**Supporting Information**

# Site-Specific Incorporation of p-propargyloxyphenylalanine in a Cell-free Environment for Direct Protein-Protein Click Conjugation

Bradley C. Bundy* and James R. Swartz*[†‡]

*Department of Chemical Engineering, Stanford University, Stanford, California 94305-5025
[†]Department of Bioengineering, Stanford University, Stanford, California, 94305-5444

CONTENTS

**Sequence 1. pY71 vector sequence.** Unessential sequences were removed from the pK7 vector *(1)* to generate the pY71 vector as shown by the sequence below that is bolded, italicized, underlined, and highlighted yellow. The italicized and highlighted green CTGCAG in the first line was added to pY71. The T7 promoter and terminator sequences are underlined and highlighted blue. The kanR sequence is underlined, italicized, and highlighted light blue. The bolded sequence is the sequence of *E. coli* DHFR containing the amber stop codon substitution at V10 and streptag encoding sequence at the 3' end. The NdeI and SalI restriction sites used for cloning into pY71 directly flank the DHFR gene sequence and are highlighted red.

GGATC*CTGCAG*TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCAC
CGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGG
CTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTC
AAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCG
GTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTG
AGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGT
ATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCG
TCAGGGGGGCGGAGCCTATGGAAA***AACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT***
***GCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTAC***
***CGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGC***
***GAGGAAGCGGAAGAAGCTCGCACGCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTC***
***ATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAA***
***TGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTG***
***TGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTA***CGAATTCA
GATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAA
ATAATTTTGTTTAACTTTAAGAAGGAGATATACAT**ATG**ATCAGTCTGATTGCGGCGTTAGCGTA
**GGATCGCGTTATCGGCATGGAAAACGCCATGCCGTGGAACCTGCCTGCCGATCTCGCCTGGTTT**
**AAACGCAACACCTTAAATAAACCCGTGATTATGGGCCGCCATACCTGGGAATCAATCGGTCGTC**
**CGTTGCCAGGACGCAAAAATATTATCCTCAGCAGTCAACCGGGTACGGACGATCGCGTAACGTG**
**GGTGAAGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGACGTACCAGAAATCATGGTGATTGGC**
**GGCGGTCGCGTTTATGAACAGTTCTTGCCAAAAGCGCAAAAACTGTATCTGACGCATATCGACG**
**CAGAAGTGGAAGGCGACACCCATTTCCCGGATTACGAGCCGGATGACTGGGAATCGGTATTCAG**
**CGAATTCCACGATGCTGATGCGCAGAACTCTCACAGCTATTGCTTTGAGATTCTGGAGCGGCGG**
**TGGAGCCATCCGCAGTTCGAAAAATAA**GTCGAC**CGGCTGCTAACAAAGCCCGAAAGGAAGCTGA**
GTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTG
AGGGGTTTTTTGCTGAAAG***GAGGAACTATATCCGGATAACCTCGAGCTGCAGGGCATGCAAGCT***
***TGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCG***
***CCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCT***
***TCCCAACAGTTGCGCAGCCTGAATGGCGAATGCGATTTATTCAACAAAGCCGCCGTCCCGTCAA***
***GTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAA***CCAATTCTGATTA*GAAAAACTCATCGA*
*GCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCG*
*TTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGG*
*TCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGT*
*TATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCAT*
*TTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACC*
*AAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGAC*

*AATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTC*
*ACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGT*
*AACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCA*
*GCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAG*
*AAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACA*
*TTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCTTCG*
*AGCAAGACGTTTCCCGTTGAATATGGCTCAT*AACACCCCTTGTATTACTGTTTATGTAAGCAGA
CAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGAC
ACAACGT***GGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCATCTTC***
***CCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAATCACCAACTGGCCCACCTACAA***
***CAAAGCTCTCATCAACCGTGGCTCCCTCACTTTCTGGCTGGATGATGGGGCGATTCAGGCCTGG***
***TATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCTCGACGGATCGTTCCACTGAGCGTCAG***
***ACCCCGTAGAAAAGATCAAAGGATCTTC***

**Sequence 2. Superfolder GFP protein sequence used in this study.** Mutations made to the superfolder GFP sequence used in this study compared to the GFP sequence (Protein Accession Number AAG29466) *(2)* are highlighted below. Green highlighting indicates "superfolder" or "enhanced" GFP mutations *(3)*. Red highlighting indicates mutations made to help GFP fold properly when fragmented (specifically reported on "Opt 1-10" and "11-M3" sequences) *(4)*. The blue and grey highlighted residues are insertions which extend the loop before the final beta sheet. The highlighted grey threonine 216 was substituted for *p*-azido-L-phenylalanine (pAz) or *p*-propargyloxyphenylalanine (pPa) to generate pAzGFP and pPaGFP. In addition, the final 8 unessential residues of the GFP were removed and the c-terminal streptag was added (yellow highlighting).

