Parallel positive selections, deep sequencing and analysis for reprogramming aminoacyl-tRNA synthetase specificity

Jason Chin (chin@mrc-lmb.cam.ac.uk)
MRC Laboratory of Molecular Biology
Michael Shaofei Zhang
MRC Laboratory of Molecular Biology
Simon Brunner
MRC Laboratory of Molecular Biology
Nicolas Huguenin-Dezot
MRC Laboratory of Molecular Biology
Alexandria Deliz Liang
MRC Laboratory of Molecular Biology
Wolfgang Schmied
MRC Laboratory of Molecular Biology
Daniel Rogerson
MRC Laboratory of Molecular Biology

Method Article

Keywords: non-natural amino acids, genetic code expansion, parallel positive selection, deep sequencing, aminoacyl-tRNA synthetase

Posted Date: May 30th, 2017

DOI: https://doi.org/10.1038/protex.2017.046

Abstract

Site-specific incorporation of non-natural amino acids into proteins, via genetic code expansion, forms the basis of powerful approaches to get access to proteins with site-specific modifications both in vitro and in vivo. The current methods for discovering aminoacyl-tRNA synthetase/tRNA pairs depend on several rounds of positive and negative selections, which have certain limitations. Here we describe a protocol to address the problem by combining parallel positive selections with deep sequencing and analytical methods from differential expression analysis to create a scalable approach for synthetase discovery.

Introduction

Genetic code expansion, using orthogonal aminoacyl-tRNA synthetase/tRNA pairs, forms the basis of powerful approaches for synthesizing proteins bearing defined post-translational modifications and complements protein semi-synthesis. Serial rounds of positive and negative selection are currently used to discover synthetases with altered specificity. Despite the clear utility of current approaches, they have certain limitations: i) it is not possible to follow the enrichment of individual clones through the selection to identify when substantial enrichment has taken place or which clones are substantially enriched, ii) the activity and specificity of isolated synthetase clones is restricted by the dynamic range of the negative selection; indeed, it is possible that the most active clones are deleted from the gene pool by negative selection, even though the activity of the synthetase they encode in the presence of an added unnatural amino acid may be sufficient to outcompete endogenous amino acid incorporation; iii) it is not possible to directly identify variants of a synthetase that have specificity with respect to many other unnatural substrates, and iv) the individual clones isolated at the end of the selection are picked randomly without regard to their relative enrichment or sequence. Our ability to find hits is a function of the number of clones we examine and the frequency of the hit in the library following selection. This is in turn is a function of the library size, the power of the selection steps to enrich the desired hits, the unknown frequency of hits in the library, and stochastic events. To address the limitations of current selection approaches we proposed selecting synthetases using parallel positive selections in the presence and absence of an non-natural amino acid, coupled to deep sequencing and statistical analyses that compare sequencing read counts for each synthetase clone in the presence and absence of a non-natural amino acid. Here we describe a protocol, which details the process to selected aminoacyl synthetase/tRNACUA pairs that incorporate Nε-\(\(1-\(6-nitrobenzo\[d]\[1,3]dioxol-5-yl)ethoxy)carbonyl)-L- lysine \(photocaged lysine, PCK), an amino acid that has been extensively used to optically control diverse protein functions.

Reagents

• PrimeSTAR Max Mix \(2x) \(Clontech, R045A) • All DNA primers are ordered from Sigma Custom DNA Oligos • Gel Loading Dye, Purple \(6x) \(NEB, B7024S) • TAE Buffer \(50x) \(Thermo Fisher, B49) • SYBR Safe DNA Gel Stain \(Thermo Fisher, S33102) • QIAquick Gel Extraction Kit \(Qiagen, 28704) • Qiaprep Spin Miniprep Kit \(Qiagen, 27106) • QIAGEN Plasmid plus Midi kit \(Qiagen, 12945) • All restrictive

