RESEARCH ARTICLE

Enhanced Directed Evolution in Mammalian Cells Yields a Hyperefficient Pyrrolysyl tRNA for Noncanonical Amino Acid Mutagenesis

Delilah Jewel, [a] Rachel E. Kelemen, [a] Rachel L. Huang, [a] Zeyu Zhu, [b] Bharathi Sundaresh, [b] Kaitlin Malley, [a] Quan Pham, [a] Conor Loynd, [a] Zeyi Huang, [a] Tim van Opijnen, [b] [c] and Abhishek Chatterjee[a]*

[a] D. Jewel, Dr. R.E. Kelemen, R. Huang, K. Malley, Q. Pham, C. Loynd, Z. Huang, Prof. A. Chatterjee
Department of Chemistry,
Boston College
2609 Beacon Street, Chestnut Hill, Massachusetts 02467
E-mail: abhishek.chatterjee@bc.edu

[b] Dr. Z. Zhu, B. Sundaresh, Prof. T.v. Opijnen
Department of Biology
Boston College
Chestnut Hill, MA 02467, USA

[c] Prof. T.v. Opijnen
Broad Institute of MIT and Harvard,
Cambridge, MA 02142, USA

Supporting information for this article is given via a link at the end of the document.

Abstract: Heterologous tRNAs used for noncanonical amino acid (ncAA) mutagenesis in mammalian cells typically show poor activity. We recently introduced a virus-assisted directed evolution strategy (VADER) that can enrich improved tRNA mutants from naïve libraries in mammalian cells. However, VADER was limited to processing only a few thousand mutants; the inability to screen a larger sequence space precluded the identification of highly active variants with distal synergistic mutations. Here, we report VADER2.0, which can process significantly larger mutant libraries. It also employs a novel library design, which maintains base-pairing between distant residues in the stem regions, allowing us to pack a higher density of functional mutants within a fixed sequence space. VADER2.0 enabled simultaneous engineering of the entire acceptor stem of *M. mazei* pyrrolysyl tRNA ($tRNA^{Pyl}$), leading to a remarkably improved variant, which facilitates more efficient incorporation of a wider range of ncAAs, and enables facile development of viral vectors and stable cell-lines for ncAA mutagenesis.

Introduction

Site-specific incorporation of noncanonical amino acids (ncAAs) with unique chemical functionalities into proteins expressed in mammalian cells is a powerful technology with numerous applications.[1] This is achieved using an engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pair, which typically incorporates the ncAA in response to a nonsense codon. To avoid cross-reaction with endogenous aaRS/tRNA pairs, the ncAA-selective pair is imported into mammalian cells from bacteria or archaea.[1] However, such heterologous tRNAs are often poorly active in the new host, likely due to their suboptimal interaction with non-native components from the host cell, including tRNA biogenesis and processing machinery, elongation factors, and the ribosome.[2]

Traditionally, the lack of intrinsic efficiency of these tRNAs has been compensated for through overexpression.[2-3] It has been shown that hundreds of tRNA-expressing genes are needed per cell to achieve efficient ncAA incorporation.[2a, 2c, 3] Although this can be accomplished in readily transfected cell lines through transient transfection of plasmids encoding multiple copies of the tRNA gene, this requirement makes it more challenging to apply this technology to cells that are not easily transfected. It also creates significant complications for the development of mammalian cells with stably integrated the ncAA incorporation machinery, which is essential for both commercial and basic science applications. Previous studies have shown that the development of such stable cell lines requires the integration of >400 tRNA genes per cell for efficient ncAA incorporation, which is inherently challenging to perform, and the resulting cell lines often show poor long-term stability and significant loss of ncAA incorporation activity over time.[3-4] Access to more efficient ncAA-incorporating tRNAs is crucial to overcome these challenges.

Our understanding of factors that directly contribute to the suboptimal activity of heterologous tRNAs in mammalian cells is quite limited, which makes it challenging to address this issue rationally. In *E. coli*, directed evolution has been used with much success to engineer enhanced ncAA-incorporating tRNAs.[5] However, similar directed evolution approaches in mammalian cells have not been available. Although semi-rational approaches have been employed with some success to improve the activity of orthogonal suppressor tRNAs in mammalian cells, such approaches are difficult to generalize.[6] Recently, we reported a novel platform for virus-assisted directed evolution of tRNAs

