SYNTHETIC BIOLOGY TOOLS FOR TARGETED INCORPORATION OF NON-CANONICAL AMINO ACIDS INTO CELLULAR PROTEINS

Thesis by

Alborz Mahdavi

In Partial Fulfillment of the Requirements for the degree

of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2015

(Defended December 16th, 2014)

This thesis is dedicated to my family, friends, and mentors.

ACKNOWLEDGEMENTS

This work could not have been accomplished without the help of many individuals; I am indebted to all of you who made my graduate career more fulfilling and successful. I would like to first thank my parents, who have made many sacrifices to ensure I had superb educational and cultural opportunities. Their strength is my inspiration and my success belongs to them.

I am most grateful to my PhD advisor, Prof. David Tirrell, the best PhD advisor in the world. I am so fortunate to have learned from this brilliant scientist and amazing human being. Some of my best memories at Caltech are scientific and life lessons that I learned from Dave through many conversations. Hopefully through the course of my career I can prove that I am deserving of all the efforts that Dave invested in making me a better person and a great scientist. Working in Dave's laboratory was an amazing experience. Dave cares so much about his students and their development as an individual and a scientist. I can't think of a better person to work for and learn from. Dave's clarity of thought, acuity, perceptiveness, and modesty are personal qualities that I strive to develop in myself. He is one of my most important role models and I hope to learn much more from him as I progress through my career.

I am grateful to my thesis committee for their contributions to this work and for being so helpful and approachable throughout my graduate career. I am grateful to Prof. Sarkis Mazmanian for always being so supportive, for scientific discussions, and in particular for providing critical support in our collaboration to develop a new method for the study of host-pathogen interactions. I am grateful to Prof. Michael Elowitz for encouraging me to focus on important problems. Prof. Ali Hajimiri was so helpful throughout my PhD, always approachable, and extremely supportive in our collaborations. I am grateful to Prof. Frances Arnold for her support throughout my PhD; I not only learned a lot about protein engineering and directed evolution from her, but was also fortunate to have the opportunity to collaborate with her research group on the development of protein-based materials.

Perhaps the most important aspect of being at Caltech is the opportunity to work with, and learn from, so many talented and inspiring individuals. I am grateful for my interactions with all the Tirrell group members, including Janek Szychowski, Jeremiah Johnson, Bradley Olson, Wenbin Zhang, Ayelet Lesman, Cole DeForest, Brett Babin, John Bagert, Larry Dooling, Kat Fang, Samuel Ho, Seth Lieblich, Andy Lim, Peter Rapp, Michael Srienc, Shannon Stone, Kai Yuet, Weslee

Glenn, Beverly Lu, Nicholas Ball, Maren Buck, Tamara Ursem, Jim van Deventer, Graham Hamblin, Mark van Eldijk, Frank Truong, Caglar Tanrikulu and Kimberly Beatty, and Anne Hormann. Many thanks in particular to Weslee Glenn for help with editing sections of this thesis.

I would also like to thank the bioengineering graduate students of my entry year at Caltech, including Jonathan Sternberg, William Dempsey, Sungwook Woo, Indira Wu, Nakul Reddy, Hesham Azizgolshani, and Niema Pahlevan, as well as Mohsen Chitsaz and Geoff Lovely in the biophysics program. I have had some incredible collaborators during my PhD studies. For work reported in chapter II in collaboration with the Mazmanian lab I would like to thank Prof. Olaf Schneewind at the University of Chicago and Arya Khosravi and Sara McBride at Caltech. For chapter III of this thesis I would like to thank Prof. Jonathan Silberg at Rice University and Prof. Rustem Ismagilov and Songzi Kou at Caltech. For publications outlined in appendices C, E, G, and F, I would like to thank Prof. Arthur Riggs at City of Hope. For our collaboration with Hajimiri group, outlined in appendix B, I would like to thank Prof. Hua Wang. I would like to thank Prof. Wenbin Zhang and Prof. Fei Sun for the work outlined in appendix D, which was a collaboration with the Arnold group at Caltech. I would also like to thank the proteome exploration laboratory of the Beckman institute at Caltech. In particular I would like to thank Prof. Sonja Hess for all her support, and PEL members Robert Graham, Mike Sweredoski, Annie Moradian, and Geoff Smith as well as Prof. Raymond Deshaies. I would also like to thank Shirley Pease at the transgenic mouse facility and Fred Tan in the Elowitz lab for helpful discussions regarding mouse embryonic stem cell work outlined in chapter V. I am very grateful to Prof. Robert Langer, Prof. Ali Khademhosseini, Prof. Jeff Karp, Prof. Utkan Dmirci, Prof. Mehmet Fatih Yanik, and Prof. Mikhail Shapiro for their support during my MIT years. I am also indebted to my mentors in Toronto, including Prof. Parham Aarabi, Prof. Peter Zandstra, and Prof. Michael Sefton. I am very grateful to Prof. Anand Asthagiri and Prof. Niles Pierce, who ensured my success upon entering Caltech, and a big thanks to Linda Scott in bioengineering. I have had the pleasure of working with, and learning from, several superb undergraduate students who conducted their summer research with me. Thank to Granton Jindal, Cathy Dong, and Megan Lo for being such great SURF students. I am also indebted to numerous science fiction writers and futurists whose books inspired me and will continue to serve as inspiration for the future.

Thanks to the National Science and Engineering Research Council of Canada (NSERC) post-graduate scholarship and the Rosen Fellowship for financial support.

ABSTRACT

Proteins mediate many essential functions in cells, and methods to profile cellular proteins are of great interest for biological discovery. Whereas all of the cells in an organism share the same genome, the landscape of proteins (the proteome) varies between different cell types and over the lifetime of the organism. Rapid progress in mass spectrometers is enabling the detailed analysis of cellular proteomes. Whereas better instruments increase coverage, throughput, and measurement precision, new chemical reporters, metabolic tags, and synthetic biology techniques are required to enhance the specificity and spatiotemporal resolution of protein labeling and detection. This work introduces methods for cell-selective proteome analysis through the incorporation of non-canonical amino acids into newly synthesized proteins.

Chapter I provides an overview of current technologies for translational profiling and proteomic analysis in cells. Strategies for the residue-specific incorporation of non-canonical amino acids and bioorthogonal non-canonical amino acid tagging are discussed. Chapter II introduces a new approach for the identification of secreted bacterial proteins from infected host cells using non-canonical amino acid labeling. This work demonstrates an application of cell-selective proteome labeling. Selectivity is achieved through controlled expression of a mutant aminoacyl tRNA synthetase (aaRS) enzyme that enables the metabolic incorporation of a non-canonical amino acid.

Ideally, the activity of multiple genes should be used to genetically control the extent of proteome labeling in cells. This is useful because many cell states are characterized by the activity of multiple genes and identified based on the expression of several proteins. Therefore chapter III introduces a novel approach to control proteome labeling as a function of multiple promoters using a genetically encoded AND gate based on a bisected methionyl-tRNA synthetase, a class I aaRS. Cellular protein labeling occurs only upon activation of two different promoters that drive expression of the N- and C-terminal fragments of this bisected aaRS. The utility of this tool is demonstrated by the selective labeling of proteins in subpopulations of bacterial cells in a laminar-flow microfluidic channel.

Chapter IV extends the cell-selective incorporation of non-canonical amino acids from bacterial systems to mammalian cells by introducing a mutant mammalian methionyl-tRNA synthetase for cell-targeted proteome labeling. This enzyme is genetically encoded and can be conditionally activated for time-resolved and cell-targeted proteome analysis in a variety of different mammalian cell types. Chapter V uses this enzyme for lineage-specific proteomic analysis of mouse embryonic stem cells during differentiation to cardiac and mesoderm lineages. This approach for lineage-

specific protein labeling enables the unbiased and comprehensive analysis of proteomic changes that occur during stem cell differentiation and cell-fate commitment.

Appendices A-G provide brief summaries of publications and research efforts during my PhD that are not directly related to this thesis. These publications are the result of a number of collaborations that I have been fortunate to be involved with during my graduate research.

The technologies and methods introduced in this thesis provide versatile tools for the comprehensive and unbiased detection and identification of newly synthesized proteins in complex multicellular systems. Time-resolved, genetically encoded, and spatially defined non-canonical amino acid incorporation enables the identification of proteins involved in cell-cell interactions and the proteins made during specific cell states.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Figure *Page*

CHAPTER I

Targeted Approaches for Comprehensive Profiling of Newly Synthesized Proteins in Cells

Abstract

Technological developments in DNA sequencers and mass spectrometers have enabled the study of cellular functions with unprecedented coverage, throughput, and precision. Cells have evolved sophisticated machinery to accurately control protein synthesis, and oftentimes mRNA levels are an inadequate proxy for protein production. Therefore, comprehensive methods to analyze protein synthesis are being developed. In this review, we present an overview of these approaches, focusing specifically on the subset of methods that enable the investigation of protein synthesis in a cell-selective manner. We provide examples of how these tools are being used to answer different biological questions at multi-cellular and sub-cellular scales.

Introduction

Proteins mediate many essential functions in cells, and methods to profile cellular proteins are of great interest for biological discovery. Whereas all of the cells in an organism share the same genome, the landscape of proteins (i.e, the proteome) varies between different cell types and over the lifetime of the organism. Therefore, proteome dynamics are often cell type specific and spatially and temporally regulated. As such, measures of protein synthesis in various cell types can be very useful for biological discovery and inform new approaches towards understanding and engineering the regulation of cellular functions. Understanding cellular regulatory mechanisms is also important because disruption in protein homeostasis is often the root cause of many diseases, including a large number of neurodegenerative diseases,[1,2] metabolic diseases,[3,4] developmental abnormalities, and inherited disorders.[5]

Transcriptomics, the identification and quantitation of mRNAs, provides a readout of actively transcribed genes and is often used as a proxy for protein synthesis. Analysis of mRNA levels has enabled investigators to develop a global perspective of gene activity in different cells and to monitor the induction of different genes in response to various stimuli. Modified transcriptional profiling methods, such as tagging mRNAs with thiouracil (TU-tagging), enable the cell specific profiling of actively transcribed genes[6]. However, a more sophisticated understanding of proteomic changes in cells requires post-transcriptomic approaches that measure translation products and track protein synthesis rather than mRNA levels. Notably, various control mechanisms influence translation of mRNAs to the extent that mRNA levels poorly correspond to protein levels[7-9]. Moreover, tracking mRNAs provides little information about proteomic changes resulting from localization (i.e., sites of protein synthesis), transport, and post-translational modifications. Indeed, these mechanisms have profound effects on protein function and are increasingly the targets of new molecular therapies[10,11]. Therefore, methods are needed to accurately detect and measure protein synthesis rather than transcripts. Traditional techniques such as genetic screens and custom assays to assess small libraries of candidate proteins can be laborious and non-comprehensive. Advances in DNA sequencers and mass spectrometers are rapidly changing the way protein synthesis is being studied. The comprehensive analysis of thousands of proteins in biological samples is becoming the norm rather than the exception. Technological advances in instrumentation are an important enabling factor for these approaches. For example, high-resolution liquid chromatography tandem mass spectrometry (LC-MS/MS) instruments can routinely identify thousands of proteins in a given sample and provide high-quality information about co- or post-translational modifications to those proteins. These

approaches provide unbiased and comprehensive datasets that are ideally suited to machine-learning and data-analytics techniques that are being increasingly used in the biological sciences.

