypeptide sequence comprises SEQ ID NO: 44, SEQ ID NO: 320, and/or SEQ ID NO: 526.

[0168] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from any polynucleotide sequence provided herein, or a complement thereof, or a polynucleotide sequence encoding any of the variant enzyme polypeptides provided herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an enzyme polypeptide comprising an amino acid sequence that has one or more residue differences as compared to a reference sequence.

[0169] In some embodiments, an isolated polynucleotide encoding any of the engineered enzyme polypeptides herein is manipulated in a variety of ways to facilitate expression of the enzyme

polypeptide. In some embodiments, the polynucleotides encoding the enzyme polypeptides comprise expression vectors where one or more control sequences is present to regulate the expression of the enzyme 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 utilized. Techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art. In some embodiments, the control sequences include among others, promoters, leader sequences, polyadenylation sequences, propeptide sequences, signal peptide sequences, and transcription terminators. In some embodiments, suitable promoters are selected based on the host cells selection. For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present disclosure, include, but are not limited to 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, but are not limited to 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 8:423-488 [1992]).

[0170] In some embodiments, the control sequence is also a suitable transcription terminator sequence (i.e., a sequence recognized by a host cell to terminate transcription). In some embodiments, the terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the enzyme polypeptide. Any suitable terminator which is functional in the host cell of choice finds use in the present invention. 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).

[0171] In some embodiments, the control sequence is also a suitable leader sequence (i.e., a non-translated region of an mRNA that is important for translation by the host cell). In some embodiments, the leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the enzyme polypeptide. Any suitable leader sequence that is functional in the host cell of choice find use in the present invention. 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 are 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).

[0172] In some embodiments, the control sequence is also a polyadenylation sequence (i.e., 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 suitable polyadenylation sequence which is functional in the host cell of choice finds use in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells include, but are not limited to 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 known (See e.g., Guo and Sherman, Mol. Cell. Biol., 15:5983-5990 [1995]).

[0173] In some embodiments, the control sequence is also a signal peptide (i.e., a 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). In some embodiments, the 5' end of the coding sequence of the nucleic acid sequence inherently contains a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, in some embodiments, the 5' end of the coding sequence contains a signal peptide coding region that is foreign to the coding sequence. Any suitable signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice finds use for expression of the engineered polypeptide(s). Effective signal peptide coding regions for bacterial host cells are the signal peptide coding regions include, but are not limited to those 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., 57:109-137 [1993]). In some embodiments, 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.

[0174] In some embodiments, the control sequence is also 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." 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 may be obtained from any suitable source, including, but 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.

[0175] 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.

[0176] In another aspect, the present invention is directed to a recombinant expression vector comprising a polynucleotide encoding an engineered enzyme 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 herein are joined together to produce recombinant expression vectors which include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the enzyme polypeptide at such sites. Alternatively, in some embodiments, the nucleic acid sequence of the present invention is expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In some embodiments involving the creation of 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.

[0177] The recombinant expression vector may be any suitable vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and bring about the expression of the enzyme polynucleotide sequence. The choice of the vector typically depends 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.

[0178] 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 is one in which, when introduced into the host cell, it is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, in some embodiments, 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, and/or a transposon is utilized.

[0179] In some embodiments, the expression vector 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 auxotrophs, 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 filamentous fungal host cells include, but are not limited to, amdS (acetamidase; e.g., from *A. nidulans* or *A. orzyae*), argB (ornithine carbamoyltransferases), bar (phosphinothricin acetyltransferase; e.g., from S. *hygroscopicus*), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate decarboxylase; e.g., from *A. nidulans* or *A. orzyae*), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof.

[0180] In another aspect, the present invention provides a host cell comprising at least one polynucleotide encoding at least one engineered enzyme polypeptide of the present invention, the polynucleotide(s) being operatively linked to one or more control sequences for expression of the engineered enzyme enzyme(s) in the host cell. Host cells suitable 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* or *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 also include various *Escherichia coli* strains (e.g., W3110 (ΔfhuA) and BL21). 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, and or tetracycline resistance.

[0181] In some embodiments, the expression vectors of the present invention contain an element(s) that permits integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome. In some embodiments involving integration into the host cell genome, the vectors rely on the nucleic acid sequence encoding the polypeptide or any other element of the vector for integration of the vector into the genome by homologous or nonhomologous recombination.

[0182] In some alternative embodiments, the expression vectors contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements preferably contain a sufficient number of nucleotides, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

[0183] For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are P15A ori or the origins of replication of plasmids pBR322, pUC19, pACYCl77 (which plasmid has the P15A ori), or pACYC184 permitting replication in *E. coli*, and pUB110, pE194, or pTA1060 permitting replication in *Bacillus*. Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. The origin of replication may be one having a mutation which makes its functioning temperature-sensitive in the host cell (See e.g., Ehrlich, Proc. Natl. Acad. Sci. USA 75:1433 [1978]).

[0184] In some embodiments, more than one copy of a nucleic acid sequence of the present invention is inserted into the host cell to increase production of the gene product. An increase in the copy number of the nucleic acid sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the nucleic acid sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the nucleic acid sequence, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

[0185] Many of the expression vectors for use in the present invention are commercially available. Suitable commercial expression vectors include, but are not limited to the p3xFLAGTM™ expression

vectors (Sigma-Aldrich Chemicals), which include a CMV promoter and hGH polyadenylation site for expression in mammalian host cells and a pBR322 origin of replication and ampicillin resistance markers for amplification in *E. coli*. Other suitable expression vectors include, but are not limited to pBluescriptII SK(-) and pBK-CMV (Stratagene), and plasmids derived from pBR322 (Gibco BRL), pUC (Gibco BRL), pREP4, pCEP4 (Invitrogen) or pPoly (See e.g., Lathe et al., Gene 57:193-201 [1987]).

[0186] Thus, in some embodiments, a vector comprising a sequence encoding at least one variant pantothenate kinase is transformed into a host cell in order to allow propagation of the vector and expression of the variant pantothenate kinase(s). In some embodiments, the variant pantothenate kinases are post-translationally modified to remove the signal peptide, and in some cases, may be cleaved after secretion. In some embodiments, the transformed host cell described above is cultured in a suitable nutrient medium under conditions permitting the expression of the variant pantothenate kinase(s). Any suitable medium useful for culturing the host cells finds use in the present invention, including, but not limited to minimal or complex media containing appropriate supplements. In some embodiments, host cells are grown in HTP media. Suitable media are available from various commercial suppliers or may be prepared according to published recipes (e.g., in catalogues of the American Type Culture Collection).

[0187] In another aspect, the present invention provides host cells comprising a polynucleotide encoding an improved pantothenate kinase polypeptide provided herein, the polynucleotide being operatively linked to one or more control sequences for expression of the pantothenate kinase enzyme in the host cell. Host cells for use in expressing the pantothenate kinase 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*, *Bacillus megaterium*, *Lactobacillus kefir*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* or *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. Appropriate culture media and growth conditions for the above-described host cells are well known in the art.

[0188] Polynucleotides for expression of the pantothenate kinase 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. Various methods for introducing polynucleotides into cells are known to those skilled in the art.

[0189] In some embodiments, the host cell is a eukaryotic cell. Suitable eukaryotic host cells include, but are not limited to, fungal cells, algal cells, insect cells, and plant cells. Suitable fungal host cells include, but are not limited to, Ascomycota, Basidiomycota, Deuteromycota, Zygomycota, Fungi imperfecti. In some embodiments, the fungal host cells are yeast cells and filamentous fungal cells.

The filamentous fungal host cells of the present invention include all filamentous forms of the subdivision Eumycotina and Oomycota. Filamentous fungi are characterized by a vegetative mycelium with a cell wall composed of chitin, cellulose and other complex polysaccharides. The filamentous fungal host cells of the present invention are morphologically distinct from yeast.

[0190] In some embodiments of the present invention, the filamentous fungal host cells are of any suitable genus and species, including, but not limited to *Achlya*, *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis, Cephalosporium*, *Chrysosporium*, *Cochliobolus*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Coprinus*, *Coriolus*, *Diplodia*, *Endothis*, *Fusarium*, *Gibberella*, *Gliocladium*, *Humicola*, *Hypocrea*, *Myceliophthora*, *Mucor*, *Neurospora*, *Penicillium*, *Podospora*, *Phlebia*, *Piromyces*, *Pyricularia*, *Rhizomucor*, *Rhizopus*, *Schizophyllum*, *Scytalidium*, *Sporotrichum, Talaromyces*, *Thermoascus*, *Thielavia*, *Trametes*, *Tolypocladium*, *Trichoderma*, *Verticillium*, and/or *Volvariella*, and/or teleomorphs, or anamorphs, and synonyms, basionyms, or taxonomic equivalents thereof.

[0191] In some embodiments of the present invention, the host cell is a yeast cell, including but not limited to cells of *Candida*, *Hansenula*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Kluyveromyces*, or *Yarrowia* species. In some embodiments of the present invention, the yeast cell is *Hansenula polymorpha*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Saccharomyces diastaticus*, *Saccharomyces norbensis*, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia kodamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria, Pichia quercuum, Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia angusta*, *Kluyveromyces lactis*, *Candida albicans*, or *Yarrowia lipolytica*.

[0192] In some embodiments of the invention, the host cell is an algal cell such as *Chlamydomonas* (e.g., *C. reinhardtii*) and *Phormidium* (P. sp. ATCC29409).