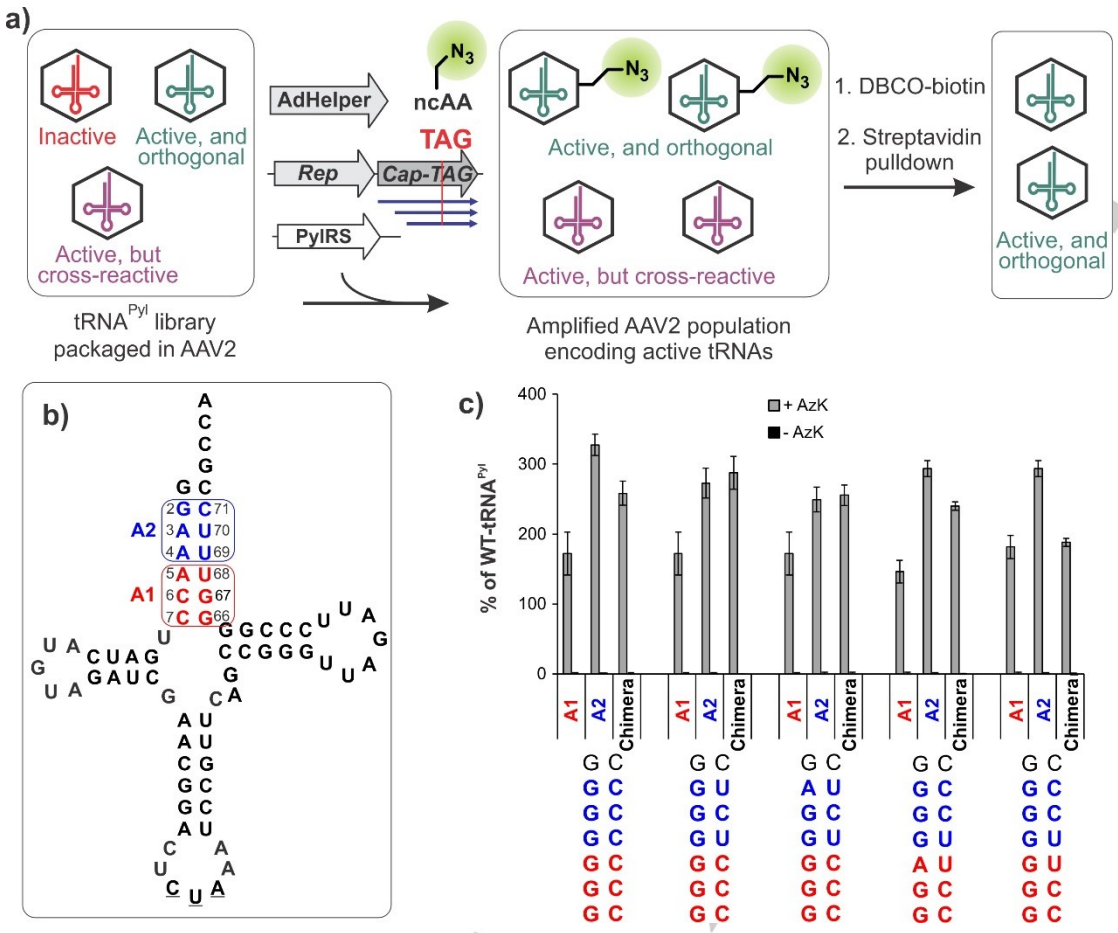


Figure 1. (a) The VADER selection scheme. AAV encoding a synthetic suppressor tRNA library is used to infect mammalian cells at low multiplicity of infection (MOI). Active and orthogonal tRNAs enable viral replication by suppressing a TAG codon in the capsid gene, and introduce an azide-ncAA at a surface-exposed site of the progeny virus, which are selectively isolated by bioorthogonal attachment of biotin and streptavidin pulldown. (b) Structure of tRNA^{Pyl} highlighting the acceptor stem residues randomized to generate the A1 and A2 libraries. (c) Chimeras constructed by combining beneficial mutations from A1 or A2 libraries do not show further enhancement in activity. The activity of each tRNA^{Pyl} variant was tested using the EGFP-39TAG reporter and normalized relative to the WT-tRNA^{Pyl}.

(VADER) in mammalian cells (Figure 1a), which employs a double-sieve selection scheme to select for orthogonal tRNA mutants with higher activity from synthetic libraries.[2a] The VADER protocol subjects adeno-associated virus (AAV) particles encoding a suppressor tRNA library to selective replication in mammalian cells. The gene encoding the AAV capsid proteins (Cap) is mutated to introduce a TAG nonsense codon at a surface-exposed permissive site and supplied *in trans*.[7] Virions encoding active tRNA mutants are able to produce full-length capsid protein through nonsense suppression and produce progeny virus, while those with inactive tRNAs fail to replicate and are depleted from the library. Synthetic tRNA libraries may contain mutants that cross-react with a host aaRS and survive the selective replication step. To remove these undesired mutants, VADER introduces an ncAA with an azide handle at a surface-exposed site of the AAV capsid and uses it to attach a biotin group through click chemistry, enabling selective isolation of such virions.[7b, 7c] AAV particles encoding orthogonal tRNA mutants that selectively incorporate the azide-ncAA survive this selection, while those encoding cross-reactive tRNA mutants that charge a canonical amino acid are not modified with a biotin and are depleted. Subjecting synthetic libraries of the *M. mazei* pyrrolysyl tRNA (tRNA^{Pyl}) to VADER, and analyzing the enrichment patterns of the mutants using next-generation sequencing (NGS), which provides deep sequence coverage of the library and quantifies the abundance of each mutant before and after the selection, allowed us to find mutants with improved activity in mammalian cells.

However, VADER was significantly limited by the small size of mutant libraries it can process,[2a] which restricted us to randomizing only short segments (three base pairs) of the tRNA stems at a time. The inability to simultaneously randomize larger segments precludes the identification of distant mutations that work synergistically to greatly improve function. Here we report VADER2.0, which overcomes this limitation using a two-pronged approach. First, we significantly expanded the processing capacity of VADER by systematically optimizing its workflow, allowing the use of libraries that are an order of magnitude larger. Additionally, we developed a novel library generation method (correlated randomization) in which distal residues that base-pair with each other are randomized together to retain base pairing, enabling us to cover a significantly larger set of functional mutants within a fixed sequence space. These advances enabled the simultaneous randomization of the entire acceptor stem of tRNA^{Pyl}, leading to the identification of a mutant that shows much

higher activity relative to its predecessors. This improved activity facilitated more efficient incorporation of a broad group of structurally diverse ncAAs in mammalian cells. It provided a particularly impressive advantage under conditions where there are practical limits on the number of tRNA genes that can be introduced per cell, such as the development of mammalian cells stably expressing the ncAA incorporation machinery. Given the widespread use of the pyrrolysyl platform for ncAA mutagenesis in mammalian cells,[1a, 1d, 8] the development of this enhanced tRNA^{Pyl} mutant is expected to be broadly beneficial. In addition, the enhancements to the VADER protocol described in this work, which allow for the selection of larger and smarter libraries, will significantly expand its capabilities for the directed evolution of tRNAs and other biomolecules in mammalian cells.

Results and Discussion

Previously, we focused our engineering efforts on the acceptor and the T-stem of tRNA^{Pyl}, as these directly interface with numerous components of the cellular translation machinery, and because similar efforts have been highly successful for tRNA engineering performed previously in *E. coli*.[5] Due to the limits on the library size that VADER could handle, each stem was randomized as two separate six-nucleotide segments, creating four sub-libraries of 4,096 mutants each (Figure 1b). Using VADER, we identified mutants from both acceptor-stem sub-libraries that show up to 1.5 or 3-fold improvement relative to wild-type, but the T-stem libraries failed to yield improved mutants.[2a] Although this was encouraging, we were interested in exploring if it was possible to further improve the activity of tRNA^{Pyl}. Since each sub-library of the acceptor stem separately yielded beneficial mutants, we wondered if these could be combined for even higher activity. We generated several such chimeric tRNA^{Pyl} variants by combining sequences from some of the most active mutants (3-fold and 1.5-fold relative to wild-type tRNA^{Pyl}) isolated from each sub library (Figure 1c). Unfortunately, none of the chimeric mutants showed further improvement, indicating that the beneficial mutations identified in the sub-libraries cannot typically be combined to produce a greater increase in activity (Figure 1d). If synergistic mutations indeed exist in the two halves of the acceptor stem, finding them would require simultaneous randomization of the entire stem. However, the established VADER workflow did not have the capacity to process such a large library.

