US 20230272363A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2023/0272363 A1**

Voladri et al. (43) **Pub. Date: Aug. 31, 2023**

(54) **PENICILLIN-G ACYLASES**

(71) Applicant: **Codexis, Inc.**, Redwood City, CA (US)

(72) Inventors: **Rama Voladri**, Pleasanton, CA (US); **Christopher Michael Micklitsch**, Philadelphia, PA (US); **Oscar Alvizo**, Fremont, CA (US); **Jovana Nazor**, Milpitas, CA (US); **Da Duan**, Foster City, CA (US); **Judy Victoria Antonio Viduya**, Greenbrae, CA (US); **Stephan Jenne**, Foster City, CA (US); **Chihui An**, Scotch Plains, NJ (US); **Keith Allen Canada**, Freehold, NJ (US); **Paul N. Devine**, Tinton Falls, NJ (US); **Iman Farasat**, Rahway, NJ (US); **Anna Fryszkowska**, New York, NY (US); **Katrina W. Lexa**, Santa Rosa, CA (US); **Robert Kevin Orr**, Cranford, NJ (US)

(21) Appl. No.: **18/189,091**

(22) Filed: **Mar. 23, 2023**

Related U.S. Application Data

(63) Continuation of application No. 16/616,041, filed on Nov. 22, 2019, now Pat. No. 11,643,642, filed as application No. PCT/US2018/034073 on May 23, 2018.

(60) Provisional application No. 62/525,404, filed on Jun. 27, 2017, provisional application No. 62/527,199, filed on Jun. 30, 2017.

Publication Classification

(51) **Int. Cl.**
C12N 9/84 (2006.01)

(52) **U.S. Cl.**
CPC ***C12N 9/84*** (2013.01); *C12Y 305/01011* (2013.01)

(57) **ABSTRACT**

The present invention provides engineered penicillin G acylase (PGA) enzymes, polynucleotides encoding the enzymes, compositions comprising the enzymes, and methods of using the engineered PGA enzymes.

Specification includes a Sequence Listing.

PENICILLIN-G ACYLASES

[0001] The present application is a continuation of co-pending U.S. patent application Ser. No. 16/616,041, filed Nov. 22, 2019, which is a national stage application filed under 35 USC §371 and claims priority to PCT International Application No. PCT/US2018/034073, filed May 23, 2018, which claims priority to US Prov. Pat. Appln. Ser. No. 62/525,404, filed Jun. 27, 2017, and to US Prov. Pat. Appln. Ser. No. 62/527,199, filed Jun. 30, 2017, which are hereby incorporated by reference in their entirety for all purposes.

FIELD OF THE INVENTION

[0002] The present invention provides engineered penicillin G acylase (PGA) enzymes, polynucleotides encoding the enzymes, compositions comprising the enzymes, and methods of using the engineered PGA enzymes.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0003] The official copy of the Sequence Listing is submitted concurrently with the specification as a XML file, with a file name of "CX2-163WO1UC1 ST26.xml", a creation date of Mar. 23, 2023, and a size of 249 kilobytes. The Sequence Listing filed is part of the specification and is incorporated in its entirety by reference herein.

BACKGROUND OF THE INVENTION

[0004] Penicillin G acylase (PGA) (penicillin amidase, EC 3.5.1.11) catalyzes the cleavage of the amide bond of the penicillin G (benzylpenicillin) side chain. The enzyme is used commercially in the manufacture of 6-amino-penicillanic acid (6-APA) and phenyl-acetic acid (PAA). 6-APA is a key compound in the industrial production of semi-synthetic β-lactam antibiotics such as amoxicillin, ampicillin and cephalexin. The naturally occurring PGA enzyme shows instability in commercial processes, requiring immobilization on solid substrates for commercial applications. PGA has been covalently bonded to various supports and PGA immobilized systems have been reported as useful tools for the synthesis of pure optical isomers. Attachment to solid surfaces, however, leads to compromised enzyme properties, such as reduced activity and/or selectivity, and limitations to solute access. Moreover, although attachment to solid substrates allows capture of enzymes and reuse in additional processing cycles, the stability of the enzyme is such that such applications may be limited. The enzymatic catalysis by PGA of penicillin G to 6-APA is regiospecific (it does not cleave the lactam amide bond) and stereospecific. The production of 6-APA constitutes perhaps the largest utilization of enzymatic catalysis in the production of pharmaceuticals. The enzymatic activity of PGA, associated with the phenacetyl moiety, allows the stereospecific hydrolysis of a rich variety of phenacetyl derivatives of primary amines as well as alcohols.

SUMMARY OF THE INVENTION

[0005] The present invention provides engineered penicillin G acylase (PGA) enzymes, polynucleotides encoding the enzymes, compositions comprising the enzymes, and methods of using the engineered PGA enzymes.

[0006] The present invention provides engineered penicillin G acylase enzymes capable of removing the A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups from an insulin-dimer to produce a free insulin-dimer, wherein the penicillin G acylase is at least about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some embodiments, the present invention provides engineered penicillin G acylases capable of removing the A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups from insulin-dimer to produce free insulin-dimer, wherein the penicillin G acylase is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some additional embodiments, the present invention provides engineered penicillin G acylases capable of removing the A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups from insulin to produce free insulin-dimer, wherein the penicillin G acylase comprises SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some further embodiments, the penicillin G acylase comprises at least one mutation as provided in Table(s) 6.1, 7.1, 8.1, 8.2, 9.1, 10.1, and/or 11.1.

[0007] The present invention provides engineered penicillin G acylase enzymes capable of removing the A1/A1'-di-phenyl acetate protecting groups from an insulin-dimer to produce a free insulin-dimer, wherein the penicillin G acylase is at least about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some embodiments, the present invention provides engineered penicillin G acylases capable of removing the A1/A1'-di-phenyl acetate protecting groups from insulin-dimer to produce free insulin-dimer, wherein the penicillin G acylase is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some additional embodiments, the present invention provides engineered penicillin G acylases capable of removing the A1/B1/A1'/B1'-di-phenyl acetate protecting groups from insulin to produce free insulin-dimer, wherein the penicillin G acylase comprises SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some further embodiments, the penicillin G acylase comprises at least one mutation as provided in Table(s) 6.1, 7.1, 8.1, 8.2, 9.1, 10.1, and/or 11.1.

[0008] The present invention also provides engineered penicillin G acylase variants having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:6, and at least one substitution at a position selected from positions 54, 62, 115, 125, 127, 127, 185, 253, 254, 254/255, 254/255/370, 255, 256, 257, 257, 260, 268, 322, 325, 348, 369, 370, 372, 373, 377, 378, 384, 384/513/536, 388, 389, 391, 435, 461, 517, 530, 554, 556, 557, 559, 560, 600/623, 623, 624, 626, 627, 705, 706, 707, 723, 740, 748, and 752, wherein said positions are numbered with reference to SEQ ID NO:6. In some embodiments, the engineered penicillin G acylase variants have at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO:6, and or more substitutions selected from 54C, 62G, 115A/P, 125L/T, 127S/V, 185V, 253K/V, 254T, 254W/255G, 254W/255G/370I, 255L, 255M/Q/T/Y, 256Q, 257I, 257V, 260A/P, 268S/V, 322P, 325G, 348C, 348Q, 369L, 369P, 369V, 369W,

370F/G/S, 372A/H/L, 373F/M, 377P, 378H, 384A, 384F/513Q/536M, 384G/L, 388T, 389L, 391P/S, 435R, 461A, 517L/P, 530C/Y, 554A/E/P/V, 556G, 557G/S, 559P/S, 560I, 600T/623V, 623A/G/R/W, 624A, 626G, 627G/H, 705G/P, 706G, 707S, 723A/G, 740L, 748G, and 752E. In some additional embodiments, the the engineered penicillin G acylase variants have at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO:6, and or more substitutions selected from G54C, T62G, T115A/P, N125L/T, 1127S/V, N185V, L253K/V, F254T, F254W/A255G, F254W/A255G/W370I, A255L/M/Q/T/Y, F256Q, L257I/V, G260A/P, D268S/V, K322P, S325G, N348C/Q, K369L/P/V/W, W370F/G/S, S372A/H/L, A373F/M, E377P, T378H, T384A, T384F/P513Q/L536M, T384G, T384L, N388T, I389L, V391P/S, S435R, G461A, A517L/P, S530C/Y, Q554A/E/P/V, Q556G, L557G/S, Q559P/S, T560I, M600T/D623V, D623A/G/R/W, I624A, Q626G, N627G/H, T705G/P, 5706G, E7075, K723A/G, S740L, R748G, and Y752E

[0009] The present invention also provides engineered penicillin G acylase variants having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:8, and at least one substitution set selected from positions 103/370/444/706/766, 103/369/370/442/444/536/556/766, 103/369/370/444, 103/369/370/444/556/706/766, 103/369/370/444/765/766, 103/369/370/765/766, 257/362/384/451, 257/362/384/451/723, 362/451/705, 369/370, 369/370/444/706/766, 369/370/556/766, 369/370/388/444/556/766, 369/370/444, 369/370/444/556/766, 369/370/556, 369/370/556/765, 369/370/556/766, 369/370/766, 369/370/444/556, 369/370/444/556/612/766, 369/370/444/556/706/765, 369/370/444/706/765/766, 372/373/384/513/560, 372/384/451/705, 372/384/560/705, 384/451/560/705/723, 384/451/705/723, 451/560/705/723, and 451/705/723, wherein said positions are numbered with reference to SEQ ID NO:8. In some additional embodiments, the engineered penicillin G acylase variants comprise a substitution set is selected from 103V/370F/444S/706G/766G, 103V/369W/370F/442I/444S/536M/556G/766G, 103V/369W/370F/444S, 103V/369W/370F/444S/556G/706G/766G, 103V/369W/370F/444S/765P/766G, 103V/369W/370F/765P/766G, 257V/362V/384A/451R, 257V/362V/384L/451R/723L, 362V/451R/705D, 369P/370F, 369P/370F/444S/706G/766G, 369P/370F/556G/766G, 369V/370F/388T/444S/556G/766G, 369V/370F/444S, 369V/370F/444S/556G/766G, 369V/370F/556G, 369V/370F/556G/765P, 369V/370F/556G/766G, 369V/370F/766G, 369W/370F/444S/556G, 369W/370F/444S/556G/612A/766G, 369W/370F/444S/556G/706G/765P, 369W/370F/444S/706G/765P/766G, 372A/373M/384L/513Q/560G, 372A/384L/451R/705D, 372A/384L/560G/705D, 384A/451R/560G/705D/723L, 384L/451R/705D/723L, 451R/560G/705D/723L, and 451R/705D/723L, wherein said positions are numbered with reference to SEQ ID NO:8. In some additional embodiments, the engineered penicillin G acylase variants comprise a substitution set selected from K103V/I370F/G444S/S706G/H766G, K103V/K369W/I370F/V442I/G444S/L536M/Q556G/H766G, K103V/K369W/I370F/G444S, K103V/K369W/I370F/G444S/Q556G/S706G/H766G, K103V/K369W/I370F/G444S/H765P/H766G, K103V/K369W/I370F/H765P/H766G, L257V/A362V/T384A/A451R, L257V/A362V/T384L/A451R/K723L, A362V/A451R/T705D, K369P/I370F, K369P/1370F/G444S/S706G/H766G, K369P/1370F/Q556G/H766G, K369V/I370F/N388T/G444S/Q556G/H766G, K369V/I370F/G444S, K369V/I370F/G444S/Q556G/H766G, K369V/I370F/Q556G, K369V/I370F/Q556G/H765P, K369V/I370F/Q556G/H766G, K369V/I370F/H766G, K369W/I370F/G444S/Q556G, K369W/I370F/G444S/Q556G/V612A/H766G, K369W/I370F/G444S/Q556G/5706G/H765P, K369W/I370F/G444S/S706G/H765P/H766G, S372A/A373M/T384L/P513Q/T560G, S372A/T384L/A451R/T705D, S372A/T384L/T560G/T705D, T384A/A451R/T560G/T705D/K723L, T384L/A451R/T705D/K723L, A451R/T560G/T705D/K723L, and A451R/T705D/K723L, wherein said positions are numbered with reference to SEQ ID NO:8

[0010] The present invention also provides engineered penicillin G acylase variants having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:34, and at least one substitution at a position selected from positions 55, 275, 403, 482, 496, 541, 616, 619/664, 622, 639, 664, 747, and 759, wherein said positions are numbered with reference to SEQ ID NO:34. In some embodiments the engineered penicillin G acylase variants comprise at least one substitution or substitution set selected from 55V, 275E, 403T, 482A/S, 496K, 541A, 616G, 619N/664G, 622R, 639G, 664G, 747G, and 759N, wherein said positions are numbered with reference to SEQ ID NO:34. In some embodiments of the engineered penicillin G acylase variants comprise at least one substitution or substitution set selected from L55V, P275E, D403T, E482A/S, P496K, Q541A, A616G, K619N/A664G, K622R, S639G, A664G, A747G, and Q759N, wherein said positions are numbered with reference to SEQ ID NO:34.

[0011] The present invention also provides engineered penicillin G acylase variants comprising at least one substitution set selected from 103/372/373/557, 253/322/369/623, 253/254/322/369/623, 253/254/369/391/623/723, 253/254/369/619/623/723, 253/254/369/623/723, 253/254/373/623/723, 253/254/255/369/623/723, 253/254/369, 253/322/369/373/723, 253/369/623/723, 253/373/623, 253/254/255/322/369/619/723, 260/372/373/556, 260/372/373/556/557/559, 322/369, 322/369/373/723, 322/369/623/723, and 369/373/556, wherein said positions are numbered with reference to SEQ ID NO:34. In some further embodiments, the engineered penicillin G acylase variants comprise a substitution set is selected from 103V/372S/373F/557G, 253H/322T/369W/623G, 253H/254Q/322T/369W/623G, 253H/254Q/369W/391A/623G/723A, 253H/254Q/369W/619R/623G/723A, 253H/254Q/369W/623G/723A, 253H/254Q/373L/623G/723A, 253H/254S/255V/369W/623S/723A, 253H/254S/369W, 253H/322T/369W/373W/723A, 253H/369W/623G/723A, 253H/373L/623S, 253S/254S/255V/322T/369W/619R/723A, 260S/372S/373F/556G, 260S/372S/373F/556G/557V/559S, 322T/369W, 322T/369W/373W/723A, 322T/369W/623G/723A, and 369W/373F/556G, wherein said positions are numbered with reference to SEQ ID NO:34. In some additional embodiments, the penicillin G acylase variants comprise a substitution set selected from K103V/A372S/M373F/L557G, L253H/K322T/K369W/D623G, L253H/W254Q/K322T/K369W/D623G, L253H/W254Q/K369W/V391A/D623G/K723A, L253H/W254Q/K369W/K619R/D623G/K723A, L253H/W254Q/K369W/D623G/K723A, L253H/W254Q/M373L/D623G/K723A, L253H/W254S/G255V/K369W/D623S/K723A, L253H/W254S/K369W, L253H/K322T/K369W/M373W/K723A,

L253H/K369W/D623G/K723A, L253H/M373L/D623S, L253S/W254S/G255V/K322T/K369W/K619R/K723A, G260S/A372S/M373F/Q556G, G260S/A372S/M373F/Q556G/L557V/Q559S, K322T/K369W, K322T/K369W/M373W/K723A, K322T/K369W/D623G/K723A, and K369W/M373F/Q556G, wherein said positions are numbered with reference to SEQ ID NO:34.

[0012] The present invention also provides engineered penicillin G acylase variants having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:46, and at least one substitution or substitution set at a position selected from positions 9/25/103/253/348/444/557/623, 9/103/322/391/444/557/623, 9/103/253/322/348/444/556/557/623, 25/103/241/253/322/348/444/556/557/623, 28, 71, 77, 103/257/260/322/348/384/444/556/623, 103/257/260/322/348/444/557, 103/322/348/373/444/556/557, 103/260/322/348/444/556/557/623, 103/260/322/348/444/623, 103/260/322/348/373/391/444/556/557/623, 111, 128, 129, 131, 146/309/556/619/748, 176/233/373/619/664, 176/373/482/569, 176/373/482/622/664, 176/373/482/622, 176/373/482/569/622/623/764, 176/482, 225/304/309/556/557/619/748, 225/304/322/494/496/616/619/664/747/756, 225/304/494/616/619/664/747/759, 233/275/482/569/664, 233/275/482/619, 233/373/482/622/664, 304/496/616/619/664/747/756/759, 322/348/373/391/444/556/623, 322/348/444/557, 369, 369/764, 370, 373/482/569/619/764S, 373, 379D, 380, 389, 451, 471, 482/623, 494/496/616/619/664, 616, 617, 619, 622, 626, and 705, wherein said positions are numbered with reference to SEQ ID NO:46. In some embodiments, the engineered penicillin G acylase variants comprise at least one substitution or substitution set selected from 9K/25V/103V/253S/348A/444S/557G/623D, 9K/103V/322T/391A/444S/557G/623D, 9K/103V/253S/322T/348A/444S/556G/557G/623D, 25V/103V/241K/253S/322T/348A/444S/556G/557G/623D, 28A/C/Q/S, A71C/F/G,L, 77T/V, 103V/257V/260S/322T/348A/384T/444S/556G/623D, 103V/257V/260S/322T/348A/444S/557G, 103V/322T/348A/373A/444S/556G/557G, 103V/260S, 322T/348A/444S/556G/557G/623D, 103V/260S/322T/348A/444S/623D, 103V/260S/322T/348A/373A/391A/444S/556G/557G/623D, 111S, 128H, 129E, 131D, 146M/309D/556N/619S/748A, 176S/233E/373A/619N/664R, 176S/373F/482A/569W, 176S/373F/482A/622F/664G, 176S/373F/482A/622V, 1765/373F/482C/569W/622C/623D/7645, 176S/482A, 225K/304C/309V/556N/557R/619S/748A, 225T/3041/322T/494E/496N/616G/619N/664G/747S/756P, 225T/3041/494E/616G/619N/664G/747P/759E, 233E/275E/482A/569W/664G, 233E/275E/482C/619N 233E/373F/482A/622V/664G, 3041/496K/6165/619N/664E/747P/756P/759E, 322T/348A/373A/391A/444S/556G/623D, 322T/348A/444S/557G, 369A/E/L, 369, 369L/764G, 369V, I370M/Q, 373A/482C/569W/619N/764S, 373G, 379D/S, 380D, 389V, 451H, 471V 482S/623D, 494E/496K/616S/619N/664E, 616D/E/G/N/Q/T, 617W, 619A/H/L/P/S/V, 622I/V, 626D/E, and 705N, wherein said positions are numbered with reference to SEQ ID NO: 46. In some additional embodiments, the engineered penicillin G acylase variants comprise at least one substitution or substitution set selected from N9K/G25V/K103V/H253S/N348A/G444S/L557G/G623D, N9K/K103V/K322T/V391A/G444S/L557G/G623D, N9K/K103V/H253S/K322T/N348A/G444S/Q556G/L557G/G623D, G25V/K103V/N241K/H253S/K322T/N348A/G444S/Q556G/L557G/G623D, N28A/C/Q/S, A71C/F/G/L, I77T/V, K103V/L257V/G260S/K322T/N348A/L384T/G444S/Q556G/G623D, K103V/L257V/G2605/K322T/N348A/G444S/L557G, K103V/K322T/N348A/M373A/G444S/Q556G/L557G, K103V/G260S/K322T/N348A/G444S/Q556G/L557G/G623D, K103V/G260S/K322T/N348A/G444S/G623D, K103V/G260S/K322T/N348A/M373A/V391A/G444S/Q556G/L557G/G623D, G111S, K128H, T129E, T131D, K146M/N309D/Q556N/K619S/R748A, T176S/Q233E/M373A/K619N/A664R T176S/M373F/E482A/Y569W, T176S/M373F/E482A/K622F/A664G, T176S/M373F/E482A/K622V, T176S/M373F/E482C/Y569W/K622C/G623D/A764S, T176S/E482A, L225K/K304C/N309V/Q556N/L557R/K619S/R748A, L225T/K304I/K322T/N494E/P496N/A616G/K619N/A664G/A747S/F756P, L225T/K304I/N494E/A616G/K619N/A664G/A747P/Q759E, Q233E/P275E/E482A/Y569W/A664G, Q233E/P275E/E482C/K619N, Q233E/M373F/E482A/K622V/A664G, K304I/P496K/A616S/K619N/A664E/A747P/F756P/Q759E, K322T/N348A/M373A/V391A/G444S/Q556G/G623D, K322T/N348A/G444S/L557G, W369A/E/L, W369L/A764G, W369V, I370M/Q, M373A/E482C/Y569W/K619N/A764S, M373G, T379D/S, Q380D, I389V, A451H, R471V, E482S/G623D, N494E/P496K/A616S/K619N/A664E, A616D/E/G/N/Q/T, F617W, K619A/H/L/P/S/V, K622I, K622V, Q626D/E, and T705N, wherein said positions are numbered with reference to SEQ ID NO:46.