endonucleases are purchased from NEB • T4 DNA Ligase \(NEB, M0202S) • MegaX DH10B Electrocomp Cells \(Thermo Fisher, C640003) • KAPA HiFi HotStart ReadyMix \(Kapa Biosystems, KK2502) • E-Gel SizeSelect Agarose Gel 2% \(Thermo Fisher, G661002) • KAPA Library Quantification Kits \(Kapa Biosystems, KK4824) • PhiX Control Kit v3 \(illumina, FC-110-3001) • MiSeq Reagent Kit v2 \(300 cycle) \(illumina, MS-102-2002) • Ni-NTA Agarose \(Qiagen, 30210) • BugBuster Protein Extraction Reagent \(Millipore, 70854-1000ML) • All the other chemicals and reagents are purchased from Sigma-Aldrich • qPCR Dilution Buffer: 200 ml consist of: 2 ml Tris-Cl pH = 8.0, 100 µl Tween20, 200 µl 4M NaOH • His-tag Lysis Buffer: To 100 ml BugBuster protein extraction reagent Add 2 ml 1 M Tris-HCl pH = 8.0, 10 ml 5 M NaCl, 20 mM imidazole 1 mM β-mercaptoethanol • His-tag Washing Buffer: 50 mM Tris-HCl pH = 8.0 500 mM NaCl 20 mM imidazole 1 mM β-mercaptoethanol • His-tag Eluting Buffer: 50 mM Tris-HCl pH = 8.0 50 mM NaCl 300 mM imidazole 1 mM DTT

Equipment

• Tprofessional Thermo Cycler \(Biometra TRIO) • Mupid One Electrophoresis System \(Clontech, AD160) • Dark Reader Transilluminator \(Clarechemical) • Avanti J-26 XP Centrifuge \(Beckman Coulter) • Eppendorf Centrifuge 5424 \(Eppendorf) • Gene Pulser®/MicroPulser™ Electroporation Cuvettes, 0.2 cm gap \(Bio-rad, #1652086) • Eppendorf® Electroporator 2510 \(Sigma, Z618969) • Eppendorf BioPhotometer Plus \(Eppendorf) • E-Gel® iBase™ Power \(Thermo Fisher, G6400UK) • ViiA7 Real-Time PCR system \(Thermo Fisher) • MiSeq Sequencer \(illumina) • Phenomenex Jupiter C4 column \(150 × 2 mm, 5 µm) • Agilent 1200 LC-MS system coupled with a 6130 Quadrupole spectrometer

Procedure

Library construction 1. To randomize codons coding for M. barkeri PylS residues A267, Y271, and L274, order forward PCR primers from Sigma with degenerative nucleotides NNK for the codon 267, 271, and 274 2. Setup PCR reactions PrimeSTAR Max mix \(2x) 500 µl pKW-PylS \(50 ng/µl) 1 µl Forward primer \(50 µM) 10 µl Reverse primer \(50 µM) 10 µl H2O 479 µl Total 1000 µl 3. Distribute the PCR mixture into 20 x 50 µl reactions, and run PCR reactions \(25 cycles) 98 °C 30 sec 98 °C 10 sec 55 °C 15 sec 72 °C 72 °C 5 min 10 min 4 °C ∞ 4. Pool the PCR reactions, add 200 µl 6x DNA loading dye, and mix thoroughly. Prepare 1% agarose gel in 0.5% TAE buffer and add SYBR Safe DNA dye to a final concentration of 0.01%. Separate the PCR products by electrophoresis at 90V for 25 min 5. With the aid of blue light transilluminator, cut the gel containing the desired PCR products, and extract the DNA using QIAquick Gel Extraction Kit following the manufacturer's instructions. Elute the DNA with 10% EB buffer diluted in H2O 6. Setup restriction digestion reaction and incubate at 37 °C for 4 hrs NEB Cutsmart buffer \(10x) 100 µl DNA from Step 5 600 µl BsaI-HF 30 µl DpnI-HF 10 µl H2O 260 µl Total 1000 µl 7. Add 5 ml PB buffer into the restrictive digestion reaction, mix well, and purify using QIAprep Spin Miniprep Columns. Elute the DNA with a total of 120 µl 10% EB \(diluted in H2O) 8. Setup T4 ligation reaction and incubate at 16°C for 16 hrs T4 ligase buffer \(10x) 100 µl Column purified DNA from step 7 120 µl T4 DNA ligase 10 µl H2O 770 µl Total 1000 µl 9. Add 100 µl 3M sodium acetate \(pH = 5.2) into the DNA ligation reaction; add 3 ml