Proteomics platforms have rapidly advanced in the last decade. Advances in sequencing and proteomics have and will continue to beat Moore's law in the years to come.[12-14] These improvements are complemented by the development of new chemical and synthetic biology tools that are specifically designed to leverage and utilize the capabilities of new instruments. Whereas instrumentation advances have enabled increased coverage, throughput, and precision, better chemical reporters and metabolic tags have enhanced the specificity and spatiotemporal resolution of labeling and detection. Here we provide an overview of new approaches to profile protein synthesis in cells. We also highlight a selective subset of chemical and synthetic techniques that enable the identification and quantitation of newly made proteins in different cell types. Notably, these techniques enable study of protein synthesis with varied spatial and temporal resolution in multicellular contexts as well as sub-cellular scales. We first review recent advances in translational profiling approaches that enable cell targeted and organelle-specific profiling. We then provide an overview of amino acid metabolic labeling methods to analyze newly synthesized proteins and highlight the use of bio-orthogonal chemistries to develop chemical reporters and improve quantitative proteomics capabilities. We close by providing an outlook on future directions to enhance the labeling, detection, and quantitation of proteome dynamics in cells that will ultimately facilitate biological discovery and inform subsequent engineering efforts.

Translational Profiling Approaches

mRNA translation often occurs in polyribosomes (often referred to as polysomes), the centers of protein production in cells.[15] Polysomes consist of mRNA-bound ribosomes and RNA-binding proteins that repress or activate translation and influence levels of protein synthesis. Polysome complexes are associated with mRNA sequences that are being translated at any given time in the cell. Sequencing RNA-bound polysomes enables the identification of proteins that are actively being synthesized. Translation profiling has been used in a series of techniques to study protein synthesis *in vitro* and *in vivo*.[16]

Polysome Profiling

Polysome profiling involves separating mRNAs based on the degree of ribosome loading. In this technique, mRNAs are fractionated based on the density of bound ribosomes, converted to

cDNAs and then sequenced (Figure 1.1A).[17] Polysome profiling has been applied to a variety of different cell types including mammalian,[18] plant,[19] and bacterial cells.[20] Polysome profiling provided some of the early examples of how genome-wide and comprehensive translation analysis can provide mechanistic insight that would otherwise not be obtained from single-transcript studies. For example, analysis of *L. lactis* demonstrated that transcript-specific translation changes occur during exponential growth [20]. Gilbert and coworkers studied the effects of environmental stress on protein translation in yeast, showing that glucose starvation leads to the removal of mRNAs from polysomes and into P-bodies. They further showed that reactivation of translation upon glucose addition only occurs in a selective subset of transcripts, wherein transcripts whose translation is rate-limiting become transiently paused in rapidly changing cell environments.[21] Notably, however, polysome profiling is not cell-selective. Moreover, it is difficult to distinguish between translationally active ribosomes and stalled ribosomes engaged with inactive mRNAs.[22]

Translating Ribosome Affinity Purification

Whereas polysome profiling enables actively translating mRNAs to be tracked, the technique lacks cell-specificity. Although cell sorting can be combined with polysome profiling to achieve cell selectivity, manipulation of cells during sorting may influence cellular functions or alter the transient profile of actively translating mRNAs[23-25]. Therefore, a logical extension of polysome profiling has been the use of genetically encoded and epitope-tagged ribosomes for affinity purification.[26] Tagging ribosomes with epitopes in cells of interest enables the selective isolation of polysomes from those cells. Translating ribosome affinity purification (TRAP)[27] is a method in which ribosomal protein L10a, a component of the 60S ribosomal unit, is fused to EGFP (Figure 1.1B). Heintz and coworkers used tissue-selective expression of this construct from bacterial artificial chromosomes to restrict labeling to specific neuronal subpopulations in the mouse brain. They affinity purified the tagged polysomes using an anti-GFP antibody to show that different subpopulations of neurons have different translation profiles. In addition to its use in the translation profiling of neurons[28] and other tissues in the mouse[29,30], this approach has been used in Zebrafish,[31,32] Drosophila,[33] and plants,[34-36] and was also modified for cre-lox conditional activation in mice.[24] A similar strategy called *in vivo* ribosome tagging (Ribo-Tag) uses epitope-tagged ribosomal subunits. In Ribo-Tag, a mouse model was developed that expresses a Cre-activated hemagglutinin-tagged ribosomal protein L22 for tissue-selective expression and affinity purification of epitope-tagged polysomes.[37,38] Although both TRAP and Ribo-Tag can be used for cell-selective translation profiling, these techniques do not provide information regarding the

positions of ribosomes on the mRNA transcripts or the number of ribosomes on each transcript. Ribosome stalling confounds these results and requires that other techniques be developed to determine the positions of ribosomes on active mRNA transcripts.

Ribosome Profiling

Weissman and coworkers developed ribosome profiling by digesting mRNA transcripts with nucleases, which results in ribosome-protected mRNA fragments that delineate the precise positioning of ribosomes along the mRNA sequence (Figure 1.1c).[39] High-throughput sequencing of ribosome-protected mRNA fragments provides a measure of ribosome density for each mRNA, which corresponds to the overall rate of translation. Because ribosome profiling provides nucleotide resolution of ribosomal positioning, this technique has enabled the identification of many upstream open reading frames, non-AUG start sites, and stop-codon read-throughs.[40] This approach has been used to determine ribosome pausing and biased codon usage.[41] The strategy was used for a pulse-chase type of analysis and applied to mouse embryonic stem cells to identify ribosome pause sites and a class of short polycistronic ribosome-associated coding RNAs that encode small proteins.[42] Ribosome profiling has been used to study plants,[43,44] Drosophila,[45] host-pathogen interactions,[46,47] and translation in subcellular compartments.[48] Enhanced subcellular translation profiling was recently achieved by proximity-specific ribosome profiling, which utilizes an encoded Avi-tagged ribosomal subunit and an organelle-targeted and spatially-localized biotin ligase, BirA, to tag organelle-associated ribosomes. This technique ultimately permits the subcellular identification of spatially defined protein synthesis in cells.[49,50] Ribosome footprints provide a detailed view of translation in cells. Proximity-specific ribosome profiling provides a valuable tool in the study of localized protein synthesis. Notably, ribosome profiling amplifies transcripts, provides excellent coverage, and enables the detection of low abundance transcripts. However, as with any technique, certain applications require workarounds due to the inherent limitations of the method. For example, restricting the labeling time is important in identifying newly synthesized proteins with good temporal resolution. In most ribosome profiling applications translation inhibitors such as harringtonine and cycloheximide are used. However, the ability to deliver such chemicals to cells, particularly for *in vivo* applications, may limit the temporal resolution of such analyses. Moreover, analysis of mRNA fragments does not directly tag newly made polypeptides and therefore cannot be used to identify and quantitate proteins that are transported between cellular compartments or between different cell types and tissues. Additionally, translation profiling cannot be used to track newly made proteins that may undergo

various post-translational modifications. For these types of analysis, methods are needed to directly detect newly synthesized proteins and nascent polypeptides chains.

Protein Synthesis Detection Strategies

Labeling and detecting newly synthesized proteins requires chemical reporters that are metabolically incorporated into the protein sequence. For example, modified amino acids that can be incorporated into proteins are ideally suited for this type of analysis. Isotopically labeled amino acids provided one of the first examples of metabolic protein labeling in cells.[51,52] Substrate mimics that are incorporated into polypeptide chains can also serve as protein synthesis reporters. Advances in bio-orthogonal chemistries- chemical reactions that can be carried out in cellular environments with specificity and selectivity- have enabled the development of chemical biology based probes that can be used to investigate protein synthesis in cells (Figure 1.2). In particular, click chemistry has enabled the development of a large number of bio-conjugations that can be deployed in various chemical biology applications. For a review of bio-orthogonal reactions and bio-conjugations, we refer the reader to recent reviews by the Finn,[53] Bertozzi,[54-56] and Chin[57] groups. Here we focus on labeling strategies that are specifically designed for the analysis of protein synthesis.

Labeling Nascent Polypeptide Chains

Inhibitors of protein synthesis constitute a class of chemical reporters that are used to detect nascent polypeptide chains. For example, labeling with puromycin, an inhibitor of protein synthesis that can be incorporated into nascent polypeptide chains, yields truncated proteins that can be detected and identified.[58,59] Puromycin is an aminonucleoside antibiotic that resembles the 3'tail of a charged (aminoacylated) tRNA; the ribosome incorporates puromycin into the C-terminus of a growing polypeptide chain where it stops translation.[60] Several different approaches have been devised based on affinity tagging with puromycin in order to detect and identify translation products . For example, Roberts and coworkers developed puromycin variants containing fluorescent and biotin conjugates to detect nascent polypeptide chains.[61] In a similar manner, Pierre and coworkers developed a method called surface sensing of translation (SunSET), which relies upon the immunodetection of puromycin (Figure 1.3A).[62] Stein and colleagues developed a method called puromycin-associated nascent chain proteomics (PUNCH-P) (Figure 1.3B), where they deploy biotin-conjugated puromycin for affinity enrichment and subsequently use mass spectrometry to identify nascent polypeptides.[63] Salic and coworkers developed an alkyne analogue of puromycin (Figure 1.2, structure **1**) that enables the

chemoselective and bioorthogonal tagging of nascent polypeptides with affinity tags (Figure 1.3C).[64] Although puromycin-based detection is not a cell-selective technique, these methods represent an important class of chemical strategies that can be used to detect protein synthesis. Utilizing an inhibitor of protein synthesis to profile translation may seem counter-intuitive; however the advent of this approach highlights the importance of developing good chemical biology tools and comprehensive methods to analyze newly made proteins.