[0013] The present invention also provides engineered penicillin G acylase variants having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:54, and at least one substitution or substitution set at a position selected from positions 28/71/128/176/619/664, 28/71/128/176/626/753, 71, 71/176/233/260/275/482/619/759, 71/176/233/482, 71/176/260/451/619, 71/176/275/482, 71/176/275/664, 71, 71/128/176/373/482/496/619, 71/128/176/373/482/496/569, 71/128/176/482/496, 71/128/176/496/664, 71/128/373/482/664/753, 71/176/233/260/451/482/664/759, 71/176/233/373/482/569/753, 71/176/260/275/482/557/759, 71/176/260/482, 71/176/260/482/557/619/664, 71/176/373/626/664/753, 71/176/451/482/619/759, 71/176/482, 71/176/482/619/664/759, 71/233/260/482/557/759, 71/260/451/482/664/759, 71/373/756, 82, 122, 126, 128/176/233/373/482/626/753, 128/176/233/496/664/753, 128/176/373/482/664, 128/176/373/496/753, 176/233/260/275/482/664/759, 176/233/451/482/619/664/759, 180, 184, 472, 496, 658, 679, 686, and 739, wherein said positions are numbered with reference to SEQ ID NO:54. In some embodiments, the engineered penicillin G acylase variants comprise at least one substitution or substitution set selected from 28A/71L/128H/176S/619N/664E, 28A/71L/128H/176S/626D/753C, 71F/176S/233E/260G/275C/482S/619N/759D, 71F/176S/233E/482A, 71F/176S/260G/451H/619V, 71F/176S/275C/482S, 71F/176S/275E/664D, 71F/G/I/L, 71L/128H/176S/373A/482C/496K/619S, 71L/128H/176S/373A/482S/496K/569C, 71L/128H/176S/482S/496K, 71L/128H/176S/496K/664E, 71L/128H/373A/482S/664E/753C, 71L/176S/233E/260G/451H/482S/664C/759E, 71L/176S/233E/373A/482C/569C/753C, 71L/176S/260G/275C/482A/557G/759E, 71L/176S/260G/482A, 71L/176S/260G/482A/557G/619P/664D, 71L/176S/373A/626E/664E/753C, 71L/176S/451H/482A/619V/759E, 71L/176S/482A, 71L/176S/482A/619P/664D/759E, 71L/233E/260G/482A/557G/759E, 71L/260G/451H/482A/664D/759E, 71L/373A/756C, 71M/V,

82V, 122M, 126L, 128H/176S/233E/373A/482S/626E/753C, 128H/176S/233E/496K/664E/753C, 128H/176S/373A/482S/664E, 128H/176S/373A/496K/753C, 176S/233E/260G/275E/482C/664E/759D, 176S/233E/451H/482S/619N/664C/759D, 180F, 184A/F, 472F/V, 496K, 658C, 679L, 686A, and P739D/S, wherein said positions are numbered with reference to SEQ ID NO:54. In some additional embodiments, the engineered penicillin G acylase variants comprise at least one substitution or substitution set selected from N28A/A71L/K128H/T176S/K619N/A664E, N28A/A71L/K128H/T176S/Q626D/P753C, A71F/T1765/Q233E/S260G/P275C/E482S/K619N/Q759D, A71F/T1765/Q233E/E482A, A71F/T176S/S260G/A451H/K619V, A71F/T176S/P275C/E482S, A71F/T176S/P275E/A664D, A71F/G/I/L, A71L/K128H/T176S/M373A/E482C/P496K/K619S, A71L/K128H/T176S/M373A/E482S/P496K/Y569C, A71L/K128H/T176S/E482S/P496K, A71L/K128H/T176S/P496K/A664E, A71L/K128H/M373A/E482S/A664E/P753C, A71L/T176S/Q233E/S260G/A451H/E482S/A664C/Q759E, A71L/T176S/Q233E/M373A/E482C/Y569C/P753C, A71L/T176S/S260G/P275C/E482A/L557G/Q759E, A71L/T176S/S260G/E482A, A71L/T176S/S260G/E482A/L557G/K619P/A664D, A71L/T176S/M373A/Q626E/A664E/P753C A71L/T176S/A451H/E482A/K619V/Q759E, A71L/T176S/E482A, A71L/T176S/E482A/K619P/A664D/Q759E, A71L/Q233E/S260G/E482A/L557G/Q759E, A71L/S260G/A451H/E482A/A664D/Q759E, A71L/M373A/F756C, A71M/V, L82V, L122M, V126L, K128H/T176S/Q233E/M373A/E482S/Q626E/P753C, K128H/T176S/Q233E/P496K/A664E/P753C, K128H/T176S/M373A/E482S/A664E, K128H/T176S/M373A/P496K/P753C, T176S/Q233E/S260G/P275E/E482C/A664E/Q759D, T176S/Q233E/A451H/E482S/K619N/A664C/Q759D, Y180F, V184A/F, H472FN, P496K, W658C, F679L, P686A, and P739D/S, wherein said positions are numbered with reference to SEQ ID NO:54.

[0014] The present invention also provides engineered penicillin G acylase variants having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:74, and at least one substitution or substitution set at a position selected from positions 71/352, 71/451/748, 71/353/357/451/705/748, 71/451/556/705/748, 77/176/712, 111/176/352, 176, 176/233, 176/233/352, 176/233/557/619/759, 176/233/759, 176/275/348/557/759, 176/275/569/759, 176/275/557/759, 176/275, 176/348/557/569/616, 176/352, 176/361, 176/482/616/759, 176/557/616, 176/557/708, 176/557/569/616/708, 176/557/569/616, 176/557/619, 176/569/616/619/759, 176/616, 176/616/619, 176/616, and 176/Q759, wherein said positions are numbered with reference to SEQ ID NO:74. In some embodiments, the engineered penicillin G acylase variants comprise at least one substitution or substitution set selected from 71C/352S, 71C/451H/748A, 71F/353A/357A/451H/705N/748A, 71F/451H/556N/705N/748A, 77T/176S/712V, 111S/176S/352S, 176S, 176S/233E, 176S/233E/352S, 176S/233E/557G/619G/759D, 176S/233E/759E, 176S/275C/348M/557G/759D, 176S/275C/569W/759D, 176S/275E/557G/759E, 176S/275E, 176S/348M/557G/569W/616G, 176S/352S, 176S/361T, 176S/482C/616G/759E, 176S/557G/616N, 176S/557G/708L, 176S/557G/569W/616G/708L, 176S/557G/569W/616T, 176S/557G/619G, 176S/569W/616G/619S/759D, 176S/616G, 176S/616S, 176S/616G/619R, 176S/616T, and 176S/759D, wherein said positions are numbered with reference to SEQ ID NO:74. In some additional embodiments, the engineered penicillin G acylase variants comprise at least one substitution or substitution set selected from L71C/T352S, L71C/A451H/R748A, L71F/S353A/R357A/A451H/T705N/R748A, L71F/A451H/Q556N/T705N/R748A, I77T/T176S/A712V, G111S/T176S/T352S, T176S, T176S/Q233E, T176S/Q233E/T352S, T176S/Q233E/L557G/K619G/Q759D, T176S/Q233E/Q759E, T176S/P275C/A348M/L557G/Q759D, T176S/P275C/Y569W/Q759D, T176S/P275E/L557G/Q759E, T176S/P275E, T176S/A348M/L557G/Y569W/A616G, T176S/T352S, T176S/A361T, T176S/S482C/A616G/Q759E, T176S/L557G/A616N, T176S/L557G/I708L, T176S/L557G/Y569W/A616G/I708L, T176S/L557G/Y569W/A616T, T176S/L557G/K619G, T176S/Y569W/A616G/K619S/Q759D, T176S/A616G, T176S/A616S, T176S/A616G/K619R, T176S/A616T, and T176S/Q759D, wherein said positions are numbered with reference to SEQ ID NO:74.

[0015] The present invention also provides engineered penicillin G acylase variants wherein said engineered penicillin G acylases comprise a histidine tag. In some embodiments, the histidine tag is present at the C-terminus of said engineered penicillin G acylase variants.

[0016] The present invention also provides engineered penicillin G acylase variants comprising a polypeptide sequence set forth in variant numbers 1-308. In some additional embodiments, the engineered penicillin G acylase variants comprise a polypeptide sequence selected from SEQ ID NOS: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88.

[0017] The present invention also provides engineered penicillin G acylase variants that are capable of deacylating protected insulin. In some embodiments, the protected insulin comprises di-protected insulin, while in some embodiments, the protected insulin comprises tetra-protected insulin. In some further embodiments, the protected insulin is a dimer, while in some alternative embodiments, the protected insulin is a tetramer.

[0018] The present invention also provides engineered polynucleotide sequences encoding the engineered penicillin G acylase variants provided herein. In some embodiments, the engineered polynucleotide sequences encode the penicillin G acylases set forth in variant numbers 1-308. In some embodiments, the engineered polynucleotide sequence comprises a polynucleotide sequence that is at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to a sequence selected from SEQ ID NOS: 3, 5, 7, 33, 45, 53, 73 and 87.

[0019] The present invention also provides a penicillin G acylase variants encoded by polynucleotide sequences having at least about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more sequence identity to at least one sequence selected from SEQ ID NOS:3, 5, 7, 33, 45, 53, 73 and 87.

[0020] In some embodiments, a penicillin G acylase variant encoded by a polynucleotide sequence has at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to a sequence selected from SEQ ID NOS:3, 5, 7, 33, 45, 53, 73 and 87. In some embodiments, the penicillin G acylase variant is encoded by a polynucleotide sequence selected from SEQ ID NOS: 3, 5, 7, 33, 45, 53, 73 and 87.

[0021] The present invention also provides vectors comprising the engineered polynucleotide sequences encoding penicillin G acylase variants provided herein. In some embodiments, the vectors comprise the polynucleotide sequences provided herein (e.g., SEQ ID NOS: 3, 5, 7, 33, 45, 53, 73 and 87). The present invention also provides host cells comprising the vectors provided herein (e.g., vectors comprising the polynucleotide sequences of SEQ ID NOS: 3, 5, 7, 33, 45, 53, 73 and 87). In some embodiments, the vectors further comprise at least one control sequence. In some embodiments, the control sequence is a promoter. In some further embodiments, the promoter is a heterologous promoter. It is not intended that the present invention be limited to control sequences comprising promoters, as any suitable and/or desirable control sequence finds use in the present invention.

[0022] The present invention further provides host cells comprising at least one of the vectors provided herein. In some embodiments, the host cell is eukaryotic, while in some alternative embodiments, the host cell is prokaryotic. In some embodiments, the host cell is *E. coli.*

[0023] The present invention also provides methods for production of the engineered penicillin G acylase variants provided herein, comprising culturing a host cell comprising a vector and/or polynucleotide encoding at least one engineered penicillin G acylase variant, under conditions that said engineered penicillin G acylase variant is produced by said host cell. In some embodiments, the methods further comprise the step of recovering the engineered penicillin G acylase variant produced by the host cell.

[0024] The present invention also provides methods for producing free insulin, comprising i) providing at least one engineered penicillin G acylase variant provided herein and a composition comprising insulin-dimer comprising A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups; and ii) exposing the engineered penicillin G acylase variant to the insulin-dimer comprising A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups, under conditions such that the engineered penicillin G acylase variant removes the A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups and free insulin is produced.

[0025] The present invention also provides methods for producing free insulin, comprising i) providing at least one engineered penicillin G acylase variant provided herein and a composition comprising insulin-dimer comprising A1/A1'-di-phenyl acetate protecting groups; and ii) exposing the engineered penicillin G acylase variant to the insulin-dimer comprising A1/A1'-di-phenyl acetate protecting groups, under conditions such that the engineered penicillin G acylase variant removes the A1/A1'-di-phenyl acetate protecting groups and free insulin is produced.

[0026] The present invention also provides methods for producing free insulin-dimer, comprising: i) providing at least one engineered penicillin G acylase provided herein, and insulin-dimer comprising A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups; and ii) exposing the engineered penicillin G acylase to the insulin-dimer comprising A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups, under conditions such that the engineered penicillin G acylase removes the A1/B1/A1'/B1'-tetra-phenyl acetate protecting groups and free insulin-dimer is produced. In some embodiments of the methods, the penicillin G acylase is at least about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some embodiments of the methods, the penicillin G acylase is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some further embodiments of the methods, the penicillin G acylase comprises SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some embodiments, the engineered penicillin G acylase produces more than 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more free insulin-dimer. The present invention also provides compositions comprising free insulin-dimer produced according to the method(s) provided herein.

[0027] The present invention also provides methods for producing free insulin-dimer, comprising: i) providing at least one engineered penicillin G acylase variant provided herein, and insulin-dimer comprising A1/A1'-di-phenyl acetate protecting groups; and ii) exposing the engineered penicillin G acylase variant to the insulin-dimer comprising A1/A1'-di-phenyl acetate protecting groups, under conditions such that the engineered penicillin G acylase variant removes the A1/A1'-di-phenyl acetate protecting groups and free insulin-dimer is produced. In some embodiments of the methods, the engineered penicillin G acylase variant is at least about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some embodiments of the methods, the engineered penicillin G acylase variant is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some further embodiments of the methods, the engineered penicillin G acylase variant comprises SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some embodiments, the engineered penicillin G acylase variant produces more than 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more free insulin-dimer. The present invention also provides compositions comprising free insulin-dimer produced according to the method(s) provided herein.

[0028] The present invention also provides methods for producing free insulin, comprising i) providing at least one engineered penicillin G acylase variant provided herein, and a composition comprising insulin comprising a A1/A1'-di-phenyl acetate protecting groups; and ii) exposing the engineered penicillin G acylase variant to the insulin comprising A1/A1'-di-phenyl acetate protecting groups, under conditions such that the engineered penicillin G acylase variant removes the A1/A1'-di-phenyl acetate protecting groups and free insulin is produced. In some embodiments of the methods, the engineered penicillin G acylase variant is at least about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some embodiments of the methods, the engineered penicillin G acylase variant is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88. In some further embodiments of the methods, the engineered penicillin G acylase variant comprises SEQ ID NOs: 2, 4, 6, 8, 34, 46, 54, 74 and/or 88.

In some embodiments, the engineered penicillin G acylase variant produces more than 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more free insulin. The present invention also provides compositions comprising free insulin produced according to the method(s) provided herein.

DESCRIPTION OF THE INVENTION

[0029] The present invention provides engineered penicillin G acylases (PGA; i.e., pencillin G acylase variants) that are capable of cleaving penicillin to phenylacetic acid and 6-aminopenicillanic acid (6-APA), which is a key intermediate in the synthesis of a large variety of β-lactam antibiotics. In particular, the present invention provides engineered PGAs that are capable of removing the A1/B1/A1/B1' tetra-phenyl acetate protecting groups from an insulin-dimer to produce a free insulin-dimer. In some additional embodiments, the PGAs are capable of removing the A1/A1' di-phenyl acetate protecting groups from an insulin dimer to produce free insulin.

[0030] Generally, naturally occurring PGAs are heterodimeric enzymes composed of an alpha subunit and a beta-subunit. Wild-type PGA is naturally synthesized as a pre-pro-PGA polypeptide, containing an N-terminal signal peptide that mediates translocation to the periplasm and a linker region connecting the C-terminus of the alpha subunit to the N-terminus of the beta subunit. Proteolytic processing leads to the mature heterodimeric enzyme. The intermolecular linker region can also function in promoting proper folding of the enzyme. The PGAs provided herein are based on the PGA from *Kluyvera citrophila* in which various modifications have been introduced to generate improved enzymatic properties as described in detail below.

[0031] For the descriptions provided herein, the use of the singular includes the plural (and vice versa) unless specifically stated otherwise. For instance, the singular forms "a", "an" and "the" include plural referents unless the context clearly indicates otherwise. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0032] 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."

[0033] Both the foregoing general description, including the drawings, and the following detailed description are exemplary and explanatory only and are not restrictive of this disclosure. Moreover, the section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described.

Definitions

[0034] As used herein, the following terms are intended to have the following meanings. In reference to the present disclosure, 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. In addition, all patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

[0035] Unless otherwise indicated, the practice of the present invention involves conventional techniques commonly used in molecular biology, fermentation, microbiology, and related fields, which are known to those of skill in the art. Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. Indeed, it is intended that the present invention not be limited to the particular methodology, protocols, and reagents described herein, as these may vary, depending upon the context in which they are used. The headings provided herein are not limitations of the various aspects or embodiments of the present invention that can be had by reference to the specification as a whole. Accordingly, the terms defined below are more fully defined by reference to the specification as a whole.