[0193] In some other embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells include, but are not limited to Gram-positive, Gram-negative and Gram-variable bacterial cells. Any suitable bacterial organism finds use in the present invention, including but not limited to *Agrobacterium*, *Alicyclobacillus*, *Anabaena*, *Anacystis, Acinetobacter, Acidothermus, Arthrobacter, Azobacter, Bacillus, Bifidobacterium, Brevibacterium, Butyrivibrio, Buchnera, Campestris, Camplyobacter, Clostridium, Corynebacterium, Chromatium, Coprococcus, Escherichia, Enterococcus, Enterobacter, Erwinia, Fusobacterium, Faecalibacterium, Francisella, Flavobacterium, Geobacillus, Haemophilus, Helicobacter, Klebsiella, Lactobacillus, Lactococcus, Ilyobacter, Micrococcus, Microbacterium, Mesorhizobium, Methylobacterium, Methylobacterium, Mycobacterium, Neisseria, Pantoea, Pseudomonas, Prochlorococcus, Rhodobacter, Rhodopseudomonas, Rhodopseudomonas, Roseburia, Rhodospirillum, Rhodococcus, Scenedesmus, Streptomyces, Streptococcus, Synecoccus, Saccharomonospora, Staphylococcus, Serratia, Salmonella, Shigella, Thermoanaerobacterium, Tropheryma, Tularensis, Temecula, Thermosynechococcus, Thermococcus, Ureaplasma, Xanthomonas, Xylella, Yersinia* and *Zymomonas*.

In some embodiments, the host cell is a species of *Agrobacterium*, *Acinetobacter*, *Azobacter*, *Bacillus*, *Bifidobacterium*, *Buchnera*, *Geobacillus*, *Campylobacter*, *Clostridium*, *Corynebacterium*, *Escherichia*, *Enterococcus*, *Erwinia*, *Flavobacterium*, *Lactobacillus*, *Lactococcus*, *Pantoea*, *Pseudomonas*, *Staphylococcus*, *Salmonella*, *Streptococcus*, *Streptomyces*, or *Zymomonas*. In some embodiments, the bacterial host strain is non-pathogenic to humans. In some embodiments the bacterial host strain is an industrial strain. Numerous bacterial industrial strains are known and suitable in the present invention. In some embodiments of the present invention, the bacterial host cell is an *Agrobacterium* species (e.g., *A. radiobacter*, *A. rhizogenes*, and *A. rubi*). In some embodiments of the present invention, the bacterial host cell is an *Arthrobacter* species (e.g., *A. aurescens*, *A. citreus*, *A. globiformis*, *A. hydrocarboglutamicus*, *A. mysorens*, *A. nicotianae*, *A. paraffineus*, *A. protophonniae*, *A. roseoparqffinus*, *A. sulfureus*, and *A. ureafaciens)*. In some embodiments of the present invention, the bacterial host cell is a *Bacillus* species (e.g., *B. thuringensis*, *B. anthracis*, *B. megaterium*, *B. subtilis*, *B. lentus*, *B. circulans*, *B. pumilus*, *B. lautus*, *B.coagulans, B. brevis, B. firmus, B. alkaophius*, *B. licheniformis*, *B. clausii*, *B. stearothermophilus*, *B. halodurans,* and *B. amyloliquefaciens*). In some embodiments, the host cell is an industrial *Bacillus* strain including but not limited to *B. subtilis*, *B. pumilus*, *B. licheniformis*, *B. megaterium, B. clausii*, *B. stearothermophilus,* or *B. amyloliquefaciens.* In some embodiments, the *Bacillus* host cells are *B. subtilis, B. licheniformis*, *B. megaterium, B. stearothermophilus,* and/or *B. amyloliquefaciens.* In some embodiments, the bacterial host cell is a *Clostridium* species *(*e.g., *C. acetobutylicum*, *C. tetani* E88, *C. lituseburense*, *C. saccharobutylicum*, *C. perfringens*, and *C. beijerinckii*). In some embodiments, the bacterial host cell is a *Corynebacterium* species (e.g., *C. glutamicum* and *C. acetoacidophilum*). In some embodiments the bacterial host cell is an *Escherichia* species (e.g., *E. coli*). In some embodiments, the host cell is *Escherichia coli* W3110. In some embodiments, the bacterial host cell is an *Erwinia* species (e.g., *E. uredovora*, *E. carotovora*, *E. ananas*, *E. herbicola*, *E. punctata*, and *E. terreus*). In some embodiments, the bacterial host cell is a *Pantoea* species (e.g., *P. citrea,* and *P. agglomerans).* In some embodiments the bacterial host cell is a *Pseudomonas* species (e.g., *P. putida*, *P. aeruginosa*, *P. mevalonii*, and *P. sp.* D-0l 10). In some embodiments, the bacterial host cell is a *Streptococcus* species (e.g., *S. equisimiles*, *S. pyogenes*, and *S. uberis*). In some embodiments, the bacterial host cell is a *Streptomyces* species (e.g., *S. ambofaciens*, *S. achromogenes*, *S. avermitilis*, *S. coelicolor*, *S. aureofaciens, S. aureus*, *S. fungicidicus, S. griseus*, and *S. lividans*). In some embodiments, the bacterial host cell is a *Zymomonas* species (e.g., *Z. mobilis*, and *Z. lipolytica*).

[0194] Many prokaryotic and eukaryotic strains that find use in the present invention are readily available to the public from a number of culture collections such as American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

[0195] In some embodiments, host cells are genetically modified to have characteristics that improve protein secretion, protein stability and/or other properties desirable for expression and/or secretion of a protein. Genetic modification can be achieved by genetic engineering techniques and/or classical microbiological techniques (e.g., chemical or UV mutagenesis and subsequent selection). Indeed, in some embodiments, combinations of recombinant modification and classical selection techniques are used to produce the host cells. Using recombinant technology, nucleic acid molecules can be introduced, deleted, inhibited or modified, in a manner that results in increased yields of pantothenate kinase variant(s) within the host cell and/or in the culture medium. For example, knockout of Alp1 function results in a cell that is protease deficient, and knockout of pyr5 function results in a cell with a pyrimidine deficient phenotype. In one genetic engineering approach, homologous recombination is used to induce targeted gene modifications by specifically targeting a gene *in vivo* to suppress expression of the encoded protein. In alternative approaches, siRNA, antisense and/or ribozyme technology find use in inhibiting gene expression. A variety of methods are known in the art for reducing expression of protein in cells, including, but not limited to deletion of all or part of the gene encoding the protein and site-specific mutagenesis to disrupt expression or activity of the gene product. (See e.g., Chaveroche et al., Nucl. Acids Res., 28:22 e97 [2000]; Cho et al., Molec. Plant Microbe Interact., 19:7-15 [2006]; Maruyama and Kitamoto, Biotechnol Lett., 30:1811-1817 [2008]; Takahashi et al., Mol. Gen. Genom., 272: 344-352 [2004]; and You et al., Arch. Microbiol.,191:615-622 [2009], all of which are incorporated by reference herein). Random mutagenesis, followed by screening for desired mutations also finds use (See e.g., Combier et al., *FEMS Microbiol. Lett.*, **220**:141–8 [2003]; and Firon et al., *Eukary. Cell* **2**:247–55 [2003], both of which are incorporated by reference).

[0196] Introduction of a vector or DNA construct into a host cell can be accomplished using any suitable method known in the art, including but not limited to calcium phosphate transfection, DEAE-dextran mediated transfection, PEG-mediated transformation, electroporation, or other common techniques known in the art. In some embodiments, the *Escherichia coli* expression vector pCK100900i (See, US Pat. No. 9,714,437, which is hereby incorporated by reference herein) finds use.

[0197] In some embodiments, the engineered host cells (i.e., "recombinant host cells") of the present invention are cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the pantothenate kinase polynucleotide. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and are well-known to those skilled in the art. As noted, many standard references and texts are available for the culture and production of many cells, including cells of bacterial, plant, animal (especially mammalian) and archebacterial origin.

[0198] In some embodiments, cells expressing the variant pantothenate kinase polypeptides of the invention are grown under batch or continuous fermentations conditions. Classical "batch

fermentation" is a closed system, wherein the compositions of the medium are set at the beginning of the fermentation and is not subject to artificial alternations during the fermentation. A variation of the batch system is a "fed-batch fermentation" which also finds use in the present invention. In this variation, the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is likely to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Batch and fed-batch fermentations are common and well known in the art. "Continuous fermentation" is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation systems strive to maintain steady state growth conditions. Methods for modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

[0199] In some embodiments of the present invention, cell-free transcription/translation systems find use in producing variant pantothenate kinase(s). Several systems are commercially available and the methods are well-known to those skilled in the art.