100% ethanol into the reaction, mix thoroughly 10. Incubate for 4 hrs at RT, and further incubate for 16 hrs at -20°C 11. Centrifuge at 20,000 g for 30 min, and carefully decant the supernatant; centrifuge briefly again and use pipette to remove the remaining liquid. Resuspend the DNA pellet in 10% EB \(diluted in H2O) 12. Snap freeze the resuspended DNA in liquid nitrogen and concentrate by lyophilization to 20 µl 13. Transform the concentrated DNA into MegaX DH10B electrocompetent cells by electroporation and recover in pre-warmed SOC medium for 70 min at 37°C 14. Transfer the recovered cells into fresh 250 ml LB medium containing 50 µg/ml kanamycin. Incubate the culture at 37°C for 6 hrs 15. Centrifuge at 7,000 g for 10 min to pellet the cells 16. Extract the plasmids by QIAGEN Plasmid Plus Midi Kit following the manufacturer's instructions 17. To randomize codons coding for M. barkeri PylS residue C313, order forward PCR primers from Sigma with degenerate nucleotides NNK for the codon 313. 18. Setup PCR reactions PrimeSTAR Max mix \(2x) 500 µl Plasmid from Step 16 \(50 ng/µl) 1 µl Forward primer \(50 µM) 10 µl Reverse primer \(50 µM) 10 µl H2O 479 µl Total 1000 µl 19. Repeat Step 3-16 for the randomization of codon 313 20. To randomize codons coding for M. barkeri PylS residue M241, order forward PCR primers from Sigma with degenerative nucleotides NNK for the codon 241 21. Setup PCR reactions PrimeSTAR Max mix \(2x) 500 µl Plasmid from Step 19 \(50 ng/µl) 1 µl Forward primer \(50 µM) 10 µl Reverse primer \(50 µM) 10 µl H2O 479 µl Total 1000 µl 22. Repeat Step 3-16 for the randomization of codon M241. The plasmids extracted at the end is the PylS library used in the following selection steps Parallel positive selection 23. To transform pREP_PylT_CAT\(112TAG) into DH10B cells, aliquot 50 µl ice-cold MegaX DH10B electrocompetent cells into 2 mm electroporation cuvette. Add 0.5 µl pREP_PylT_CAT\(112TAG) \(50 ng/µl) into the cells and electroporate with the setting of 2,500 V. Immediately add 1 ml SOC into the cuvette and incubate at 37°C for 60 min with constant shaking at 220 rpm 24. Plate 30 µl recovered cells onto LB agar plate containing 25 µg/ml tetracyclin. Air-dry the plate and incubate the plate at 37°C overnight 25. Pick a single colony from the plate and inoculate into 5 ml LB medium containing 25 µg/ml tetracyclin. Incubate at 37°C overnight with constant shaking at 220 rpm 26. Dilute the overnight culture into 1 L fresh LB medium with 25 µg/ml tetracyclin and incubate at 37°C with constant shaking at 220 rpm. Monitor the OD600 by spectrophotometer 27. When cell density reaches around OD600 = 0.5-0.6, transfer the flask into ice water and cool down for 20 min. Pre-cool the centrifuge to 4°C at the same time. Keep the cells below 4°C from now on 28. Centrifuge at 4,000 g for 12 min, and carefully decant the supernatant LB medium without disturbing the cell pellet. Resuspend the cells with 1 L ice-cold water 29. Centrifuge at 4,000 g for 12 min, carefully decant the supernatant, and resuspend the cells with 1 L ice-cold 10% glycerol 30. Centrifuge at 4,000 g for 12 min, carefully decant the supernatant, and resuspend the cells in 50 ml ice-cold 10% glycerol 31. Centrifuge at 4,000 g for 12 min, carefully decant the supernatant, and resuspend the cells in 1 ml ice-cold 10% glycerol 32. Aliquot 3 x 90 µl cells into 2 mm electroporation cuvette. Add 2 µl PylS library from Step 22 \(200 ng/µl) into each cuvette and electroporate with the setting of 2,500 V. Immediately add 1 ml SOC into each cuvette and incubate at 37°C for 60 min with constant shaking at 220 rpm. Snap freeze the rest of the competent cells in liquid nitrogen. These cells can maintain their competency for at least 6 months at -80°C 33. After incubation, transfer the recovered cells from each cuvette into three flasks of 250 ml fresh LB medium containing 50 µg/ml kanamycin and 25 µg/ml tetracyclin. incubate the three cultures at 37°C overnight with constant shaking at 220 rpm 34. From each of the three overnight cultures \(#1, #2, and #3), take 5

