

Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient *In Vivo* Incorporation of Structurally Divergent Non-Native Amino Acids

Jeffrey K. Takimoto^{1,†}, Nikki Dellas^{1,†}, Joseph P. Noel^{1,2} & Lei Wang^{1,*}

¹The Jack H. Skirball Center for Chemical Biology & Proteomics, ² Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California, 92037, U.S.A.

Supporting Information

DNA sequences:

Sperm whale myoglobin (Myo4TAG)

```
atgactagtgttctgtaggaaggtgaatggcagctggttctgcatggttgggctaagttgaagctgacg
tcgctggatcatgggcaggacatcttgattcgactgttcaaactctcatccggaaactctggaaaaattcga
tcggttcaaacatctgaaaactgaagctgaaatgaaagcttctgaagatctgaaaaacatgggtgttacc
gtgtaactgcccaggtgctatccttaagaaaaaggcatcatgaagctgagctcaaaccgcttgac
aatcgcatgctactaaacataagatcccgatcaaaactggaattcatctctgaagcgatcatccatgt
tctgcattctagacatccaggtgacttcgggtgctgacgctcaggggtgctatgaacaaagctctcgagctg
ttccgtaaagatatcgctgctaagtaacaaagaactgggttaccagggatctcatcaccatcaccatcact
aa
```

GFP (182TAG)

```
atgactagtgtgagcaagggcgaggagctgttcaccgggggtggtgcccatcctggtcgagctggacggcg
acgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacct
gaagttcatctgcaccaccggcaagctgcccgtgccctggcccacctcgtgaccacctgacctacggc
gtgcagtgcttcagccgctaccccgaccacatgaagcagcagcacttcttcaagtccgccatgccgaag
gctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagaccgcgccgaggtgaagtt
cgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctg
gggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggca
tcaaggccaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactagcagca
gaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcaccctcagcctg
agcaagacccccaacgagaagcgcgatcacatggctcctgctggagttcgtgaccgcccgggatcactc
tcggcatggacgagctgtacaagagatctcatcaccatcaccatcactaa
```

MmPy1RS (Y384F)

```
atggataaaaagcctctgaacactctgatttctgacccggctctgtggatgtcccgcaccggcaccatcc
acaaaatcaaacaccatgaagtttagcgttccaaaatctacattgaaatggcttgccggcgatcacctggt
tgtcaacaactccggttcttctcgtaccgctcgcgactgcgccaccacaaatctcgcaaaacctgcaaa
cgttgccgtgttagcgtgagggacctgaacaaattcctgaccaaagctaaccgaggtcagacctccgtaa
aagtgaaaggtagtaagcgtccgaccgctactaaaaaggctatgccaaaaagcgtggcccgtgcccggaa
acctctggaaaaacaccgagggcggtcaggtcaaccatccgggttctaaatcttctccggcgatcccagtg
tccaccaagaatctgttccgtaccagcaagcgtgtctaccagcatttagcagcatttctaccgggtgcta
ccgcttctgcgctggtaaaaggtaacactaacccgattactagcatgtctgcaccgggtacaggcaagcgc
cccagctctgactaaatcccagacggaccgtctggaggtgctgctgaaccacaaaggatgaaatctctctg
aacagcggcaagccttccgtagctggaagcgcagctgctgctcgtcgtaaaaaggatctgcaacaga
tctacgctgaggaacgcgagaactatctgggtaagctggagcgcgaaattactcgcttcttctcgtggatcg
```

cggtttcctggagatcaaactctccgattctgattccgctggaatacattgaacgatggggcatcgataat
gataccgaactgtctaaacagatcttccgtgtggataaaaacttctgtctgcgtccgatgctggccccga
acctgtacaactatctgcgtaaaactggaccgtgccctgccggaccgatcaaaatcttcgagatcgggcc
ttgctaccgtaagagtcgacggttaagagcaccctggaagaattcaccatgctgaacttctgccagatg
ggtagcgggttgacgcgtgaaaacctggaatccattatcaccgacttctgaatcacctgggtatcgatt
tcaaaattgttggtgacagctgtatggtgtttggcgatacgcctggatggtatgcacggcgatctggagct
gtcttccgcagtagtgggccaatcccgcctggatcgtgagtggggtatcgacaaaccttggatcggtgcg
ggttttggctctggagcgtctgctgaaagtaaaacacgacttcaagaacatcaaacgtgctgcacgttccg
agtcctattacaatggtatttctactaacctgtaa