[0200] The present invention provides methods of making variant pantothenate kinase polypeptides or biologically active fragments thereof. In some embodiments, the method comprises: providing a host cell transformed with a polynucleotide encoding an amino acid sequence that comprises at least about 70% (or at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%) sequence identity to SEQ ID NO: 4, SEQ ID NO: 44, SEQ ID NO: 320, and/or SEQ ID NO: 526, and comprising at least one mutation as provided herein; culturing the transformed host cell in a culture medium under conditions in which the host cell expresses the encoded variant pantothenate kinase polypeptide; and optionally recovering or isolating the expressed variant pantothenate kinase polypeptide, and/or recovering or isolating the culture medium containing the expressed variant pantothenate kinase polypeptide. In some embodiments, the methods further provide optionally lysing the transformed host cells after expressing the encoded pantothenate kinase polypeptide and optionally recovering and/or isolating the expressed variant pantothenate kinase polypeptide from the cell lysate. The present invention further provides methods of making a variant pantothenate kinase polypeptide comprising cultivating a host cell transformed with a variant pantothenate kinase polypeptide under conditions suitable for the production of the variant pantothenate kinase polypeptide and recovering the variant pantothenate kinase polypeptide. Typically, recovery or isolation of the pantothenate kinase polypeptide is from the host cell culture medium, the host cell or both, using protein recovery techniques that are well known in the art, including those described herein. In some embodiments, host cells are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells

employed in expression of proteins can be disrupted by any convenient method, including, but not limited to freeze-thaw cycling, sonication, mechanical disruption, and/or use of cell lysing agents, as well as many other suitable methods well known to those skilled in the art.

[0201] Engineered pantothenate kinase enzymes expressed in a host cell can be recovered from the cells and/or the culture medium using any one or more of the techniques known in the art for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available under the trade name CelLytic B™ (Sigma-Aldrich). Thus, in some embodiments, the resulting polypeptide is recovered/isolated and optionally purified by any of a number of methods known in the art. For example, in some embodiments, the polypeptide is isolated from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, chromatography (e.g., ion exchange, affinity, hydrophobic interaction, chromatofocusing, and size exclusion), or precipitation. In some embodiments, protein refolding steps are used, as desired, in completing the configuration of the mature protein. In addition, in some embodiments, high performance liquid chromatography (HPLC) is employed in the final purification steps. For example, in some embodiments, methods known in the art, find use in the present invention (See e.g., Parry et al., Biochem. J., 353:117 [2001]; and Hong et al., Appl. Microbiol. Biotechnol., 73:1331 [2007], both of which are incorporated herein by reference). Indeed, any suitable purification methods known in the art find use in the present invention.

[0202] Chromatographic techniques for isolation of the pantothenate kinase polypeptide include, but are not limited to reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., are known to those skilled in the art.

[0203] In some embodiments, affinity techniques find use in isolating the improved pantothenate kinase enzymes. For affinity chromatography purification, any antibody which specifically binds the pantothenate kinase polypeptide may be used. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., may be immunized by injection with the pantothenate kinase. The pantothenate kinase polypeptide may be attached to a suitable carrier, such as BSA, by means of a side chain functional group or linkers attached to a side chain functional group. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (Bacillus Calmette Guerin) and *Corynebacterium parvum*.

[0204] In some embodiments, the pantothenate kinase variants are prepared and used in the form of cells expressing the enzymes, as crude extracts, or as isolated or purified preparations. In some embodiments, the pantothenate kinase variants are prepared as lyophilisates, in powder form (e.g., acetone powders), or prepared as enzyme solutions. In some embodiments, the pantothenate kinase variants are in the form of substantially pure preparations.

[0205] In some embodiments, the pantothenate kinase polypeptides are attached to any suitable solid substrate. Solid substrates include but are not limited to a solid phase, surface, and/or membrane. Solid supports include, but are not limited to 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 the substrate 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. A plurality of supports can be configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments.

[0206] In some embodiments, immunological methods are used to purify pantothenate kinase variants. In one approach, antibody raised against a variant pantothenate kinase polypeptide (e.g., against a polypeptide comprising SEQ ID NO: 4, SEQ ID NO: 44, SEQ ID NO: 320, and/or SEQ ID NO: 526, and/or an immunogenic fragment thereof) using conventional methods is immobilized on beads, mixed with cell culture media under conditions in which the variant pantothenate kinase is bound, and precipitated. In a related approach, immunochromatography finds use.

[0207] In some embodiments, the variant pantothenate kinases are expressed as a fusion protein including a non-enzyme portion. In some embodiments, the variant pantothenate kinase sequence is fused to a purification facilitating domain. As used herein, the term "purification facilitating domain" refers to a domain that mediates purification of the polypeptide to which it is fused. Suitable purification domains include, but are not limited to metal chelating peptides, histidine-tryptophan modules that allow purification on immobilized metals, a sequence which binds glutathione (e.g., GST), a hemagglutinin (HA) tag (corresponding to an epitope derived from the influenza hemagglutinin protein; See e.g., Wilson et al., Cell 37:767 [1984]), maltose binding protein sequences, the FLAG epitope utilized in the FLAGS extension/affinity purification system (e.g., the system available from Immunex Corp), and the like. One expression vector contemplated for use in the compositions and methods described herein provides for expression of a fusion protein comprising a polypeptide of the invention fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography; See e.g., Porath *et al.*, Prot. Exp. Purif., 3:263-281 [1992]) while the enterokinase

cleavage site provides a means for separating the variant pantothenate kinase polypeptide from the fusion protein. pGEX vectors (Promega) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (e.g., glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

[0208] Accordingly, in another aspect, the present invention provides methods of producing the engineered enzyme polypeptides, where the methods comprise culturing a host cell capable of expressing a polynucleotide encoding the engineered enzyme polypeptide under conditions suitable for expression of the polypeptide. In some embodiments, the methods further comprise the steps of isolating and/or purifying the enzyme polypeptides, as described herein.

[0209] Appropriate culture media and growth conditions for host cells are well known in the art. It is contemplated that any suitable method for introducing polynucleotides for expression of the enzyme polypeptides into cells will find use in the present invention. Suitable techniques include, but are not limited to electroporation, biolistic particle bombardment, liposome mediated transfection, calcium chloride transfection, and protoplast fusion.

[0210] Various features and embodiments of the present invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXPERIMENTAL

[0211] 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.

[0212] In the experimental disclosure below, the following abbreviations apply: M (molar); mM (millimolar), uM and μM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg (micrograms); L and l (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); um 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 Centigrade); RT and rt (room temperature); RH (relative humidity); CV (coefficient of variability); CAM and cam (chloramphenicol); PMBS (polymyxin B sulfate); IPTG (isopropyl β-D-l-thiogalactopyranoside); LB (Luria 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

E. coli Expression Hosts Containing Recombinant PanK Genes

[0213] The initial engineered PanK enzyme used to produce the variants of the present invention, SEQ ID NO: 4, was cloned into the expression vector pCK110900 (See, FIG. 3 of US Pat. Appln. Publn. No. 2006/0195947) operatively linked to the lac promoter under control of the lacI repressor. The expression vector also contains the P15a origin of replication and the chloramphenicol resistance gene. The resulting plasmids were transformed into *E. coli* W3110, using standard methods known in the art. The transformants were isolated by subjecting the cells to chloramphenicol selection, as known in the art (See e.g., US Pat. No. 8,383,346 and WO2010/144103).

EXAMPLE 2

Preparation of HTP PanK-Containing Wet Cell Pellets

[0214] *E. coli* cells containing recombinant PanK-encoding genes from monoclonal colonies were inoculated into 180µl LB containing 1% glucose and 30 µg/mL chloramphenicol (CAM) in the wells of 96-well, shallow-well microtiter plates. The plates were sealed with O_2-permeable seals, and cultures were grown overnight at 30°C, 200 rpm, and 85% humidity. Then, 10µl of each of the cell cultures were transferred into the wells of 96-well, deep-well plates containing 390 mL TB and 30 µg/mL CAM. The deep-well plates were sealed with O_2-permeable seals and incubated at 30°C, 250 rpm, and 85% humidity until OD_{600} 0.6-0.8 was reached. The cell cultures were then induced by IPTG to a final concentration of 1 mM and incubated overnight under the same conditions as originally used. The cells were then pelleted using centrifugation at 4,000 rpm for 10 min. The supernatants were discarded, and the pellets were frozen at -80°C prior to lysis.

EXAMPLE 3

Preparation of HTP PanK-Containing Cell Lysates

[0215] First, 200 µl lysis buffer containing 20 mM potassium phosphate buffer, pH 7.5, 1 mg/mL lysozyme, and 0.5 mg/mL PMBS were added to the cell paste in each well, produced as described in Example 2. The cells were lysed at room temperature for 2 hours with shaking on a bench top shaker.

The plate was then centrifuged for 15 min at 4,000 rpm and 4°C. The clear supernatants were used in biocatalytic reactions to determine their activity levels.

EXAMPLE 4

Preparation of Lyophilized Lysates from Shake Flask (SF) Cultures

[0216] Selected HTP cultures grown as described above were plated onto LB agar plates with 1% glucose and 30 μg/ml CAM and were grown overnight at 37°C. A single colony from each culture was transferred to 6 ml of LB with 1% glucose and 30μg/ml CAM. The cultures were grown for 18 h at 30°C, 250 rpm, and subcultured approximately 1:50 into 250 ml of TB containing 30 μg/ml CAM, to a final OD_{600} of 0.05. The cultures were grown for approximately 195 minutes at 30°C and 250 rpm, to an OD_{600} between 0.6-0.8 and induced with 1 mM IPTG. The cultures were then grown for 20 h at 30°C and 250 rpm. The cultures were centrifuged at 4,000 rpm for 20 min. The supernatant was discarded, and the pellets were resuspended in 30 ml of 25 mM potassium phosphate buffer, pH 7.5. The cells were pelleted (4,000 rpm for 20 min) and frozen at -80°C for 120 minutes. Frozen pellets were resuspended in 30 ml of 25 mM potassium phosphate buffer, pH 7.5, and lysed using a Microfluidizer system (Microfluidics) at 18,000 psi. The lysates were pelleted (10,000 rpm for 60 min), and the supernatants were frozen and lyophilized to generate shake flake (SF) enzymes.