Isotopically Labeled Amino Acids for Proteomics

Isotopically labeled amino acids, particularly ^{35}S-methionine and cysteine, have enabled the detection of protein synthesis in cells. Based on this concept, Mann and coworkers developed a method titled stable isotope labeled amino acids in cell culture (SILAC) by utilizing heavy arginine and lysine.[65] SILAC differs from radiolabeling in that it enables relative quantitation of proteins between samples. SILAC is among the first chemical reporter technologies specifically designed for use with modern mass spectrometers. Other notable quantitative proteomics approaches in this category include ICAT[66] and iTRAQ.[67] However, in contrast to iTRAQ and ICAT, which use isotopically labeled tags that react with the side chains of amino acids, SILAC metabolically incorporates isotopically labeled amino acids into nascent proteins. In SILAC, cells grown in normal amino acids and under one set of conditions are compared with cells grown in the presence of a heavy amino acid (often arginine and lysine) and under a different set of conditions. The mass shift associated with isotope labeling between samples is detected on a mass spectrometer and used to provide relative quantification between protein levels in samples (Figure 1.3D). Similarly, pulsing in heavy amino acids and subsequently analyzing the samples by mass spectrometry can be used to identify newly made proteins.[68] Pulsed-SILAC (pSILAC), a modified version of SILAC, utilizes a pulse with isotopically labeled amino acids,[69] which obviates the need to continuously culture cells in heavy amino acids. Importantly, pSILAC can be used to detect proteins synthesized in response to different stimuli. SILAC has been applied to a variety of organisms including yeast,[70] plants,[71] Drosophila,[72] and C. *elegans.*[73] SILAC has also been used to label whole mice for quantitative proteomics.[74,75] SILAC has been modified to study cell-cell interactions (trans-SILAC)[76] and to identify secreted proteins in an approach called secretome-derived isotope tags (SDIT).[77] SILAC can also be combined with various chemical biology-based labeling approaches that are discussed below. Notably, even though SILAC can be combined with fractionation techniques for subcellular proteomics,[78] this approach does not label newly made proteins in a cell-selective fashion. Moreover, the affinity enrichment of low abundance, newly synthesized, proteins is not possible through this method because isotopically

labeled amino acids do not contain orthogonal chemical handles. However, several new techniques have been devised recently to address the lack of cell-specificity in current proteomic analyses. For example, a method titled isolation of nuclei tagged in specific cell types (INTACT)[79] has been devised for cell-specific proteomic analysis using purified nuclei labeled *in vivo* with biotin.[80] Miller and coworkers developed cell type-specific labeling using amino acid precursors (CTAP),[81] a method that utilizes genetically encoded enzymes that convert isotopically labeled lysine precursors to uncaged lysine, which can be incorporated in cells expressing these enzymes (Figure 1.3E).[81] Only cells expressing these enzymes are able to uncage the lysine precursors. Therefore, the controlled expression of this enzyme enables the cell-selective labeling of proteins with isotopically labeled lysine. This technique was used to investigate cell-cell interactions in continuous culture.[81,82] Although such approaches confer cell selectively for proteomic analyses, the ability to detect newly synthesized proteins of low abundance proteins in complex cellular mixtures is limited by the lack of suitable enrichment tags. Additional methods are needed to address these limitations.

Non-Canonical Amino Acids and Chemical Biology-Inspired Methods

Whereas isoptopically labeled amino acids provide mass signatures that can be detected for proteomic identification and quantitation, non-canonical amino acids (ncAAs) provide chemical handles that can be selectively tagged for the detection and affinity purification of newly synthesized proteins. Enrichment simplifies subsequent analyses by enabling the separation of newly synthesized proteins from abundant pre-existing proteins, thereby greatly reducing sample complexity. Non-canonical amino acids are ideal for tracking protein dynamics because they are readily taken up by cells through amino acid transporters, their small size diminishes interference with protein function, and they can be covalently tagged, which permits more stringent affinity purification. Taking advantage of these desirable characteristics, bioorthogonal non-canonical amino acid tagging (BONCAT) was developed to label, detect, and identify newly synthesized proteins.[83] BONCAT utilizes residue-specific incorporation of non-canonical amino acids. Specifically, a proteogenic amino acid, such as methionine, is replaced by its non-canonical counterpart throughout the proteome during protein synthesis. Labeling is achieved without *a priori* knowledge of protein identity, and the concentration of the non-canonical amino acid can be used to control the extent of incorporation. The first example of BONCAT utilized the non-canonical amino acid Aha (Figure 1.2A, **2**), a methionine surrogate, to label newly synthesized proteins. Aha incorporation does not alter the proteome.[84] Moreover, the azide side chain of Aha can be tagged with alkyne-functionalized affinity reagents through click chemistry,[85] allowing

separation of newly synthesized proteins from pre-existing proteins by affinity chromatography (Figure 1.3F). After separation, proteins are identified by tandem mass spectrometry (MS). Temporal control of labeling is achieved though addition of Aha to the growth medium (Figure 1.4A). However, because all cells incorporate this non-canonical amino acid, it is not ideally suited for cell-selective proteomics. Cell-selective BONCAT was developed by deploying non-canonical amino acids that require a mutant enzyme [aminoacyl tRNA synthetase (aaRS)] to be charged to an endogenous tRNA. The controlled expression of the mutant aaRS enables selective incorporation of non-canonical amino acids into cells of interest (Figure 1.4B).

Initial examples of cell-selective BONCAT were demonstrated using the azide-functionalized methionine surrogate Anl,[86] and it was later also performed with ncAAs containing an alkyne functionality (Figure 1.2B, **5-6**).[87] In principle, the same approach can be used with non-methionine surrogates (e.g., phenylalanine or leucine analogues) (Figure 1.2B, **7-12**). Such an expanded set of ncAAs can be used to label multiple subpopulations of cells with distinct sets of reactions (Figure 1.4C). For example, two different mutant aaRSs were used to incorporate the non-canonical amino acids Anl (Figure 1.2A, **4**) and Pra (Figure 1.2A, **5**) into distinct bacterial subpopulations.[88] An expanded set of ncAAs enables the use of a wide range of chemical reactions to tag proteins. Alkyne- and azide-functionalized ncAAs can be tagged using the copper-catalyzed azide-alkyne cycloaddition reaction (Figure 1.2C); azide functionalized ncAAs can also be tagged through copper-free reactions with alkyne or alkynyl probes (Figure 1.2D).[89-91] Copper-free reactions with strain-promoted cyclooctynes are particularly useful for live-cell and *in vivo* applications,[56,90,92] and also well-suited for affinity enrichment and proteomic analysis.[93] ncAAs containing aryl-halides (Figure 1.2A, **9, 10**) can be tagged using modified Suzuki-Miyaura cross-coupling reactions (Figure 1.2E).[94] Ketone-containing ncAAs (Figure 1.2A, **11**,**12**) can be reacted with hydrazide functionalized tags (Figure 1.2F).[95] Chin and coworkers recently incorporated a ncAA containing 1,3 disubstituted cyclopropene, which was reacted with tetrazine tags using the inverse electron demand Diels-Alder reaction (Figure 1.2G).[96] The incorporation of this ncAA required co-expression of a mutant aaRS and its cognate tRNA; the approach was titled stochastic orthogonal recoding of translation with chemoselective modification (SORT-M). SORT-M differs from BONCAT in that BONCAT only requires expression of a mutant aaRS in cells of interest. While both tetrazine-cyclopropene and strained cyclooctyne-azide reactions are fast enough for most proteomic applications, strained cyclooctynes and cyclopropenes may react with thiols and result in non-specific labeling;[97] therefore, additional care must be taken to protect free thiols and reduce non-specific labeling when using these chemistries. Taken together, these

examples demonstrate that a variety of chemical coupling strategies have been developed for BONCAT, and the choice of a particular chemistry depends on the availability of reagents, ease of use, and the subsequent proteomic analysis. To date, Aha has been the most widely used ncAA for BONCAT applications (Figure 1.2A, **2**). However, we anticipate that cell-selective BONCAT using Anl (Figure 1.2A, **4**) will be increasingly used for various proteomics applications requiring cell selectivity and tissue specificity.

Non-Canonical Amino Acids and Quantitative Proteomics

One of the important advantages of BONCAT is that it can be easily interfaced with other protein labeling strategies, including quantitative proteomic labeling methods such as SILAC. In a combined BONCAT SILAC approach, proteins are simultaneously labeled with both a non-canonical amino acid for BONCAT labeling and an isotopically labeled amino acids for SILAC labeling. After BONCAT enrichment and shotgun proteomics, quantitative proteomic search engines such as MaxQuant[98] are used to look for SILAC pairs in the mass spectra, enabling quantitation of newly made proteins across different conditions (Figure 1.4D). BONCAT reduces sample complexity and enables the detection of low-abundance proteins.[84] Lu *et al*. used this approach to study the effects of microRNAs on proteome dynamics in cancer cells.[99] Krigsveld and coworkers used BONCAT to quantitatively analyze secreted proteins. Notably, BONCAT enabled them to analyze the secretome of different cell lines even in the presence of abundant serum proteins.[100,101] Acuto and coworker used quantitative non-canonical amino acid tagging (QuaNCAT) to analyze primary cells.[102] While all of these studies used Aha for BONCAT labeling, we anticipate that recent progress in cell-selective BONCAT will allow the combination of SILAC with cell- and tissue-targeted BONCAT for quantitative proteome and secretome analysis in *in vitro* and *in vivo* cell populations. BONCAT can also be combined with other quantitative proteomic approaches. For example, one particularly promising approach is dimethyl labeling, which is a peptide labeling strategy based on the reductive amination of the N-terminus and the ε-amino groups of lysines.[103] Dimethyl labeling is inexpensive, commercially available, and easy to use. Importantly, dimethyl labeling does not require the incorporation of istopically labeled amino acids. We anticipate that dimethyl labeling could be easily combined with BONCAT to provide an additional approach for cell-selective quantitative proteomics (Figure 1.4D).