[0036] Nonetheless, in order to facilitate understanding of the present invention, a number of terms are defined below. 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.

[0037] 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).

[0038] As used herein and in the appended claims, the singular "a", "an" and "the" include the plural reference unless the context clearly dictates otherwise. Thus, for example, reference to a "host cell" includes a plurality of such host cells.

[0039] Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation and amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[0040] As used herein, the terms "protein", "polypeptide", and "peptide" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids.

[0041] As used herein, "polynucleotide" and "nucleic acid" refer to two or more nucleosides that are covalently linked together. The polynucleotide may be wholly comprised ribonucleosides (i.e., an RNA), wholly comprised of 2' deoxyribonucleotides (i.e., a DNA) or mixtures of ribo- and 2' deoxyribonucleosides. 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 (e.g., inosine, xanthine, hypoxanthine, etc.). Preferably, such modified or synthetic nucleobases will be encoding nucleobases.

[0042] 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 T 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, 5x Denhart's solution, 5×SSPE, 0.2% SDS at 42° C., followed by washing in 0.1xSSPE, 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 0.1×SSC containing 0.1% SDS at 65° C. Other high stringency hybridization conditions, as well as moderately stringent conditions, are known to those of skill in the art.

[0043] 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.

[0044] 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. In some embodiments, the polynucleotides encoding the PGA enzymes may be codon optimized for optimal production from the host organism selected for expression. 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 PGAs enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0045] As used herein, "preferred, optimal, high codon usage bias codons" refers interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (See e.g., GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, John Peden, University of Nottingham; McInerney, Bioinform., 14:372-73 [1998]; Stenico et al., Nucleic Acids Res., 222:437-46 [1994]; and Wright, Gene 87:23-29 [1990]). Codon usage tables are available for a growing list of organisms (See e.g., Wada et al., Nucleic Acids Res., 20:2111-2118 [1992]; Nakamura et al., Nucl. Acids Res., 28:292 [2000]; Duret, et al., supra; Henaut and Danchin, "*Escherichia coli* and Salmonella," Neidhardt, et al. (eds.), ASM Press, Washington D.C., [1996], p. 2047-2066. 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., Uberbacher, Meth. Enzymol., 266:259-281 [1996]; Tiwari et al., Comput. Appl. Biosci., 13:263-270 [1997]).

[0046] As used herein, "control sequence" is defined herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present invention. Each control sequence may be native or foreign to the polynucleotide of interest. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator.

[0047] As used herein, "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.

[0048] As used herein, "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 control sequence may comprise an appropriate promoter 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.

[0049] As used herein, "naturally occurring" and "wild-type" refers to the form found in nature. For example, a naturally occurring or wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

[0050] As used herein, "non-naturally occurring," "engineered," and "recombinant" when used in the present disclosure with reference to (e.g., 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 material is identical to naturally occurring material, 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.

[0051] As used herein, "percentage of sequence identity," "percent identity," and "percent identical" refer to comparisons between polynucleotide sequences or polypeptide sequences, 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 is 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. Determination of optimal alignment and percent sequence identity is performed using the BLAST and BLAST 2.0 algorithms (See e.g., Altschul et al., J. Mol. Biol. 215: 403-410 [1990]; and Altschul et al., Nucl. Acids Res., 25:3389-3402 [1977]). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website.

[0052] Briefly, the BLAST analyses involve first identifying high scoring sequence pairs (HSPs) by identifying short words of length Win the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as, the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (See e.g., Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 [1989]).

[0053] Numerous other algorithms are available and known in the art that function similarly to BLAST in providing percent identity for two sequences. Optimal alignment of sequences for comparison can be conducted using any suitable method known in the art (e.g., by the local homology algorithm of Smith and Waterman, Adv. Appl. Math. 2:482 [1981]; by the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol. 48:443 [1970]; by the search for similarity method of Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85:2444 [1988]; and/or by computerized implementations of these algorithms [GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package]), or by visual inspection, using methods commonly known in the art. Additionally, determination of sequence alignment and percent sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison Wisc.), using the default parameters provided.

[0054] 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 and 89 to 95 percent sequence identity, 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 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 preferred embodiments, residue positions that are not identical differ by conservative amino acid substitutions.

[0055] As used herein, "reference sequence" refers to a defined sequence to which another sequence is compared. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, or the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a sequence (i.e., a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptide are typically performed by comparing sequences of the two polynucleotides over a comparison window to identify and compare local regions of sequence similarity. The term "reference sequence" is not intended to be limited to wild-type sequences, and can include engineered or altered sequences. For example, in some embodiments, a "reference sequence" can be a previously engineered or altered amino acid sequence.

[0056] As used herein, "comparison window" refers to a conceptual segment of at least about 20 contiguous nucleo-

tide positions or amino acids 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.

[0057] 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 refers 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 PGA, 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. As used herein, a reference to a residue position, such as "Xn" as further described below, is to be construed as referring to "a residue corresponding to", unless specifically denoted otherwise. Thus, for example, "X94" refers to any amino acid at position 94 in a polypeptide sequence.

[0058] As used herein, "improved enzyme property" refers to a PGA that exhibits an improvement in any enzyme property as compared to a reference PGA. For the engineered PGA polypeptides described herein, the comparison is generally made to the wild-type PGA enzyme, although in some embodiments, the reference PGA can be another improved engineered PGA. Enzyme properties for which improvement is desirable include, but are not limited to, enzymatic activity (which can be expressed in terms of percent conversion of the substrate at a specified reaction time using a specified amount of PGA), chemoselectivity, thermal stability, solvent stability, pH activity profile, cofactor requirements, refractoriness to inhibitors (e.g., product inhibition), stereospecificity, and stereoselectivity (including enantioselectivity).

[0059] As used herein, "increased enzymatic activity" refers to an improved property of the engineered PGA 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 PGA) as compared to the reference PGA enzyme. Exemplary methods to determine enzyme activity are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m, V_{mas} or k_{cat}, changes of which can lead to increased enzymatic activity. Improvements in enzyme activity can be from about 1.5 times the enzymatic activity of the corresponding wild-type PGA enzyme, to as much as 2 times. 5 times, 10 times, 20 times, 25 times, 50 times, 75 times, 100 times, or more enzymatic activity than the naturally occurring PGA or another engineered PGA from which the PGA polypeptides were derived. In specific embodiments, the engineered PGA enzyme exhibits improved enzymatic activity in the range of 1.5 to 50 times, 1.5 to 100 times greater than that of the parent PGA enzyme. It is understood by the skilled artisan that the activity of any enzyme is diffusion limited such that the catalytic turnover rate cannot exceed the diffusion rate of the substrate, including any required cofactors. The theoretical maximum of the diffusion limit, or k_{cat}/K_m, is generally about 10^8 to 10^9 (M^{-1} s^{-1}). Hence, any improvements in the enzyme activity of the PGA will have an upper limit related to the diffusion rate of the substrates acted on by the PGA enzyme. PGA activity can be measured by any one of standard assays used for measuring the release of phenylacetic acid upon cleavage of penicillin G, such as by titration (See e.g., Simons and Gibson, Biotechnol. Tech.,13:365-367 [1999]). In some embodiments, the PGA activity can be measured by using 6-nitrophenylacetamido benzoic acid (NIPAB), which cleavage product 5-amino-2-nitro-benzoic acid is detectable spectrophotometrically (λmax=405 nm). Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0060] As used herein, "increased enzymatic activity" and "increased activity" refer to an improved property of an engineered enzyme, 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 PGA) as compared to a reference enzyme as described herein. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m, V_{max} or k_{cat}, changes of which can lead to increased enzymatic activity. In some embodiments, the PGA enzymes provided herein frees insulin by removing tri-phenyl acetate protecting groups from specific residues of insulin. Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when enzymes in cell lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

[0061] As used herein, "conversion" refers to the enzymatic transformation of a substrate to the corresponding product.

[0062] As used herein "percent conversion" refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, for example, the "enzymatic activity" or "activity" of a PGA polypeptide can be expressed as "percent conversion" of the substrate to the product.

[0063] As used herein, "chemoselectivity" refers to the preferential formation in a chemical or enzymatic reaction of one product over another.

[0064] As used herein, "thermostable" and "thermal stable" are used interchangeably to refer to a polypeptide that is resistant to inactivation when exposed to a set of temperature conditions (e.g., 40-80° C.) for a period of time (e.g., 0.5-24 hrs) compared to the untreated enzyme, thus retaining a certain level of residual activity (e.g., more than 60% to 80%) after exposure to elevated temperatures.

[0065] As used herein, "solvent stable" refers to the ability of a polypeptide to maintain similar activity (e.g., more than e.g., 60% to 80%) after exposure to varying concentrations (e.g., 5-99%) of solvent (e.g., isopropyl alcohol, tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butylacetate, methyl tert-butylether, etc.) for a period of time (e.g., 0.5-24 hrs) compared to the untreated enzyme.

[0066] As used herein, "pH stable" refers to a PGA 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.

[0067] As used herein, "thermo- and solvent stable" refers to a PGA polypeptide that is both thermostable and solvent stable.

[0068] 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).

[0069] As used herein, "acidic amino acid or residue" refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK 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).

[0070] As used herein, "basic amino acid or residue" refers to a hydrophilic amino acid or residue having a side chain exhibiting a pK 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).

[0071] 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).

[0072] 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) 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).

[0073] 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).

[0074] 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.

[0075] 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).

[0076] 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.

[0077] 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 a-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).

[0078] 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).

[0079] 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. 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 X40 as compared to SEQ ID

NO:2" refers to a difference of the amino acid residue at the polypeptide position corresponding to position 40 of SEQ ID NO:2. Thus, if the reference polypeptide of SEQ ID NO:2 has a histidine at position 40, then a "residue difference at position X40 as compared to SEQ ID NO:2" refers to an amino acid substitution of any residue other than histidine at the position of the polypeptide corresponding to position 40 of SEQ ID NO:2. 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, the present disclosure 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 disclosure 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 "f" (e.g., X192A/G). The present disclosure includes engineered polypeptide sequences comprising one or more amino acid differences that include either/or both conservative and non-conservative amino acid substitutions. The amino acid sequences of the specific recombinant carbonic anhydrase polypeptides included in the Sequence Listing of the present disclosure include an initiating methionine (M) residue (i.e., M represents residue position 1). The skilled artisan, however, understands that this initiating methionine residue can be removed by biological processing machinery, such as in a host cell or in vitro translation system, to generate a mature protein lacking the initiating methionine residue, but otherwise retaining the enzyme's properties. Consequently, the term "amino acid residue difference relative to SEQ ID NO:2 at position Xn" as used herein may refer to position "Xn" or to the corresponding position (e.g., position (X−1)n) in a reference sequence that has been processed so as to lack the starting methionine.

[0080] As used herein, the phrase "conservative amino acid substitutions" refers to the interchangeability of residues having similar side chains, 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 a hydroxyl side chain is substituted with another amino acid with a 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. Exemplary conservative substitutions are provided in Table 1.

TABLE 1

Exemplary Conservative Amino Acid Substitutions

Residue	Potential Conservative Substitutions
A, L, V, I	Other aliphatic (A, L, V, I) Other non-polar (A, L, V, I, G, M)
G, M	Other non-polar (A, L, V, I, G, M)
D, E	Other acidic (D, E)
K, R	Other basic (K, R)
N, Q, S, T	Other polar
H, Y, W, F	Other aromatic (H, Y, W, F)
C, P	Non-polar

[0081] As used herein, the phrase "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.

[0082] As used herein, "deletion" refers to modification of 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 polypeptide while retaining enzymatic activity and/or retaining the improved properties of an engineered 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.

[0083] As used herein, "insertion" refers to modification of the polypeptide by addition of one or more amino acids to the reference polypeptide. In some embodiments, the improved engineered PGA enzymes comprise insertions of one or more amino acids to the naturally occurring PGA polypeptide as well as insertions of one or more amino acids to engineered PGA polypeptides. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

[0084] 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 PGAs listed in the Tables provided in the Examples.

[0085] As used herein, "fragment" refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can typically have about 80%, about 90%, about 95%, about 98%, or about 99% of the full-length PGA polypeptide, for example the polypeptide of SEQ ID NO:2. In some embodiments, the fragment is "biologically active" (i.e., it exhibits the same enzymatic activity as the full-length sequence).

[0086] As used herein, "isolated polypeptide" refers to a polypeptide that is substantially separated from other contaminants that naturally accompany it (e.g., proteins, lipids, and polynucleotides). The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (e.g., host cell or in vitro synthesis). The improved PGA enzymes may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the engineered PGA polypeptides of the present disclosure can be an isolated polypeptide.

[0087] As used herein, "substantially pure polypeptide" refers to a composition in which the polypeptide species is the predominant species present (i.e., on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure engineered PGA polypeptide composition comprises about 60% or more, about 70% or more, about 80% or more, about 90% or more, about 91% or more, about 92% or more, about 93% or more, about 94% or more, about 95% or more, about 96% or more, about 97% or more, about 98% or more, or about 99% of all macromolecular species by mole or % weight present in the composition. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated improved PGA polypeptide is a substantially pure polypeptide composition.

[0088] As used herein, when used in reference to a nucleic acid or polypeptide, the term "heterologous" refers to a sequence that is not normally expressed and secreted by an organism (e.g., a wild-type organism). In some embodiments, the term encompasses a sequence that comprises two or more subsequences which are not found in the same relationship to each other as normally found in nature, or is recombinantly engineered so that its level of expression, or physical relationship to other nucleic acids or other molecules in a cell, or structure, is not normally found in nature. For instance, a heterologous nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged in a manner not found in nature (e.g., a nucleic acid open reading frame (ORF) of the invention operatively linked to a promoter sequence inserted into an expression cassette, such as a vector). In some embodiments, "heterologous polynucleotide" refers to any polynucleotide that is introduced into a host cell by laboratory techniques, and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

[0089] As used herein, "suitable reaction conditions" refer to those conditions in the biocatalytic reaction solution (e.g., ranges of enzyme loading, substrate loading, cofactor loading, temperature, pH, buffers, co-solvents, etc.) under which a PGA polypeptide of the present disclosure is capable of releasing free insulin by removing tri-phenyl acetate protecting groups. Exemplary "suitable reaction conditions" are provided in the present disclosure and illustrated by the Examples.

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

[0091] As used herein, "substrate" in the context of a biocatalyst mediated process refers to the compound or molecule acted on by the biocatalyst.

[0092] As used herein "product" in the context of a biocatalyst mediated process refers to the compound or molecule resulting from the action of the biocatalyst.

[0093] 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.

[0094] As used herein "acylase" and "acyltransferases" are used interchangeably to refer to enzymes that are capable of transferring an acyl group from a donor to an acceptor to form esters or amides. The acylase mediated reverse reaction results in hydrolysis of the ester or amide.

[0095] As used herein, "penicillin G" and "benzylpenicillin" refer to the antibiotic also known as (2S,5R,6R)-3,3-dimethyl-7-oxo-6-(2-phenylacetamido)-4-thia-1-azabicyclo [3.2.0]heptane-2-carboxylic acid ($C_{16}H_{18}N_2O_4S$). It is primarily effective against Gram-positive organisms, although some Gram-negative organisms are also susceptible to it.

[0096] As used herein, "penicillin G acylase" and "PGA" are used interchangeably to refer to an enzyme having the capability of mediating cleavage of penicillin G (benzylpenicillin) to phenylacetic acid (PHA) and 6-aminopenicillanic acid (6-APA). In some embodiments, PGA activity can be based on cleavage of model substrates, for instance the cleavage of 6-nitro-3-(phenylacetamide)benzoic acid to phenylacetic acid and 5-amino-2-nitro-benzoic acid. PGAs are also capable of carrying out the reverse reaction of transferring an acyl group of an acyl donor to an acyl acceptor. PGAs as used herein include naturally occurring (wild type) PGAs as well as non-naturally occurring PGA enzymes comprising one or more engineered polypeptides generated by human manipulation. The wild-type PGA gene is a heterodimer consisting of alpha subunit (23.8 KDa) and beta subunit (62.2KDa) linked by a spacer region of 54 amino acids. Due to the presence of the spacer region, an auto-processing step is required to form the active protein.

[0097] As used herein, "acyl donor" refers to that portion of the acylase substrate which donates the acyl group to an acyl acceptor to form esters or amides.

[0098] As used herein, "acyl acceptor" refers to that portion of the acylase substrate which accepts the acyl group of the acyl donor to form esters or amides.

[0099] As used herein, "α-chain sequence" refers to an amino acid sequence that corresponds to (e.g., has at least

85% identity to) the residues at positions 27 to 235 of SEQ ID NO: 2. As used herein, a single chain polypeptide can comprise an "a-chain sequence" and additional sequence(s).

[0100] As used herein, "β-chain sequence" refers to an amino acid sequence that corresponds to (e.g., has at least 85% identity to) residues at positions 290 to 846 of SEQ ID NO:2. As used herein, a single chain polypeptide can comprise an "β-chain sequence" and additional sequence(s).

[0101] As used herein, "derived from" when used in the context of engineered PGA enzymes, identifies the originating PGA enzyme, and/or the gene encoding such PGA enzyme, upon which the engineering was based. For example, the engineered PGA enzyme of SEQ ID NO: 88 was obtained by artificially evolving, over multiple generations the gene encoding the *K. citrophila* PGA alpha-chain and beta-chain sequences of SEQ ID NO:6. Thus, this engineered PGA enzyme is "derived from" the naturally occurring or wild-type PGA of SEQ ID NO: 6.

[0102] As used herein, "insulin" refers to the polypeptide hormone produced by the beta-cells of the pancreas in normal individuals. Insulin is necessary for regulating carbohydrate metabolism, by reducing blood glucose levels. Systematic deficiency of insulin results in diabetes. Insulin is comprised of 51 amino acids and has a molecular weight of approximately 5800 daltons. Insulin is comprised of two peptide chains (designated "A" and "B"), containing one intrasubunit and two intersubunit disulfide bonds. The A chain is composed of 21 amino acids and the B chain is composed of 30 amino acids. The two chains form a highly ordered structure, with several alpha-helical regions in both the A and B chains. Isolated chains are inactive. In solution, insulin is either a monomer, dimer, or hexamer. It is hexameric in the highly concentrated preparations used for subcutaneous injection, but becomes monomeric as it is diluted in body fluids. The definition is intended to encompass proinsulin and any purified isolated polypeptide having part or all of the structural conformation and at least one of the biological properties of naturally-occurring insulin. It is further intended to encompass natural and synthetically-derived insulin, including glycoforms, as well as analogs (e.g., polypeptides having deletions, insertions, and/or substitutions).