EXAMPLE 5

Improvements over SEQ ID NO: 4 in the Phosphorylation Activity of 2-Ethynyl Glycerol

[0217] SEQ ID NO: 4 was selected as the parent enzyme after screening variants disclosed in US Pat. Appln. No. 16/460,147 for the phosphorylation activity of ethynyl glycerol substrate. Libraries of engineered genes were produced using well established techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. Each variant was screened in a 50 μL reaction that comprised of 50 mg/mL 2-ethynylglycerol (EGO) substrate, 0.1 mol% ATP, 0.25 mg/mL acetate kinase (SEQ ID NO: 632), 1.25 eq $(NH_4)_2$ SEQ ID NO: 4 was selected as the parent enzyme after screening variants disclosed in US Pat. Appln. No. 16/460,147 for the phosphorylation activity of ethynyl glycerol substrate. Libraries of engineered genes were produced using well established techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. Each variant was screened in a 50 μL reaction that comprised of 50 mg/mL 2-ethynylglycerol (EGO) substrate, 0.1 mol% ATP, 0.25 mg/mL acetate kinase (SEQ ID NO: 632), 1.25 eq $(NH_4)_2$acetyl phosphate (AcP), 10 mM $MgCl_2$ in 50 mM Bis-Tris buffer, pH 7.5 and 32x diluted soluble lysate for 15 hours at 30°C. The 96-well plates were heat-sealed and incubated in a Thermotron® shaker at 600 rpm to produce a chromophore containing species and

enable simple reaction monitoring. The reaction samples were derivatized using tripotassium 5,5',5'' –[2,2',2''-nitrilotris(methylene-tris(1H-benzimidazole-2,1-diyl)]tripentanoate hydrate ((BimC4A)3) at the following conditions 5 equivalents (eq) benzyl azide, 5 mol% Copper sulfate, 7.5 mol% (BimC4A)3, 20 mol% sodium ascorbate, 9:1 water : DMSO to achieve click chemistry. Post reaction, 5 uL of reaction were combined with 220 uL of (BimC4A)3 derivatization solution in new 96-well plates. The click chemistry reaction was incubated for 1 hour at 45°C. The samples were then filtered by centrifugation using 0.22 micron 96-well filter plates in preparation for analysis by UPLC-UV (Example 9) or RapidFire MS (Example 10).

[0218] Activity relative to SEQ ID NO: 4 (Activity FIOP) was calculated as the percent conversion of the product formed by the variant over the percent conversion produced by SEQ ID NO: 4 and shown in Table 5.1. The percent conversion was calculated by dividing the area of the product peak by the sum of the areas of the substrate, product and impurities/side product peaks as observed by the UPLC-UV analysis (Example 9).

Table 5.1 Activity of Variants Relative to SEQ ID NO: 4

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4)	Activity FIOP 1 (Relative to SEQ ID NO: 4)
5/6	K154H	+++
7/8	H141L	+
9/10	I287L	+
11/12	Y250F	+
13/14	I251L	+
15/16	K289A	+
17/18	K154R	+
19/20	I103V	+
21/22	K297V	+
23/24	E290G	+
25/26	T135S	+
27/28	I193V	+
29/30	Q247T	+
31/32	S216G	+
33/34	Y238F	+
35/36	K289S	+
37/38	T248S	+
39/40	T248A	+
41/42	K297R	+
43/44	I103V/T135S/H141L	+++
45/46	T135S/H141L/Y238F	++
47/48	A315L	+
49/50	I103V/H141L/S216G	+
51/52	V33I	+
53/54	N79S	+
55/56	T135S/I193V/Y238F/K289A/M291K	+
57/58	I103V/H141L/Y238F	+
59/60	S293L	+
61/62	R129S	+

Table 5.1 Activity of Variants Relative to SEQ ID NO: 4		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 4)**	**Activity FIOP 1 (Relative to SEQ ID NO: 4)**
63/64	T135S	+
65/66	R129G	+
67/68	A120T	+
69/70	R96P	+
71/72	V227T	+
73/74	D195A	+
75/76	I103V/T135S/S216G	+
77/78	P194H	+
79/80	R34A	+
81/82	M19L	+
83/84	H222G	+
85/86	I103V/H141L/I193V/S216G/Y238F	+
87/88	I103V/T135S	+
89/90	I103V/H141L	+
91/92	T135S/R323C	+
93/94	M218C	+
95/96	I193V/S216G	+
97/98	E63L	+
99/100	K274A	+
101/102	T135S/I251L/R323C	+
103/104	A31R	+
105/106	D197G	+
107/108	T135S/H141L/Q298A/R323C	+
109/110	R82T	+
111/112	T286D	+
113/114	N37F	+
115/116	S77L	+
117/118	N73G	+
119/120	T135S/D219G	+
121/122	H141L/K297V	+
123/124	D195G	+
125/126	D44W	+
127/128	M291G	+
129/130	P221E	+
131/132	R130Y	+
133/134	T17V	+
135/136	D35R	+
137/138	Q85S	+
139/140	R124G	+
141/142	P126S	+
143/144	N79A	+
145/146	P126V	+
147/148	I46K	+
149/150	H141L/I193V/I251L/R323C	+
151/152	T135S/T248A/I251L/M291K/K297R	+
153/154	F74L	+
155/156	H128S	+
157/158	D197V	+
159/160	S308A	+

Table 5.1 Activity of Variants Relative to SEQ ID NO: 4		
SEQ ID NO: (nt/aa)	**Amino Acid Differences (Relative to SEQ ID NO: 4)**	**Activity FIOP 1 (Relative to SEQ ID NO: 4)**
161/162	P221S	+
163/164	K198R	+
165/166	H141L/I251L/R323C	+
167/168	P126Q	+
169/170	V145I	+
171/172	G91C	+
173/174	H141L/R323C	+
175/176	T135S/Y238F/M291K	+
177/178	H141L/I251L	+
179/180	N79Y	+
181/182	I251L	+
183/184	G94S	+
185/186	H225Y	+
187/188	I103V/T135S/M291K	+
189/190	V233C	+
191/192	L80E	+
193/194	Q95G	+
195/196	H164P	+
197/198	R323C	+
199/200	I103V/S216G/Y238F	+
201/202	D35S	+
203/204	I103V	+
205/206	A120L	+
207/208	I76V	+
209/210	D26A	+
211/212	D26T	+
[1] Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4 and defined as follows: "+" > than 1-fold but less than 2.0-fold increased activity; "++" > than 2.0-fold but less than 4-fold increased activity; "+++" > than 4-fold increased activity		

EXAMPLE 6

Improvements over SEQ ID NO: 44 in the Phosphorylation Activity of 2-Ethynyl Glycerol

[0219] SEQ ID NO: 44 was selected as the parent enzyme for this round of directed evolution. Libraries of engineered genes were produced using well established techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. Each variant was screened in a 50 µL reaction that comprised of 50 mg/mL 2-ethynylglycerol (EGO) substrate, 0.1 mol% ATP, 0.25 mg/mL acetate kinase (SEQ ID NO: 632), 1.25 eq $(NH_4)_2AcP$, 10 mM $MgCl_2$ in 50 mM Bis-Tris buffer, pH 7.0 (final master mix pH 6.8) and 64x diluted soluble lysate for 15 hours at 30°C. The 96-well plates were heat-sealed and incubated in a Thermotron® shaker at 600 rpm. To produce a chromophore containing species and enable simple reaction monitoring, the reaction samples were derivatized using tripotassium 5,5',5'' –[2,2',2''-nitrilotris(methylene-tris(1H-benzimidazole-2,1-diyl)]tripentanoate hydrate ((BimC4A)3) at

the following conditions 5 eq benzyl azide, 5 mol% Copper sulfate, 7.5 mol% (BimC4A)3, 20 mol% sodium ascorbate, 9:1 water : DMSO to achieve click chemistry. Post reaction, 5 uL of reaction were combined with 220 uL of (BimC4A)3 derivatization solution in new 96-well plates. The click chemistry reaction was incubated for 1 hour at 45°C. The samples were then filtered by centrifugation using 0.22 micron 96-well filter plates in preparation for analysis by UPLC-UV (Example 9) or RapidFire MS (Example 10).

[0220] Activity relative to SEQ ID NO: 44 (Activity FIOP) was calculated as the click product peak intensity, of each variant, per click product peak intensity of SEQ ID NO: 44 using Analytical Method 10.1, and the results are shown in Table 6.1.