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGT
DFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSTQTVLSKDPNEKGTRDHMVLHEYVNAAGITWSHPQFEK

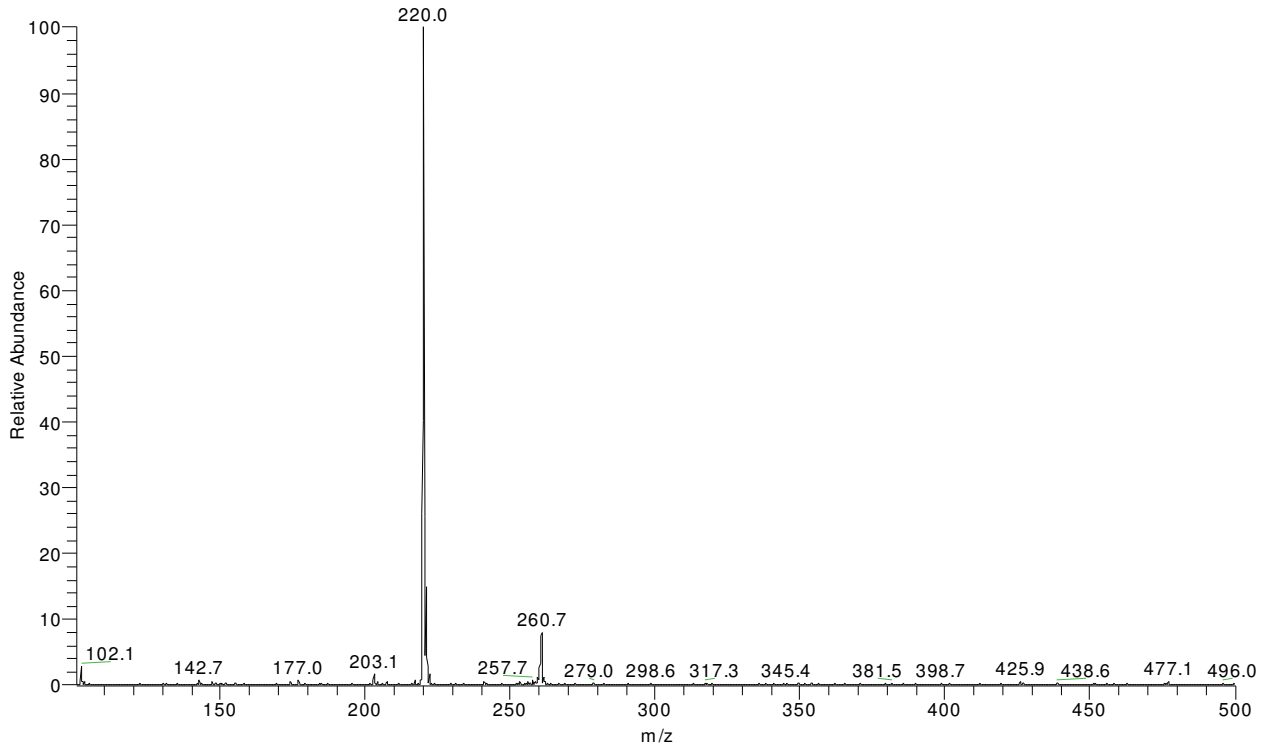


**Figure S1. Characterization of *p*-propragyloxyphenylalanine.** Electrospray mass spectra (M+1) of *p*-propargyloxyphenylalanine synthesized as explained previously *(5)*. The major peak is at m/z 220.0 (expected $C_{12}H_{13}NO_3$ (M+1) m/z of 220.09). The sample was prepared in 50:50 water:acetonitrile before analysis with a Thermo Fisher Scientific Surveyor HPLC and LCQ ion trap mass spectrometer.