[0103] Insulin contains three nucleophilic amines that can potentially react with a phenylacetate-donor and be deprotected by PGA. These residues include a Lys on the B-chain at position 29 (B29) and two N-terminal free amines, Gly on the A-chain at position 1 (Al) and Phe on the B-chain at position 1 (B1). An insulin dimer is composed of two insulin molecules are linked through a linker at the Lys on the B-chain at position 29 (B29) of insulin. In some embodiments, the insulin dimers comprise two molecules of insulin linked by covalent bonds, hydrogen bonds, electrostatic interactions, and/or hydrophobic interactions. It is not intended that the present invention be limited to any specific means of dimerization nor any specific positions of dimerization. PGA has previously been reported to catalyze hydrolysis of N-phenylacetate-protected peptides and insulin with exclusive selectivity for the phenylacetate amide bond, leaving the rest of the peptide bonds of the protein intact (Brtnik et al., Coll. Czech. Chem. Commun., 46 (8), 1983-1989 [1981]; and Wang et al. Biopolym. 25 (Suppl.), S109-S114 [1986]).

[0104] As used herein, "tetra-phenyl acetate protecting group," refers to an insulin-dimer molecule in which the four primary amines at the A1/A1' and B1/B1' positions that are protected with a phenyl acyl group.

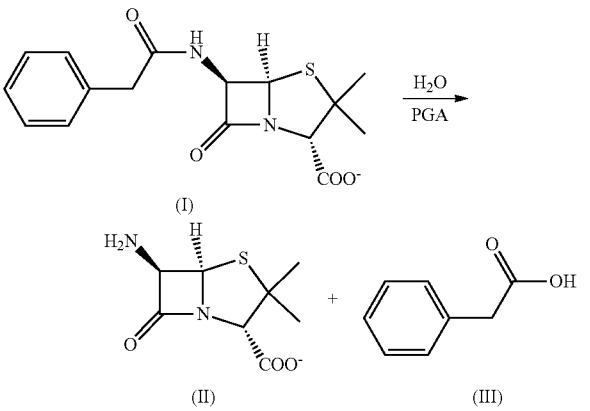
[0105] As used herein, "di-phenyl acetate protecting group" refers to an insulin-dimer molecule in which the two primary amines at the A1, and A1' positions that are protected with a phenyl acyl group.

Penicillin G Acylases

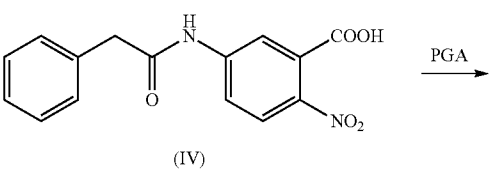
[0106] Penicillin acylase was first described from *Penicillium chrysogenum* Wisc. Q176 by Sakaguchi and Murao (Sakaguchi and Murao, J. Agr.Chem. Soc. Jpn., 23:411 [1950]). Penicillin G acylase is a hydrolytic enzyme that acts on the side chains of penicillin G, cephalosporin G, and related antibiotics to produce the β-lactam antibiotic intermediates 6-amino penicillanic acid and 7-amino des-acetoxy cephalosporanic acid, with phenyl acetic acid as a common by-product. These antibiotic intermediates are among the potential building blocks of semi-synthetic antibiotics, such as ampicillin, amoxicillin, cloxacillin, cephalexin, and cefatoxime.

[0107] As indicated above, penicillin G acylases (PGA) are characterized by the ability to catalyze the hydrolytic cleavage of penicillin G, with a conjugate base of structural formula (I), to 6-amino penicillanic acid, with a conjugate base of structural formula (II), and phenylacetic acid of structural formula (III), as shown in Scheme 1:

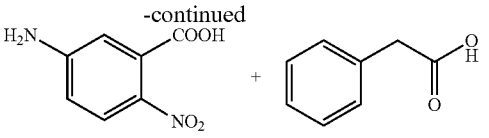
Scheme 1



[0108] While not being bound by theory, substrate specificity appears associated with recognition of the hydrophobic phenyl group while a nucleophile, which in some PGAs is a serine residue at the N-terminus of the beta-chain acts as the acceptor of beta-lactam and a variety of other groups, such as beta-amino acids. PGAs can also be characterized by the ability to cleave a model substrates analogous to penicillin G, for instance cleavage of 6-nitro-3-(phenylacetamido)benzoic acid (NIPAB) of structural formula (IV), as shown in Scheme 2:



-continued



(V) (III)

to phenylacetic acid of structural formula (III) and 5-amino-2-nitro-benzoic acid of structural formula (V) (See e.g., Alkema et al., Anal. Biochem., 275:47-53 [1999]). Because the 5-amino-2-nitro-benzoic acid is chromogenic, the substrate of formula (IV) provides a convenient way of measuring PGA activity. In addition to the foregoing reactions, PGAs can also be used in the kinetic resolution of DL-tert leucine for the preparation of optically pure tert leucine (See e.g., Liu et al., Prep. Biochem. Biotechnol., 36:235-41 [2006]).

[0109] The variant PGAs of the present invention are derived from based on the enzyme obtained from the organism *Kluyvera citrophila* (*K citrophila*). As with PGAs from other organisms, the PGA of *K. citrophila* is a heterodimeric enzyme comprised of an alpha-subunit and a beta-subunit that is generated by proteolytic processing of a pre-pro-PGA polypeptide. Removal of a signal peptide and a spacer peptide produces the mature heterodimer (See e.g., Barbero et al., Gene 49:69-80 [1986]). The amino acid sequence of the naturally occurring pre-pro-PGA polypeptide of *K. citrophila* is publicly available (See e.g., Genbank accession No. P07941, [gi:129551]) and is provided herein as SEQ ID NO:2. The alpha-chain sequence of the naturally occurring *K. citrophila* PGA corresponds to residues 27 to 235 of SEQ ID NO:2. The beta-chain sequence of the naturally occurring K citrophila PGA corresponds to residues 290 to 846 of SEQ ID NO:2. Residues 1 to 26 of SEQ ID NO:2 correspond to the signal peptide and residues 236-289 of SEQ ID NO:2 correspond to the linking propeptide, both of which are removed to generate the naturally occurring mature PGA enzyme which is a heterodimer comprising an a-chain subunit and a β-chain subunit.

[0110] In some embodiments, the present invention provides engineered PGA polypeptides with amino acid sequences that have at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to SEQ ID NOS:4, 6, 8, 34, 46, 54, 74, and/or 88.

[0111] In some embodiments, the present invention provides insulin-specific deacylation biocatalysts suitable for commercial use. Directed evolution was used to develop efficient acylase variants capable of deprotecting the A1/A1'/B1/B1'-tetra-phenyl acetate protecting groups and A1/A1'-di-phenyl acetate protecting groups on the insulin-dimer and generate greater than 99% of the free insulin-dimer. As indicated in the Examples, after six rounds of evolution variants were produced that generate more than 99% free insulin in less than 5 hrs. at 0.15g/L enzyme and 30 g/L A1/A1'di-phenyl acetate protected dimer loading. The activity of one variant (SEQ ID NO: 88) was shown to be ~300-fold improved over the initial backbone tested. Indeed, the PGA variants provided herein are capable of accepting a wide range of acyl groups, exhibit increased solvent stability, and improved thermostability, as compared to the wild-type PGA. In some embodiments, the variant PGAs provided herein lack the spacer region. Thus, the auto-processing step is not required in order to produce active enzymes of interest.

[0112] The present invention also provides polynucleotides encoding the engineered PGA polypeptides. 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. Expression constructs containing a heterologous polynucleotide encoding the engineered PGA polypeptides can be introduced into appropriate host cells to express the corresponding PGA polypeptide.

[0113] Because of the knowledge of the codons corresponding to the various amino acids, availability of a protein sequence provides a description of all the polynucleotides capable of encoding the subject. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons allows an extremely large number of nucleic acids to be made, all of which encode the improved PGA enzymes disclosed herein. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the protein. In this regard, the present disclosure specifically contemplates each and every possible variation of polynucleotides that could be made by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide disclosed herein, including the amino acid sequences presented in the Tables in Examples 6-11.

[0114] In various embodiments, the codons are preferably selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria are used to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells.

[0115] In some embodiments, all codons need not be replaced to optimize the codon usage of the PGA polypeptides since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon optimized polynucleotides encoding the PGA enzymes may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full length coding region.

[0116] In some embodiments, the polynucleotide comprises a nucleotide sequence encoding a PGA polypeptide with an amino acid sequence that has at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to the alpha-chain and/or beta-chain any of the reference engineered PGA polypeptides described herein. Accordingly, in some embodiments, the polynucleotide encodes an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to at least one reference alpha- and beta-chain sequence (e.g., SEQ ID NOS: 4, 6, 8, 34, 46, 54, 74, and/or 88). In some embodiments, the polynucleotide encodes an alpha- and/or beta-chain amino acid sequence of SEQ ID NO: 4, 6, 8, 34, 46, 54, 74, and/or 88.

[0117] In some embodiments, the polynucleotide comprises a nucleotide sequence encoding a PGA polypeptide with an amino acid sequence that has at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to SEQ ID NO: 4, 6, 8, 34, 46, 54, 74, and/or 88. Accordingly, in some embodiments, the polynucleotide encodes an amino acid sequence that is at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 4, 6, 8, 34, 46, 54, 74, and/or 88.

[0118] In some embodiments, the isolated polynucleotide encoding an improved PGA polypeptide is manipulated in a variety of ways to provide for improved activity and/or expression of the polypeptide. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

[0119] For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed. Mutagenesis and directed evolution methods are well known in the art (See e.g., U.S. Pat. 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 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

[0120] ; 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).

[0121] In some embodiments, the variant PGA acylases of the present invention further comprise additional sequences that do not alter the encoded activity of the enzyme. For example, in some embodiments, the variant PGA acylases are linked to an epitope tag or to another sequence useful in purification.

[0122] In some embodiments, the variant PGA acylase polypeptides of the present invention are secreted from the host cell in which they are expressed (e.g., a yeast or filamentous fungal host cell) and are expressed as a preprotein including a signal peptide (i.e., an amino acid sequence linked to the amino terminus of a polypeptide and which directs the encoded polypeptide into the cell secretory pathway).

[0123] In some embodiments, the signal peptide is an endogenous *K. citrophila* PGA acylase signal peptide. In some additional embodiments, signal peptides from other *K. citrophila* secreted proteins are used. In some embodiments, other signal peptides find use, depending on the host cell and other factors. Effective signal peptide coding regions for filamentous fungal host cells include, but are not limited to, the signal peptide coding regions obtained from *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, *Humicola lanuginosa* lipase, and *T. reesei* cellobiohydrolase II. Signal peptide coding regions for bacterial host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NC1B 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* β-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. In some additional embodiments, other signal peptides find use in the present invention (See e.g., Simonen and Palva, Microbiol. Rev., 57: 109-137 [1993], incorporated herein by reference). Additional useful signal peptides for yeast host cells include those from the genes for *Saccharomyces cerevisiae* alpha-factor, Saccharomyces cerevisiae SUC2 invertase (See e.g., Taussig and Carlson, Nucl. Acids Res., 11:1943-54 [1983]; SwissProt Accession No. P00724; and Romanos et al., Yeast 8:423-488 [1992]). In some embodiments, variants of these signal peptides and other signal peptides find use. Indeed, it is not intended that the present invention be limited to any specific signal peptide, as any suitable signal peptide known in the art finds use in the present invention.

[0124] In some embodiments, the present invention provides polynucleotides encoding variant PGA acylase polypeptides, and/or biologically active fragments thereof, as described herein. In some embodiments, the polynucleotide is operably linked to one or more heterologous regulatory or control sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. In some embodiments, expression constructs containing a heterologous polynucleotide encoding a variant PGA acylase is introduced into appropriate host cells to express the variant PGA acylase.

[0125] Those of ordinary skill in the art understand that due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding variant PGA acylase poly-

peptides of the present invention exist. For example, the codons AGA, AGG, CGA, CGC, CGG, and CGU all encode the amino acid arginine. Thus, at every position in the nucleic acids of the invention where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described above without altering the encoded polypeptide. It is understood that "U" in an RNA sequence corresponds to "T" in a DNA sequence. The invention contemplates and provides each and every possible variation of nucleic acid sequence encoding a polypeptide of the invention that could be made by selecting combinations based on possible codon choices.

[0126] As indicated above, DNA sequence encoding a PGA may also be designed for high codon usage bias codons (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. A codon whose frequency increases with the level of gene expression is typically an optimal codon for expression. In particular, a DNA sequence can be optimized for expression in a particular host organism. A variety of methods are well-known in the art for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis (e.g., using cluster analysis or correspondence analysis,) and the effective number of codons used in a gene. 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, as is well-known in the art. Polynucleotides encoding variant PGAs can be prepared using any suitable methods known in the art. Typically, oligonucleotides are individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase-mediated methods) to form essentially any desired continuous sequence. In some embodiments, polynucleotides of the present invention are prepared by chemical synthesis using, any suitable methods known in the art, including but not limited to automated synthetic methods. For example, in the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In some embodiments, double stranded DNA fragments are then obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence. There are numerous general and standard texts that provide methods useful in the present invention are well known to those skilled in the art.

[0127] The engineered PGAs can be obtained by subjecting the polynucleotide encoding the naturally occurring PGA to mutagenesis and/or directed evolution methods, as discussed above. Mutagenesis may be performed in accordance with any of the techniques known in the art, including random and site-specific mutagenesis. Directed evolution can be performed with any of the techniques known in the art to screen for improved variants including shuffling. Other directed evolution procedures that find use include, but are not limited to staggered extension process (StEP), in vitro recombination, mutagenic PCR, cassette mutagenesis, splicing by overlap extension (SOEing), ProSARTM directed evolution methods, etc., as well as any other suitable methods. In some embodiments, the variant PGA polypeptide or polynucleotide is a circularly permuted sequence that may or may not comprise a linker sequence. Such circularly permuted variants are produced using any suitable method known in the art (See e.g., U.S. Pat. Nos. 8,034,579, 8,338,138, and 9,428,563, incorporated herein by reference).

[0128] The clones obtained following mutagenesis treatment are screened for engineered PGAs having a desired improved enzyme property. Measuring enzyme activity from the expression libraries can be performed using the standard biochemistry technique of monitoring the rate of product formation. Where an improved enzyme property desired is thermal stability, enzyme activity may be measured after subjecting the enzyme preparations to a defined temperature and measuring the amount of enzyme activity remaining after heat treatments. Clones containing a polynucleotide encoding a PGA are then isolated, sequenced to identify the nucleotide sequence changes (if any), and used to express the enzyme in a host cell.

[0129] When the sequence of the engineered polypeptide is known, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides of the invention can be prepared by chemical synthesis (e.g., using the classical phosphoramidite method described by Beaucage et al., Tet. Lett., 22:1859-69 [1981], or the method described by Matthes et al., EMBO J., 3:801-05 [1984], as it is typically practiced in automated synthetic methods). According to the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In addition, essentially any nucleic acid can be obtained from any of a variety of commercial sources (e.g., The Midland Certified Reagent Company, Midland, Tex., The Great American Gene Company, Ramona, Calif., ExpressGen Inc. Chicago, Ill., Operon Technologies Inc., Alameda, Calif., and many others).

[0130] The present invention also provides recombinant constructs comprising a sequence encoding at least one variant PGA, as provided herein. In some embodiments, the present invention provides an expression vector comprising a variant PGA polynucleotide operably linked to a heterologous promoter. In some embodiments, expression vectors of the present invention are used to transform appropriate host cells to permit the host cells to express the variant PGA protein. Methods for recombinant expression of proteins in fungi and other organisms are well known in the art, and a number of expression vectors are available or can be constructed using routine methods. In some embodiments, nucleic acid constructs of the present invention comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), and the like, into which a nucleic acid sequence of the invention has been inserted. In some

embodiments, polynucleotides of the present invention are incorporated into any one of a variety of expression vectors suitable for expressing variant PGA polypeptide(s). Suitable vectors include, but are not limited to chromosomal, non-chromosomal and synthetic DNA sequences (e.g., derivatives of SV40), as well as bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated virus, retroviruses, and many others. Any suitable vector that transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host finds use in the present invention.

[0131] In some embodiments, the construct further comprises regulatory sequences, including but not limited to a promoter, operably linked to the protein encoding sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art. Indeed, in some embodiments, in order to obtain high levels of expression in a particular host it is often useful to express the variant PGAs of the present invention under the control of a heterologous promoter. In some embodiments, a promoter sequence is operably linked to the 5' region of the variant PGA coding sequence using any suitable method known in the art. Examples of useful promoters for expression of variant PGAs include, but are not limited to promoters from fungi. In some embodiments, a promoter sequence that drives expression of a gene other than a PGA gene in a fungal strain finds use. As a non-limiting example, a fungal promoter from a gene encoding an endoglucanase may be used. In some embodiments, a promoter sequence that drives the expression of a PGA gene in a fungal strain other than the fungal strain from which the PGAs were derived finds use. Examples of other suitable promoters useful for directing the transcription of the nucleotide constructs of the present invention in a filamentous fungal host cell 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, incorporated herein by reference), 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), promoters such as cbh1, cbh2, egl1, egl2, pepA, hfb1, hfb2, xyn1, amy, and glaA (See e.g., Nunberg et al., Mol. Cell Biol., 4:2306 -2315 [1984]; Boel et al., EMBO J., 3:1581-85 [1984]; and European Patent Appin. 137280, all of which are incorporated herein by reference), and mutant, truncated, and hybrid promoters thereof.

[0132] In yeast host cells, useful promoters include, but are not limited to those from the genes for *Saccharomyces cerevisiae* enolase (eno-1), *Saccharomyces cerevisiae* galactokinase (gall), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *S. cerevisiae* 3-phosphoglycerate kinase. Additional useful promoters useful for yeast host cells are known in the art (See e.g., Romanos et al., Yeast 8:423-488 [1992], incorporated herein by reference). In addition, promoters associated with chitinase production in fungi find use in the present invention (See e.g., Blaiseau and Lafay, Gene 120243-248 [1992]; and Limon et al., Curr. Genet., 28:478-83 [1995], both of which are incorporated herein by reference).

[0133] For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present disclosure, include but are not limited to the promoters obtained from the *E. coli* lac operon, *E. coli* trp operon, bacteriophage lambda, *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]).

[0134] In some embodiments, cloned variant PGAs of the present invention also have a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator that is functional in the host cell of choice finds use in the present invention. Exemplary transcription terminators for filamentous fungal host cells include, but are not limited to those 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 (See e.g., U.S Pat. No. 7,399,627, incorporated herein by reference). In some embodiments, exemplary terminators for yeast host cells include those 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 well-known to those skilled in the art (See e.g., Romanos et al., Yeast 8:423-88 [1992]).