Table 6.1 Activity of Variants Relative to SEQ ID NO: 44

SEQ ID NO: (nt/aa)	**Amino Acid Differences** (Relative to SEQ ID NO: 44)	**Activity FIOP** [1] (Relative to SEQ ID NO: 44)
213/214	D35R/Q85S/M218C/H225Y	+++
215/216	D35R/M218C/S308A	+++
217/218	D35R/D197V/K274A	+++
219/220	D35R/F74L/M218C	++
221/222	M19L/D35R/D197G/M218C/L275M	++
223/224	T20G/H164S/N179V/S308M/E320K	++
225/226	T20G/E59R/A120V/H164S/T263M/K282A	++
227/228	T20G/S308M	++
229/230	F74L/L141H	++
231/232	S77L/M218C	++
233/234	E59R/A120V	++
235/236	T17V/E63L/R96P	++
237/238	T20G/H164S	+
239/240	M19L/Q85S/V145I/D197G/M218C/S308A	+
241/242	T20G/K282A	+
243/244	T17V/P126V/R129S/R130Y/H164P	+
245/246	T17V/I46K/H128S/R129G/R130Y/H164P/I193V	+
247/248	T17V/I46K/E63L/R96P	+
249/250	T263M	+
251/252	H164S/P221C/S308M	+
253/254	M19L/L141H	+
255/256	T17V/P126Q/R129G/H164P	+
257/258	T20G/E59R/N93G	+
259/260	I46K/H164P	+
261/262	D35R/L141H/D195G/M218C/K274A	+
263/264	E59R	+
265/266	T17V/E63L/H164P	+
267/268	M19L/F74L/V145I	+
269/270	T17V/I46K/E63L	+
271/272	V33I/D35S/I76V/L80E/P126S/P221E/H222G/V227T	+
273/274	T17V/R34A/R96P/H128S/R129G/H164P	+
275/276	L141H/D195G/D197G/M218C	+

Table 6.1 Activity of Variants Relative to SEQ ID NO: 44

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 44)	Activity FIOP [1] (Relative to SEQ ID NO: 44)
277/278	A31R/I76V/N79A/A120T/P126S/S293L	+
279/280	A120V/H164S/K282A/E321L	+
281/282	A31R/V33I/D35S/A120T/P126S/V227T	+
283/284	T20G/E59A/N93G/L141H/E321L	+
285/286	D35R/V145I/T286D	+
287/288	S308A	+
289/290	R34A/I46K/E63L/H164P	+
291/292	T17V/I46K/E63L/R96P	+
293/294	D35R	+
295/296	T17V/E63L/R96P/H164P	+
297/298	T17V/I46K/H164P	+
299/300	T17V/E63L/N79S/H128S/R129G/R130Y/H164P	+
301/302	M218C	+
303/304	H225Y	+
305/306	Q16R/N93G/A120V/P221C/S308M	+
307/308	T17V/N79S/H128S/R129S/R130Y	+
309/310	H128S/R129G/R130Y	+
311/312	D35S/I76V/L80E/P126S	+
313/314	R34A/N79S/P126Q/H128S/R130Y	+
315/316	I46K/N79S	+
317/318	T17V/P126V	+
319/320	V33I/D35S/P126S/H222G/V227T	+
321/322	D35S/G94S/A120T/P126S/S293L	+
323/324	T17V/E63L	+
325/326	E63L/H164P	+
327/328	N79S/P126V/H164P	+
329/330	D26T/V33I/L80E/P126S/M291G/S293L	+
331/332	R34A/H164P	+
333/334	M19L/S77L/L141H/M218C/H225Y	+
335/336	G94S/P126S/S293L	+
337/338	I46K/I193V	+
339/340	V33I/N79A/P126S/P221E/V227T/S293L	+
341/342	I46K/H128S/R129G/H164P	+
343/344	V33F/D35S/P126S/S293P	+
345/346	P126V/R129G/H164P	+
347/348	T17V/E63L/N79S	+
349/350	A32S/I46K	+
351/352	L80E/P126S/P221E/V227T	+
353/354	T17V/E63L/I193V	+
355/356	D35R/S77L/M218C	+
357/358	D26T/A31R/V33I/D35S/I76V/N79A/G94S/P221E/S293L	+
359/360	V33I/I76V/N79A	+
361/362	I46K/R96P/H164P	+
363/364	R34A/I46K/N79S/H164P	+
365/366	R34A/E63L/P126V/R129S/I193V	+
367/368	I46K/E63L	+

Table 6.1 Activity of Variants Relative to SEQ ID NO: 44

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 44)	Activity FIOP [1] (Relative to SEQ ID NO: 44)
369/370	K274A	+
371/372	A31R/V33I/D35S/I76V/L141H/S293L	+
373/374	E63L	+
375/376	M19L/L141H/H225Y	+
377/378	A31R/V33I/A120T	+
379/380	D26T/A31R/V33I/I76V/N79A	+
381/382	V33I/D35S/P126S/K198R/S293L	+
383/384	A31R/V33I/I76V/N79A/L80E/G94S/M291G	+
385/386	Q16R/L141H/N179V/K282A	+
387/388	S308M	+
389/390	D26T/A31R/A120T	+
391/392	I46K/H164P/I193V	+
393/394	I76V/N79A/L80E/A120T	+
395/396	T17V/I46K/E63L/H164P	+
397/398	F74L/Q85S/V145I	+
399/400	A31R/I76V/N79A/S293L	+
401/402	I46K/R129G	+
403/404	D26T/V33I/D35S/P126S/H222G/M291G	+
405/406	D35R/F74L/S77L/D197V/H225Y	+
407/408	R34A/I46K/E63L	+
409/410	P126V/H128S/R129G/I193V	+
411/412	T17V/R34A/I46K	+
413/414	A31R/V33I/D35S/I76V/N79A/G91C/M291G	+
415/416	E63L/I193V	+
417/418	V33I/D35S/P221E/H222G	+
419/420	D35R/Q85S/M218C/K274A	+
421/422	D26T/D35S/A120T/H222G/V227T/S293L	+
423/424	D35S/I76V	+
425/426	T17V	+
427/428	T17V/E63L/R129S/R130Y	+
429/430	V33I/I76V/N79A/P126S	+
431/432	D26T/A31R/V33I/I76V/N79A/G94S	+
433/434	R34A/I46K/N79S/R96P/P126V	+
435/436	A31R/D35S/L80E/G94S/P126S/L141H/S293L	+
437/438	T17V/I46K/P126V/H128S/R129S/H164P	+
439/440	I46K/N79S/P126V/I193V	+
441/442	V33I/D35S/I76V/N79A/A120T/L141H/K198R/P221E	+
443/444	G94S/Q95G/A120T/M291G	+
445/446	N79S/R96P/P126Q/R129S	+
447/448	V33I/D35S	+
449/450	D35R/Q85S/L141H/M218C	+
451/452	T17V/I46K	+
453/454	G94S/S293L	+
455/456	R34A/I46K/H164P	+
457/458	R34A/I46K	+
459/460	A31R/V33I/D35S/P126S	+
461/462	H164S	+

Table 6.1 Activity of Variants Relative to SEQ ID NO: 44

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 44)	Activity FIOP [1] (Relative to SEQ ID NO: 44)
463/464	R34A	+
465/466	M19L/D35R/Q85S/K274A	+
467/468	D26T/D35S/N79A/A120T/P126S/P194H/K198R	+
469/470	T17V/R34A	+
471/472	A31R/V33I/I76V/N79A/Q95G	+
473/474	D26T/V33I/G94S/S293L	+
475/476	D35R/K274A	+
477/478	D35R/Q85S	+
479/480	A120V	+
481/482	D26T/A31R/V33I/D35S/A120T/L141H/S293L	+
483/484	D35S/I76V/N79A/L80E/P126S	+
485/486	M19L/F74L/S77L/L141H/M218C	+
487/488	S77L/Q85S/V145I	+
489/490	D26T/D35S/P126S	+
491/492	V33I/D35S/G94S/P126S/L141H	+
493/494	M19L	+
495/496	V33I/D35S/I76V/L80E/S293L	+
497/498	R34A/I46K/R48C	+
499/500	I46K	+

[1] Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 44 and defined as follows: "+" > than 1-fold but less than 2.0-fold increased activity; "++" > than 2.0-fold but less than 4-fold increased activity; "+++" > than 4-fold increased activity

EXAMPLE 7

Improvements over SEQ ID NO: 320 in the Phosphorylation Activity of 2-Ethynyl Glycerol and Stability of the PanK enzyme

[0221] SEQ ID NO: 320 was selected as the parent enzyme for this round of directed evolution. Libraries of engineered genes were produced using well established techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. Each variant was screened in a 50 μL reaction that comprised of 50 mg/mL 2-ethynylglycerol (EGO) substrate, 0.1 mol% ATP, 0.125 mg/mL acetate kinase (SEQ ID NO: 634), 1.25 eq $(NH_4)_2AcP$, 10 mM MgCl2 in 50 mM Bis-Tris buffer, pH 7.5 (final master mix pH 8.0) and 64x diluted soluble lysate for 15 hours at 30°C. For stability testing, the soluble lysates, generated as described in Example 3, were pre-incubated at 40°C for 1.5 hours and then diluted 32x for screening. The 96-well plates were heat-sealed and incubated in a Thermotron® shaker at 600 rpm. To produce a chromophore containing species and enable simple reaction monitoring, the reaction samples were derivatized using tripotassium 5,5',5'' –[2,2',2''-nitrilotris(methylene-tris(1H-benzimidazole-2,1-diyl)]tripentanoate hydrate ((BimC4A)3) at the following conditions 5 eq benzyl

azide, 5 mol% Copper sulfate, 7.5 mol% (BimC4A)3, 20 mol% sodium ascorbate, 9:1 water : DMSO to achieve click chemistry. Post reaction, 5 uL of reaction were combined with 220 uL of (BimC4A)3 derivatization solution in new 96-well plates. The click chemistry reaction was incubated for 1 hour at 45°C. The samples were then filtered by centrifugation using 0.22 micron 96-well filter plates in preparation for analysis by UPLC-UV (Example 9) or RapidFire MS (Example 10).

