



(51) International Patent Classification:

C12P 21/02 (2006.01) A61K 38/00 (2006.01)
C12P 21/00 (2006.01)

(21) International Application Number:

PCT/US2018/040692

(22) International Filing Date:

03 July 2018 (03.07.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/529,320 06 July 2017 (06.07.2017) US
62/535,090 20 July 2017 (20.07.2017) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **LIU, David, R.**; 3 Whitman Circle, Lexington, MA 02420 (US). **BRYSON, David, Irby Jr.**; 127 King Street, Apartment 1, Dorchester, MA 02122 (US).

(74) Agent: **MACDONALD, Kevin**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

(54) Title: EVOLUTION OF tRNA SYNTHETASES

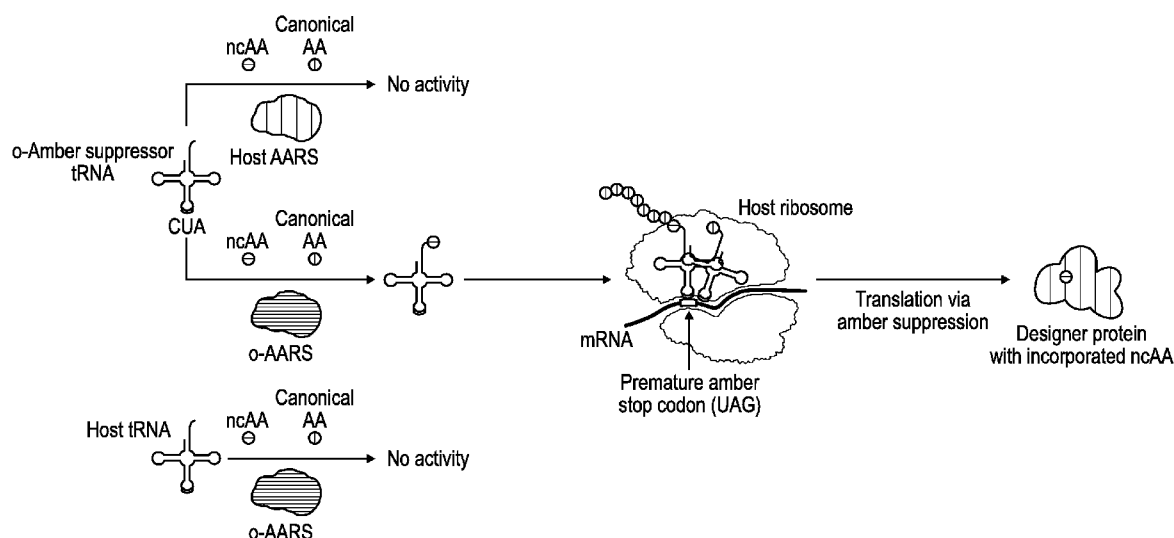


FIG. 1

(57) Abstract: The disclosure provides amino acid sequence variants of orthogonal aminoacyl-tRNA synthetases (AARSs) having increased activity and selectivity compared to previous AARSs, and methods of producing the same.



EVOLUTION OF tRNA SYNTHETASES

RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. provisional patent applications, U.S.S.N. 62/535,090, filed July 20, 2017, entitled “EVOLUTION OF tRNA SYNTHETASES”, and U.S.S.N. 62/529,320, filed July 6, 2017, entitled “EVOLUTION OF tRNA SYNTHETASES”, which are incorporated herein by reference.

FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant numbers N66001-12-C-4207, awarded by the Defense Advanced Research Projects Agency; EB022376, GM118062, AI119813, GM022854 and GM106621, awarded by the National Institutes of Health; and FG02-98ER2031, awarded by the Department of Energy. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0003] The directed evolution of orthogonal aminoacyl-tRNA synthetases (AARSs) enables genetic code expansion through the site-specific installation of non-canonical amino acids into proteins. Traditional laboratory evolution techniques typically produce AARSs with greatly reduced activity (often ~1,000-fold lower) and poor amino acid selectivity compared to their wild-type counterparts, limiting their utility.

[0004] Although researchers have evolved many AARSs to incorporate non-canonical amino acids (ncAAs) into proteins, several outstanding challenges limit their utility and generality. Laboratory evolution of AARSs with altered amino acid specificity typically relies on three to five rounds of sequential positive and negative selections from an AARS library containing either partially or fully randomized residues in the amino acid-binding pocket. The limited number of rounds of selection typically conducted in AARS evolution campaigns reflects the effort required to complete each round of evolution, which is on the order of one week or longer. A consequence of conducting relatively few rounds of selection on libraries that focus mutagenesis on and around the amino acid-binding pocket is that laboratory-evolved AARSs routinely emerge with suboptimal properties, including ~1,000-fold reduced activity (k_{cat}/K_M) compared to their wild-type counterparts, and modest selectivity for the target ncAA

over endogenous amino acids that can require compensation with high concentrations of ncAA and expression in minimal media, lowering protein yields. The modest enzymatic efficiency and selectivity of many laboratory-evolved AARSs are longstanding challenges that limit the production and purity of expressed proteins containing ncAAs.

SUMMARY OF THE INVENTION

[0005] In some aspects, the disclosure relates to evolved AARSs that increase the utility of orthogonal translation systems and establish the capability of rapidly and continuously evolving orthogonal AARSs with high activity and amino acid specificity. The disclosure is based, in part, on the discovery that positive and negative phage-assisted continuous evolution (PACE) selections produce highly active and selective orthogonal AARSs through hundreds of generations on rapid time scales. For example, as described in the Examples section, continuous evolution of a pyrrolysyl-tRNA synthetase (PylRS), in some embodiments, improved enzymatic efficiency ($k_{\text{cat}}/K_M^{\text{tRNA}}$) up to 45-fold compared to the wild-type enzyme.

[0006] In some aspects, the disclosure relates to the discovery that PACE unexpectedly generated highly active, split-PylRS variants produced as two mutually dependent polypeptide fragments, recapitulating natural PylRS homologs. It was observed that simultaneous positive and negative selection PACE over 48 h greatly improved the selectivity of a promiscuous tyrosyl-tRNA synthetase variant for site-specific incorporation of *p*-iodo-l-phenylalanine, rejecting *p*-nitro-l-phenylalanine.

[0007] Accordingly, in some aspects, the disclosure provides pyrrolysyl-tRNA synthetase (PylRS) protein variants. In some embodiments, a PylRS protein variant described herein comprises a nucleic acid sequence or an amino acid sequence that is at least 90% identical to a *Methanosarcina* PylRS or a fragment thereof (*e.g.*, the N-terminal domain of a *Methanosarcina* PylRS or the C-terminal domain of a *Methanosarcina* PylRS), for example *M. bakeri* PylRS (*e.g.*, SEQ ID NO: 6) or *M. mazei* PylRS (*e.g.*, SEQ ID NO: 7).

[0008] In some aspects, the disclosure provides tyrosyl-tRNA synthetase (TyrRS) proteins variant capable of incorporating a *p*-iodo-l-phenylalanine into a protein. In some embodiments, a TyrRS protein variant described herein comprises a nucleic acid sequence or an amino acid sequence that is at least 90% identical to a *Methanocaldococcus jannaschii* (*M. jannaschii*) TyrRS (*MjTyrRS*), for example SEQ ID NO: 24.

[0009] In some embodiments, a PylRS comprises an amino acid sequence that is at least 90% identical to the amino acid sequence of a wild-type *M. barkeri* or *M. mazei* PylRS (*e.g.*, the amino acid sequence set forth in SEQ ID NO: 20 or 21), or a fragment thereof (*e.g.*, amino acids 1-149 of SEQ ID NO: 20 or amino acids 185-454 of SEQ ID NO: 21).

[0010] In some aspects, the disclosure provides a pyrrolysyl-tRNA synthetase (PylRS) protein variant having an N-terminal domain amino acid substitution present in at least one of the following positions: V31, T56, H62, or A100 (*e.g.*, relative to SEQ ID NO: 20 or 21). In some embodiments, the amino acid substitution is V31I, T56P, H62Y, or A100E.

[0011] In some aspects, the disclosure provides a chimeric pyrrolysyl-tRNA synthetase (PylRS) protein variant comprising: a first portion comprising amino acid residues 1-149 of *Methanosarcina barkeri* PylRS (SEQ ID NO: 20); and a second portion comprising amino acid residues 185-454 of *Methanosarcina mazei* PylRS (SEQ ID NO: 21), wherein the first portion or the second portion comprises at least one of the amino acid substitutions set forth in Tables 2-6.

[0012] In some embodiments, the chimeric protein variant comprises an amino acid substitution in at least one of the following positions: V31, T56, H62, or A100. In some embodiments, the amino acid substitution is V31I, T56P, H62Y, A100E, or any combination thereof.

[0013] In some aspects, the disclosure provides a tyrosyl-tRNA synthetase (TyrRS) protein variant having an amino acid substitution present in at least one of the following positions: L69 or V235. In some embodiments, the amino acid substitution is L69F, V235I, or L69F and V235I.

[0014] In some aspects, the disclosure provides an isolated nucleic acid comprising a sequence represented by any one of SEQ ID NO: 5-19. In some aspects, the disclosure provides a protein encoded by an isolated nucleic acid comprising a sequence represented by any one of SEQ ID NO: 5-19.

[0015] In some aspects, the disclosure relates to a host cell comprising a tRNA synthetase protein variant as described by the disclosure. In some aspects, the disclosure relates to an isolated nucleic acid as described by the disclosure. In some embodiments, the tRNA synthetase protein variant is orthogonal to the host cell (*e.g.*, not expressed naturally in the host cell). In some embodiments, a host cell is a bacterial cell. In some embodiments, a bacterial cell is an *E. coli* cell.

[00016] In some aspects, the disclosure relates to a selection system comprising: a first container housing a selection phagemid as described by the disclosure; a second container housing a positive selection system as described by the disclosure; and, optionally, a third container housing negative selection system as described by the disclosure.

[00017] In some embodiments, a selection system further comprises a container housing one or more bacterial cells. In some embodiments, the bacterial cells are *E. coli* cells.

[00018] In some aspects, the disclosure relates to methods of using tRNA synthetase protein variants described by the disclosure. In some embodiments, the disclosure relates to methods for aminoacylation of a tRNA, the methods comprising, contacting a tRNA encoding an amber codon with a tRNA synthetase protein variant as described by the disclosure in the presence of a non-canonical amino acid.

[00019] In some embodiments, a non-canonical amino acid is a pyrolysine or a p-iodo-L-phenylalanine. In some embodiments, the tRNA is contacted with the tRNA synthetase inside a cell.

[00020] In some aspects, the disclosure relates to methods for incorporating a non-canonical amino acid into a peptide, the method comprising expressing in a cell: (i) an mRNA transcript, wherein the transcript comprises an amber codon at a position in which a non-canonical amino acid (ncAA) is to be translated; (ii) a tRNA capable of incorporating the ncAA; (iii) a tRNA synthetase protein variant as described herein, wherein (i), (ii), and (iii) are expressed in the presence of the ncAA.

[00021] In some embodiments, the non-canonical amino acid is a pyrolysine or a p-iodo-L-phenylalanine. In some embodiments, the tRNA synthetase is orthogonal to the cell. In some embodiments, the cell is an *E. coli* cell.

[00022] The summary above is meant to illustrate, in a non-limiting manner, some of the embodiments, advantages, features, and uses of the technology disclosed herein. Other embodiments, advantages, features, and uses of the technology disclosed herein will be apparent from the Detailed Description, the Drawings, the Examples, and the Claims

BRIEF DESCRIPTION OF DRAWINGS

[00023] FIG. 1 shows an overview of orthogonal translation in biological systems. The orthogonal amber suppressor tRNA is not recognized by any of the cell's endogenous AARS enzymes, but is selectively aminoacylated by the orthogonal AARS with the desired ncAA. The

charged amber suppressor tRNA decodes 'UAG' stop codons during translation of the protein of interest, enabling site-specific incorporation of the ncAA into proteins made by the cell.

[00024] FIGs. 2A-2C show an overview of PACE positive selections for the continuous evolution of AARS activity. FIG. 2A shows strategies for linking AARS activity to the expression of gene III, which encodes the pIII protein required for phage to be infectious. In strategy 1, AARS-catalyzed aminoacylation of an amber suppressor tRNA enables translation of full-length T7 RNAP from a transcript containing a premature amber stop codon. T7 RNAP subsequently drives expression of gene III from the T7 promoter (P_{T7}). In strategy 2, amber suppressor tRNA aminoacylation permits full-length translation of pIII from gene III mRNA containing a premature stop codon. FIG. 2B shows host-cell plasmids used to implement both selection strategies. The accessory plasmid (AP) encodes gene III and the amber suppressor tRNA. The complementary plasmid (CP) encodes T7 RNAP controlled by the phage-shock promoter (P_{psp}), which is induced only upon phage infection. The mutagenesis plasmid (MP) increases the rate of evolution during PACE through arabinose-induced production of mutagenic proteins. The selection phage (SP) encodes all phage genes except gene III, which is replaced by the evolving AARS gene. FIG. 2C shows a diagram of PACE with selection strategy 1 plasmids shown. SPs capable of catalyzing aminoacylation of the amber suppressor tRNA result in production of pIII protein from gene III of the AP in host *E. coli*. Under continuous dilution in the fixed-volume vessel (the "lagoon"), phage that are capable of triggering the production of pIII propagate faster than the rate of dilution, resulting in the continuous enrichment of SPs encoding active AARS variants.

[00025] FIG. 3 shows Non-canonical amino acids in this study. (1) p-nitro-L-phenylalanine, (2) N ϵ -(tert-butoxycarbonyl)-L-lysine, (3) L-pyrrolysine, (4) N ϵ -acetyl-L-lysine, (5) 3-iodo-L-phenylalanine, and (6) p-iodo-L-phenylalanine.

[00026] FIGs. 4A-4C show optimization of the T7 RNAP-mediated PACE positive selection for aminoacylation. Two amber stop codons in T7 RNAP are required to make reporter expression completely dependent on orthogonal translation of full-length T7 RNAP. FIG. 4A shows a luciferase reporter assay for optimizing the position and number of TAG stop codons in T7 RNAP. The orthogonal AARS (inducible with IPTG) charges the ncAA onto the amber suppressor tRNA, enabling translation of full-length T7 RNAP (inducible with anhydrotetracycline, ATc). Production of T7 RNAP results in subsequent expression of the luciferase reporter gene, *luxAB*. FIG. 4B shows using *p*-NFRS to site-specifically incorporate *p*-

NF at two positions (Ser12TAG + Ser203TAG) in T7 RNAP provided optimal reporter signal that was dependent on orthogonal AARS activity (+IPTG, +ncAA) and on expression of T7 RNAP (+ATc). FIG. 4C shows using chPylRS, reporter signal resulting from site-specific incorporation of BocK into T7 RNAP(Ser12TAG + Ser203TAG) suggests broad ncAA-tolerance at both sites of ncAA installation. Each value and error bar in b and c reflects the mean and s.d. of at least three independent biological replicates.

[00027] FIGs. 5A-5D show non-continuous propagation of SP in positive selections designed for PACE. To confirm activity dependence of phage propagation for each of the two positive selections (suppression of stop codons in T7 RNAP or in gene III), phage titers resulting from 16 h of propagation in batch culture were compared for SP expressing *p*-NFRS (FIG. 5A, FIG. 5C) or chPylRS (FIG. 5B, FIG. 5D). In each experiment, equal amounts of SP encoding the AARS of interest were used to infect cultures of S1030 host cells harboring the required PACE AP and CP plasmids in the presence or absence of the ncAA. Controls representing the starting titers for each set of experiments were prepared by diluting the same amount of SP into media lacking cells. Results indicate that selection stringency increases as the number of stop codons is increased in T7 RNAP (FIGs. 5A-5B) or gIII (FIGs. 5C-5D).

[00028] FIGs. 6A-6B show positive selections for aminoacylation support activity-dependent, continuous propagation in PACE. FIG. 6A shows that, in lagoon 1 (L1) supplemented with 1 mM *p*-NF, SP-*p*-NFRS propagates for 48 h of PACE using the selection based on amber suppression of two stop codons in T7 RNAP. SP-Kan, which lacks AARS activity, however, rapidly washed out of lagoon 2 (L2) by the first time point (16 h) under identical conditions. FIG. 6B shows phage were propagated for 30 h of PACE in the presence of 1 mM *p*-NF starting from a 1:1 mixture of SP-*p*-NFRS and SP-MBP-TEV using the selection based on amber suppression of a single stop codon in gene III. Activity-dependent phage titers and PCR analysis of phage taken from each time point sampled during PACE confirmed that SP-*p*-NFRS propagated exclusively while SP-MBP-TEV rapidly washed out.

[00029] FIGs. 7A-7B show evolution of AARS activity during mock PACE. FIG. 7A shows *p*-NFRS was challenged to aminoacylate the amber suppressor tRNA in the absence of its cognate ncAA substrate, *p*-NF, over 48 h of positive selection PACE conducted in two separate lagoons (L1 and L2). Enhanced mutagenesis from the MP was supplied in L2 only. Phage titers of L2 (green) rapidly increased after 16 h, while titers in L1 (magenta) were relatively stable throughout the evolution. FIG. 7B shows mutations in PACE-evolved clones and the relative

amino acid substrate specificities of clones from L2. Relative aminoacylation activity was compared in the PACE host strain, S1030, by measuring luminescence signal resulting from amber suppression of a premature stop codon at position 361 of a luciferase gene (*luxAB*). More coding mutations were obtained in phage isolates from L2, in which the MP provided enhanced mutagenesis, and every characterized L2 mutant emerged from PACE with increased activity on endogenous amino acids (no ncAA) compared to the progenitor enzyme, *p*-NFRS. Each value and error bar in b reflects the mean and s.d. of at least three independent biological replicates.

[00030] FIGs. 8A-8E show continuous evolution and characterization of chimeric pyrrolysyl-tRNA synthetase (chPylRS) variants with enhanced aminoacylation activity. FIG. 8A shows PACE was performed in three segments designed to gradually increase selection stringency. The first two segments (Pyl-1 and Pyl-2) used the selection requiring amber suppression of two stop codons in T7 RNAP, and the final segment (Pyl-3) used the selection requiring direct amber suppression of stop codons in gene III. The number of stop codons in the gene required for each selection and the concentration of BocK substrate are shown above the phage titer graph. Dotted lines (black) indicate transfer of evolved phage from the end of each PACE segment into the subsequent segment. Triangles indicate convergence toward the specified mutations. FIGs. 8B-8C show the relative expression of luciferase containing BocK at position 361 resulting from aminoacylation by progenitor enzyme, chPylRS, compared to evolved variants from the end of PACE segment Pyl-1 (FIG. 8B) or compared to variants containing only the consensus mutations from the end of each PACE segment (FIG. 8C). Labels correspond to PACE segments in FIG. 8A. FIG. 8D shows the relative efficiency of multisite, BocK incorporation into sfGFP resulting from aminoacylation by chPylRS variants with or without beneficial mutations discovered in PACE (V31I, T56P, H62Y, and A100E; IPYE). FIG. 8E shows the relative efficiency of AcK incorporation at position 2 of sfGFP resulting from aminoacylation by AcK3RS variants with or without transplanted mutations from PACE. Each value and error bar in b-e reflects the mean and s.d. of at least three independent biological replicates.

[00031] FIGs. 9A-9F show mutations emerging from PACE enhance the activity of PylRS variants on their target ncAA. FIG. 9A shows contributions toward improved activity from consensus mutations in chPylRS generated during PACE segments Pyl-1 and Pyl-2. FIGs. 9C-9D show transplantation of the activity-enhancing PACE mutations V31I, T56P, H62Y, and A100E (IPYE) into *M. barkeri* (Mb) or *M. Mazei* (Mm) PylRS greatly improved the expression

levels of luciferase containing the ncAA BocK at position 361 (FIG. 9B) and the expression levels of sfGFP containing a BocK at position 2 (FIG. 9C) or position 151 (FIG. 9D). FIGs. 9E-9F show transplantation of the 'IPYE' mutations into multiple variants of AcK3RS (FIG. 9E) or into the chimeric IFRS (FIG. 9F) improved expression of luciferase containing the ncAA residue at position 361. Each value and error bar in b-e reflects the mean and s.d. of at least three independent biological replicates.

[00032] FIGs. 10A-10D show ESI-MS analysis of purified sfGFP containing up to three BocK residues produced by chPylRS(IPYE). Analysis of purified wild type sfGFP (FIG. 10A) or sfGFP containing one (FIG. 10B), two (FIG. 10C) or three (FIG. 10D) BocK residues produced by chPylRS(IPYE) in the presence of 1 mM ncAA. BocK substitutions in sfGFP were made in response to premature amber stop codons at positions 39 (1xTAG), 39 and 151 (2xTAG), or 39, 135, and 151 (3xTAG). Protein was expressed in TOP10 cells in LB media. The major peak in each of the spectra was in agreement with the calculated mass of BocK incorporation. In each of the spectra containing BocK, a minor peak corresponding to an unclipped N-terminal methionine was also observed (calculated mass + 131.19 Da).

[00033] FIG. 11 shows ESI-MS analysis of purified sfGFP containing an AcK residue at position 2 produced by chAcK3RS(IPYE) in the presence of 1 mM AcK. Protein was expressed in TOP10 cells in LB media, and the major peak at found at 27,812.58 Da was in agreement with the calculated value (27,812.3 Da).

[00034] FIGs. 12A-12C show characterization of split variants of chPylRS emerging from PACE. Evolved split variants of chPylRS require the 'IPYE' tetramutation to retain high activity. Aminoacylation is dependent on both the N- and C-terminal fragments of the chPylRS variants shown. FIGs. 12A-12B show the relative expression sfGFP containing three premature stop codons at positions 39, 135, and 151 (sfGFP(3xTAG)) was compared in the presence or absence of 1 mM BocK for the six, split proteins containing the 'IPYE' tetramutation (FIG. 12A) or with variants lacking the tetramutation (FIG 12B). FIG. 12C shows the relative expression of sfGFP(Asn39TAG) in the presence of the unsplit chPylRS(IPYE) was compared to expression in the presence of the N-terminal fragments of split2 (NTerm.S2), split3 (NTerm.S3), or split6 (NTerm.S6) or the C-terminal fragment (CTerm) that would result from reinitiation at Met-107. Each value and error bar reflects the mean and s.d. of four independent biological replicates.

[00035] FIG. 13 shows characterization of the S326I mutation emerging from lagoon 2 during the Pyl-3 segment of PACE. The relative activity of split1, split2, and split3 containing the additional mutation, S326I, were compared to variants lacking the mutation and to the full-length chPylRS(IPYE). Each variant was used to produce sfGFP(3xTAG) containing three premature stop codons at positions 39, 135, and 151. Each value and error bar represents the s.d. of four independent biological replicates.

[00036] FIGs. 14A-14B show Western blot analysis of full-length and split chPylRS variants from PACE. FIG. 14A shows the chPylRS variants were N-terminally tagged with c-Myc and C-terminally tagged with 6xHis to enable two-color detection of the expressed proteins in order to characterize translation of stop codon-containing mutants that arose during PACE. FIG. 14B shows Western blot analysis of the protein lysates expressed in BL21 star DE3 cells indicated that the full-length variants chPylRS and PylRS(IPYE) were expressed with the N- and C-termini intact, but the presence of an internal start site also promotes alternative expression of the truncated, C-terminal fragment. Each of the split variants (split2, split3, and split 6) are expressed as two, distinct N- and C-terminal fragments indicating termination of translation at the premature stop codon and reinitiation at an internal start site.

[00037] FIGs. 15A-15D show ESI-MS analysis of affinity-tagged Ni-NTA-purified chPylRS variants from PACE. The evolved synthetases, chPylRS(IPYE) (FIG. 15A), Spit2 (FIG. 15B), Split3 (FIG. 15C), and Split6 (FIG. 15D) were labeled with an N-terminal c-Myc-tag and a C-terminal 6xHis-tag and purified over Ni-NTA resin prior to ESI-MS analysis. In the split variants of chPylRS, the N-terminal fragment is lost upon affinity purification. Protein was expressed in BL21 star DE3 cells in LB media. The major peaks in each spectra were in agreement with the calculated mass of the full-length enzyme, chPylRS(IPYE) (FIG. 15A), or the C-terminal fragment resulting from reinitiation at position Met-107 (FIGs. 15A-15D).

[00038] FIG. 16 shows alignment of PylRS sequences from multiple organisms and from PACE variants. Activity enhancing mutations from PACE and premature stop codons (*) that emerged in each of the split variants are shown. Note that the activity-enhancing A100E mutation became A100S in Split1 and Split 2 due to the frameshift. Split3, Split4, and Split6 each lack the A100E mutations because they terminate earlier in the sequence. Arrows denote the PylSn and PylSc gene products of the *D. hafniense* strains.

[00039] FIGs. 17A-17E show evolution of AARS variants from dual positive- and negative-selection PACE with greatly improved amino acid specificity. FIG. 17A shows strategy

for linking undesired aminoacylation to gene III-neg expression, which encodes the pIII-neg protein. When undesired aminoacylation occurs in the negative selection, pIII-neg is produced, impeding progeny phage infectivity. In the absence of undesired aminoacylation, only pIII is produced, resulting in infectious phage progeny. Negative-selection stringency is modulated by ATc concentration. FIG. 17B shows host-cell plasmids used to implement the negative selection. FIG. 17C shows a diagram of dual-selection PACE using simultaneous positive and negative selections. Evolving phage are continuously cross-seeded between positive and negative selection lagoons at a 50-fold dilution. FIG. 17D shows the relative site-specific incorporation efficiency of either endogenous amino acids (no ncAA), p $\rightarrow$ -NF, or p-IF at position 39 of sfGFP resulting from aminoacylation by p-NFRS, p-IFRS, or evolved variants from PACE (Iodo.1, Iodo.5, Iodo.7, and Iodo.8). FIG. 17E shows the predicted position of mutations evolved during dual-selection PACE. The shown crystal structure is the p-NFRS protein sequence aligned to pdb:2AG636, which is the crystal structure of an AARS that has the identical protein sequence of p-IFRS and is bound to the ncAA substrate, p-bromo-L-phenylalanine. The shaded spheres in the crystal structure correspond to the mutations in the table to the left. Active-site residues within a 5 Å radius around the ncAA substrate are colored gray. Each value and error bar in d reflects the mean and s.d. of at least three independent biological replicates.

[00040] FIGs. 18A-18B show an overview of the PACE negative selection for AARS activity using the dominant-negative variant of pIII (pIII-neg). FIG. 18A shows a diagram of PACE negative selection plasmids. PACE host cells (S1030) are cotransformed with the negative-selection accessory plasmid (AP-) and a negative-selection complementary plasmid (CP-). When an SP infects the negative selection host, production of pIII protein from gene III is induced from the phage shock promoter (P_{psp}) of the AP-. If the AARS encoded by the SP can catalyze aminoacylation under the conditions of the negative selection (e.g., in the absence of ncAA), full-length T7 RNAP is produced from the AP- through amber suppression of amber stop codons at position 12 and 203 of the T7 RNAP gene. When full-length T7 RNAP is produced, expression of gene III-neg is induced from the T7 promoter (P_{T7}) of the CP- resulting in production of the dominant-negative pIII-neg protein. The infectivity of progeny phage decreases with the amount of pIII-neg in the host cell. Expression levels of the T7 RNAP gene on the AP- are also controlled by an ATc-inducible promoter (P_{tet}), allowing the negative selection to be turned on or off during PACE. FIG. 18B shows a diagram of inputs and outputs of the AND logic gate created by the PACE negative selection. The dominant-negative pIII-neg

protein is produced only in the presence of both aminoacylation activity and ATc. In the absence of either negative-selection input, progeny phage are infectious and carry forward the encoded AARS into the subsequent round of evolution in PACE.

[00041] FIGs. 19A-19C show validation of the PACE negative selection. FIG. 19A shows mock PACE experiments were performed in parallel to demonstrate that the negative selection is dependent on both aminoacylation activity and the concentration of ATc. In lagoon 1 (L1), SP-p-NFRS was propagated in the absence of substrate amino acid (–p-NF) to determine the maximum concentration of ATc that could be tolerated without decreasing the rate of phage propagation when aminoacylation does not occur. In lagoon 2 (L2), SP-p-NFRS and SP-MBP-TEV were both propagated in the presence of the p-NFRS substrate (+p-NF) to determine the minimum concentration of ATc that would support negative selection when aminoacylation does occur. FIG. 19B shows activity-dependent titers were measured to detect the relative amount of active SP-p-NFRS present in the lagoons at each sampled time point of PACE. In L1, the maximum concentration of ATc (broken gray line) that did not affect phage propagation was 30 ng/mL. In L2 (magenta line), the minimum concentration of ATc that induced negative selection against aminoacylation was 10 ng/mL. FIG. 19C shows PCR analysis of phage from each sampled time point of L2 confirms that the inactive SP-MBP-TEV was selectively enriched from a 1000:1 excess of SP-p-NFRS at time points that correspond to ATc concentrations between 10 and 30 ng/mL (16–40 h of PACE).

[00042] FIGs. 20A-20D show the previously evolved AARS, p-NFRS, accepts multiple amino acid substrates. ESI-MS analysis of purified wild type sfGFP (FIG. 20A) or sfGFP(Asn39TAG) expressed with p-NFRS in the presence of 1 mM p-NF (FIG. 20B), no ncAA (FIG. 20C), or 1 mM p-IF (FIG. 20D) demonstrates that p-NFRS accepts Phe, p-NF, and p-IF. Protein was expressed in BL21 star DE3 cells in LB media. FIG. 20B shows a peak corresponding to incorporation of p-NF into sfGFP was observed at 27,918.09 Da (calculated: 27,918.31 Da). FIGs. 20B-20C show peaks corresponding to incorporation of Phe were found at 27,873.01 Da and 27873.09 Da, respectively, (calculated: 27,873.32 Da) from expression in the presence or absence of 1 mM p-NF. FIG> 20C shows a peak corresponding to incorporation of p-IF into sfGFP was found at 27,999.04 Da (calculated: 27,999.22 Da). Minor peaks in each spectrum correspond to an unclipped N-terminal methionine (calculated mass +131.19 Da).

[00043] FIGs. 21A-21B show dual-selection PACE of the polyspecific MjTyrRS variant, p-NFRS, to evolve selective activity on p-IF. FIG. 21A shows a diagram of chemostats and

lagoons during dual-selection PACE. DRM media supplemented with 4 mM p-NF was pumped into the negative selection lagoon and DRM media supplemented with 1 mM p-IF was pumped into the positive selection lagoon. Host cell cultures from each chemostat were pumped through the corresponding lagoons that were supplemented with required inducers (ATc and arabinose). The opposing lagoons were coupled such that material was continuously exchanged ('cross-seeded') between each lagoon at a 50-fold slower flow rate (gray arrows) with respect to the flow rate from the chemostats through each lagoon (black arrows). FIG. 21B shows a plot of phage titers measured from samples taken at the indicated time points from each lagoon during PACE. Positive selection was conducted exclusively for the first 24 h of the experiment, and dual-selection began at the 24-h time point by cross-seeding phage between the opposing lagoons. The flow rate from the chemostats through the lagoons (broken gray line) was doubled after the two lagoons were coupled, and the flow rate of cross-seeded material was adjusted to maintain 50-fold dilution into the opposing selections.

[00044] FIGs. 22A-22B show non-continuous counterselections to isolate p-IF-selective evolved AARS variants after dual-selection PACE. FIG. 22A shows two counterselections were performed in parallel without enhanced mutagenesis (no MP) on the evolved pool of SP sampled from the negative-selection lagoon at the end of dual-selection PACE. Negative selections were performed in batch culture to non-continuously propagate phage lacking unwanted AARS activity on canonical amino acids and p-NF. The stringent negative selection (left) was performed in host cells containing an AP⁻:CP⁻ pair in which the ATc-inducible promoter driving expression of T7 RNAP(Ser12TAG, Ser203TAG) on AP⁻ (FIG. 18) was replaced with the strong, PproD constitutive-promoter1. A less stringent negative selection was performed (right) using an AP⁻:CP⁻ pair in which the weaker PproA constitutive-promoter1 was upstream of T7 RNAP(Ser12TAG, Ser203TAG). SPs that propagated overnight in the non-continuous negative selection were isolated and used to infect positive-selection host cells to conduct activity-dependent plaque assays in the presence of p-NF or p-IF. Plaques that formed in the presence of the desired amino acid, p-IF, were isolated and subjected to DNA sequencing. FIG. 22B shows data from parallel counterselections. The enrichment factor reports the number of activity-dependent plaques that formed in 1 mM p-IF divided by the number of plaques that formed in 1 mM p-NF. Mutants marked "RBS mutation" indicate that the ribosome-binding site (RBS) driving translation of the AARS was mutated; these clones were not further characterized.

[00045] FIGs. 23A-23B show the PACE-evolved Iodo.5 variant is highly selective for the desired substrate, p-IF. ESI-MS analysis of purified sfGFP from expression of sfGFP(Asn39TAG) with p-IFRS (FIG. 23A) or Iodo.5 (FIG. 23B) in LB media supplemented with both 1 mM p-NF and 1 mM p-IF demonstrates that each AARS enzyme selectively incorporates p-IF. (a, b) A peak corresponding to incorporation of p-IF into sfGFP was found at 27,999.52 Da and 27,999.45 Da, respectively (calculated: 27,999.22 Da). Incorporation of p-NF into sfGFP was calculated to have a mass of 27,918.31 Da (dashed line).

DEFINITIONS

[00046] The term “phage-assisted continuous evolution (PACE),” as used herein, refers to continuous evolution that employs phage as viral vectors. The general concept of PACE technology has been described, for example, in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Application, U.S. Patent No. 9,023,594, issued May 5, 2015, International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015, and International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, the entire contents of each of which are incorporated herein by reference.

[00047] The term “continuous evolution,” as used herein, refers to an evolution process, in which a population of nucleic acids encoding a gene to be evolved is subjected to multiple rounds of (a) replication, (b) mutation, and (c) selection to produce a desired evolved version of the gene to be evolved that is different from the original version of the gene, for example, in that a gene product, such as, *e.g.*, an RNA or protein encoded by the gene, exhibits a new activity not present in the original version of the gene product, or in that an activity of a gene product encoded by the original gene to be evolved is modulated (increased or decreased). The multiple rounds can be performed without investigator intervention, and the steps (a)-(c) can be carried out simultaneously. Typically, the evolution procedure is carried out *in vitro*, for example, using cells in culture as host cells. In general, a continuous evolution process provided herein relies on a system in which a gene encoding a gene product of interest is provided in a nucleic acid vector that undergoes a life-cycle including replication in a host cell and transfer to another host

cell, wherein a critical component of the life-cycle is deactivated and reactivation of the component is dependent upon an activity of the gene to be evolved that is a result of a mutation in the nucleic acid vector.

[00048] The term “vector,” as used herein, refers to a nucleic acid that can be modified to encode a gene of interest and that is able to enter into a host cell, mutate and replicate within the host cell, and then transfer a replicated form of the vector into another host cell. Exemplary suitable vectors include viral vectors, such as retroviral vectors or bacteriophages, and conjugative plasmids. Additional suitable vectors will be apparent to those of skill in the art based on the instant disclosure.

[00049] The term “viral vector,” as used herein, refers to a nucleic acid comprising a viral genome that, when introduced into a suitable host cell, can be replicated and packaged into viral particles able to transfer the viral genome into another host cell. The term viral vector extends to vectors comprising truncated or partial viral genomes. For example, in some embodiments, a viral vector is provided that lacks a gene encoding a protein essential for the generation of infectious viral particles. In suitable host cells, for example, host cells comprising the lacking gene under the control of a conditional promoter, however, such truncated viral vectors can replicate and generate viral particles able to transfer the truncated viral genome into another host cell. In some embodiments, the viral vector is a phage, for example, a filamentous phage (*e.g.*, an M13 phage). In some embodiments, a viral vector, for example, a phage vector, is provided that comprises a gene of interest to be evolved.

[00050] The term “phage,” as used herein interchangeably with the term “bacteriophage,” refers to a virus that infects bacterial cells. Typically, phages consist of an outer protein capsid enclosing genetic material. The genetic material can be ssRNA, dsRNA, ssDNA, or dsDNA, in either linear or circular form. Phages and phage vectors are well known to those of skill in the art and non-limiting examples of phages that are useful for carrying out the methods provided herein are λ (Lysogen), T2, T4, T7, T12, R17, M13, MS2, G4, P1, P2, P4, Phi X174, N4, Φ 6, and Φ 29. In certain embodiments, the phage utilized in the present invention is M13. Additional suitable phages and host cells will be apparent to those of skill in the art and the invention is not limited in this aspect. For an exemplary description of additional suitable phages and host cells, see Elizabeth Kutter and Alexander Sulakvelidze: *Bacteriophages: Biology and Applications*. CRC Press; 1st edition (December 2004), ISBN: 0849313368; Martha R. J. Clokie and Andrew M. Kropinski: *Bacteriophages: Methods and Protocols, Volume 1*:

Isolation, Characterization, and Interactions (Methods in Molecular Biology) Humana Press; 1st edition (December, 2008), ISBN: 1588296822; Martha R. J. Clokie and Andrew M. Kropinski: *Bacteriophages: Methods and Protocols, Volume 2: Molecular and Applied Aspects (Methods in Molecular Biology)* Humana Press; 1st edition (December 2008), ISBN: 1603275649; all of which are incorporated herein in their entirety by reference for disclosure of suitable phages and host cells as well as methods and protocols for isolation, culture, and manipulation of such phages).

[00051] The term “accessory plasmid,” as used herein, refers to a plasmid comprising a gene required for the generation of infectious viral particles under the control of a conditional promoter. In the context of continuous evolution of genes, transcription from the conditional promoter of the accessory plasmid is typically activated, directly or indirectly, by a function of the gene to be evolved. Accordingly, the accessory plasmid serves the function of conveying a competitive advantage to those viral vectors in a given population of viral vectors that carry a version of the gene to be evolved able to activate the conditional promoter or able to activate the conditional promoter more strongly than other versions of the gene to be evolved. In some embodiments, only viral vectors carrying an “activating” version of the gene to be evolved will be able to induce expression of the gene required to generate infectious viral particles in the host cell, and, thus, allow for packaging and propagation of the viral genome in the flow of host cells. Vectors carrying non-activating versions of the gene to be evolved, on the other hand, will not induce expression of the gene required to generate infectious viral vectors, and, thus, will not be packaged into viral particles that can infect fresh host cells.

[00052] The term “helper phage,” as used herein, interchangeable with the terms “helper phagemid” and “helper plasmid,” refers to a nucleic acid construct comprising a phage gene required for the phage life cycle, or a plurality of such genes, but lacking a structural element required for genome packaging into a phage particle. For example, a helper phage may provide a wild-type phage genome lacking a phage origin of replication. In some embodiments, a helper phage is provided that comprises a gene required for the generation of phage particles, but lacks a gene required for the generation of infectious particles, for example, a full-length pIII gene. In some embodiments, the helper phage provides only some, but not all, genes for the generation of infectious phage particles. Helper phages are useful to allow modified phages that lack a gene for the generation of infectious phage particles to complete the phage life cycle in a host cell. Typically, a helper phage will comprise the genes for the generation of infectious phage particles

that are lacking in the phage genome, thus complementing the phage genome. In the continuous evolution context, the helper phage typically complements the selection phage, but both lack a phage gene required for the production of infectious phage particles.

[00053] The term “selection phage,” as used herein interchangeably with the term “selection plasmid,” refers to a modified phage that comprises a nucleic acid sequence encoding a tRNA synthetase to be evolved, and lacks a full-length gene encoding a protein required for the generation of infectious phage particles. For example, some M13 selection phages provided herein comprise a nucleic acid sequence encoding a gene to be evolved, *e.g.*, under the control of an M13 promoter, and lack all or part of a phage gene encoding a protein required for the generation of infectious phage particles, *e.g.*, gI, gII, gIII, gIV, gV, gVI, gVII, gVIII, gIX, or gX, or any combination thereof. For example, some M13 selection phages provided herein comprise a nucleic acid sequence encoding a tRNA synthetase protein to be evolved, *e.g.*, under the control of an M13 promoter, and lack all or part of a gene encoding a protein required for the generation of infective phage particles, *e.g.*, the gIII gene encoding the pIII protein.