[0135] In some embodiments, a suitable leader sequence is part of a cloned variant PGA sequence, which is a nontranslated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice finds use in the present invention. Exemplary leaders for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells include, but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[0136] In some embodiments, the sequences of the present invention also comprise a polyadenylation sequence, which is a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice finds

use in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells include, but are not limited to those obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are known in the art (See e.g., Guo and Sherman, Mol. Cell. Biol., 15:5983-5990 [1995].

[0137] In some embodiments, the control sequence comprises a signal peptide coding region encoding an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not naturally contain a signal peptide coding region.

[0138] Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

[0139] Effective signal peptide coding regions for bacterial host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NCIB 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]).

[0140] 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.

[0141] Useful signal peptides for yeast host cells include, but are not limited to genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are known in the art (See e.g., Romanos et al., [1992], supra).

[0142] In some embodiments, the control sequence comprises a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active PGA polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from 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).

[0143] 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.

[0144] In some embodiments, regulatory sequences are also used to allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which 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, as examples, the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[0145] Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene, which is amplified in the presence of methotrexate, and the metallothionein genes, which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the PGA polypeptide of the present invention would be operably linked with the regulatory sequence.

[0146] Thus, in additional embodiments, the present invention provides recombinant expression vectors comprising a polynucleotide encoding an engineered PGA polypeptide or a variant thereof, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. In some embodiments, the various nucleic acid and control sequences described above are joined together to produce a recombinant expression vector that may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, in some embodiments, the nucleic acid sequences are expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[0147] The recombinant expression vector comprises any suitable vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the 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. In some embodiments, the vectors are linear or closed circular plasmids.

[0148] In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome). In some embodiments, the vector contains any means for assuring self-replication. Alternatively, in some other embodiments, upon being introduced into the host cell, the vector is integrated into the genome and replicated together with the chromosome(s) into which

it has been integrated. Furthermore, in additional 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, or a transposon find use.

[0149] In some embodiments, the expression vector of the present invention 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 antimicrobials or heavy metals, prototrophy to auxotrophs, and the like. Any suitable selectable markers for use in a filamentous fungal host cell find use in the present invention, including, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferase), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof. Additional markers useful in host cells such as *Aspergillus*, include but are not limited to the amdS and pyrG genes of *Aspergillus nidulans* or *Aspergillus oryzae*, and the bar gene of *Streptomyces hygroscopicus*. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. 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.

[0150] 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 non-homologous recombination.

[0151] 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.

[0152] 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, pACYC177 (which plasmid has the P15A ori), or pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, or pAM□1 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]).

[0153] 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.

[0154] 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 p3xFLAG™ 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].

[0155] Thus, in some embodiments, a vector comprising a sequence encoding at least one variant PGA is transformed into a host cell in order to allow propagation of the vector and expression of the variant PGA(s). In some embodiments, the variant PGAs 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 PGA(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).

[0156] In another aspect, the present invention provides host cells comprising a polynucleotide encoding an improved PGA polypeptide provided herein, the polynucleotide being operatively linked to one or more control sequences for expression of the PGA enzyme in the host cell. Host cells for use in expressing the PGA 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.

[0157] Polynucleotides for expression of the PGA 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.

[0158] 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.

[0159] 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.

[0160] 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.*

[0161] 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).

[0162] 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-01 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*).

[0163] 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 and Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

[0164] 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 PGA 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. Micriobiol.,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).

[0165] 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 U.S. Pat. Appln. Publn. 2006/0195947, which is hereby incorporated by reference herein) find use.

[0166] 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 PGA 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.

[0167] In some embodiments, cells expressing the variant PGA 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 is 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.

[0168] In some embodiments of the present invention, cell-free transcription/translation systems find use in producing variant PGA(s). Several systems are commercially available and the methods are well-known to those skilled in the art.

[0169] The present invention provides methods of making variant PGA 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, 6, 8, 34, 46, 54, 74, and/or 88, 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 PGA polypeptide; and optionally recovering or isolating the expressed variant PGA polypeptide, and/or recovering or isolating the culture medium containing the expressed variant PGA polypeptide. In some embodiments, the methods further provide optionally lysing the transformed host cells after expressing the encoded PGA polypeptide and optionally recovering and/or isolating the expressed variant PGA polypeptide from the cell lysate. The present invention further provides methods of making a variant PGA polypeptide comprising cultivating a host cell transformed with a variant PGA polypeptide under conditions suitable for the production of the variant PGA polypeptide and recovering the variant PGA polypeptide. Typically, recovery or isolation of the PGA 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.

[0170] Engineered PGA 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.

[0171] Chromatographic techniques for isolation of the PGA 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.

[0172] In some embodiments, affinity techniques find use in isolating the improved PGA enzymes. For affinity chromatography purification, any antibody which specifically binds the PGA 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 PGA. The PGA 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.*

[0173] In some embodiments, the PGA 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 PGA variants are prepared as lyophilisates, in powder form (e.g., acetone powders), or prepared as enzyme solutions. In some embodiments, the PGA variants are in the form of substantially pure preparations.

[0174] In some embodiments, the PGA 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.

[0175] In some embodiments, immunological methods are used to purify PGA variants. In one approach, antibody raised against a variant PGA polypeptide (e.g., against a polypeptide comprising any of SEQ ID NOS: 4, 6, 8, 34, 46, 54, 74, and/or 88, and/or an immunogenic fragment thereof) using conventional methods is immobilized on beads, mixed with cell culture media under conditions in which the variant PGA is bound, and precipitated. In a related approach, immunochromatography finds use.

[0176] In some embodiments, the variant PGAs are expressed as a fusion protein including a non-enzyme portion. In some embodiments, the variant PGA 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 PGA 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.

EXPERIMENTAL

[0177] Various features and embodiments of the disclosure are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

[0178] In the experimental disclosure below, the following abbreviations apply: ppm (parts per million); M (molar); mM (millimolar), uM and μM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg (micrograms); L and 1 (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); um and μm (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); ° C. (degrees Centigrade); RT (room temperature); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); aa (amino acid); TB (Terrific Broth; 12 g/L bacto-tryptone, 24 g/L yeast extract, 4 mL/L glycerol, 65 mM potassium phosphate, pH 7.0, 1 mM $MgSO_4$); LB (Luria broth); CAM (chloramphenicol); PMBS (polymyxin B sulfate); IPTG (isopropyl thiogalactoside); TFA (trifluoroacetic acid); CHES (2-cyclohexylamino)ethanesulfonic acid; acetonitrile (MeCN); dimethylsulfoxide (DMSO); dimethylacetamide (DMAc); HPLC (high performance liquid chromatography); UPLC (ultra performance liquid chromatography); FIOPC (fold improvement over positive control); HTP (high throughput); MWD (multiple wavelength detector); UV (ultraviolet); Codexis (Codexis, Inc., Redwood City, Calif.); Sigma-Aldrich (Sigma-Aldrich, St. Louis, Mo.); Millipore (Millipore, Corp., Billerica Mass.); Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, Mich.); Daicel (Daicel, West Chester, Pa.); Genetix (Genetix USA, Inc., Beaverton, Oreg.); Molecular Devices (Molecular Devices, LLC, Sunnyvale, Calif.); Applied Biosystems (Applied Biosystems, part of Life Technologies, Corp., Grand Island, N.Y.), Agilent (Agilent Technologies, Inc., Santa Clara, Calif.); Thermo Scientific (part of Thermo Fisher Scientific, Waltham, Mass.); (Infors; Infors-HT, Bottmingen/Basel, Switzerland); Corning (Corning, Inc., Palo Alto, Calif.); and Bio-Rad (Bio-Rad Laboratories, Hercules, Calif.); Microfluidics (Microfluidics Corp., Newton, Mass.); Waters (Waters Corp., Milford, Mass.).

EXAMPLE 1

E. coli Expression Hosts Containing Recombinant PGA Genes

[0179] The initial PGA enzymes used to produce the variant enzymes of the present invention were obtained from variants disclosed in co-owned US Pat. Appin. Publ. No. 2016/0326508, incorporated herein by reference in its entirety and for all purposes. The PGA panel plate comprises a collection of engineered PGA polypeptides that have improved properties, as compared to the wild-type *Kluyvera citrophila* PGA. The wild type PGA gene is a heterodimer consisting of an alpha subunit (23.8 KDa) and a beta subunit (62.2KDa) that are linked by 54aa spacer region. Due to the presence of the spacer region, an autoprocessing step is required to form the active protein. During the development of the present invention, the wild-type gene was modified to eliminate the spacer region, thus eliminating the auto processing step. The PGA panel plate (Codexis) contains PGA variants that lack the spacer region (See e.g., US Pat. Appin. Publn. 2010/0143968, which is incorporated herein by reference in its entirety and for all purposes). A C-terminal histidine was added to the genes and the PGA-encoding genes were cloned into the expression vector pCK110900 (See US Pat. Appin. Publn. No. 2006/0195947 and 2016/0244787, both of which are incorporated herein by reference in their entireties and for all purposes), operatively linked to the lac promoter under control of the lacI repressor. The expression vector also contains the P15a origin of replication and a 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., U.S. Pat. No. 8,383,346 and WO2010/144103, each of which is incorporated herein by reference in its entirety and for all purposes).

EXAMPLE 2

Preparation of HTP PGA-Containing Wet Cell Pellets

[0180] *E. coli* cells containing recombinant PGA-encoding genes from monoclonal colonies were inoculated into 180 μl LB containing 1% glucose and 30 μg/mL chloramphenicol into 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, 10 μl 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 4000 rpm for 10 min. The supernatants were discarded and the pellets frozen at −80° C. prior to lysis.

EXAMPLE 3

Preparation of HTP PGA-Containing Cell Lysates

[0181] First, 2000 lysis buffer containing 50 mM Tris-HCl buffer, pH 7.5, 1 mg/mL lysozyme, and 0.5 mg/mL PMBS was 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 4000 rpm and 4° C. The clear supernatants were then used in biocatalytic reactions to determine their activity levels.

EXAMPLE 4

Preparation of Lyophilized Lysates from Shake Flask (SF) Cultures

[0182] Selected HTP cultures grown as described above were plated onto LB agar plates with 1% glucose and 30

kg/ml CAM, and grown overnight at 37° C. A single colony from each culture was transferred to 6 ml of LB with 1% glucose and 30 μg/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., 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., 250 rpm. The cultures were centrifuged 4000 rpm×20 min. The supernatant was discarded, and the pellets were resuspended in 30 ml of 20 mM TRIS-HCl, pH 7.5. The cells were pelleted (4000 rpm x 20 min) and frozen at −80° C. for 120 minutes. Frozen pellets were resuspended in 30 ml of 20 mM TRIS-HCl pH 7.5, and lysed using a Microfluidizer® processor system (Microfluidics) at 18,000 psi. The lysates were pelleted (10,000 rpm×60 min) and the supernatants were frozen and lyophilized to generate shake flake (SF) enzymes.

[0183] The activity of selected shake flask PGA variants was evaluated based on the efficiency of the variants in removing the four/two phenyl acetate groups chemically attached to the A1/A1' (glycine), and B1/B1' (phenylalanine), residues of an insulin-dimer. Reactions using shake flask powders were carried out in 2 mL 96-well plates. In this assay, 200 μL reactions solutions consisting of 10-30 g/L tetra-protected insulin dimer (A1,A1',B1,B1'-tetraphenylacetimido-insulin-dimer) or di-protected insulin dimer (A1, A1'-diphenylacetimido- insulin dimer), 0.15-3 g/L shake flask powder, 0.2 M triethanolamine (TEoA) buffer, pH 8.5, and 20-30% (v/v) DMSO were prepared. The reaction plates were sealed with a heat seal and incubated at 30° C. and agitated at 300 RPM in a Thermotron® shaker (2 mm throw, model # AJ185, Infors) for 25 h. Three 20 μL aliquots of each reaction were taken at 45 min, 4.75 and 25 h, and quenched 1:1 with MeCN, then diluted 1:5 with deionized water. The samples were analyzed by UPLC using parameters in Tables 12.1, 12.2 and 12.3.

EXAMPLE 5

Evaluation of Shake Flask Powders of SEQ ID NO: 4 and SEQ ID NO: 6 on Tetra-Protected Insulin Tethered-Dimer

[0184] In order to assess activity differences between SEQ ID NO: 4 and the engineered variant SEQ ID NO: 6, which contains a C-terminal polyhistidine affinity tag (His-tag, HT), reactions using shake flask powders (See, Example 4) were carried out in 96-well plates with 2 mL wells. In these assays, 200 μL reaction solutions consisting of 14.5 g/L tetra-protected insulin dimer substrate (A1, A1', B1,B1'-tetraphenylacetimido-insulin tethered-dimer), 2.5 g/L shake flask powder, 0.2 M triethanolamine (TEoA) buffer, pH 8.5, and 20% (v/v) DMSO were prepared. Reaction solution-containing plates were sealed with a heat seal and incubated at 30° C. and agitated at 300 RPM in a Thermotron® shaker (2 mm throw, model # AJ185, Infors) for 25 h. Three 20 μL aliquots of each reaction were taken at 45 min, 4.75 and 25 h, and quenched 1:1 with MeCN, then diluted 1:5 with deionized water. The samples were analyzed by UPLC using a Waters Cortecs® C18 column and the method described in Table 12.1. The activity was determined by comparing percent conversion (n=3) to the insulin dimer (product). The results are presented in Table 5.1.

TABLE 5.1

Activity Assessment of SEQ ID NO: 4 Compared to SEQ ID NO: 6

	% Conversion		
Enzyme	45 min	4.75 h	25 h
SEQ ID NO: 4	58%	71%	83%
SEQ ID NO: 6	3%	35%	78%

EXAMPLE 6

Improvements in the Deacylation of Insulin Compared to SEQ ID NO: 6 in High Throughput Screening

[0185] SEQ ID NO: 6 was selected as the next parent enzyme, based on the results described in Example 5. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and 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.

[0186] HTP reactions were carried out in 96 well deep-well plates containing 200 μL of 10 g/L tetraprotected insulin dimer substrate, 200 mM TEoA buffer, pH 8.5, 20% DMSO and 10 μL HTP lysate. The HTP plates were heat sealed and incubated in Thermotron® shakers at 30° C., 300 rpm, for 18 hours. The reactions were quenched with 200 μL MeCN and mixed for 5 minutes using a bench top shaker. The plates were then centrifuged at 4000 rpm for 5 minutes, diluted 24× into water, and injected onto an UPLC for analysis using the parameters in Table 12.1.

[0187] The percent conversion relative to SEQ ID NO:6 (Percent Conversion FIOP) was calculated as the percent conversion of the product formed by the variant over the percent conversion produced by SEQ ID NO: 6 The results are shown in Table 6.1. The percent conversion was quantified by dividing the area of the product peak by the sum of the areas of the substrate, and product as determined by UPLC analysis.

TABLE 6.1

Activity of Deacylating Variants Relative to SEQ ID NO: 6

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 6)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 6
4	13/14	A373M	++
5		K369W	++

TABLE 6.1-continued

Activity of Deacylating Variants Relative to SEQ ID NO: 6

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 6)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 6
6		L253V	++
7		K369V	++
8		L257V	++
9	7/8	F254W/A255G/W370I	++
10		T115A	++
11		K369L	++
12		Q626G	++
13		F254T	++
14		D623W	++
15		D268S	++
16		V391S	++
17		T560I	++
18		D623A	++
19		N348Q	++
20		N627G	++
21		Q554P	++
22		M600T/D623V	++
23		S706G	++
24		V391P	++
25	15/16	K369P	++
26		A255M	++
27	9/10	F254W/A255G	++
28		Q554V	++
29		S740L	++
30		N185V	++
31		S530C	++
32		Y752E	+
33		A255Y	+
34		T115P	+
35		N348C	+
36		G260P	+
37		W370S	+
38		L253K	+
39		Q556G	+
40	11/12	W370F	+
41		N388T	+
42		I624A	+
43		Q554A	+
44		T384L	+
45		I127S	+
46		Q559S	+
47		W370G	+
48		N125L	+
49		N125T	+
50		T705P	+
51		S372A	+
52		E377P	+
53		I389L	+
54		L557G	+
55		A373F	+
56		E707S	+
57		T384F/P513Q/L536M	+
58		R748G	+
59		F256Q	+
60		A517P	+
61	17/18	T384A	+
62		L557S	+
63		D623R	+
64		Q554E	+
65		T384G	+
66		K723G	+
67	19/20	A255Q	+
68		D268V	+
69		Q559P	+
70		S435R	+
71		A255T	+
72		K723A	+
73		G260A	+
74		T705G	+
75		N627H	+
76		L257I	+

TABLE 6.1-continued

Activity of Deacylating Variants Relative to SEQ ID NO: 6

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 6)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 6
77		S530Y	+
78		K322P	+
79		A517L	+
80		G54C	+
81		I127V	+
82		T62G	+
83		G461A	+
84		S325G	+
85		S372H	+
86		D623G	+
87		A255L	+
88		T378H	+
89		S372L	+

[1]Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 6 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase; "++" > than 2.5-fold but less than 5-fold.

EXAMPLE 7

Improvements in the Deacylation of Insulin Compared to SEQ ID NO: 8 in High Throughput Screening

[0188] SEQ ID NO: 8 was selected as the next parent enzyme, based on the results described in Example 6. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and 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.

[0189] HTP reactions were carried out in 96 well deep-well plates containing 200 μL of 10 g/L tetraprotected insulin dimer substrate, 200 mM TEoA buffer, pH 8.5, 30% DMSO and 10 μL HTP lysate. The HTP plates were heat sealed and incubated in Thermotron® at 30° C., 300 rpm, for 18 hours. The reactions were quenched with 200 μL MeCN and mixed for 5 minutes using a bench top shaker. The plates were then centrifuged at 4000 rpm for 5 minutes, diluted 24× into water, and injected onto an UPLC for analysis using the parameters in Table 12.1.

[0190] The percent conversion relative to SEQ ID NO:8 (Percent Conversion FIOP) was calculated as the percent conversion of the product formed by the variant over the percent conversion produced by SEQ ID NO: 8 The results are shown in Table 7.1. The percent conversion was quantified by dividing the area of the product peak by the sum of the areas of the substrate, and product peaks as determined by HPLC analysis.