ml and pellet the cells by centrifuging at 4,000 g for 12 min. Extract the plasmids from each cell pellet using Qiagen Miniprep Kit following the manufacturer's instructions. Label them as Pre-1, Pre-2, and Pre-3. 35. Prepare six 200 ml fresh LB medium with 50 µg/ml kanamycin and 25 µg/ml tetracyclin. Add 1 mM photo-caged lysine \(PCK) into three of the cultures \(+1, +2, and +3). Label the rest three cultures as -1, -2, and -3. 36. Transfer 2 ml overnight culture #1 into either +1 culture and -1 culture. Do the same dilution also for culture #2 and #3. Incubate at 37°C with constant shaking at 220 rpm. Monitor the OD600 by spectrophotometer 37. Prepare six LB agar plate \(25 cm x 25 cm square plate) containing 25 µg/ml kanamycin, 12.5 µg/ml tetracyclin, and 100 µg/ml chloramphenicol. Add 1 mM PCK into three LB agar plates. Label six plates as -1, -2, -3, +1, +2, and +3 accordingly 38. When the cultures in Step 36 reach around OD600 = 0.6, pipette 2 ml culture and spread evenly onto the corresponding LB plate. Air-dry the plates and incubate at 37°C for 40 hrs 39. To collect cells from each plate, add 50 ml PBS to each plate, use L-shape plastic spreader to carefully dissociate the colonies. Transfer the dissociated cells into 50-ml tubes and pellet cells by centrifugation at 7,000 g for 10 min 40. Extract the plasmids from each cell pellet using QIAGEN Plasmid Plus Midi Kit following the manufacturer's instructions. Label the six plasmid extracts as Post -1, Post -2, Post -3, Post +1, Post +2, and Post +3, respectively Sample preparation for deep sequencing 41. Order DNA primers to amplify the target region on the PylS plasmid for deep sequencing. From 5' to 3' both the forward and the reverse primer consisted of the following: 1) Illumina adapter \(forward 'P5' primer sequence: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT, reverse 'P7' primer sequence: CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT), 2) four degenerate nucleotides \(NNNN), 3) hexanucleotide barcode sequence to allow for sample multiplexing 1 \(the following table), and 4) primer binding sites specific to the sequences flanking the library sites \(F: CGGCGGAATATGTGGAAC, R: CAGCCGCTGCCCATT) Barcode ID Sequence Fwd1 ACCTCA Fwd2 CCTTCT Fwd3 CTACCT Fwd4 CTCACA Fwd5 CTCGAA Fwd6 GGAGAA Fwd7 GGATCT Fwd8 TCGACT Fwd9 AACACC Fwd10 GAGTCA Fwd11 GTAACC Rev1 ATGTCC Rev2 CAACTC Rev3 CACAGT Rev4 CGTAGA 42. Setup PCR reactions to amplify the region of interest on PylS plasmid KAPA HiFi HotStart ReadyMix \(2x) 10 µl Plasmid extracted from Step 34 and 40 1 µl Forward primer \(50 µM) 0.3 µl Reverse primer \(50 µM) 0.3 µl H2O 8.4 µl Total 20 µl Use primers with different barcode \(listed in step 41) combinations for different sample amplification: Sample Forward barcode Reverse Barcode Pre 1 Fwd1 Rev1 Pre 2 Fwd2 Rev1 Pre 3 Fwd3 Rev1 Post -1 Fwd4 Rev1 Post -2 Fwd5 Rev1 Post -3 Fwd6 Rev1 Post +1 Fwd7 Rev1 Post +2 Fwd8 Rev1 Post +3 Fwd9 Rev1 43. Run PCR reactions \(18 cycles) 95 °C 3 min 95 °C 20 sec 60 °C 15 sec 72 °C 72 °C 1 min 10 min 4 °C ∞ 44. Purify PCR products by gel extraction using the E-Gel SizeSelect Agarose Gel 2% kit following the manufacturer's instructions. In brief, insert the gel into the E-Gel iBase Power System, load PCR reactions into wells and start electrophoresis using the built-in power supply. Monitor the DNA migration using the built-in blue light transilluminator. Pause the run when the desired bands migrate into the wells for extraction. Transfer the well contents to eppendorf tubes. 45. Dilute the extracted PCR products 1 in 106 in qPCR Dilution Buffer. Perform sequential dilution by diluting 10 µl extracted PCR products in 990 µl buffer. Repeat the dilution twice more. Keep the diluted DNA samples on ice and use within two hours 46. Setup qPCR reactions with Kapa Illumina Complete Kit Universal KAPA SYBR Fast qPCR Master Mix \(2x) 5 µl Diluted template DNA \(from Step 45) 4 µl ROX