MmPylT (*M. mazei* tRNA^{Pyl}_{CUA})

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccgggggttccgc
ca

MmOmeRS

atggataaaaagcctctgaacactctgatttctgcgaccgggtctgtggatgtcccgcaccggcaccatcc
acaaaatcaaacaccatgaagttagccgttccaaaatctacattgaaatggcttgcggcgatcacctggt
tgtcaacaactcccgttcttctcgtaccgctcgcgcaactgcgccaccacaaatcgcaaaacctgcaaa
cgttgcgctgtagcgatgaggacctgaacaaattcctgaccaaagctaacgaggatcagacctccgtaa
aagtgaaggtagtaagcgtccgaccgtaactaaaaaggctatgccaaaagcgtggcccggtgccccgaa
acctctggaaaacaccgagggcggtcaggctcaaccatccgggttctaaatcttctccggcgatcccagtg
tccaccaagaatctgtttccgtaccagcaagcgtgtctaccagcatttagcagcatttctaccggtgcta
ccgcttctgcgctggtaaaaggtaacactaaccgattactagcatgtctgcaccggtagcaggcaagcgc
cccagctctgactaaatcccagacggaccgtctggaggtgctgctgaaccacaaaggatgaaatctctctg
aacagcggcaagccttccgtgagctggaaagcgagctgctgtctcgtcgtaaaaaggatctgcaacaga
tctacgctgaggaacgcgagaactatctgggtaagctggagcgcgaaattactcgttcttctggtggatcg
cggtttctcctggagatcaaactctccgattctgattccgctggaatacattgaacgatggggcatcgataat
gataccgaactgtctaaacagatcttccgtgtggataaaaacttctgtctgcgtccgatgctgactccga
acctgtacaactatcttctgtaaaactggaccgtgccctgccggaccgatcaaaatcttcgagatcgggtcc
ttgctaccgtaagagtcgacggttaagagcaccctggaagaattcaccatgctggttttctggcagatg
ggtagcgggttgacgcgtgaaaacctggaatccattatcaccgacttctgaatcacctgggtatcgatt
tcaaaattgttggtgacagctgtatggtgtttggcgatacgcctggatggtatgcacggcgatctggagct
gtcttccgcattggtgggccaatcccgcctggatcgtgagtggggtatcgacaaaccttggatcggtgcg
ggttttggctctggagcgtctgctgaaagtaaaacacgacttcaagaacatcaaacgtgctgcacgttccg
agtcctattacaatggtatttctactaacctgtaa

Primers for cloning:

GFP182TAG

DJ654 5' -GCA GAT CTC TTG TAC AGC TCG TCC ATG CC- 3'
DJ655 5' -ATG ACT AGT GTG AGC AAG GGC GAG G- 3'

Myo4TAG

JT762 5' -GAG ACT AGT GTT CTG TAG GAA GGT GAA TGG CAG CTG GTT
C- 3'
JT763 5' -GAG GGA TCC CTG GTA ACC CAG TTC TTT GTA CTT AGC
AG- 3'

pBK5-MmPylRS

JT552 5' -GTG TAC ACA TAT GGA TAA AAA GCC TCT GA- 3'
JT632 5' -GCC GTG TTA GCG ATG AGG ACC TGA ACA AAT TCC TGA
CC- 3'