[00054] The term “mutagenesis plasmid,” as used herein, refers to a plasmid comprising a gene encoding a gene product that acts as a mutagen. In some embodiments, the gene encodes a DNA polymerase lacking a proofreading capability. In some embodiments, the gene is a gene involved in the bacterial SOS stress response, for example, a UmuC, UmuD', or RecA gene. In some embodiments, the gene is a GATC methylase gene, for example, a deoxyadenosine methylase (dam methylase) gene. In some embodiments, the gene is involved in binding of hemimethylated GATC sequences, for example, a seqA gene. In some embodiments, the gene is involved with repression of mutagenic nucleobase export, for example emrR. Mutagenesis plasmids (also referred to as mutagenesis constructs) are described, for example by International Patent Application, PCT/US2016/027795, filed April 16, 2016, published as WO2016/168631 on October 20, 2016, the entire contents of which are incorporated herein by reference.

[00055] The term “nucleic acid,” as used herein, refers to a polymer of nucleotides. The polymer may include natural nucleosides (*i.e.*, adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine), nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 7-deazaadenosine,

7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, 4-acetylcytidine, 5-(carboxyhydroxymethyl)uridine, dihydrouridine, methylpseudouridine, 1-methyl adenosine, 1-methyl guanosine, N6-methyl adenosine, and 2-thiocytidine), chemically modified bases, biologically modified bases (*e.g.*, methylated bases), intercalated bases, modified sugars (*e.g.*, 2'-fluororibose, ribose, 2'-deoxyribose, 2'-O-methylcytidine, arabinose, and hexose), or modified phosphate groups (*e.g.*, phosphorothioates and 5'-N-phosphoramidite linkages).

[00056] The term “protein,” as used herein, refers to a polymer of amino acid residues linked together by peptide bonds. The term, as used herein, refers to proteins, polypeptides, and peptide of any size, structure, or function. Typically, a protein will be at least three amino acids long. A protein may refer to an individual protein or a collection of proteins. Inventive proteins preferably contain only natural amino acids, although non-natural amino acids (*i.e.*, compounds that do not occur in nature but that can be incorporated into a polypeptide chain; see, for example, cco.caltech.edu/~dadgrp/Unnatstruct.gif, which displays structures of non-natural amino acids that have been successfully incorporated into functional ion channels) and/or amino acid analogs as are known in the art may alternatively be employed. Also, one or more of the amino acids in an inventive protein may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein may also be a single molecule or may be a multi-molecular complex. A protein may be just a fragment of a naturally occurring protein or peptide. A protein may be naturally occurring, recombinant, or synthetic, or any combination of these.

[00057] The term “evolved synthetase protein,” as used herein, refers to a tRNA synthetase protein variant that is expressed by a gene of interest (*e.g.*, a gene encoding a wild-type synthetase protein, such as a wild-type pyrrolysyl-tRNA synthetase) that has been subjected to continuous evolution, such as PACE or SE-PACE. Examples of evolved synthetase proteins include but are not limited to evolved aminoacyl-tRNA synthetases (AARSs), such as evolved pyrrolysyl-tRNA synthetase (PylRS) proteins and evolved tyrosyl-tRNA synthetase (TyrRS) proteins.

[00058] The term “wild-type pyrrolysyl-tRNA synthetase (PylRS)” refers to the amino acid sequence of a pyrrolysyl-tRNA synthetase (PylRS) protein as it naturally occurs in the genome of the host from which it is derived. Examples of a wild-type PylRS proteins include *Methanosarcina barkeri* PylRS (*MbPylRS*), which is represented by the amino acid sequence set

forth in SEQ ID NO: 20 or the amino acid sequence of NCBI Accession Number WP_011305865.1, or *Methanosarcina mazei* PylRS (*MmPylRS*), which is represented by the amino acid sequence set forth in SEQ ID NO: 21 or the amino acid sequence of NCBI Accession Number WP_011033391.1.

[00059] The term “wild-type tyrosyl-tRNA synthetase (TyrRS)” refers to the amino acid sequence of a tyrosyl-tRNA synthetase (TyrRS) protein as it naturally occurs in the genome of the host from which it is derived. Examples of a wild-type TyrRS proteins include *M. jannaschii* TyrRS (*MjTyrRS*), which is represented by the amino acid sequence set forth in SEQ ID NO: 24 or the amino acid sequence of NCBI Accession Number WP_010869888.1.

[00060] The term “pyrrolysyl-tRNA synthetase (PylRS) protein variant” refers to a PylRS protein having one or more amino acid variations introduced into the amino acid sequence, *e.g.*, as a result of application of the PACE method, as compared to the amino acid sequence of a naturally-occurring or wild-type PylRS protein. Amino acid sequence variations may include one or more mutated residues within the amino acid sequence of the PylRS protein variant, *e.g.*, as a result of a change in the nucleotide sequence encoding the protein that results in a change in the codon at any particular position in the coding sequence, the deletion of one or more amino acids (*e.g.*, a truncated protein), the insertion of one or more amino acids, or any combination of the foregoing. In some embodiments, the N- or C-terminal domain of a PylRS variant is a variant of a naturally-occurring PylRS from an organism, that does not occur in nature. In some embodiments, a PylRS variant or PylRS N- or C-terminal domain variant is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring PylRS (or corresponding PylRS domain) from an organism.

[00061] The term “tyrosyl-tRNA synthetase (TyrRS) protein variant” refers to a TyrRS protein having one or more amino acid variations introduced into the amino acid sequence, *e.g.*, as a result of application of the PACE method, as compared to the amino acid sequence of a naturally-occurring or wild-type TyrRS protein. Amino acid sequence variations may include one or more mutated residues within the amino acid sequence of the TyrRS protein variant, *e.g.*, as a result of a change in the nucleotide sequence encoding the protein that results in a change in the codon at any particular position in the coding sequence, the deletion of one or more amino acids (*e.g.*, a truncated protein), the insertion of one or more amino acids, or any combination of the foregoing. In some embodiments, a TyrRS variant is a variant of a naturally-occurring

TyrRS from an organism, or a variant of an evolved TyrRS that does not occur in nature. In some embodiments, a TyrRS variant or TyrRS N- or C-terminal domain variant is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring TyrRS (or corresponding TyrRS domain) from an organism.

[00062] The term “recombinant” as used herein in the context of proteins or nucleic acids refers to proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence.

[00063] The term “mutation,” as used herein, refers to a substitution of a residue within a sequence, *e.g.*, a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

[00064] The term “fusion protein” as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an “amino-terminal fusion protein” or a “carboxy-terminal fusion protein,” respectively. A protein may comprise different domains, for example, a nucleic acid binding domain (*e.g.*, the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage domain or a catalytic domain of a nucleic-acid editing protein. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th

ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

[00065] The term “chimeric protein” refers to a fusion protein in which the first protein portion and the second protein portion are derived from different species. For example, in some embodiments, a chimeric PylRS protein comprises an N-terminal domain of a *MbPylRS* (*e.g.*, *MbPylRS* amino acids 1-149 as set forth in SEQ ID NO: 20) and a C-terminal domain from a *MmPylRS* (*e.g.*, *MmPylRS* amino acids 185-454 as set forth in SEQ ID NO: 21).

[00066] The term “host cell,” as used herein, refers to a cell that can host, replicate, and transfer a phage vector useful for a continuous evolution process as provided herein. In embodiments where the vector is a viral vector, a suitable host cell is a cell that can be infected by the viral vector, can replicate it, and can package it into viral particles that can infect fresh host cells. A cell can host a viral vector if it supports expression of genes of viral vector, replication of the viral genome, and/or the generation of viral particles. One criterion to determine whether a cell is a suitable host cell for a given viral vector is to determine whether the cell can support the viral life cycle of a wild-type viral genome that the viral vector is derived from. For example, if the viral vector is a modified M13 phage genome, as provided in some embodiments described herein, then a suitable host cell would be any cell that can support the wild-type M13 phage life cycle. Suitable host cells for viral vectors useful in continuous evolution processes are well known to those of skill in the art, and the disclosure is not limited in this respect.

[00067] The term “stop codon”, as used herein, refers to a three-nucleotide sequence that is present within messenger RNA (mRNA) and typically functions to terminate protein translation. Examples of stop codons include the DNA sequences “TAG” or “UAG” (also referred to as an “amber codon”), “TAA” or “UAA” (also referred to as an “ochre” codon), and “TGA” or “UGA” (also referred to as an “opal” or “umber” codon). In some embodiments, a tRNA synthetase protein variant, for example a PylRS protein variant, is evolved to recognize one or more stop codons and allow protein translation to read through the codon to produce a full-length protein. In some embodiments, a PylRS protein variant is evolved to enable a tRNA to insert a pyrrolyine amino acid at protein position encoded a canonical stop codon (*e.g.*, an amber stop codon) of an mRNA. In some embodiments, a TyrRS protein variant is evolved to enable a tRNA to insert a p-iodo-L-phenylalanine amino acid at protein position encoded a canonical stop codon (*e.g.*, an amber stop codon) of an mRNA.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS

[00068] Some aspects of this disclosure provide tRNA synthetase variants (*e.g.*, PylRS protein variants, TyrRS protein variants, *etc.*) and methods, compositions, and systems for producing the same. In some embodiments, the disclosure relates to the use of phage-assisted continuous evolution (PACE) to produce tRNA synthetase protein variants. In some embodiments, tRNA synthetase protein variants described by the disclosure exhibit improved activity (*e.g.*, improved incorporation of target non-canonical amino acids (ncAAs) into tRNAs) or amino acid specificity (*e.g.*, charging of preferred ncAAs) relative to the wild-type or variant tRNA synthetase protein from which they are derived. Some aspects of this disclosure provide fusion proteins, such as chimeric PylRS protein variants comprising an N-terminal domain of a *MbPylRS* protein or *MbPylRS* protein variant and a C-terminal domain of a *MmPylRS* protein or *MmPylRS* protein variant.

PylRS Protein Variants

[00069] Some aspects of the disclosure relate to tRNA synthetase protein variants. The disclosure is based, in part, on certain tRNA synthetase protein variants (*e.g.*, PylRS variants, *etc.*) that are orthogonal (*e.g.*, with respect to a non-archaeobacterial cell, for example an *E. coli* cell) and are characterized by increased activity and amino acid specificity relative to wild-type tRNA synthetase proteins (*e.g.*, the PylRS protein from which the variant was evolved). In some embodiments, tRNA synthetase protein variants described by the disclosure are characterized by improved (*e.g.*, increased) charging activity (*e.g.*, binding of a non-canonical amino acid to a tRNA via aminoacylation) relative to wild-type tRNA synthetase proteins (*e.g.*, the PylRS protein from which the variant was evolved).

[00070] The tRNA synthetase protein variants described by the disclosure are typically derived from a wild-type PylRS protein and have at least one variation in the amino acid sequence of the variant protein as compared to the amino acid sequence of the cognate wild-type tRNA synthetase protein. In some embodiments, a tRNA synthetase protein variant has at least one variation in its encoding nucleic acid sequence that results in a change in the amino acid sequence present within a cognate wild-type tRNA synthetase protein. The variation in amino acid sequence generally results from a mutation, insertion, or deletion in a DNA coding sequence. Mutation of a DNA sequence can result in a nonsense mutation (*e.g.*, a transcription termination codon (TAA, TAG, or TGA) that produces a truncated protein), a missense mutation

(*e.g.*, an insertion or deletion mutation that shifts the reading frame of the coding sequence), or a silent mutation (*e.g.*, a change in the coding sequence that results in a codon that codes for the same amino acid normally present in the cognate protein, also referred to sometimes as a synonymous mutation). In some embodiments, mutation of a DNA sequence results in a non-synonymous (*i.e.*, conservative, semi-conservative, or radical) amino acid substitution.

[00071] The tRNA synthetase protein can be any tRNA synthetase protein known in the art. In some embodiments, a tRNA synthetase protein variant is a pyrrolysyl-tRNA synthetase (PylRS) protein variant. In some embodiments, a wild-type PylRS protein is a *M. bakeri* PylRS (*MbPylRS*) protein. In some embodiments, a *MbPylRS* is represented by the amino acid sequence set forth in NCBI Accession Number WP_011305865.1 or SEQ ID NO: 20.

Additional PylRS proteins are described, for example, in Wan et al. (2014) *Biochim Biophys Acta* 1844(6):1059-1070.

[00072] In some aspects, the disclosure relates to chimeric tRNA synthetase proteins. In some embodiments, a chimeric PylRS protein or chimeric PylRS protein variant (*e.g.*, a chimeric PylRS protein that has been subjected to PACE) comprises an N-terminal domain from a first PylRS protein and a C-terminal domain from a second PylRS protein. In some embodiments, an N-terminal domain comprises amino acids 1-149 of a PylRS protein, for example *MbPylRS* (amino acids 1-149 of SEQ ID NO: 20). In some embodiments, a C-terminal domain comprises amino acids 185-454 of a PylRS protein, for example *MmPylRS* (amino acids 185-454 of SEQ ID NO: 21). Examples of chimeric PylRS proteins and chimeric PylRS protein variants are described by the nucleic acid sequences set forth in SEQ ID NOs: 5, 8, and 11-17.

[00073] In some embodiments, a PylRS protein variant and a wild-type PylRS protein (*e.g.*, *MbPylRS* or *MmPylRS*) are from about 50% to about 99.9% identical, about 55% to about 95% identical, about 60% to about 90% identical, about 65% to about 85% identical, or about 70% to about 80% identical at the amino acid sequence level. In some embodiments, a PylRS protein variant comprises an amino acid sequence that is at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or at least 99.9% identical to the amino acid sequence of a wild-type PylRS protein (*e.g.*, *MbPylRS* or *MmPylRS*). In some embodiments, amino acid sequence identity is based on an alignment against a reference sequence (*e.g.*, a wild-type PylRS protein, for example, SEQ ID NO: 20 or 21) by NCBI Constraint-based Multiple Alignment Tool (COBALT), using the following parameters; Alignment Parameters: Gap penalties-11,-1 and End-Gap penalties-5,-1, CDD Parameters: Use

RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on, and Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

[00074] In some embodiments, a PylRS protein variant is about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, 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 about 99.9% identical to a wild-type PylRS protein (*e.g.*, *MbPylRS*, *MmPylRS*, or a chimeric PylRS).

[00075] The amount or level of variation between a wild-type PylRS protein and a PylRS protein variant can also be expressed as the number of mutations present in the amino acid sequence encoding the PylRS protein variant relative to the amino acid sequence encoding the wild-type PylRS protein. In some embodiments, an amino acid sequence encoding a PylRS protein variant comprises between about 1 mutation and about 100 mutations, about 10 mutations and about 90 mutations, about 20 mutations and about 80 mutations, about 30 mutations and about 70 mutations, or about 40 and about 60 mutations relative to an amino acid sequence encoding a wild-type PylRS protein (*e.g.*, *MbPylRS*, *MmPylRS*, or a chimeric PylRS). In some embodiments, an amino acid sequence encoding a PylRS protein variant comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 mutations relative to an amino acid sequence encoding a wild-type PylRS protein (*e.g.*, *MbPylRS*, *MmPylRS*, or a chimeric PylRS protein). In some embodiments, an amino acid sequence of a PylRS protein variant comprises more than 100 mutations relative to an amino acid sequence of a wild-type PylRS protein.

[00076] Particular combinations of mutations present in an amino acid sequence encoding a PylRS protein variant can be referred to as the “genotype” of the PylRS protein variant. For example, a PylRS protein variant genotype may comprise the mutations V31I, T56P, H62Y, and A100E, relative to a wild-type PylRS protein (*e.g.*, SEQ ID NO: 20 or 21).

TyrRS Protein Variants

[00077] In some aspects, the disclosure relates to tyrosyl-tRNA synthetase (TyrRS) protein variants. The disclosure is based, in part, on certain tRNA synthetase protein variants (*e.g.*, TyrRS variants, *etc.*) that are orthogonal (*e.g.*, with respect to a non-archaeobacterial cell,

for example an *E. coli* cell) and are characterized by increased activity and amino acid specificity relative to wild-type tRNA synthetase proteins (*e.g.*, the TyrRS protein from which the variant was evolved). In some embodiments, tRNA synthetase protein variants described by the disclosure are characterized by improved (*e.g.*, increased) charging activity (*e.g.*, binding of a non-canonical amino acid to a tRNA via aminoacylation) relative to wild-type tRNA synthetase proteins (*e.g.*, the TyrRS protein from which the variant was evolved).

[00078] The tRNA synthetase protein variants described by the disclosure are typically derived from a wild-type TyrRS protein and have at least one variation in the amino acid sequence of the variant protein as compared to the amino acid sequence of the cognate wild-type tRNA synthetase protein. In some embodiments, a tRNA synthetase protein variant has at least one variation in its encoding nucleic acid sequence that results in a change in the amino acid sequence present within a cognate wild-type tRNA synthetase protein. The variation in amino acid sequence generally results from a mutation, insertion, or deletion in a DNA coding sequence. Mutation of a DNA sequence can result in a nonsense mutation (*e.g.*, a transcription termination codon (TAA, TAG, or TGA) that produces a truncated protein), a missense mutation (*e.g.*, an insertion or deletion mutation that shifts the reading frame of the coding sequence), or a silent mutation (*e.g.*, a change in the coding sequence that results in a codon that codes for the same amino acid normally present in the cognate protein, also referred to sometimes as a synonymous mutation). In some embodiments, mutation of a DNA sequence results in a non-synonymous (*i.e.*, conservative, semi-conservative, or radical) amino acid substitution.

[00079] The TyrRS protein can be any TyrRS protein known in the art. In some embodiments, a tRNA synthetase protein variant is tyrosyl-tRNA synthetase (TyrRS) protein variant. In some embodiments, a wild-type TyrRS protein is a *M. jannaschii* TyrRS (*MjTyrRS*) protein. In some embodiments, a *MjTyrRS* is represented by the amino acid sequence set forth in SEQ ID NO: 24 or NCBI Accession Number WP_010869888.1. Additional TyrRS proteins are described, for example, in Bedouelle H. Tyrosyl-tRNA Synthetases. In: Madame Curie Bioscience Database [Internet]. Austin, TX : Landes Bioscience; 2000-2013.

[00080] In some embodiments, a TyrRS protein variant and a wild-type TyrRS protein (*e.g.*, *MjTyrRS*) are from about 50% to about 99.9% identical, about 55% to about 95% identical, about 60% to about 90% identical, about 65% to about 85% identical, or about 70% to about 80% identical at the amino acid sequence level. In some embodiments, a TyrRS protein variant comprises an amino acid sequence that is at least 50%, at least 60%, at least 70%, at least

80%, at least 90%, at least 95%, at least 99%, or at least 99.9% identical to the amino acid sequence of a wild-type TyrRS protein (*e.g.*, *MjTyrRS*). In some embodiments, amino acid sequence identity is based on an alignment against a reference sequence (*e.g.*, a wild-type TyrRS protein, for example, SEQ ID NO: 24) by NCBI Constraint-based Multiple Alignment Tool (COBALT), using the following parameters; Alignment Parameters: Gap penalties-11,-1 and End-Gap penalties-5,-1, CDD Parameters: Use RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on, and Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

[00081] In some embodiments, a TyrRS protein variant is about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, 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 about 99.9% identical to a wild-type TyrRS protein (*e.g.*, *MjTyrRS*).

[00082] The amount or level of variation between a wild-type TyrRS protein and a TyrRS protein variant can also be expressed as the number of mutations present in the amino acid sequence encoding the TyrRS protein variant relative to the amino acid sequence encoding the wild-type TyrRS protein. In some embodiments, an amino acid sequence encoding a TyrRS protein variant comprises between about 1 mutation and about 100 mutations, about 10 mutations and about 90 mutations, about 20 mutations and about 80 mutations, about 30 mutations and about 70 mutations, or about 40 and about 60 mutations relative to an amino acid sequence encoding a wild-type TyrRS protein (*e.g.*, *MjTyrRS*). In some embodiments, an amino acid sequence encoding a TyrRS protein variant comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 mutations relative to an amino acid sequence encoding a wild-type TyrRS protein (*e.g.*, *MjTyrRS*). In some embodiments, an amino acid sequence of a TyrRS protein variant comprises more than 100 mutations relative to an amino acid sequence of a wild-type TyrRS protein.

[00083] Particular combinations of mutations present in an amino acid sequence encoding a PylRS protein variant can be referred to as the “genotype” of the PylRS protein variant. For example, a TyrRS protein variant genotype may comprise the mutations L69F and V235I, relative to a wild-type TyrRS protein (*e.g.*, SEQ ID NO: 24).

Methods of Use

[00084] Some aspects of this disclosure provide methods of using the tRNA synthetase protein variants provided herein. For example, some aspects of this disclosure provide methods comprising contacting a tRNA with a tRNA synthetase protein variant as described by the disclosure (*e.g.*, a PylRS protein variant or a TyrRS protein variant), in the presence of a cognate non-canonical amino acid, for example pyrrolysine (in the case of PylRS) or *p*-iodo-L-phenylalanine (in the case of TyrRS), under conditions under which the tRNA synthetase protein variant “charges” (binds) the non-canonical amino acid to the tRNA.

[00085] In some embodiments, the tRNA, tRNA synthetase protein variant, and the non-canonical amino acid are contacted to one another in a cell, for example a bacterial cell. In some embodiments, the cell in which the tRNA, tRNA synthetase protein variant, and the non-canonical amino acid are contacted to one another does not naturally express the tRNA synthetase protein variant or the tRNA synthetase protein from which the variant is derived (*e.g.*, the tRNA synthetase protein variant is orthogonal to the cell). In some embodiments, the cell is a non-archaeobacteria cell, for example an *E. coli* cell.

[00086] Methods described by the disclosure are useful, in some embodiments, for charging (*e.g.*, binding) a transfer RNA (tRNA) with a non-canonical amino acid by an aminoacylation reaction. In some embodiments, the activity (*e.g.*, catalytic efficiency of aminoacylation) of a tRNA synthetase protein variant described by the disclosure is between about 2-fold and about 50-fold increased relative to a wild-type tRNA synthetase enzyme. In some embodiments, the activity of a tRNA synthetase protein variant is increased about 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, 30-fold, 40-fold, or 50-fold relative to the activity of a wild-type tRNA synthetase protein. In some embodiments, the activity of a tRNA synthetase protein variant is increased more than 50-fold relative to a wild-type tRNA synthetase protein.

[00087] In some embodiments, tRNA synthetase protein variants described by the disclosure are characterized by improved substrate specificity (*e.g.*, reduced incorporation of off-target or undesirable amino acids into a tRNA) relative to a wild-type tRNA synthetase protein. In some embodiments, substrate specificity of a tRNA synthetase protein variant described by the disclosure is increased about 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-

fold, 15-fold, 20-fold, 25-fold, 30-fold, 40-fold, or 50-fold relative to the activity of a wild-type tRNA synthetase protein.

[00088] In some aspects, the disclosure relates to orthogonal translation systems (OTSs) that allow non-canonical amino acids (ncAAs) to be site-specifically incorporated into recombinant proteins, for example during translation of the recombinant protein in a cell or *in vitro*. Accordingly, in some embodiments, the disclosure provides a method for incorporating a non-canonical amino acid (ncAA) into a peptide, the method comprising expressing in a cell containing an ncAA: 1) a tRNA synthetase protein variant (*e.g.*, PylRS or TyrRS protein variant); 2) a tRNA capable of incorporating the ncAA; and 3) a nucleic acid sequence encoding a protein, wherein the nucleic acid sequence comprises a codon that is recognized (*e.g.*, bound) by the tRNA.

[00089] In some embodiments, the tRNA synthetase protein variant is orthogonal to the cell in which it is being expressed. In some embodiments, the cell is an *E. coli* cell. In some embodiments, the nucleic acid sequence is an mRNA sequence. In some embodiments, the nucleic acid sequence (*e.g.*, mRNA sequence) comprises an amber codon (UAG) that is recognized by the tRNA synthetase.

[00090] In some embodiments, the ncAA is a pyrolysine or a *p*-iodo-L-phenylalanine. In some embodiments, the ncAA is introduced into the culture media surrounding the cell prior to being contained by the cell. In some embodiments, the tRNA synthetase protein variant, tRNA, and the nucleic acid sequence are expressed in the cell prior to the cell containing the ncAA (*e.g.*, the ncAA is added to the cell after expression of the tRNA synthetase protein variant and the nucleic acid sequence).

Vectors and Systems

[00091] Some aspects of this disclosure provide expression constructs encoding gene products that select for a desired physiochemical characteristic or desired function of an evolved tRNA synthetase protein, such as PylRS or TyrRS in a host cell, *e.g.*, in a bacterial host cell. In some embodiments, a PACE selection system comprises one or more gene products encoded by a nucleic acid (*e.g.*, an isolated nucleic acid). In some embodiments, one or more nucleic acids that are operably linked comprise an expression construct. Expression constructs are sometimes also referred to as vectors. In some embodiments, the expression constructs are plasmids.

[00092] In some embodiments, a PACE selection system for production of an evolved tRNA synthetase protein comprises one or more positive selection plasmids. In some embodiments, at least one of the positive selection plasmids is an accessory plasmid (AP). In some embodiments, a positive selection AP comprises a nucleic acid sequence encoding an amber suppressor tRNA and a gene III (*e.g.*, pIII protein). In some embodiments, a positive selection system comprises a complementary plasmid (CP) that encodes T7 RNAP controlled by the phage-shock promoter (P_{psp}), which is induced only upon phage infection. In some embodiments, a positive selection system comprises a mutagenesis plasmid (MP) that increases the rate of evolution during PACE through arabinose-induced production of mutagenic proteins. In some embodiments, the selection phage (SP) encodes all phage genes except gene III, which is replaced by the evolving AARS gene (*e.g.*, gene encoding the tRNA synthetase protein to be evolved). FIGs. 2A-2C provide schematics of positive selection systems described by the disclosure.

[00093] In some embodiments, a PACE selection system for production of an evolved tRNA synthetase protein comprises one or more negative selection plasmids. In some embodiments, one or more of the negative selection plasmids is a negative accessory plasmid (AP-) and a negative complementary plasmid (CP-). In some embodiments, a negative accessory plasmid comprises one or more nucleic acid sequences encoding gene III under the control of a P_{PSP} promoter, an amber suppressor tRNA, and a T7RNA polymerase that comprises amber stop codons and is under the control of a Tet promoter. In some embodiments, a negative complementary plasmid comprises a nucleic acid sequence encoding a dominant-negative variant of gene III (*e.g.*, pIII-neg) under the control of a T7 promoter.

[00094] Without wishing to be bound by any theory, when an SP infects the negative selection host, production of pIII protein from gene III is induced from the phage shock

promoter (P_{psp}) of the AP⁻. If the AARS encoded by the SP (*e.g.*, tRNA synthetase protein to be evolved) can catalyze aminoacylation under the conditions of the negative selection (*e.g.*, in the absence of ncAA), full-length T7 RNAP is produced from the AP⁻ through amber suppression of amber stop codons at position 12 and 203 of the T7 RNAP gene. When full-length T7 RNAP is produced, expression of gene III-neg is induced from the T7 promoter (PT7) of the CP⁻ resulting in production of the dominant-negative pIII-neg protein. The infectivity of progeny phage decreases with the amount of pIII-neg in the host cell. Expression levels of the T7 RNAP gene on the AP⁻ are also controlled by an ATc-inducible promoter (Ptet), allowing the negative selection to be turned on or off during PACE. Non-limiting examples of negative selection PACE systems are described in FIGs. 18A-18B.

[00095] In some embodiments, a selection system for production of an evolved tRNA synthetase protein comprises one or more positive selection plasmids and one or more negative selection plasmids as described herein, and also may be referred to as a “dual-selection system”. In some embodiments, a dual-selection system is described by FIGs. 17A-17C and further in the Examples section below.

[00096] In some aspects, the disclosure relates to methods of evolving a tRNA synthetase protein variant. In some embodiments, the methods comprise the steps of (i) introducing a selection phagemid comprising a gene encoding a tRNA synthetase to be evolved into a flow of bacterial host cells through a lagoon, the host cells comprise phage genes required to package the selection phagemid into infectious phage particles, wherein at least one gene required to package the selection phagemid into infectious phage particles is expressed in response to expression of the gene to be evolved in the host cell, and wherein the flow rate of the host cells through the lagoon permits replication of the phagemid, but not of the host cells, in the lagoon; (ii) replicating and mutating the phagemid within the flow of host cells; and (iii) isolating a phagemid comprising a mutated gene encoding an evolved tRNA synthetase protein variant from the flow of cells.

EXAMPLES

Example 1

[00097] **General methods.** PCR and all cloning steps were performed in HyClone water (GE Healthcare Life Sciences). In all other experiments, water was purified by a MilliQ

purification system (EMD Millipore). PCR was performed with Q5 Hot Start High-Fidelity DNA polymerase (New England Biolabs) when unmodified primers were used, and Phusion U Hot Start DNA polymerase (Thermo Fisher Scientific) was used when deoxyuridine-containing primers were required for USER cloning. Plasmids and selection phage were prepared using isothermal assembly with Gibson Assembly 2x Master Mix (New England Biolabs), USER cloning with USER enzyme (New England Biolabs), or ligation cycling reaction with Ampligase (Epicentre). Genes were either synthesized from gBlock gene fragments (Integrated DNA Technologies) or PCR amplified from native sources. Chimeric PylRS, MbPylRS, and MmPylRS were obtained from pTECH plasmid sources. Premature stop codons and single point mutations were placed into genes using the Q5 Site-Directed Mutagenesis kit (New England Biolabs). The gene encoding p-NFRS was PCR amplified from the pEVOL plasmid, which was generously provided to us by P. Schultz of the Scripps Research Institute. DNA vector amplification was performed using TOP10, Mach1 (Thermo Fisher Scientific) or NEB 5-alpha F' Iq (New England Biolabs) cells. All Sanger sequencing of plasmids and SPs was performed from DNA samples that had been amplified using the Illustra Templiphi 100 Amplification Kit (GE Healthcare Life Sciences). All ncAAs were purchased from Chem-Impex International except for 4-nitro-L-phenylalanine (Nanjing Pharmatechs) and 4-iodo-L-phenylalanine (Astatech, Inc.).

[00098] Non-continuous phage propagation. S1030 cells (25 μ L) were electroporated with the accessory plasmid of interest and a complementary plasmid, when required (Table 1). Transformed cells recovered 1 h in SOC media (New England Biolabs) at 37 °C while shaking. Recovered cells were plated on LB agar (United States Biologicals) containing the antibiotics required for plasmid maintenance and grew 20 h at 37 °C. Single colonies of the transformed cells were picked and grown for 16 h in a 37 °C shaker at 230 rpm using 3 mL of Davis rich media (DRM) containing antibiotics. The saturated cultures were diluted 1,000-fold into 3 mL of identical media or media supplemented with 1 mM ncAA where noted. The diluted cultures were grown at 37 °C while shaking to mid-log phase (absorbance at 600 nm (A_{600}) = 0.5–0.7). Once the desired cell density was reached, the cultures were inoculated with selection plasmid (SP) to provide a desired starting titer of $\sim 1 \times 10^5$ pfu/mL. A dilution reference was also prepared by diluting an identical volume of SP into media containing no cells. All cultures and dilution references were shaken for 16 h at 37 °C. The resulting saturated cultures were centrifuged 8

min at 3,000g, and the supernatant was filtered using a 0.22 µm cellulose acetate, Spin-X centrifuge tube filter (Costar), and the samples were stored at 4 °C.

Table 1- Plasmids

Plasmid Name	Class (resistance)	Origin	ORF1		ORF2		ORF3		PACE Experiments
			Prom	[RBS] ² Genes	Prom	Genes	Prom	[RBS] Genes	
pDB007(+)	AP (carb ^H)	SC101	P _{T7}	[SD8] gIII, luxAB	P _{PtrcK}	tyrT ^{Upt} _{CUA}	—	—	p-NFRS
pDB021CH(+)	AP (carb ^H)	SC101	P _{T7}	[SD8] gIII, luxAB	P _{PtrcK}	pylT	—	—	Pyl-1, Pyl-2
pDB026a	AP (carb ^H)	SC101	P _{psp}	[SD8] gIII(P29*), luxAB	P _{PtrcK}	tyrT ^{Upt} _{CUA}	—	—	
pDB026b	AP (carb ^H)	SC101	P _{psp}	[SD8] gIII(P83*), luxAB	P _{PtrcK}	tyrT ^{Upt} _{CUA}	—	—	
pDB026c	AP (carb ^H)	SC101	P _{psp}	[SD8] gIII(T177*), luxAB	P _{PtrcK}	tyrT ^{Upt} _{CUA}	—	—	
pDB026d	AP (carb ^H)	SC101	P _{psp}	[SD8] gIII(Y184*), luxAB	P _{PtrcK}	tyrT ^{Upt} _{CUA}	—	—	
pDB026e	AP (carb ^H)	SC101	P _{psp}	[SD8] gIII(P29*, Y184*), luxAB	P _{PtrcK}	tyrT ^{Upt} _{CUA}	—	—	
pDB026f	AP (carb ^H)	SC101	P _{psp}	[SD8] gIII(P29*, P83*, Y184*), luxAB	P _{PtrcK}	tyrT ^{Upt} _{CUA}	—	—	
pDB026g	AP (carb ^H)	SC101	P _{psp}	[SD8] gIII(P29*, P83*, T177*, Y184*), luxAB	P _{PtrcK}	tyrT ^{Upt} _{CUA}	—	—	
pJC175e	AP (carb ^H)	SC101	P _{psp}	[SD8] gIII, luxAB	—	—	—	—	
pDB038	AP (spec ^H)	ColE1	P _{psp}	[SD8] gIII(P29*), luxAB	P _{PtrcK}	pylT	—	—	Pyl-3
pDB038a	AP (spec ^H)	ColE1	P _{psp}	[SD8] gIII(P29*, Y184*), luxAB	P _{PtrcK}	pylT	—	—	Pyl-3
pDB038b	AP (spec ^H)	ColE1	P _{psp}	[SD8] gIII(P29*, P83*, Y184*), luxAB	P _{PtrcK}	pylT	—	—	Pyl-3
pDB007ns2a	AP ⁻ (carb ^H)	SC101	P _{psp}	[SD8] gIII	P _{PtrcK}	tyrT ^{Upt} _{CUA}	P _{tet}	[SD4] T7RNAP(S12*, S203*)	p-NFRS
pDB036a	AP ⁻ (carb ^H)	SC101	P _{psp}	[SD8] gIII	P _{PtrcK}	tyrT ^{Upt} _{CUA}	P _{prcD}	[SD4] T7RNAP(S12*, S203*)	Countersel.
pDB036d	AP ⁻ (carb ^H)	SC101	P _{psp}	[SD8] gIII	P _{PtrcK}	tyrT ^{Upt} _{CUA}	P _{prcA}	[SD4] T7RNAP(S12*, S203*)	Countersel.
pDB023f	CP (spec ^H)	ColE1	P _{psp}	[SD8] T7RNAP(S12*, S203*)	—	—	—	—	Pyl-1, Pyl-2
pDB023f1	CP (spec ^H)	ColE1	P _{psp}	[SD4] T7RNAP(S12*, S203*)	—	—	—	—	p-NFRS
pDB023k	CP (spec ^H)	ColE1	P _{psp}	[SD8] T7RNAP(S12*, S203*, S527*)	—	—	—	—	
pDB016	CP ⁻ (spec ^H)	ColE1	P _{T7}	[SD8] gIII-neg	—	—	—	—	p-NFRS, Countersel.
DP4	DP (chlor ^H)	cloDF13	P _{psp}	dnaQ926, dam, seqA	P _C	araC	P _{psp-tet}	[sd8] gIII	Pyl-1, Pyl-2, p-NFRS
DP6	DP (chlor ^H)	cloDF13	P _{psp}	dnaQ926, dam, seqA, emrR, ugi, cda1	P _C	araC	P _{psp-tet}	[sd8] gIII	Pyl-3
pBAD-sfGFP	EP (carb ^H)	pBR322	P _{BAD}	sfGFP-6xHis variant	P _C	araC	—	—	
pDB005x(-)	EP (carb ^H)	SC101	P _{lacZ}	[SD8] chPylRS	P _{PtrcK}	pylT	P _{T7}	[SD8] luxAB	
pDB007xb(-)	EP (carb ^H)	SC101	P _{lacZ}	[SD8] p-NFRS	P _{PtrcK}	tyrT ^{Upt} _{CUA}	P _{T7}	[SD8] luxAB	
pDB027c	EP (carb ^H)	SC101	P _{BAD}	[SD8] luxAB(Y361*), [SD8] M/TyrRS variant	P _{PtrcK}	tyrT ^{Upt} _{CUA}	P _C	araC	
pDB032c	EP (carb ^H)	SC101	P _{BAD}	[SD8] luxAB(Y361*), [SD8] PylRS variant	P _{PtrcK}	pylT	P _C	araC	
pDB059c	EP (carb ^H)	SC101	P _{BAD}	[SD8] luxAB(Y361*)	P _C	araC	—	—	
pDB070	EP (chlor ^H)	p15A	P _{tet}	M/TyrRS variant	P _{PtrcK}	tyrT ^{Upt} _{CUA}	P _{PN25}	TetR	
pTECH-Ack3RS	EP (chlor ^H)	p15A	P _{pp}	Ack3RS variant	P _{PtrcK}	pylT	—	—	
pTECH-PylRS	EP (chlor ^H)	p15A	P _{pp}	PylRS variant	P _{PtrcK}	pylT	—	—	
pET28b(+)-sfGFP	EP (Kan ^R)	pBR322	P _{T7}	sfGFP-6xHis variant	P _I	LacI	—	—	
pDB009a	EP (spec ^H)	ColE1	P _{tet}	[SD8] wt T7 RNAP	—	—	—	—	
pDB009b	EP (spec ^H)	ColE1	P _{tet}	[SD8] T7 RNAP(S12*)	—	—	—	—	
pDB009c	EP (spec ^H)	ColE1	P _{tet}	[SD8] T7 RNAP(S203*)	—	—	—	—	
pDB009d	EP (spec ^H)	ColE1	P _{tet}	[SD8] T7 RNAP(S527*)	—	—	—	—	
pDB009f	EP (spec ^H)	ColE1	P _{tet}	[SD8] T7 RNAP(S12*, S203*)	—	—	—	—	
pDB009g	EP (spec ^H)	ColE1	P _{tet}	[SD8] T7 RNAP(Y250*)	—	—	—	—	

pDB009h	EP (spec ^H)	ColE1	P _{1st}	[SD8] T7 RNAP(Y312*)	—	—	—	—	
pDB009i	EP (spec ^H)	ColE1	P _{1st}	[SD8] T7 RNAP(Y250*, Y312*)	—	—	—	—	
pDB009j	EP (spec ^H)	ColE1	P _{1st}	[SD8] T7 RNAP(S12*, S527*)	—	—	—	—	
pDB060-Ack3RS	EP (spec ^H)	ColE1	P _{1pp}	Ack3RS variant	P _{PraK}	pyIT	—	—	
pDB060-IFRS	EP (spec ^H)	ColE1	P _{1pp}	IFRS variant	P _{PraK}	pyIT	—	—	
pDB060-PyIRS	EP (spec ^H)	ColE1	P _{1pp}	PyIRS variant	P _{PraK}	pyIT	—	—	
MP4	MP (chlor ^H)	cloDF13	P _{ppp}	dnaQ926, dam, seqA	P _C	araC	—	—	Pyl-2, <i>p</i> -NFRS
SP-Kan	SP (kan ^H)	M13 f1	P _{gIII}	Kan	—	—	—	—	
SP-chPyIRS	SP (none)	M13 f1	P _{gIII}	[SD4] chPyl	—	—	—	—	Pyl-1
SP-MBP-TEV	SP (none)	M13 f1	P _{gIII}	[SD8] MBP-TEV	—	—	—	—	
SP- <i>p</i> -NFRS	SP (none)	M13 f1	P _{gIII}	[SD4] <i>p</i> -NFRS	—	—	—	—	<i>p</i> -NFRS

[00099] Plaque assay. S1030 cells transformed with the appropriate plasmids were grown in 2xYT liquid media (United States Biologicals) supplemented with antibiotics required for plasmid maintenance to $A_{600} = 0.6$ – 0.8 . Phage supernatant was serially diluted at 10-fold or 100-fold increments yielding either eight or four total samples, respectively, including undiluted sample. For each phage sample, 100 μ L of cells were combined with 10 μ L of phage. Within 2 min from phage infection, 950 μ L of 55 °C top agar (7 g/L bacteriological agar in 2xYT; no antibiotics) was added and mixed with the phage-infected cells by gentle pipetting once up and down while avoiding formation of bubbles. The final mixtures were plated onto quartered Petri plates that had been previously poured with 1.5 mL of bottom agar (15 g/L bacteriological agar in 2xYT; no antibiotics). Once the overlaid agar congealed, the plates were incubated 16 h at 37 °C to allow plaque formation. When plaque formation was dependent on orthogonal AARS activity, 1 mM ncAA was also added to all liquid and solid media when denoted. When clonal-phage isolates were required, well separated plaques were picked from plates and grown individually at 37 °C while shaking in 3 mL of DRM supplemented with 1 mM ncAA of interest where required. The resulting saturated cultures were pelleted at 3,000g for 8 min, and the phage supernatant was sterile filtered and stored at 4 °C for further analysis.