TABLE 7.1

Activity of Deacylating Variants Relative to SEQ ID NO: 8

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 8)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 8
90	29/30	K103V/K369W/I370F/G444S/Q556G/S706G/H766G	++
91	25/26	K369W/I370F/G444S/Q556G/V612A/H766G	++
92		K369W/I370F/G444S/S706G/H765P/H766G	++
93	31/32	K369V/I370F/N388T/G444S/Q556G/H766G	++
94	27/28	K369W/I370F/G444S/Q556G/S706G/H765P	++
95		K369V/I370F/Q556G	++
96		K369W/I370F/G444S/Q556G	++
97		K369V/I370F/G444S/Q556G/H766G	++
98		K369P/I370F/Q556G/H766G	++
99		K103V/K369W/I370F/G444S/H765P/H766G	++
100		K369V/I370F/Q556G/H766G	++
101		K103V/K369W/I370F/V442I/G444S/L536M/Q556G/H766G	++
102		T384A/A451R/T560G/T705D/K723L	++
103		K103V/I370F/G444S/S706G/H766G	++
104		A451R/T560G/T705D/K723L	++
105		K369V/I370F/G444S	++
106		A451R/T705D/K723L	++
107		K369V/I370F/H766G	++
108		L257V/A362V/T384A/A451R	++
109		T384L/A451R/T705D/K723L	++
110		S372A/T384L/A451R/T705D	++
111		K369P/I370F/G444S/S706G/H766G	++
112		K103V/K369W/I370F/G444S	++
113		K369V/I370F/Q556G/H765P	++
114		L257V/A362V/T384L/A451R/K723L	++
115		K369P/I370F	++
116		K103V/K369W/I370F/H765P/H766G	++
117	33/34	S372A/A373M/T384L/P513Q/T560G	++
118	21/22	A362V/A451R/T705D	++
119	23/24	S372A/T384L/T560G/T705D	+

[1]Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 8 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase; "++" > than 2.5-fold but less than 5-fold; "+++" > than 5-fold.

EXAMPLE 8

Improvements in the Deacylation of Insulin Compared to SEQ ID NO: 34 in High Throughput Screening in 30% DMSO

[0191] SEQ ID NO: 34 was selected as the next parent enzyme, based on the results described in Example 7. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and 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.

[0192] For Table 8.1, HTP reactions were carried out in 96 well deep-well plates containing 2004 of 20 g/L diprotected insulin dimer substrate (A1, A1'-diphenylacetimido-insulin tethered-dimer), 200 mM TEoA buffer, pH 8.5, 30% DMSO and 2.54 HTP lysate. The HTP plates were heat sealed and incubated in Thermotron® shakers at 30° C., 300 rpm, for 3.5 hours. The reactions were quenched with 1:5 DMAc and mixed for 5 min using a bench top shaker. The plates were then centrifuged at 4000 rpm for 5 min, and injected onto an UPLC for analysis using the parameters in Table 12.2.

[0193] For Table 8.2, HTP reactions were carried out in 96 well deep-well plates containing 200 of 20 g/L Diprotected Insulin Dimer substrate, 200 mM TEoA buffer, pH 8.5, 20% DMSO and 2.54 HTP lysate. The HTP plates were heat sealed and incubated in Thermotron® shakers at 30° C., 300 rpm, for 5 hours. The reactions were quenched with 1:5 DMAc and mixed for 5 min using a bench top shaker. The plates were then centrifuged at 4000 rpm for 5 min, and injected onto an UPLC for analysis using the parameters in Table 12.2.

[0194] The percent conversion relative to SEQ ID NO:34 (Percent Conversion FIOP) was calculated as the percent conversion of the product formed by the variant over the percent conversion produced by SEQ ID NO: 34 The results are shown in Tables 8.1 and 8.2. The percent conversion was quantified by dividing the area of the product peak by the sum of the areas of the substrate, and product peaks as determined by UPLC analysis.

TABLE 8.1

Activity of Deacylating Variants Relative to SEQ ID NO: 34

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 34)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 34
120		D403T	+
121		P275E	+
122		A664G	+
123		A747G	+
124		K622R	+
125		Q541A	+
126		Q759N	+
127		L55V	+
128		E482A	+
129		P496K	+
130		A616G	+
131		E482S	+
132		S639G	+
133		K619N/A664G	+

[1]Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 34 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase; "++" > than 2.5-fold but less than 5-fold; "+++" > than 5-fold.

TABLE 8.2

Activity of Deacylating Variants Relative to SEQ ID NO: 34

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 34)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 34
134		L253H/K369W/D623G/K723A	++
135		L253H/W254Q/K322T/K369W/D623G	++
136	41/42	K103V/A372S/M373F/L557G	++
137	37/38	G260S/A372S/M373F/Q556G/L557V/Q559S	+
138		L253H/K322T/K369W/M373W/K723A	+
139		L253H/W254Q/K369W/K619R/D623G/K723A	+
140		L253H/K322T/K369W/D623G	+
141	39/40	G260S/A372S/M373F/Q556G	+
142	47/48	L253H/W254S/K369W	+
143	43/44	L253H/W254Q/K369W/V391A/D623G/K723A	+
144	35/36	K369W/M373F/Q556G	+
145	45/46	L253H/W254Q/K369W/D623G/K723A	+
146		L253S/W254S/G255V/K322T/K369W/K619R/K723A	+
147		L253H/M373L/D623S	+
148		K322T/K369W/D623G/K723A	+
149		L253H/W254Q/M373L/D623G/K723A	+
150		K322T/K369W/M373W/K723A	+
151		L253H/W254S/G255V/K369W/D623S/K723A	+
152		K322T/K369W	+

[1]Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 34 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase; "++" > than 2.5-fold but less than 5-fold; "+++" > than 5-fold.

EXAMPLE 9

Improvements in the Deacylation of Insulin Compared to SEQ ID NO: 46 in High Throughput Screening

[0195] SEQ ID NO: 46 was selected as the next parent enzyme, based on the results described in Example 8. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and 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.

[0196] HTP reactions were carried out in 96 well deep-well plates containing 200 μL of 30 g/L diprotected insulin dimer substrate, 200 mM TEoA buffer, pH 8.5, 30% DMSO and 10 μL HTP lysate. The HTP plates were heat sealed and incubated in Thermotron® shakers at 30° C., 300 rpm, for 5 hours. The reactions were quenched with 1:5 DMAc and mixed for 5 min using a bench top shaker. The plates were then centrifuged at 4000 rpm for 5 min, and injected onto an UPLC for analysis using the parameters in Table 12.2 or 12.3.

[0197] The percent conversion relative to SEQ ID NO:46 (Percent Conversion FIOP) was calculated as the percent conversion of the product formed by the variant over the percent conversion produced by SEQ ID NO: 46 The results are shown in Table 9.1. The percent conversion was quantified by dividing the area of the product peak by the sum of the areas of the substrate, and product peaks as determined by UPLC analysis.

TABLE 9.1

Activity of Deacylating Variants Relative to SEQ ID NO: 46

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 46)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 46
153		A71F	++
154		M373A/E482C/Y569W/K619N/A764S	++
155		T176S/M373F/E482A/K622V	++
156		Q233E/M373F/E482A/K622V/A664G	++
157		T176S/M373F/E482A/K622F/A664G	++
158		Q233E/P275E/E482C/K619N	++
159		T176S/M373F/E482A/Y569W	++
160		K103V/G260S/K322T/N348A/G444S/Q556G/L557G/G623D	++
161		T176S/E482A	++
162		K146M/N309D/Q556N/K619S/R748A	++
163		T176S/Q233E/M373A/K619N/A664R	++
164		Q233E/P275E/E482A/Y569W/A664G	++
165	51/52	K103V/L257V/G260S/K322T/N348A/L384T/G444S/ Q556G/G623D	++
166	55/56	N9K/K103V/H253S/K322T/N348A/G444S/Q556G/L557G/ G623D	++
167		N9K/G25V/K103V/H253S/N348A/G444S/L557G/G623D	++
168		E482S/G623D	++
169		K304I/P496K/A616S/K619N/A664E/A747P/F756P/Q759E	++
170	57/58	K103V/G260S/K322T/N348A/M373A/V391A/G444S/ Q556G/L557G/G623D	++
171		N9K/K103V/K322T/V391A/G444S/L557G/G623D	++
172		L225T/K304I/N494E/A616G/K619N/A664G/A747P/ Q759E	++
173		N494E/P496K/A616S/K619N/A664E	++
174		L225K/K304C/N309V/Q556N/L557R/K619S/R748A	++
175		T176S/M373F/E482C/Y569W/K622C/G623D/A764S	++
176		G25V/K103V/N241K/H253S/K322T/N348A/G444S/ Q556G/L557G/G623D	++
177		L225T/K304I/K322T/N494E/P496N/A616G/K619N/ A664G/A747S/F756P	++
178		K103V/L257V/G260S/K322T/N348A/G444S/L557G	++
179	53/54	K103V/G260S/K322T/N348A/G444S/G623D	++
180	59/60	K322T/N348A/M373A/V391A/G444S/Q556G/G623D	++
181	61/62	K103V/K322T/N348A/M373A/G444S/Q556G/L557G	+
182	49/50	K322T/N348A/G444S/L557G	+
183		A71F	++
184		K619P	+
185		K619V	+
186		A71C	+
187		K619A	+
188		A71L	+
189		K619H	+
190		K619S	+
191		T705N	+
192		K128H	+
193		A71G	+
194		Q626E	+
195		F617W	+
196		A616D	+
197		W369A	+
198		M373G	+
199		K619L	+
200		N28A	+
201		T129E	+
202		A616E	+
203		I370M	+
204		W369L/A764G	+
205		A616N	+
206		I370Q	+
207		A451H	+
208		W369L	+
209		I389V	+
210		K622V	+
211		T379S	+
212		I77V	+
213		K622I	+
214		Q380D	+
215		G111S	+
216		N28S	+

TABLE 9.1-continued

Activity of Deacylating Variants Relative to SEQ ID NO: 46

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 46)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 46
217		A616G	+
218		N28Q	+
219		W369E	+
220		T131D	+
221		W369V	+
222		I77T	+
223		R471V	+
224		Q626D	+
225		A616T	+
226		T379D	+
227		A616Q	+
228		N28C	+

[1]Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 46 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase; "++" > than 2.5-fold but less than 5-fold; "+++" > than 5-fold.

EXAMPLE 10

Improvements in the Deacylation of Insulin Compared to SEQ ID NO: 54 in High Throughput Screening

[0198] SEQ ID NO: 54 was selected as the next parent enzyme, based on the results described in Example 9. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and 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.

[0199] HTP reactions were carried out in 96 well deep-well plates containing 200 μL of 30 g/L diprotected insulin dimer substrate, 200 mM TEoA buffer, pH 8.5, 30% DMSO and 2.5 μL HTP lysate. The HTP plates were heat sealed and incubated in Thermotron® shakers at 30° C., 300 rpm, for 5 hours. The reactions were quenched with 1:5 DMAc and mixed for 5 min using a bench top shaker. The plates were then centrifuged at 4000 rpm for 5 min, and injected onto an UPLC for analysis using the parameters in Table 12.3.

[0200] The percent conversion relative to SEQ ID NO:54 (Percent Conversion FIOP) was calculated as the percent conversion of the product formed by the variant over the percent conversion produced by SEQ ID NO: 54. The results are shown in Table 10.1. The percent conversion was quantified by dividing the area of the product peak by the sum of the areas of the substrate, and product peaks as determined by HPLC analysis.

TABLE 10.1

Activity of Deacylating Variants Relative to SEQ ID NO: 54

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 54)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 54
229	73/74	A71L/K128H/M373A/E482S/A664E/P753C	+++
230	75/76	A71L/K128H/T176S/E482S/P496K	+++
231	71/72	A71L/K128H/T176S/M373A/E482C/P496K/K619S	+++
232		A71F/T176S/P275C/E482S	+++
233	69/70	K128H/T176S/M373A/E482S/A664E	+++
234		A71L/T176S/E482A/K619P/A664D/Q759E	+++
235		A71L/T176S/A451H/E482A/K619V/Q759E	+++
236		A71L/K128H/T176S/M373A/E482S/P496K/Y569C	+++
237		A71L/T176S/E482A	+++
238		A71L/Q233E/S260G/E482A/L557G/Q759E	+++
239		K128H/T176S/Q233E/M373A/E482S/Q626E/P753C	+++
240	63/64	A71L/T176S/S260G/P275C/E482A/L557G/Q759E	+++
241		A71L/K128H/T176S/P496K/A664E	+++
242		K128H/T176S/Q233E/P496K/A664E/P753C	+++
243		A71F/T176S/S260G/A451H/K619V	+++
244	67/68	A71L/T176S/Q233E/S260G/A451H/E482S/A664C/Q759E	++
245	65/66	A71F/T176S/Q233E/S260G/P275C/E482S/K619N/Q759D	++
246		A71L/T176S/M373A/Q626E/A664E/P753C	++
247		A71F/T176S/P275E/A664D	++
248		A71L/T176S/Q233E/M373A/E482C/Y569C/P753C	++
249		A71L/T176S/S260G/E482A	++
250		N28A/A71L/K128H/T176S/Q626D/P753C	++
251		K128H/T176S/M373A/P496K/P753C	++

TABLE 10.1-continued

Activity of Deacylating Variants Relative to SEQ ID NO: 54

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 54)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 54
252		A71L/T176S/S260G/E482A/L557G/K619P/A664D	++
253		N28A/A71L/K128H/T176S/K619N/A664E	++
254		T176S/Q233E/A451H/E482S/K619N/A664C/Q759D	++
255		A71F/T176S/Q233E/E482A	++
256		A71L/M373A/F756C	++
257		A71L/S260G/A451H/E482A/A664D/Q759E	++
258		T176S/Q233E/S260G/P275E/E482C/A664E/Q759D	++
259		A71M	++
260		A71F	++
261		A71G	+
262		A71L	+
263		Y180F	+
264		A71I	+
265		L122M	+
266		L82V	+
267		P739D	+
268		A71V	+
269		W658C	+
270		F679L	+
271		P496K	+
272		V184F	+
273		V184A	+
274		P739S	+
275		H472F	+
276		H472V	+
277		P686A	+
278		V126L	+

[1]Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 54 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase; "++" > than 2.5-fold but less than 5-fold; "+++" > than 5-fold.

EXAMPLE 11

Improvements in the Deacylation of Insulin Compared to SEQ ID NO: 74 in High Throughput Screening

[0201] SEQ ID NO: 74 was selected as the next parent enzyme, based on the results described in Example 10. Libraries of engineered genes were produced using well-established techniques (e.g., saturation mutagenesis and 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.

[0202] HTP reactions were carried out in 96 well deep-well plates containing 200 μL of 30 g/L diprotected insulin dimer substrate, 200 mM TEoA buffer, pH 9, 30% DMSO and 1.25 μL HTP lysate.

[0203] The HTP plates were heat sealed and incubated in Thermotron® shakers at 35 ° C., 300 rpm, for 5 hours. The reactions were quenched with 1:5 DMAc and mixed for 5 min using a bench top shaker. The plates were then centrifuged at 4000 rpm for 5 min, and injected onto an UPLC for analysis using the parameters in Table 12.3.

[0204] The percent conversion relative to SEQ ID NO:74 (Percent Conversion FIOP) was calculated as the percent conversion of the product formed by the variant over the percent conversion produced by SEQ ID NO: 74 The results are shown in Table 11.1. The percent conversion was quantified 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 UPLC analysis.

TABLE 11.1

Activity of Deacylating Variants Relative to SEQ ID NO: 74

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 74)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 74
279	77/78	T176S/P275C/Y569W/Q759D	+++
280	81/82	T176S/L557G/Y569W/A616T	+++
281	83/84	T176S/Q759D	++
282		T176S/A616S	++
283		T176S/A348M/L557G/Y569W/A616G	++
284		T176S/L557G/Y569W/A616G/I708L	++
285		T176S	++
286		T176S/P275C/A348M/L557G/Q759D	++

TABLE 11.1-continued

Activity of Deacylating Variants Relative to SEQ ID NO: 74

Variant NO:	SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 74)	Deacylation Percent Conversion (FIOP)[1] Relative to SEQ ID NO: 74
287		T176S/Q233E	++
288		T176S/Q233E/Q759E	++
289	85/86	T176S/T352S	++
290		T176S/P275E	++
291	89/90	L71C/A451H/R748A	++
292		T176S/L557G/K619G	++
293		L71F/S353A/R357A/A451H/T705N/R748A	++
294	87/88	G111S/T176S/T352S	++
295		T176S/A616G/K619R	++
296		T176S/L557G/I708L	++
297		T176S/A616T	++
298		T176S/Q233E/T352S	++
299	79/80	L71F/A451H/Q556N/T705N/R748A	++
300		T176S/P275E/L557G/Q759E	++
301		T176S/Q233E/L557G/K619G/Q759D	++
302		T176S/A616G	++
303		T176S/L557G/A616N	++
304		T176S/A361T	++
305		T176S/Y569W/A616G/K619S/Q759D	++
306		L71C/T352S	++
307		T176S/S482C/A616G/Q759E	++
308		I77T/T176S/A712V	++

[1]Levels of increased activity or selectivity were determined relative to the reference polypeptide of SEQ ID NO: 74 and defined as follows: "+" > than 1.2-fold but less than 2.5-fold increase; "+" > than 2.5-fold but less than 5-fold; "+++" > than 5-fold.

EXAMPLE 12

Analytical Detection of Insulin Dimer and its Deacylated Products

[0205] Data described in Examples 5-11 were collected using analytical methods in Tables 12.1, 12.2, 12.3. The methods provided herein all find use in analyzing the variants produced using the present invention. However, it is not intended that the methods described herein are the only methods applicable to the analysis of the variants provided herein and/or produced using the methods provided herein.

TABLE 12.1

Analytical Method

Instrument	Thermo Scientific Vanquish ™ UPLC
Column	Waters Cortecs ® C18, 2.7 × 50 mm, 1.6 μM
Mobile Phase	Gradient I (A: 0.1% TFA in water; B: 0.1% TFA in MeCN)

Time(min)	% A
0.000	69
1.500	50
1.550	5
1.950	5
2.000	69

Flow Rate	1.000 mL/min
Run Time	2.400 min
Product Elution order	Insulin dimer (0.77 min); A1-acylated insulin dimer (0.83 min); B1-acylated insulin dimer (0.91 min); A1,A1'-diacylated insulin dimer (0.99 min); B1,B1'-diacylated insulin dimer (1.1 min); A1, A1', B1-triacylated insulin dimer & A1, B1, B1'-triacylated insulin dimer (1.15 min); A1,A1',B1,B1'-tetraacylated insulin dimer (1.25 min).
Column Temperature	40° C.
Injection Volume	1.0 μL
Detection	UV 218 nm; Detector: MWD-Data Collection Rate: 20 Hz

TABLE 12.2

Analytical Method	
Instrument	Thermo Scientific Vanquish ™ UPLC
Column	Waters Cortecs ® C18, 2.7 × 50 mm, 1.6 μM
Mobile Phase	Gradient I (A: 0.1% TFA in water; B: 0.1% TFA in MeCN)

Time(min)	% A
0.000	69
0.900	50
0.950	5
1.300	5
1.350	100
1.650	69

Flow Rate	0.800 mL/min
Run Time	2.000 min
Product	Insulin dimer (0.75 min); A1-acylated insulin dimer
Elution order	(0.82 min); A1,A1'-diacylated insulin dimer (0.89 min).
Column Temperature	40° C.
Injection Volume	1.0 μL
Detection	UV 218 nm; Detector: MWD-Data Collection Rate: 20 Hz

TABLE 12.3

Analytical Method	
Instrument	Thermo Scientific Vanquish ™ UPLC
Column	Thermo Hypersil ™ Gold C18, 2.1 × 50 mm, 1.9 μM
Mobile Phase	Gradient I (A: 0.1% TFA in water; B: 0.1% TFA in MeCN)

Time(min)	% A
0.000	70
0.900	53
0.950	5
1.300	5
1.350	70

Flow Rate	0.950 mL/min
Run Time	1.500 min
Product	Insulin dimer (0.61 min); A1-acylated insulin dimer (0.65
Elution order	min); A1,A1'-diacylated insulin dimer (0.69 min).
Column Temperature	40° C.
Injection Volume	1.0 μL
Detection	UV 218 nm; Detector: MWD-Data Collection Rate: 20 Hz

[0206] All publications, patents, patent applications and other documents cited in this application are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference for all purposes.