Low 0.2 µl H2O 0.8 µl Total 10 µl Setup triplicates reactions for each template DNA sample. Also setup reactions for DNA standards provided in the kit. 47. Run qPCR reactions \(40 cycles) 95 °C 5 min 95 °C 30 sec 60 °C 45 sec 48. Fit CT values against log-transformed DNA standard concentrations. Use the resulting the linear regression function to infer the concentrations of the other DNA samples with the respective CT values. 49. Use water to dilute each DNA sample to a final concentration of 2 nM. 50. Mix same volume of 2 nM DNA from each sample. Freeze and keep at -20°C before use Deep Sequencing Setup 51. Follow the Illumina MiSeq Reagent Kit v2 \(300-cycle) instructions to prepare 2 nM DNA adapter-tagged DNA for next-generation sequencing. In brief, mix 10 µl DNA sample \(pooled 2 nM sample from Step 50) with 10 µl freshly prepared 0.2 N NaOH. Briefly vortex and centrifuge at 250 g for 1 min. Incubate for 5 min at RT 52. Add 10 µl 200 mM Tris-HCl pH = 7. Briefly vortex and centrifuge at 250 g for 1 min 53. Dilute with 970 µl HT buffer to a final DNA concentration of 20 pM 54. Mix 600 µl DNA sample from Step 53 with 400 µl HT buffer to further dilute DNA to 12 pM 55. For PhiX control prepration, dilute the 10 nM PhiX Sequencing Control v3 to 2 nM in H2O. Perform Step 51-54 to further prepare the 12 pM PhiX control. 56. Spike in 10% PhiX by mixing 900 µl 12 pM DNA sample from Step 54 with 100 µl 12 pM PhiX control from Step 55. Load the combined 1 ml sample into MiSeq Reagent Kit v2 \(300-cycle). Set up the rest of deep sequencing run according to manufacturer's instructions. Perform deep sequencing on a MiSeq Sequencer. Data Analysis 57. Process raw deep sequencing reads using bespoke scripts written in the programming language Python. Each paired end read was first trimmed to 141 nucleotides. Paired end read 1 was concatenated with the reverse-complement of paired end read 2. The assembly was matched using regular expressions against a pattern specifying the expected sequence string with degeneracy specified for the sequence barcode and for each library site. For matching sequences, the degenerate sites were extracted and collected into a comma-separated file together with the detected barcode ID 58. Construct an m x n counts table N of m unique mutants in n conditions. Each entry kij corresponded to mutant i in condition j. To identify the m mutants and their corresponding read counts, all unique rows in the comma-separated file were collected and their occurrences were counted. 59. Implement the DESeq algorithm 2 in the Matlab R2014a \(Mathworks) software using instructions provided online \(http://uk.mathworks.com/help/bioinfo/examples/identifying-differentially-expressed-genes-from-rna-seq-data.html). Use the DESeq algorithm to identify sequences that are enriched in Post +1/+2/+3 relative to Post -1/-2/-3. Correct the resulting p-values for multiple testing using the procedure by Benjamini and Hochberg, yielding the False Discovery Rate \(FDR) 3. Phenotypic validation 60. For validation of the hits, order DNA primers to mutate codon 267/271/274 on the wild type PylS sequence to the mutant sequences that are statistically significant. Setup PCR reactions PrimeSTAR Max mix \(2x) 25 µl pKW-PylS-PylT \(50 ng/µl) 0.5 µl Forward primer \(50 µM) 0.5 µl Reverse primer \(50 µM) 0.5 µl H2O 23.5 µl Total 50 µl 61. Repeat step 3-8 to purify, digest, and ligate the PCR products 62. Transform the concentrated DNA to MegaX DH10B electrocompetent cells by electroporation and recover in pre-warmed SOC medium for 60 min at 37°C 63. Spread 60 µl recovered cells onto LB agar plate containing 50 µg/ml kanamycin and incubate the plates at 37°C overnight 64. Pick three colonies from each plate and inoculate individually into 5 ml fresh LB medium containing 50 µg/ml kanamycin and incubate at 37°C overnight with constant shaking at 220 rpm 65. Pellet the cells by centrifuging at 4,000 g for 12 min. Extract the plasmids from each cell pellet using Qiagen Miniprep Kit following the manufacturer's