Non-Canonical Amino Acids for Cell-Selective and Subcellular-Specific Proteomic Labeling

The original implementation of BONCAT utilized Aha, which allows for detection of newly synthesized proteins at specified time points. This protein labeling strategy has been used to study the kinetics of nucleosome turnover by metabolically labeling histones,[104,105] and to visualize newly synthesized proteins in neuronal cells,[106-108] zebrafish,[109] C. *elegans*,[110] Xenopus embryos,[111] and environmental microbes.[112] Restricting protein labeling through the controlled expression of mutant aaRS enzymes has been used to achieve cell-selective BONCAT. Cell-selective BONCAT has been used to restrict protein labeling to pathogenic microbes in the presence of host cells and to identify the virulence factors that bacterial pathogens secrete into mammalian cells. Because the host cells do not incorporate the non-canonical amino acid, only the pathogen proteins are labeled, permitting them to be tagged and identified even inside host cells.[113] Promoter-controlled expression of mutant aaRS enzymes is another strategy by which selective BONCAT can be achieved. For example, a mutant methionyl tRNA synthetase was placed under the control of a redox sensitive promoter in the SoxRS regulon. This strategy enabled redox-activated labeling of proteins in bacteria (Figure 1.5A).[114] There are also instances in which one would like to activate protein labeling as a function of multiple promoters for context-specific BONCAT. To solve this problem, a split mutant methionyl tRNA synthetase was developed that allows incorporation of the non-canonical amino acid Anl only when both fragments of the enzyme are present in cells. In this approach, labeling occurs only upon activation of two different promoters that drive the expression of the N- and C-terminal fragments of this bisected synthetase.[115] Using multiple promoters to control proteomic labeling expands the possibility of questions that BONCAT can address by improving the spatiotemporal resolution of the technique. Additional approaches towards subcellular-specific proteome labeling are being developed. For example, Schuman and coworkers used BONCAT to detect dendrite localized protein synthesis in neurons (Figure 1.5B).[106] Specialized methods to target protein labeling to specific organelles are also being developed. For example, Ting and coworkers recently developed a method called APEX using an engineer peroxidase to analyze proteins at the interface of the mitochondrial intermembrane.[116] Selective proteome labeling methods can potentially be complemented with the targeted delivery of chemical reporters to cells of interest.[117] Taken together, we anticipate that such approaches will increase the diversity of chemical toolsets available for the study of protein synthesis in specific cell types and subcellular compartments.

Future Directions

Technological advancements in DNA sequencers and mass spectrometers will continue to increase the coverage, throughput, and precision of measurements. In conjunction, the development of better chemical reporters and metabolic tags will enhance the specificity and spatiotemporal resolution of labeling and improve the detection of biological molecules. Advances in chemical reporters will continue to push the frontiers of protein detection by moving increasingly towards both *in vivo* applications and sub-cellular resolution. Elucidating the proteins that are synthesized in specific subcellular compartments and determining how they are modified will enhance our understanding of translation.[118] As improved instruments enable the analysis of more complex samples, we anticipate that chemical proteomic strategies such as BONCAT will be increasingly deployed for *in vivo* applications where temporal and context-dependent proteome profiling is required. Such strategies will also be utilized to analyze secreted proteins (i.e., the 'secretome' of various cell types). Progress towards combining quantitative proteomic platforms, such as multiple reaction monitoring methods (MRM),[119] with chemical reporter technologies will continue to enhance our understanding of proteome dynamics by providing quantitative measurements of proteomic changes in various cells. Current snapshots of cellular protein synthesis will evolve towards time-series and comprehensive datasets that provide a more complete picture of cellular proteomes over organisms' lifetimes.

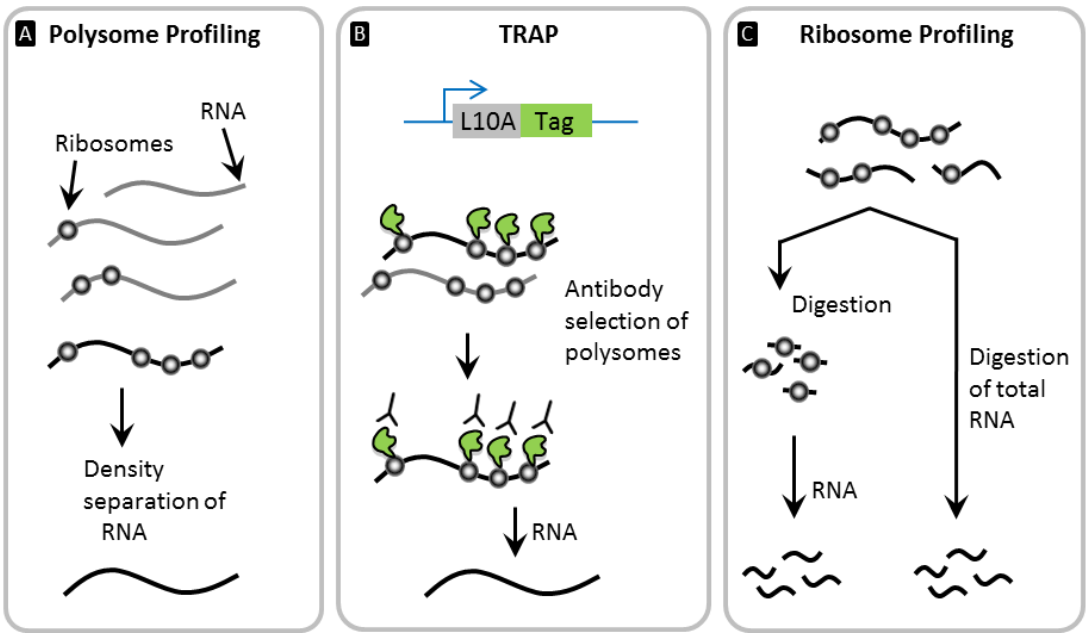


Figure 1.1. Methods for detecting and quantitating protein synthesis based on the translation state of cells and the sequencing of ribosome-bound mRNAs. A. Polysome profiling uses density separation of ribosome-bound mRNAs to identify mRNAs that are actively translated in cells. B. Translating Ribosome Affinity Purification (TRAP) uses epitope-tagged ribosomes to affinity enrich mRNAs that are being translated in cells. C. Ribosome profiling utilizes the nuclease digestion of ribosome-protected mRNAs to produce mRNA sequences that correspond to the precise position of ribosomes on the mRNA and therefore provide a snapshot of active translation in cells.

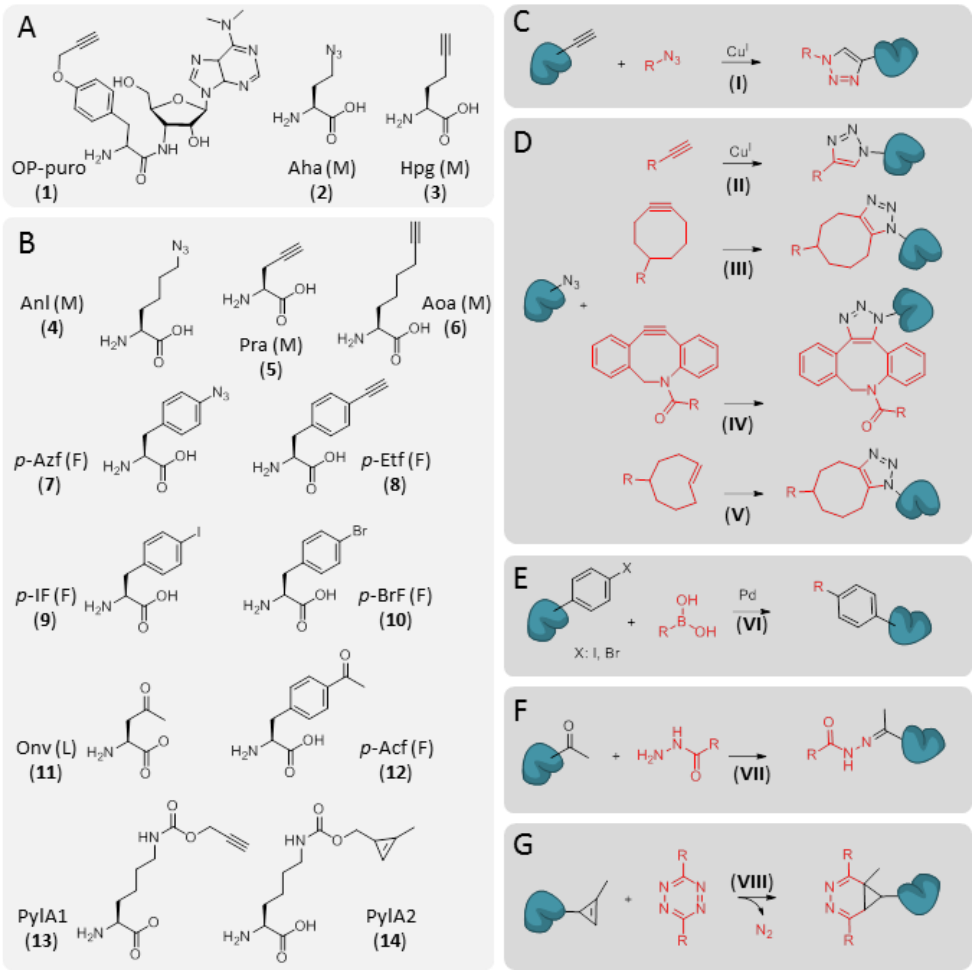


Figure 1.2. Chemical reporters that are incorporated into newly synthesized proteins. Chemically reactive functional groups and reactions discussed in this review. A. Chemical reporters that are incorporated into newly proteins based on puromycin addition to nascent polypeptide chains and incorporation of non-canonical amino acids. B. Non-canonical amino acids that are incorporated into newly synthesized proteins in a cell-selective manner. C. Copper-catalyzed azide-alkyne cycloaddition between alkyne functionalized proteins and azide tags. D. Azide functionalized proteins can be tagged through the copper-catalyzed azide alkyne cycloaddition reaction (II), or through strained cyclooctynes (II, III) or alkynyl probes (IV). E. Aryl-halide functionalized proteins can be tagged using palladium-catalyzed modified Suzuki-Miyaura cross-coupling reactions. F. Ketone-containing proteins can be tagged with hydrazide functionalized tags. G. Proteins labeled with chemical tags containing 1,3 disubstituted cyclopropene functional groups can be reacted with tetrazines using the inverse electron demand Diels-Alder reaction.