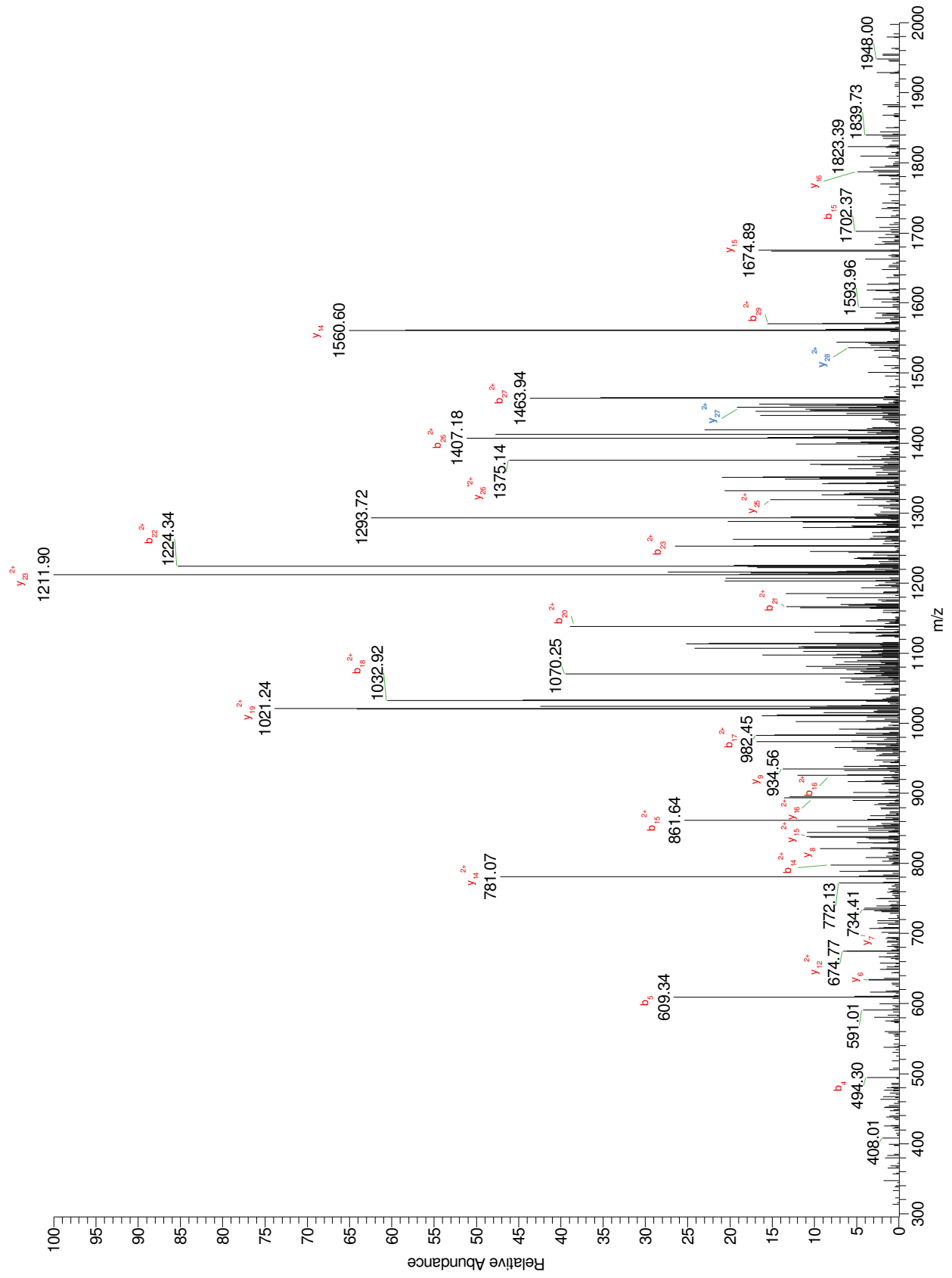
JT633 5' -GGT CAG GAA TTT GTT CAG GTC CTC ATC GCT AAC ACG
GC- 3'
JT634 5' -GAA ACT GCA GCC ACC CGC CTG GTT ACA GGT TAG TAG
AAA TAC CAT TGT AAT AGG- 3'

