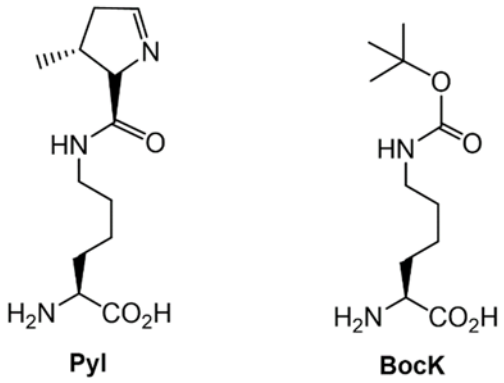
SUPPLEMENTARY INFORMATION

Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase

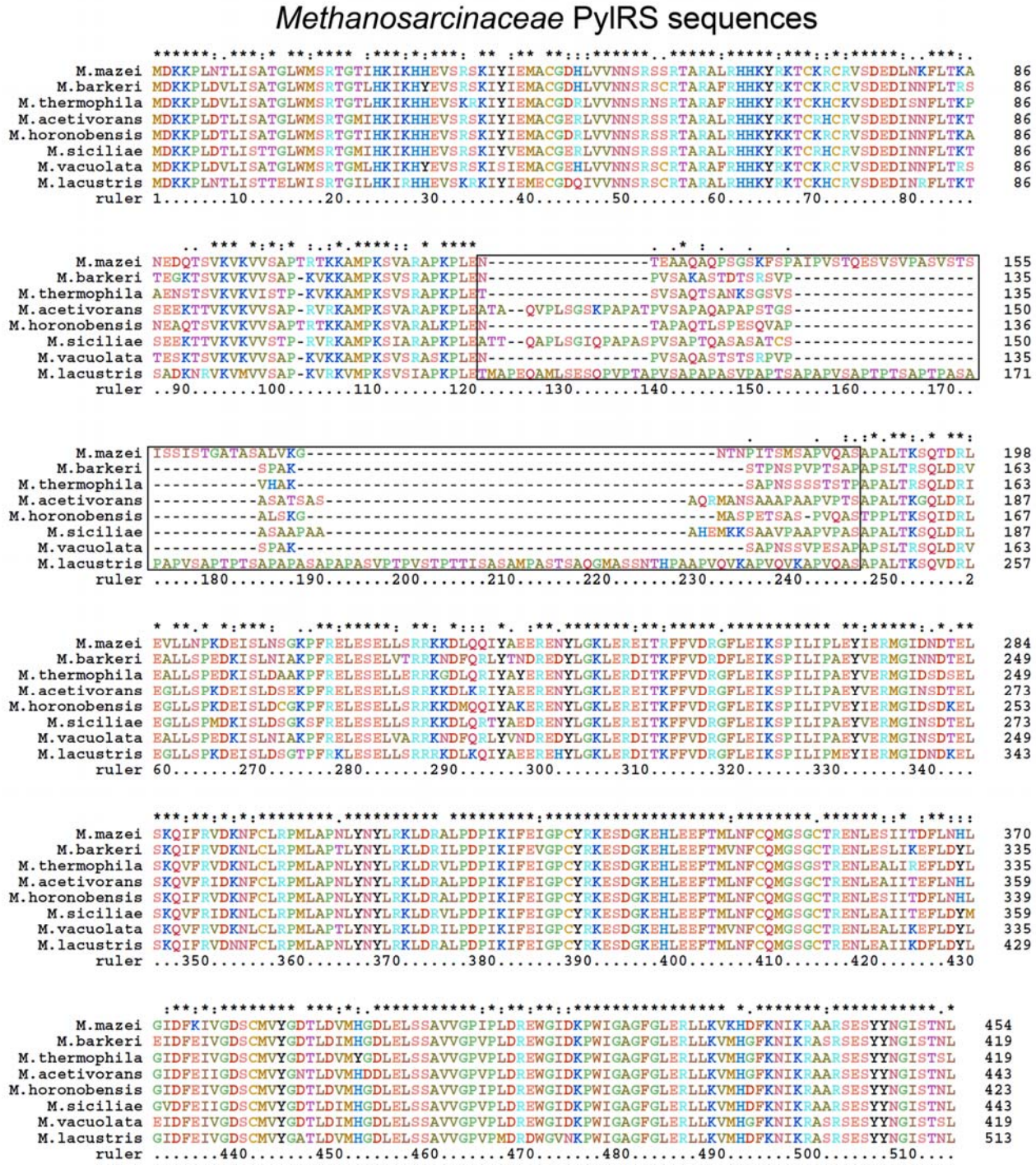
Tateki Suzuki[a] , Corwin Miller[a,f], Li-Tao Guo[a], Joanne M. L. Ho[a], David I. Bryson[c], Yane-Shih Wang[a,g], David R. Liu[c,d,e], and Dieter Söll[a,b,*]

Departments of [a]Molecular Biophysics & Biochemistry, and [b]Chemistry, Yale University, New Haven, CT, USA
[c]Department of Chemistry and Chemical Biology, Harvard University
[d]Howard Hughes Medical Institute, Cambridge, MA, USA
[e]Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA
[f]Current address: DS Therapeutics, Houston, TX, 77056, USA
[g]Current address: Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan

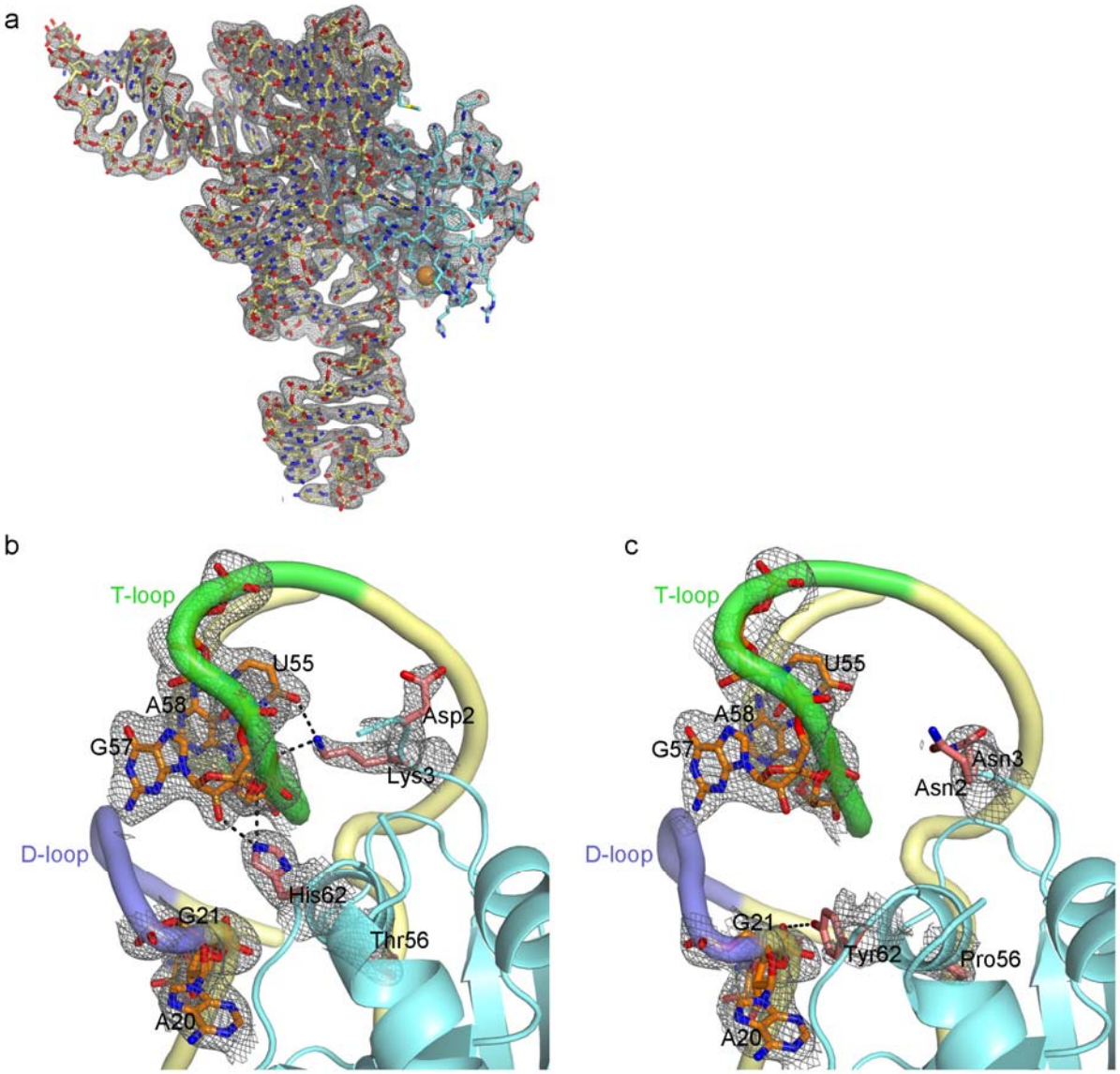
SUPPLEMENTARY RESULTS



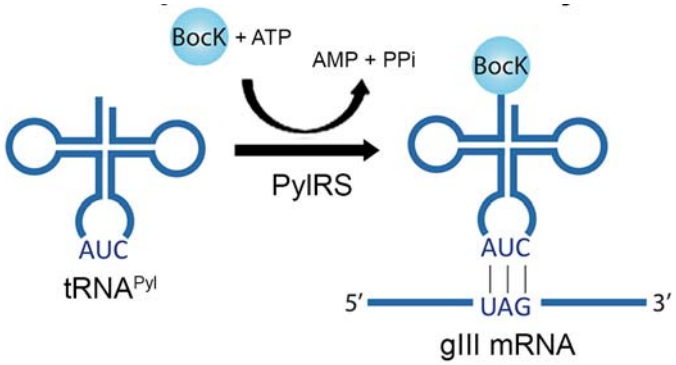
Supplementary Figure 1 | Non-canonical amino acid chemical structures. L-pyrrolysine (**Pyl**) and N^{ε}-(tert-butoxycarbonyl)-L-lysine (**BocK**).



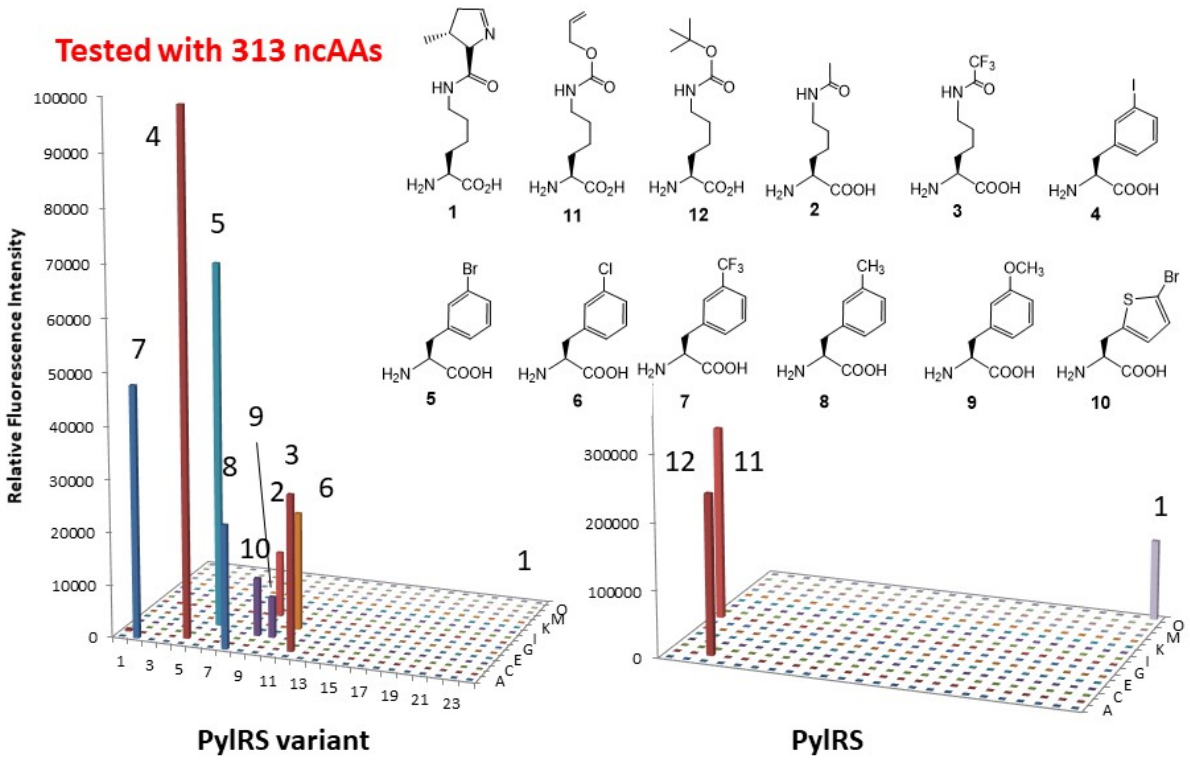
Supplementary Figure 2 | Alignment of eight archaeal *Methanosarcina* PylRS sequences. From *M. mazei*, *M. barkeri*, *M. thermophila*, *M. acetivorans*, *M. horonobensis*, *M. siciliae*, *M. vacuolata*, and *M. lacustris*. The sequence alignment was generated by Clustal X[1]. The linker regions of these PylRSs are framed.