instructions 66. Confirm the mutation by sequencing the plasmid 67. Repeat Steps 60-66 to introduce mutations on codon 241 and codon 313 if necessary 68.Transform pBAD_sfGFP\(150TAG)-His6 to MegaX DH10B electrocompetent cells by electroporation and recover in pre-warmed SOC medium for 60 min at 37°C 69. Spread 30 µl recovered cells onto LB agar plate containing 25 µg/ml tetracycline and incubate the plates at 37°C overnight 70. Pick a single colony and inoculate into 5 ml LB medium containing 25 µg/ml tetracycline and incubate at 37°C overnight with constant shaking at 220 rpm 71. Dilute the overnight culture into fresh 1 L LB medium. Prepare electrocompetent DH10B cells by repeating steps 27-31 72. Aliquot 50 µl competent cells into 2 mm electroporation cuvette. Add 1 µl PylS plasmid containing wild type or mutated sequence from Step 66 into each cuvette and electroporate with the setting of 2,500 V. Immediately add 1 ml SOC into each cuvette and incubate at 37°C for 60 min with constant shaking at 220 rpm. Snap freeze the rest of the competent cells in liquid nitrogen. These cells can maintain their competency for at least 6 months at -80°C 73. After incubation, spread 60 µl recovered cells onto an LB agar plate containing 25 µg/ml tetracycline and 50 µg/ml kanamycin. Incubate the plates at 37°C overnight 74. Pick a single colony from each plate and inoculate into 5 ml fresh LB medium containing 25 µg/ml tetracycline and 50 µg/ml kanamycin. Incubate at 37°C overnight with constant shaking at 220 rpm 75. Dilute 500 µl overnight culture into 25 ml fresh LB medium containing 10 µg/ml tetracycline, 25 µg/ml kanamycin, 0.2% \(w/v) L-Arabinose, and 1 mM PCK. Incubate at 37°C for 16hrs with constant shaking at 220 rpm 76. Pellet down cells by centrifugation at 7,000 g for 10 min 77. Resuspend cells in 1 ml His-tag Lysis Buffer and transfer to 1.7 ml eppendorf tube. Rotate end-by-end for 15 min at RT 78. Centrifuge at 20,000 g for 20 min 79. Transfer the supernatant to a new tube, add 30 µl Ni-NTA resin pre-washed three times with His-tag Lysis Buffer. Rotate end-by-end for 10 min at RT 80. Centrifuge at 1,000 g for 1 min. Carefully remove supernatant without disturbing the resin 81. Resuspend resin with 1 ml His-tag washing Buffer. Rotate end-by-end a few times by hand 82. Repeat Step 80-81 for four more times 83. Centrifuge at 1,000 g for 1 min. Carefully remove supernatant without disturbing the resin. Resuspend resin with 100 µl His-tag Elution Buffer. Rotate end-by-end a few times by hand. Centrifuge again at 1,000 g for 1 min. 84. Transfer the supernatant to a new tube. Estimate the protein concentration by Nanodrop. 85. Dilute the protein to the final concentration of 0.1 mg/ml with H_2O and inject 5 µl to electrospray ionization mass spectrometer. Use 0.2 % formic acid in H_2O as buffer A, and 0.2 % formic acid in acetonitrile \(MeCN) as buffer B. Samples were injected into Phenomenex Jupiter C4 column \(150 × 2 mm, 5 µm) and subsequently into the mass spectrometer using a fully automated system. 86. Theoretical average molecular weight of proteins with unnatural amino acids was calculated by first computing the theoretical molecular weight of wild-type protein using an online tool \(http://www.peptidesynthetics.co.uk/tools/), and then manually correcting for the theoretical molecular weight of unnatural amino acids.

References

1. Faircloth, B. C. & Glenn, T. C. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. PLoS ONE 7, e42543 \(2012). 2. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106 \(2010). 3. Benjamini, Y.

& Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society Series B \(… \(1995). doi:10.2307/2346101