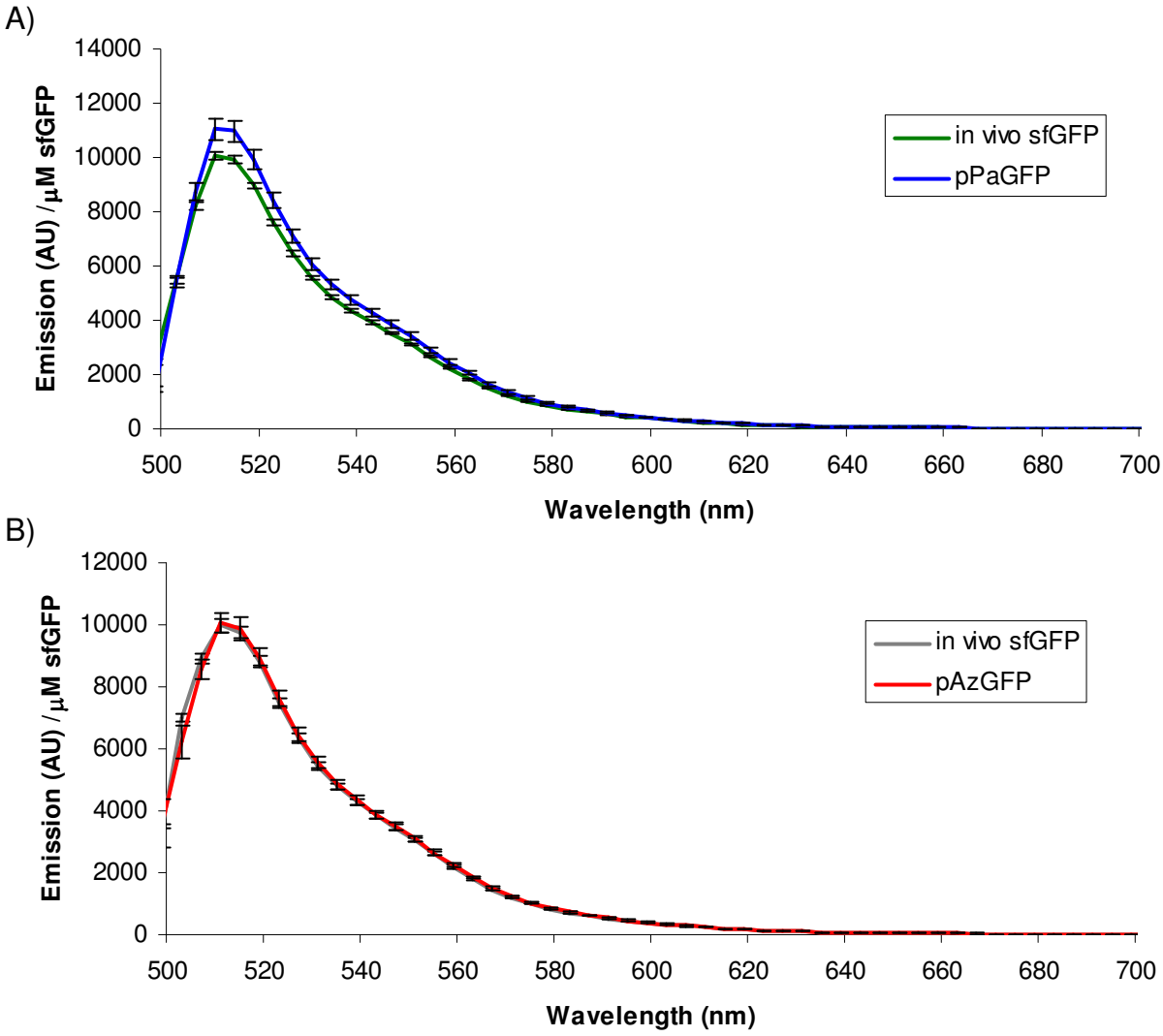


**Figure S2. Comparison of the emission profiles of pPaGFP (A) and pAzGFP (B) relative to *in vivo*-produced sfGFP.** All samples were affinity purified prior to analysis. The spectra was obtained using a NanoDrop 3300 Fluorimeter (Thermo Scientific, Waltham, MA) with the blue light source excitation setting and 2 μL of GFP sample at concentrations ranging from 0.5 to 7.9 μM. As expected for sfGFP the emission peaks at 511nm. Also, the incorporation of pPa or pAz does not appear to negatively affect sfGFP activity as the emission profile of pPaGFP and pAzGFP align well with sfGFP. The purified pPaGFP sample appears to have a slightly higher activity than sfGFP (9% increase), which is likely due to variation in sample preparation and concentration measurements rather than the incorporation of pPa. Error bars represent standard deviation from multiple measurements ($n \geq 6$).