[000100] Phage-assisted continuous evolution of aminoacyl-tRNA synthetases. In general, the PACE apparatus—including host-cell strains, lagoons, chemostats, and media—was used as previously described, for example in WO2010/028347. All liquid and solid media contained antibiotics required for plasmid maintenance unless indicated otherwise. To prepare each PACE strain, the accessory plasmid (AP), complementary plasmid (CP), and MP or drift plasmid (DP) of interest were cotransformed into electrocompetent S1030 cells, which recovered for 1 h in SOC medium without antibiotics (New England Biolabs). The recovered

transformants were plated onto 2xYT agar containing 0.4% glucose to prevent induction of mutagenesis prior to PACE, and colonies were grown for 16–20 h in a 37 °C incubator. Three colonies were picked and resuspended in DRM. A portion of each suspension was tested for arabinose sensitivity as previously described, and the remainder was used to inoculate liquid cultures in DRM, which were subsequently grown for 16 h in a 230 rpm shaker at 37 °C.

[000101] Each PACE chemostat was prepared by diluting an arabinose-sensitive overnight culture into 40 mL or 80 mL of DRM, which was supplemented with ncAA where noted, and the chemostats grew at 37 °C while stirring with a magnetic stir bar. Once the culture reached an approximate cell density of $A_{600} = 1.0$, fresh DRM (supplemented with ncAA where noted) was used to continuously dilute the chemostat culture at a dilution rate of 1.6 chemostat volumes per h while maintaining a constant culture volume as previously described.

[000102] Lagoons flowing from the chemostats were continuously diluted using the indicated flow rates while maintaining a 25-mL constant volume by adjusting the height of the needle drawing waste out of each lagoon. All lagoons were supplemented with 25 mM arabinose from a syringe pump to induce mutagenesis from the MP or DP, unless otherwise indicated. Arabinose supplementation began at least two hours prior to phage infection to insure cells were maximally induced at the start of each experiment. Lagoons were also supplemented with anhydrotetracycline (ATc), where noted, to induce either genetic drift (mutagenesis under weak or no selective pressure) or negative selection depending on the nature of the host-cell plasmids.

[000103] Samples of evolving SP pools were taken periodically at indicated time points from the waste line of each lagoon. Collected samples were centrifuged at 10,000g for 2 min, and the supernatant was passed through a 0.22 μm filter and stored at 4 °C for subsequent analysis. Phage titers were determined by plaque assays using S1059 cells (containing the phage-responsive pJC175e plasmid to report total phage titer) and untransformed S1030 cells (reporting cheaters from unwanted recombination of gene III into the SP) for all collected samples. Activity-dependent plaque assays were performed for mock selection PACE experiments, using S1030 cells cotransformed with the AP and CP used in the host cells of the corresponding experiment. Mock selections were also monitored by PCR performed on phage aliquots using primers DB212 (5'-CAAGCCTCAGCGACCGAATA; SEQ ID NO: 1) and DB213 (5'-GGAAACCGAGGAAACGCAA; SEQ ID NO: 2), which anneal to regions of the phage backbone flanking the gene of interest.

[000104] Evolution of chPylRS (Pyl-1). Host cells cotransformed with pDB021CH(+), pDB023f, and DP4 were maintained in an 80 mL chemostat using media containing 1 mM BocK. At the beginning of PACE, genetic drift was induced (200 ng/mL ATc) in a lagoon that was flowing from the chemostat at 1 lagoon volume per h. The lagoon was infected with 10^8 pfu of SP-chPylRS to start the experiment. ATc supplementation was adjusted to 20 ng/mL at 16 h of PACE to slowly reduce the amount of genetic drift, and ATc supplementation was stopped at 24 h. The lagoon flow rate was increased to 2 volumes per h at 40 h of PACE to increase selection stringency for the remainder of the experiment, which ended at 120 h.

[000105] Continuation of chPylRS evolution (Pyl-2). Three preparations of media were used, which contained different concentrations of *N*ε-(*tert*-butoxycarbonyl)-L-lysine (BocK) (DRM-A: 1 mM BocK; DRM-B: 0.5 mM BocK; DRM-C: 0.25 mM BocK). Host cells cotransformed with pDB021CH(+), pDB023f, and MP4 were maintained in a 40 mL chemostat containing DRM-A at the start of the experiment. A single lagoon was flowed from the chemostat at 1 lagoon volume per h, and the experiment was initiated by infecting the lagoon with 100 μ L (2×10^4 pfu) of the evolved pool of SP collected from the 120-h end point of Pyl-1. To increase the selection stringency during the experiment, the media being pumped into the chemostat was changed to DRM-B at 42 h of PACE and was changed to DRM-C at 69 h. The experiment ended at 168 h.

[000106] Continuation of chPylRS evolution (Pyl-3). PACE was conducted in two separate lagoons, L1 and L2, and a concentration of 1 mM BocK was maintained throughout the experiment. Selection stringency was increased during the experiment by modulating the lagoon flow rate and altering the ratio of host cells in the lagoons to increase the number of amber suppression events required to produce full-length pIII during translation (Host-A: pDB038 and DP6; Host-B: pDB038a and DP6; Host-C: pDB038b and DP6). Host cells were maintained separately in three, 40 mL chemostats (C1–C3, respectively), and each chemostat was individually prepared and coupled to both lagoons, as needed, over the course of the experiment to minimize media waste and to minimize the total growth time of each chemostat culture.

[000107] At the start of the experiment L1 and L2 were continuously diluted with Host-A from C1 at a rate of 0.5 lagoon volumes per h, and genetic drift was induced only in L1 (100 ng/mL ATc). Each lagoon was infected with 10^8 pfu of clonal-phage isolate SP-Pyl2.288-2, which was isolated from the Pyl-2 segment (Table 2; bold residues responsible for enhancing activity of chPylRS). The flow rate from C1 through each lagoon was increased to 1 lagoon

volume per h at 41 h of PACE. At the 91-h mark of the experiment, L1 and L2 were fed a 1:1 mixture of Host-A:Host-B supplied from C1 and C2, respectively, and the flow through each lagoon was maintained at 1 lagoon volume per h. At the 120-h mark, 100% Host-B was flowed to each lagoon at 0.5 lagoon volumes per h, and the flow rate was later doubled to 1 lagoon volume per h at 136 h of PACE. At the 162 h mark, L1 and L2 were fed a 1:1 mixture of Host-B:Host-C supplied from C2 and C3, respectively, and the flow through each lagoon was maintained at 1 lagoon volume per h. ATc supplementation to L1 was stopped at 184 h of PACE to end genetic drift. At the 190-h time point of PACE, 100% Host-C was flowed to each lagoon at 0.5 lagoon volumes per h for the remainder of the experiment, which was stopped at 209 h.

Table 2

chPylRS Residue	Pyl-2.162.1-5					Pyl-2.189.1-5					Pyl-2.288.1-5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
D2															E
D7	E														
A12						G									
V31		I	I	I	I	I	I	I	I	I	I	I	I	I	I
T56		P	P	P	P	P	P	P	P	P	P	P	P	P	P
H62							Y								
E77						K									
T91	S														
A100											E	E	E	E	E
K104														E	
R113		H													
L118						M									
A150		V													
R217									S						
D257		G	G	G	G		G	G	G	G	G	G	G	G	G
N259				S				S							
L266								I					I		
P282								S							
I327					M		M								
G336											E				
D338							E								

[000108] Evolution of p-NFRS with dual selection. The media of the positive selection contained 1 mM *p*-IF, and media of the negative selection contained 4 mM *p*-NF. Three host-cell strains (Host-A: pDB007(+), pDB023f1, and DP4; Host-B: pDB007(+), pDB023f1, and MP4; Host-C: pDB007(+), ns2a, pDB016, and MP4) were used, and were maintained separately in three, 80 mL chemostats (C1–C3), respectively. Host-A was pumped into a positive-selection

lagoon (L1-pos) at a flow rate of 1 lagoon volume per h, and genetic drift was induced by supplementation with 200 ng/mL ATc to the lagoon. L1-pos was infected with 10^8 pfu of SP-*p*-NFRS to initiate the experiment, and supplementation of ATc was stopped at 24 h to end genetic drift. Concomitant with the end of genetic drift, C1 was disconnected from L1-pos, and the lagoon was connected to C2 (containing Host-B), which was pumped into L1-pos at 1 lagoon volume per h. Also at this time, L1-pos was cross coupled to a second lagoon (L2-neg), which was being continuously flowed with negative-selection Host-C from C3. The maximum level of negative-selection stringency from Host-C was maintained by supplementing L2-neg with 30 ng/mL ATc. Cross coupling of the opposing selection lagoons was accomplished using two lines of Masterflex Microbore two-stop tubes (silicone; platinum cured; 0.89 mm ID) (Cole-Parmer), which each had a dead volume of 1 mL. One of the tubes was used to transfer small volumes of culture from L1-pos into L2-neg, and the second tube transferred material in the opposing direction. Material was peristaltically pumped through the cross-coupling lines with a Masterflex L/S Standard Digital Drive (Cole-Parmer) equipped with a Masterflex L/S 8-channel multichannel pump head for microbore tubing (Cole-Parmer). The flow rate through each cross-coupling line was initially set to 0.5 mL/h to maintain a 50-fold dilution of the transferred material into the opposing lagoons. The flow rate through L1-pos and L2-neg was doubled to 2 lagoon volumes per h at 28 h, and flow through the cross-coupling lines was adjusted to 1 mL/h to maintain 50-fold dilution of transferred material in each direction. The experiment ended at 48 h.

[000109] Luciferase assay. S1030 cells (25 μ L) were electroporated with the appropriate plasmid(s) and recovered in SOC media (New England Biolabs) for 1 h while shaking at 37 °C. Transformed cells were plated and grown overnight at 37 °C on LB agar containing the antibiotics required for plasmid maintenance. Single colonies were used to inoculate 2–3 mL of DRM containing antibiotics and were grown overnight at 37 °C while shaking at 230 rpm. The saturated overnight cultures were diluted 100-fold in a 96-well deep well plate using 1 mL of DRM containing the required antibiotic and supplemented with 1 mM ncAA where denoted. The plate was shaken at 37 °C for 2 h at 230 rpm and then supplemented with the indicated concentration of isopropyl- β -D-thiogalactosidase (IPTG), anhydrotetracycline (ATc), or 1 mM arabinose—depending on the nature of the plasmids—to induce protein expression. The plate continued to incubate with shaking at 37 °C for an additional 2–3 h until maximum luminescence signal was observed. Each luminescence measurement was taken on 150 μ L of

each culture, which had been transferred to a 96-well black wall, clear bottom plate (Costar). The A_{600} and luminescence measurements from each well were taken using an Infinite M1000 Pro microplate reader (Tecan). Background A_{600} measurements were taken on wells containing media only. The raw luminescence value from each well was divided by the background-subtracted A_{600} value of the corresponding well to provide the luminescence value normalized to cell density. All variants were assayed in at least biological triplicate, and error bars represent the standard deviation of the independent measurements.

[000110] sfGFP assay. In assays of *Mj*TyrRS variants (*p*-NFRS, *p*-IFRS, and PACE-evolved), a pDB070 plasmid containing the AARS of interest and a pET28b(+) containing the superfolder GFP (sfGFP) of interest were cotransformed into chemically competent BL21 Star (DE3) cells (Thermo Fisher Scientific). The transformed cells recovered in SOC (New England Biolabs) for 1 h while shaking at 37 °C and were then plated and grown overnight at 37 °C on LB agar containing 50 µg/mL kanamycin and 25 µg/mL chloramphenicol. Single colonies were used to inoculate 2 mL of LB media (United States Biologicals) containing antibiotics and were grown overnight at 37 °C while shaking at 230 rpm. The saturated overnight cultures were diluted 100-fold in a 96-well deep well plate using 600 µL of LB media containing the required antibiotic and were grown at 37 °C to a cell density of $A_{600} = 0.3$ while shaking at 230 rpm. AARS expression was induced by addition of LB media (200 µL) containing antibiotics and the additional components to provide each well with a final concentration of 200 ng/mL anhydrotetracycline (ATc) and 1 mM ncAA where indicated. Incubation continued until cultures reached a cell density of $A_{600} = 0.5$. Each well was then supplemented with 1 mM isopropyl-β-D-thiogalactosidase (IPTG) to induce sfGFP expression.

[000111] In assays of PylRS variants (including AcK3RS variants), a pTECH plasmid containing the AARS of interest and a pBAD plasmid containing the sfGFP of interest were cotransformed into chemically competent TOP10 cells (Thermo Fisher Scientific). The transformed cells recovered in SOC (New England Biolabs) for 1 h while shaking at 37 °C and were then plated and grown overnight at 37 °C on LB agar containing 100 µg/mL carbenicillin and 25 µg/mL chloramphenicol. Single colonies were used to inoculate 2–3 mL of LB media (United States Biologicals) containing antibiotics and were grown overnight at 37 °C while shaking at 230 rpm. The saturated overnight cultures were diluted 100-fold in a 96-well deep well plate using 500 µL of LB media containing the required antibiotic. The plate was shaken at 37 °C for 3 h at 230 rpm and an additional 500 µL of LB was added containing antibiotics and

additional components to provide each well with a final concentration of 1 mM ncAA where denoted and 1.5 mM arabinose to induce expression of sfGFP.

[000112] For all experiments, the cultures incubated with shaking at 37 °C for an additional 16 h after induction of sfGFP, and 150 µL of each culture was transferred to a 96-well black wall, clear bottom plate (Costar). The A_{600} and fluorescence (excitation = 485 nm; emission = 510 nm; bandwidth of excitation and emission = 5 nm) readings from each well were taken using an Infinite M1000 Pro microplate reader (Tecan). Background A_{600} and background fluorescence measurements were taken on wells containing LB media only. The background-subtracted fluorescence value from each well was divided by the background-subtracted A_{600} value of the same well to provide the fluorescence value normalized to cell density. All variants were assayed in at least biological triplicate, and error bars represent the standard deviation of the independent measurements.

[000113] Protein expression and purification of sfGFP. Expression of His-tagged sfGFP was performed using the plasmids, cell strains, and antibiotic concentrations described in the methods for sfGFP assays. Saturated overnight cultures were prepared from single colonies of cotransformed cells, which were diluted 1,000-fold into 300 mL of LB media containing 1 mM ncAA where denoted and were grown while shaking at 230 rpm at 37 °C. Once the cultures utilizing a pDB070 plasmid grew to a cell density of $A_{600} = 0.3$, AARS expression was induced by supplementing with anhydrotetracycline (ATc) to a final concentration of 200 ng/mL, and incubation was continued. Cultures utilizing a pTECH plasmid did not require this step as the AARS was expressed constitutively. Once cultures grew to a cell density of $A_{600} = 0.5$, sfGFP expression was induced by supplementation with a final concentration of 1 mM isopropyl- β -D-thiogalactosidase (IPTG) for cultures utilizing a pET28b(+) plasmid or a final concentration of 1 mM arabinose for cultures utilizing a pBAD plasmid. Incubation with shaking at 37 °C continued for an additional 16 h after induction of sfGFP expression. Cells were harvested by centrifugation at 5,000g for 10 min at 4 °C, and the resulting pellets were resuspended in B-PER II Bacterial Protein Extraction Reagent (Thermo Fisher Scientific) containing EDTA-free protease inhibitor cocktail (Roche). The soluble fraction of the cell lysates were diluted by an equal volume of equilibration buffer (20 mM Tris (pH 7.4), 10 mM imidazole, 300 mM NaCl) and were separately loaded onto a column containing 2 mL of HisPur Ni-NTA resin (Thermo Fisher Scientific) that had been pre-washed with two bed-volumes of equilibration buffer. The resin was washed with two bed-volumes of wash buffer (20 mM Tris (pH 7.4), 25 mM

imidazole, 300 mM NaCl) and protein was then eluted in 3 mL of elution buffer (20 mM Tris (pH 7.4), 250 mM imidazole, 300 mM NaCl). The purified protein was dialyzed against 20 mM Tris (pH 7.4), 150 mM NaCl, 5 mM EDTA, 1 mM 2-mercaptoethanol (BME), and 10% glycerol. Purified protein was stored at -80 °C until analysis.

[000114] AARS variant expression and purification for aminoacylation assays. The genes of chPylRS variants and *Mj*TyrRS variants were cloned into pET15a and transformed into BL21(DE3) (New England Biolabs) cells for expression. Cells were grown in 500 mL of LB media supplemented with 100 µg/mL ampicillin at 37 °C to an A_{600} of 0.6–0.8, and protein expression was induced by addition of 0.5 mM IPTG (chPylRS variants) or 1 mM IPTG (*Mj*TyrRS variants). Cells were incubated at 30°C for an additional 6 h (chPylRS variants) or 4 h (*Mj*TyrRS variants) and harvested by centrifugation at 5,000g for 10 min at 4 °C. The cell pellet was resuspended in 15 mL of lysis buffer (50 mM Tris (pH 7.5), 300 mM NaCl, 20 mM imidazole), and cells were lysed by sonication. The crude extract was centrifuged at 20,000g for 30 min at 4 °C. The soluble fraction was loaded onto a column containing 2 mL of Ni-NTA resin (Qiagen) previously equilibrated with 20 mL of lysis buffer. The column was washed with 20 mL of lysis buffer. The bound protein was then eluted with 2 mL of 50 mM Tris (pH 7.5), 300 mM NaCl, 300 mM imidazole. The purified protein was dialyzed with 50 mM HEPES-KOH (pH 7.5), 50 mM KCl, 1 mM DTT and 50% glycerol and stored at -80 °C for further studies.

[000115] Purification of c-Myc-chPylRS-6xHis variants. The chPylRS variants were cloned into the pTech plasmid using insertion primers that incorporate the N-terminal c-Myc sequence (MEQKLISEEDL-; SEQ ID NO: 3) and the C-terminal 6xHis sequence (-GSHHHHHH; SEQ ID NO: 4). BL21 star (DE3) cells (Thermo Fisher Scientific) transformed with the appropriate pTech plasmids were grown in LB media (United States Biologicals) supplemented with 25 µg/mL chloramphenicol. For each variant, a saturated overnight culture was prepared from a single colony, and a 1:100 dilution of culture was made into 5 mL of fresh LB media containing chloramphenicol. The starter culture grew at 37 °C while shaking at 230 rpm until the cell density reached $A_{600} = 0.3$. The starter culture was then used to inoculate a 1 L culture of LB media containing chloramphenicol, which continued to incubate while shaking for an additional 16 h. Cells were harvested by centrifugation at 5,000g for 10 min at 4 °C, and cell pellets were resuspended in lysis buffer (20 mM Tris (pH 7.4), 300 mM NaCl, 10 mM imidazole, and EDTA-free protease inhibitor cocktail (Roche)). The cells were lysed by

sonication on ice, and the crude extract was centrifuged at 15,000g for 15 min at 4 °C. Lysates were loaded onto columns containing 2 mL of HisPur Ni-NTA resin (Thermo Fisher Scientific) that had been pre-washed with two bed-volumes of equilibration buffer. The resin was washed with 10 bed-volumes of wash buffer (20 mM Tris (pH 7.4), 25 mM imidazole, 300 mM NaCl) and protein was then eluted in 3 mL of elution buffer (20 mM Tris (pH 7.4), 250 mM imidazole, 300 mM NaCl). The purified protein was dialyzed against 20 mM Tris (pH 7.4), 150 mM NaCl, 5 mM EDTA, 1 mM dithiothreitol. Purified protein was stored in 20% glycerol at -80 °C until analysis.

[000116] Western blot analysis of c-Myc-chPylRS-6xHis variants. Cell lysates (30 µL) of expressed protein were combined with 25 µL of XT Sample Buffer (Bio-Rad), 5 µL of 2-mercaptoethanol, and 40 µL water. The samples were heated at 70 °C for 10 min and 7.5 µL of prepared sample was loaded per well of a Bolt Bis-Tris Plus Gel (Thermo Fisher Scientific). Precision Plus Protein Dual Color Standard (4 µL) Bio-Rad was used as the reference ladder. The loaded gel was run at 200V for 22 min in 1x Bolt MES SDS running buffer (Thermo Fisher Scientific). The gel was transferred to a PVDF membrane using the iBlot 2 Gel Transfer Device (Thermo Fisher Scientific). The membrane was blocked for 1 h at room temperature in 50% Odyssey blocking buffer (PBS) (Li-Cor) and was then soaked 4 times for 5 min in PBS containing 0.1% Tween-20 (PBST). The blocked membrane was soaked with primary antibodies (rabbit anti-6xHis (1:1,000 dilution) (Abcam, ab9108) and mouse anti-c-Myc (1:7,000 dilution) (Sigma-Aldrich, M4439)) in 50% Odyssey buffer (PBS) containing 0.2% Tween-20 for 4 h at room temperature. The membrane was washed four times in PBST, and then soaked for 1 h in the dark at room temperature with secondary antibodies (donkey anti-mouse 800CW (1:20,000 dilution) (Li-Cor) and goat anti-rabbit 680RD (1:20,000 dilution) (Li-Cor)) in Odyssey buffer containing 0.01% SDS, 0.2% Tween-20. The membrane was washed 4 times in PBST and finally rinsed with PBS. The membrane was scanned using an Odyssey Imaging System (Li-Cor).

[000117] LCMS analysis of intact purified proteins. Purified protein samples were diluted to 10 µM in dialysis buffer lacking reducing agent or glycerol prior to analysis on an Agilent 6220 ESI-TOF mass spectrometer equipped with an Agilent 1260 HPLC. Separation and desalting was performed on an Agilent PLRP-S Column (1,000A, 4.6 x 50 mm, 5 µm). Mobile the phase A was 0.1% formic acid in water and mobile phase B was acetonitrile with 0.1% formic acid. A constant flow rate of 0.250 mL/min was used. Ten microliters of the protein

solution was injected and washed on the column for the first 3 min at 5% B, diverting non-retained materials to waste. The protein was then eluted using a linear gradient from 5% B to 100% B over 7 min. The mobile phase composition was maintained at 100% B for 5 min and then returned to 5% B over 1 minute. The column was then re-equilibrated at 5%B for the next 4 min. Data was analyzed using Agilent MassHunter Qualitative Analysis software (B.06.00, Build 6.0.633.0 with Bioconfirm). The charge state distribution for the protein produced by electrospray ionization was deconvoluted to neutral charge state using Bioconfirm's implementation of MaxEnt algorithm, giving a measurement of average molecular weight. The average molecular weight of the proteins were predicted using ExPASy Compute pI/Mw tool (http://web.expasy.org/compute_pi/), and each calculation was adjusted for chromophore maturation in sfGFP and any ncAA substitutions.

[000118] Amber suppressor tRNA preparation. Template plasmid containing the tRNA^{Pyl} or tRNA_{CUA}^{Tyr/Opt} gene was purified with the plasmid maxi kit (Qiagen). The plasmid containing tRNA^{Pyl} (100 µg) was digested with BstNI (New England Biolabs). The tRNA_{CUA}^{Tyr/Opt} gene was amplified by PCR. The BstNI digested template DNA or PCR product was purified by phenol chloroform extraction, followed by ethanol precipitation and dissolved in double distilled water. A His-tagged T7 RNA polymerase was purified over a column of Ni-NTA resin according to manufacturer's instructions (Qiagen). The transcription reaction (40 mM Tris (pH 8); 4 mM each of UTP, CTP, GTP, and ATP at pH 7.0; 22 mM MgCl₂; 2 mM spermidine; 10 mM DTT; 6 µg pyrophosphatase (Roche Applied Science); 60 µg/mL of DNA template, approximately 0.2 mg/mL T7 RNA polymerase) was performed in 10 mL reactions overnight at 37 °C. The tRNA was purified on 12% denaturing polyacrylamide gel containing 8 M urea and TBE buffer (90 mM Tris, 90 mM boric acid, 2 mM EDTA). UV shadowing was used to illuminate the pure tRNA band, which was excised and extracted three times with 1M sodium acetate pH 5.3 at 4 °C. The tRNA extractions were then ethanol precipitated, dissolved in RNase-free distilled water, pooled, and finally desalted using a Biospin 30 column (BioRad). The tRNA was refolded by heating to 100 °C for 5 min and slow cooling to room temperature. At 65 °C, MgCl₂ was added to a final concentration of 10 mM to aid folding. A His-tagged CCA adding enzyme was purified over column of Ni-NTA resin according to manufacturer's instructions (Qiagen). 16 µM refolded tRNA in 50 mM Tris (pH 8.0), 20 mM MgCl₂, 5 mM DTT, and 50 µM NaPPi was incubated at room temperature for 1 h with approximately 0.2 mg/mL CCA-adding enzyme and 1.6 µCi/µL of (α-³²P)-labeled ATP (PerkinElmer). The sample

was phenol/chloroform extracted and then passed over a Bio-spin 30 column (Bio-Rad) to remove excess ATP.

[000119] Aminoacylation assay. A 20 μ L aminoacylation reaction contained the following components for chPylRS variants: 50 mM HEPES-KOH (pH 7.2), 25 mM KCl, 10 mM MgCl_2 , 5 mM DTT, 10 mM ATP, 25 $\mu\text{g/mL}$ pyrophosphatase (Roche Applied Science), 10 mM amino acids, 500 nM PylRS variants, 5 μM unlabeled tRNA^{Pyl} , and 100 nM ^{32}P -labeled tRNA^{Pyl} . A 20 μ L aminoacylation reaction contained the following components for *Mj*TyrRS variants: 50 mM Tris-HCl (pH 7.5), 1 mM DTT, 10 mM MgCl_2 , 10 mM ATP, 20 μM unlabeled $\text{tRNA}_{\text{CUA}}^{\text{Tyr/Opt}}$, 3 μM ^{32}P -labeled $\text{tRNA}_{\text{CUA}}^{\text{Tyr/Opt}}$, 2 μM *Mj*TyrRS variants. Various concentrations of ATP (1–100 μM), BocK (0.1–10 mM), Pyl (5–500 μM), Phe (0.1–3.2 mM), *p*-NF (1–32 mM), *p*-IF (1–32 mM), and tRNA (0.05–5 μM) were used to determine K_M values for corresponding substrates. Time points were taken at 5 min, 20 min and 60 min by removing 2 μL aliquots from the reaction and immediately quenching the reaction into an ice-cold 3 μL quench solution (0.66 $\mu\text{g}/\mu\text{L}$ nuclease P1 (Sigma) in 100 mM sodium citrate (pH 5.0)). For each reaction, 2 μL of blank reaction mixture (containing no enzyme) was added to the quench solution as the start time point. The nuclease P1 mixture was then incubated at room temperature for 30 min and 1 μL aliquots were spotted on PEI-cellulose plates (Merck) and developed in running buffer containing 5% acetic acid and 100 mM ammonium acetate. Radioactive spots for AMP and AA-AMP (representing free tRNA and aminoacyl-tRNA, respectively) were separated and then visualized and quantified by phosphorimaging using a Molecular Dynamics Storm 860 phosphorimager (Amersham Biosciences). The ratio of aminoacylated tRNA to total tRNA was determined to monitor reaction progress.

Example 2

[000120] The development of orthogonal translation systems (OTSs) that allow non-canonical amino acids (ncAAs) to be site-specifically incorporated into recombinant proteins has enabled researchers to dramatically expand the genetic code. More than 200 ncAAs have been installed into designer proteins using OTSs in prokaryotes, eukaryotic cells, and even in whole animals. The most common strategy for genetic code expansion *in vivo* requires three key components. An unused or rarely used codon (typically the TAG nonsense codon) is placed into a gene's coding sequence at the position(s) of desired ncAA incorporation. An orthogonal tRNA (o-tRNA) that is not recognized by host endogenous aminoacyl-tRNA synthetases

(AARSs) decodes the nonsense codon during translation. Lastly, an orthogonal AARS is required, which is typically a variant that researchers have evolved to selectively aminoacylate the o-tRNA, but not endogenous tRNAs, with the target ncAA (FIG. 1). This third component must be generated for each different ncAA of interest, and evolving a tailor-made orthogonal AARS is by far the most challenging and labor-intensive requirement of this strategy.

[000121] Although researchers have evolved many AARSs to incorporate ncAAs into proteins, several outstanding challenges limit their utility and generality. Laboratory evolution of AARSs with altered amino acid specificity typically relies on three to five rounds of sequential positive and negative selections from an AARS library containing either partially or fully randomized residues in the amino acid-binding pocket. The limited number of rounds of selection typically conducted in AARS evolution campaigns reflects the effort required to complete each round of evolution, which is on the order of one week or longer. A consequence of conducting relatively few rounds of selection on libraries that focus mutagenesis on and around the amino acid-binding pocket is that laboratory-evolved AARSs routinely emerge with suboptimal properties, for example ~1,000-fold reduced activity (k_{cat}/K_M) compared to their wild-type counterparts, and modest selectivity for the target ncAA over endogenous amino acids that can require compensation with high concentrations of ncAA and expression in minimal media, lowering protein yields. The modest enzymatic efficiency and selectivity of many laboratory-evolved AARSs are longstanding challenges that limit the production and purity of expressed proteins containing ncAAs.

[000122] This example describes phage-assisted continuous evolution (PACE) selections that enable the laboratory evolution of orthogonal AARSs over hundreds of generations of mutation, selection, and replication on practical time scales. AARS PACE was performed over 268 generations to evolve pyrrolysyl-tRNA synthetase (PylRS) variants that acquired up to a 45-fold improvement in enzymatic efficiency ($k_{\text{cat}}/K_M^{\text{tRNA}}$) compared to the parent PylRS. The enabling mutations from PACE were also successfully transplanted into other PylRS variants without requiring further evolution, resulting in up to 9.7-fold higher expression of ncAA-containing protein when introduced into a previously reported PylRS-derived synthetase, AcK3RS. Interestingly, PACE also gave rise to unexpected mutations that split PylRS into mutually dependent N- and C-terminal fragments that maintained high activity and specificity when co-expressed, mimicking naturally occurring split PylRS homologs. In addition, a promiscuous mutant *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS) was

evolved into a variant with >23-fold higher selectivity for the desired amino acid, *p*-iodo-L-phenylalanine, over the undesired substrate, *p*-nitro-L-phenylalanine, in 48 h of PACE.

Together, these results establish a rapid and effective approach to improve the catalytic efficiency and alter the amino acid specificity of AARS enzymes.

[000123] Development of a positive PACE selection for AARS activity. PACE has enabled the rapid laboratory evolution of diverse classes of proteins including polymerases, proteases, genome-editing agents, and insecticidal proteins. Two strategies by which aminoacylation of an orthogonal amber suppressor tRNA would induce pIII production through amber suppression are described. In the first strategy, amber suppression of premature stop codons in the T7 RNA polymerase (T7 RNAP) gene allows translation of full-length T7 RNAP, which transcribes gene III from an upstream T7 promoter. This approach results in pIII production in an amplified manner since each amber suppression event can give rise to many gene III transcripts. In the second, more stringent strategy, amber suppression of premature stop codons in gene III results in direct translation of full-length pIII without amplification (FIGs. 2A-2C).

[000124] To implement the first selection strategy, permissive residues in T7 RNAP that would not inhibit enzymatic activity when mutated to a wide variety of amino acids were identified. The number of amber codons are needed in the T7 RNAP gene to make full-length translation of the polymerase completely dependent on orthogonal translation was also identified. Amber mutations were installed in the T7 RNAP gene at Ser-12, Ser-203, Tyr-250, Tyr-312, and Ser-527, positions predicted from the crystal structure that avoid perturbation of RNA polymerization or DNA binding. Suppression with *p*-nitro-L-phenylalanine (*p*-NF) (FIG. 3) at combinations of these sites using a previously evolved *Mj*TyrRS variant (*p*-NFRS) revealed that a minimum of two amber stop codons were required for transcriptional activation by T7 RNAP to become fully dependent on both the AARS and the ncAA substrate (FIGs. 4A-4B). Similar results were observed with site-specific installation of *Nε*-(*tert*-butoxycarbonyl)-L-lysine (BocK) using a chimeric PylRS (chPylRS), comprising residues 1–149 of *Methanosarcina barkeri* PylRS (*Mb*PylRS) and residues 185–454 of *Methanosarcina mazei* PylRS (*Mm*PylRS) (FIG. 4C).

[000125] To test the ability of this selection to support phage propagation, selection phage (SP) expressing either *p*-NFRS (SP-*p*-NFRS) or chPylRS (SP-chPylRS) were propagated non-continuously in cultures of host *E. coli* cells harboring an accessory plasmid (AP) and complementary plasmid (CP) that together expressed the requisite amber suppressor tRNA, T7

RNAP(S12TAG, S203TAG), and gene III downstream of a T7 promoter. It was observed that SP propagation in these cultures was dependent on the presence of a matched ncAA substrate (FIGs. 5A-5B). Together, these results validate the PACE selection strategy based on amplified expression of gene III through amber suppression of two or more stop codons in T7 RNAP.

[000126] To implement the second, more stringent selection strategy based on direct amber suppression of premature stop codons in gene III, amber mutations were installed at positions Pro-29, Pro-83, Thr-177, or Tyr-184 of gene III. These residues were chosen because they are predicted to be uninvolved in pIII binding to the host cell TolA protein or to the host cell F pilus. The N-terminal signal peptide of pIII, which spans residues 1–18, was not targeted for amber suppression as this region is required for insertion of pIII into the host inner membrane. The ability of this selection to support phage propagation was investigated by challenging selection phage expressing either *p*-NFRS (SP-*p*-NFRS) or chPylRS (SP-chPylRS) to propagate non-continuously in cultures of host *E. coli* cells harboring an accessory plasmid that expressed the requisite amber suppressor tRNA and gene III containing one or more premature stop codons. It was observed that each of the mutated positions in pIII were permissive to ncAA incorporation, and the presence of a single premature stop codon in the coding sequence of pIII was sufficient to make robust phage propagation dependent on AARS activity from SP-*p*-NFRS or SP-chPylRS (FIGs. 5C-5D). Collectively, these developments identify positions in T7 RNAP and pIII that tolerate a range of amino acid side chains, and thereby establish two strategies to link AARS activity to phage infectivity through amber suppression of premature stop codons in T7 RNAP or in gene III.

[000127] Next, whether PACE positive selection for aminoacylation based on amber suppression of stop codons in T7 RNAP could support activity-dependent phage propagation in the continuous flow format of PACE was investigated. 48-h mock PACE selections were conducted. It was observed that SP-*p*-NFRS propagated at high phage titer levels in a lagoon supplemented with *p*-NF substrate without further adaptation. In a separate control lagoon, SP expressing a kanamycin resistance gene rather than an AARS were unable to propagate in the positive selection and rapidly washed out (FIG. 6A). It was also observed that the PACE positive selection based on direct amber suppression of stop codons in gene III supported activity-dependent phage propagation in continuous flow. In a single, 30-h mock PACE using this selection strategy, active SP-*p*-NFRS was highly enriched starting from a 1:1 input-mixture of SP-*p*-NFRS and an SP expressing an unrelated gene (FIG. 6B). Together, these results

confirmed that both selection strategies were capable of supporting phage propagation in PACE in a manner dependent on orthogonal AARS activity.

[000128] To demonstrate that an AARS with little or no starting activity on a target amino acid could evolve new activity to propagate in the positive selection, an additional mock PACE experiment in which SP-*p*-NFRS was challenged to evolve acceptance of endogenous amino acids was performed by propagating the phage in Davis rich media (DRM) that was not supplemented with *p*-NF using the selection requiring amber suppression of stop codons in T7 RNAP. Under these conditions, high titers of SP-*p*-NFRS were dependent on induction of a mutagenesis plasmid (MP), MP4, that enhances the rate of mutagenesis in the host *E. coli* (FIG. 7A). This observation indicates that mutation of the AARS was required in order for SP to propagate when the cognate amino acid substrate was unavailable. Sanger sequencing analysis of clonal phage from the experiments confirmed that more mutations accumulated in the gene encoding *p*-NFRS when MP4 was induced. Additionally, the evolved mutants, but not the starting *p*-NFRS, displayed strong aminoacylation activity in the absence of *p*-NF using a luciferase reporter of amber suppression (FIG. 7B), confirming that the AARS evolved to accept one or more canonical amino acids during PACE. Together, results of these experiments validated the positive selection for continuously evolving an orthogonal AARS.

[000129] Continuous evolution of catalytically enhanced PylRS variants. PylRS from archaeobacteria are preferred evolutionary starting points for genetic code expansion efforts due to their tRNA orthogonality in a range of hosts. Wild-type PylRS variants, however, are generally hampered by poor catalytic efficiency, which is typically further diminished as an undesired consequence of traditional laboratory evolution.

[000130] chPylRS—the chimera of residues 1–149 of *Mb*PylRS and residues 185–454 of *Mm*PylRS—was evolved to have improved aminoacylation activity over 497 h of PACE in three segments (FIG. 8A). The PylRS substrate analog, BocK, was used in the evolution of PylRS rather than the natural cognate substrate, L-pyrrolysine, which is not readily available. In the first two segments of PACE (Pyl-1 and Pyl-2), SP-chPylRS was evolved using the less stringent selection requiring amber suppression of T7 RNAP. During Pyl-1, the flow rate was modulated to increase selection stringency. The pool of phage surviving Pyl-1 was further evolved in Pyl-2, which challenged SP-chPylRS to propagate as the ncAA substrate concentration was incrementally reduced. The final PACE segment, Pyl-3, was conducted in two lagoons using the more stringent selection strategy, and the number of amber stop codons in gene III was

incrementally increased from one to three to increase demands on PylRS efficiency. This approach gradually increased selection stringency over the 497-h evolution of chPylRS, and emerging variants had survived on average 268 generations of mutation, selection, and replication.

[000131] Clonal SP isolates from the 120-h endpoint of Pyl-1 were sequenced and revealed mutations throughout the PylRS gene with strong convergence toward a pair of mutations in PylRS: V31I and T56P. Sequencing of clonal isolates from the second PACE segment (Pyl-2) revealed strong convergence toward D257G at 162 h, and full convergence on A100E by the end of Pyl-2 (288 h). The additionally stringent conditions of the Pyl-3 segment selected for complete convergence toward mutation H62Y in all sequenced clones from 408 h of PACE (Tables 3-6).

Table 3: Summary of mutations observed in PACE segment Pyl-1

chPylRS Residue	Pyl-1.120.1-8							
	1	2	3	4	5	6	7	8
D7			A					
V31					F			
41A						E		
T56								
A100		T						
S127	P			P				
A152	V							
D257		G						
G343		D						

Mutations in chPylRS from the Pyl-1 segment were determined by Sanger sequencing of eight clonal SP isolates from 120 h of total PACE. Only coding mutations are shown. Shaded mutations were shown to be responsible for enhancing the activity of chPylRS.

Table 4: Summary of mutations observed in PACE segment Pyl-2

chPylRS Residue	Pyl-2.162.1-5					Pyl-2.189.1-5					Pyl-2.288.1-5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
D2														E	
D7	E														
A12						G									
V31															
T56															
H62															
E77						K									
T91	S														
A100															
K104															
R113		H												E	
L118						M									

Mutations in chPylRS from the Pyl-2 segment were determined by Sanger sequencing of five clonal SP isolates from 162 h, 189 h, and 288 h of total PACE. Only coding mutations are shown. Shaded mutations were shown to be responsible for enhancing the activity of chPylRS.