[0207] While various specific embodiments have been illustrated and described, it will be appreciated that various changes can be made without departing from the spirit and scope of the invention(s).

SEQUENCE LISTING

```
Sequence total quantity: 90
SEQ ID NO: 1          moltype = DNA  length = 2541
FEATURE               Location/Qualifiers
misc_feature          1..2541
                      note = Penicillin G Acylase From Kluyvera citrophila
source                1..2541
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 1
atgaaaaata gaaatcgtat gatcgtgaac ggtattgtga cttccctgat ctgttgttct  60
agcctgtcag cgctggcggc aagcccgcca accgaggtta agatcgttcg cgatgaatac  120
ggcatgccgc atatttacgc cgatgatacc tatcgactgt tttacggcta tggctacgtg  180
gtggcgcagg atcgcctgtt ccagatggaa atggcgcgcc gcagtactca ggggaccgtc  240
tccgaggtgc tgggcaaagc attcgtcagt tttgataaag atattcgcca gaactactgg  300
ccggattcta ttcgcgcgca gatagcttcc ctctccgctg aggataaatc cattctgcag  360
ggctatgccg atggcatgaa tgcgtggatc gataaagtga acgccagccc cgataagctg  420
ttaccccagc agttctccac ctttggtttt aaacccaagc attgggaacc gtttgatgtg  480
gcgatgattt ttgtcggcac catggcgaac cggttttctg acagcaccag cgaaattgat  540
aacctggcgc tgctgacggc gctaaaagat aaatacggca agcagcaggg catggcggtc  600
tttaaccagc tgaaatggct ggttaatcct tccgcgccaa ccaccattgc ggcgcgggaa  660
agcgcctatc cgctgaagtt tgatctgcaa aacacgcaaa cggcggcgct gctgccgcgc  720
tacgaccagc cggcaccgat gctcgaccgc ccggcaaaag ggaccgatgg cgcgctgctg  780
gcgctgaccg ccgatcagaa ccgggaaact atcgccgcgc agttcgcgca aagcggcgct  840
aacggcctgg ctggctaccc gaccactagc aatatgtggg tgattggcaa aaacaaagcc  900
caggatgcga aggccattat ggtcaatggg ccgcagtttg gttggtatgc gccggcgtac  960
acctacggta tcggcctgca cggcgcgggc tatgacgtca ccggcaatac gccgtttgcc  1020
tatccgggcc tcgtttttgg tcacaacggc accatttcat ggggatccac cgccggtttt  1080
ggtgatgatg tcgatatctt tgccgaaaaa ctttccgccg agaagccggg ctattaccag  1140
cataacggcg agtgggtgaa gatgttgagc cgcaaggaga ctattgcggt caaagacggc  1200
cagccggaga cctttaccgt ttggcgcacg ctgcacggca acgtcattaa aaccgatact  1260
gcgacgcaga ccgcctatgc caaagcgcgc gcctgggatg gcaaagaggt ggcgtccctg  1320
ctggcgtgga cgcaccagat gaaggccaaa aactggccgg agtggacgca gcaggcggcc  1380
aaacaggcgc tgaccattaa ctggtactac gccgatgtga acggcaatat cggctatgtg  1440
cataccggcg cctatccgga tcgccagccc ggccacgacc cgcgtttgcc ggttcccggc  1500
actggaaaat gggactggaa agggttgctg tcgtttgatt tgaatccgaa agtgtataac  1560
ccgcagtcgg gctatatcgc caactggaac aactcgccgc aaaaagacta cccggcctct  1620
gatctgttcg cgttcctgtg gggcggtgcg gatcgagtta ctgagatcga cacgatcctc  1680
```

-continued

```
gataagcaac cgcgcttcac cgccgatcag gcgtgggatg tgatccgcca aaccagccgt  1740
cgggatctca acctgcggtt gttcttaccg gcgctgaagg acgccaccgc gaacctggcg  1800
gaaaacgatc cgcgccgcca actggtggat aaactggcga gctgggacgg tgaaaacctt  1860
gtcaacgatg acggaaaaac ctatcagcaa ccgggatcgg cgattcttaa cgcctggctg  1920
accagcatgc tcaagcgcac ggtggttgcc gcggtcccag cgccgtttgg caagtggtac  1980
agcgccagtg gctatgaaac cacccaggac gggccaaccg gctcgctgaa catcagcgtg  2040
ggggcgaaaa tcctctacga agctctgcag ggtgataagt cgccaatccc gcaggcggtc  2100
gatctgtttg gcgggaaacc gcagcaggaa gtgatactgg cggcgctgga cgacgcttgg  2160
cagacgctgt caaaacgcta cggtaacgac gtcaccggct ggaaaacccc tgccatggcg  2220
cttaccttcc gggccaataa cttcttcggc gtgccgcagg cggcagcaaa agaggcgcgt  2280
catcaggcgg agtaccagaa ccgcggtacg gaaaacgaca tgattgtctt ctcaccgacg  2340
tcgggtaacc gcccggttct tgcctgggat gtggtggcgc cggggcaaag cggttttatc  2400
gcgccggatg gcaaagccga taagcactat gacgatcagc tgaaaatgta cgagagcttt  2460
ggccgtaaat cgctgtggtt aacgcctcag gacgttgacg agcacaaaga gtctcaggaa  2520
gtgctgcagg tacagcgcta a                                            2541

SEQ ID NO: 2           moltype = AA  length = 846
FEATURE                Location/Qualifiers
REGION                 1..846
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..846
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 2
MKNRNRMIVN GIVTSLICCS SLSALAASPP TEVKIVRDEY GMPHIYADDT YRLFYGYGYV  60
VAQDRLFQME MARRSTQGTV SEVLGKAFVS FDKDIRQNYW PDSIRAQIAS LSAEDKSILQ  120
GYADGMNAWI DKVNASPDKL LPQQFSTFGF KPKHWEPFDV AMIFVGTMAN RFSDSTSEID  180
NLALLTALKD KYGKQQGMAV FNQLKWLVNP SAPTTIAARE SAYPLKFDLQ NTQTAALLPR  240
YDQPAPMLDR PAKGTDGALL ALTADQNRET IAAQFAQSGA NGLAGYPTTS NMWVIGKNKA  300
QDAKAIMVNG PQFGWYAPAY TYGIGLHGAG YDVTGNTPFA YPGLVFGHNG TISWGSTAGF  360
GDDVDIFAEK LSAEKPGYYQ HNGEWVKMLS RKETIAVKDG QPETFTVWRT LHGNVIKTDT  420
ATQTAYAKAR AWDGKEVASL LAWTHQMKAK NWPEWTQQAA KQALTINWYY ADVNGNIGYV  480
HTGAYPDRQP GHDPRLPVPG TGKWDWKGLL SFDLNPKVYN PQSGYIANWN NSPQKDYPAS  540
DLFAFLWGGA DRVTEIDTIL DKQPRFTADQ AWDVIRQTSR RDLNLRLFLP ALKDATANLA  600
ENDPRRQLVD KLASWDGENL VNDDGKTYQQ PGSAILNAWL TSMLKRTVVA AVPAPFGKWY  660
SASGYETTQD GPTGSLNISV GAKILYEALQ GDKSPIPQAV DLFGGKPQQE VILAALDDAW  720
QTLSKRYGND VTGWKTPAMA LTFRANNFFG VPQAAAKEAR HQAEYQNRGT ENDMIVFSPT  780
SGNRPVLAWD VVAPGQSGFI APDGKADKHY DDQLKMYESF GRKSLWLTPQ DVDEHKESQE  840
VLQVQR                                                            846

SEQ ID NO: 3           moltype = DNA  length = 2295
FEATURE                Location/Qualifiers
misc_feature           1..2295
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2295
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 3
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgcgcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt tcgcgttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggtcc cagcgccgtt tggtaagtgg tacagcgcca gtggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
```

-continued

```
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgtaa                                                   2295

SEQ ID NO: 4           moltype = AA  length = 764
FEATURE                Location/Qualifiers
REGION                 1..764
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..764
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 4
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLFAFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKW YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTA                   764

SEQ ID NO: 5           moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 5
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgcgcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt tcgcgttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggtcc cagcgccgtt tggtaagtgg tacagcgcca gtggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
```

-continued

```
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 6           moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 6
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLFAFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKW YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 7           moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 7
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggtaagatc tacagcgcct cgggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
```

-continued

```
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 8           moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 8
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKI YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 9           moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 9
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgcgcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtagtt  1080
gccgcggttc cagcgccgtt tggtaagtgg tacagcgcca gtggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
```

-continued

```
acgcaaacgg cgcaccatca ccatcaccat taa                                2313

SEQ ID NO: 10          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 10
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKW YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 11          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 11
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgcgcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt tcgcgttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggtcc cagcgccgtt tggtaagttt tacagcgcca gtggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                                2313

SEQ ID NO: 12          moltype = AA  length = 770
FEATURE                Location/Qualifiers
```

-continued

```
REGION                  1..770
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..770
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 12
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLFAFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKF YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 13           moltype = DNA  length = 2313
FEATURE                 Location/Qualifiers
misc_feature            1..2313
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..2313
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 13
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgcgcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt tcgcgttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggtcc cagcgccgtt tggtaagtgg tacagcatga gtggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 14           moltype = AA  length = 770
FEATURE                 Location/Qualifiers
REGION                  1..770
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..770
```

-continued

```
                     mol_type = protein
                     organism = synthetic construct
SEQUENCE: 14
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLFAFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKW YSMSGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 15          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 15
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgcgcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt tcgcgttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggtcc cagcgccgtt tggtccgtgg tacagcgcca gtggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 16          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 16
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
```

-continued

```
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLFAFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGPW YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 17          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 17
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgcgcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt tcgcgttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggtcc cagcgccgtt tggtaagtgg tacagcgcca gtggctatga aaccacccag  1140
gacgggccag ctggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 18          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 18
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLFAFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
```

-continued

```
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKW YSASGYETTQ DGPAGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 19          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 19
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgcgcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt tccagttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggtcc cagcgccgtt tggtaagtgg tacagcgcca gtggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 20          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 20
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLFQFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKW YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
```

-continued

```
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 21          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 21
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgtggttc cagcgccgtt tggtaagatc tacagcgcct cgggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg aggcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagacg  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcgacagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 22          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 22
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AVVPAPFGKI YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM RLTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSDSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770
```

-continued

SEQ ID NO: 23 moltype = DNA length = 2313
FEATURE Location/Qualifiers
misc_feature 1..2313
note = Variant of Penicillin G Acylase From Kluyvera citrophila
source 1..2313
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 23
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat 60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg 120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac 180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa 240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg 300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc 360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg 420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc 480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac 540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag 600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg 660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg 720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttcct gtggggcggt 780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat 840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta 900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg 960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggtaagatc tacgcggcgt cgggctatga aaccacccag 1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagggt 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt 1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcgacagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcaccatca ccatcaccat taa 2313

SEQ ID NO: 24 moltype = AA length = 770
FEATURE Location/Qualifiers
REGION 1..770
note = Variant of Penicillin G Acylase From Kluyvera citrophila
source 1..770
mol_type = protein
organism = synthetic construct
SEQUENCE: 24
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN 60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR 120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY 180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW 240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL 300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV 360
AAVPAPFGKI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ 420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG 480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP 540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM 600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID 660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSDSEIDN LALLTALKDK 720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH 770

SEQ ID NO: 25 moltype = DNA length = 2313
FEATURE Location/Qualifiers
misc_feature 1..2313

-continued

```
                     note = Variant of Penicillin G Acylase From Kluyvera
                      citrophila
source               1..2313
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 25
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggttc tacagcgcct cgggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtagggtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggcgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcatggaca ccatcaccat taa                               2313

SEQ ID NO: 26        moltype = AA  length = 770
FEATURE              Location/Qualifiers
REGION               1..770
                     note = Variant of Penicillin G Acylase From Kluyvera
                      citrophila
source               1..770
                     mol_type = protein
                     organism = synthetic construct
SEQUENCE: 26
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGWF YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EALGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHGHHHH             770

SEQ ID NO: 27        moltype = DNA  length = 2313
FEATURE              Location/Qualifiers
misc_feature         1..2313
                     note = Variant of Penicillin G Acylase From Kluyvera
                      citrophila
source               1..2313
                     mol_type = other DNA
```

-continued

```
                     organism = synthetic construct
SEQUENCE: 27
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggttc tacagcgcct cgggctatga aaccacccag  1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtagggtt ggatcagacc  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccggcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcctcatca ccatcaccat taa                               2313

SEQ ID NO: 28          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 28
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGWF YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTGEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAPHHHHH             770

SEQ ID NO: 29          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 29
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
```

-continued

```
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgttg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggttggttc tacagcgcct cgggctatga aaccacccag 1140
gacgggccaa ccggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcacca gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtagggtt ggatcagacc 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt 1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcaccggcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcatggaca ccatcaccat taa                              2313

SEQ ID NO: 30          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 30
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGWF YSASGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTGEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHGHHHH             770

SEQ ID NO: 31          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 31
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
```

-continued

```
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttcct gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggtgtcttc tacagcgcct cgggctatga aaccacccag 1140
gacgggccaa ccggctcgct gaccatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcacca gctggaaaac ccctgccatg gcgcttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgccgg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtagggtt ggatcagacc 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt 1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcatggaca ccatcaccat taa                              2313

SEQ ID NO: 32          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 32
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGVF YSASGYETTQ DGPTGSLTIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAPDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGLDQT EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHGHHHH             770

SEQ ID NO: 33          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 33
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
```

-continued

```
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttctt gtggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggtaagatc tacgcgatgt cgggctatga aaccacccag 1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagggt 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt 1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcaccatca ccatcaccat taa                              2313

SEQ ID NO: 34          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 34
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 35          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 35
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttctt gtggggcggc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
```

-continued

```
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggttggatc tacgcgtttt cgggctatga aaccacccag 1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaggttt agatcagggt 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt 1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcaccatca ccatcaccat taa                              2313

SEQ ID NO: 36          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 36
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGWI YAFSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 37          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 37
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttctt gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
```

-continued

```
gccgcggttc cagcgccgtt tggtaagatc tactcgtttt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaggtgt tgattcgggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 38          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 38
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKI YSFSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGVDSG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 39          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 39
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttctt gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggtaagatc tactcgtttt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
```

-continued

```
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaggttt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcacgagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaaac agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 40          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 40
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKI YSFSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 41          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 41
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgttg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatctgt ggggcttctt gtggggcggc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggtaagatc tactcgtttt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
```

-continued

```
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacaagg tgatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcacgagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaaac agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 42          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 42
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDLWGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGKI YSFSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQGDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 43          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 43
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcatc aaggtttcct ttggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatc tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gcgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
```

-continued

```
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaaggca ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 44          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 44
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS AGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKGIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 45          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 45
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcatc aaggtttcct ttggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatc tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaaggca ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
```

-continued

```
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 46          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 46
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKGIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 47          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 47
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcatt cgggtttcct ttggggcggt  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gataaactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct taacgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatc tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcaccg gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaagttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcaagc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
```

-continued

```
acgcaaacgg cgcaccatca ccatcaccat taa                                2313

SEQ ID NO: 48          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 48
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHSGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DKLASWDGEN LVNDDGKTYQ QPGSAILNAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTGWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGKQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 49          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 49
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcggc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacaggg ggatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaaggca ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                                2313

SEQ ID NO: 50          moltype = AA  length = 770
FEATURE                Location/Qualifiers
```

-continued

```
REGION                  1..770
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..770
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 50
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQGDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKGIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 51           moltype = DNA  length = 2313
FEATURE                 Location/Qualifiers
misc_feature            1..2313
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..2313
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 51
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcgt gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcgatgt cgggctatga aaccacccag  1140
gacgggccaa caggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaggttt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 52           moltype = AA  length = 770
FEATURE                 Location/Qualifiers
REGION                  1..770
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..770
```

-continued

```
                     mol_type = protein
                     organism = synthetic construct
SEQUENCE: 52
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFVWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPTGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 53          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 53
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 54          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 54
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
```

-continued

```
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 55          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 55
agcaatatgt gggtgattgg caaaaaaaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgattctc aaggtttcct gtggggcggc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaggtgg ggatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 56          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 56
SNMWVIGKKK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDSQGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
```

-continued

```
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGGDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 57          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 57
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gcgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaggtgg ggatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 58          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 58
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS AGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
```

-continued

```
QDVDEHKESQ EVLQVGGDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 59          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 59
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcaagg agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcggc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag 1140
gacgggccac tcggctcgct gaacatcagc gcgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaggttt agatcagggt 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt 1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 60          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 60
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRKETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS AGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVGLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770
```

-continued

SEQ ID NO: 61 moltype = DNA length = 2313
FEATURE Location/Qualifiers
misc_feature 1..2313
note = Variant of Penicillin G Acylase From Kluyvera citrophila
source 1..2313
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 61
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat 60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg 120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac 180
ggcaccattt catggggatc caccgccggt gccggtgata gcgtcgatat ctttgccgaa 240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg 300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc 360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg 420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc 480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac 540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag 600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg 660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg 720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcggc 780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat 840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta 900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg 960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag 1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acggaaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaggtgg ggatcagggt 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt 1860
gataaaggca ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcaccatca ccatcaccat taa 2313

SEQ ID NO: 62 moltype = AA length = 770
FEATURE Location/Qualifiers
REGION 1..770
note = Variant of Penicillin G Acylase From Kluyvera citrophila
source 1..770
mol_type = protein
organism = synthetic construct
SEQUENCE: 62
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN 60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR 120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY 180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW 240
NNSPQKDYPA SDHQGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL 300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV 360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ 420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG 480
TENNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP 540
QDVDEHKESQ EVLQVGGDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM 600
ARRSTQGTVS EVLGKAFVKF DKGIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID 660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK 720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH 770