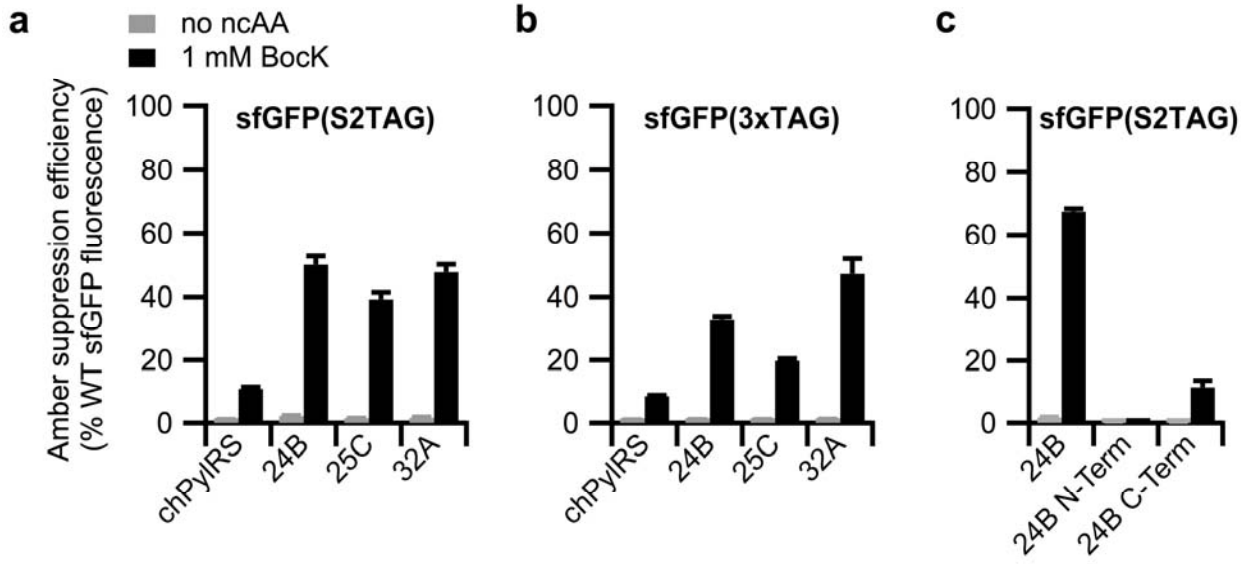
Supplementary Figure 3 | The structures of PylRS and a PANCE-evolved variant in complex with tRNA^{Pyl} with the electron density map. (**a**) Overview of the $2F_o$-F_c map for MmPylRS NTD • tRNA^{Pyl} complex contoured at 1σ. MmPylRS NTD (cyan) and tRNA^{Pyl} (yellow) are shown in stick model. Zn^{2+} is shown as an orange sphere. (**b**) The simulated annealing composite omit map of the interface between MmPylRS NTD and tRNA^{Pyl} contoured at 1σ. (**c**) The simulated annealing composite omit map of the interface between PANCE-evolved variant 32A NTD and tRNA^{Pyl} contoured at 1σ. Evolved residues in PANCE (pink) and nucleotides interacting with the NTD of each PylRS variant (orange) are shown in stick. Polar interactions between PylRS residues and tRNA nucleotides are shown by a black dashed line.



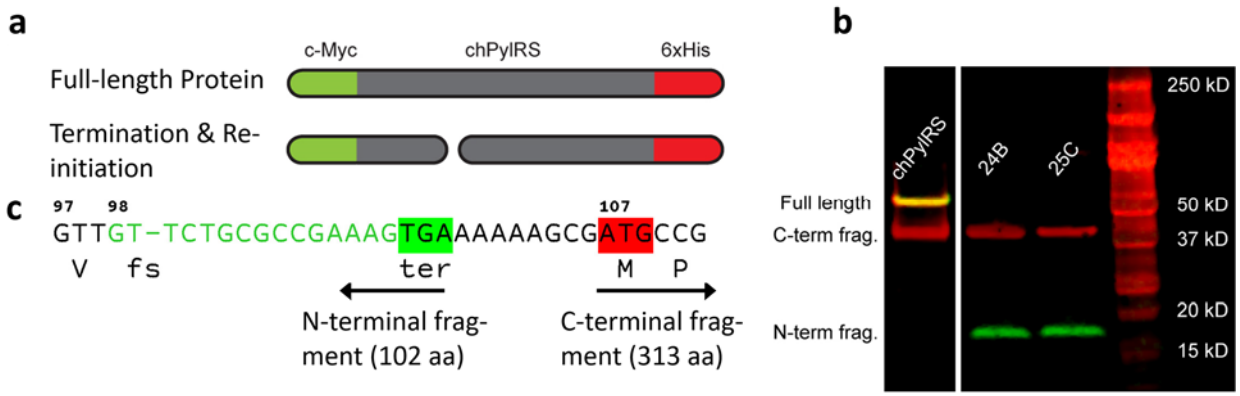
Supplementary Figure 4 | PylRS facilitates read-through of M13 *gIII*. PylRS forms BocK-$tRNA^{Pyl}_{am}$, which is then recruited to the ribosome by EF-Tu, enabling read-through of UAG codons by BocK.