Table 5: Summary of mutations observed in lagoon 1 of PACE segment Pyl-3

48

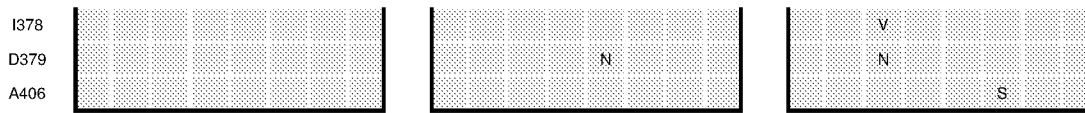
A406



Mutations in chPylRS from lagoon 1 (L1) of the Pyl-3 segment were determined by Sanger sequencing of eight clonal SP isolates from 408 h, 450 h, and 497 h of total PACE. Only coding mutations are shown. Shaded mutations were shown to be responsible for enhancing the activity of chPylRS. Mutations denoted by a star indicate stop codons that resulted in split-protein variants in which translation reinitiates at the position corresponding to Met-107 of chPylRS (**M'**).

Table 6: Summary of mutations observed in lagoon 2 of PACE segment Pyl-3

chPylRS Residue	Pyl3-L2.408.1-8								Pyl3-L2.450.1-8								Pyl3-L2.497.1-8							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
V8									G							G								
T20							P			P										P			P	
I26				V																				
H28																Y								
V31																								
D44																								
H45									R													G		
S53						F																		
T56																								
A59				T																				
H62																								
R73																	H							
D78																				E				
N80															D									
T91									A															
S92																								
V93																								
K94																								
S99																								
A100																								
P101								R																
V103								*																
M107								M'						M'										
S112													P											
E119																								
N120																								
A126								T																
T130												P												
N143				K																				
P147																								
P153																								
P234																								
E236								G																
D257	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
K258														E										
R321																								
S326												N												
G343						D																		
V367																								



Mutations in chPylRS from lagoon 2 (L2) of the Pyl-3 segment were determined by Sanger sequencing of eight clonal SP isolates from 408 h, 450 h, and 497 h of total PACE. Only coding mutations are shown. Shaded mutations were shown to be responsible for enhancing the activity of chPylRS. Mutations denoted by a star indicate stop codons that resulted in split-protein variants in which translation reinitiates at the position corresponding to Met-107 of chPylRS (**M'**).

[000132] Each of the Pyl-1 variants exhibited improved aminoacylation activity in a luciferase reporter of amber suppression with BocK (FIG. 8B). Comparison of the consensus mutations acquired in each segment of PACE showed that the two combined mutations from Pyl-1 increased luciferase signal 8.5-fold compared to the progenitor chPylRS, and the additional two mutations from Pyl-2 improved amber suppression signal 21-fold. The variant containing all consensus mutations from the three segments of PACE provided 24-fold improved amber suppression signal compared to chPylRS while maintaining substrate specificity (FIG. 8C). Further analysis of the consensus mutations acquired in the first two segments of PACE demonstrated that D257G did not significantly contribute to enhance the activity of chPylRS (FIG. 9A). Therefore, the tetramutant comprising V31I, T56P, H62Y, and A100E was responsible for the large improvement in apparent activity.

[000133] BocK was incorporated at up to three positions in sfGFP to compare the relative activity of chPylRS to the tetramutant variant, chPylRS(IPYE), containing the activity-enhancing mutations from PACE (FIG. 8D and FIGs. 10A-10D). Expression of sfGFP containing three BocK residues was improved nearly 4-fold by chPylRS(IPYE) compared to chPylRS. Biochemical characterization of chPylRS(IPYE) using BocK confirmed that the k_{cat} improved 8.7-fold, and the K_M for tRNA^{Pyl} substrate improved 5.7-fold, such that the catalytic efficiency (k_{cat}/K_M^{tRNA}) of the evolved variant was enhanced 45-fold compared to chPylRS (Table 7). These findings indicate that the increased apparent activity of the tetramutant results from catalytic enhancement of chPylRS, rather than solely from non-catalytic improvements such as enhanced protein expression or stability. The outcome of these experiments demonstrates that PACE positive selection is highly effective at improving the activity of an AARS commonly used for genetic code expansion.

Table 7: Kinetic parameters of chPylRS variants containing mutations from PACE.

PylRS variant	k_{cat} , $s^{-1} \times 10^{-3}$	K_M^{ATP} , μM	K_M^{tRNA} , μM	K_M^{BocK} , mM	k_{cat}/K_M^{tRNA} , $\mu M^{-1} \cdot s^{-1} \times 10^{-3}$	Relative catalytic efficiency
---------------	-------------------------------------	-----------------------	------------------------	-------------------	---	-------------------------------

chPylRS	11.88 ± 0.18	2.54 ± 0.16	0.26 ± 0.07	1.03 ± 0.05	45.69	1
V31I, T56P	73.15 ± 1.01	5.74 ± 0.20	0.10 ± 0.02	0.82 ± 0.18	731.50	15.9
V31P, T56P, A100E	110.23 ± 4.65	3.45 ± 1.19	0.13 ± 0.03	0.91 ± 0.08	847.92	18.4
V31P, T56P, H62Y, A100E	103.87 ± 2.37	3.96 ± 0.52	0.05 ± 0.01	1.13 ± 0.23	2,077.40	45.2

[000134] The activity-enhancing coding mutations discovered through PACE were localized exclusively in the N-terminal domain of chPylRS, which is involved in tRNA binding and is typically not targeted for mutagenesis in traditional laboratory evolution efforts. These changes occur at positions conserved in *MbPylRS* and *MmPylRS*, and in some embodiments, the PACE mutations may generally improve the activity of other natural and engineered PylRS homologs. Amber suppression assays using several different reporters demonstrated that the activity of the *MbPylRS*(IPYE) variant was dramatically enhanced while the *MmPylRS*(IPYE) variant was also improved, albeit more modestly (FIGs. 9B-9D). To test whether the beneficial mutations also enhance evolved PylRS enzymes, the four PACE-derived mutations were transplanted into AcK3RS, which was previously evolved to accept *N*ε-acetyl-L-lysine (AcK). *MbAcK3RS*, *MmAcK3RS*, and chimeric AcK3RS (chAcK3RS) variants containing the four mutations each exhibited increased expression of reporter proteins up to 9.7-fold compared to their unmodified PylRS counterparts, without sacrificing amino acid selectivity (FIG. 8E and FIGs. 9E and 11). Reporter expression was also enhanced more than 5-fold when the mutations were transplanted into the PylRS-derived IFRS, which was previously evolved to charge 3-iodo-L-phenylalanine (3-IF) (FIG. 9F). Collectively, these results show that the beneficial mutations discovered exclusively in the N-terminal domain of chPylRS substantially enhance activity in all six additional PylRS variants tested.

Unexpected Evolution of Split PylRS Enzymes

[000135] Although there was no strong convergence toward new beneficial coding mutations between the 408-h and 497-h time points, 13 of 16 (81%) of the sequenced SP isolates from the two lagoons of Pyl-3 acquired a surprising frameshift in their coding sequences by 497 h. Of the thirteen affected clones, 12 of these contained a single frameshift at one of four different locations in chPylRS (Tables 5 and 6). In each case, the shifted reading frame in the chPylRS gene produced a premature ochre (TAA) or opal (TGA) stop codon resulting in a

truncated protein of 93, 99, or 102 residues. In addition, one of the 13 affected isolates from the 497-h time point contained an in-frame ochre stop codon at position Lys-90, resulting in a truncated protein of 89 residues. Downstream of the premature stop codon in every case is a Met codon at canonical position 107 of chPylRS. In some embodiments, protein synthesis reinitiates from Met-107 resulting in a split chPylRS. In assays of amber suppression, the split chPylRS(IPYE) variants exhibited comparable apparent activity as the full-length chPylRS(IPYE) enzyme. Activity of the split variants was strictly dependent on the presence of both fragments (FIG. 12A, FIG. 12C, and Tables 8-9). In contrast, split chPylRS variants lacking the PACE-evolved coding mutations in their N-terminal fragment resulted in significant loss of activity (FIG. 12B), which may explain why the split enzyme was not observed in PACE until the four activity-enhancing mutations were acquired. Results from western blot analysis and ESI-MS analysis of split variants confirmed translational reinitiation from Met-107 (FIGs. 14-15). The prevalence of the split PylRS variants suggest a fitness advantage to the split constructs during PACE, although the molecular basis of this potential advantage is currently unknown.

Table 8: Kinetic parameters of chPylRS variants using L-pyrrolysine substrate.

PylRS variant	k_{cat} , $s^{-1} \times 10^{-3}$	K_M^{Pyl} , μM
chPylRS	33.24 ± 2.74	21.03 ± 0.15
V31I, T56P, A100E	289.16 ± 11.45	18.42 ± 0.69

Table 9: Kinetic parameters of the fusions of split chPylRS variants from PACE.

AARS variant	k_{cat} , $s^{-1} \times 10^{-3}$	K_M^{BocK} , mM	K_M^{tRNA} , μM
Fused Split2	20 ± 1	1.68 ± 0.19	3.62 ± 0.51
Fused Split3	33 ± 3	4.90 ± 0.92	3.84 ± 0.34
Fused Split6	19 ± 0.2	1.00 ± 0.05	3.61 ± 0.38

[000136] The evolution of a split PylRS variant in PACE appears to mirror the evolution of PylRS in nature, as PylRS homologs in certain bacteria are expressed from two separate genes (*pylSc* and *pylSn*). The *D. hafniense* *pylSn* encodes a 110-residue polypeptide that is homologous to the N-terminal region of archaeal PylRS, and an alignment of PylSn to the N-terminal split PylRS evolved in PACE shows that they terminate near the same location (FIG. 16). These observations together demonstrate the ability of PACE to evolve unexpected changes in protein topology.

[000137] Development and validation of AARS negative selections in PACE. While positive selection PACE was able to greatly increase the activity of PylRS, the evolution of AARSs to recognize non-cognate substrates requires the use of negative selections to minimize activity on endogenous amino acids. A PACE negative selection that links tRNA aminoacylation to the *inhibition* of phage propagation was developed. A dominant-negative variant of pIII (pIII-neg) was used as the basis of a PACE negative selection for RNA polymerase activity and DNA binding activity. Because pIII-neg poisons the infectivity of emergent phage, variants possessing undesired activity are unable to effectively propagate and are gradually washed out from the evolving pool of SP under constant dilution.

[000138] In the PACE negative selection for aminoacylation, amber suppression of two stop codons in T7 RNAP(S12TAG, S203TAG) allows transcriptional activation of the gene encoding pIII-neg. Amber suppression in this context thus results in expression of pIII-neg and reduced progeny phage infectivity (FIGs. 17A-17B, and FIGs. 18A-18B). Mock PACE negative selections with SP-*p*-NFRS confirmed negative selection against AARS activity. SP-*p*-NFRS in the presence of *p*-NF quickly washed out of PACE lagoons under negative selection, whereas an SP lacking any AARS activity propagated robustly under the same conditions (FIGs. 19A-19C). These findings established a PACE negative selection against undesired aminoacylation activity.

Continuous evolution of an AARS with greatly improved amino acid selectivity.

[000139] Homogeneity of ncAA incorporation is often crucial for downstream applications, as it is usually impractical or impossible to purify proteins containing the desired ncAA substitution from mixtures containing undesired amino acids at the position(s) of interest. The amino acid selectivity of evolved AARSs is therefore a critical determinant of their utility. The laboratory-evolved *Mj*TyrRS variant, *p*-NFRS, selectively charges *p*-NF in minimal media, but overnight expression in LB media demonstrated that *p*-NFRS also efficiently charges Phe in the presence or in the absence of 1 mM *p*-NF (FIGs. 20A-20C). Additionally, *p*-NFRS is a polyspecific enzyme, as it efficiently charges *p*-iodo-L-phenylalanine (*p*-IF) in addition to *p*-NF (FIG. 20D). The ability of coupled PACE positive and negative selections to generate a highly specific AARS by evolving *p*-NFRS to charge *p*-IF selectively was investigated.

[000140] Opposing positive and negative selections were coupled continuously by constantly exchanging small volumes of material from opposing PACE lagoons, which allows the pool of AARS variants to be evolved in both selections simultaneously, rather than

performing iterative counterselections (FIG. 17C). This strategy's effectiveness relies on (1) the only actively replicating element in the selection lagoons is the SP, (2) the comparatively small number of host cells that are diverted into the opposing selection should not greatly affect either selection due to the much larger population of correct host cells being continuously infused, and (3) any contaminating ncAA diverted into the opposing selection would be diluted to a very low concentration that would be insufficient to support effective aminoacylation. In some embodiments, coupling the opposing selections lagoons provides an opportunity for SP variants capable of propagating in both selections—*i.e.*, those AARS variants that evolved high amino acid selectivity—to outcompete variants able to propagate exclusively in one of the opposing selections.

[000141] SP-*p*-NFRS was evolved for 24 h of positive selection PACE toward *p*-IF followed by 24 h of coupled positive selection with negative selection against the undesired ncAA, *p*-NF (FIGs. 21A-21B). SPs that acquired preferential activity toward *p*-IF in PACE were isolated from the evolved pool using a single round of non-continuous counterselections. To enrich variants possessing little to no activity on the undesired ncAA, endpoint SPs from the PACE negative selection were challenged to propagate non-continuously on negative-selection host cells in media containing 4 mM *p*-NF. The resulting SPs were then challenged with positive-selection host cells in the presence of the desired substrate, 1 mM *p*-IF, for their ability to promote formation of activity-dependent plaques (the result of phage propagation in semi-solid media) and eight of the resulting plaques were sequenced (FIGs. 22A-22B).

[000142] Of the eight sequenced phage isolates, four acquired no new mutations in the AARS gene, but instead emerged from PACE with weakened ribosome binding sites driving AARS expression. Each of the remaining four SP variants contained one or more coding mutations and demonstrated a strong preference for charging *p*-IF over *p*-NF (FIGs. 17D-17E and Table 10). The best performing PACE-evolved variant, Iodo.5, which contained mutations L69F and V235I with respect to *p*-NFRS, matched the amino acid specificity of a previously reported *Mj*TyrRS variant, *p*-IFRS, that was evolved to charge *p*-IF through positive and negative selection on agar plates. Based on our limit of detection in the assay, expression of sfGFP(Asn39TAG) using variant Iodo.5 was >23-fold higher with *p*-IF than with *p*-NF.

Table 10: Kinetic parameters of MjTyrRS variants containing mutations from PACE.

AARS variant	ncAA	$k_{\text{cat}}, \text{s}^{-1} \times 10^{-3}$	$K_{\text{M}}^{\text{ncAA}}, \text{mM}$	$k_{\text{cat}}/K_{\text{M}}^{\text{ncAA}}, \text{mM}^{-1} \cdot \text{s}^{-1} \times 10^{-3}$	Relative catalytic efficiency
<i>p</i> -NFRS	<i>p</i> -NF	1.40 ± 0.05	3.68 ± 0.29	0.38	1.00
<i>p</i> -NFRS	<i>p</i> -IF	0.87 ± 0.11	2.23 ± 0.46	0.39	1.03
<i>p</i> -NFRS	Phe	0.14 ± 0.003	0.16 ± 0.03	0.875	2.3
Iodo.5	<i>p</i> -NF	ND	ND	ND	ND
Iodo.5	<i>p</i> -IF	ND	ND	ND	ND
Iodo.1	<i>p</i> -IF	1.60 ± 1.27	5.65 ± 1.82	0.28	0.74
Iodo.7	<i>p</i> -IF	0.21 ± 0.03	0.92 ± 0.22	0.23	0.61
Iodo.8	<i>p</i> -IF	1.00 ± 0.10	3.80 ± 0.84	0.26	0.68

[000143] The protein sequences of *p*-IFRS and *p*-NFRS differ by only a single amino acid; *p*-NFRS contains Asn160 and *p*-IFRS contains His160. It is possible that His160 also emerged in PACE but was not isolated. We further tested Iodo.5 by expressing the sfGFP reporter in LB media containing both 1 mM *p*-NF and 1 mM *p*-IF in a single culture. Intact protein mass spectrometry of the resulting purified protein revealed the desired mass corresponding to incorporation of *p*-IF with only trace *p*-NF incorporation and no detectable incorporation of Phe at the site of interest (FIGs. 23A-23B). These results establish that PACE can rapidly evolve a highly selective AARS from a polyspecific variant in 48 h with no library cloning.

EQUIVALENTS AND SCOPE

[000144] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the embodiments described herein. The scope of the present disclosure is not intended to be limited to the above description, but rather is as set forth in the appended claims.

[000145] Articles such as “a,” “an,” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include “or” between two or more members of a group are considered satisfied if one, more than one, or all of the group members are present, unless indicated to the contrary or otherwise evident from the context. The disclosure of a group that includes “or” between two or more group members provides embodiments in which exactly one member of the group is present, embodiments in which more than one members of the group are present, and embodiments in

which all of the group members are present. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[000146] It is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitation, element, clause, or descriptive term, from one or more of the claims or from one or more relevant portion of the description, is introduced into another claim. For example, a claim that is dependent on another claim can be modified to include one or more of the limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of making or using the composition according to any of the methods of making or using disclosed herein or according to methods known in the art, if any, are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[000147] Where elements are presented as lists, *e.g.*, in Markush group format, it is to be understood that every possible subgroup of the elements is also disclosed, and that any element or subgroup of elements can be removed from the group. It is also noted that the term “comprising” is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where an embodiment, product, or method is referred to as comprising particular elements, features, or steps, embodiments, products, or methods that consist, or consist essentially of, such elements, features, or steps, are provided as well. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[000148] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in some embodiments, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. For purposes of brevity, the values in each range have not been individually spelled out herein, but it will be understood that each of these values is provided herein and may be specifically claimed or disclaimed. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can

assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

[000149] In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

SEQUENCES

>SEQ ID NO: 5; DNA sequence of **chPylRS**. Bolded codons (Val-31, Thr-56, His-62, and Ala-100) were mutated to 'ATT', 'CCC', 'TAT', and 'GAG', respectively, in the 'IPYE' variant of the enzyme (SEQ ID NO: 25, 26).

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
GCACAAGATCAAGCACTATGAG**GTT**TCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
TGGTTGTGAACAACTCTCGTTCTTGTTCGT**ACC**GCACGTGCATTCCGT**CAT**CATAAAATACCGTAAAACC
TGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACCTCCTGACCCGTTCTACCGAAGGCAAAAC
CTCTGTTAAAGTTAAAGTTGTTTCT**GCG**CCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGTG
CGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCT
CCGCGCAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCCA
GACTGACAGGCTTGAAGTCTGTTAAACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTCA
GGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAG
GGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTCTGGAAA
TAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTT
TCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAA
CTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGAATAGGCCCATGCTACA
GAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACCTTCTGCCAGATGGGATC
GGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTCA
AGATCGTAGGCGATTCTGTCATGGTCTATGGGGATACCCTTGATGTAATGCACGGAGACCTGGAACCTT
TCCTCTGCAGTAGTCGGACCCATAACCGCTTGACCGGGAATGGGGTATTGATAAACCTTGATAGGGGC
AGGTTTCGGACTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGG
TCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

>SEQ ID NO: 6; DNA sequence of **MbPylRS**. Bolded codons (Val-31, Thr-56, His-62, and Ala-100) were mutated to 'ATT', 'CCC', 'TAT', and 'GAG', respectively, in the 'IPYE' variant of the enzyme (SEQ ID NO: 27, 28).

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
GCACAAGATCAAGCACTATGAG**GTT**TCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
TGGTTGTGAACAACTCTCGTTCTTGTTCGT**ACC**GCACGTGCATTCCGT**CAT**CATAAAATACCGTAAAACC
TGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACCTCCTGACCCGTTCTACCGAAGGCAAAAC
CTCTGTTAAAGTTAAAGTTGTTTCT**GCG**CCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGTG
CGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCT
CCGCGCAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCGCCGGCGCCGTCTCTGACCCGTTCTCA
GCTGGATCGTGTTGAAGCGCTGCTGTCTCCGGAAGATAAAATCTCTCTGAACATCGCGAAACCGTTCC
GTGAACCTGGAATCTGAACCTGGTTACCCGTCGTAAAAACGATTTCCAGCGTCTGTACACCAACGATCGT
GAAGACTACCTGGGTAAACTGGAACGTGACATCACCAAAATCTTCGTTGACCGTGATTTCCCTGGAAAT
CAAATCTCCGATCCTGATCCCGGCGGAATACGTTGAACGTATGGGTATCAACAACGATACCGAACTGT
CTAAACAGATCTTCCGTGTTGATAAAAAACCTGTGCCTGCGTCCGATGCTGGCGCCGACCCCTGTACAAC
TATCTGCGTAAACTGGATCGTATCCTGCCGGACCCGATCAAAATCTTCGAAGTTGGTCCGTGCTACCG
TAAAGAATCTGACGGTAAAGAACACCTGGAAGAGTTACCATGGTGAACCTTCTGCCAGATGGGTTCT
GGTTGCACCCGTGAGAACCTGGAATCTCTGATCAAAGAATTTCTGGACTACCTGGAAATCGACTTCGA
AATCGTTGGTGACTCCTGCATGGTGTACGGTGATACCCTGGACATCATGCACGGTGACCTGGAACCTGT
CTTCTGCGGTTGTTGGTCCGGTTCCGCTGGATCGTGAATGGGGTATCGACAAACCGTGGATCGGTGCG
GGTTTCGGTCTGGAACGTCTGCTGAAAGTTATGCACGGTTTCAAAAACATCAAACGTGCGTCTCGTTC
TGAATCTTACTACAACGGTATCTCTACCAACCTGTAA

>SEQ ID NO: 7; DNA sequence of **MmPylRS**. Bolded codons (Val-31, Thr-56, His-62, and Ala-100) were mutated to 'ATT', 'CCC', 'TAT', and 'GAG', respectively, in the 'IPYE' variant of the enzyme (SEQ ID NO: 29, 30).

ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGTCCAGGACCGGAACAA
TTCATAAAATAAACACACGAA**GTC**TCTCGAAGCAAAATCTATATTGAAATGGCATGCGGAGACCA

CCTTGTGTGTAACAACCTCCAGGAGCAGCAGG**ACT**GCAAGAGCGCTCAGG**CAC**CACAAATACAGGAA
 GACCTGCAAACGCTGCAGGGTTTCGGATGAGGATCTCAATAAGTTCCCTCACAAAGGCAAACGAAGAC
 CAGACAAGCGTAAAAAGTCAAGGTCGTTTCT**CCC**CCTACCAGAACGAAAAAGGCAATGCCAAAATCCG
 TTGCGAGAGCCCCGAAACCTCTTGAGAATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTT
 TCACCTGCGATACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCAATATCA
 AGCATTTCTACAGGAGCAACTGCATCCGCACTGGTAAAAGGGAATACGAATCCCATTACATCCATGTC
 TGCCCCCTGTTTCAGGCAAGTGCCCCCGCACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTGTAA
 ACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCTCTCT
 CGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAATTATCTGGGGAAACTCGAG
 CGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTT
 GAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACA
 AGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAACCTACCTGCGCAAGCTTGACAGGGCC
 CTGCCTGATCCAATAAAAAATTTTGAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAAC
 ACCTCGAAGAGTTTACCATGCTGAACCTTCTGCCAGATGGGATCGGGATGCACACGGGAAAATCTTGA
 AAGCATAATTACGGACTTCCTGAACACCTGGGAATTGATTTCAAGATCGTAGGCGATTCTTGCATGG
 TCTATGGGGATACCTTGATGTAATGCACGGAGACCTGGAACCTTTCCTCTGCAGTAGTCGGACCCATA
 CCGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGGCAGGTTTCGGGCTCGAACGCCTTCT
 AAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATT
 TCTACCAACCTGTAA

>SEQ ID NO: 8; DNA sequence of **chAcK3RS**. Bolded codons (Val-31, Thr-56, His-62, and Ala-100) were mutated to 'ATT', 'CCC', 'TAT', and 'GAG', respectively, in the 'IPYE' variant of the enzyme (SEQ ID NO: 31, 32).

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
 GCACAAGATCAAGCACTATGAG**GTT**TCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
 TGGTTGTGAACAACTCTCGTTCTTGTCTG**ACCG**CACGTGCATTCCGT**CAT**CATAAATACCGTAAAACC
 TGCAAACGTTGTCGTGTTTCTGGTGAAGATATCAACAACTTCCTGACCCGTTCTACCGAAGGCAAAAC
 CTCTGTAAAGTTAAAGTTGTTTCT**CCG**CCGAAAGTGAAAAAAGCGATGCCGAAATCTGTTTCTCGTG
 CGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCT
 CCGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCCA
 GACTGACAGGCTTGAAGTCCTGTAAACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTCA
 GGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAG
 GGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAA
 TAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTT
 TCAAAACAGATCTTCAGGGTTGACAAGAATTCTGCCTGAGACCCATGATGGCTCCAAACATTTTAA
 CTACGCTCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGAATAGGCCCATGCTACA
 GAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACCTCTTTCAGATGGGATC
 GGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACACCTGGGAATTGATTTCA
 AGATCGTAGGCGATTCTTGCATGGTCTATGGGGATACCTTGATGTAATGCACGGAGACCTGGAACCTT
 TCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGGC
 AGGTTTTCGGACTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGG
 TCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

>SEQ ID NO: 9; DNA sequence of **MbAcK3RS**. Bolded codons (Val-31, Thr-56, His-62, and Ala-100) were mutated to 'ATT', 'CCC', 'TAT', and 'GAG', respectively, in the 'IPYE' variant of the enzyme (SEQ ID NO: 33, 34).

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
 GCACAAGATCAAGCACTATGAG**GTT**TCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
 TGGTTGTGAACAACTCTCGTTCTTGTCTG**ACCG**CACGTGCATTCCGT**CAT**CATAAATACCGTAAAACC
 TGCAAACGTTGTCGTGTTTCTGGTGAAGATATCAACAACTTCCTGACCCGTTCTACCGAAGGCAAAAC
 CTCTGTAAAGTTAAAGTTGTTTCT**CCG**CCGAAAGTGAAAAAAGCGATGCCGAAATCTGTTTCTCGTG
 CGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCT

CCGGCGAAATCTACCCCGAACTCTCCGGTTCGACCTCTGCGCCGGCGCCGTCTCTGACCCGTTCTCA
 GCTGGATCGTGTGTAAGCGCTGCTGTCTCCGGAAGATAAAATCTCTCTGAACATCGCGAAACCGTTCC
 GTGAACTGGAATCTGAACTGGTTACCCGTCGTAAAAACGATTTCCAGCGTCTGTACACCAACGATCGT
 GAAGACTACCTGGGTAAACTGGAACGTGACATCACCAAATTTCTTCGTTGACCGTGATTTCTGGAAT
 CAAATCTCCGATCCTGATCCCGGCGGAATACGTTGAACGTATGGGTATCAACAACGATACCGAACTGT
 CTAAACAGATCTTCCGTGTTGATAAAAAACCTGTGCCTGCGTCCGATGATGGCGCCGACCATTTTTAAC
 TATGCTCGTAAACTGGATCGTATCCTGCCGGACCCGATCAAAATCTTCGAAGTTGGTCCGTGCTACCG
 TAAAGAATCTGACGGTAAAGAACACCTGGAAGAGTTCACCATGGTGAACCTCTTTTCAGATGGGTTCG
 GTTGCACCCGTGAGAACCTGGAATCTCTGATCAAAGAATTTCTGGACTACCTGGAAATCGACTTCGAA
 ATCGTTGGTGACTCCTGCATGGTGTACGGTGATACCCTGGACATCATGCACGGTGACCTGGAACGTGC
 TTCTGCGGTTGTTGGTCCGGTTCGCTGGATCGTGAATGGGGTATCGACAAACCGTGATCGGTGCGG
 GTTTCGGTCTGGAACGTCTGCTGAAAGTTATGCACGGTTTCAAAAACATCAAACGTGCGTCTCGTTCT
 GAATCTTACTACAACGGTATCTCTACCAACCTGTAA

>SEQ ID NO: 10; DNA sequence of **MmAcK3RS**. Bolded codons (Val-31, Thr-56, His-62, and Ala-100) were mutated to 'ATT', 'CCC', 'TAT', and 'GAG', respectively, in the 'IPYE' variant of the enzyme (SEQ ID NO: 35, 36).

ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGTCCAGGACCGGAACAA
 TTCATAAAATAAAACACCACGAA**GTC**TCTCGAAGCAAAATCTATATTGAAATGGCATGCGGAGACCA
 CCTTGTGTGTAACAACCTCCAGGAGCAGCAGG**ACT**GCAAGAGCGCTCAGG**CAC**CACAAATACAGGAA
 GACCTGCAAACGCTGCAGGGTTTCGGGTGAGGATCTCAATAAGTTCTCTACAAAGGCAAACGAAGAC
 CAGACAAGCGTAAAAGTCAAGGTCTTTCT**CCC**CCTACCAGAACGAAAAAGGCAATGCCAAAATCCG
 TTGCGAGAGCCCCGAAACCTCTTGAGAATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTT
 TCACCTGCGATACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCAATATCA
 AGCATTCTACAGGAGCAACTGCATCCGCACTGGTAAAAGGGAATACGAATCCCATTACATCCATGTC
 TGCCCCTGTTACAGGCAAGTGCCCCCGCACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCCCTGTTAA
 ACCCAAAAGATGAGATTTCCTGAATTCCCGCAAGCTTTCAGGGAGCTTGAGTCCGAATTGCTCTCT
 CGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAATTATCTGGGGAAACTCGAG
 CGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTT
 GAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACA
 AGAACTTCTGCCTGAGACCCATGATGGCTCCAAACATTTTAACTACGCTCGCAAGCTTGACAGGGCC
 CTGCCTGATCCAATAAAAAATTTTGAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAAC
 ACCTCGAAGAGTTTACCATGCTGAACCTTCTTTCAGATGGGATCGGGATGCACACGGGAAAATCTTGAA
 AGCATAATTACGGACTTCTGAACCACCTGGGAATTGATTTCAAGATCGTAGGCGATTCTGTCATGGT
 CTATGGGGATACCTTGATGTAATGCACGGAGACCTGGAACCTTCTCTGCAAGTAGTCGGACCCATAC
 CGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGGCGAGTTTCGGGCTCGAACGCCTTCTA
 AAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTC
 TACCAACCTGTAA

>SEQ ID NO: 11; DNA sequence of **chIFRS**. Bolded codons (Val-31, Thr-56, His-62, and Ala-100) were mutated to 'ATT', 'CCC', 'TAT', and 'GAG', respectively, in the 'IPYE' variant of the enzyme (SEQ ID NO: 37, 38).

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
 GCACAAGATCAAGCACTATGAG**GTT**TCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
 TGGTTGTGAACAACCTCTCGTTCTTGTGCT**ACCG**CACGTGCATTCCGT**CAT**CATAAATACCGTAAAC
 TGCAAACGTTGTCGTGTTCTGACGAAGATATCAACAACCTTCTGACCCGTTCTACCGAAGGCAAAC
 CTCTGTTAAAGTTAAAGTTGTTTCT**GGG**CCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGTG
 CGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCT
 CCGCGAAACTCTACCCGAACCTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCCA
 GACTGACAGGCTTGAGTCCGTGTTAAACCCAAAGATGAGATTTCCTGAATTCCGGGAAGCTTTCA
 GGGAGCTTGAGTCCGAATTGCTCTCTCGCAAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAG
 GGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAA
 TAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTT
 TCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAA
 CTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGAATAGGCCCATGCTACA

GAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGCTCGTTCATTTCAGATGGGATC
GGGATGTACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTC
AGATCGTAGGCGATTCTGTCATGGTCTATGGGGATACCTTGATGTAATGCACGGAGACCTGGAACTT
TCCTCTGCAGTAGTCGGACCCATAACCGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGGC
AGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGG
TCCGAGTCTTACTATAACGGGATTCTACCAACCTGTAA

>SEQ ID NO: 12; DNA sequence of PACE-evolved chPylRS variant, **Split1**. The in-frame, premature stop codon of the split enzyme is underlined, and the position of translational reinitiation, corresponding to Met-107 of chPylRS, is italicized. In the **Split1** variant, bolded codons were reverted back to 'GTT', 'ACC', 'CAT', and 'GCG', respectively.

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
GCACAAGATCAAGCACTATGAGATTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
TGGTTGTGAACAACTCTCGTTCTTGTCTGCGTCCCGCACGTGCATTCCGTTATCATAAATACCGTAAAACC
TGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACCTCCTGACCCGTTCTACCGAAGGCAAAAC
CTCTGTTAAAGTTAAAGCTGTTCTGAGCCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGTG
CGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCT
CCGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCCA
GACTGACAGGCTTGAAGTCTGTTAAACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTCA
GGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAG
GGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAA
TAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTT
TCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAA
CTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGAATAGGCCCATGCTACA
GAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACCTTCTGCCAGATGGGATC
GGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTC
AGATCGTAGGCGATTCTGTCATGGTCTATGGGGATACCTTGATGTAATGCACGGAGACCTGGAACTT
TCCTCTGCAGTAGTCGGACCCATAACCGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGGC
AGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGG
TCCGAGTCTTACTATAACGGGATTCTACCAACCTGTAA

>SEQ ID NO: 13; DNA sequence of PACE-evolved chPylRS variant, **Split2**. The in-frame, premature stop codon of the split enzyme is underlined, and the position of translational reinitiation, corresponding to Met-107 of chPylRS, is italicized. In the **Split2** variant, bolded codons were reverted back to 'GTT', 'ACC', 'CAT', and 'GCG', respectively.

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
GCACAAGATCAAGCACTATGAGATTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
TGGTTGTGAACAACTCTCGTTCTTGTCTGCGTCCCGCACGTGCATTCCGTTATCATAAATACCGTAAAACC
TGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACCTCCTGACCCGTTCTACCGAAGGCAAAAC
CTCTGTTAAAGTTAAAGTTGTTCTGAGCCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGTGC
GCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCTC
CGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCCAG
ACTGACAGGCTTGAAGTCTGTTAAACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTTCA
GGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAG
GAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAAT
AAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTTT
CAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAA
TACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGAATAGGCCCATGCTACAG
AAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACCTTCTGCCAGATGGGATCG
GGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTC
GATCGTAGGCGATTCTGTCATGGTCTATGGGGATACCTTGATGTAATGCACGGAGACCTGGAACTTT
CCTCTGCAGTAGTCGGACCCATAACCGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGGC

AGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGG
TCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

>SEQ ID NO: 14; DNA sequence of PACE-evolved chPylRS variant, **Split3**. The in-frame, premature stop codon of the split enzyme is underlined, and the position of translational reinitiation, corresponding to Met-107 of chPylRS, is italicized. In the **Split3** variant, bolded codons were reverted back to 'GTT', 'ACC', 'CAT', and 'GCG', respectively.

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
GCACAAGATCAAGCACTATGAGATTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
TGGTTGTGAACAACTCTCGTTCTTGTTCGTCCCGCACGTGCATTCCGTTATCATAAATACCGTAAAACC
TGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACCTCCTGACCCGTTCTACCGAAGGCAAAAC
CTCTGTTAAAGTTAAAGTTGTTTTCTGAGCCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGT
GCGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTC
TCCGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCC
AGACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTC
AGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAA
GGGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAA
ATAAAATCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACT
TTCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACA
ACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGTAAATAGGCCCATGCTAC
AGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACCTTCTGCCAGATGGGAT
CGGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTT
AAGATCGTAGGCGATTTCCTGCATGGTCTATGGGGATACCCTTGATGTAATGCACGGAGACCTGGAAC
TTCTCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTGATAAACCTTGATAGGGG
CAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAG
GTCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

>SEQ ID NO: 15; DNA sequence of PACE-evolved chPylRS variant, **Split4**. The in-frame, premature stop codon of the split enzyme is underlined, and the position of translational reinitiation, corresponding to Met-107 of chPylRS, is italicized. In the **Split4** variant, bolded codons were reverted back to 'GTT', 'ACC', 'CAT', and 'GCG', respectively.

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
GCACAAGATCAAGCACTATGAGATTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
TGGTTGTGAACAACTCTCGTTCTTGTTCGTCCCGCACGTGCATTCCGTTATCATAAATACCGTAAAACC
TGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACCTCCTGACCCGTTCTACCGAAGGCTAAAC
CTCTGTTAAAGTTAAAGTTGTTTTCTGAGCCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGTG
CGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTC
CCGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCCA
GACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTC
GGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAG
GGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAA
TAAAATCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTT
TCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAA
CTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTGTAAATAGGCCCATGCTACA
GAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACCTTCTGCCAGATGGGATC
GGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTCA
AGATCGTAGGCGATTTCCTGCATGGTCTATGGGGATACCCTTGATGTAATGCACGGAGACCTGGAAC
TTCTCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTGATAAACCTTGATAGGGG
AGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAG
TCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

>SEQ ID NO: 16; DNA sequence of PACE-evolved chPylRS variant, **Split5**. The in-frame, premature stop codon of the split enzyme is underlined, and the position of translational reinitiation, corresponding to Met-107 of chPylRS, is italicized. In the **Spit5** variant, bolded codons were reverted back to 'GTT', 'ACC', 'CAT', and 'GCG', respectively.

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
GCACAAGATCAAGCACTATGAGATTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
TGGTTGTGAACAACTCTCGTTCTTGTCTGCGCGCACGTGCATTCCGTTATCATAAATACCGTAAAACC
TGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACCTCCTGACCCGTTCTACCGAAGGCAAAAC
CTCTGTATAAGTTAAAGTTGTTTCTGAGCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGTGC
GCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCTC
CGGCGAAATCTACCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCCAG
ACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTCCCTGAATTCGGGCAAGCCTTTTCAG
GGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGG
GAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAAT
AAAATCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACTTT
CAAAACAGATCTTCAGGGTTGACAAGAAGTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAAC
TACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTTGAAATAGGCCCATGCTACAG
AAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACTTCTGCCAGATGGGATCG
GGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTCAA
GATCGTAGGCGATTCTGTCATGGTCTATGGGGATACCCTTGATGTAATGCACGGAGACCTGGAACCTT
CCTCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGGC
AGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGG
TCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

>SEQ ID NO: 17; DNA sequence of PACE-evolved chPylRS variant, **Split6**. This split enzyme contained several in-frame, premature stop codons (underlined) between the frameshift and the position of translational reinitiation, corresponding to Met-107 of chPylRS italicized. In the **Spit6** variant, bp; ded codons were reverted back to 'GTT', 'ACC', 'CAT', and 'GCG', respectively.

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCT
GCACAAGATCAAGCACTATGAGATTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATC
TGGTTGTGAACAACTCTCGTTCTTGTCTGCGCGCACGTGCATTCCGTTATCATAAATACCGTAAAACC
TGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACCTCCTGACCCGTTCTACCGAAGGCAAAAC
CCTCTGTATAAGTTAAAGTTGTTTCTGAGCGCGAAAGTGAAAAAGCGATGCCGAAATCTGTTTCTCGT
GCGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTC
TCCGCGGAAATCTACCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCACTTACGAAGAGCC
AGACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTCCCTGAATTCGGGCAAGCCTTTC
AGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAA
GGGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAA
ATAAAATCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACCGAACT
TTCAAAACAGATCTTCAGGGTTGACAAGAAGTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACA
ACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAAAATTTTTGAAATAGGCCCATGCTAC
AGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACTTCTGCCAGATGGGAT
CGGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTCA
AAGATCGTAGGCGATTCTGTCATGGTCTATGGGGATACCCTTGATGTAATGCACGGAGACCTGGAACCT
TTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTGATAAACCTGGATAGGGG
CAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAG
GTCCGAGTCTTACTATAACGGGATTTCTACCAACCTGTAA

>SEQ ID NO: 18; DNA sequence of **p-NFRS** (amino acid sequence SEQ ID NO: 39).

ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGCGAGGAAGAGTTAAGAGAG
GTTTTAAAAAAGATGAAAAATCTGCTCTGATAGGTTTTGAACCAAGTGGTAAAATACATTTAGGGCA
TTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTGGATTTGATATAATTATATTGTTGGCTG

ATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATAACAA
 AAAAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATATGTTTATGGAAGTTCGTTCCAGCTTGATAAGG
 ATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGAGCAAGAAGGAGTATGGA
 ACTTATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTAAATC
 CTCTTAATTATGAGGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATACACATGTTAGC
 AAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTACACAACCCTGTCTTAACGGGTTTGGATGGAGAAG
 GAAAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG
 ATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGATAGCTAAATACT
 TCCTTGAATATCCTTTAACCATAAAAAAGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTAT
 GAGGAGTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGCGCTTAAAAAATGCTGTAGCTG
 AAGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGATTATAA

>SEQ ID NO: 19; DNA sequence of *p-IFRS* (amino acid sequence SEQ ID NO: 40).

ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGCGAGGAAGAGTTAAGAGAG
 GTTTTAAAAAAGATGAAAAATCTGCTCTGATAGGTTTTGAACCAAGTGGTAAAATACATTTAGGGCA
 TTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTGGATTTGATATAATTATATTGTTGGCTG
 ATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATAACAA
 AAAAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATATGTTTATGGAAGTTCGTTCCAGCTTGATAAGG
 ATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGAGCAAGAAGGAGTATGGA
 ACTTATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTAAATC
 CTCTTCATTATGAGGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATACACATGTTAGC
 AAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTACACAACCCTGTCTTAACGGGTTTGGATGGAGAAG
 GAAAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG
 ATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGATAGCTAAATACT
 TCCTTGAATATCCTTTAACCATAAAAAAGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTAT
 GAGGAGTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGCGCTTAAAAAATGCTGTAGCTG
 AAGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGATTATAA

>SEQ ID NO: 20; Amino acid sequence of MbPylRS

MDKKPLDVLISATGLWMSRTGTLHKIKHYEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKTCK
 RCRVSEDEINFLTRSTEGKTSVKVKVVSAPKVKKAMPKSVSRAPKPLENPVSAKASTDTSRSPSPAKST
 PNSPVPTSAPAPSLTRSQDRVEALLSPEDKISLNIAPFRELESELVTRRKNDQRLYTDNDREDYLGKLERD
 ITKFFVDRDFLEIKSPILPAEYVERMGINDTELSKQIFRVDKNLCLRPMLAPTLNYLRLKLDRLPDPIKIFE
 VGPCYRKESDGKEHLEEFMTVNFCQMSGCTRENLESLIKEFLDYLEIDFEIVGDSCMVYGDTLDIMHGDL
 ELSSAVVGPVPLDREWIDKWPWIGAGFGLERLLKVMHGFKNIKRASRSSESYNGISTNL

>SEQ ID NO: 21; Amino acid sequence of MmPylRS

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCK
 RCRVSEDEDLNFKLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPV
 STQESVSVPASVSTSISSISTGATASALVKGNTNPITSMSPVQASAPALTKSQTDRLLEVLLNPKDEISLNSG
 KPFRELESELLSRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQI
 FRVDKNFCLRPMLAPNLNYLRLKLDRLPDPIKIFEIGPCYRKESDGKEHLEEFMTNFCQMSGCTRENLE
 ESIITDFLNHLGIDFKIVGDSCMVYGDTLDMHGDLLESSAVVGPIPLDREWIDKWPWIGAGFGLERLLKVK
 HDFKNIKRAARSESYNGISTNL

>SEQ ID NO: 22; Amino acid sequence of p-NFRS

MDEFEMIKRNTSEIIEEELREVLKKDEKSALIGFEPGSKIHLGHYLQIKKMIDLQNAFGDIILLADLHAYLN
 QKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSSFLDKDYTLNVYRLALKTTLKRARRSMELIAREDENP
 KVAEVIYPIQMVNPLNYEGVDVAVGGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIA
 VDDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRPEKFGDLTVNSYEELESFLKKNELHPMR
 LKNAVAEELIKILEPIRKRL

SEQ ID NO: 23; Amino acid sequence p-IFRS

MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPGKIHLGHYLQIKKMIDLQNAGFDIILLADLHAYLN
QKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSSFQLDKDYTLNVYRLALKTTTLKRARRSMELIAREDENP
KVAEVIYPIMQVNPLHYEGVDVAVGGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIA
VDDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESFLFKNKELHPMR
LKNAVAEELIKILEPIRKRL

SEQ ID NO: 24; Amino acid sequence of *M. jannaschii* TyrRS

MDEFEMIKRNTSEIISEEELREVLKKDEKSAYIGFEPGKIHLGHYLQIKKMIDLQNAGFDIILLADLHAYLN
QKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTTLKRARRSMELIAREDENP
KVAEVIYPIMQVNDIHYLGVDVAVGGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIA
VDDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESFLFKNKELHPMD
LKNAVAEELIKILEPIRKRL

CLAIMS

What is claimed is:

1. A tRNA synthetase protein variant comprising an amino acid sequence that is at least 90% identical to SEQ ID NO: 20 or SEQ ID NO: 21, and includes at least one mutation at a position selected from V31, T56, H62, and A100.
2. The tRNA synthetase protein variant of claim 1, wherein the at least one mutation is V31I, T56P, H62Y, A100E, or A100S.
3. The tRNA synthetase protein variant of claim 1 or 2, comprising mutations at V31, T56, H62, and A100.
4. The tRNA synthetase protein variant of claim 3, wherein the mutations are V31I, T56P, H62Y, or A100E.
5. The tRNA synthetase protein variant of any one of claims 1 to 3, wherein the nucleic acid sequence encoding the amino acid sequence comprises one or more premature stop codons.
6. A pyrrolysyl-tRNA synthetase (PylRS) protein variant having an N-terminal domain amino acid substitution present in at least one of the following positions: V31, T56, H62, or A100.
7. The PylRS protein variant of claim 6, wherein the amino acid substitution is V31I, T56P, H62Y, or A100E.
8. The PylRS protein variant of claim 6 or 7, comprising the following amino acid substitutions: V31I, T56P, H62Y, and A100E.
9. A chimeric pyrrolysyl-tRNA synthetase (PylRS) protein variant comprising:

(i) a first portion comprising amino acid residues 1-149 of *Methanosarcina barkeri* PylRS (SEQ ID NO: 20); and

(ii) a second portion comprising amino acid residues 185-454 of *Methanosarcina mazei* PylRS (SEQ ID NO: 21), wherein the first portion or the second portion comprises at least one of the amino acid substitutions set forth in Tables 2-6.

10. The chimeric PylRS variant of claim 9, wherein the variant protein comprises an amino acid substitution in at least one of the following positions: V31, T56, H62, or A100.

11. The chimeric PylRS variant of claim 10, wherein the amino acid substitution is V31I, T56P, H62Y, A100E, or any combination thereof.

12. A tRNA synthetase protein variant comprising an amino acid sequence that is at least 90% identical to SEQ ID NO: 24, and includes at least one mutation at a position selected from L69 and V235.

13. The tRNA synthetase protein variant of claim 12, wherein the at least one mutation is L69F or V235I.

14. The tRNA synthetase protein variant of claim 12 or 13, comprising mutations at L69 and V235.

15. The tRNA synthetase protein variant of claim 14, wherein the mutations are L69F and V235I.

16. A tyrosyl-tRNA synthetase (TyrRS) protein variant having an amino acid substitution present in at least one of the following positions: L69 or V235.

17. The TyrRS protein variant of claim 16, wherein the amino acid substitution is L69F, V235I, or L69F and V235I.

18. An isolated nucleic acid comprising the sequence represented by any one of SEQ ID NO: 5-19.
19. A protein encoded by the isolated nucleic acid of claim 18.
20. A method for aminoacylation of a tRNA, the method comprising, contacting a tRNA encoding an amber codon with a tRNA synthetase protein variant of any one of claims 1 to 17 in the presence of a non-canonical amino acid.
21. The method of claim 20, wherein the non-canonical amino acid is a pyrrolysine or a *p*-iodo-L-phenylalanine.
22. The method of claim 20 or 21, wherein the tRNA is contacted with the tRNA synthetase inside a cell.
23. A method for incorporating a non-canonical amino acid into a peptide, the method comprising expressing in a cell:
- (i) an mRNA transcript, wherein the transcript comprises an amber codon at a position in which a non-canonical amino acid (ncAA) is to be translated;
 - (ii) a tRNA capable of incorporating the ncAA;
 - (iii) a tRNA synthetase protein variant of any one of claims 1 to 17,
- wherein (i), (ii), and (iii) are expressed in the presence of the ncAA.
24. The method of claim 23, wherein the non-canonical amino acid is a pyrrolysine or a *p*-iodo-L-phenylalanine.
25. The method of claim 23 or 24, wherein the tRNA synthetase is orthogonal to the cell, optionally wherein the cell is an *E. coli* cell.

1/41

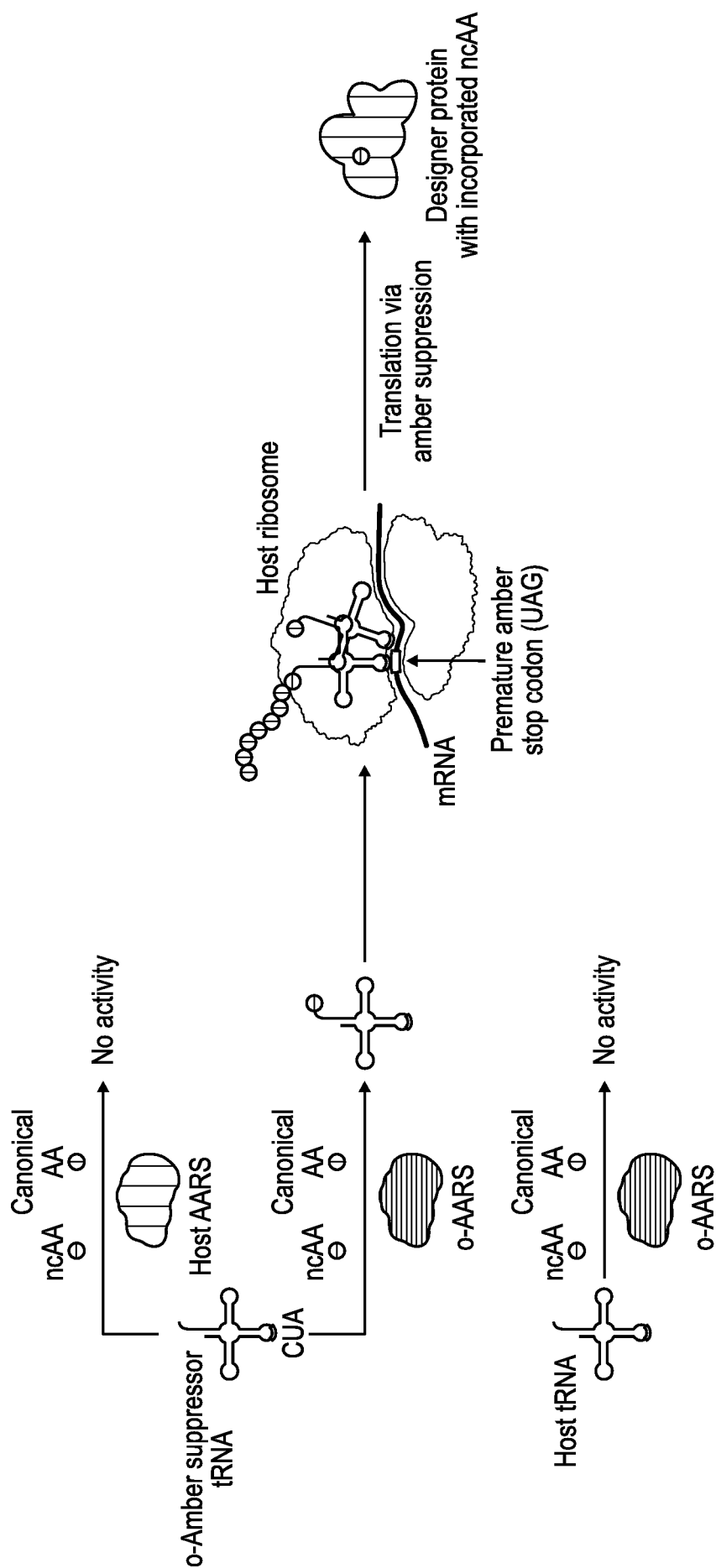


FIG. 1

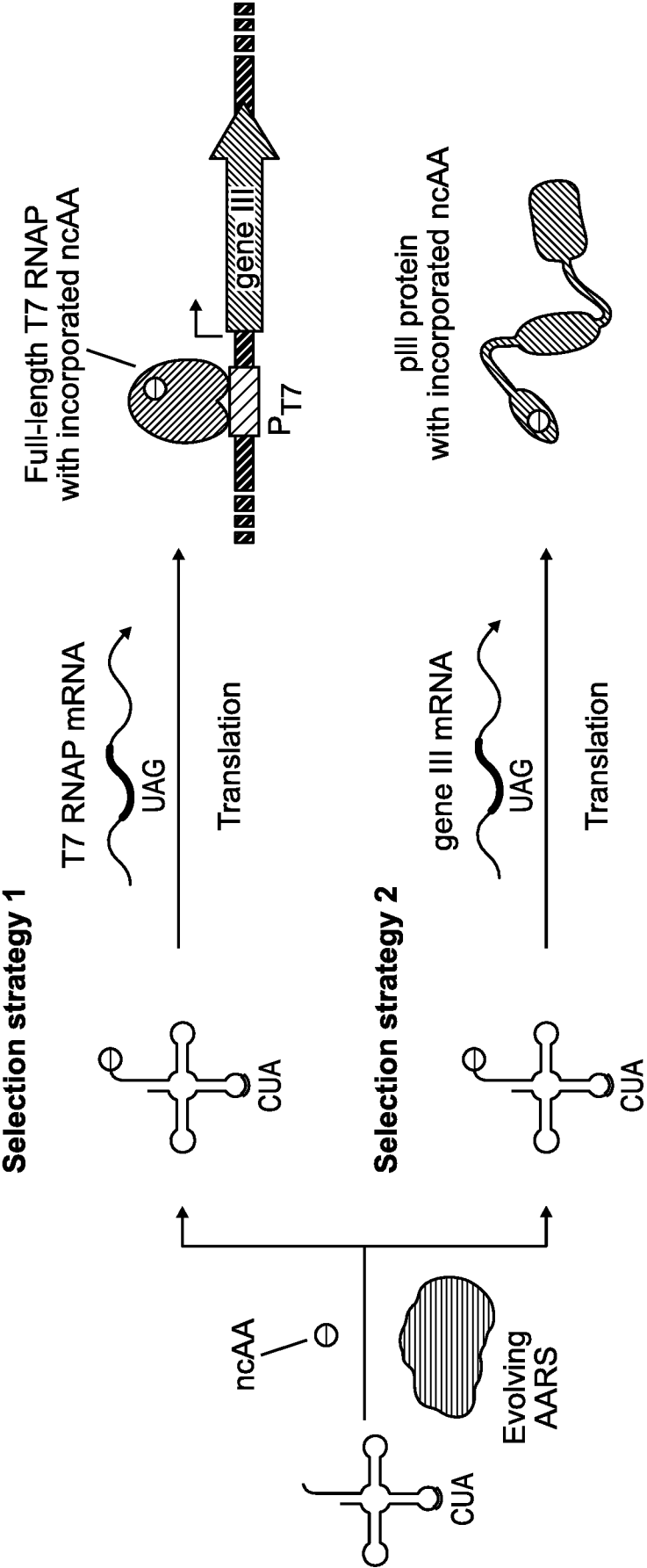


FIG. 2A

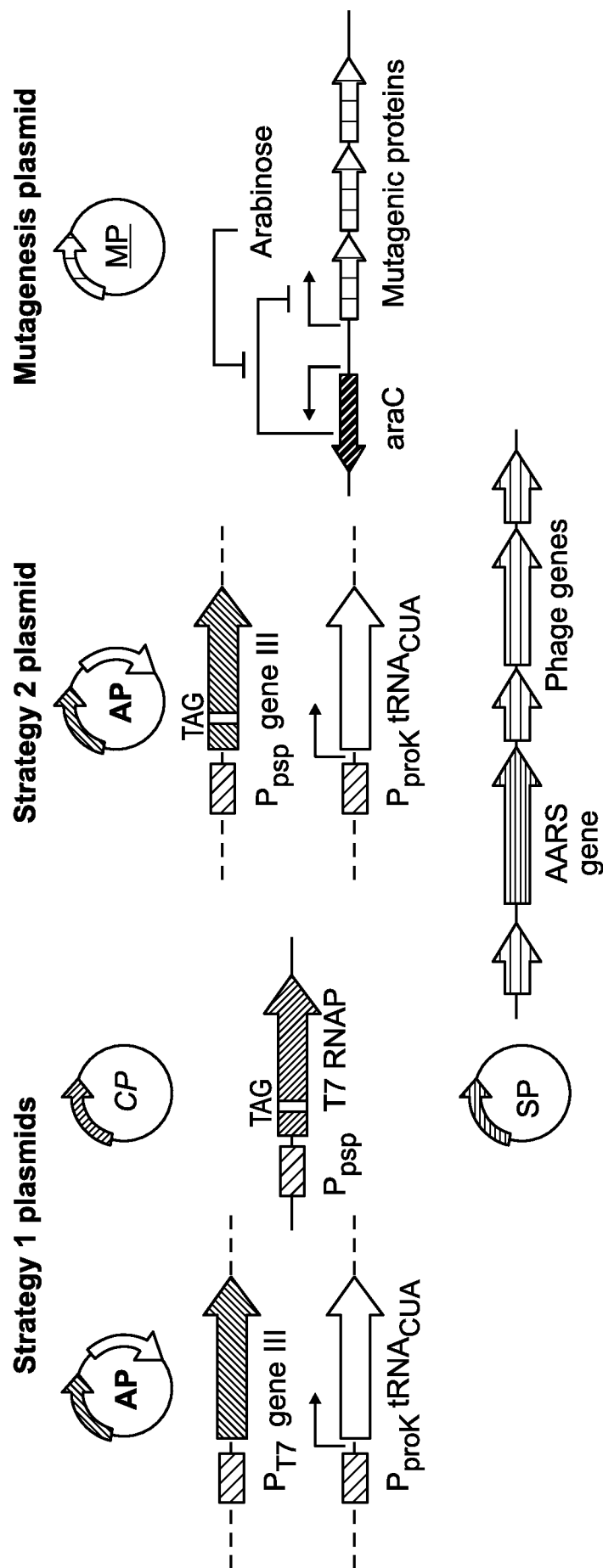


FIG. 2B

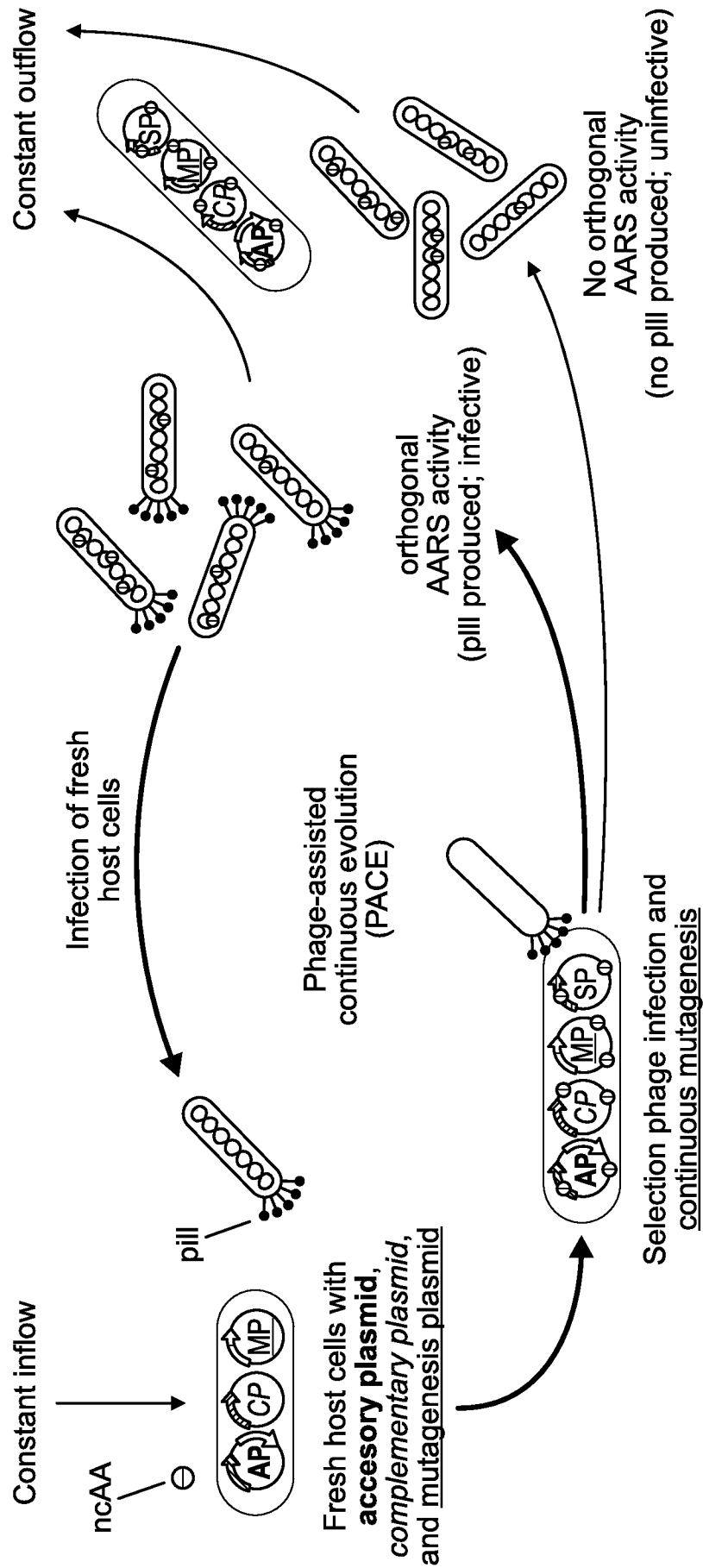


FIG. 2C

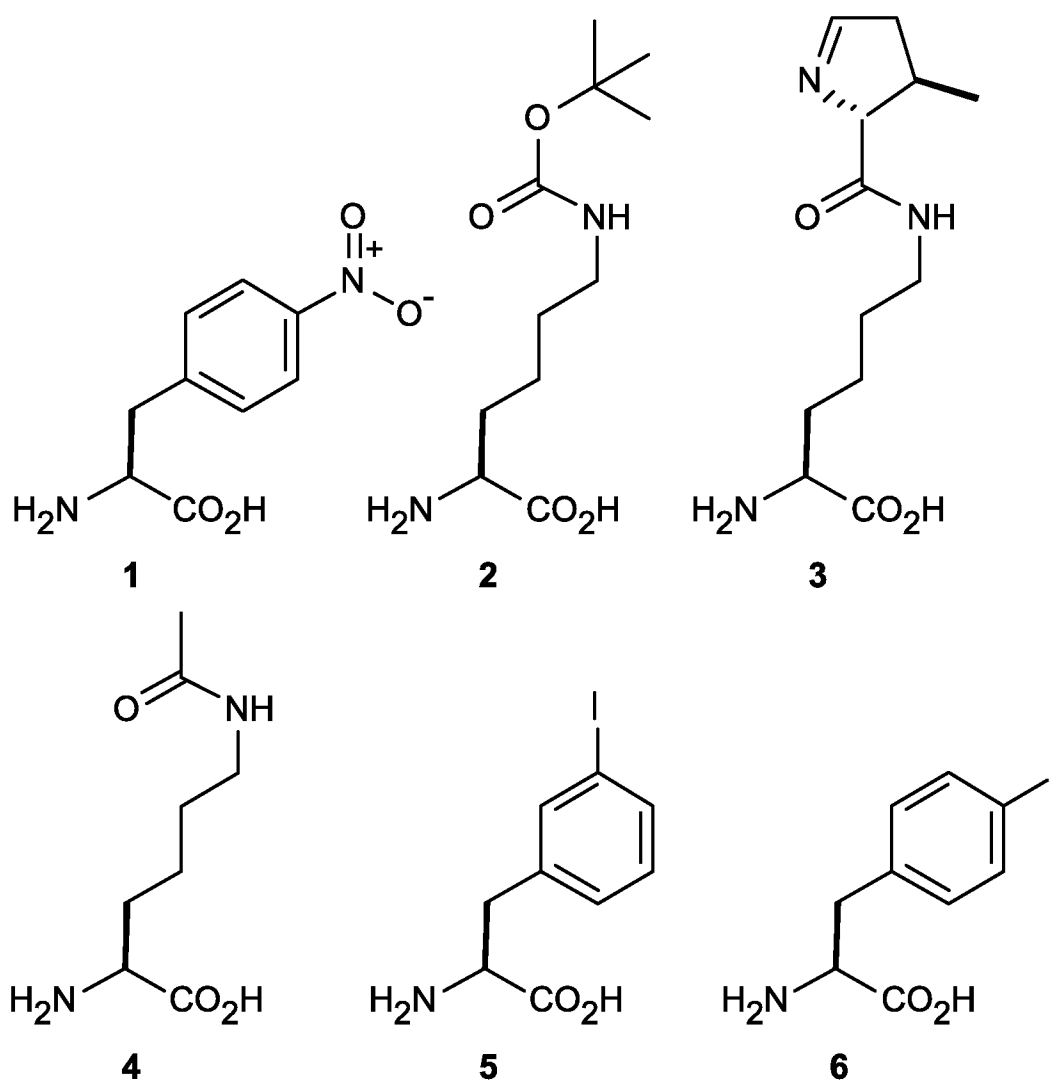


FIG. 3

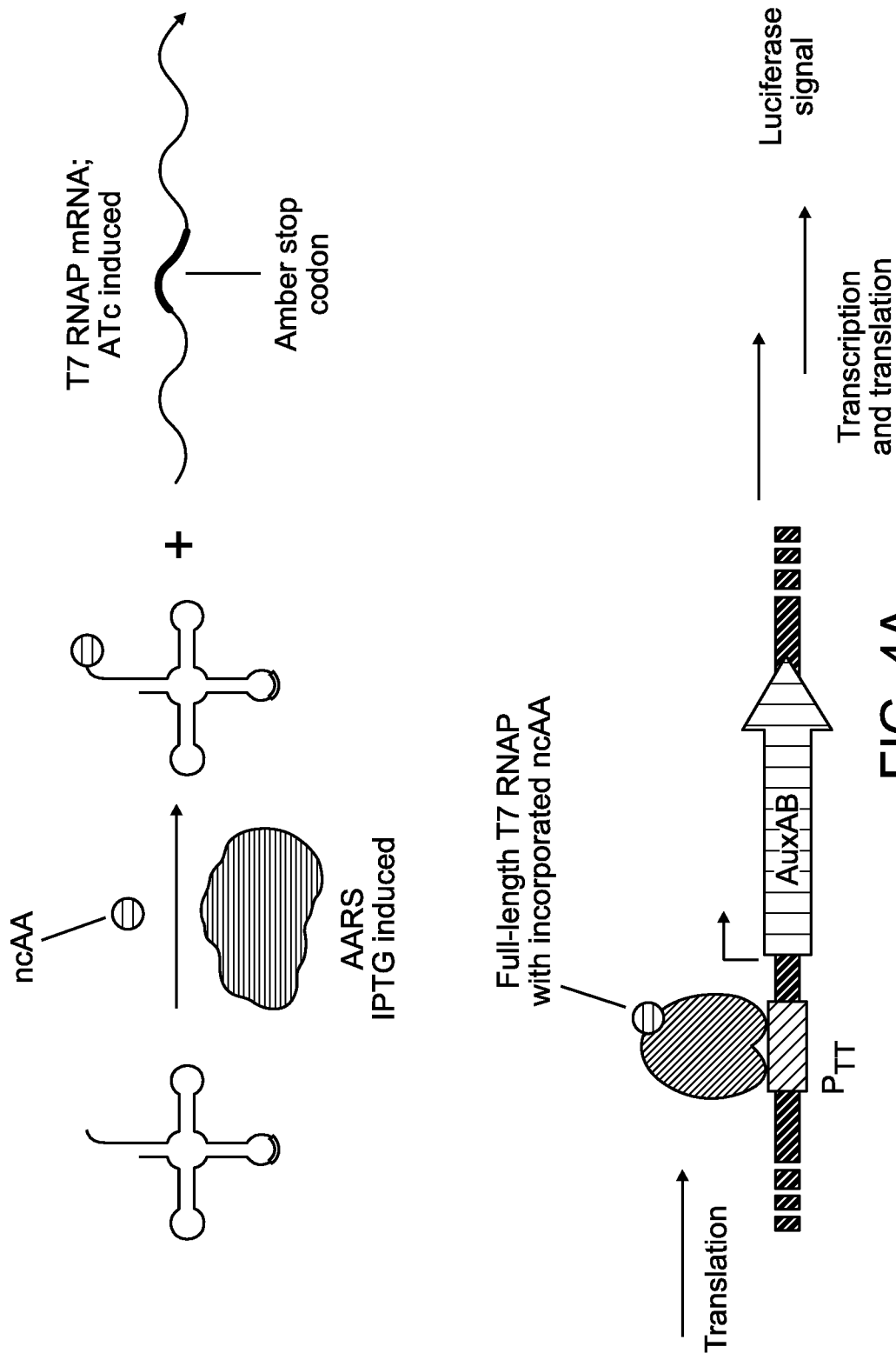


FIG. 4A

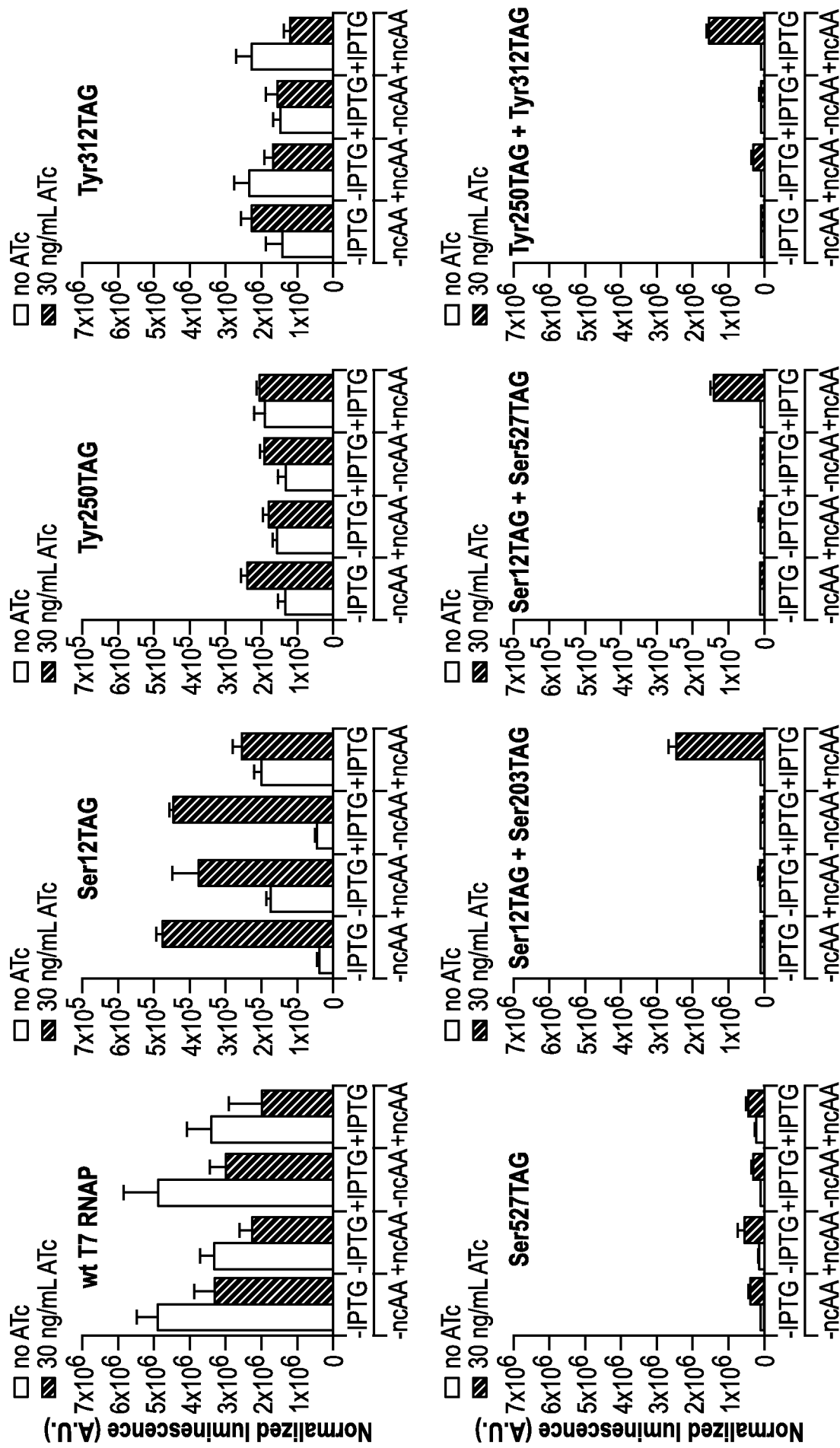


FIG. 4B

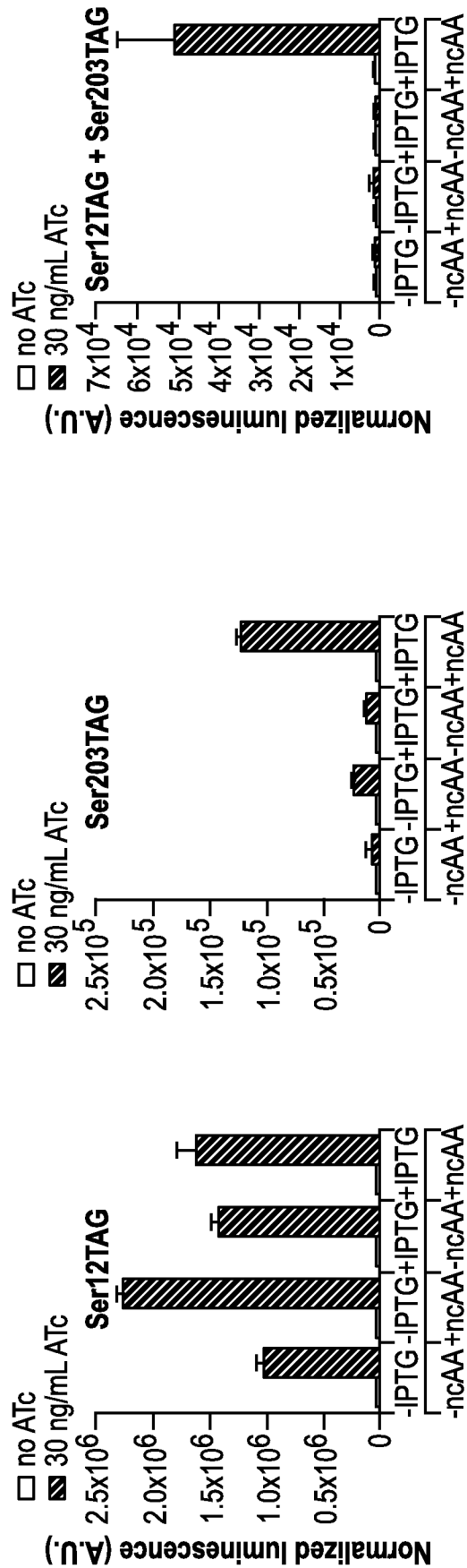


FIG. 4C

9/41

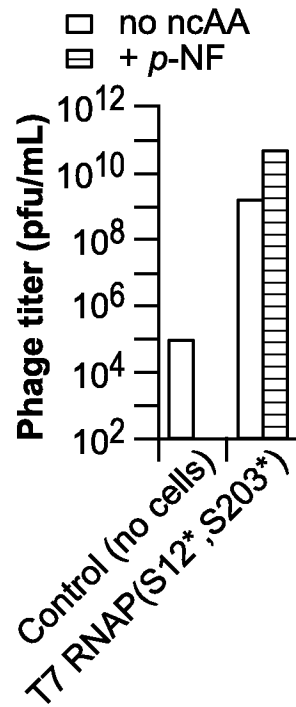


FIG. 5A

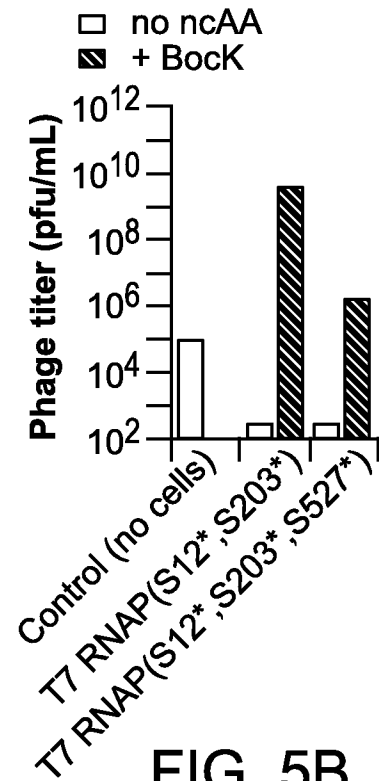


FIG. 5B

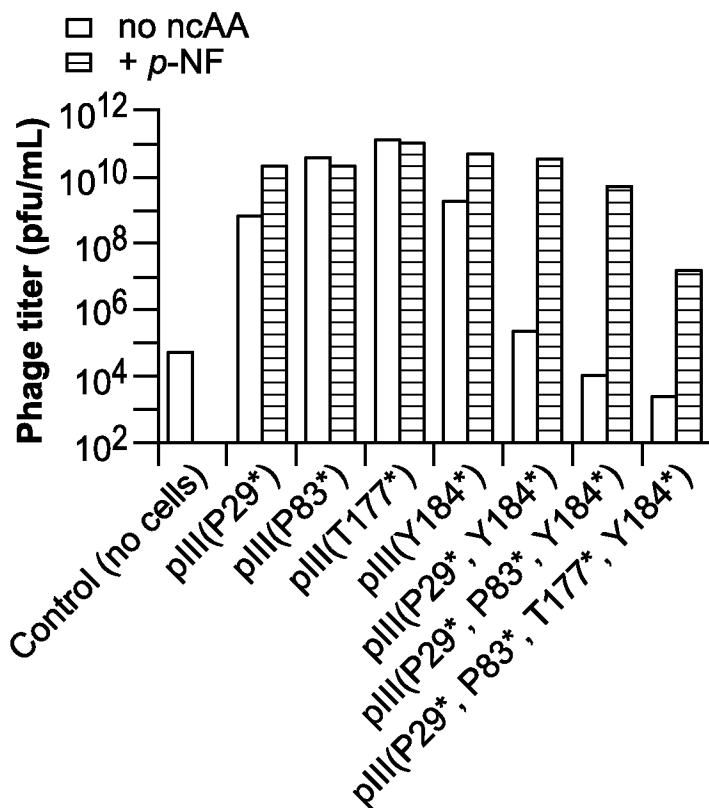


FIG. 5C

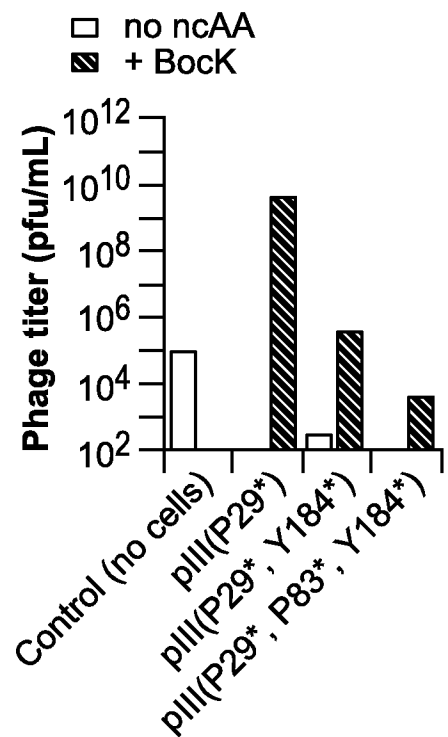


FIG. 5D

10/41

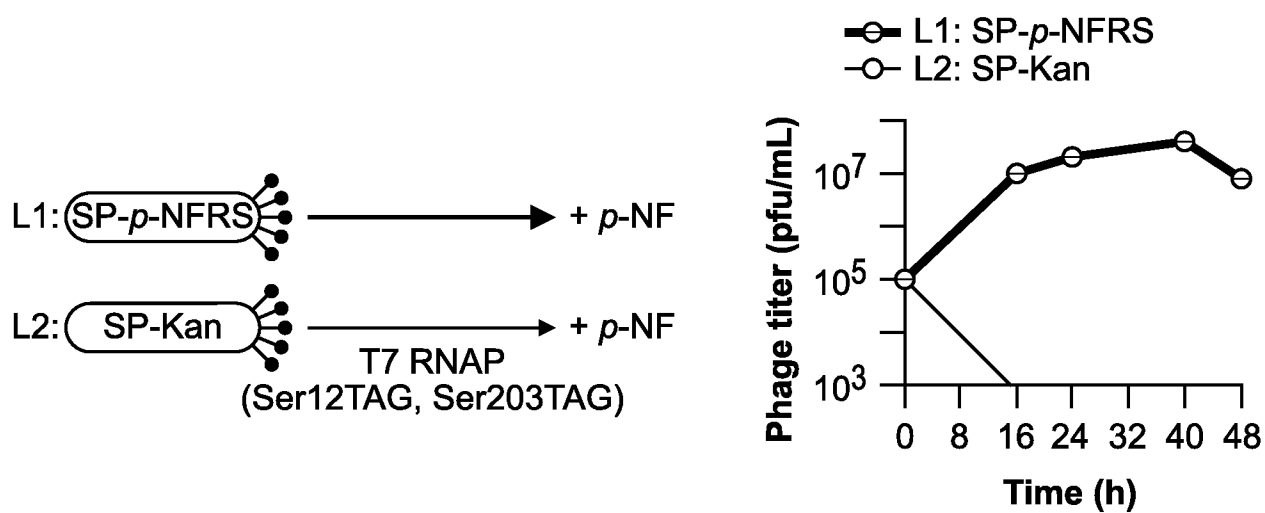


FIG. 6A

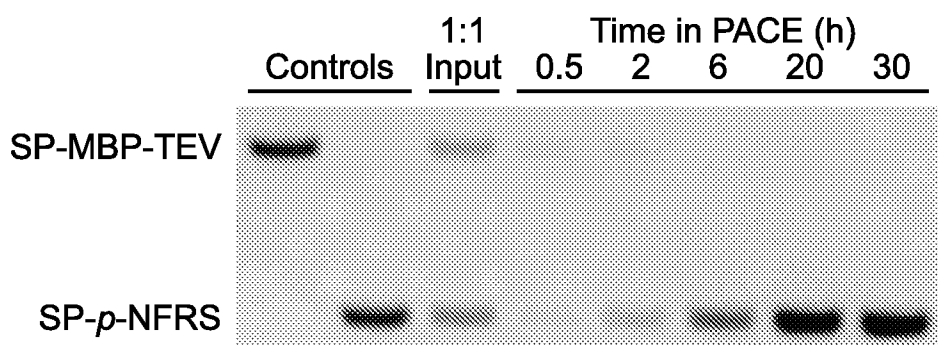
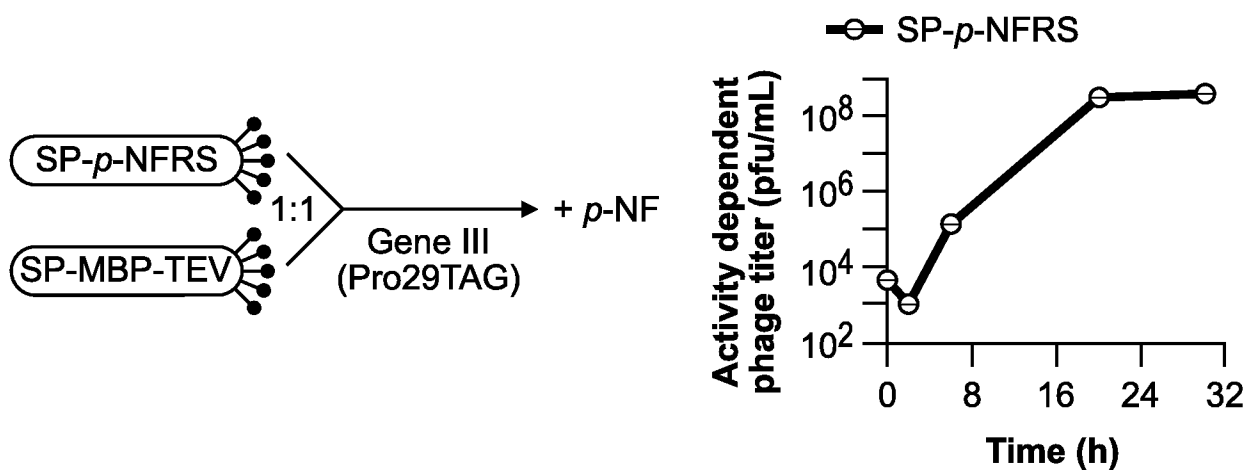