We systematically optimized various aspects of the VADER protocol to expand the upper limit of the library size that it can process. Key changes in the workflow that contributed to this improvement include: 1) The use of a more processive DNA polymerase to amplify the library after selection. This addressed a critical issue that contributed to reduced efficiency, where two distal randomized regions of the tRNA were frequently scrambled within the amplified product. 2) The use of golden gate assembly to generate the tRNA library, which significantly increased its quality relative to traditional restriction cloning. 3) Careful optimization of the capsid biotinylation/capture/release protocol to maximize the recovery. Using the optimized VADER protocol, it was possible to enrich an AAV encoding an active tRNA from its mixture with another AAV, containing an inactive tRNA, in 100,000-fold excess (Figure S1). This mock selection experiment suggests that it should now be possible to process a library size of up to 10^5 using the optimized VADER workflow. This capacity is close to the theoretical upper limit, given that the selection was performed using approximately 3 x 10^7 mammalian cells, which must be sparsely infected (<10%) to prevent coinfection with multiple different mutants, and a large excess (5-10 fold) of individual infections are needed to cover the theoretical diversity of the library. If needed, a higher library capacity can be accessed by scaling up the number of cells used in the selection.

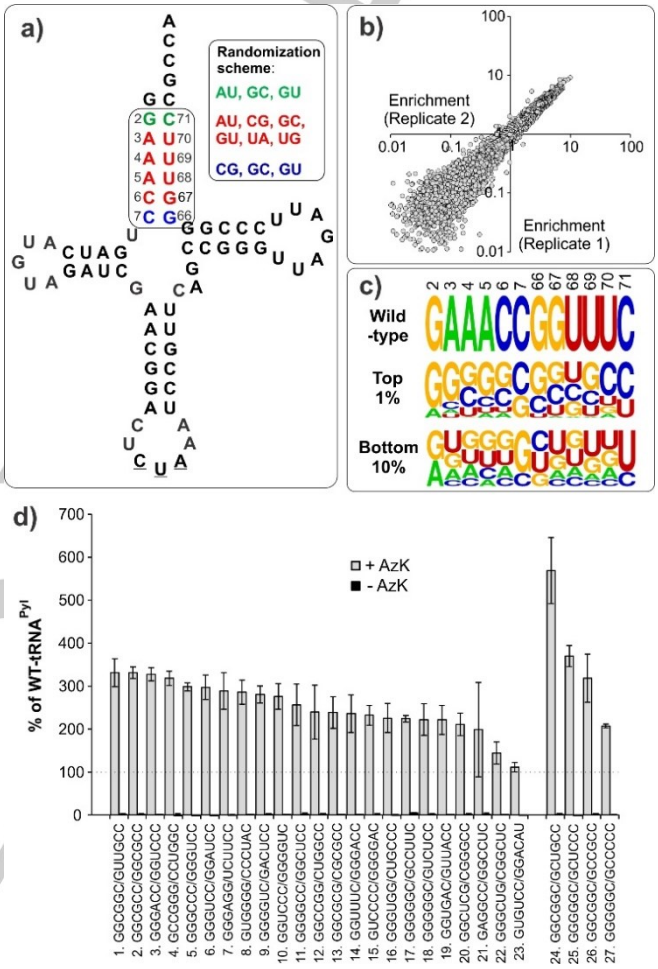


Figure 2. (a) The structure of tRNA^{Pyl} showing the strategy to generate a pairwise-randomized acceptor stem library. (b) Enrichment of each mutant in the library upon subjecting it to the VADER selection was measured in two replicates and plotted against each other. (c) Consensus sequence of top 1% (most enriched) sequences and bottom 10% (most depleted) sequences are shown. The sequence of the WT-tRNA^{Pyl} is provided as reference. (d) Activities of tRNA^{Pyl} mutants selected based on their strong enrichment profiles (1-23 are most enriched by average; 24-27 are highly enriched sequences that match the consensus sequence of the top 1% shown in panel c). The activity of each tRNA^{Pyl} variant was tested by co-transfecting these into HEK293T cells together with MbPylRS and EGFP-39TAG in the presence or absence of 1 mM AzK. Expression of EGFP-39TAG was measured in cell-free extract, normalized relative to wild-type mCherry expression (encoded in the tRNA-expressing plasmid; Figure S4), and normalized relative to the activity of wild-type tRNA^{Pyl}.

Taking advantage of the higher capacity, we constructed a 65,536-member tRNA^{Pyl} library by randomizing four contiguous base pairs in its acceptor stem, focusing around the previous A2 sub-library that yielded the most active variants (Figure S2a). This