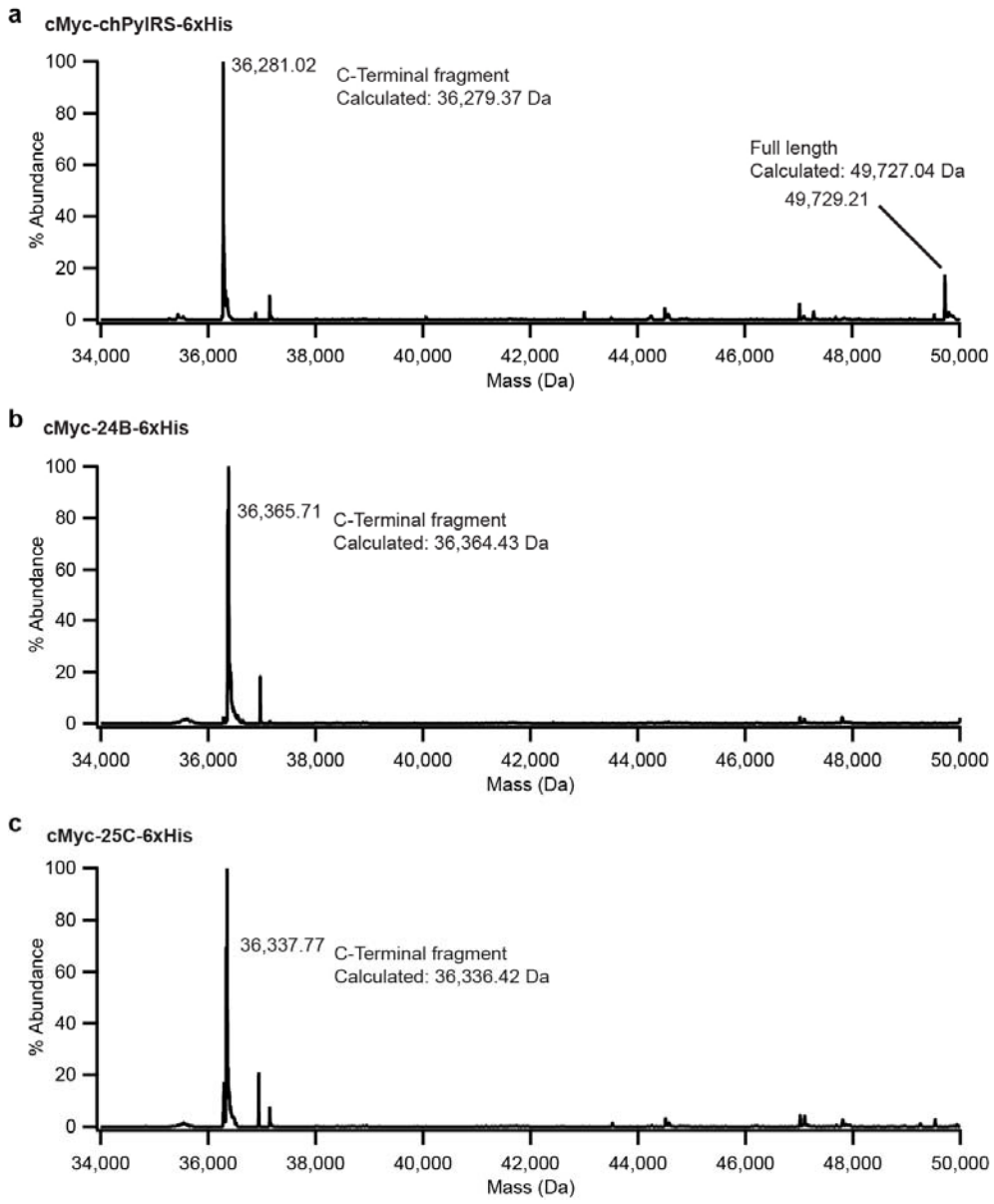
Supplementary Figure 5 | Range of substrate specificity of two MmPylRS enzymes. Wild-type MmPylRS and MmPylRS-L301M/Y306L/L309A/C348F[2]. Suppression of the sfGFP-UAG2 gene by the library of ncAA-tRNA^{Pyl} was measured by fluorescence intensity. Note that the PylRS variant is much less active than wild-type PylRS. A library of 313 ncAAs (see Supplementary Table 2 in reference[3]) was tested. Fluorescence signals from the incorporation of ncAAs **1–12** are labeled. Well A1 is a control without added ncAA. Inserted chemical structures of ncAAs used in this figure: **1**, L-pyrrolysine (Pyl); **2**, N$^{\varepsilon}$-acetyl-L-Lys (AcK); **3**, N$^{\varepsilon}$-trifluoroacetyl-L-Lys (CF3-AcK); **4**, 3-iodo-L-Phe (3-I-Phe); **5**, 3-bromo-L-Phe (3-Br-Phe); **6**, 3-chloro-L-Phe (3-Cl-Phe); **7**, 3-trifluoromethyl-L-Phe (3-CF3-Phe); **8**, 3-methyl-L-Phe (3-Me-Phe); **9**, 3-methoxyl-L-Phe (3-MeO-Phe); **10**, 3-bromo-L-ThA (3-Br-ThA); **11**, N$^{\varepsilon}$-allyloxycarbonyl-L-Lys (AllocK); **12**, N$^{\varepsilon}$-t-butyloxycarbonyl-L-Lys (BocK).