FIG. 6B

11/41

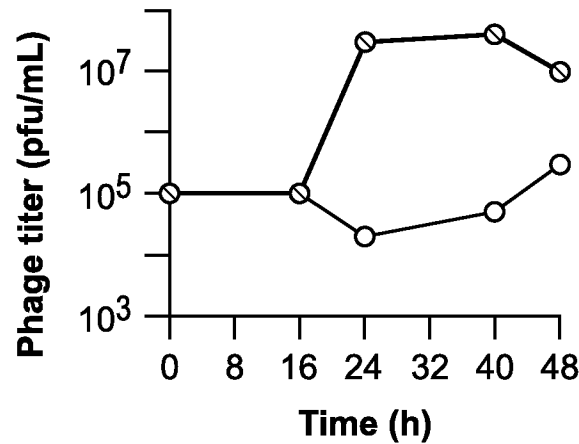
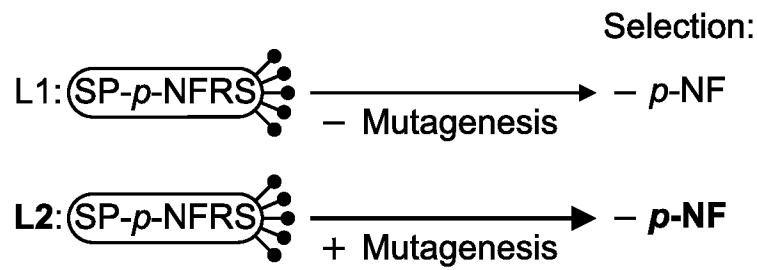


FIG. 7A

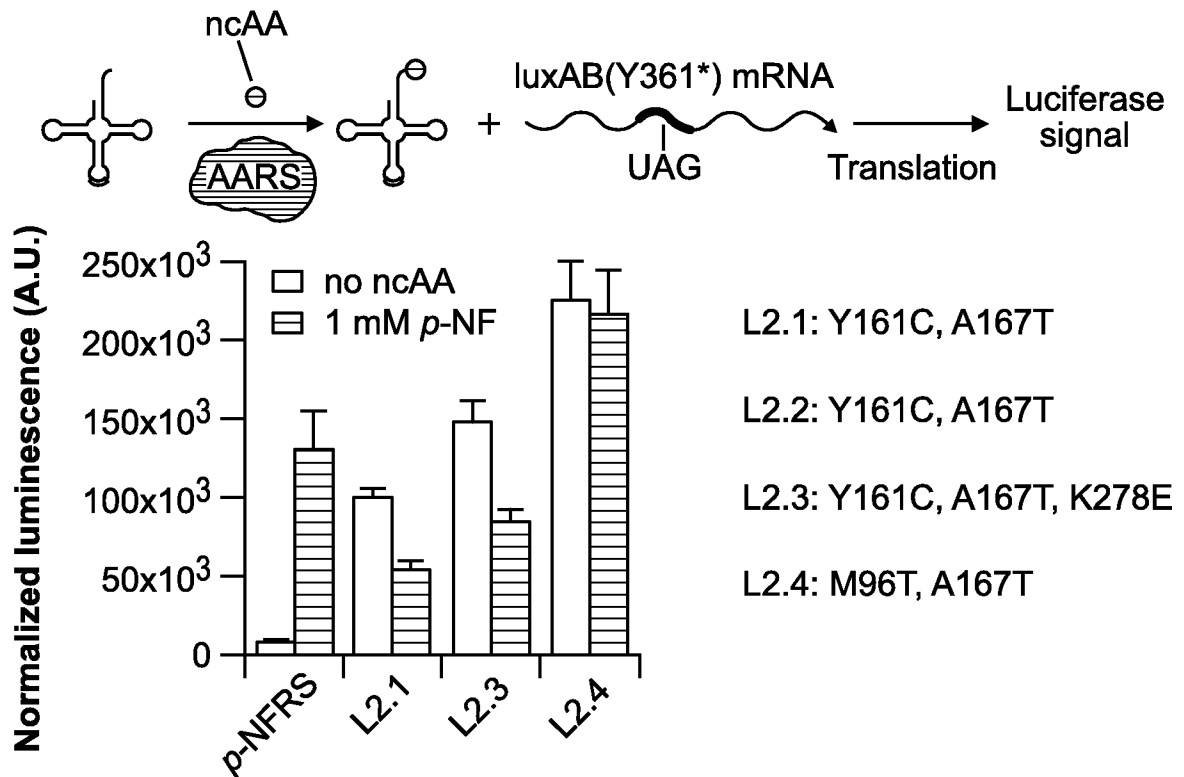


FIG. 7B

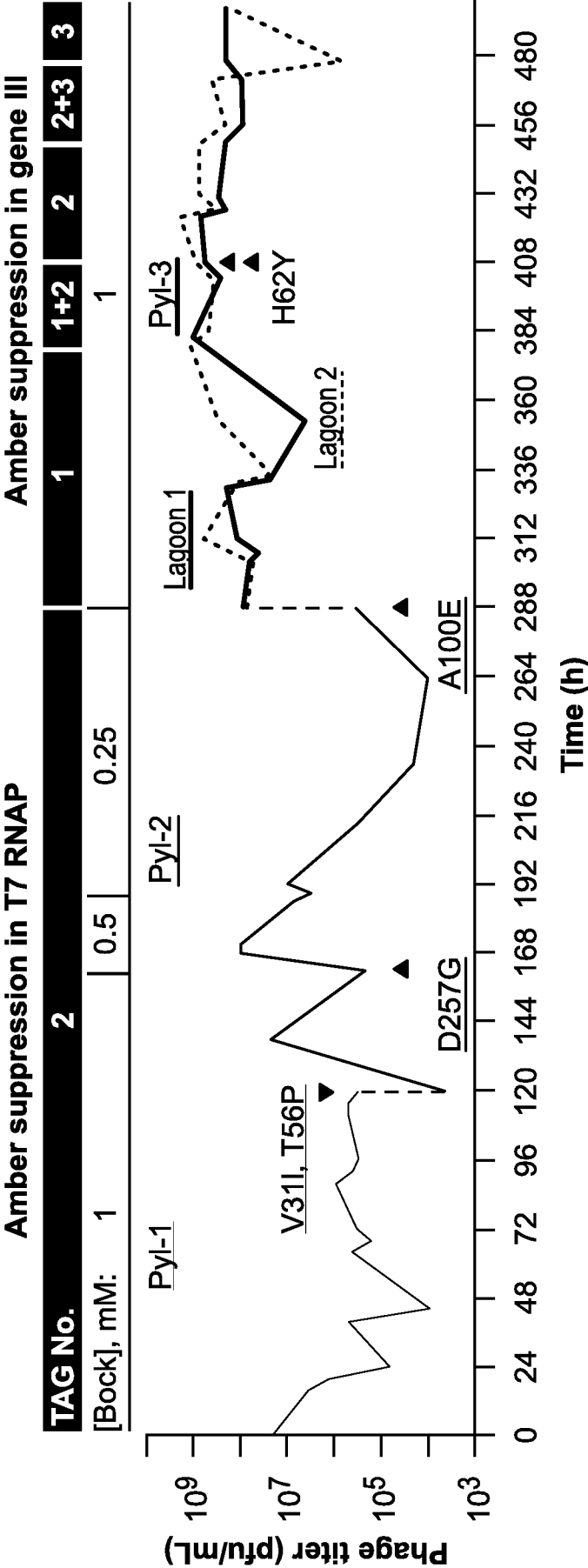


FIG. 8A

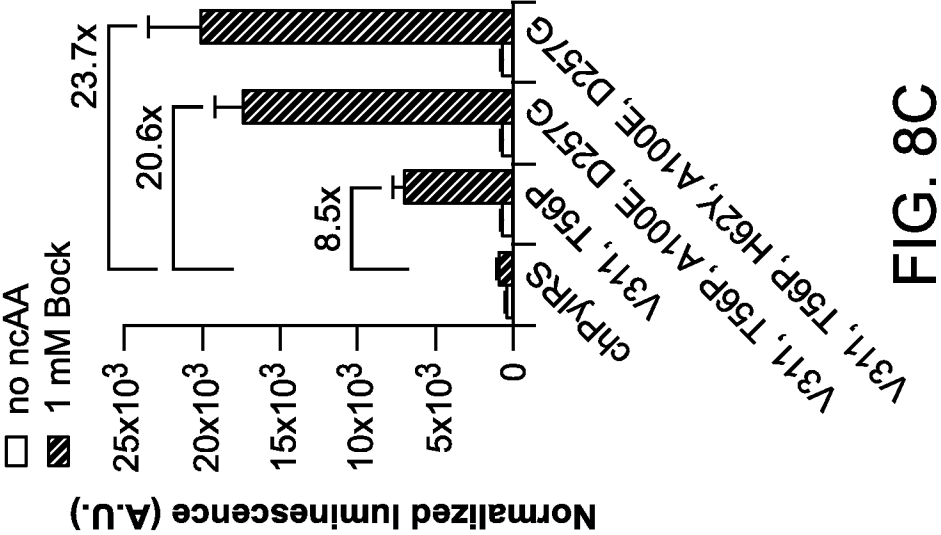


FIG. 8C

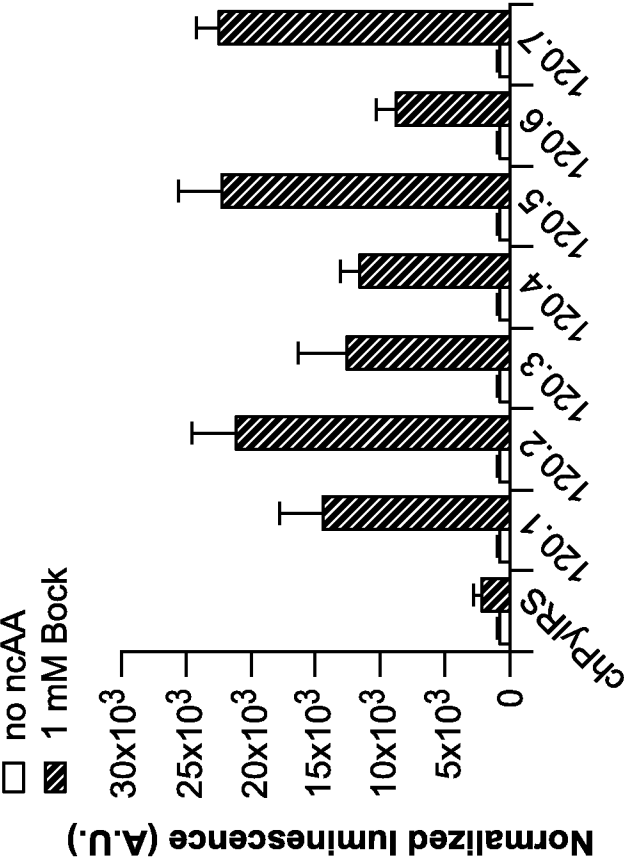


FIG. 8B

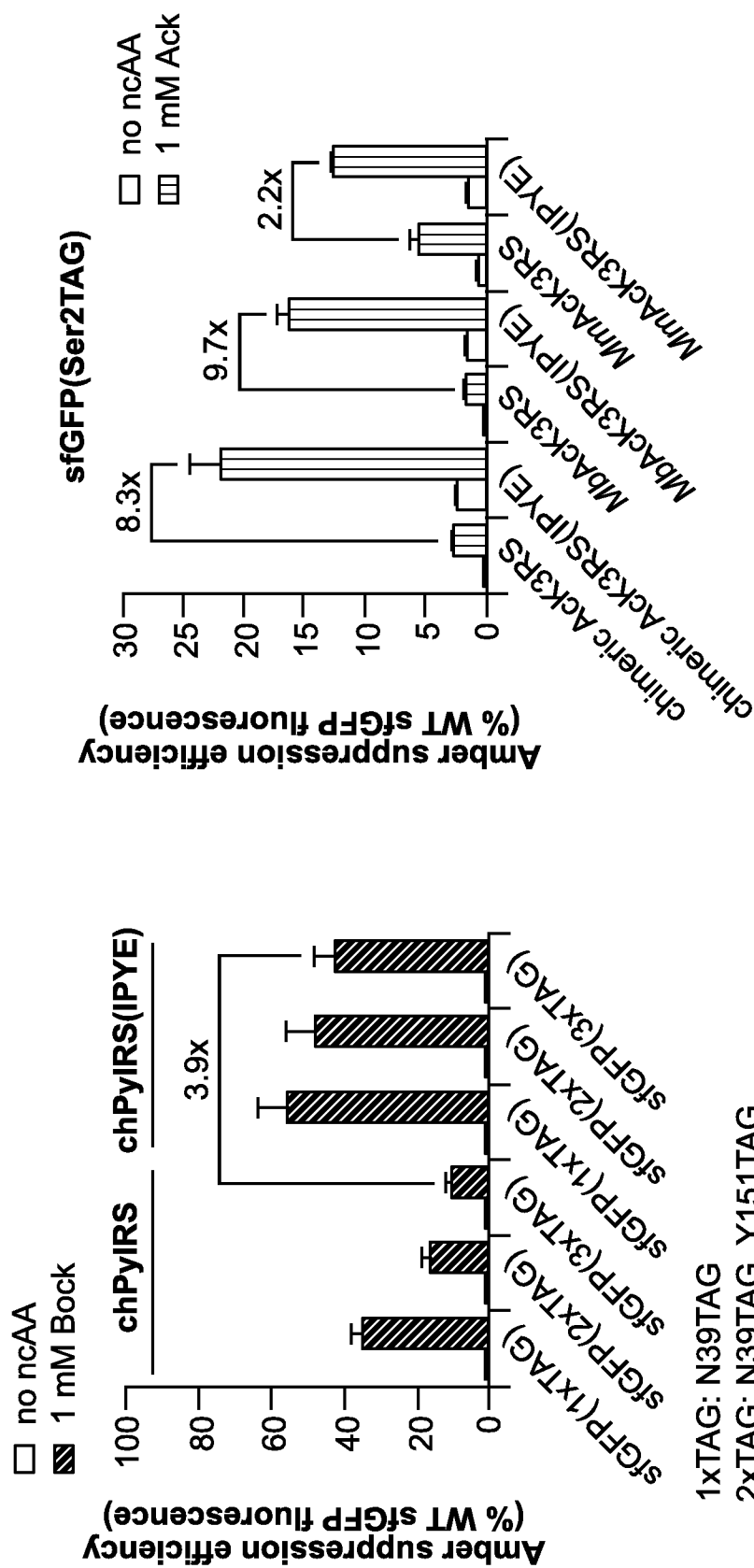


FIG. 8D

FIG. 8E

15/41

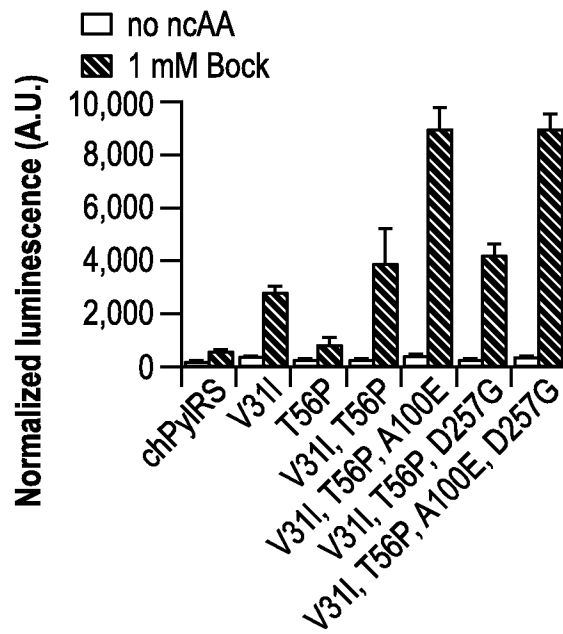


FIG. 9A

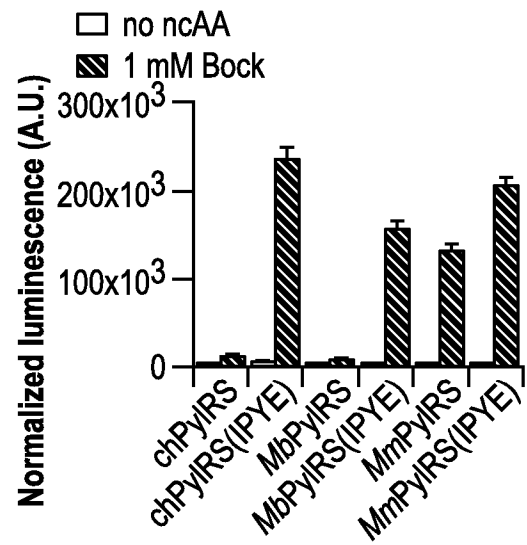


FIG. 9B

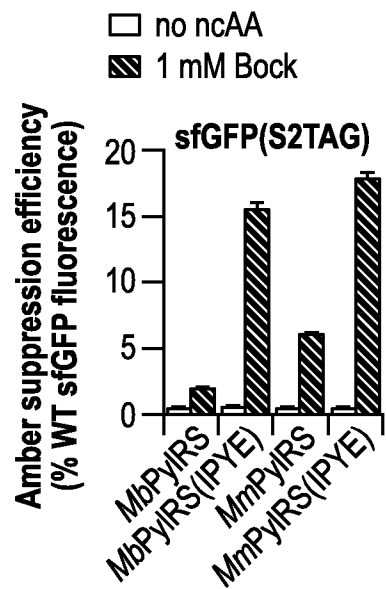


FIG. 9C

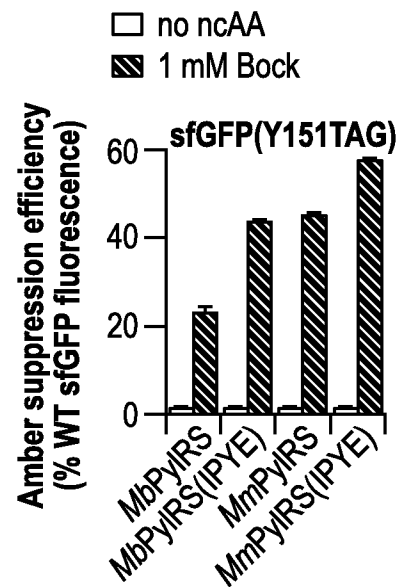


FIG. 9D

16/41

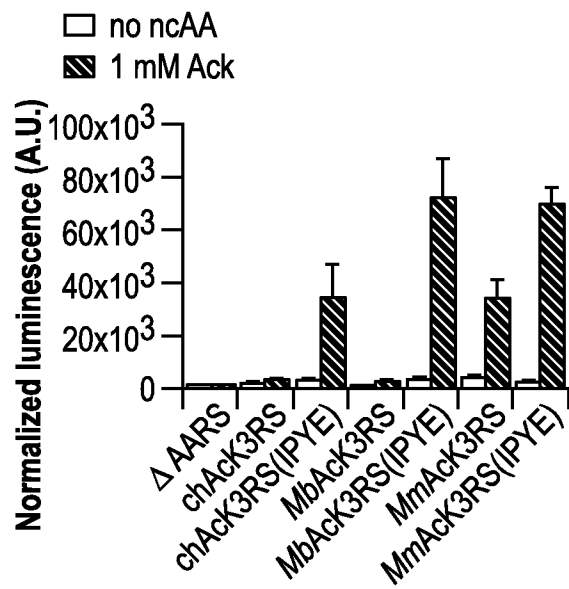


FIG. 9E

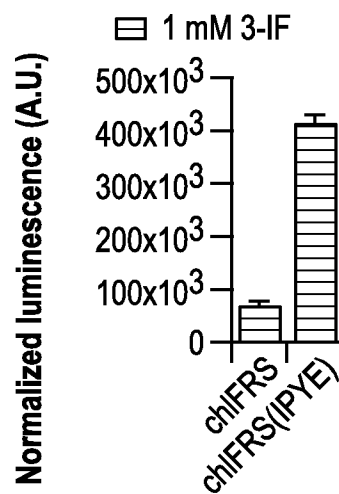
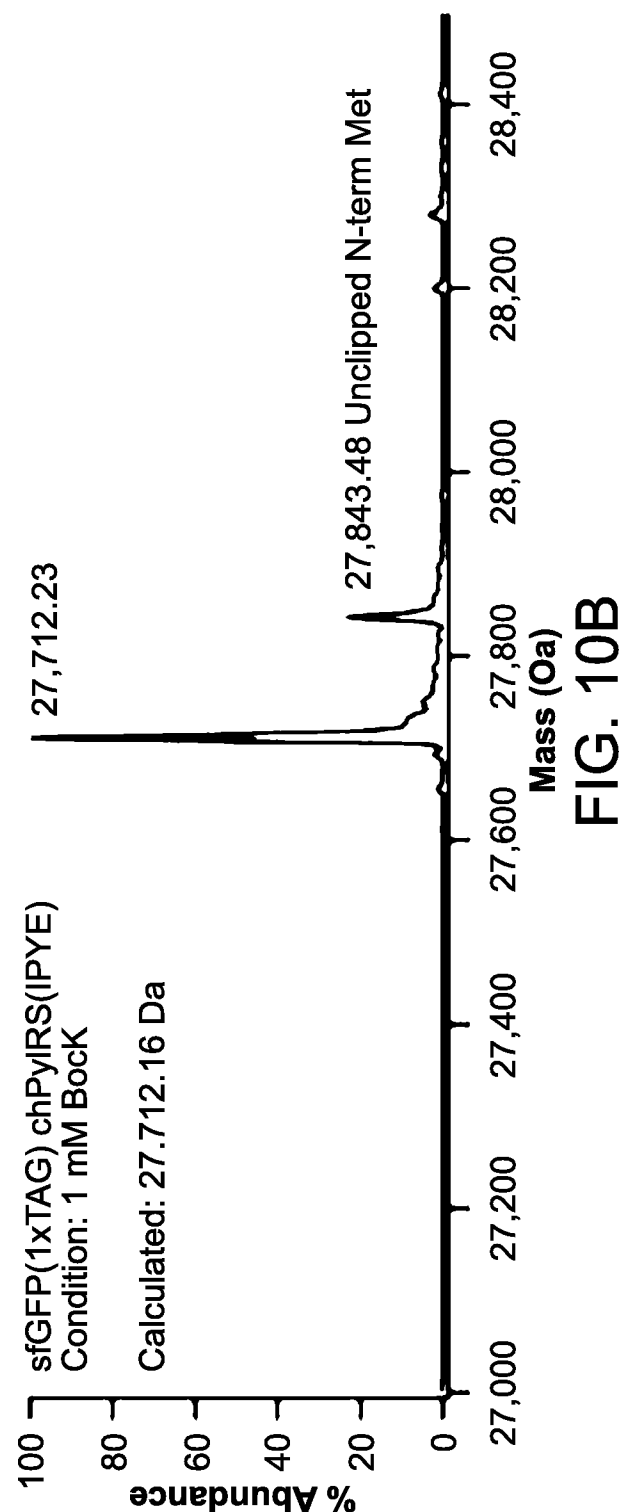
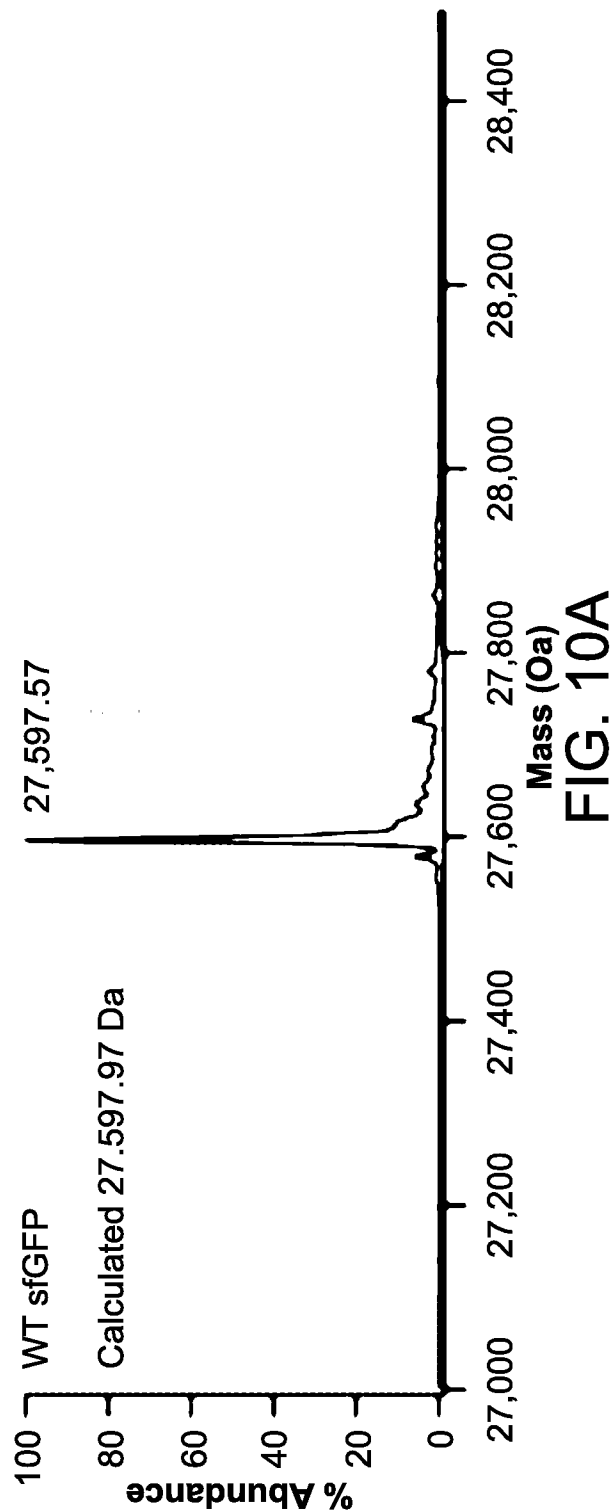
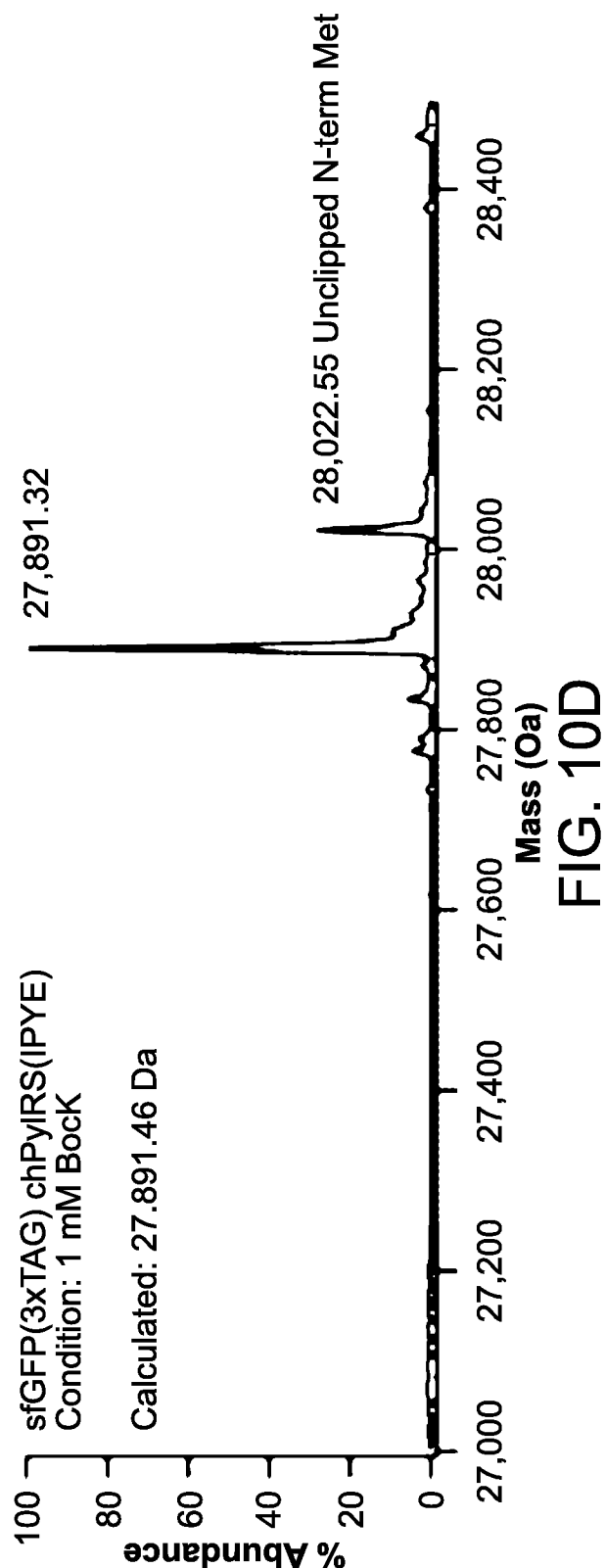
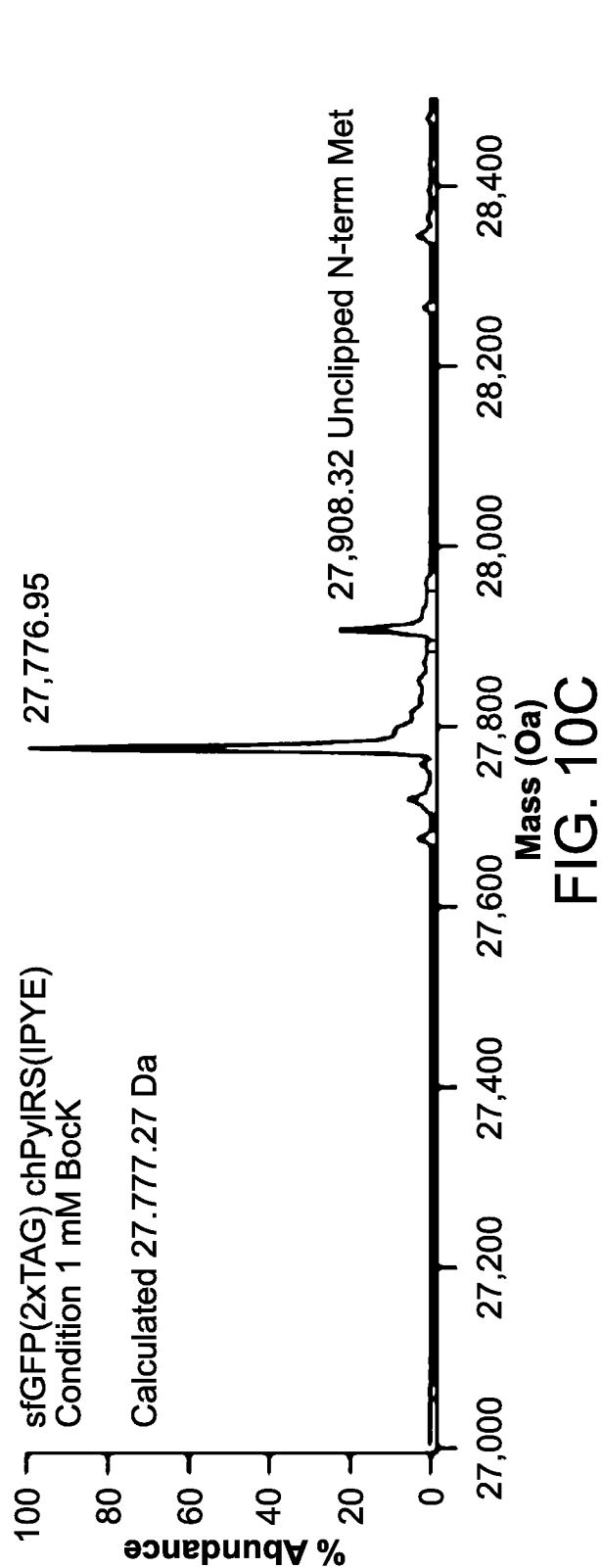