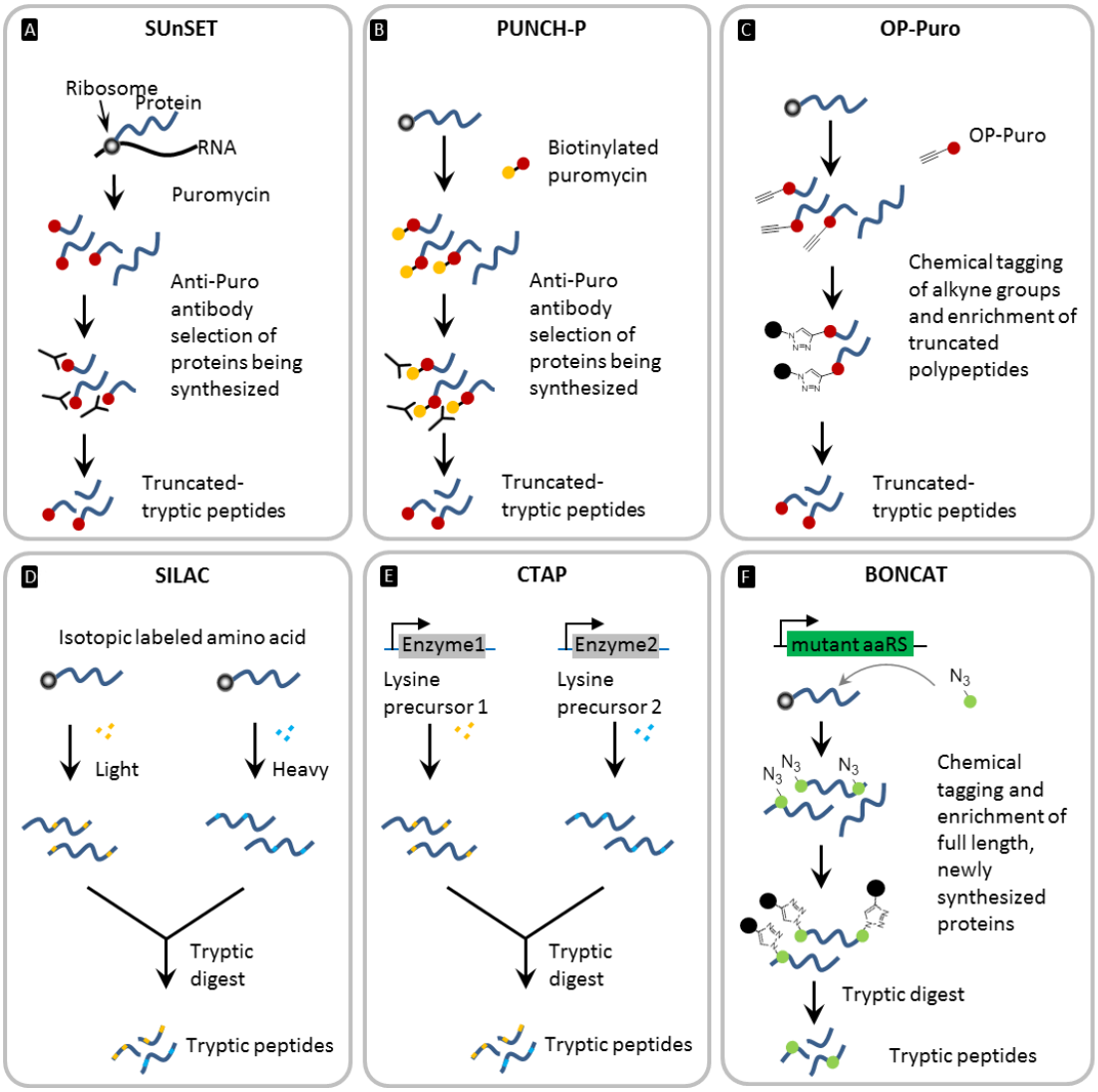


Figure 1.3. Strategies for labeling nascent polypeptides and identifying newly synthesized proteins. A. Surface sensing of translation (SunSET) utilizes the immunodetection of puromycin, which is incorporated into the C-terminus of nascent polypeptide chains. B. Puromycin-associated nascent chain proteomics (PUNCH-P) utilizes a biotin-conjugated puromycin tag for the affinity enrichment of nascent polypeptides. C. O-propargyl-puromycin (OP-puro) is an alkyne functionalized puromycin that enables the chemoselective tagging of puromycin in nascent polypeptide chains. D. Stable isotope labeled amino acids in cell culture (SILAC) uses the incorporation of isotope labeled amino acids for relative proteomic quantification of newly made proteins between samples. E. Cell-type specific labeling using amino acid precursors (CTAP) utilizes genetically encoded enzymes to uncage isotopically labeled lysine precursors that are incorporate into newly synthesized proteins. Only cells that express the enzymes can metabolically incorporate the lysine precursors. F. Bioorthogonal non-canonical amino acid

tagging method (BONCAT) uses incorporation of non-canonical amino acids for the bioorthogonal and chemoselective labeling, detection, and identification of newly synthesized proteins in cells. Cell-selective BONCAT utilizes genetically encoded enzymes for the incorporation of non-canonical amino acids and proteomic labeling.

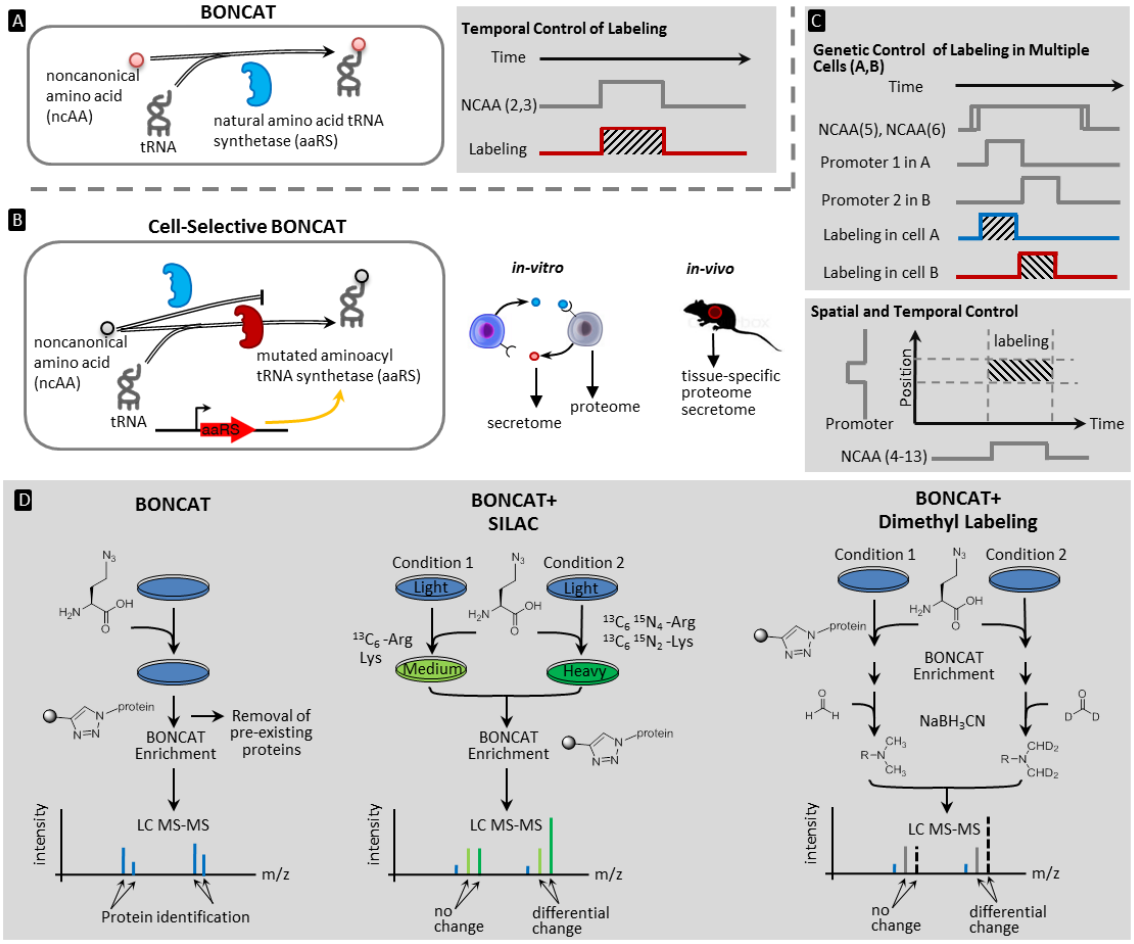


Figure 1.4. Approaches for labeling, identifying, and quantifying newly synthesized proteins by metabolically incorporating non-canonical amino acids. A. BONCAT utilizes the cell's endogenous translation machinery to incorporate non-canonical amino acids and label newly synthesized proteins. The addition of non-canonical amino acids at time-windows of interest enables temporal control over protein labeling. B. Cell-selective BONCAT utilizes a mutant aaRS enzyme to incorporate non-canonical amino acids. Only cells expressing the mutant aaRS incorporate the non-canonical amino acid; the proteome and secretome of these cells can be labeled and identified using mass spectrometry. C. Cell-selective BONCAT can be used to label distinct subsets of cells based on activation of different promoters that drive the expression of mutant aaRS enzymes. This approach provides spatial and temporal control of proteome labeling. D. Combining BONCAT with quantitative proteomic labeling methods such as SILAC or dimethyl labeling enables the identification and relative quantification of newly synthesized proteins between samples. In SILAC+BONCAT, SILAC labeling is performed at the same time as non-canonical amino acid labeling; BONCAT is used to chemically tag and enrich newly made

proteins. In the BONCAT+dimethyl labeling approach, BONCAT is used to affinity enrich newly made proteins; after tryptic digestion, tryptic peptides are labeled using dimethyl labeling for quantitative proteomics.