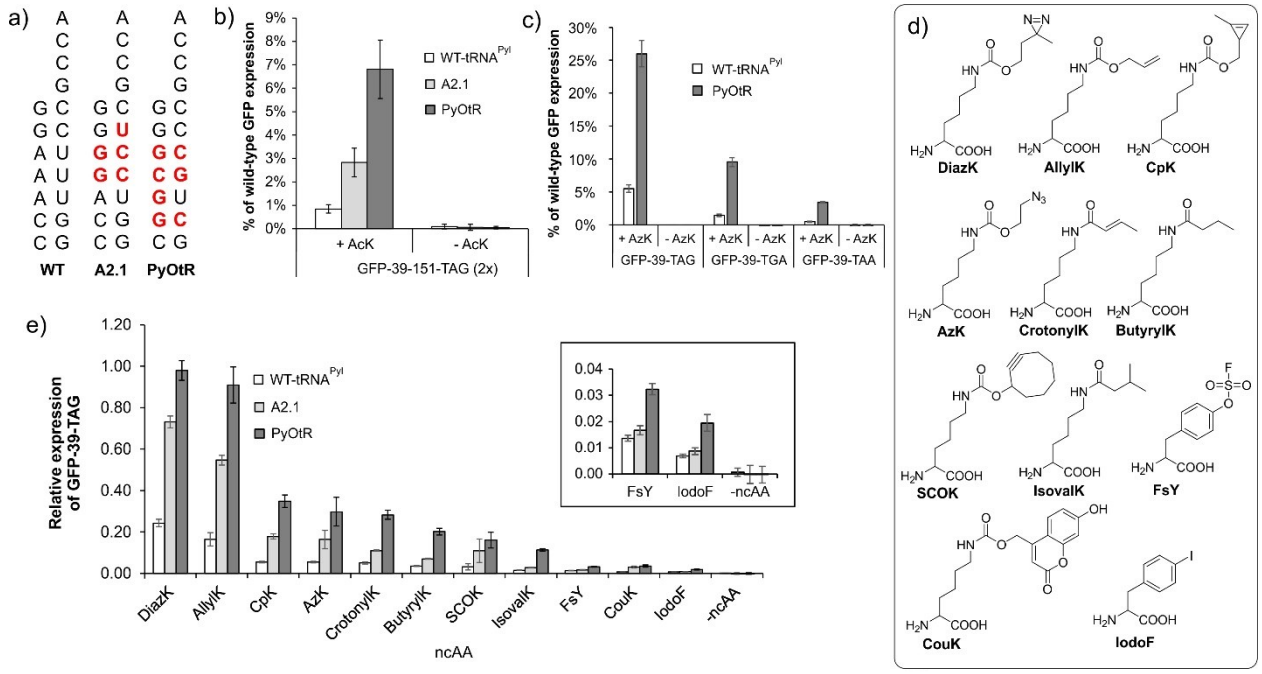


Figure 3. (a) Sequences of the acceptor stems of tRNA^{Pyl}-WT, A2.1, and PyOtR. (b) Expression of EGFP-39TAG-151TAG using the WT, A2.1, or PyOtR tRNA with AcK-selective MbPylRS, in the presence and absence of AcK. (c) Expression of EGFP-39TAG, EGFP-39TGA, and EGFP-39TAA reporters using tRNA$_{CUA}^{Pyl}$, tRNA$_{UCA}^{Pyl}$, and tRNA$_{UUA}^{Pyl}$ (WT and PyOtR), respectively, and MbPylRS in the presence and absence of AzK. (d) The structures of ncAAs used in the experiment in panel e. (e) Incorporation efficiency of various ncAAs using tRNA^{Pyl}-WT, A2.1, and PyOtR. In each case, the activity was measured by co-transfecting the tRNA-plasmid into HEK293T cells with plasmids encoding the appropriate MbPylRS and EGFP-39TAG in the presence or absence of 1 mM ncAA. Expression of EGFP-39TAG, normalized relative to mCherry expression (encoded in the tRNA-expressing plasmid; Figure S4), and normalized relative to the highest activity.

library was cloned into an AAV vector with 21-fold coverage, packaged into the AAV particles, and subjected to a single round of VADER in duplicate. The library was characterized by Illumina sequencing before and after the selection, which was used to quantify the degree of enrichment/depletion for each mutant upon selection (Figure S2b, S2c). All theoretical mutants were observed in the input library, and 99.99% of mutants (65,533/65,536) were observed in the selection output, attesting to the excellent library coverage during library construction and VADER workflow, respectively (Figure S2b, S2c). We have previously demonstrated that the activity of a tRNA mutant typically correlates with its degree of enrichment in VADER.[2a] We generated and characterized several variants, based on high enrichment as well as randomly picked survivors post-selection, and tested their relative activities using an EGFP-39TAG reporter (Figure S2d). Although many of these mutants showed improved activity relative to WT-tRNA^{Pyl}, with the most active mutant showing >3-fold improvement, none were significantly better than the most active mutant A2.1 that we previously identified from the three-base pair acceptor stem library.[2a]

Although the optimized VADER workflow significantly increased the library size that can be processed, it did not translate to being able to simultaneously randomize a significantly larger segment of the acceptor stem. This is because each additional base pair included for simultaneous randomization increases the library size by a factor of 16. The traditional strategy for generating a site-saturation library of a tRNA stem region involves independent randomization of each complementary strand using synthetic randomized primers. However, full randomization of two residues that form a base pair creates 16 possible mutants, of which only 6 retain base pairing (including the G:U wobble pair).[9a] Consequently, the large majority of such a library is composed of mutants that do not retain base pairing and are likely inactive. Creating a library through pairwise randomization, which is devoid of such unpaired mutations in the stem region, will drastically reduce the library size without compromising the functional diversity. We envisioned achieving this using the oligo-pool synthesis technology,[9b] which can generate a large number of distinct single-stranded DNA sequences long enough to cover the entire tRNA sequence. Using this approach, a new tRNA^{Pyl} library was designed in which the 3rd – 6th base pairs of the acceptor stem were randomized to all six base pairs (including the G:U wobble pair); the 2nd and the 7th base pair were partially randomized, based on trends observed from previous VADER selections of acceptor stem sub libraries; and the 1st base pair was kept unchanged, as it is an important identity element in PylRS:tRNA^{Pyl} interaction (Figure 2a).[10] Although we rationally restricted the library size (11,664 mutants) to minimize synthesis costs, complete correlated randomization of six base pairs would generate a library size (46,656 mutants) that can still be processed by our optimized VADER protocol. The traditional randomization approach for the same segment would create a drastically larger library ($1.6x10^7$), underscoring the advantage of our correlated randomization strategy.