FIG. 9F





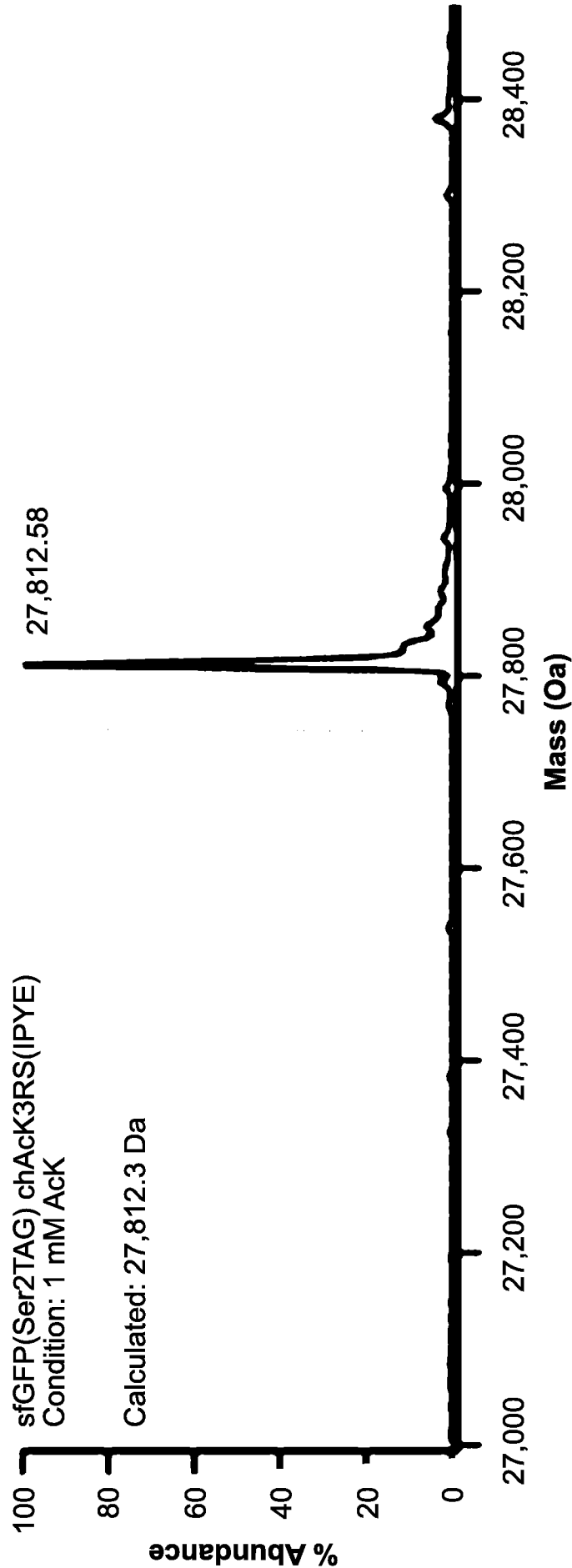


FIG. 11

20/41

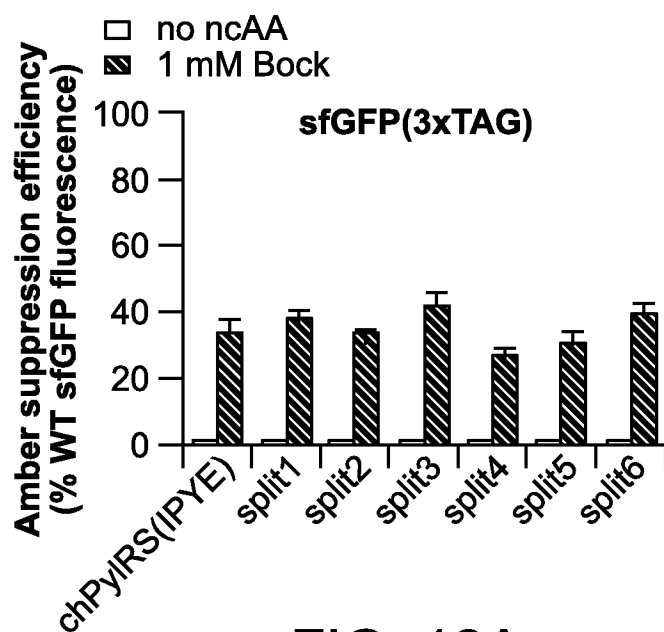


FIG. 12A

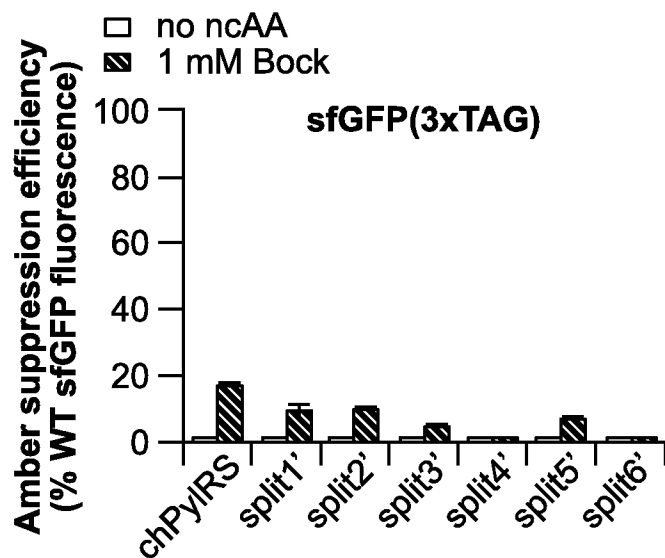


FIG. 12B

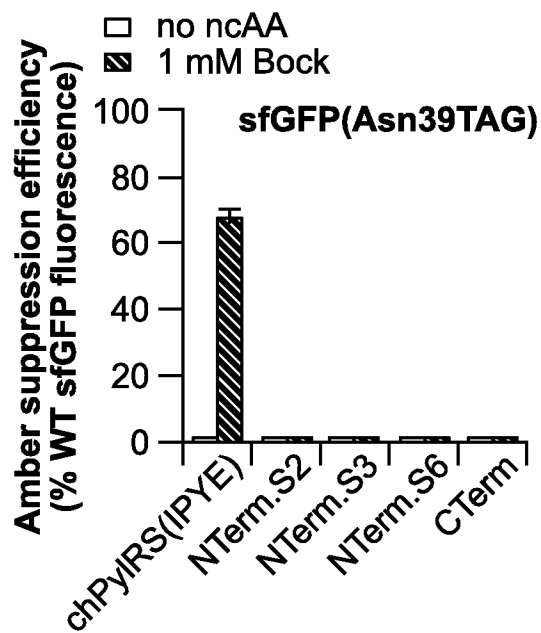


FIG. 12C

21/41

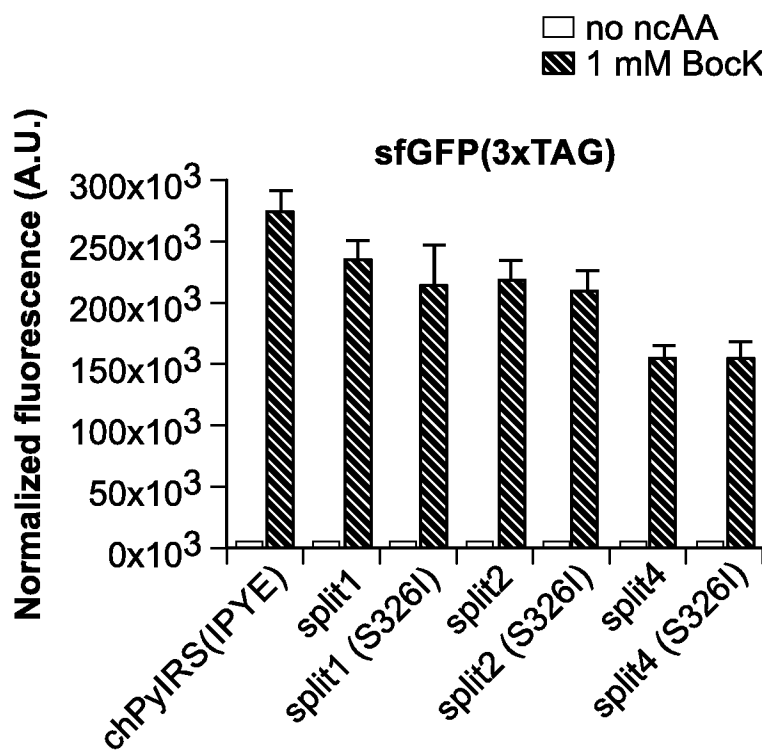


FIG. 13

22/41

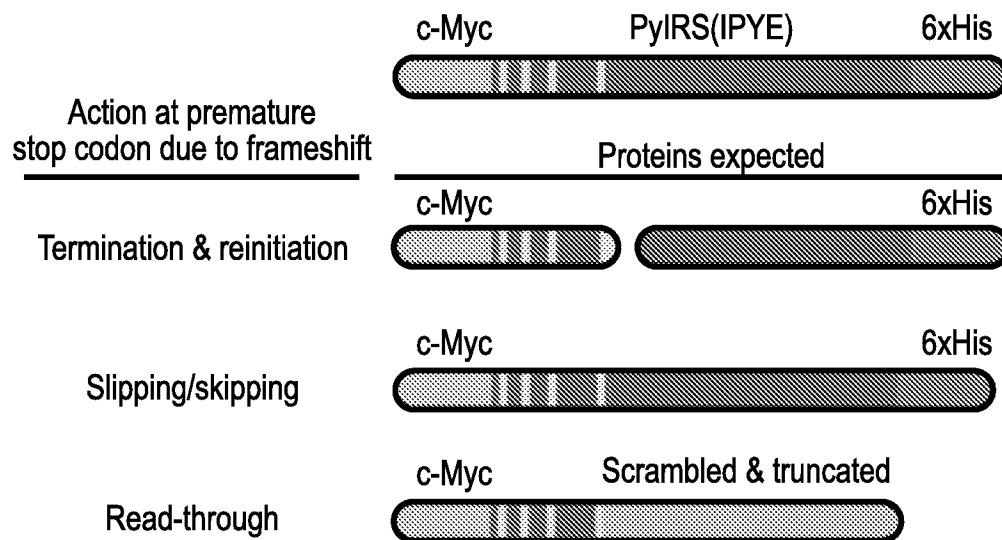


FIG. 14A

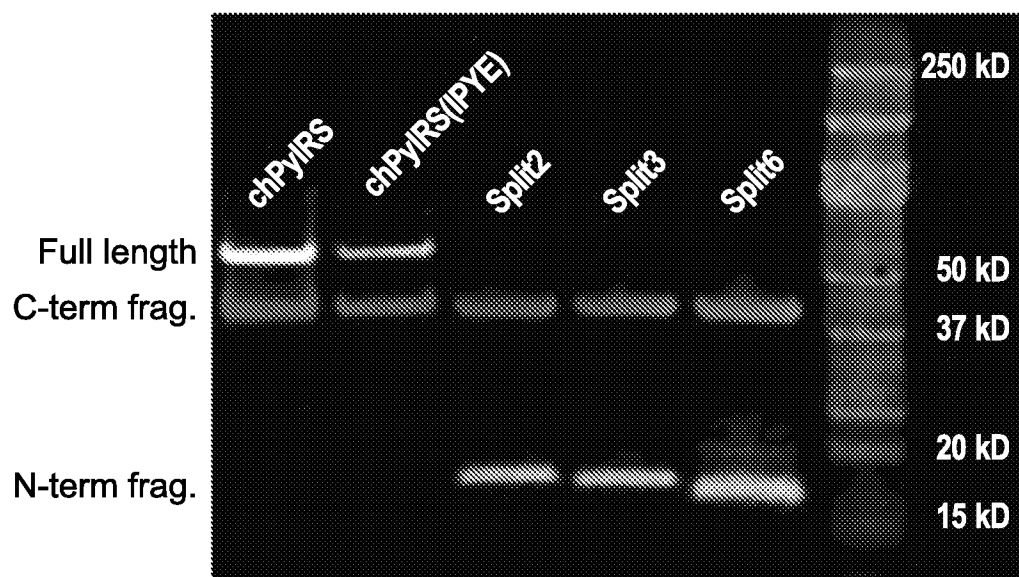


FIG. 14B

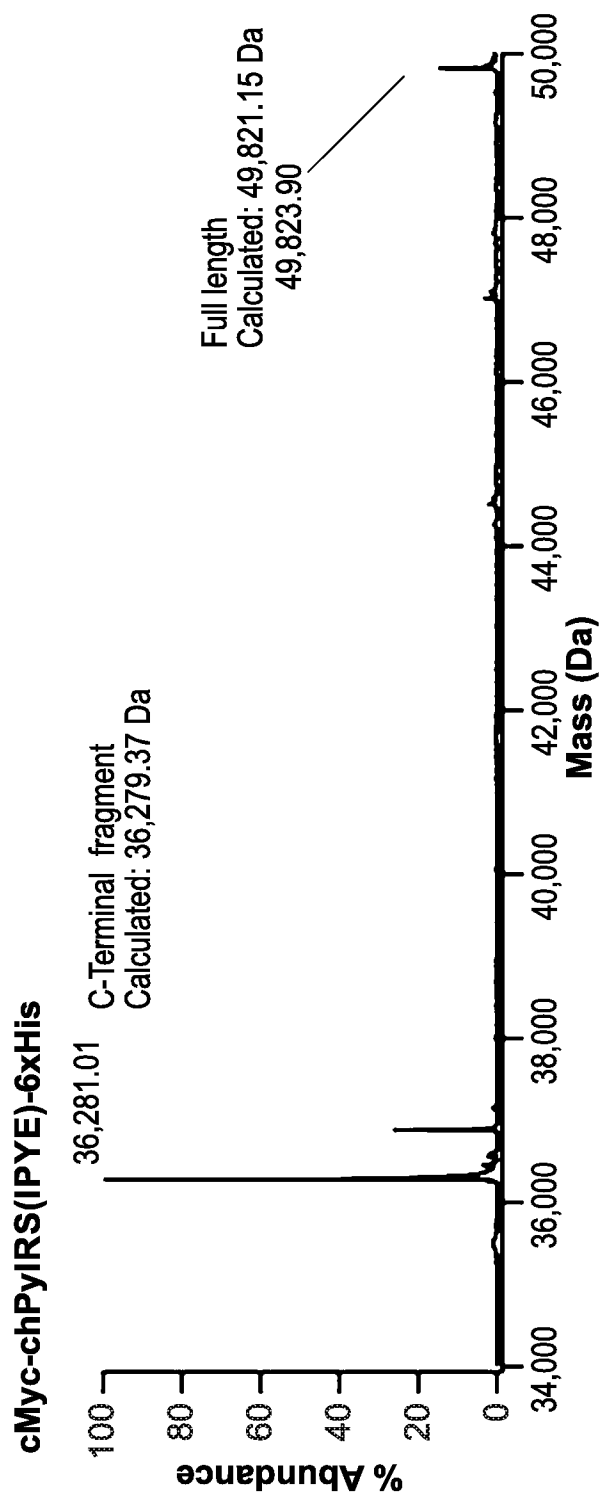


FIG. 15A

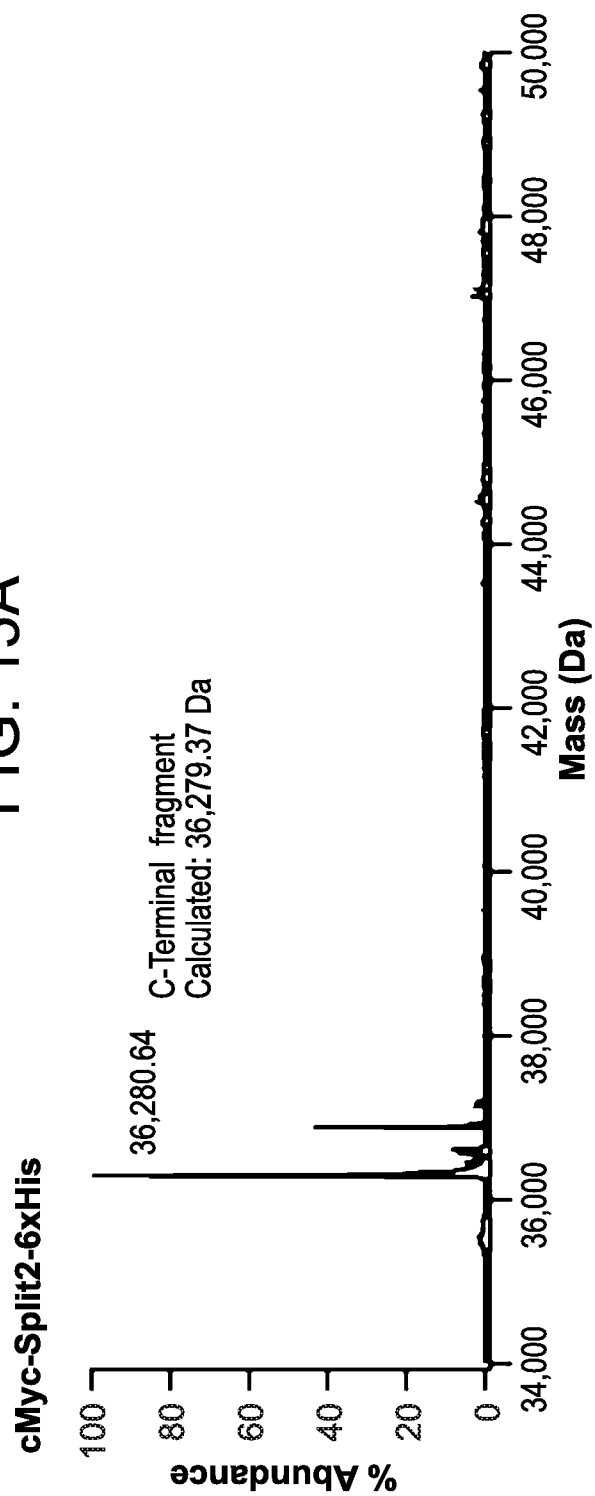


FIG. 15B

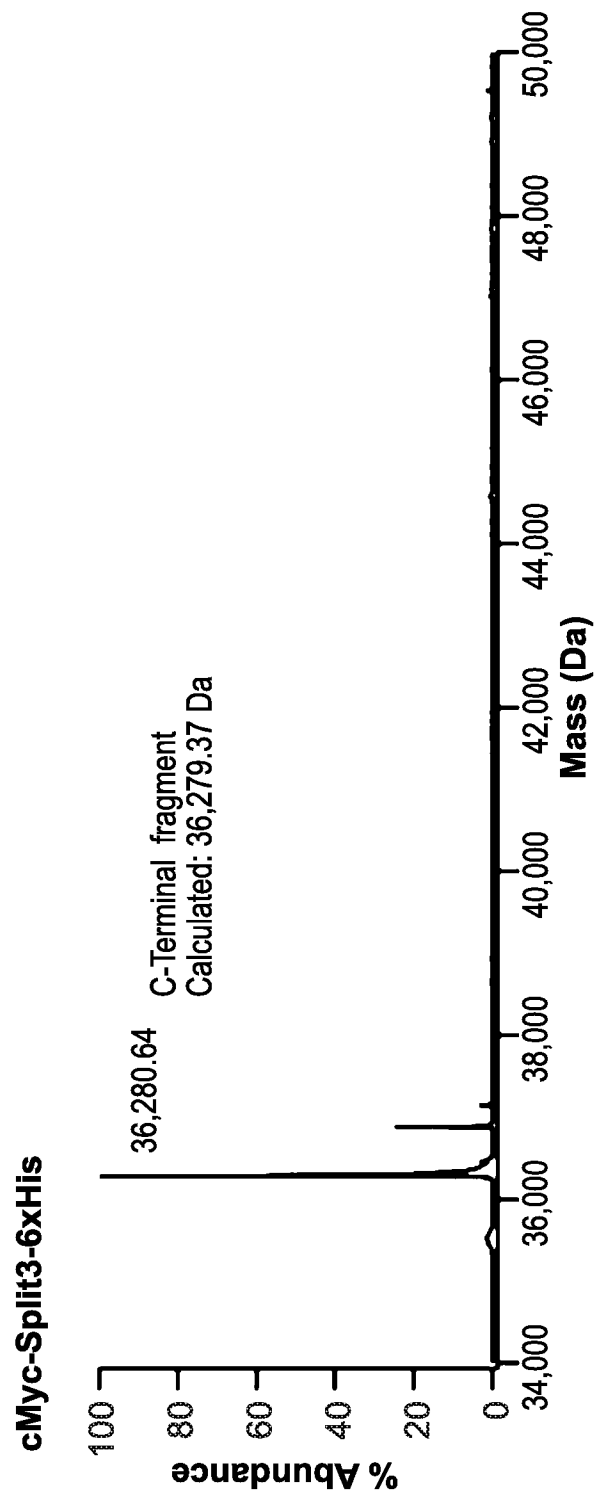


FIG. 15C



FIG. 15D

M.barkeri_F 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEVSRSKIYIEMACGDHLVNNNSRSCRTARAFRHHKY
M.mazei 1 -----MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVNNNSRSTRARALRHHKY
chPylRS 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEVSRSKIYIEMACGDHLVNNNSRSCRTARAFRHHKY
chPylRS (IPYE) 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEISRSKIYIEMACGDHLVNNNSRSCRPARAFRYHKY
Split1 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEISRSKIYIEMACGDHLVNNNSRSCRPARAFRYHKY
Split2 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEISRSKIYIEMACGDHLVNNNSRSCRPARAFRYHKY
Split3 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEISRSKIYIEMACGDHLVNNNSRSCRPARAFRYHKY
Split4 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEISRSKIYIEMACGDHLVNNNSRSCRPARAFRYHKY
Split5 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEISRSKIYIEMACGDHLVNNNSRSCRPARAFRYHKY
Split6 1 -----MDKKPLDVLI SATGLWMSRTGTLHKIKHKEISRSKIYIEMACGDHLVNNNSRSCRPARAFRYHKY
D.hafniense_PCP 1 -----
D.hafniense_Y51 1 MRGVSQASEEKKRYRKNVDFN LVEKIKLWPSRSGTLHG IKA MTRRGNTAEI VTHCNRRF I IYNS KHSRAARWL RNKLH
D.hafniense_DCB2 1 MRGVSQASEEKKRYRKNVDFN LVEKIKLWPSRSGTLHG IKA MTRRGNTAEI VTHCNRRF I IYNS KHSRAARWL RNKLH

M.barkeri_F 66 RKTCKRCRVSEDED INNFLTRSTEGKTSVKVKVVSAPK-VKKAMPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
M.mazei 66 RKTCKRCRVSEDED LNKFLT KANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQKQPSGSKFSPAIPVSTQES
chPylRS 66 RKTCKRCRVSEDED INNFLTRSTEGKTSVKVKVVSAPK-VKKAMPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
chPylRS (IPYE) 66 RKTCKRCRVSEDED INNFLTRSTEGKTSVKVKVVS~~EPK~~-VKKAMPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
Split1 66 RKTCKRCRVSEDED INNFLTRSTEGKTSVKKAVLSRK*-----MPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
Split2 66 RKTCKRCRVSEDED INNFLTRSTEGKTSVKVKVLSRK*-----MPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
Split3 66 RKTCKRCRVSEDED INNFLTRSTEGKTSVKVKVF*-----MPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
Split4 66 RKTCKRCRVSEDED INNFLTRSTEG*-----MPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
Split5 66 RKTCKRCRVSEDED INNFLTRSTEGKTSVKVKVVSERK*-----MPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
Split6 66 RKTCKRCRVSEDED INNFLTRSTEGKTLIC*-----MPKSVSRAPKPLEN PVSAKASTDTSRSVPSP-----
D.hafniense_PCP 1 -----
D.hafniense_Y51 81 FGVCPC~~CR~~IP~~EW~~KLQKYSSIVMSQHYGSHL*-----
D.hafniense_DCB2 81 FGVCPC~~CR~~IP~~EW~~KLQKYSSIVMSQHYGSHL*-----

↓ Dh.PylSn

FIG. 16

M.barkeri_F	138	-----AKSTPNSPVPTSAPAPSLTRSQIDRVEALLSPED-----KISLNI AKPFREL
M.mazei	146	VSPASVSTSISSITGATASALVKGNTNPIITMSAPVQASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
chPylRS	138	-----AKSTPNSPVPTSASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
chPylRS(IPYE)	138	-----AKSTPNSPVPTSASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
Split1	134	-----AKSTPNSPVPTSASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
Split2	134	-----AKSTPNSPVPTSASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
Split3	131	-----AKSTPNSPVPTSASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
Split4	121	-----AKSTPNSPVPTSASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
Split5	134	-----AKSTPNSPVPTSASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
Split6	125	-----AKSTPNSPVPTSASAPALTKSQTDRL EVLLNPKD -----EISLNSGKPFREL
D.hafniense_PCP	1	-----MFLTRRD PPL -----SSEWTKVQYQRLKELNASGEQLEMGS DALS SRDRA FQGI
D.hafniense_Y51	111	-----MDRIDH TD SKFVQAGETPVLPA TF MFLTRRD PPL -----SSEWTKVQYQRLKELNASGEQLEMGS DALS SRDRA FQGI
D.hafniense_DCB2	111	-----MSEWTKVQYQRLKELNASGEQLEMGS DALS SRDRA FQGI
		 Dh.PylSc
M.barkeri_F	185	ESELL TRR KKND FQRL Y TND REDYLGKLERDITKFFVDRD FLE IKSPI LPAEY VERMGINND TELSKQIFRVDKNLC LRP
M.mazei	220	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEY IERMGIDND TELSKQIFRVDKNF CLRP
chPylRS	185	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
chPylRS(IPYE)	185	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
Split1	181	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
Split2	181	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
Split3	178	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
Split4	168	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
Split5	181	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
Split6	172	ESELL SR KKDLQ QI YAEERENYL GKLER EITRFFVDRGF E IKSPI LPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
D.hafniense_PCP	50	EHQLMSQCK RHLEQ LR TV KHRPALLE E EKLAKALHQ Q GFVQV VTPTIT ITKSALAKMTIGEDHPLFSQ VFWLDG KKKCLRP
D.hafniense_Y51	184	EHQLMSQCK RHLEQ LR TV KHRPALLE E EGLAKALHQ Q GFVQV VTPTIT ITKSALAKMTIGEDHPLFSQ VFWLDG KKKCLRP
D.hafniense_DCB2	151	EHQLMSQCK RHLEQ LR TV KHRPALLE E EGLAKALHQ Q GFVQV VTPTIT ITKSALAKMTIGEDHPLFSQ VFWLDG KKKCLRP

FIG. 16 (cont.)

M.barkeri_F	265	MLAP ^T LYNLRKLDRIILDPPIKIFEVGPCYRKESDGKEHLEEFF ^T MVNFCQMGSGCTRE--NLES ^L LIKEFLDYLE-IDFEI
M.mazei	300	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
chPylRS	265	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
chPylRS(IPYE)	265	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
Split1	261	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
Split2	261	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
Split3	258	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
Split4	248	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
Split5	261	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
Split6	252	MLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFF ^T MLNFCQMGSGCTRE--NLESII ^T TD ^T FLNHLG-IDFKI
D.hafniense_PCP	130	MLAPNLYTLWRELERLWDKPIRIFEIGTCYRKESQGAQH ^L NEFT ^T MLNLT ^{EL} GT ^T PLEERHQRLEDMARWVLEAAGIREFEL
D.hafniense_Y51	264	MLAPNLYTLWRELERLWDKPIRIFEIGTCYRKESQGAQH ^L NEFT ^T MLNLT ^{EL} GT ^T PLEERHQRLEDMARWVLEAAGIREFEL
D.hafniense_DCB2	231	MLAPNLYTLWRELERLWDKPIRIFEIGTCYRKESQGAQH ^L NEFT ^T MLNLT ^{EL} GT ^T PLEERHQRLEDMARWVLEAAGIREFEL

M.barkeri_F	342	VGDSCMVYGD ^T LDIMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M HGFKNIKRA ^S RSESYNGISTNL-
M.mazei	377	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
chPylRS	342	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
chPylRS(IPYE)	342	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
Split1	338	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
Split2	338	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
Split3	335	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
Split4	325	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
Split5	338	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
Split6	329	VGDSCMVYGD ^T LDVMHGDLELSSAVVGPVPLDREWGIDK ^P WIGAGFGLERLLKV ^M KHDFKNIKRA ^S RSESYNGISTNL-
D.hafniense_PCP	210	VTESSVYGD ^T VDVMKGDLELASGAMGPHFLDEKWEI ^F DPWVGLGFGLE ^R LLMIREG ^T QHVQSMARSL ^S YLDGVRLNIN
D.hafniense_Y51	344	VTESSVYGD ^T VDVMKGDLELASGAMGPHFLDEKWEI ^F DPWVGLGFGLE ^R LLMIREG ^T QHVQSMARSL ^S YLDGVRLNIN
D.hafniense_DCB2	311	VTESSVYGD ^T VDVMKGDLELASGAMGPHFLDEKWEI ^F DPWVGLGFGLE ^R LLMIREG ^T QHVQSMARSL ^S YLDGVRLNIN

FIG. 16 (cont.)