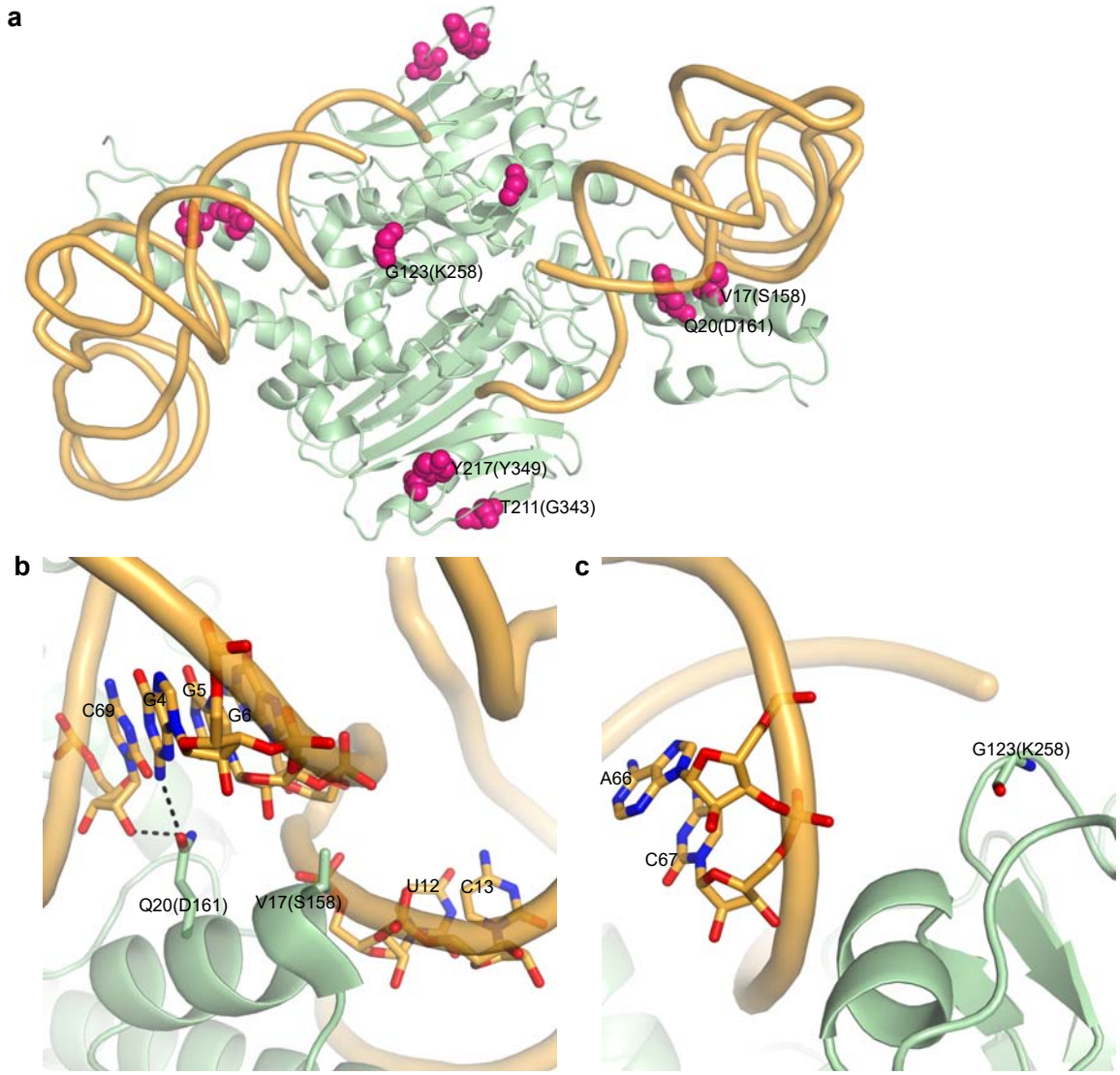
Supplementary Figure 6 | sfGFP expression facilitated by PANCE-evolved chPylRS variants. The relative expression of (**a**) sfGFP containing a single UAG codon at position 2 (Ser) (**b**) sfGFP(3xTAG) containing three UAG codons at positions 39 (Asn), 135 (Asn), and 151 (Tyr) was compared in the presence or absence of 1 mM BocK. (**c**) The relative expression of sfGFP containing a single UAG codon at position 2 (Ser) mediated by either the split enzyme, 24B or the individual N- or C-terminal fragment of 24B in the presence or absence of 1 mM BocK. Each value and error bar reflects the mean and s.d. of four independent replicates.



Supplementary Figure 7 | Western blot analysis of full-length and split PANCE-evolved chPylRS variants. (**a**) 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 to characterize translation of stop codon-containing mutants that arose during PANCE. (**b**) Western blot analysis of the protein lysates expressed in BL21 star (DE3) cells. Wild-type chPylRS is expressed as a full-length protein, but the presence of this internal start site also generates the 313 aa C-terminal fragment. The split variants (24B and 25C) are expressed as distinct N- and C-terminal fragments. (**c**) Nucleotide sequence of genes encoding 25C from codon 97 (V) to codon 108 (P). fs, denotes the position of the -1 frameshift (deletion of T); ter, indicates the UGA stop codon releasing the N-terminal fragment. Initiation of the C-terminal fragment occurs at codon 107 (M).



Supplementary Figure 8 | ESI-MS analysis of Ni-NTA-purified affinity-tagged PANCE-evolved chPylRS variants. (**a-c**) The PylRS variants, chPylRS (**a**), 24B (**b**), and 25C (**c**) containing an N-terminal c-Myc-tag and a C-terminal 6xHis-tag analyzed by ESI-MS analysis. In the split variants of chPylRS, the N-terminal fragment is lost upon affinity purification over Ni-NTA resin. Protein was expressed in BL21 star (DE3) cells in LB medium. The major peaks in each spectrum were in agreement with the calculated mass of the full-length enzyme, chPylRS (**a**), or the C-terminal fragment resulting from reinitiation at codon 107 (M) (**b-c**).