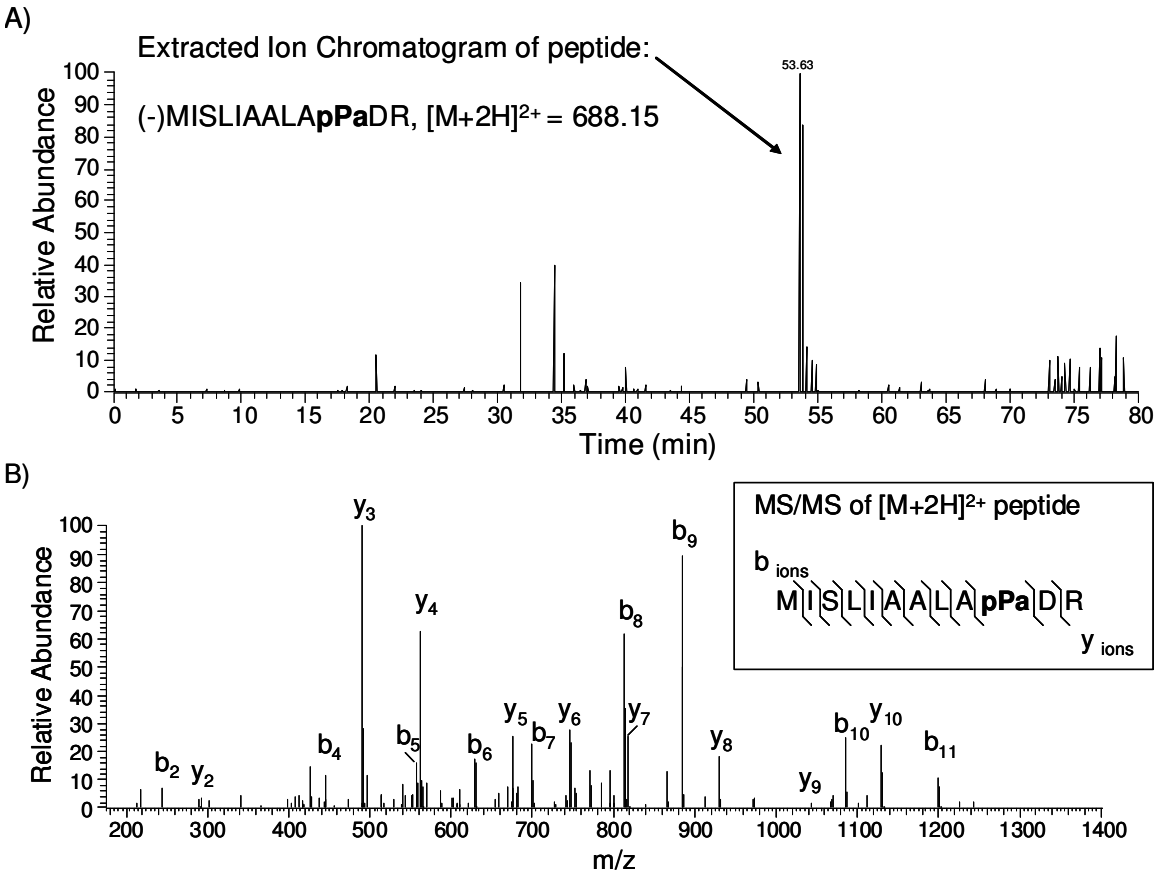


**Figure S3**. **Tryptic digestion of pPaDHFR followed by LC-MS. A)** The elution time of the peptide containing pPa. **B)** The MS/MS of the peptide containing pPa verifying incorporation of pPa site-specifically at the specified position.