This new library was packaged into an AAV2 vector and subjected to the VADER selection scheme in duplicate. The library underwent Illumina sequencing before and after selection, and the enrichment factors for each mutant from each replicate was plotted against each other (Figure 2b, S3b). Over 99.9% of the library population (11,658/11,664 mutants) were identified in our sequencing analysis, which further revealed reproducible enrichment of a fraction of mutants (Figure 2b, S3). Sequence analysis of the most enriched mutants revealed a significant departure from the wild-type sequence, with a strong preference

for G residues on the 5'-side of the acceptor stem (Figure 2c). Not surprisingly, the least enriched sequences had an abundance of G:U pairs, which are typically detrimental in tRNA stem regions.[5a, 11] A large number of mutants showed higher enrichment relative to tRNA^{Pyl} wild-type, highlighting the plasticity of the tRNA^{Pyl} acceptor stem (Figure S3b). Given the large number of potential hits that showed strong enrichment, we pursued two different strategies to select the ones to be individually characterized. In addition to testing all of the top 23 sequences showing the most enrichment, we also generated four mutants based on the consensus sequence of the top 1% most enriched hits. The activity of each mutant was characterized using an EGFP-39TAG reporter in HEK293T cells (Figure 2d, S4). Gratifyingly, all of the mutants tested showed activities higher than wild-type tRNA^{Pyl}, and the average improvement was substantially higher than what was previously observed with mutants from the two acceptor stem sub-libraries of tRNA^{Pyl}; only 7 out of the 43 sequences tested from A1 and A2 sub libraries (Figure 1b) showed >2-fold increased activity relative to WT-tRNA^{Pyl},[2a] whereas 24 out of the 27 mutants tested from the correlated randomization library (Figure 2a) were at least 2-fold more active. This underscores the importance of simultaneously engineering the entire stem, instead of doing it in short segments. The most active tRNA^{Pyl} mutant (the Pyrrolysyl Optimized tRNA; PyOtR) shows approximately 5.5-fold improvement relative to the wild-type, which is significantly higher than any other mutant we have identified thus far. PyOtR has a G:C-rich sequence, with a higher abundance of G on the 5' side, consistent with the consensus sequence of high-performing mutants (Figure 2c, 3a). It also has a single G:U base pair. Although the presence of this wobble pair in the tRNA stem regions is typically detrimental for tRNA activity,[5a, 11] in this particular case, the G:U pair is necessary for enhanced activity; another otherwise identical tRNA^{Pyl} mutant, in which this G:U pair is replaced with G:C (#26; Figure 2d) shows significantly attenuated activity.

We benchmarked the activity of PyOtR in the context of several different ncAA mutagenesis applications that have been traditionally challenging. First, we used an EGFP reporter harboring two TAG codons, at sites 39 and 151, respectively. Incorporation of N-acetyllysine (AcK),[12] which is one of the weaker ncAA substrates available through the pyrrolysyl platform, into this reporter was evaluated using different tRNA^{Pyl} variants. PyOtR facilitated the highest level of reporter expression, showing approximately 7-fold or 2.5-fold improvements relative to the wild-type tRNA^{Pyl} and A2.1, respectively, suggesting that it can significantly improve the scope of multisite ncAA incorporation applications (Figure 3b). It is also known that tRNA^{Pyl} shows significantly lower suppression efficiency of the TAA and TGA nonsense codons relative to TAG.[2a, 13] Using GFP reporters encoding the appropriate nonsense codon at the 39 position, we found that PyOtR facilitates >6-fold higher efficiency of TGA and TAA suppression relative to wild-type tRNA^{Pyl} (Figure 3c). This would be particularly beneficial for site-specific incorporation of two distinct ncAAs, each of which must be assigned to a different nonsense codon.[13-14] The pyrrolysyl platform has been engineered to genetically encode numerous ncAAs with different structures.[1a, 1d, 8] We found that PyOtR facilitates improved incorporation of a diverse group of ncAAs relative to WT-tRNA^{Pyl} or A2.1 (Figure 3d, 3e). Improved efficiency of PyOtR was also observed in several other mammalian cell lines (Figure S5). We have previously used the pyrrolysyl system to incorporate ncAAs into the AAV capsid,[7] which provides an enabling approach to probe and engineer its properties. The ability to do so is important given the tremendous potential of AAV for human gene therapy. However, using this platform, we got significantly lower ncAA-AAV titers, relative to wild-type AAV.[7] When we replaced the WT-tRNA^{Pyl} in our existing ncAA-AAV packaging system with PyOtR, the packaging efficiency was improved by nearly 3-fold, further highlighting its advantage (Figure S6).

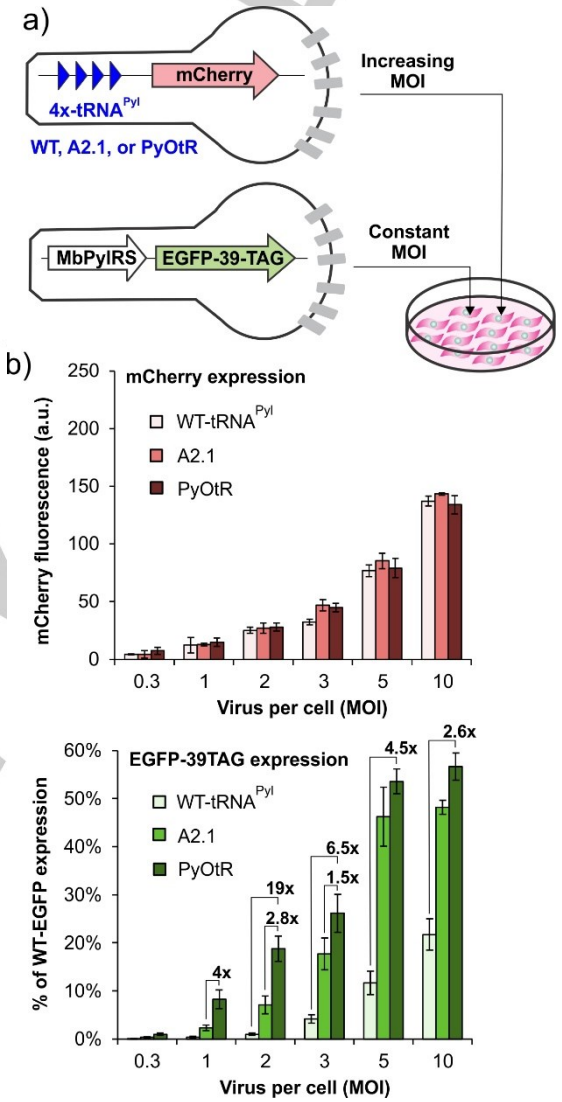


Figure 4. (a) Mammalian cell-optimized baculovirus vectors and the scheme of the experiment for panel b. (b) Expression of EGFP-39TAG as increasing MOI of tRNA-BacMam vector is used (bottom). Expression of mCherry (top) in the same experiments shows equivalent delivery of both tRNAs, which increases proportionately with higher MOI. Expression of the fluorescent reporters is measured in HEK293T cell-free extract. These experiments were performed using MbPylRS (Y349F) and AzK. Figure S8 shows representative images.