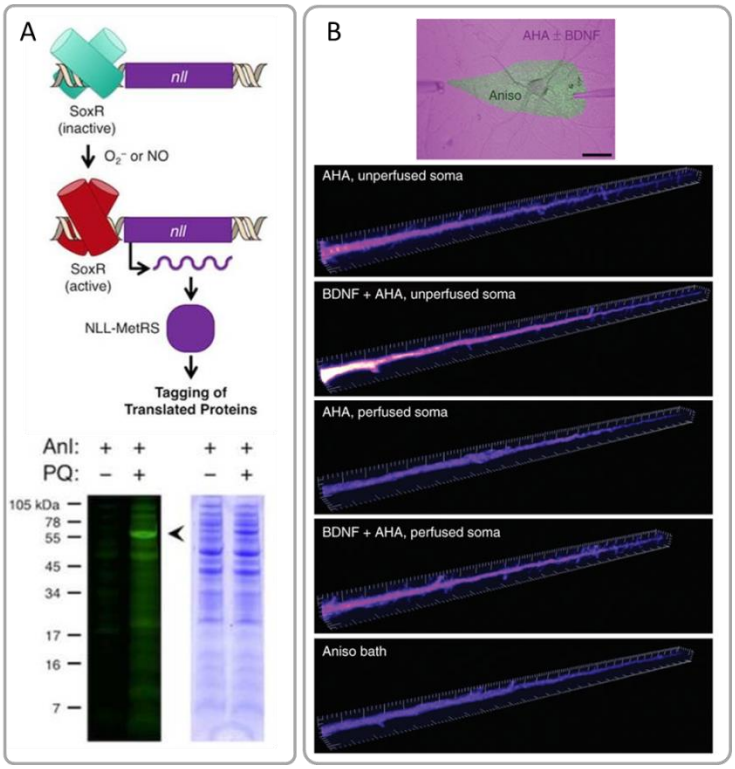


Figure 1.5. Non-canonical amino acids for the genetically encoded, cell-specific, and subcellular-localized labeling of newly synthesized proteins. A. State-selective BONCAT utilizes promoter-controlled expression of a mutant aaRS to activate the labeling of newly synthesized proteins with non-canonical amino acids. In this example, a SoxR promoter is activated in the presence of oxidizing conditions and turns on Anl incorporation in newly made proteins. PQ is the superoxide-generating agent paraquat. Proteome labeling was only detected in the presence of paraquat.[114] B. Subcellular labeling of newly synthesized proteins in neuronal dendrites. Somata of neurons were perfused with different combinations of the non-canonical amino acid Aha, protein synthesis inhibitor anisomycin, and brain-derived neurotrophic factor (BDNF), which induces the protein synthesis dependent enhancement of synaptic strength and dendritic translation. Fluorescence confocal microscopy images at the bottom show detection of newly synthesized proteins in neuronal dendrites using an alkyne-functionalized dye that selectively labels Aha residues.[106]

References

(1) Narayan, P.; Ehsani, S.; Lindquist, S. *Nat Chem Biol* **2014**, *10*, 911.

(2) Hetz, C.; Mollereau, B. *Nat. Rev. Neurosci.* **2014**, *15*, 233.

(3) Wang, W.; Nag, S.; Zhang, X.; Wang, M. H.; Wang, H.; Zhou, J.; Zhang, R. *Med. Res. Rev.* **2014**.

(4) Zhang, Y.; Nicholatos, J.; Dreier, J. R.; Ricoult, S. J.; Widenmaier, S. B.; Hotamisligil, G. S.; Kwiatkowski, D. J.; Manning, B. D. *Nature* **2014**, *513*, 440.

(5) Scheper, G. C.; van der Knaap, M. S.; Proud, C. G. *Nat. Rev. Genet.* **2007**, *8*, 711.

(6) Gay, L.; Miller, M. R.; Ventura, P. B.; Devasthali, V.; Vue, Z.; Thompson, H. L.; Temple, S.; Zong, H.; Cleary, M. D.; Stankunas, K.; Doe, C. Q. *Genes Dev.* **2013**, *27*, 98.

(7) Anderson, L.; Seilhamer, J. *Electrophoresis* **1997**, *18*, 533.

(8) Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. *Mol. Cell. Biol.* **1999**, *19*, 1720.

(9) Maier, T.; Guell, M.; Serrano, L. *FEBS Lett.* **2009**, *583*, 3966.

(10) Goyama, S.; Huang, G.; Kurokawa, M.; Mulloy, J. C. *Oncogene* **2014**, *0*.

(11) Hetz, C.; Chevet, E.; Harding, H. P. *Nat. Rev. Drug Discov.* **2013**, *12*, 703.

(12) Cox, J.; Mann, M. *Cell* **2007**, *130*, 395.

(13) Church, G. M.; Regis, E. *Regenesis : how synthetic biology will reinvent nature and ourselves*; Basic Books: New York, 2012.

(14) Kurzweil, R. *The age of spiritual machines : when computers exceed human intelligence*; Viking: New York, 1999.

(15) Balagopal, V.; Parker, R. *Curr. Opin. Cell Biol.* **2009**, *21*, 403.

(16) Kapeli, K.; Yeo, G. W. *Frontiers in neuroscience* **2012**, *6*, 144.

(17) Zong, Q.; Schummer, M.; Hood, L.; Morris, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10632.

(18) Morin, P., Jr.; Storey, K. B. *Int. J. Dev. Biol.* **2009**, *53*, 433.

(19) Long, T. A. *Curr. Opin. Plant Biol.* **2011**, *14*, 325.

(20) Picard, F.; Milhem, H.; Loubiere, P.; Laurent, B.; Cocaign-Bousquet, M.; Girbal, L. *BMC Genomics* **2012**, *13*, 528.

(21) Arribere, J. A.; Doudna, J. A.; Gilbert, W. V. *Mol. Cell* **2011**, *44*, 745.

(22) Sivan, G.; Kedersha, N.; Elroy-Stein, O. *Mol. Cell. Biol.* **2007**, *27*, 6639.

(23) Okaty, B. W.; Sugino, K.; Nelson, S. B. *J. Neurosci.* **2011**, *31*, 6939.

(24) Hupe, M.; Li, M. X.; Gertow Gillner, K.; Adams, R. H.; Stenman, J. M. *Nucleic acids research* **2014**, *42*, e14.

(25) Sarioglu, Y.; Ilhan, M.; Kayaalp, S. O. *Arch. Int. Physiol. Biochim.* **1988**, *96*, 35.

(26) Inada, T.; Winstall, E.; Tarun, S. Z., Jr.; Yates, J. R., 3rd; Schieltz, D.; Sachs, A. B. *RNA* **2002**, *8*, 948.

(27) Heiman, M.; Schaefer, A.; Gong, S.; Peterson, J. D.; Day, M.; Ramsey, K. E.; Suarez-Farinas, M.; Schwarz, C.; Stephan, D. A.; Surmeier, D. J.; Greengard, P.; Heintz, N. *Cell* **2008**, *135*, 738.

(28) Grange, P.; Bohland, J. W.; Okaty, B. W.; Sugino, K.; Bokil, H.; Nelson, S. B.; Ng, L.; Hawrylycz, M.; Mitra, P. P. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 5397.

(29) Liu, J.; Krautzberger, A. M.; Sui, S. H.; Hofmann, O. M.; Chen, Y.; Baetscher, M.; Grgic, I.; Kumar, S.; Humphreys, B. D.; Hide, W. A.; McMahon, A. P. *J. Clin. Invest.* **2014**, *124*, 1242.

(30) Zhou, P.; Zhang, Y.; Ma, Q.; Gu, F.; Day, D. S.; He, A.; Zhou, B.; Li, J.; Stevens, S. M.; Romo, D.; Pu, W. T. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 15395.

(31) Lam, P. Y.; Harvie, E. A.; Huttenlocher, A. *PLoS One* **2013**, *8*, e84436.

(32) Fang, Y.; Gupta, V.; Karra, R.; Holdway, J. E.; Kikuchi, K.; Poss, K. D. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 13416.

(33) Thomas, A.; Lee, P. J.; Dalton, J. E.; Nomie, K. J.; Stoica, L.; Costa-Mattioli, M.; Chang, P.; Nuzhdin, S.; Arbeitman, M. N.; Dierick, H. A. *PLoS One* **2012**, *7*, e40276.

(34) Wang, Y.; Jiao, Y. *Methods Mol Biol* **2014**, *1110*, 323.

(35) Jiao, Y.; Meyerowitz, E. M. *Mol. Syst. Biol.* **2010**, *6*, 419.

(36) Mustroph, A.; Zanetti, M. E.; Jang, C. J.; Holtan, H. E.; Repetti, P. P.; Galbraith, D. W.; Girke, T.; Bailey-Serres, J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18843.

(37) Sanz, E.; Yang, L.; Su, T.; Morris, D. R.; McKnight, G. S.; Amieux, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 13939.

(38) Sanz, E.; Evanoff, R.; Quintana, A.; Evans, E.; Miller, J. A.; Ko, C.; Amieux, P. S.; Griswold, M. D.; McKnight, G. S. *PLoS One* **2013**, *8*, e66179.

(39) Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R.; Weissman, J. S. *Science* **2009**, *324*, 218.

(40) Brar, G. A.; Yassour, M.; Friedman, N.; Regev, A.; Ingolia, N. T.; Weissman, J. S. *Science* **2012**, *335*, 552.

(41) Li, G. W.; Oh, E.; Weissman, J. S. *Nature* **2012**, *484*, 538.

(42) Ingolia, N. T.; Lareau, L. F.; Weissman, J. S. *Cell* **2011**, *147*, 789.

(43) Zoschke, R.; Watkins, K. P.; Barkan, A. *Plant Cell* **2013**, *25*, 2265.

(44) Liu, M. J.; Wu, S. H.; Wu, J. F.; Lin, W. D.; Wu, Y. C.; Tsai, T. Y.; Tsai, H. L.; Wu, S. H. *Plant Cell* **2013**, *25*, 3699.

(45) Dunn, J. G.; Foo, C. K.; Belletier, N. G.; Gavis, E. R.; Weissman, J. S. *eLife* **2013**, *2*, e01179.

(46) Jensen, B. C.; Ramasamy, G.; Vasconcelos, E. J.; Ingolia, N. T.; Myler, P. J.; Parsons, M. *BMC Genomics* **2014**, *15*, 911.

(47) Liu, X.; Jiang, H.; Gu, Z.; Roberts, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 11928.

(48) Rooijers, K.; Loayza-Puch, F.; Nijtmans, L. G.; Agami, R. *Nat Commun* **2013**, *4*, 2886.

(49) Williams, C. C.; Jan, C. H.; Weissman, J. S. *Science* **2014**, *346*, 748.