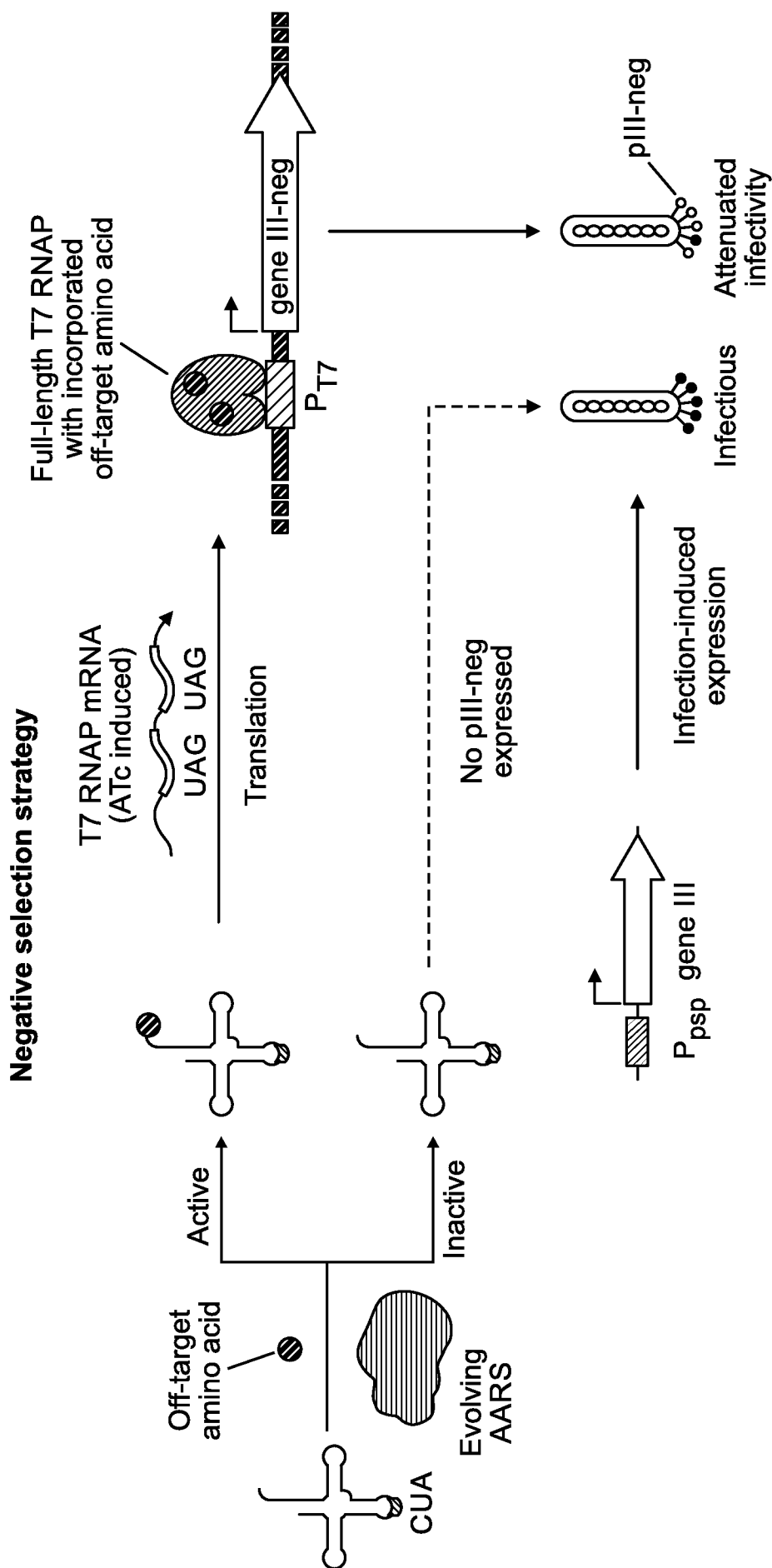


FIG. 17A

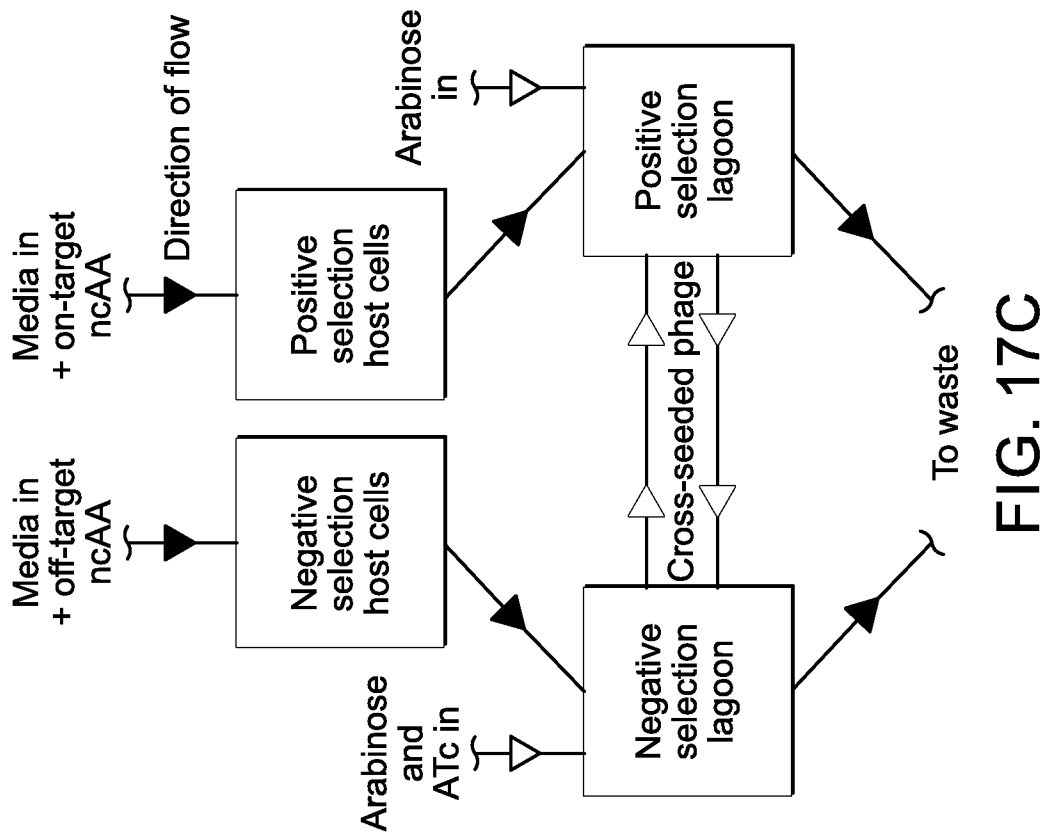


FIG. 17C

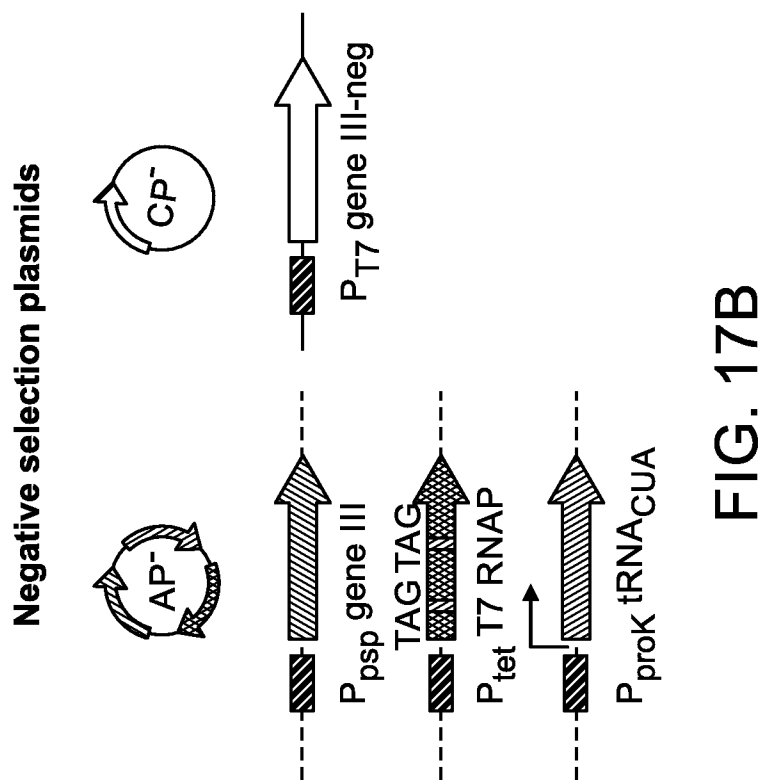


FIG. 17B

Mutations relative to p-NFRS

p-IFRS	N160H
Iodo.1	S107P
Iodo.5	L69F, V235I
Iodo.7	G163C, N211K
Iodo.8	S207A

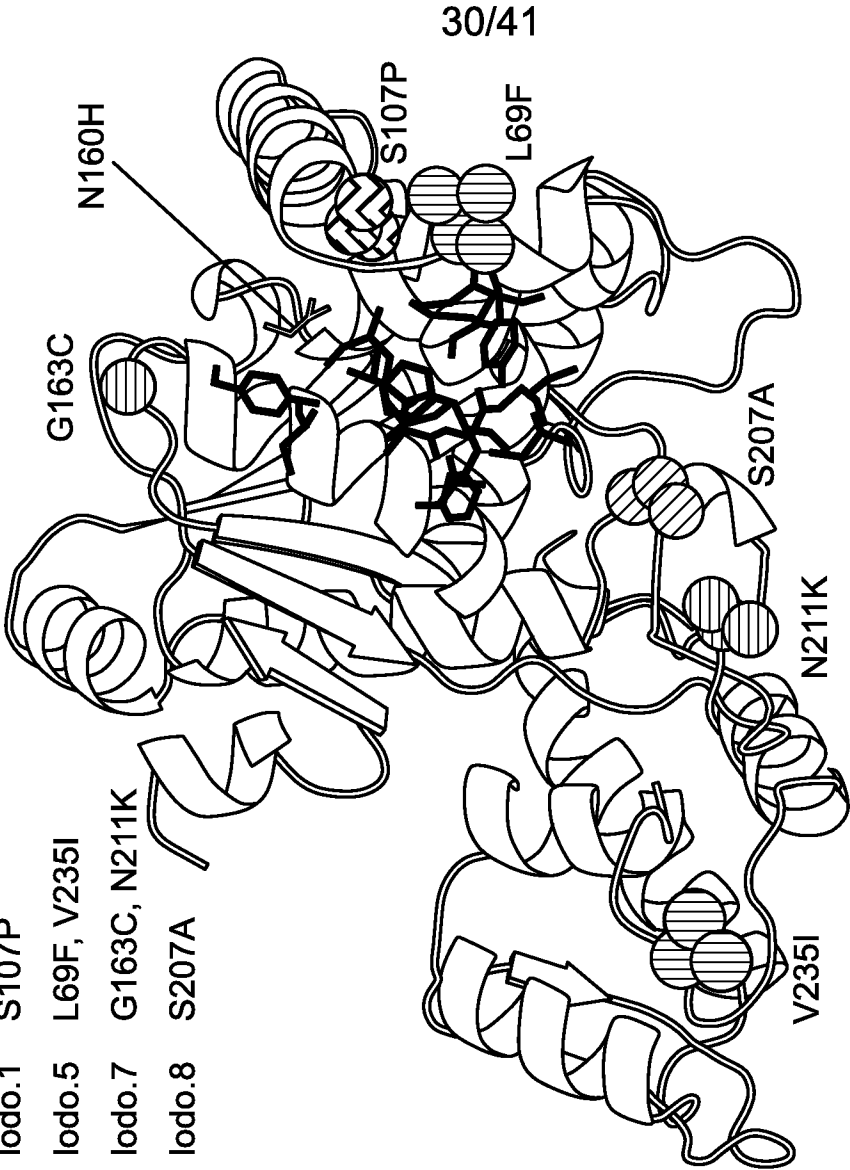


FIG. 17E

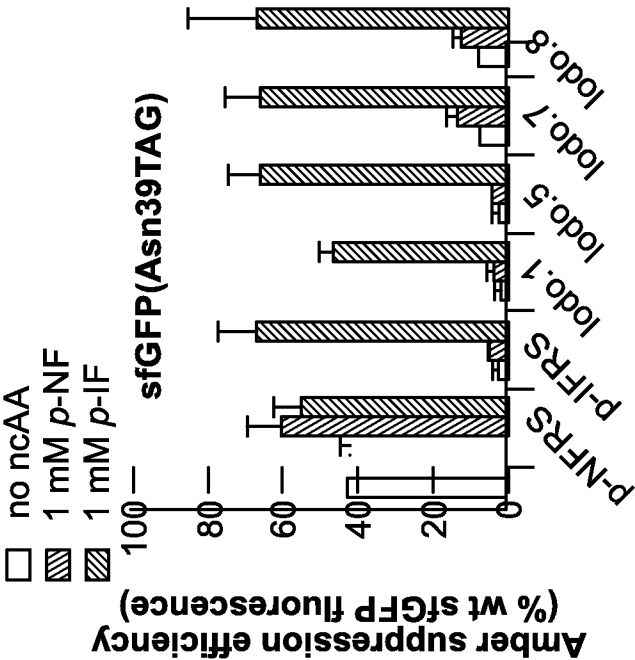


FIG. 17D

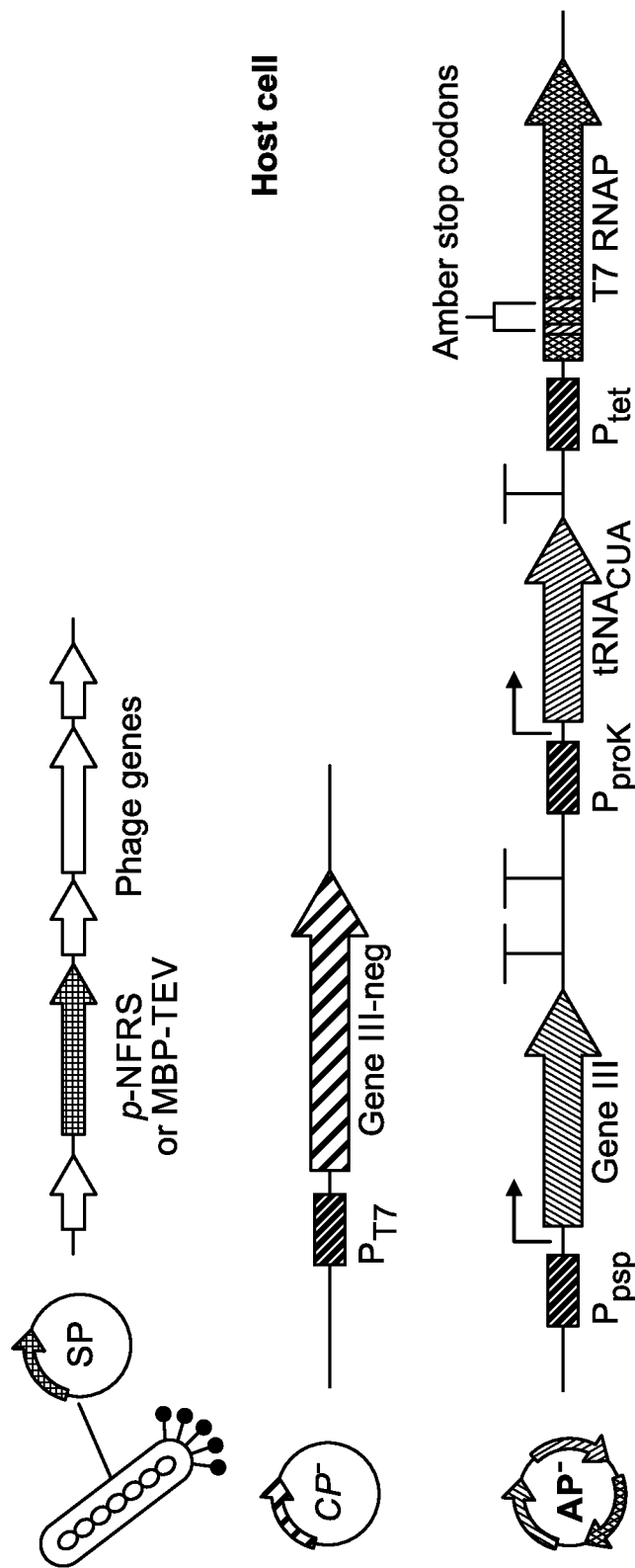


FIG. 18A

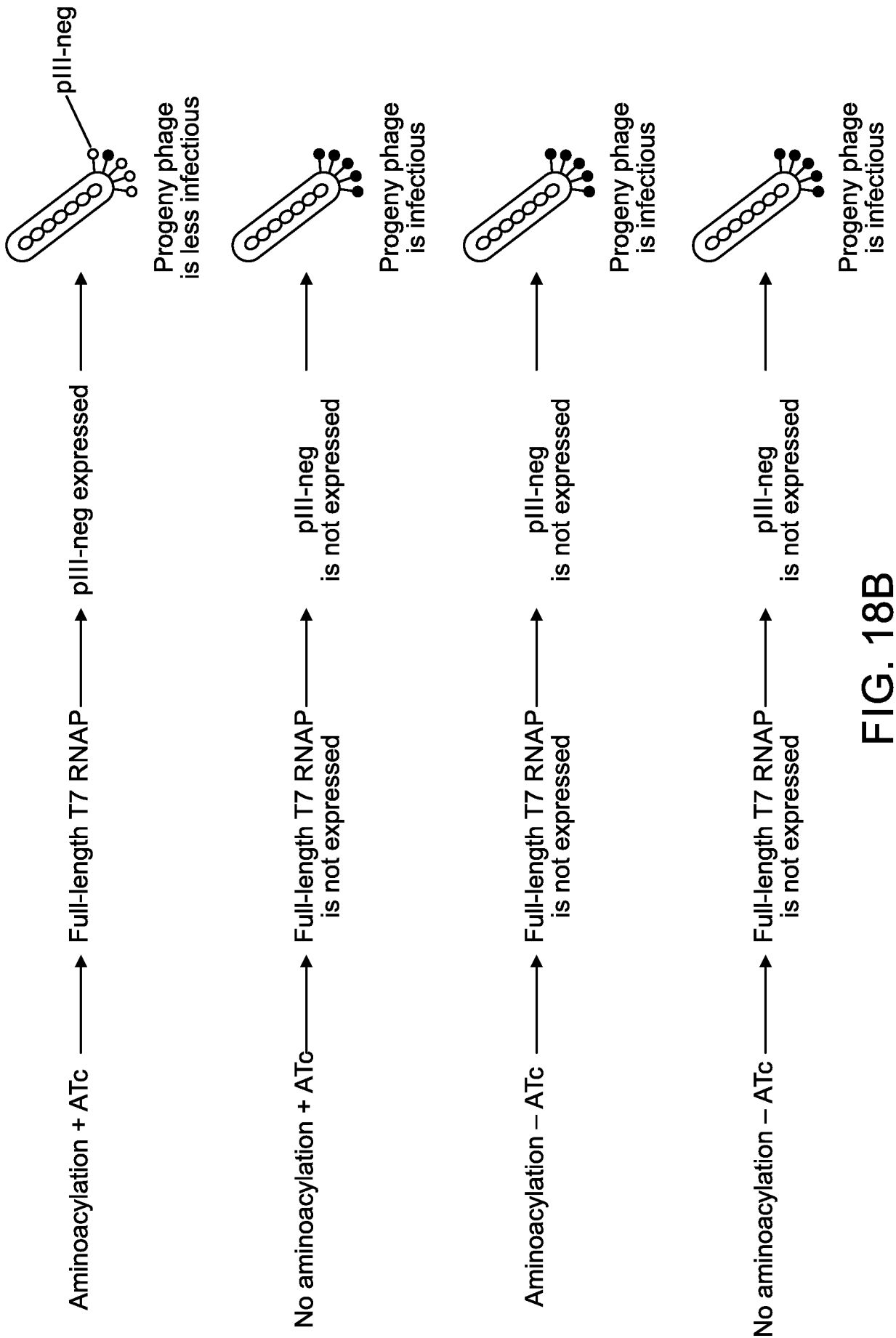


FIG. 18B

33/41

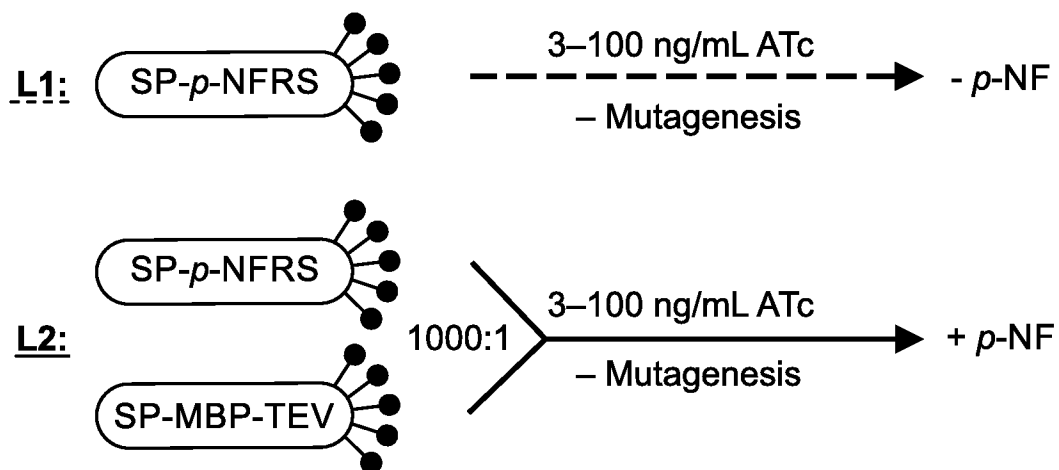


FIG. 19A

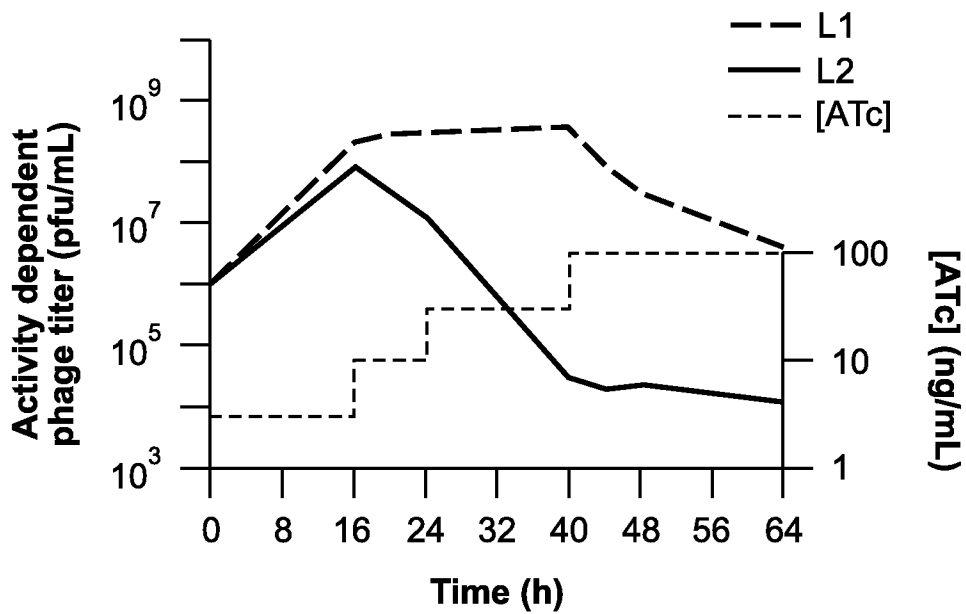


FIG. 19B

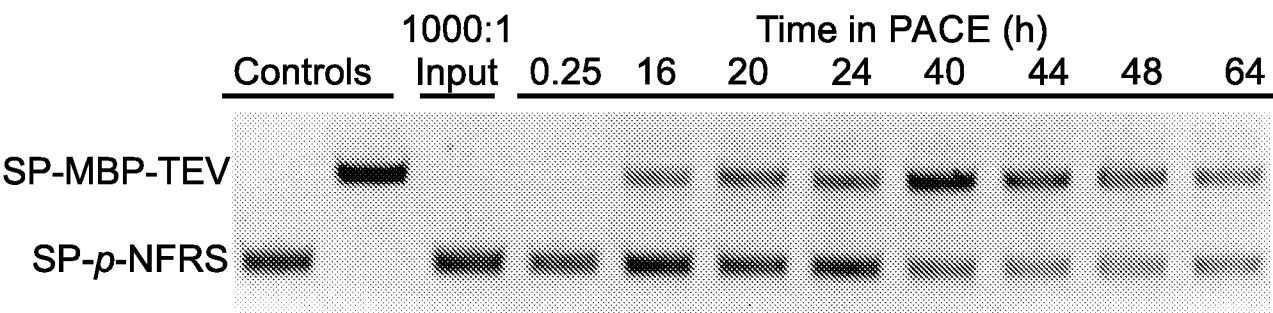


FIG. 19C

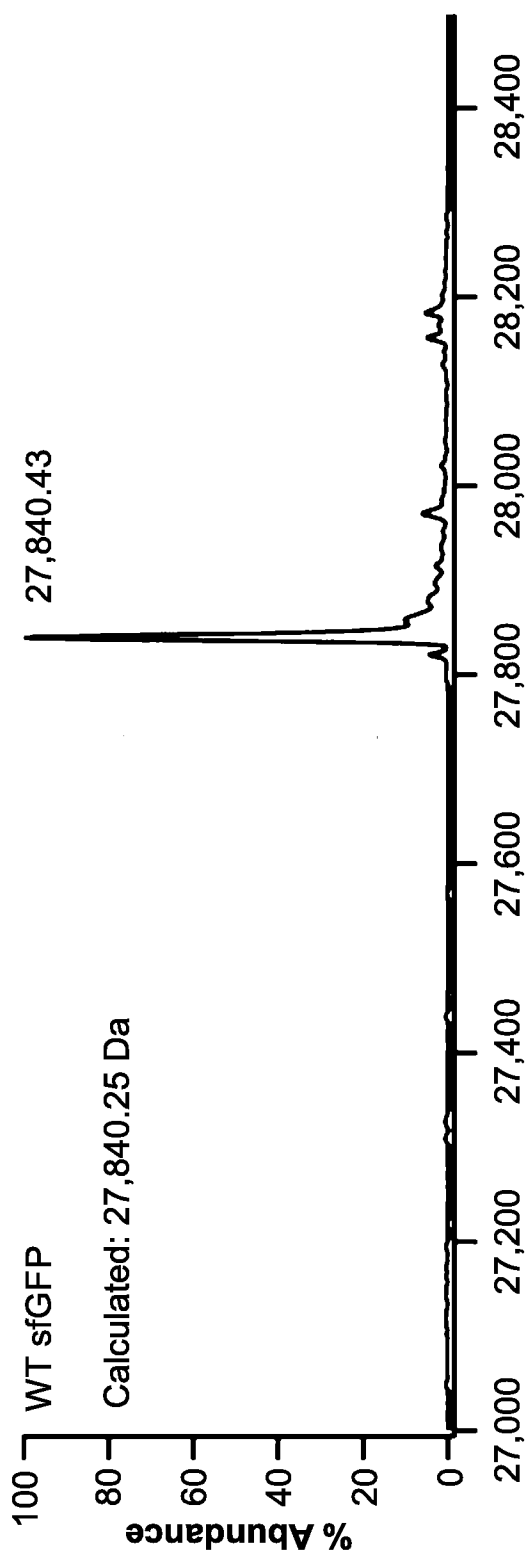


FIG. 20A

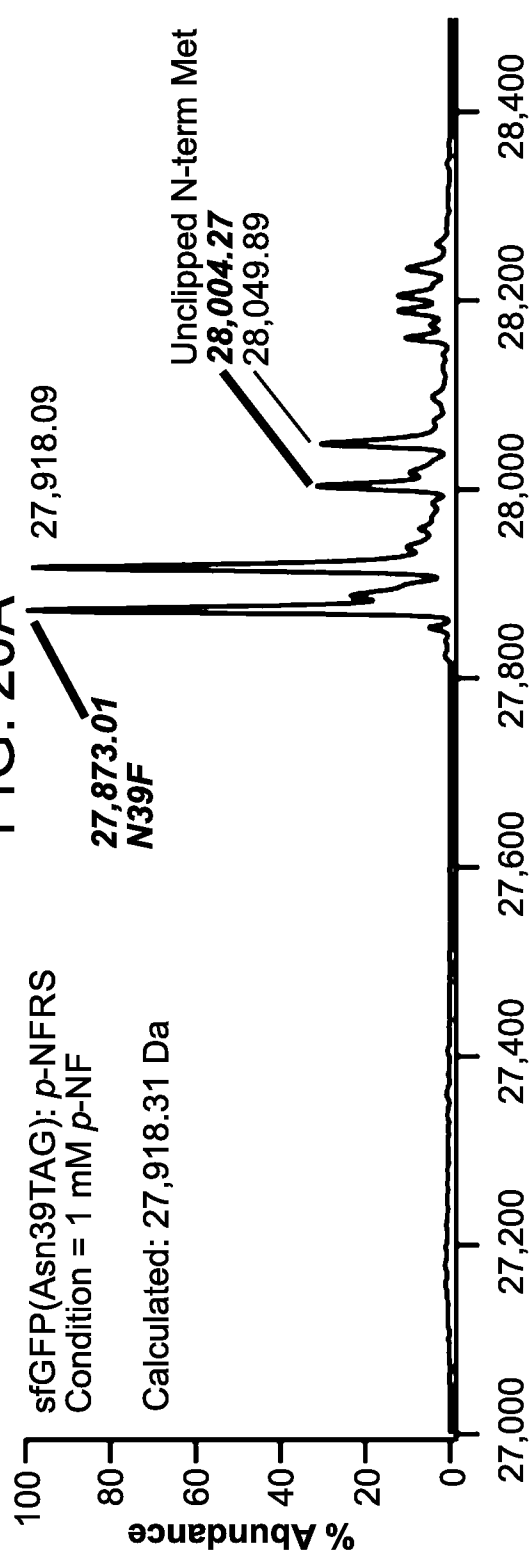


FIG. 20B

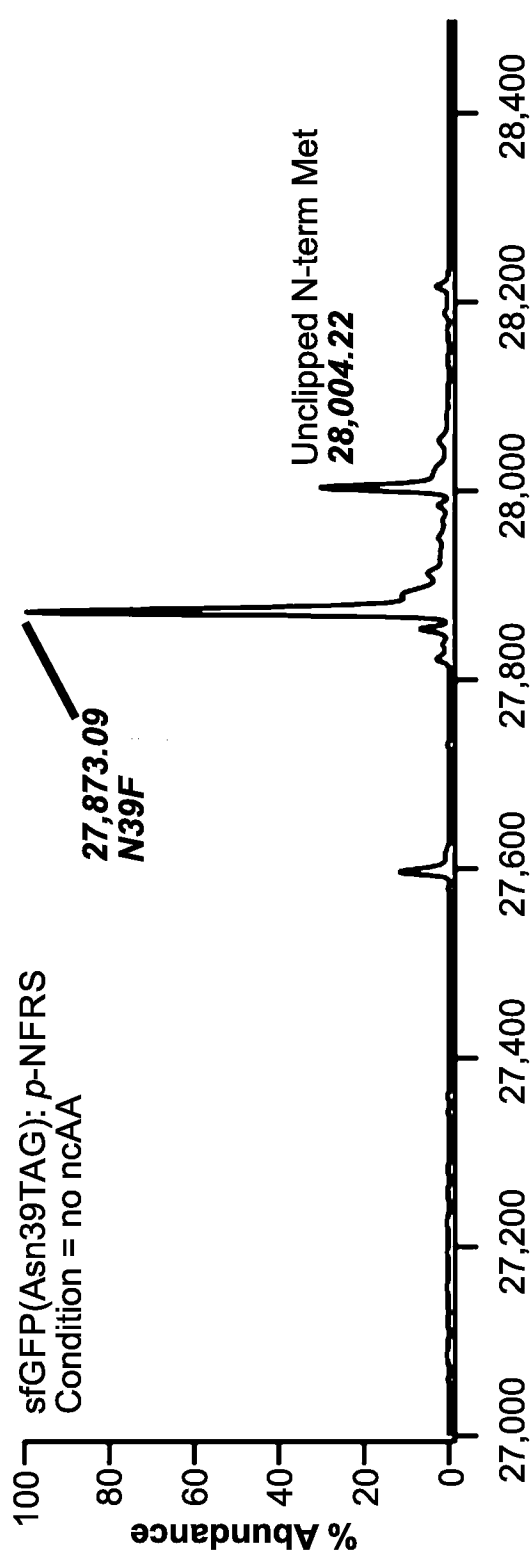


FIG. 20C

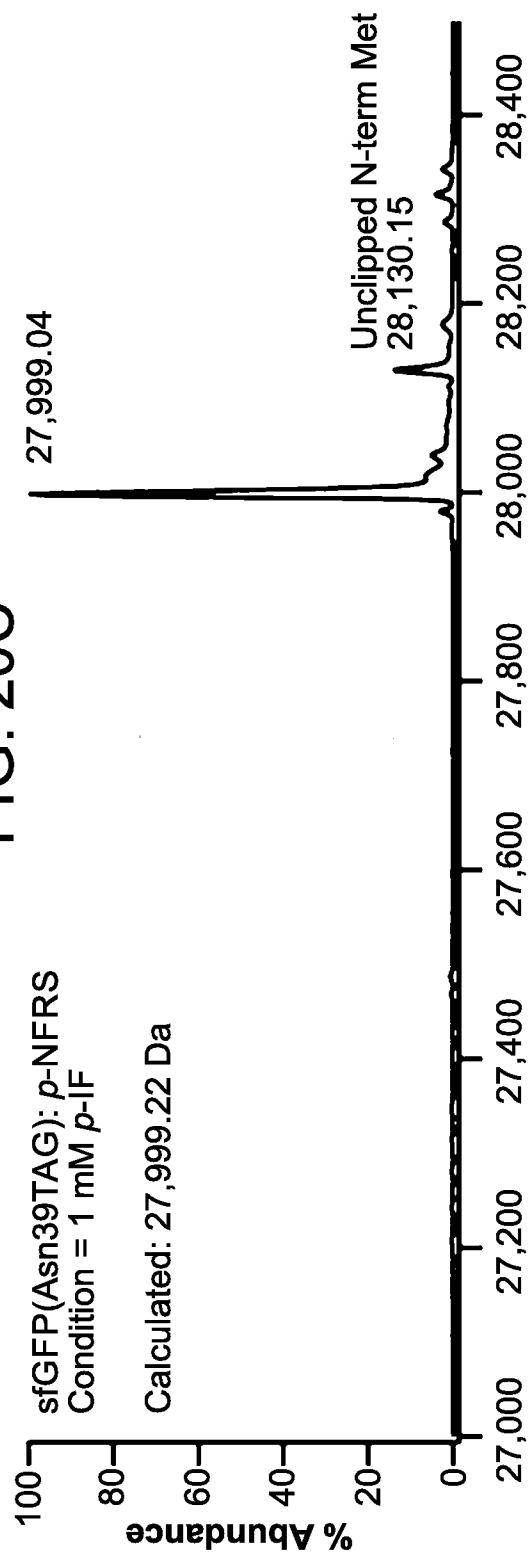


FIG. 20D

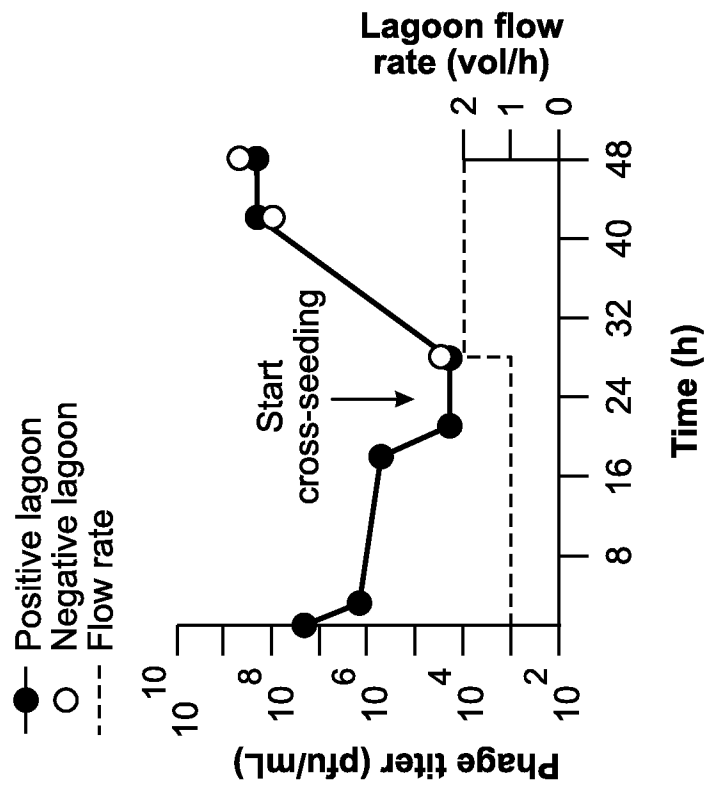


FIG. 21B

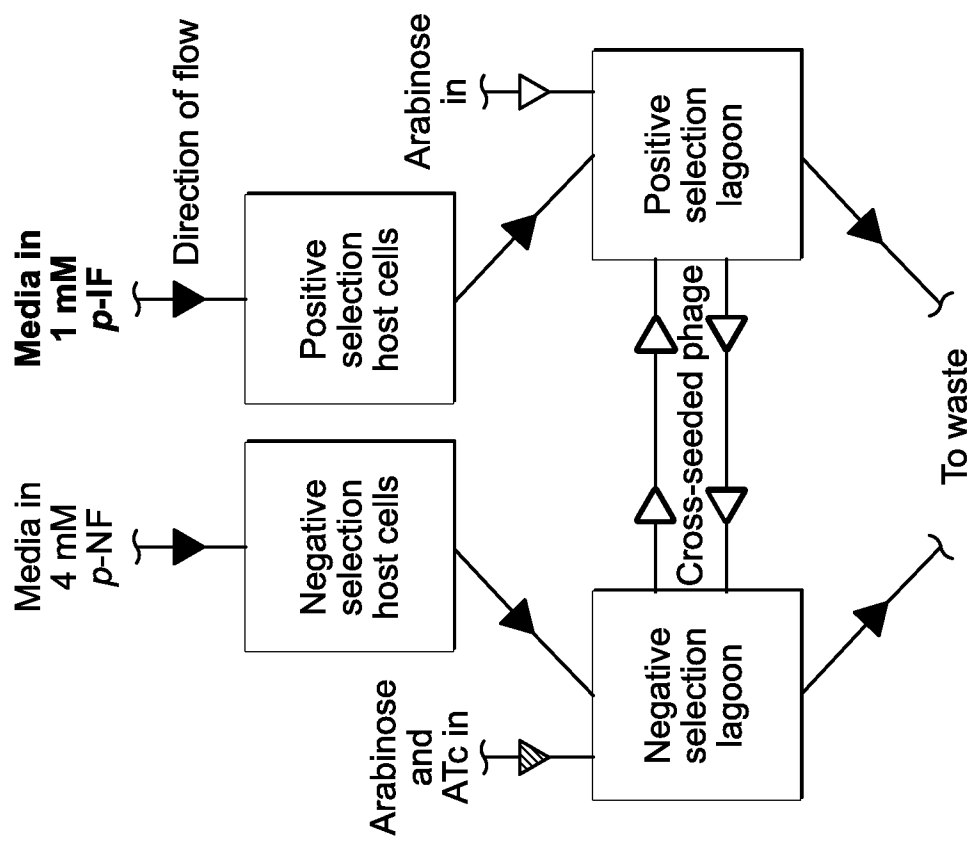


FIG. 21A

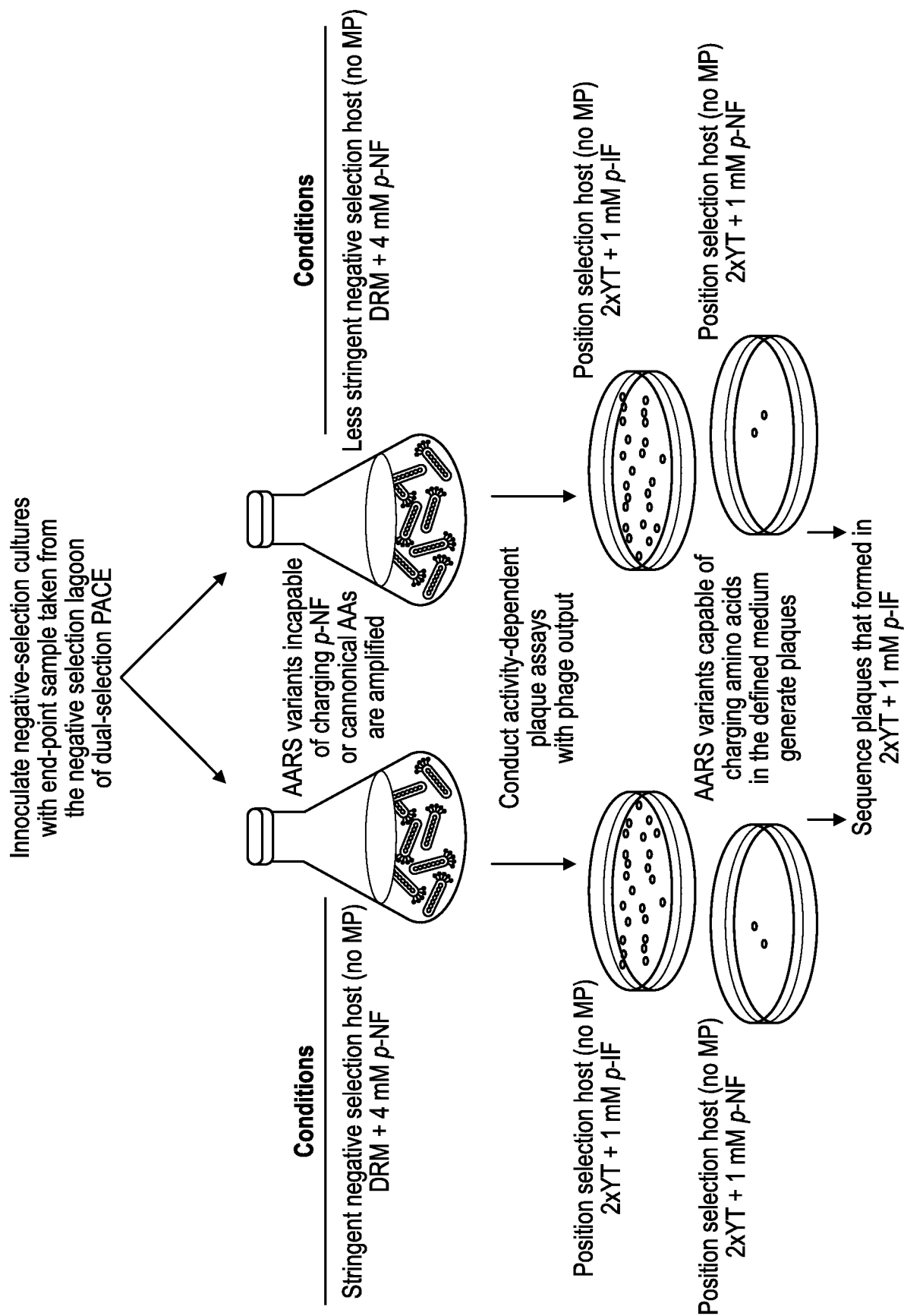


FIG. 22A

Non-continuous negative selection	Enrichment Factor (pfu in <i>p</i> -IF / pfu in <i>p</i> -NF)	Clonal phage isolate	Mutations
Stringent	267	lodo.1	S107P
		lodo.2	<i>RBS mutation</i>
		lodo.3	<i>RBS mutation</i>
		lodo.4	<i>RBS mutation</i>
Less stringent	233	lodo.5	L69F, V235I
		lodo.6	<i>RBS mutation</i>
		lodo.7	G163C, N211K
		lodo.8	S207A

FIG. 22B

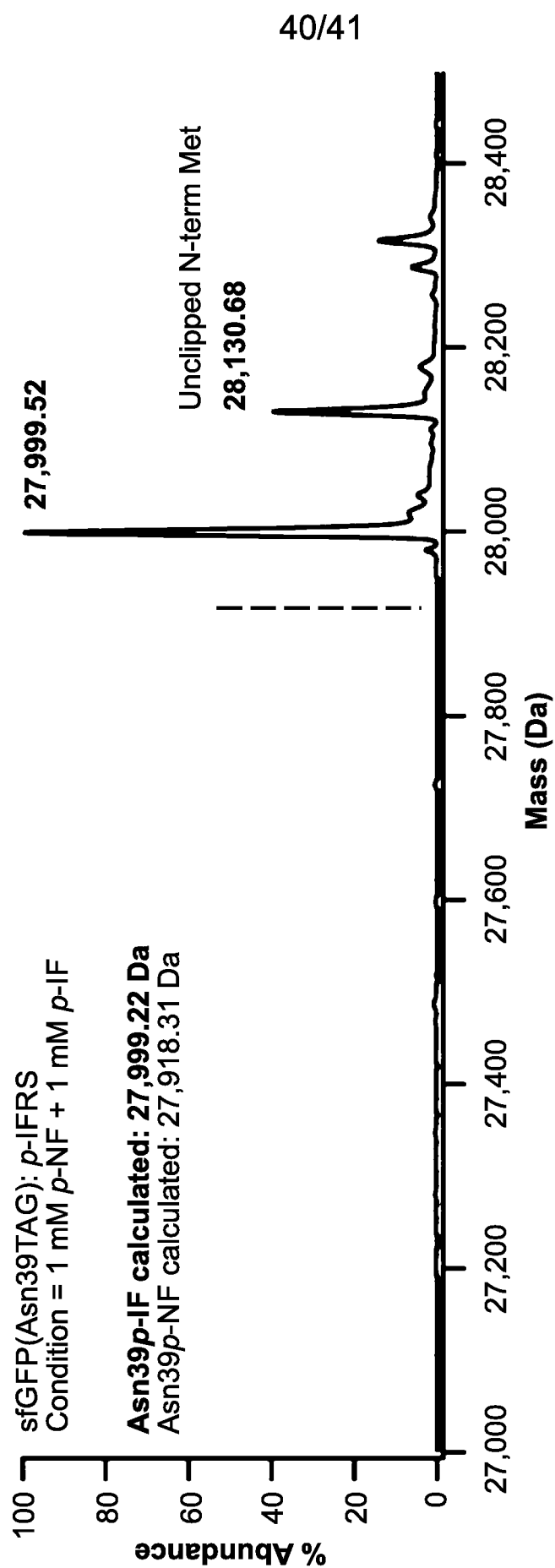


FIG. 23A

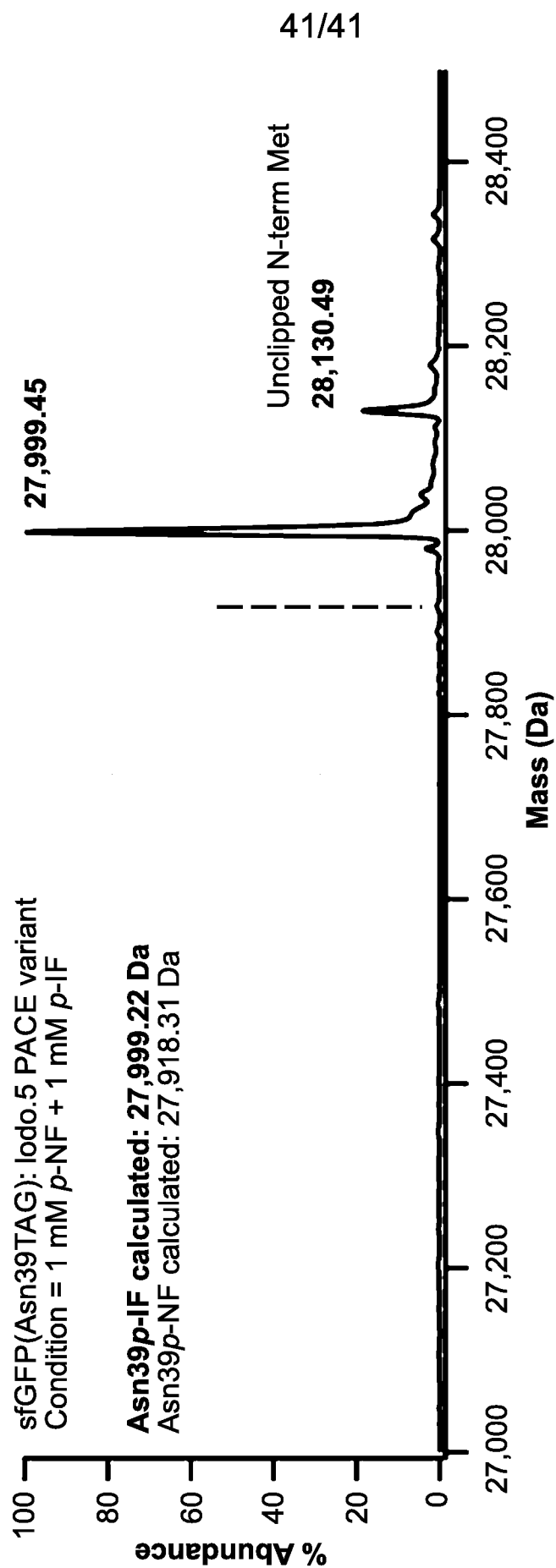


FIG. 23B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US18/40692

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12P 21/02, 21/00; A61K 38/00 (2018.01)

CPC - C12P 21/02, 21/00; A61K 38/00

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2015/193897 A1 (B. G. NEGEV TEGECHOLOGIES AND APPLICATIONS LTD., AT BEN-GURION UNIVERSITY) 23 December 2015; Abstract; Paragraphs [001], [0019], [0042]; Claims 3, 4	1, 2, 6, 7
Y	US 2006/166319 A1 (CHAN, MK, et al.) 27 July 2006; Abstract; Claim 7; SEQ ID NO: 10	1, 2, 6, 7
Y	WO 2011/039518 A2 (MEDICAL RESEARCH COUNCIL) 07 April 2011; Abstract; Paragraph [0029]; Page 9 of 43, AZG75773 standard protein amino acid sequence	1, 2, 6, 7
Y	US 2015/0259721 A1 (ALLOYZYNE, INC.) 17 September 2015; Abstract; Paragraph [0083]; SEQ ID NO: 3	1, 2, 6, 7



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

23 October 2018 (23.10.2018)

Date of mailing of the international search report

15 NOV 2018

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US18/40692

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 5, 20-25
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

This International Searching Authority found multiple inventions in this international application, as follows:

---Please See within Supplemental Page---

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

---Please See Supplemental Page---

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/US18/40692

***-Continued from Box No. III: Observations Where Unity of Invention is Lacking:

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

Groups I+, Claims 1-4, 6-19 and SEQ ID NO: 20 with V31I mutation are directed toward tRNA synthetase protein variants, fragments thereof, and chimeric proteins which comprise at least one amino acid substitution in the N-terminal domain, C-terminal domain or both domains.

The tRNA synthetase protein variants will be searched to the extent that they encompass SEQ ID NO: 20 with a V31I mutation (first exemplary tRNA synthetase variant). Applicant is invited to elect additional tRNA synthase sequence(s), (including amino acid sequence(s) and/or, where applicable, encoding nucleic acid sequence(s)) with specified SEQ ID NO: for each, or with specified substitution(s) at specified site(s) of a SEQ ID NO: such that the sequence of each elected species is fully specified (i.e. no optional or variable residues or substituents), to be searched. Additional tRNA synthase sequence(s) will be searched upon the payment of additional fees. It is believed that claims 1 (in-part) 2 (in-part), 6 (in-part), and 7 (in-part) encompass this first named invention and thus these claims will be searched without fee to the extent that they encompass SEQ ID NO: 20 with a V31I mutation (tRNA synthetase protein variant). Applicants must specify the claims that encompass any additionally elected tRNA synthetase sequence(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be SEQ ID NO: 21 with a T56P mutation (tRNA synthetase protein variant).

No technical features are shared between the polypeptide sequences of Groups I+ and, accordingly, these groups lack unity a priori.

Additionally, even if Groups I+ were considered to share the technical features of: a tRNA synthetase protein variant comprising an amino acid sequence that is at least 90% identical to a defined sequence, and includes at least one mutation, further wherein said variant is a pyrrolysyl-tRNA synthetase (PylRS) protein variant having an N-terminal domain amino acid substitution; a chimeric pyrrolysyl-tRNA synthetase (PylRS) protein variant comprising: (i) a first portion comprising amino acid residues 1-149 of Methanosarcina barkeri PylRS; and (ii) a second portion comprising amino acid residues 185-454 of Methanosarcina mazei PylRS, wherein the first portion or the second portion comprises at least one amino acid substitution; however, these shared technical features are previously disclosed by WO 2015/193897 A1 to B.G. Negev Technologies and Applications at Ben Gurion University (hereinafter 'Negev Technologies').

Negev Technologies discloses a tRNA synthetase protein variant (tRNA synthetase mutant; abstract; paragraph [0042]) comprising an amino acid sequence that is at least 90% identical to a defined sequence (tRNA synthetase having a rare or non-natural amino acid; abstract; paragraph [001]), and includes at least one mutation (amber or ochre mutant; abstract; paragraph [0039]), further wherein said variant is a pyrrolysyl-tRNA synthetase (PylRS) protein variant (paragraphs [0028], [0042], [00111]) having an N-terminal domain amino acid substitution (mutation at amino acid position Y37; paragraph [0075]). Negev Technologies also discloses chimeric tRNA synthetases (fusion proteins or conjugates of tRNA synthetases; paragraphs [00113], [00273]) wherein the tRNA synthetases are derived from Methanosarcina barkeri PylRS and Methanosarcina mazei PylRS (paragraphs [0012], [00155]), and wherein the mutation occurs within residues 185-454 of the tRNA synthetase protein (amino acid substitution at Q195; paragraph [0075]), but Negev Technologies does not disclose a chimeric pyrrolysyl-tRNA synthetase (PylRS) protein variant comprising: (i) a first portion comprising amino acid residues 1-149 of Methanosarcina barkeri PylRS; and (ii) a second portion comprising amino acid residues 185-454 of Methanosarcina mazei PylRS, wherein the first portion or the second portion comprises at least one amino acid substitution. However, it would have been obvious to one of ordinary skill in the art to have modified the disclosure of Negev Technologies to provide a chimeric pyrrolysyl-tRNA synthetase (PylRS) protein variant comprising: (i) a first portion comprising amino acid residues 1-149 of Methanosarcina barkeri PylRS; and (ii) a second portion comprising amino acid residues 185-454 of Methanosarcina mazei PylRS, wherein the first portion or the second portion comprises at least one amino acid substitution, because Negev Technologies discloses mutations in both the N-terminal (residues 1-149) and C-terminal (residues 185-454) domains of PylRS proteins from both Methanosarcina barkeri and Methanosarcina mazei and the ability to conjugate tRNA synthetase domains which could be combined to produce said chimeric pyrrolysyl-tRNA synthetase (PylRS) protein variant.

Since none of the special technical features of the Groups I+ inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Negev Technologies reference, unity of invention is lacking.