Northern blot analysis of WT-tRNA^{Pyl} and PyOtR revealed similar expression levels when identical plasmids encoding these tRNA variants were transfected into mammalian cells (Figure S7), suggesting that the enhanced activity of PyOtR is due to its superior intrinsic activity, and not elevated expression levels. We have previously shown that transient transfection leads to significant overexpression of the tRNA, which obscures the advantage of a highly efficient variant relative to a moderately

active one: The activity of the former saturates at higher expression level, allowing the latter to close the efficiency gap.[2a]

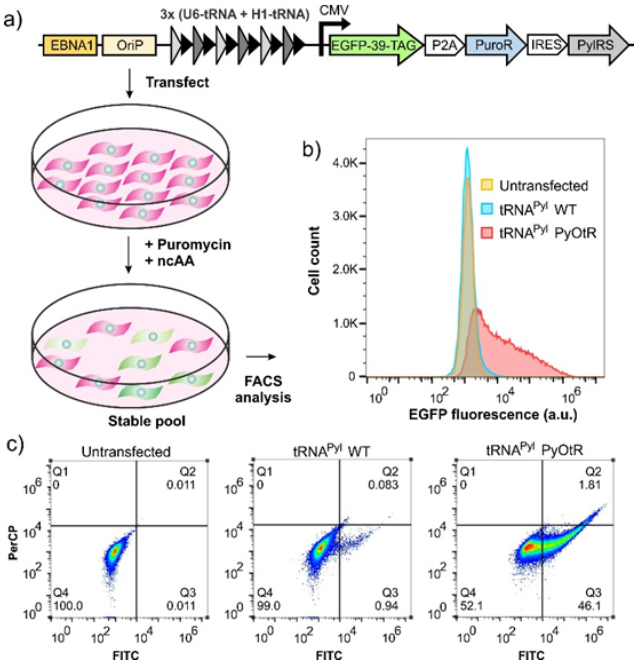


Figure 5. (a) Map of the EBV construct, as well as the experimental scheme, used to generate stable mammalian cell lines expressing MbPylRS, tRNA^{Pyl} (WT or PyOtR), and an EGFP-39TAG reporter. FACS histogram (b) and dot-plot (c) analysis (EGFP fluorescence) of the stable pool generated upon selection for both WT-tRNA^{Pyl} and PyOtR, as well as an untransfected control.

We have established an assay to better compare the activities of tRNA variants across a range of expression levels, which is achieved by controlled delivery using a mammalian-optimized baculovirus (BacMam) vector.[2a, 2c, 15] This assay uses two different BacMam vectors encoding: 1) four copies of the tRNA variant being tested and a wild-type mCherry reporter, which is used in increasing MOI (multiplicity of infection; virus:cell ratio), and 2) a copy of MbPylRS and an EGFP-39TAG reporter, which is used at a constant MOI (Figure 4a). The mCherry expression levels correlate with the number of tRNA-encoding virus delivered, while the EGFP-39TAG expression levels report the efficiency of TAG suppression. A virus encoding a wild-type EGFP was also separately tested at the same MOI as the 2nd virus, to estimate the TAG suppression efficiency relative to a sense codon. Using this assay, we compared the activity of WT-tRNA^{Pyl}, A2.1, and PyOtR across various expression levels. At all MOIs tested, the three different tRNA-viruses showed similar levels of mCherry expression, confirming comparable delivery (Figure 4b, S8). PyOtR facilitated the highest levels of EGFP-39TAG expression at the lowest MOI (Figure 4b, S8); even at MOI of 1, it showed 8% TAG suppression, while A2.1 showed only 2% TAG suppression, and the suppression efficiency for WT-tRNA^{Pyl} was below the detection level. At MOI 2, PyOtR showed 19-fold and 2.8-fold higher activity relative to WT, and A2.1, respectively. The TAG suppression efficiency for PyOtR got saturated at ~50% between MOIs 5 and 10, indicating that the tRNA stops being the limiting factor at this range. In contrast, the activity of the WT-tRNA^{Pyl} continued to rise even at the highest MOI tested that is tolerated by the cells. These experiments show that PyOtR is particularly effective when there are limits on the number of tRNA genes that can be introduced in a cell. Additionally, we have previously shown that this engineered baculovirus enables efficient gene delivery into a broad range of mammalian cells, including primary cells, stem cells, neurons, as well as in animal tissues.[2c, 15] Consequently, the PyOtR-containing baculoviral vector reported here, which enables highly efficient ncAA incorporation, would be beneficial for improving the scope of this technology to mammalian cells and tissues that cannot be readily transfected.