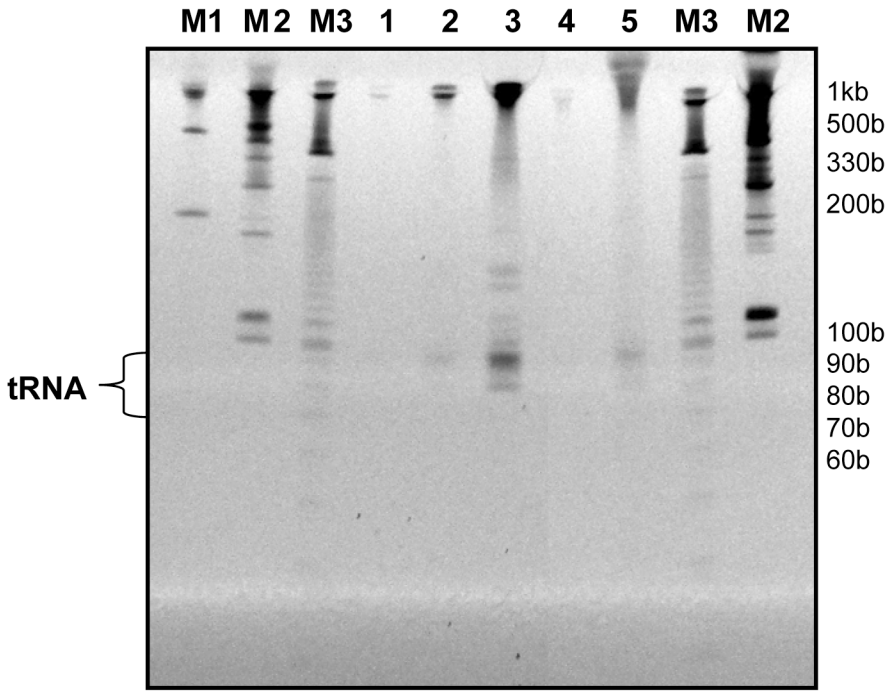


**Figure S4**. **Evaluation of the relative concentration of *Mj*tRNA$^{Tyr}_{CUA}$ in ethanol extracted nucleic acid.** UV-illuminated image of 15% TBE-Urea SDS-PAGE gel (Invitrogen, Carlsbad, CA) stained with ethidium bromide. Electrophoresis and staining was performed per the manufacturer's procedure. Image J 1.36b software (National Institutes of Health, USA) *(6)* was used to perform densitometry to determine the relative concentration of tRNA in the ethanol-extracted tRNA compared to the original *E coli* lysate used for CFPS. Lanes were loaded with the following: M1) 0.2-10kb perfect RNA marker, M2) 100bp DNA ladder, M3) 10bp DNA ladder, 1) 300x diluted ethanol-extracted tRNA, 2) 100x diluted ethanol-extracted tRNA, 3) 10x diluted ethanol-extracted tRNA, 4) 100x diluted *E coli* lysate, 5) 10x diluted *E coli* lysate. The approximate location of *E coli* tRNA is shown (75 – 95 bases). The expected length in nucleotide bases is shown on the right as estimated from the commercial markers (M1, M2, and M3). The bands at 1kb and above in lanes 1 through 5 could be due to the presence of rRNA. Although, both RNA and DNA are stained by ethidium bromide, the ethanol-extracted tRNA solution (lanes 1-3) is predominantly RNA. The absorbance at 260 nm correlates to an RNA concentration of 30 μg/μl (Standard Devation = 0.7 μg/μl) and an RNA concentration of 30 μg/μl (SD = 0.2 μg/μl) was determined by the Qubit-iT 5-100ng RNA assay (Invitrogen). A significant amount of DNA would cause a mismatch between the concentration calculations as DNA absorbs at the wavelength 260 nm but was not detected with the Qubit-iT 5-100ng RNA assay.

References:


(1) Swartz, J. R., Jewett, M. C., and Woodrow, K. M. (2004) Cell-free protein synthesis with prokaryotic combined transcription-translation, in *Recombinant Gene Expression: Reviews and Protocols* (Balbas, P., and Lorence, A., Eds.) pp 169-182, Humana Press, Totowa, NJ.

(2) Miller, W. G., Leveau, J. H., and Lindow, S. E. (2000) Improved gfp and inaZ broad-host-range promoter-probe vectors. *Mol. Plant Microbe. In. 13*, 1243-1250.

(3) Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol. 24*, 79-88.

(4) Cabantous, S., Terwilliger, T. C., and Waldo, G. S. (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol. 23*, 102-107.

(5) Deiters, A., Cropp, T. A., Mukherji, M., Chin, J. W., Anderson, J. C., and Schultz, P. G. (2003) Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J. Am. Chem. Soc. 125*, 11782-11783.

(6) Abramoff, M. D., Magelhaes, P. J., and Ram, S. J. (2004) Image processing with ImageJ. *Biophotonics Intl. 11*, 36-42.