(50) Jan, C. H.; Williams, C. C.; Weissman, J. S. *Science* **2014**, *346*, 1257521.

(51) Ngo, J. T.; Tirrell, D. A. *Acc Chem Res* **2011**, *44*, 677.

(52) Tirrell, D. In *Hierarchical Macromolecular Structures: 60 Years after the Staudinger Nobel Prize I*; Percec, V., Ed.; Springer International Publishing: 2013; Vol. 261, p 199.

(53) McKay, C. S.; Finn, M. G. *Chem. Biol.* **2014**, *21*, 1075.

(54) Shieh, P.; Bertozzi, C. R. *Organic & biomolecular chemistry* **2014**, *12*, 9307.

(55) Chang, P. V.; Bertozzi, C. R. *Chem Commun (Camb)* **2012**, *48*, 8864.

(56) Jewett, J. C.; Bertozzi, C. R. *Chem Soc Rev* **2010**, *39*, 1272.

(57) Lang, K.; Chin, J. W. *ACS. Chem. Biol.* **2014**, *9*, 16.

(58) Eggers, D. K.; Welch, W. J.; Hansen, W. J. *Mol. Biol. Cell* **1997**, *8*, 1559.

(59) Hansen, W. J.; Lingappa, V. R.; Welch, W. J. *J Biol Chem* **1994**, *269*, 26610.

(60) Nathans, D. *Fed. Proc.* **1964**, *23*, 984.

(61) Starck, S. R.; Green, H. M.; Alberola-Ila, J.; Roberts, R. W. *Chem. Biol.* **2004**, *11*, 999.

(62) Schmidt, E. K.; Clavarino, G.; Ceppi, M.; Pierre, P. *Nat Methods* **2009**, *6*, 275.

(63) Aviner, R.; Geiger, T.; Elroy-Stein, O. *Genes Dev.* **2013**, *27*, 1834.

(64) Liu, J.; Xu, Y.; Stoleru, D.; Salic, A. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 413.

(65) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol Cell Proteomics* **2002**, *1*, 376.

(66) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat Biotechnol* **1999**, *17*, 994.

(67) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Mol Cell Proteomics* **2004**, *3*, 1154.

(68) Pratt, J. M.; Petty, J.; Riba-Garcia, I.; Robertson, D. H.; Gaskell, S. J.; Oliver, S. G.; Beynon, R. J. *Mol Cell Proteomics* **2002**, *1*, 579.

(69) Schwanhausser, B.; Gossen, M.; Dittmar, G.; Selbach, M. *Proteomics* **2009**, *9*, 205.

(70) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. *Mol Cell Proteomics* **2005**, *4*, 310.

(71) Schutz, W.; Hausmann, N.; Krug, K.; Hampp, R.; Macek, B. *Plant Cell* **2011**, *23*, 1701.

(72) Sury, M. D.; Chen, J. X.; Selbach, M. *Mol Cell Proteomics* **2010**, *9*, 2173.

(73) Fredens, J.; Engholm-Keller, K.; Giessing, A.; Pultz, D.; Larsen, M. R.; Hojrup, P.; Moller-Jensen, J.; Faergeman, N. J. *Nat Methods* **2011**, *8*, 845.

(74) Kruger, M.; Moser, M.; Ussar, S.; Thievessen, I.; Luber, C. A.; Forner, F.; Schmidt, S.; Zanivan, S.; Fassler, R.; Mann, M. *Cell* **2008**, *134*, 353.

(75) Zanivan, S.; Krueger, M.; Mann, M. *Methods Mol Biol* **2012**, *757*, 435.

(76) Rechavi, O.; Kalman, M.; Fang, Y.; Vernitsky, H.; Jacob-Hirsch, J.; Foster, L. J.; Kloog, Y.; Goldstein, I. *Nat Methods* **2010**, *7*, 923.

(77) Li, R. X.; Ding, Y. B.; Zhao, S. L.; Xiao, Y. Y.; Li, Q. R.; Xia, F. Y.; Sun, L.; Lin, X.; Wu, J. R.; Liao, K.; Zeng, R. *J. Proteome Res.* **2012**, *11*, 2851.

(78) Drissi, R.; Dubois, M. L.; Boisvert, F. M. *FEBS J.* **2013**, *280*, 5626.

(79) Deal, R. B.; Henikoff, S. *Dev. Cell* **2010**, *18*, 1030.

(80) Amin, N. M.; Greco, T. M.; Kuchenbrod, L. M.; Rigney, M. M.; Chung, M. I.; Wallingford, J. B.; Cristea, I. M.; Conlon, F. L. *Development* **2014**, *141*, 962.

(81) Gauthier, N. P.; Soufi, B.; Walkowicz, W. E.; Pedicord, V. A.; Mavrakis, K. J.; Macek, B.; Gin, D. Y.; Sander, C.; Miller, M. L. *Nat Methods* **2013**, *10*, 768.

(82) Tape, C. J.; Norrie, I. C.; Worboys, J. D.; Lim, L.; Lauffenburger, D. A.; Jorgensen, C. *Mol Cell Proteomics* **2014**, *13*, 1866.

(83) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482.

(84) Bagert, J. D.; Xie, Y. J.; Sweredoski, M. J.; Qi, Y.; Hess, S.; Schuman, E. M.; Tirrell, D. A. *Mol Cell Proteomics* **2014**, *13*, 1352.

(85) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 19.

(86) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Biol.* **2009**, *5*, 715.

(87) Grammel, M.; Zhang, M. M.; Hang, H. C. *Angew. Chem. Int. Ed.* **2010**, *49*, 5970.

(88) Truong, F.; Yoo, T. H.; Lampo, T. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2012**, *134*, 8551.

(89) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596.

(90) Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A. *Chembiochem* **2010**, *11*, 2092.

(91) Liang, Y.; Mackey, J. L.; Lopez, S. A.; Liu, F.; Houk, K. N. *J. Am. Chem. Soc.* **2012**, *134*, 17904.

(92) Beatty, K. E.; Szychowski, J.; Fisk, J. D.; Tirrell, D. A. *Chembiochem* **2011**, *12*, 2137.

(93) Yuet, K. P.; Tirrell, D. A. *Ann. Biomed. Eng.* **2014**, *42*, 299.

(94) Chalker, J. M.; Wood, C. S.; Davis, B. G. *J. Am. Chem. Soc.* **2009**, *131*, 16346.

(95) Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 5652.

(96) Elliott, T. S.; Townsley, F. M.; Bianco, A.; Ernst, R. J.; Sachdeva, A.; Elsasser, S. J.; Davis, L.; Lang, K.; Pisa, R.; Greiss, S.; Lilley, K. S.; Chin, J. W. *Nat Biotechnol* **2014**, *32*, 465.

(97) Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A. *J. Am. Chem. Soc.* **2012**, *134*, 18638.

(98) Cox, J.; Mann, M. *Nat Biotechnol* **2008**, *26*, 1367.

(99) Lu, Y. Y.; Sweredoski, M. J.; Huss, D.; Lansford, R.; Hess, S.; Tirrell, D. A. *ACS. Chem. Biol.* **2014**, *9*, 334.

(100) Eichelbaum, K.; Winter, M.; Berriel Diaz, M.; Herzig, S.; Krijgsveld, J. *Nat Biotechnol* **2012**, *30*, 984.

(101) Eichelbaum, K.; Krijgsveld, J. *Methods Mol Biol* **2014**, *1174*, 101.

(102) Howden, A. J.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nat Methods* **2013**, *10*, 343.

(103) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Anal Chem* **2003**, *75*, 6843.

(104) Deal, R. B.; Henikoff, J. G.; Henikoff, S. *Science* **2010**, *328*, 1161.

(105) Deal, R. B.; Henikoff, S. *Cell cycle* **2010**, *9*, 3389.

(106) Dieterich, D. C.; Hodas, J. J.; Gouzer, G.; Shadrin, I. Y.; Ngo, J. T.; Triller, A.; Tirrell, D. A.; Schuman, E. M. *Nat. Neurosci.* **2010**, *13*, 897.

(107) Hodas, J. J.; Nehring, A.; Hoche, N.; Sweredoski, M. J.; Pielot, R.; Hess, S.; Tirrell, D. A.; Dieterich, D. C.; Schuman, E. M. *Proteomics* **2012**, *12*, 2464.

(108) Zhang, G.; Bowling, H.; Hom, N.; Kirshenbaum, K.; Klann, E.; Chao, M. V.; Neubert, T. A. *J. Proteome Res.* **2014**.

(109) Hinz, F. I.; Dieterich, D. C.; Tirrell, D. A.; Schuman, E. M. *ACS chemical neuroscience* **2012**, *3*, 40.

(110) Ullrich, M.; Liang, V.; Chew, Y. L.; Banister, S.; Song, X.; Zaw, T.; Lam, H.; Berber, S.; Kassiou, M.; Nicholas, H. R.; Gotz, J. *Nat Protoc* **2014**, *9*, 2237.

(111) Shen, W.; Liu, H. H.; Schiapparelli, L.; McClatchy, D.; He, H. Y.; Yates, J. R., 3rd; Cline, H. T. *Cell reports* **2014**, *6*, 737.

(112) Hatzenpichler, R.; Scheller, S.; Tavormina, P. L.; Babin, B. M.; Tirrell, D. A.; Orphan, V. J. *Environ. Microbiol.* **2014**, *16*, 2568.

(113) Mahdavi, A.; Szychowski, J.; Ngo, J. T.; Sweredoski, M. J.; Graham, R. L.; Hess, S.; Schneewind, O.; Mazmanian, S. K.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 433.

(114) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. *ACS. Chem. Biol.* **2012**, *7*, 1326.

(115) Mahdavi, A.; Segall-Shapiro, T. H.; Kou, S.; Jindal, G. A.; Hoff, K. G.; Liu, S.; Chitsaz, M.; Ismagilov, R. F.; Silberg, J. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2013**, *135*, 2979.

(116) Hung, V.; Zou, P.; Rhee, H. W.; Udeshi, N. D.; Cracan, V.; Svinkina, T.; Carr, S. A.; Mootha, V. K.; Ting, A. Y. *Mol. Cell* **2014**, *55*, 332.

(117) Xie, R.; Dong, L.; Huang, R.; Hong, S.; Lei, R.; Chen, X. *Angew Chem Int Ed Engl* **2014**.