SEQ ID NO: 63 moltype = DNA length = 2313
FEATURE Location/Qualifiers
misc_feature 1..2313

-continued

```
                     note = Variant of Penicillin G Acylase From Kluyvera
                      citrophila
source               1..2313
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 63
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt cttggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcggc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aatgccgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcacttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acggcaaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacaggg tgatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg ccagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctggagaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 64        moltype = AA  length = 770
FEATURE              Location/Qualifiers
REGION               1..770
                     note = Variant of Penicillin G Acylase From Kluyvera
                      citrophila
source               1..770
                     mol_type = protein
                     organism = synthetic construct
SEQUENCE: 64
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGG ADRATEIDTI LDKQCRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TANNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQGDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLEN TQTAHHHHHH             770

SEQ ID NO: 65        moltype = DNA  length = 2313
FEATURE              Location/Qualifiers
misc_feature         1..2313
                     note = Variant of Penicillin G Acylase From Kluyvera
                      citrophila
source               1..2313
                     mol_type = other DNA
```

-continued

```
                     organism = synthetic construct
SEQUENCE: 65
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt tttggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccggaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcggc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aatgccgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcacttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgtctaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaatttt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg caagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctggacaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 66          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 66
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG FGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPESGYIANW  240
NNSPQKDYPA SDHQGFLWGG ADRATEIDTI LDKQCRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVNF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLDN TQTAHHHHHH             770

SEQ ID NO: 67          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 67
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
```

-continued

```
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt cttggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat taaaaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccggaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcggc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggttggatt tacgcgatgt cgggctatga aaccacccag 1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcacca gctggaaaac ccctgccatg catcttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acgtctaaca acatgattgt cttctcaccg acgtcgggta accgcccggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt ggatcagggt 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt 1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaact gcagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tatggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctggagaac 2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 68          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 68
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIKTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPESGYIANW  240
NNSPQKDYPA SDHQGFLWGG ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM HLTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNCSPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLEN TQTAHHHHHH             770

SEQ ID NO: 69          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 69
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt gcgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
```

-continued

```
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag 1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acgtccaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt 1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg aaagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcaccatca ccatcaccat taa                              2313

SEQ ID NO: 70          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 70
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG AGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 71          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 71
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taatccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
```

-continued

```
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag 1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt 1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag 1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg 1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag 1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac 1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc 1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt 1440
acgtgcaaca acatgattgt cttctcaccg acgtcgggta accgcaaggt tcttgcctgg 1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac 1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct 1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt 1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat 1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg 1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgtttcattt 1860
gataaagata ttcgccaaaa ctactggccg gattctattc gcgcgcagat agcttccctc 1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat 1980
aaagtgaacg caagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa 2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt 2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa 2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc 2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac 2280
acgcaaacgg cgcaccatca ccatcaccat taa                              2313

SEQ ID NO: 72          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 72
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TCNNMIVFSP TSGNRKVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVSF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 73          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 73
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
```

-continued

```
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgtccaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg aaagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctattgcc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 74          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 74
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYCLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 75          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 75
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
```

-continued

```
gccgcggttc cagcgccgtt tggttggatt tacgcgatgt cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgtccaaca acatgattgt cttctcaccg acgtcgggta accgcaaggt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccaaaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg caagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctatccgc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 76          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 76
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAMSGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRKVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNASPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYPLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 77          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 77
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aatgccgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgctgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
```

-continued

```
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgagcaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatggggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg aaagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctattgcc tgaagtttga tctggataac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 78          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 78
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQCRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEWG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYCLKFDLDN TQTAHHHHHH             770

SEQ ID NO: 79          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 79
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttcggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgacctcga tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg catcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgtccaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
```

-continued

```
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtaaactt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt tgttaagttt  1860
gataaggata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg aaagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaacagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc ggctgaaagc gcctattgcc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 80          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 80
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG FGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM HLTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVNLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSNSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAAAES AYCLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 81          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 81
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgctgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgagcaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacaggg tgatcagggt  1680
gaggttaaga tcgttcgcga tgaatggggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
```

-continued

```
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaaacttt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg aaagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctattgcc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 82          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 82
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQGDQG EVKIVRDEWG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKTFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYCLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 83          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 83
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgctgcctgg ctgaccagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgtccaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg aaagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
```

-continued

```
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctattgcc tgaagtttga tctggataac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 84          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 84
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYCLKFDLDN TQTAHHHHHH             770

SEQ ID NO: 85          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 85
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgagcagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacattagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgtccaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg aaagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcagtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctattgcc tgaagtttga tctgcaaaac  2280
```

-continued

```
acgcaaacgg cgcaccatca ccatcaccat taa                                2313

SEQ ID NO: 86          moltype = AA  length = 770
FEATURE                Location/Qualifiers
REGION                 1..770
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..770
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 86
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LSSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYCLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 87          moltype = DNA  length = 2313
FEATURE                Location/Qualifiers
misc_feature           1..2313
                       note = Variant of Penicillin G Acylase From Kluyvera
                        citrophila
source                 1..2313
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 87
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt ttgggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac agccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgagcat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcaat cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgagcagca tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg gcccttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgtccaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacagtt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt cgttaaattt  1860
gataaagata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg aaagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcactagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc gcgggaaagc gcctattgcc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                                2313

SEQ ID NO: 88          moltype = AA  length = 770
FEATURE                Location/Qualifiers
```

-continued

```
REGION                  1..770
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..770
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 88
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG LGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD SQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALSINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LSSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM ALTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAARES AYCLKFDLQN TQTAHHHHHH             770

SEQ ID NO: 89           moltype = DNA  length = 2313
FEATURE                 Location/Qualifiers
misc_feature            1..2313
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..2313
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 89
agcaatatgt gggtgattgg caaaaacaaa gcccaggatg cgaaggccat tatggtcaat  60
gggccgcagt ttggttggta taacccggcg tatacctacg gtatcggcct gcacggcgcg  120
ggctatgacg tcaccggcaa tacgccgttt gcctatccgg gcctcctttt tggtcacaac  180
ggcaccattt catggggatc caccgccggt tgcggtgata gcgtcgatat ctttgccgaa  240
aaactttccg ccgagaagcc gggctattac cagcataacg gcgagtgggt gaagatgttg  300
agccgcgtag agactattgc ggtcaaagac ggccagccgg agacctttac cgtttggcgc  360
acgctgcacg gcaacgtcat tcacaccgat actgcgacgc agaccgccta tgccaaagcg  420
cgggcctggg atggcaaaga ggtggcgtcc ctgctggcgt ggacgcacca gatgaaggcc  480
aaaaactggc cggagtggac gcagcaggcg gccaaacagg cgctgaccat caactggtac  540
tacgccgatg tgaacggcaa tatcggctat gtgcataccg gcgcctatcc ggatcgccag  600
cccggccacg acccgcgttt gccggttccc ggcactggaa aatgggactg gaaagggttg  660
ctgtcgtttg atttgaatcc gaaagtgtat aacccgcagt cgggctatat cgccaactgg  720
aacaactcgc cgcaaaaaga ctacccggcc tctgatcacc aaggtttcct gtggggcagc  780
gcggatcgag cgactgagat cgacacgatc ctcgataagc aaccgcgctt caccgccgat  840
caggcgtggg atgtgatccg ccaaaccagc cgtcgggatc tcaacctgcg gttgttctta  900
ccggcgctga aggacgccac cgccaacctg gcggaaaacg atccgcgccg ccaactggtg  960
gatacactgg cgagctggga cggcgaaaac cttgtcaacg atgacggaaa aacctatcag  1020
caaccgggat cggcgattct tgcagcctgg ctgacctcga tgctcaagcg cacggtggtt  1080
gccgcggttc cagcgccgtt tggttggatt tacgcggcat cgggctatga aaccacccag  1140
gacgggccac tcggctcgct gaacatcagc gtgggggcga aaatcctcta cgaagctctg  1200
cagggtgata agtcgccaat cccgcaggcg gtcgatctgt ttggcgggaa accgcagcag  1260
gaagtaatac tggcggcgct ggacgacgct tggcagacgc tgtcaaaacg ctacggtaac  1320
gacgtcacca gctggaaaac ccctgccatg catcttacct tccgggccaa taacttcttc  1380
ggcgtgccgc aggcggcagc aaaagaggcg cgtcatcagg cggagtacca gaaccgcggt  1440
acgtccaaca acatgattgt cttctcaccg acgtcgggta accgccccgt tcttgcctgg  1500
gatgtggtgg cgccggggca aagcggtttt atcgcgcagg atggcaaagc cgataagcac  1560
tatgacgatc agctgaaaat gtacgagagc tttggccgta aatcgctgtg gttaacgcct  1620
caggacgttg acgagcacaa agagtctcag gaagtgctgc aggtacaatt agatcagggt  1680
gaggttaaga tcgttcgcga tgaatacggc atgccgcata tttacgccga tgatacctat  1740
cgactgtttt acggctatgg ctacgtggtg gcgcaggatc gcctgttcca gatggaaatg  1800
gcgcgccgca gtactcaggg gaccgtctcc gaggtgctgg gcaaagcatt tgttaagttt  1860
gataaggata ttcgccagaa ctactggccg gattctattc gcgcgcagat agcttccctc  1920
tccgctgagg ataaatccat tctgcagggc tatgccgatg gcatgaatgc gtggatcgat  1980
aaagtgaacg agagccccga taagctgtta ccccagcagt tctccacctt tggttttaaa  2040
cccaagcatt gggaaccgtt tgatgtggcg atgatttttg tcggcaccat ggcgaaccgt  2100
ttttctgaca gcaccagcga aattgataac ctggcgctgc tgacggcgct aaaagacaaa  2160
tacggcgcgc agcagggcat ggcggtcttt aaccagctga aatggctggt taatccttcc  2220
gcgccaacca ccattgcggc ggctgaaagc gcctattgcc tgaagtttga tctgcaaaac  2280
acgcaaacgg cgcaccatca ccatcaccat taa                               2313

SEQ ID NO: 90           moltype = AA  length = 770
FEATURE                 Location/Qualifiers
REGION                  1..770
                        note = Variant of Penicillin G Acylase From Kluyvera
                         citrophila
source                  1..770
```

-continued

```
                     mol_type = protein
                     organism = synthetic construct
SEQUENCE: 90
SNMWVIGKNK AQDAKAIMVN GPQFGWYNPA YTYGIGLHGA GYDVTGNTPF AYPGLLFGHN  60
GTISWGSTAG CGDSVDIFAE KLSAEKPGYY QHNGEWVKML SRVETIAVKD GQPETFTVWR  120
TLHGNVIHTD TATQTAYAKA RAWDGKEVAS LLAWTHQMKA KNWPEWTQQA AKQALTINWY  180
YADVNGNIGY VHTGAYPDRQ PGHDPRLPVP GTGKWDWKGL LSFDLNPKVY NPQSGYIANW  240
NNSPQKDYPA SDHQGFLWGS ADRATEIDTI LDKQPRFTAD QAWDVIRQTS RRDLNLRLFL  300
PALKDATANL AENDPRRQLV DTLASWDGEN LVNDDGKTYQ QPGSAILAAW LTSMLKRTVV  360
AAVPAPFGWI YAASGYETTQ DGPLGSLNIS VGAKILYEAL QGDKSPIPQA VDLFGGKPQQ  420
EVILAALDDA WQTLSKRYGN DVTSWKTPAM HLTFRANNFF GVPQAAAKEA RHQAEYQNRG  480
TSNNMIVFSP TSGNRPVLAW DVVAPGQSGF IAQDGKADKH YDDQLKMYES FGRKSLWLTP  540
QDVDEHKESQ EVLQVQLDQG EVKIVRDEYG MPHIYADDTY RLFYGYGYVV AQDRLFQMEM  600
ARRSTQGTVS EVLGKAFVKF DKDIRQNYWP DSIRAQIASL SAEDKSILQG YADGMNAWID  660
KVNESPDKLL PQQFSTFGFK PKHWEPFDVA MIFVGTMANR FSDSTSEIDN LALLTALKDK  720
YGAQQGMAVF NQLKWLVNPS APTTIAAAES AYCLKFDLQN TQTAHHHHHH              770
```

We claim:

1. An engineered polynucleotide encoding an engineered penicillin G acylase variant having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:6, and at least one substitution at a position selected from positions 54, 62, 115, 125, 127, 127, 185, 253, 254, 254/255, 254/255/370, 255, 256, 257, 257, 260, 268, 322, 325, 348, 369, 370, 372, 373, 377, 378, 384, 384/513/536, 388, 389, 391, 435, 461, 517, 530, 554, 556, 557, 559, 560, 600/623, 623, 624, 626, 627, 705, 706, 707, 723, 740, 748, and 752, wherein said positions are numbered with reference to SEQ ID NO:6.

2. The engineered polynucleotide encoding an engineered penicillin G acylase variant of claim **1**, wherein said said engineered penicillin acylase variant has 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 6, and at least one substitution selected from 54C, 62G, 115A/P, 125L/T, 127S/V, 185V, 253K/V, 254T, 254W/255G, 254W/255G/370I, 255L, 255M/Q/T/Y, 256Q, 257I, 257V, 260A/P, 2685/V, 322P, 325G, 348C, 348Q, 369L, 369P, 369V, 369W, 370F/G/S, 372A/H/L, 373F/M, 377P, 378H, 384A, 384F/513Q/536M, 384G/L, 388T, 389L, 391P/S, 435R, 461A, 517L/P, 530C/Y, 554A/E/P/V, 556G, 557G/S, 559P/S, 560I, 600T/623V, 623A/G/R/W, 624A, 626G, 627G/H, 705G/P, 706G, 707S, 723A/G, 740L, 748G, and 752E.

3. An engineered polynucleotide encoding an engineered penicillin G acylase variant having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:8, and at least one substitution set selected from positions103/370/444/706/766, 103/369/370/442/444/536/556/766, 103/369/370/444, 103/369/370/444/556/706/766, 103/369/370/444/765/766, 103/369/370/765/766, 257/362/384/451, 257/362/384/451/723, 362/451/705, 369/370, 369/370/444/706/766, 369/370/556/766, 369/370/388/444/556/766, 369/370/444, 369/370/444/556/766, 369/370/556, 369/370/556/765, 369/370/556/766, 369/370/766, 369/370/444/556, 369/370/444/556/612/766, 369/370/444/556/706/765, 369/370/444/706/765/766, 372/373/384/513/560, 372/384/451/705, 372/384/560/705, 384/451/560/705/723, 384/451/705/723, 451/560/705/723, and 451/705/723, wherein said positions are numbered with reference to SEQ ID NO:8.

4. The engineered polynucleotide encoding an engineered penicillin G acylase variant of claim **3**, wherein said substitution set is selected from 103V/370F/444S/706G/766G, 103V/369W/370F/442I/444S/536M/556G/766G, 103V/369W/370F/444S, 103V/369W/370F/444S/556G/706G/766G, 103V/369W/370F/444S/765P/766G, 103V/369W/370F/765P/766G, 257V/362V/384A/451R, 257V/362V/384L/451R/723L, 362V/451R/705D, 369P/370F, 369P/370F/444S/706G/766G, 369P/370F/556G/766G, 369V/370F/388T/444S/556G/766G, 369V/370F/444S, 369V/370F/444S/556G/766G, 369V/370F/556G, 369V/370F/556G/765P, 369V/370F/556G/766G, 369V/370F/766G, 369W/370F/444S/556G, 369W/370F/444S/556G/612A/766G, 369W/370F/444S/556G/706G/765P, 369W/370F/444S/706G/765P/766G, 372A/373M/384L/513Q/560G, 372A/384L/451R/705D, 372A/384L/560G/705D, 384A/451R/560G/705D/723L, 384L/451R/705D/723L, 451R/560G/705D/723L, and 451R/705D/723L.

5. An engineered polynucleotide encoding an engineered penicillin G acylase variant having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:34, and at least one substitution set selected from 103/372/373/557, 253/322/369/623, 253/254/322/369/623, 253/254/369/391/623/723, 253/254/369/619/623/723, 253/254/369/623/723, 253/254/373/623/723, 253/254/255/369/623/723, 253/254/369, 253/322/369/373/723, 253/369/623/723, 253/373/623, 253/254/255/322/369/619/723, 260/372/373/556, 260/372/373/556/557/559, 322/369, 322/369/373/723, 322/369/623/723, and 369/373/556, wherein said positions are numbered with reference to SEQ ID NO:34.

6. The engineered polynucleotide encoding an engineered penicillin G acylase variant of claim **5**, wherein said substitution set is selected from 103V/372S/373F/557G, 253H/322T/369W/623G, 253H/254Q/322T/369W/623G, 253H/254Q/369W/391A/623G/723A, 253H/254Q/369W/619R/623G/723A, 253H/254Q/369W/623G/723A, 253H/254Q/373L/623G/723A, 253H/254S/255V/369W/623S/723A, 253H/254S/369W, 253H/322T/369W/373W/723A, 253H/369W/623G/723A, 253H/373L/623S, 253S/254S/255V/322T/369W/619R/723A, 260S/372S/373F/556G, 260S/372S/373F/556G/557V/559S, 322T/369W, 322T/369W/373W/723A, 322T/369W/623G/723A, and 369W/373F/556G.

7. The engineered polynucleotide encoding an engineered penicillin G acylase variant of claim **1**, wherein said engineered penicillin G acylase variant comprises a histidine tag.

8. The engineered polynucleotide encoding an engineered penicillin G acylase variant of claim **7**, wherein said histidine tag is present at the C-terminus of said engineered penicillin G acylase variant.

9. The engineered polynucleotide encoding an engineered penicillin G acylase variant of claim **1**, wherein said engineered penicillin G acylase variant comprises a polypeptide sequence set forth in variant numbers 1-308.

10. The engineered polynucleotide encoding an engineered penicillin G acylase variant of claim **1**, wherein said engineered penicillin G acylase variant comprises a polypeptide sequence selected from the even-numbered sequences between SEQ ID NO:4 and SEQ ID NO:90.

11. The engineered polynucleotide sequence of claim **1**, said wherein said sequence comprises a polynucleotide sequence that is at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to a sequence selected from the odd-numbered sequences between SEQ ID NO:5 and SEQ ID NO:89.

12. A vector comprising the engineered polynucleotide sequence of claim **1**.

13. The vector of claim **12**, further comprising at least one control sequence.

14. A host cell comprising the vector of claim **13**.

15. A method for producing an engineered penicillin G acylase variant, comprising culturing said host cell of claim **14** under conditions that said engineered penicillin G acylase variant is produced by said host cell.

16. The method of claim **15**, further comprising the step of recovering said engineered penicillin G acylase variant produced by said host cell.

* * * * *