Supplementary Figure 9 | Structural insight into PANCE mutations in the chPylRS CTD. (**a**) Residues corresponding to PANCE mutations are mapped to the bacterial *D. hafniense* PylRS•tRNA^{Pyl} structure (PDB ID 2ZNI) as pink spheres. V17, Q20 (**b**), and G123 (**c**) in the *D. hafniense* CTD correspond to D161, S158 and K258 in chPylRS; nucleotides in the vicinity of these residues are shown as sticks. Residue numbers corresponding to PANCE mutations are shown in parentheses. Polar interactions between *D. hafniense* PylRS and tRNA^{Pyl} are shown in black dashed lines.

V17 (S158) lies close to the phosphate group of G6, U12 and C13 in acceptor and D stem of tRNA. Thus, the S158N mutation may increase the affinity for tRNA^{Pyl} by interacting with the corresponding nucleotides in archaeal tRNA^{Pyl}. The Oε1 of Q20 (D161) interacts with N2 of G4, while Nε2 interacts with 02' of C69 in the acceptor stem. Thus, Nδ2 of N161 may form a similar interaction with 02'of A69 of archaeal tRNA^{Pyl} thereby increasing the affinity to tRNA. G123 (K258) is close to the phosphate group of A66 or C67 in the acceptor stem. Thus, the mutant Q258 may interact with the phosphate group of corresponding residues in archaeal tRNA^{Pyl}. T211 (G343) is far from tRNA^{Pyl}; the effect of this mutation is unknown.

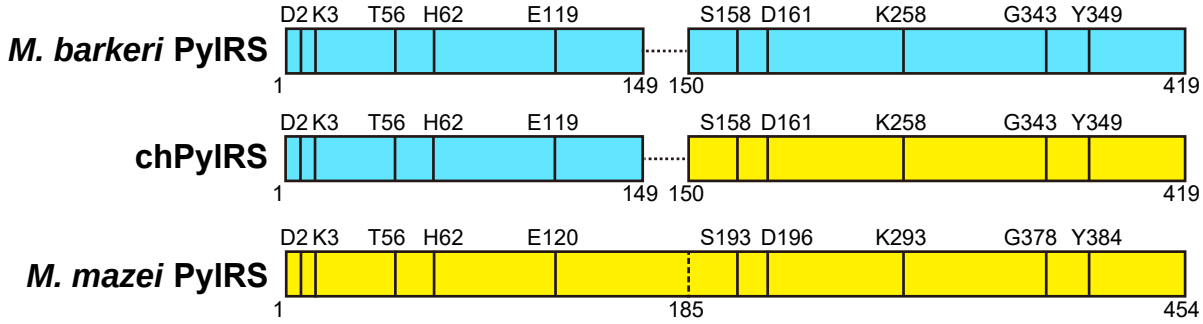
Supplementary Table 1 | Apparent kinetic parameters of PylRS and chPylRS for Pyl acylation

PylRS	ncAA	K_M (μM)	k_{cat} ($10^{-3}s^{-1}$)	k_{cat}/K_M ($10^{-3}s^{-1}\mu M^{-1}$)
MmPylRS	Pyl	21 ± 4	4.9 ± 0.2	0.23
MbPylRS	Pyl	16 ± 2	30 ± 0.4	1.9
chPylRS	Pyl	7.6 ± 0.2	11 ± 1	1.4

The three enzymes were overexpressed in BL21(DE3) cells induced by 1 mM IPTG for 4 hr at 37ºC when cell density reached A_{600} = 0.6. They were purified using an N-terminal His_6-tag by Ni-NTA. The purification of these enzymes is similar to the method described in **ONLINE METHODS** for c-Myc-chPylRS-6xHis variants, except the lysis buffer contains 1 M NaCl. The protein yield for the three enzymes is similar, around 10 mg/L of LB medium. Pure MmPylRS or MbPylRS precipitated from solution when the concentration exceeded ~2 mg/mL, while pure chPylRS could be concentrated to >15 mg/mL; this was a prerequisite for our crystallographic work on the full-length enzyme. The kinetic parameters were measured as described in **ONLINE METHODS**. The data were derived from three replicates. *M. mazei* $tRNA^{Pyl}$ was used for aminoacylation, and was produced by *in vitro* transcription. After purification as described in **ONLINE METHODS**, tRNA was refolded by heating to 100ºC for 5 min in deionized water, and slowly cooled down to room temperature for 30 min.

Supplementary Table 2 | Genotypes of mutant chPylRS variants resulting from this work.

chPylRS variant	Mutations observed											
32A	D2N	K3N	T56P	H62Y			E119K			K258Q		Y349F
32A-Nter†	D2N	K3N	T56P	H62Y								
24B		K3N	T56P	H62Y	Δt293	R100N*		S158N			G343D	
25C		K3N	T56A	H62Y	Δt293				D161N		G343D	



†32A-Nter was constructed by inserting the PANCE-evolved four N-terminal mutations into wild-type chPylRS. As mutation Δt293 caused a -1 frameshift (resulting in chain termination, and use of the AUG107 codon as translation initiation site) and production of a split protein, other mutations are shown relative to the wild-type protein sequence within the corresponding coding frame. *Mutation R100N is listed as relative to the frameshifted protein sequence (see **Supplementary Note** for detailed discussion of frameshifted region). The schematic sequence comparison of *M. barkeri* PylRS, chPylRS and *M. mazei* PylRS relates the positions of the PANCE evolved mutations in the established numbering systems of the three different PylRS enzymes.