[0222] Activity and stability relative to SEQ ID NO: 320 (Activity FIOP) were calculated as the click product peak intensity, of each variant, per click product peak intensity of SEQ ID NO: 320 using Analytical method 10.1, and the results are shown in Table 7.1.

Table 7.1 Activity and Stability of Variants Relative to SEQ ID NO: 320

SEQ ID NO: (nt/aa)	**Amino Acid Differences** (Relative to SEQ ID NO: 320)	**Activity FIOP** [1] (Relative to SEQ ID NO: 320)	**Stability FIOP** [2] (Relative to SEQ ID NO: 320)
501/502	T20G/G94S/A120T	+	+
503/504	R48C/E63L	+	+
505/506	T20G/R48C/M218C/T227V	+	+
507/508	G94S/A120T	+	+
509/510	T20G/A31R/S35D/R48C/E63L/F74L/G94S/A120T/H164P	+	+
511/512	E63L	+	+
513/514	E63L/A120T/D197G/M218C/L275M	+	+
515/516	E63L/R129S/V145I	+	+
517/518	R48C/E63L/V145I	+	+
519/520	T20G/A120T/V145I/D197G/T227V	++	+
521/522	R48C/E63L/F74L/N79S/A120T/R129G/T227V/K282A	+	+
523/524	F74L/N79A/G94S/L275M	+	+
525/526	R48C/E63L/G94S/L275M	++	+
527/528	T20G/S35D/A120T/D197G	+	+
529/530	E63L/M218C/L275M/K282A	+	+
531/532	T20G/R48C/E63L/G94S/A120T/V145I/D197G/M218C/T227V	+	+
533/534	T20G/A31R/S35D/F74L/N79A/H164S/M218C/H225Y/K282A	+	+
535/536	T20G/D197G/H225Y/T227V	+	+
537/538	Q95D	+	+
539/540	Q95E	+	
541/542	K289C		+
543/544	I33V		+
545/546	G271M	+	+
547/548	Q24H		+
549/550	P178H		+
551/552	G196I	+	+
553/554	K277R	+	+

Table 7.1 Activity and Stability of Variants Relative to SEQ ID NO: 320

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 320)	Activity FIOP [1] (Relative to SEQ ID NO: 320)	Stability FIOP [2] (Relative to SEQ ID NO: 320)
555/556	P178D	+	+
557/558	Q24V	+	++
559/560	V201G		+
561/562	P221G		++
563/564	R130L		+
565/566	I46A		+
567/568	S293Y	+	+
569/570	E321L	+	+
571/572	R149A	++	+
573/574	H164G	++	+
575/576	Q298E	++	+
577/578	K282D	+	++
579/580	S314G	++	+
581/582	Q298D	+	++
583/584	R48V	+	
585/586	G271C	+	+

[1] Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 320 and defined as follows: "+" > than 1-fold but less than 1.75-fold increased activity; "++" > than 1.75-fold

[2] Levels of increased stability were determined relative to the reference polypeptide of SEQ ID NO: 320 and defined as follows: "+" > than 1-fold but less than 1.75-fold increased activity; "++" > than 1.75-fold

EXAMPLE 8

Improvements over SEQ ID NO: 526 in the Phosphorylation Activity of 2-Ethynyl Glycerol and Stability of the PanK enzyme

[0223] SEQ ID NO: 526 was selected as the parent enzyme for this round of directed evolution. Libraries of engineered genes were produced using well established techniques (e.g., saturation mutagenesis, recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP as described in Example 2, and the soluble lysate was generated as described in Example 3. Each variant was screened in a 50 µL reaction that comprised of 50 mg/mL 2-ethynylglycerol (EGO) substrate, 0.1 mol% ATP, 0.125 mg/mL acetate kinase (SEQ ID NO: 634), 1.25 eq $(NH_4)_2AcP$, 10 mM $MgCl_2$ in 50 mM Bis-Tris buffer, pH 7.5 (final master mix pH 8.0) and 128x diluted soluble lysate for 15 hours at 30°C. For stability testing, the soluble lysates, generated as described in Example 3, were pre-incubated at 45°C for 1.5 hours and then diluted 64x for screening. The 96-well plates were heat-sealed and incubated in a Thermotron® shaker at 600 rpm. To produce a chromophore containing species and enable simple reaction monitoring, the reaction samples were derivatized using tripotassium 5,5',5'' –[2,2',2''-nitrilotris(methylene-tris(1H-

benzimidazole-2,1-diyl)]tripentanoate hydrate ((BimC4A)3) at the following conditions 5 eq benzyl azide, 5 mol% Copper sulfate, 7.5 mol% (BimC4A)3, 20 mol% sodium ascorbate, 9:1 water : DMSO to achieve click chemistry. Post reaction, 5 uL of reaction were combined with 220 uL of (BimC4A)3 derivatization solution in new 96-well plates. The click chemistry reaction was incubated for 1 hour at 45°C. The samples were then filtered by centrifugation using 0.22 micron 96-well filter plates in preparation for analysis by UPLC-UV (Example 9) or RapidFire MS (Example 10).

[0224] Activity and stability relative to SEQ ID NO: 526 (Activity FIOP) were calculated as the click product peak intensity, of each variant, per click product peak intensity of SEQ ID NO: 526 using Analytical method 10.1, and the results are shown in Table 8.1.

Table 8.1 Activity and Stability of Variants Relative to SEQ ID NO: 526

SEQ ID NO: (nt/aa)	**Amino Acid Differences** (Relative to SEQ ID NO: 526)	**Activity FIOP** [1] (Relative to SEQ ID NO: 526)	**Stability FIOP** [2] (Relative to SEQ ID NO: 526)
587/588	I33V/Q95N/A104T	++	+
589/590	P221G/K289C	++	+
591/592	Q24H/I46V/P221G	+	
593/594	Q24H/T227V	+	
595/596	I46V/Q95E/K282A	+	+
597/598	Q24H/I33V	+	+
599/600	C48V	+	+
601/602	V201G	+	+
603/604	P178D/S293Y/Q298D	+	+
605/606	Q95D/V201G/K282D/S293Y/S314G	+	+
607/608	Q24V/V201G/K282D	+	+
609/610	H164G/V201G/G271C/S314G	+	+
611/612	Q298D	+	+
613/614	Q24V/R149A/V201G/K282D	+	+
615/616	Q24V/Q95D/H164G/P178D/V201G/K282D	+	++
617/618	Q24V/P178D/V201G/G271C/K282D/S293Y/Q298E/S314G	+	+
619/620	Q24V	+	+
621/622	Q24V/R149A/P178D/K282D/S293Y/Q298E	+	+
623/624	K282D/S293Y/Q298E/S314G/E321L	+	+
625/626	V201G/K282D	+	+
627/628	Q24V/V201G/G271C/K282D/S293Y/Q298E	+	+
629/630	Q24V/R149A/K282D/S314G	+	+

[1] Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 526 and defined as follows: "+" > than 1.2-fold but less than 1.70-fold increased activity; "++" > than 1.70-fold

Table 8.1 Activity and Stability of Variants Relative to SEQ ID NO: 526			
SEQ ID NO: (nt/aa)	**Amino Acid Differences** (Relative to SEQ ID NO: 526)	**Activity FIOP** [1] (Relative to SEQ ID NO: 526)	**Stability FIOP** [2] (Relative to SEQ ID NO: 526)
[2] Levels of increased stability were determined relative to the reference polypeptide of SEQ ID NO: 526 and defined as follows: "+" > than 1.2-fold but less than 1.70-fold increased activity; "++" > than 1.70-fold			

EXAMPLE 9

Analytical Detection of (BimC4A)3 -Derived 2-Ethynylglycerol Phosphate by UPLC-UV

[0225] Data described in Example 5 was collected using the analytical method provided in Table 9.1. The method provided herein finds use in analyzing the variants produced using the present invention. However, it is not intended that present invention be limited to the methods described herein, as there are other suitable methods known in the art that are applicable to the analysis of the variants provided herein and/or produced using the methods provided herein.

Table 9.1 Analytical Method	
Instrument	Agilent 1290 – UPLC
Column	Waters Acquity HSS T3, 2.1 x 50 mm, 1.8 um
Mobile Phase	A: 0.1% formic acid in water; B: 0.1% formic acid in acetonitrile
Gradient	Time (min) %A %B 0 98 2 0.1 98 2 0.7 80 20 1.1 0 100 1.5 98 2 2.0 98 2
Flow Rate	1.25 mL/min
Run Time	2.0 min
Product Elution order	(BimC4A)3 -Derived 2-Ethynylglycerol phosphate: ~0.98 min 2-ethynylglycerol: ~1.35 min
Column Temperature	40 °C
Injection Volume	5 µL
Detection	UV 210 nm

EXAMPLE 10

Analytical Detection of (BimC4A)3 -Derived 2-Ethynylglycerol Phosphate by Rapid Fire-MS

[0226] Data described in Examples 6, 7, and 8 were collected using the analytical method provided in Table 10.1. The method provided herein finds use in analyzing the variants produced using the present invention. However, it is not intended that present invention be limited to the methods

described herein, as there are other suitable methods known in the art that are applicable to the analysis of the variants provided herein and/or produced using the methods provided herein.