Development of mammalian cell lines stably expressing the ncAA incorporation machinery is important for both biotechnology, such as for large-scale commercial production of therapeutic proteins with ncAA handles, and for advanced basic science applications, such as those in animal models.[3-4, 16] Such stable cell lines are typically generated by incorporating the aaRS/tRNA genes into the host genome, either through random integration, or via transposon-mediated integration.[3-4, 16b] An Epstein-Barr virus (EBV) based vector, which is episomally maintained in mammalian cells in a stable manner, has also been used for this purpose.[16a] However, careful analyses of stable cell lines with high ncAA incorporation efficiency have shown the need for the presence of hundreds of tRNA genes per cell.[3] Incorporation of such a large number tRNA genes is intrinsically challenging, and the resulting cell lines often show poor stability, presumably through recombination-mediated loss of tRNA copies. We hypothesized that PyOtR should enable facile development of stable ncAA-incorporating mammalian cell lines, as our results suggest that it provides robust activity from fewer genes. To test this notion, we sought to compare the efficiency of generating stable suppressor cell lines using either PyOtR or WT-tRNA^{Pyl} via the Epstein-Barr virus (EBV) based stable vector.[16a] Unlike other methods that rely on genomic integration at random sites, which may give rise to context effects that cannot be exactly replicated across experiments, the EBV vector is maintained as a stable episome and should be free from such artifacts. We generated two EBV vectors, each encoding six copies of either WT-tRNA^{Pyl} or PyOtR, as well as a CMV-driven polycistronic gene cassette expressing the EGFP-39TAG reporter, the puromycin-N-acetyltransferase (Pac), and PylRS (Figure 5a). EGFP-39TAG and Pac were translationally coupled through a P2A sequence, requiring TAG suppression to express both genes, while PylRS was independently expressed from the same transcript using an IRES. HEK293T cells were transfected with both plasmids and the resulting cells were subjected to puromycin selection. The surviving colonies were expanded under continued puromycin selection to establish polyclonal stable pools for both tRNAs (Figure S9a). FACS analysis of these stable pool revealed that only about 1% of the cells derived from WT-tRNA^{Pyl} show significantly higher fluorescence relative to untransfected control cells, whereas nearly half the polyclonal cell population derived from PyOtR shows significantly higher EGFP expression (Figure 5b, 5c). Many more cells from the PyOtR pool also showed very high reporter expression, which would aid the development of highly efficient monoclonal suppressor lines (Figure 5c). As expected, reporter expression was rapidly lost when cultured in the absence of a ncAA (Figure S9b). These observations corroborate the significant advantage of using PyOtR for developing stable ncAA-incorporating mammalian cell lines.

Conclusion

The suboptimal performance of the nonsense suppressor tRNAs in mammalian cells limits the scope of the ncAA mutagenesis technology. Here we present VADER2.0 that effectively addresses this issue by enabling the selection of enhanced tRNA mutants from large synthetic libraries in mammalian cells. Previously, engineering short segments of $tRNA^{Pyl}$ acceptor stem yielded only moderately beneficial mutations that could not be combined for additional improvement. VADER2.0 overcomes this challenge with drastically improved capacity and a more effective library design, making it possible to simultaneously engineer the entire $tRNA^{Pyl}$ acceptor stem. This enabled the identification of PyOtR, a mutant more efficient than its predecessors, despite expressing at a lower level. The advantage of PyOtR was highlighted by improved incorporation of a diverse group of ncAAs, construction of a viral vector that enables efficient ncAA mutagenesis at low MOI, and facile generation of stable mammalian cell lines with high ncAA incorporation efficiency. We have previously shown that tRNA directed evolution in a particular host cell often leads to host-specific improvements.[2a] Consequently, a mammalian cell based directed evolution platform is essential to develop engineered tRNAs that show enhanced activity in these cells. VADER2.0 offers a general platform to this end, and would enable facile directed evolution of additional suppressor tRNAs used for ncAA mutagenesis in mammalian cells, as well as of endogenous suppressor tRNAs that hold much potential for developing gene therapies targeting in-frame nonsense codons that lead to inherited diseases. Moreover, this platform can be further adapted to enable directed evolution of additional genetic parts in mammalian cells.[17]

Acknowledgements

We thank Prof. P.G. Schultz for providing the pEBV plasmids. This work was supported by National Science Foundation (MCB-1817893), and National Institutes of Health (R35GM136437 to A.C. and U01 AI124302 to T.v.O.).

Keywords: Directed evolution • genetic code expansion • nonsense suppression • tRNA engineering •

[1] a) J. W. Chin, *Nature* 2017, *550*, 53-60; b) J. S. Italia, Y. Zheng, R. E. Kelemen, S. B. Erickson, P. S. Addy, A. Chatterjee, *Biochemical Society transactions* 2017, *45*, 555-562; c) M. Manandhar, E. Chun, F. E. Romesberg, *J Am Chem Soc* 2021, *143*, 4859-4878; d) D. D. Young, P. G. Schultz, *ACS Chem Biol* 2018, *13*, 854-870.

[2] a) D. Jewel, R. E. Kelemen, R. L. Huang, Z. Zhu, B. Sundaresh, X. Cao, K. Malley, Z. Huang, M. Pasha, J. Anthony, T. van Opijnen, A. Chatterjee, *Nature Methods* **2023**, *20*, 95-103; b) W. H. Schmied, S. J. Elsässer, C. Uttamapinant, J. W. Chin, *Journal of the American Chemical Society* **2014**, *136*, 15577-15583; c) Y. Zheng, T. L. Lewis, Jr., P. Igo, F. Polleux, A. Chatterjee, *ACS synthetic biology* **2016**, *6*, 13-18.

[3] G. Roy, J. Reier, A. Garcia, T. Martin, M. Rice, J. Wang, M. Prophet, R. Christie, W. Dall'Acqua, S. Ahuja, M. A. Bowen, M. Marelli, *mAbs* **2020**, *12*, 1684749.

[4] F. Tian, Y. Lu, A. Manibusan, A. Sellers, H. Tran, Y. Sun, T. Phuong, R. Barnett, B. Hehli, F. Song, M. J. DeGuzman, S. Ensari, J. K. Pinkstaff, L. M. Sullivan, S. L. Biroc, H. Cho, P. G. Schultz, J. DiJoseph, M. Dougher, D. Ma, R. Dushin, M. Leal, L. Tchistiakova, E. Feyfant, H. P. Gerber, P. Sapra, *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 1766-1771.