Supplementary Table 3 | Crystallographic data collection and refinement statistics

PylRS•tRNA^{Pyl} complex	SeMet	MmPylRS NTD	32A NTD
PDB ID		5UD5	5V6X
Data collection			
Space group	$P3_121$	$P3_121$	$P3_121$
Cell dimensions			
a, *b*, *c* (Å)	72.1, 72.1, 239.0	72.3, 72.3, 239.0	72.6, 72.6, 238.6
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Wavelength (Å)	0.979190	0.979200	0.979200
Resolution (Å)	49.13–2.82 (2.99–2.82)*	49.22–2.35 (2.49–2.35)	49.33–2.76 (2.92–2.76)
R_{meas}	0.100 (1.339)	0.059 (1.403)	0.074 (1.312)
$<I/\sigma(I)>$	19.63 (1.96)	22.57 (1.88)	25.7 (1.88)
$CC_{1/2}$	0.999 (0.927)	0.999 (0.779)	1.000 (0.880)
Completeness (%)	99.8 (99.2)	99.7 (98.8)	99.9 (99.8)
Redundancy	11.5 (11.3)	9.8 (10.1)	9.7 (10.1)
Refinement			
Resolution (Å)		49.22–2.35	49.33–2.76
No. reflections		31207	19633
R_{work} / R_{free}		0.216/0.242	0.206/0.249
No. atoms			
Protein		1392	1368
Ligand/ion		2	2
Water		24	-
RNA		3020	3020
B-factors			
Protein		71.5	79.6
Ligand/ion		70.5	74.5
Water		58.6	-
RNA		88.4	95.7
R.m.s. deviations			
Bond lengths (Å)		0.004	0.003
Bond angles (°)		0.835	0.636

*Values in parentheses are for highest-resolution shell.

Supplementary Table 4 | Plasmids used in this work

Plasmid	Class	Origin	ORF1		ORF2		Reference
Name	(resistance)		Prom	[RBS][4] Genes	Prom	Genes	
pJC175e	AP (carb^R)	SC101	P_{psp}	[SD8] gIII, luxAB	–	–	([5])
pDB038	AP (spec^R)	ColE1	P_{psp}	[SD8] gIII(P29am), luxAB	P_{ProK}	pylT	This work
pDB038a	AP (spec^R)	ColE1	P_{psp}	[SD8] gIII(P29am, Y184am), luxAB	P_{ProK}	pylT	This work
pDB038b	AP (spec^R)	ColE1	P_{psp}	[SD8] gIII(P29am,P83am, Y184am), luxAB	P_{ProK}	pylT	This work
MP6	MP (chlor^R)	cloDF13	P_{psp}	dnaQ926, dam, seqA	P_C	araC	([6])
SP-chPylRS	SP (none)	M13 f1	P_{gIII}	[SD4] chPyl	–	–	This work

Plasmids pDB038, pDB038a, and pDB038b contain *gIII* with mutations P29am, P29am/E84am, and P29am/E84am/Y183am, respectively. Selection phage SP-chPylRS embodies a mutant M13 phage with chPylRS inserted in place of *gIII*, as described in this work.

SUPPLEMENTARY NOTE

DNA and amino acid equences of ChPylRS variants (with mutations in red)

chPylRS (ancestor)

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACG
CTGCACAAGATCAAGCACTATGAGGTTTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGAC
CATCTGGTTGTGAACAACTCTCGTTCTTGTCGTACCGCACGTGCATTCCGTCATCATAAATACCGT
AAAACCTGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACTTCCTGACCCGTTCTACCGAA
GGCAAAACCTCTGTTAAAGTTAAAGTTGTTTCTGCGCCGAAAGTGAAAAAAGCGATGCCGAAATCT
GTTTCTCGTGCGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGT
TCTGTTCCGTCTCCGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCA
CTTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTTCCCTGAAT
TCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAAGACCTGCAGCAG
ATCTACGCGGAAGAAAGGGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTG
GACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGC
ATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCC
ATGCTTGCTCCAAACCTTTACAACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAA
ATTTTTGAAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACC
ATGCTGAACTTCTGCCAGATGGGATCGGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGAC
TTCCTGAACCACCTGGGAATTGATTTCAAGATCGTAGGCGATTCCTGCATGGTCTATGGGGATACC
CTTGATGTAATGCACGGAGACCTGGAACTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGG
GAATGGGGTATTGATAAACCCTGGATAGGGGCAGGTTTCGGACTCGAACGCCTTCTAAAGGTTAAA
CACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACCAAC
CTGTAA

chPylRS (ancestor)

MDKKPLDVLISATGLWMSRTGTLHKIKHYEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYR
KTCKRCRVSDEDINNFLTRSTEGKTSVKVKVVSAPKVKKAMPKSVSRAPKPLENPVSAKASTDTSR
SVPSPAKSTPNSPVPTSASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQ
IYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRP
MLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITD
FLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVK
HDFKNIKRAARSESYYNGISTNL

Evolved variant 32A

ATG**A**ATAA**T**AAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACG
CTGCACAAGATCAAGCACTATGAGGTTTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGAC
CATCTGGTTGTGAACAACTCTCG**C**TCTTG**C**CGTC**C**CGCACGTGCATTCCGT**T**ATCATAAATACCGT
AAAACCTGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACTTCCTGACCCGTTCTACCGAA

GGCAAAACCTCTGTTAAAGTTAAAGTTGTTTCTGCGCCGAAAGTGAAAAAAGCGATGCCGAAATCT
GTTTCTCGTGCGCCGAAACCGCTGAAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGT
TCTGTTCCGTCTCCGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCA
CTTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTTCCCTGAAT
TCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAAGACCTGCAGCAG
ATCTACGCGGAAGAAAGGGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTG
GACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGC
ATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACCAGAACTTCTGCCTGAGACCC
ATGCTTGCTCCAAACCTTTACAACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAA
ATTTTTGAAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACC
ATGCTGAACTTCTGCCAGATGGGATCGGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGAC
TTCCTGAACCACCTGGGAATTGATTTCAAGATCGTAGGCGATTCCTGCATGGTCTTTGGGGATACC
CTTGATGTAATGCACGGAGACCTGGAACTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGG
GAATGGGGTATTGATAAACCCTGGATAGGGGCAGGTTTCGGACTCGAACGCCTTCTAAAGGTTAAA
CACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACCAAC
CTGTAA

Evolved variant 32A

MNNKPLDVLISATGLWMSRTGTLHKIKHYEVSRSKIYIEMACGDHLVVNNSRSCRPARAFRYHKYR
KTCKRCRVSDEDINNFLTRSTEGKTSVKVKVVSAPKVKKAMPKSVSRAPKPLKNPVSAKASTDTSR
SVPSPAKSTPNSPVPTSASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQ
IYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDQNFCLRP
MLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITD
FLNHLGIDFKIVGDSCMVFGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVK
HDFKNIKRAARSESYYNGISTNL