Table 10.1 Analytical Method						
Instrument	Agilent RapidFire					
Pump1 Buffer	0.1% Formic water LCMS grade water; 1.5 mL/min flow rate					
Pump2 Buffer	ACC Juice (50% water LCMS grade, 25% acetonitrile LCMS grade, 25% acetone GCMS grade) + 100um ammonium acetate ; 1.25 mL/min flow rate					
Pump3 Buffer	ACC Juice (50% water LCMS grade, 25% acetonitrile LCMS grade, 25% acetone GCMS grade) + 100um ammonium acetate; 1.25 mL/min flow rate					
Aqueous wash	Water					
Organic wash	Acetonitrile					
SPE cartridge	C18					
RF state 1 Aspirate	600 ms					
RF state 2 Load/Wash	3000 ms					
RF state 3 Extra Wash	0					
RF state 4 Elute	6500 ms					
RF state 5 Reequilibrate	1000 ms					
Agilent Jet Stream source parameters						
Drying gas temperature	350 °C					
Drying gas flow	13 L/min					
Nebulizer pressure	45 psi					
Sheath gas temperature	330 °C					
Sheath gas flow	12 L/min					
Capillary voltage	2500 V					
Nozzle voltage	2500 V					
Agilent 6470 Triple Quadrupole MRM parameters						
Compound	Q1	Q3	Dwell	Fragmentor	CE	CAV
(BimC4A)3 -Derived 2-Ethynylglycerol phosphate	330	231.8	109	80	10	3
(BimC4A)3 -Derived 2-Ethynylglycerol phosphate	330	214	109	80	10	3
(BimC4A)3 -Derived 2-Ethynylglycerol phosphate	330	90.8	109	80	30	3

CLAIMS

We claim:

1. An engineered pantothenate kinase comprising a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 4, SEQ ID NO: 44, SEQ ID NO: 320, and/or SEQ ID NO: 526, or a functional fragment thereof, wherein said engineered pantothenate kinase comprises at least one substitution or substitution set in said polypeptide sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4, SEQ ID NO: 44, SEQ ID NO: 320, and/or SEQ ID NO: 526.

2. The engineered pantothenate kinase of Claim 1, comprising a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 4, or a functional fragment thereof, wherein said engineered pantothenate kinase comprises at least one substitution or substitution set in said polypeptide sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4.

3. The engineered pantothenate kinase of Claim 2, wherein said engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 4, or a functional fragment thereof, and wherein said engineered pantothenate kinase comprises at least one substitution or substitution set at one or more positions selected from 103, 103/135/141, 17, 19, 26, 31, 33, 34, 35, 37, 44, 46, 63, 73, 74, 76, 77, 79, 80, 82, 85, 91, 94, 95, 96, 103/135, 103/135/216, 103/135/291, 103/141, 103/141/193/216/238, 103/141/216, 103/141/238, 103/216/238, 120, 124, 126, 128, 129, 130, 135, 135/141/238, 135/141/298/323, 135/193/238/289/291, 135/219, 135/238/291, 135/248/251/291/297, 135/251/323, 135/323, 141, 141/193/251/323, 141/251, 141/251/323, 141/297, 141/323, 145, 154, 164, 193, 193/216, 194, 195, 197, 198, 216, 218, 221, 222, 225, 227, 233, 238, 247, 248, 250, 251, 274, 286, 287, 289, 290, 291, 293, 297, 308, 315, and 323, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4.

4. The engineered pantothenate kinase of Claim 1, comprising a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 44, or a functional fragment thereof, wherein said engineered pantothenate kinase comprises at least one substitution or substitution set in said polypeptide sequence, and wherein the amino acid positions of said polypeptide sequence are

numbered with reference to SEQ ID NO: 44.

5. The engineered pantothenate kinase of Claim 2, wherein said engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 44, or a functional fragment thereof, and wherein said engineered pantothenate kinase comprises at least one substitution or substitution set at one or more positions selected from 16/93/120/221/308, 16/141/179/282, 17, 17/34, 17/34/46, 17/34/96/128/129/164, 17/46, 17/46/63, 17/46/63/96, 17/46/63/164, 17/46/126/128/129/164, 17/46/128/129/130/164/193, 17/46/164, 17/63, 17/63/79, 17/63/79/128/129/130/164, 17/63/96, 17/63/96/164, 17/63/129/130, 17/63/164, 17/63/193, 17/79/128/129/130, 17/126, 17/126/129/130/164, 17/126/129/164, 19, 19/35/85/274, 19/35/197/218/275, 19/74/77/141/218, 19/74/145, 19/77/141/218/225, 19/85/145/197/218/308, 19/141, 19/141/225, 20/59/93, 20/59/93/141/321, 20/59/120/164/263/282, 20/164, 20/164/179/308/320, 20/282, 20/308, 26/31/33/35/76/79/94/221/293, 26/31/33/35/120/141/293, 26/31/33/76/79, 26/31/33/76/79/94, 26/31/120, 26/33/35/126/222/291, 26/33/80/126/291/293, 26/33/94/293, 26/35/79/120/126/194/198, 26/35/120/222/227/293, 26/35/126, 31/33/35/76/79/91/291, 31/33/35/76/141/293, 31/33/35/120/126/227, 31/33/35/126, 31/33/76/79/80/94/291, 31/33/76/79/95, 31/33/120, 31/35/80/94/126/141/293, 31/76/79/120/126/293, 31/76/79/293, 32/46, 33/35, 33/35/76/79/120/141/198/221, 33/35/76/80/126/221/222/227, 33/35/76/80/293, 33/35/94/126/141, 33/35/126/198/293, 33/35/126/222/227, 33/35/126/293, 33/35/221/222, 33/76/79, 33/76/79/126, 33/79/126/221/227/293, 34, 34/46, 34/46/48, 34/46/63, 34/46/63/164, 34/46/79/96/126, 34/46/79/164, 34/46/164, 34/63/126/129/193, 34/79/126/128/130, 34/164, 35, 35/74/77/197/225, 35/74/218, 35/76, 35/76/79/80/126, 35/76/80/126, 35/77/218, 35/85, 35/85/141/218, 35/85/218/225, 35/85/218/274, 35/94/120/126/293, 35/141/195/218/274, 35/145/286, 35/197/274, 35/218/308, 35/274, 46, 46/63, 46/79, 46/79/126/193, 46/96/164, 46/128/129/164, 46/129, 46/164, 46/164/193, 46/193, 59, 59/120, 63, 63/164, 63/193, 74/85/145, 74/141, 76/79/80/120, 77/85/145, 77/218, 79/96/126/129, 79/126/164, 80/126/221/227, 94/95/120/291, 94/126/293, 94/293, 120, 120/164/282/321, 126/128/129/193, 126/129/164, 128/129/130, 141/195/197/218, 164, 164/221/308, 218, 225, 263, 274, and 308, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 44.

6. The engineered pantothenate kinase of Claim 1, comprising a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 320, or a functional fragment thereof, wherein said engineered pantothenate kinase comprises at least one substitution or substitution set in said polypeptide sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 320.

7. The engineered pantothenate kinase of Claim 2, wherein said engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 320, or a functional fragment thereof, and wherein said engineered pantothenate kinase comprises at least one substitution or substitution set at one or more positions selected from 20/31/35/48/63/74/94/120/164, 20/31/35/74/79/164/218/225/282, 20/35/120/197, 20/48/63/94/120/145/197/218/227, 20/48/218/227, 20/94/120, 20/120/145/197/227, 20/197/225/227, 24, 33, 46, 48, 48/63, 48/63/74/79/120/129/227/282, 48/63/94/275, 48/63/145, 63, 63/120/197/218/275, 63/129/145, 63/218/275/282, 74/79/94/275, 94/120, 95, 130, 149, 164, 178, 196, 201, 221, 271, 277, 282, 289, 293, 298, 314, and 321, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 320.

8. The engineered pantothenate kinase of Claim 1, comprising a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 526, or a functional fragment thereof, wherein said engineered pantothenate kinase comprises at least one substitution or substitution set in said polypeptide sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 526.

9. The engineered pantothenate kinase of Claim 2, wherein said engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 526, or a functional fragment thereof, and wherein said engineered pantothenate kinase comprises at least one substitution or substitution set at one or more positions selected from 24, 24/33, 24/46/221, 24/95/164/178/201/282, 24/149/178/282/293/298, 24/149/201/282, 24/149/282/314, 24/178/201/271/282/293/298/314, 24/201/271/282/293/298, 24/201/282, 24/227, 33/95/104, 46/95/282, 48, 95/201/282/293/314, 164/201/271/314, 178/293/298, 201, 201/282, 221/289, 282/293/298/314/321, and 298, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 526.

10. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered pantothenate kinase variant set forth in Table 5-1, 6-1, 7-1, and/or Table 8-1.

11. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 4.

12. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a variant engineered polypeptide set forth in SEQ ID NO: 4.

13. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 44.

14. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a variant engineered polypeptide set forth in SEQ ID NO: 44.

15. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 320.

16. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a variant engineered polypeptide set forth in SEQ ID NO: 320.

17. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 526.

18. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a variant engineered polypeptide set forth in SEQ ID NO: 526.

19. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a polypeptide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the sequence of at least one engineered pantothenate kinase variant set forth in the even numbered sequences of SEQ ID NOS: 4-630.