Library construction (A302, L309, N346, C348, V401, and W417)

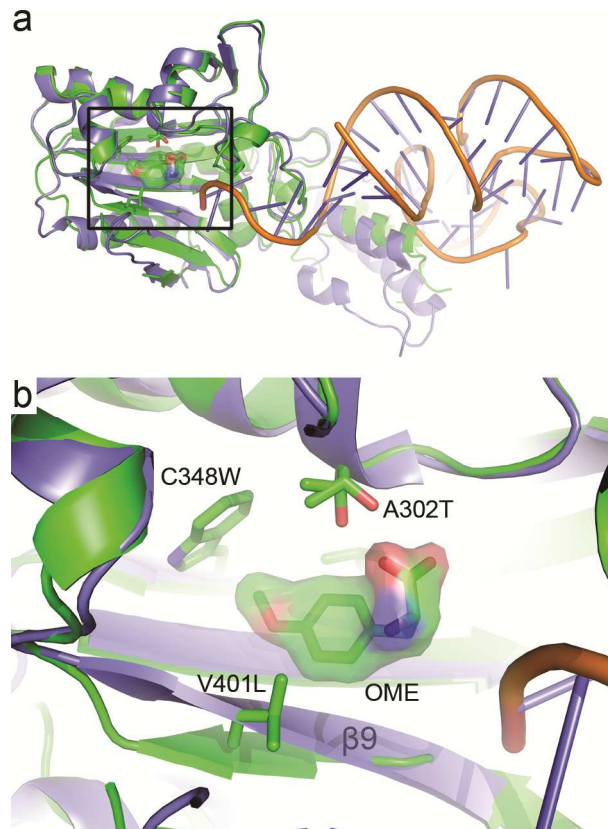
JT552 5' -GTG TAC ACA TAT GGA TAA AAA GCC TCT GA- 3'
JT620 5' -CAG GGC ACG GTC CAG TTT ACG MNN ATA GTT GTA CAG
GTT CGG MNN CAG CAT C- 3'
JT621 5' -AAA CTG GAC CGT GCC CTG- 3'
JT624 5' -GAT CCA GCG GGA TTG GGC CCA CMN NTG CGG AAG- 3'
JT634 5' -GAA ACT GCA GCC ACC CGC CTG GTT ACA GGT TAG TAG
AAA TAC CAT TGT AAT AGG- 3'
JT625 5' -GGC CCA ATC CCG CTG GAT CGT GAG TGG GGT ATC GAC
AAA CCT NNK ATC GGT GC- 3'

pMPcua-OmeRS

JT768 5' -CAC AGA TCT CCG GAA ACC TGA TCA TGT AGA TCG AAT
GGA C- 3'
JT769 5' -GAG AAG CTT CAA AAA AAG GAC CCG GAA ACC CCG GGA
ATC TAA C-3'



Supplementary Figure 1. Tandem mass spectrum of peptide ^+H_3N -HNIEDGSVQLADHXQQ NTPIGDGPVLLPDNHYLSTQSALS K -CO $_2$ H. X represents the UAG encoded site at position 182 of GFP. The sequence of the peptide containing Ome at position X was unambiguously assigned from the annotated b and y ion series.



Supplementary Figure 2. Superposition of *M. mazei* OmeRS complexed with Ome and AMP-PNP to *D. hafniense* PylRS complexed with tRNA (PDB ID 2ZNI). **a)** global view of the PylRS crystal structure (tRNA colored orange and PylRS colored purple) superposed to the OmeRS crystal structure (colored green). **b)** close-up view of the active sites for both synthetases. The Ome substrate and all four mutated residues are displayed for the OmeRS structure.