Evolved variant 24B

ATGGATAATAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACG
CTGCACAAGATCAAGCACTATGAGGTTTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGAC
CATCTGGTTGTGAACAACTCTCGTTCTTGTCGTCCCGCACGTGCATTCCGTTATCATAAATACCGT
AAAACCTGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACTTCCTGACCCGTTCTACCGAA
GGCAAAACCTCTGTTAAAGTTAAAGTTGTT-CTGAACCGAAAGTGAAAAAAGCGATGCCGAAATCT
GTTTCTCGTGCGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGT
TCTGTTCCGTCTCCGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCA
CTTACGAAGAACCAGACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTTCCCTGAAT
TCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAAGACCTGCAGCAG
ATCTACGCGGAAGAAAGGGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTG
GACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGC
ATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCC
ATGCTTGCTCCAAACCTTTACAACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAA

ATTTTTGAAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACC
ATGCTGAACTTCTGCCAGATGGGATCGGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGAC
TTCCTGAACCACCTGGGAATTGATTTCAAGATCGTAGACGATTCCTGCATGGTCTATGGGGATACC
CTTGATGTAATGCACGGAGACCTGGAACTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGG
GAATGGGGTATTGATAAACCCTGGATAGGGGCAGGTTTCGGACTCGAACGCCTTCTAAAGGTTAAA
CACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACCAAC
CTGTAA

Evolved variant 24B: N-terminal fragment

MDNKPLDVLISATGLWMSRTGTLHKIKHYEVSRSKIYIEMACGDHLVVNNSRSCRPARAFRYHKYR
KTCKRCRVSDEDINNFLTRSTEGKTSVKVKVVLNRK-

Evolved variant 24B: C-terminal fragment

MPKSVSRAPKPLENPVSAKASTDTSRSVPSPAKSTPNSPVPTSASAPALTKNQTDRLEVLLNPKDE
ISLNSGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYI
ERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHL
EEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVDDSCMVYGDTLDVMHGDLELSSAVVGPI
PLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL-

Evolved variant 25C:

ATGGATAATAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACG
CTGCACAAGATCAAGCACTATGAGGTTTCTCGTTCTAAAATCTACATCGAAATGGCGTGTGGTGAC
CATCTGGTTGTGAACAACTCTCGTTCTTGTCGTGCCGCACGTGCATTCCGTTATCATAAATACCGT
AAAACCTGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACAACTTCCTGACCCGTTCTACCGAA
GGCAAAACCTCTGTTAAAGTTAAAGTTGTT-CTGCGCCGAAAGTGAAAAAAGCGATGCCGAAATCT
GTTTCTCGTGCGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACCGACACCTCTCGT
TCTGTTCCGTCTCCGGCGAAATCTACCCCGAACTCTCCGGTTCCGACCTCTGCAAGTGCCCCCGCA
CTTACGAAGAGCCAGACTAACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGAGATTTCCCTGAAT
TCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCTCTCTCGCAGAAAAAAAGACCTGCAGCAG
ATCTACGCGGAAGAAAGGGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTG
GACAGGGGTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGC
ATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCC
ATGCTTGCTCCAAACCTTTACAACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATAAAA
ATTTTTGAAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACC
ATGCTGAACTTCTGCCAGATGGGATCGGGATGCACACGGGAAAATCTTGAAAGCATAATTACGGAC
TTCCTGAACCACCTGGGAATTGATTTCAAGATCGTAGACGATTCCTGCATGGTCTATGGGGATACC
CTTGATGTAATGCACGGAGACCTGGAACTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGG
GAATGGGGTATTGATAAACCCTGGATAGGGGCAGGTTTCGGACTCGAACGCCTTCTAAAGGTTAAA
CACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACCAAC
CTGTAA

Evolved variant 25C: N-terminal fragment

MD**N**KPLDVLISATGLWMSRTGTLHKIKHYEVSRSKIYIEMACGDHLVVNNSRSCR**A**ARAFR**Y**HKYR
KTCKRCRVSDEDINNFLTRSTEGKTSVKVKVV**LRR**K**–**

Evolved variant 25C: C-terminal fragment

MPKSVSRAPKPLENPVSAKASTDTSRSVPSPAKSTPNSPVPTSASAPALTKSQT**N**RLEVLLNPKDE
ISLNSGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYI
ERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHL
EEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIV**D**DSCMVYGDTLDVMHGDLELSSAVVGPI
PLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

Effects of frameshift mutations on scrambled regions of chPylRS. In sample 24B, mutation ∆t293 results in a frameshift beginning at residue V98. This causes in a synonymous change at residue 98, and nonsynonymous changes at residues 99-103 (SAPKV becomes LNRK-Opal). Within this frameshifted region, there are two additional mutations: c299a, and g300a. This leads to mutation R100N within the frameshifted region; in the absence of the frameshift however, these two nucleotide mutations fall in the same codon causing mutation A100E (note that the non-frameshifted wild-type sequence has an 'A' residue at position 100). As A100E has been shown[7] to bestow a benefit, it is likely that mutation A100E occurred within this lineage prior to the appearance of the frameshift. Similarly, within sample 25C, mutation ∆t293 results in a frameshift beginning at residue V98. This causes a synonymous change at residue 98, and nonsynonymous changes at residues 99-103 (SAPKV becomes LRRK-Opal).

1. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).
2. Umehara, T. et al. N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett* **586**, 729-733 (2012).
3. Ko, J.H. et al. Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS Lett* **587**, 3243-3248 (2013).
4. Ringquist, S. et al. Translation initiation in *Escherichia coli:* sequences within the ribosome-binding site. *Mol Microbiol* **6**, 1219-1229 (1992).
5. Badran, A.H. et al. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **533**, 58-63 (2016).
6. Badran, A.H. & Liu, D.R. Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. *Nat Commun* **6**, 8425 (2015).
7. Bryson, D. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat Chem Biol*, aa-bb (2017).