20. The engineered pantothenate kinase of Claim 1, wherein said engineered pantothenate kinase comprises a polypeptide sequence forth in the even numbered sequences of SEQ ID NOS: 4-630.

21. The engineered pantothenate kinase of any of Claims 1-20, wherein said engineered pantothenate kinase comprises at least one improved property compared to wild-type *E. coli* pantothenate kinase.

22. The engineered pantothenate kinase of Claim 21, wherein said improved property comprises improved activity on a substrate, as compared to a wild-type pantothenate kinase.

23. The engineered pantothenate kinase of Claim 22, wherein said substrate comprises at least one alcohol.

24. The engineered pantothenate kinase of Claim 21, wherein said improved property comprises improved production of phospho-ethynyl glycerol, as compared to a wild-type pantothenate kinase.

25. The engineered pantothenate kinase of any of Claims 1-24, wherein said engineered pantothenate kinase is purified.

26. A composition comprising at least one engineered pantothenate kinase of any of Claims 1-25.

27. A polynucleotide sequence encoding at least one engineered pantothenate kinase of any of Claims 1-24.

28. A polynucleotide sequence encoding at least one engineered pantothenate kinase, wherein said polynucleotide sequence comprises at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 3, SEQ ID NO: 43, SEQ ID NO: 319, and/or SEQ ID NO: 525, wherein the polynucleotide sequence of said engineered pantothenate kinase comprises at least one substitution at one or more positions.

29. The polynucleotide sequence of Claim 27, encoding at least one engineered pantothenate kinase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 3, or a functional fragment thereof.

30. The polynucleotide sequence of Claim 27, encoding at least one engineered pantothenate kinase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%,

95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 43, or a functional fragment thereof.

31. The polynucleotide sequence of Claim 27, encoding at least one engineered pantothenate kinase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 319, or a functional fragment thereof.

32. The polynucleotide sequence of Claim 27, encoding at least one engineered pantothenate kinase comprising at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 525, or a functional fragment thereof.

33. The polynucleotide sequence of any of Claims 27-32, wherein said polynucleotide sequence is operably linked to a control sequence.

34. The polynucleotide sequence of any of Claims 27-33, wherein said polynucleotide sequence is codon optimized.

35. The polynucleotide sequence of any of Claims 27-34, wherein said polynucleotide comprises an odd-numbered sequence of SEQ ID NOS: 3-629.

36. An expression vector comprising at least one polynucleotide sequence of any of Claims 27-35.

37. A host cell comprising at least one expression vector of Claim 36.

38. A host cell comprising at least one polynucleotide sequence of any of Claims 27-34.

39. A method of producing an engineered pantothenate kinase in a host cell, comprising culturing the host cell of Claim 37 and/or 38, under suitable conditions, such that at least one engineered pantothenate kinase is produced.

40. The method of Claim 39, further comprising recovering at least one engineered pantothenate kinase from the culture and/or host cell.

41. The method of Claim 39 and/or 40, further comprising the step of purifying said at least one engineered pantothenate kinase.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/52644

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 9/12, C12N 9/14, C12N 9/10, C12P 13/02, C12P 13/04 (2021.01)

CPC - C12N 9/1205, C12N 9/1217, C12Y 207/01033, C12N 9/1022

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---- A	US 2020/0010815 A1 (CODEXIS, INC.) 9 January 2020 (09.01.2020) Abstract; Claim 2; Claim 13; Claim 14; Claim 15; SEQ ID NO 268; para [0136]	1-2, 11-12, 19-20, (21-24)/(1-2, 11-12, 19-20) ---------- 3, (21-24)/3
A	US 7,220,561 B2 (HERMANN et al.) 22 May 2007 (22.05.2007) Abstract	3, (21-24)/3

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 December 2021 (31.12.2021)	FEB 17 2022
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Kari Rodriquez Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/52644

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/52644

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 25-27, 29-41
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
----Please see continuation in first extra sheet ----------

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-3, 11-12, 19-24, limited to engineered pantothenate kinase polypeptide with 100% identity to SEQ ID NO:4 with a single substitution at position 103

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/52644

Continuation of Box No. III. Observations where unity of invention is lacking.

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I+, Claims 1-24, directed to an engineered pantothenate kinase comprising a polypeptide sequence. The engineered pantothenate kinase polypeptide will be searched to the extent that the engineered pantothenate kinase polypeptide encompasses a polypeptide sequence having at least 100% sequence identity to SEQ ID NO: 4, wherein said engineered pantothenate kinase comprises a single substitution at position 103 in said polypeptide sequence compared to a wild type sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4 (note, these are the first claimed sequences for the engineered pantothenate kinase of the first embodiment). It is believed that claims 1-3, 11-12, 19-24 encompass this first named invention, and thus these claims will be searched without fee to the extent that the engineered pantothenate kinase polypeptide encompasses a polypeptide sequence having at least 100% sequence identity to SEQ ID NO: 4, wherein said engineered pantothenate kinase comprises a single substitution at position 103 in said polypeptide sequence compared to a wild type sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4. Additional engineered pantothenate kinase variant polypeptide(s) with additional substitution(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected engineered pantothenate kinase variant polypeptide(s) with additional substitution(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be an engineered pantothenate kinase polypeptide encompasses a polypeptide sequence having at least 100% sequence identity to SEQ ID NO: 4, wherein said engineered pantothenate kinase comprises a single substitution at position 17 in said polypeptide sequence compared to a wild type sequence, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 4 (claims 1-3, 11-12, 19-24).

Group II+, Claim 28, directed to a polynucleotide sequence encoding at least one engineered pantothenate kinase. Group II+ will be searched upon payment of additional fees. The polynucleotide sequence may be searched, for example, to encompass a polynucleotide sequence encoding at least one engineered pantothenate kinase wherein said polynucleotide sequence comprises 100% sequence identity to SEQ ID NO: 3 for an additional fee and election as such. It is believed that claim 28 reads on this exemplary invention. Additional polynucleotide sequence(s) encoding additional engineered pantothenate kinase variant(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected polynucleotide sequence(s) encoding additional engineered pantothenate kinase variant(s). Failure to clearly identify how any paid additional invention fees are to be applied to the"+" group(s) will result in only the first claimed invention to be searched. Another exemplary election would be a polynucleotide sequence encoding an engineered pantothenate kinase wherein said polynucleotide sequence comprises 100% sequence identity to SEQ ID NO: 13 (claim 28).

The inventions listed as Group I+ do not relate to a single special technical feature under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special technical features

The inventions of Groups I+ and II+ each include the special technical feature of a unique amino acid and/or nucleic acid sequence. Each amino acid sequence encodes a unique peptide, and is considered a distinct technical feature.
Additionally,
Group I+ has the special technical feature of an engineered pantothenate kinase comprising a polypeptide sequence, that is not required by Group II+.
Group II+ has the special technical feature of a polynucleotide sequence encoding at least one engineered pantothenate kinase, that is not required by Group I+.

Common technical features

No technical features are shared between the engineered pantothenate kinase amino acid sequences of Group I+ and/or polynucleotide sequences encoding at least one engineered pantothenate kinase of Group II+ and, accordingly, these groups lack unity a priori.

Additionally, even if Groups I+ and II+ were considered to share the technical features of including:
an engineered pantothenate kinase comprising a polypeptide sequence, or a functional fragment thereof, wherein said engineered pantothenate kinase comprises at least one substitution or substitution set in said polypeptide sequence;
a polynucleotide sequence encoding at least one engineered pantothenate kinase polynucleotide sequence encoding at least one engineered pantothenate kinase;
these shared technical features are previously disclosed by US 2020/0010815 A1 to Codexis, Inc., (hereinafter 'Codexis').

-------please see continuation on next sheet ----------

Continuation of Box No. III. Observations where unity of invention is lacking.

-----continued from previous sheet ----------

Codexis teaches an engineered pantothenate kinase comprising a polypeptide sequence, or a functional fragment thereof, wherein said engineered pantothenate kinase comprises at least one substitution or substitution set in said polypeptide sequence;
a polynucleotide sequence encoding at least one engineered pantothenate kinase polynucleotide sequence encoding at least one engineered pantothenate kinase (Abstract - 'The present invention provides engineered pantothenate kinase (PanK) enzymes, polypeptides having PanK activity, and polynucleotides encoding these enzymes, as well as vectors and host cells comprising these polynucleotides and polypeptides. Methods for producing PanK enzymes are also provided. The present invention further provides compositions comprising the PanK enzymes and methods of using the engineered PanK enzymes. The present invention finds particular use in the production of pharmaceutical compounds.'; Claim 2 - 'The engineered pantothenate kinase of claim 1, wherein said engineered pantothenate kinase comprises a polypeptide sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 2, or a functional fragment thereof, and wherein said engineered pantothenate kinase comprises at least one substitution or substitution set at one or more positions selected from 277/281, 54/240/277/281, 240, 240/277, 240/277/281, 240/277/281/282, 240/281, and 240/281/282, 277, and wherein the amino acid positions of said polypeptide sequence are numbered with reference to SEQ ID NO: 2.').

As the technical features were known in the art at the time of the invention, they cannot be considered special technical features that would otherwise unify the groups.

Therefore, Group I+ and II+ inventions lack unity under PCT Rule 13 because they do not share the same or corresponding special technical feature.

*Continuation of Item 4 above: claims 25-27, 29-41 are held unsearchable because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).