[5] a) J. C. Anderson, P. G. Schultz, *Biochemistry* **2003**, *42*, 9598-9608; b) J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, D. S. King, P. G. Schultz, *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 7566-7571; c) A. Chatterjee, H. Xiao, P. G. Schultz, *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 14841-14846; d) A. Chatterjee, H. Xiao, P. Y. Yang, G. Soundararajan, P. G. Schultz, *Angewandte Chemie (International ed. in English)* **2013**, *52*, 5106-5109; e) J. W. Ellefson, A. J. Meyer, R. A. Hughes, J. R. Cannon, J. S. Brodbelt, A. D. Ellington, *Nature Biotechnology* **2014**, *32*, 97; f) J. Guo, C. E. Melancon, 3rd, H. S. Lee, D. Groff, P. G. Schultz, *Angewandte Chemie (International ed. in English)* **2009**, *48*, 9148-9151; g) A. C. Maranhao, A. D. Ellington, *ACS synthetic biology* **2017**, *6*, 108-119; h) D. T. Rogerson, A. Sachdeva, K. Wang, T. Haq, A. Kazlauskaite, S. M. Hancock, N. Huguenin-Dezot, M. M. Muqit, A. M. Fry, R. Bayliss, J. W. Chin, *Nature chemical biology* **2015**, *11*, 496-503; i) K. Wang, A. Sachdeva, D. J. Cox, N. M. Wilf, K. Lang, S. Wallace, R. A. Mehl, J. W. Chin, *Nature chemistry* **2014**, *6*, 393-403; j) N. Wang, X. Shang, R. Cerny, W. Niu, J. Guo, *Scientific reports* **2016**, *6*, 21898; k) E. A. DeBenedictis, D. Söll, K. M. Esvelt, *eLife* **2022**, *11*, e76941.

[6] R. Serfling, C. Lorenz, M. Etzel, G. Schicht, T. Böttke, M. Mörl, I. Coin, *Nucleic acids research* **2018**, *46*, 1-10.

[7] a) S. B. Erickson, R. Mukherjee, R. E. Kelemen, C. J. Wrobel, X. Cao, A. Chatterjee, *Angew Chem Int Ed Engl* **2017**, *56*, 4234-4237; b) R. E. Kelemen, S. B. Erickson, A. Chatterjee, *Methods in molecular biology (Clifton, N.J.)* **2018**, *1728*, 313-326; c) R. E. Kelemen, R. Mukherjee, X. Cao, S. B. Erickson, Y. Zheng, A. Chatterjee, *Angewandte Chemie (International ed. in English)* **2016**, *55*, 10645-10649.

[8] W. Wan, J. M. Tharp, W. R. Liu, *Biochimica et biophysica acta* **2014**, *1844*, 1059-1070.

[9] a) B. Masquida, E. Westhof, *RNA* **2000**, 6, 9-15; b) S. Kosuri, G. M. Church, *Nature Methods* **2014**, *11*, 499-507.

[10] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Soll, O. Nureki, *Nature* **2009**, *457*, 1163-1167.

[11] S. W. Santoro, J. C. Anderson, V. Lakshman, P. G. Schultz, *Nucleic acids research* **2003**, *31*, 6700-6709.

[12] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nature chemical biology* **2008**, *4*, 232-234.

[13] Y. Zheng, P. S. Addy, R. Mukherjee, A. Chatterjee, *Chem Sci* **2017**, *8*, 7211-7217.

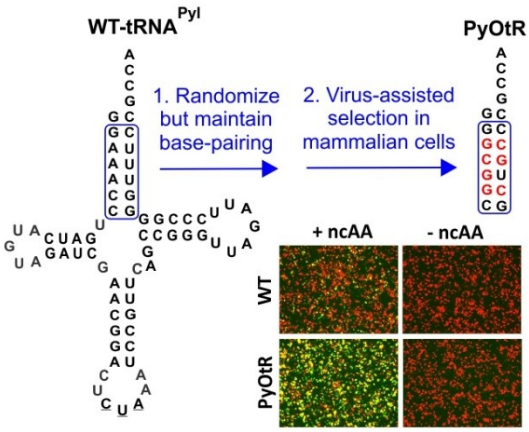
[14] a) Y. Zheng, M. J. Gilgenast, S. Hauc, A. Chatterjee, *ACS Chem Biol* **2018**, *13*, 1137-1141; b) Y. Zheng, R. Mukherjee, M. A. Chin, P. Igo, M. J. Gilgenast, A. Chatterjee, *Biochemistry* **2018**, *57*, 441-445; c) R. M. Bednar, P. A. Karplus, R. A. Mehl, *Cell chemical biology* **2023**, 30, 343; d) A. O. Osgood, Y. Zheng, S. J. S. Roy, N. Biris, M. Hussain, C. Loynd, D. Jewel, J. S. Italia, A. Chatterjee, *Angewandte Chemie International Edition* **2023**, *n/a*, e202219269.

[15] A. Chatterjee, H. Xiao, M. Bollong, H. W. Ai, P. G. Schultz, *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 11803-11808.

[16] a) S. Shao, M. Koh, P. G. Schultz, *Proceedings of the National Academy of Sciences* **2020**, *117*, 8845-8849; b) S. J. Elsässer, R. J. Ernst, O. S. Walker, J. W. Chin, *Nat Methods* **2016**, *13*, 158-164.

[17] a) C. M. Berman, L. J. Papa, 3rd, S. J. Hendel, C. L. Moore, P. H. Suen, A. F. Weickhardt, N. D. Doan, C. M. Kumar, T. G. Uil, V. L. Butty, R. C. Hoeben, M. D. Shoulders, *J Am Chem Soc* **2018**, *140*, 18093-18103; b) S. J. Hendel, M. D. Shoulders, *Nat Methods* **2021**, *18*, 346-357; c) J. G. English, R. H. J. Olsen, K. Lansu, M. Patel, K. White, A. S. Cockrell, D. Singh, R. T. Strachan, D. Wacker, B. L. Roth, *Cell* **2019**, *178*, 748-761.e717.

Entry for the Table of Contents



Insert text for Table of Contents here. Using an optimized virus-assisted directed evolution scheme in mammalian cells, an enhanced variant of the *M. mazei* pyrrolysyl tRNA, PyOtR, was developed. PyOtR enabled improved incorporation of many noncanonical amino acids (ncAAs) into proteins expressed in mammalian cells, and enabled facile generation of cell lines stably expressing the ncAA incorporation machinery.

Institute and/or researcher Twitter usernames: @AchemSynBio @ChemistryBC