(118) Reynolds, N. M.; Ling, J.; Roy, H.; Banerjee, R.; Repasky, S. E.; Hamel, P.; Ibba, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4063.

(119) Picotti, P.; Aebersold, R. *Nat Methods* **2012**, *9*, 555.

CHAPTER II

Identification of Secreted Bacterial Proteins by Non-Canonical Amino Acid Tagging

Abstract

Pathogenic microbes have evolved complex secretion systems to deliver virulence factors into host cells. Identification of these factors is critical for understanding the infection process. We report a powerful and versatile approach to the selective labeling and identification of secreted pathogen proteins. Selective labeling of microbial proteins is accomplished via translational incorporation of azidonorleucine (Anl), a methionine surrogate that requires a mutant form of the methionyl-tRNA synthetase for activation. Secreted pathogen proteins containing Anl can be tagged by the azide-alkyne cycloaddition and enriched by affinity purification. Application of the method to analysis of the type III secretion system of the human pathogen *Yersinia enterocolitica* enabled efficient identification of secreted proteins, identified distinct secretion profiles for intracellular and extracellular bacteria, and allowed determination of the order of substrate injection into host cells. This approach should be widely useful for the identification of virulence factors in microbial pathogens and for development of potential new targets for antimicrobial therapy.

This chapter has been published as:

Mahdavi, A.; Szychowski, J.; Ngo, J. T.; Sweredoski, M. J.; Graham, R. L.; Hess, S.; Schneewind, O.; Mazmanian, S. K.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 433.

Introduction

Many bacterial pathogens use elaborate secretion systems to transfer effector proteins into target cells.[1] The injected proteins disrupt host-cell functions, including cytoskeletal assembly and cytokine production, to promote infection.[2] An important step in understanding virulence mechanisms is the identification of injected and secreted bacterial proteins. Traditional methods have included genetic screens and candidate protein approaches, which can be laborious and non-comprehensive. Proteome-wide labeling strategies offer the potential to rapidly identify secreted pathogen proteins without bias and with limited prior knowledge of host-pathogen interactions.[3] We have previously developed a method (designated BONCAT – bio-orthogonal non-canonical amino acid tagging) to incorporate azide functional groups into proteins as a general strategy for the enrichment of newly synthesized cellular proteins, making it possible to elucidate the spatial and temporal character of proteomic changes. [4,5] Our initial studies used the non-canonical amino acid azidohomoalanine (Aha) (structure **2**, Fig. 2.1*A*), a methionine (Met) surrogate, to label newly synthesized proteins. [4,5] The azide side chain of Aha allows newly synthesized proteins to be tagged with alkyne-functionalized affinity reagents and separated from pre-existing proteins by affinity chromatography. After separation, proteins are identified by tandem mass spectrometry (MS). Enrichment of newly synthesized proteins reduces the complexity of the sample and facilitates identification of the proteins of interest.

Recently we showed that introduction of a mutant form of the methionyl-tRNA synthetase (designated NLL-MetRS) into *Escherichia coli* enables incorporation of the non-canonical amino acid azidonorleucine (Anl) (structure **3**, Fig. 2.1*A*) into the bacterial proteome. [6] Because Anl is not activated to any significant extent by any of the wild-type synthetases, [6] labeling is restricted to cells in which NLL-MetRS is expressed. [7] This approach has prompted recent efforts to study proteomic changes in pathogens during infection.[8] At the same time, there has been considerable interest in protein labeling strategies to study secreted pathogen proteins, most notably through the use of stable-isotope labeling of amino acids in cell culture (SILAC).[9,10] However, isotopic labeling does not allow enrichment of secreted pathogen proteins, and enrichment is important for identification of virulence factors that would otherwise be undetected among abundant host proteins. Here we show that non-canonical amino acid labeling enables enrichment of secreted virulence factors and identification of injected proteins from host cell lysates using a shotgun, bottom-up proteomics approach.

Pathogenic bacteria secrete proteins through a variety of mechanisms. Secretion via type III, type IV and type VI systems occurs by direct injection of proteins into host cells, while type II and type V secretion systems use a two-step passage through the inner and outer membranes of the pathogen. Secreted outer membrane vesicles also mediate export of a complex array of proteins.[11] We focus here on the well-characterized type III secretion system (T3SS) of *Yersinia enterocolitica*, a Gram-negative bacterium. In *Yersinia,* the majority of secreted proteins, designated Yersinia Outer Proteins (Yops), are encoded on the 70 kb virulence plasmid pYV.[2,12] In addition to encoding Yops, the plasmid encodes machinery consisting of needle-shaped structures that assemble on the bacterial surface and inject proteins into the cytoplasm of host cells. The T3SS is activated by a temperature shift from 26˚C to the host temperature (37˚C); injection is initiated upon surface contact with target cells.[13,14] The pYV virulence plasmid also encodes a low calcium response (LCR) that enables secretion of T3SS substrates into the medium in the absence of host cells.[15] As a control for type III secretion, we used a *YscU* mutant strain (designated T3SS-Mut), which is unable to secrete Yops.[16] YSCU is an inner membrane protein required for T3SS assembly and recruitment of substrates.[17]

In the study described here, NLL-MetRS was introduced to both wild-type and mutant *Yersinia* strains to enable selective Anl-labeling of bacterial proteins (Fig. 2.1*B*). Because host cells do not express the NLL-MetRS, host-cell proteins are not labeled with Anl. After addition of Anl to the infection medium, Anl-labeled proteins were tagged by copper-catalyzed cycloaddition [18] (Fig 2.1*C*) with alkyne-functionalized dyes and detected by in-gel fluorescence or by confocal fluorescence imaging of infected host cells. Similarly, enrichment of Anl-labeled proteins was performed after attachment of a cleavable affinity tag (structure **4**, Fig. 2.1*A*) that permits binding of labeled proteins to immobilized streptavidin resin and removal of unlabeled proteins. The small mass modification resulting from tagging of Anl residues is readily detected by MS, thereby facilitating identification of enriched proteins (Figs. S2.1 and 2.S2.2). In a HeLa cell infection model, we identified the *Yersinia* proteins that were secreted into the medium and injected into HeLa cells. In addition to identifying known Yops, we identified novel secreted proteins that may play important roles in *Yersinia* infection. An extension of this approach allowed us to selectively label proteins secreted by *Yersinia* that had invaded HeLa cells, and to reveal secretion of distinct subsets of virulence factors. Pulse-labeling with Anl was used to investigate the order of injection of type III substrates into HeLa cells, and provided a simple method to determine the hierarchy of injection of virulence factors. The approach described here

is not limited to the study of T3SS substrates; it can be used to examine the many different secretion systems of microbial pathogens.

Results

Labeling of the *Yersinia* Proteome and T3SS Substrates

E. coli NLL-MetRS was constitutively expressed in *Y. enterocolitica* under control of its natural promoter to enable Anl incorporation into bacterial proteins (Fig. S2.3). Proteins secreted under LCR conditions were tagged with an alkyne-functionalized tetramethylrhodamine (TAMRA) dye (Fig. S2.4) and detected by in-gel fluorescence imaging (Fig. 2.2*A*). Labeling was observed only in samples treated with Anl (Fig. 2.2*A*, lane 3); nonspecific labeling in the absence of Anl was negligible (Fig. 2.2*A*, lane 1). Lack of TAMRA labeling in the absence of Anl was not due to absence of secreted proteins, as these proteins were detected by colloidal blue staining (Fig. 2.2*B*, lane 1). The chemoselectivity of the copper-catalyzed click reaction was confirmed by these results. As expected, the T3SS-Mut strain did not secrete any labeled proteins (Fig. 2.2*B*, lanes 2 and 4). The similarity of the protein secretion profiles in the Met- and Anl-treated samples (Fig. 2.2*B*, lanes 1 and 3), and the lack of secretion by the mutant strain (lanes 2 and 4) indicate that Anl incorporation does not interfere with type III secretion. Western blot analysis with antibodies specific for YopD and YopE confirmed Yop secretion by the T3SS-Wt strain (Fig. 2.2*C*). Analysis of *Yersinia* lysates showed proteome-wide incorporation of Anl into *Yersinia* proteins (Fig. 2.2*D*).

Detection of Injected Proteins in Host Cells by Fluorescence Imaging

We examined the injection of Yops into HeLa cells, a widely-used *in vitro* model for *Yersinia* pathogenesis.[17] Yop injection results in a characteristic rounded HeLa cell morphology that can be used to track injection (Fig. S2.5).[19,20] Infections were performed with 1 mM Anl; incubation of *Yersinia* at 37˚C prior to infection increased the efficiency of injection (Fig. S2.6). Digitonin was used to lyse HeLa cells selectively; it does not cause significant disruption of *Yersinia* membranes (Fig. S2.7).[19,21] Injected T3SS substrates were labeled with alkyne-TAMRA and detected by in-gel fluorescence imaging. We observed distinct bands of labeled proteins corresponding to molecular weights of known Yops D, E, H, M, N, P, Q and LcrV (Fig. 2.3*A*, lane 1). These bands were not observed in infections with the T3SS-Mut strain (although a low level of background labeling was observed (Fig. 2.3*A*, lane 2)), and proteins injected by *Yersinia* lacking the NLL-MetRS were not labeled (Fig. 2.3*A*, lane 3). Cell-specific proteome-wide

incorporation of Anl in *Yersinia* was confirmed in these co-cultures (Fig. 2.3*A*, lanes 4 and 5). Injected proteins were easily detected by this chemical labeling approach despite the much greater abundance of host cell proteins. Western blot analysis of the same samples with antibodies for Yops D, E and H showed the presence of these Yops in infected HeLa cells; antibody staining for *Yersinia* RpoA confirmed the absence of significant *Yersinia* cell lysis (Fig. 2.3*B*). When we transferred HeLa cell lysates treated with alkyne-TAMRA to nitrocellulose membranes and probed the membranes with antibodies for Yops D and E, we found that the protein bands detected by these antibodies were also labeled with the TAMRA dye (Fig. 2.3*C*). To visualize injection, we used fluorescence confocal microscopy to detect Anl-labeled T3SS substrates in the HeLa cell cytoplasm. Anl-labeled proteins were tagged with alkyne-AlexaFluor 488. We observed increased fluorescence in the cytoplasm of HeLa cells infected with the T3SS-Wt strain as compared to infections with T3SS-Mut *Yersinia* (Figs. 2.3*D* and 2.3*E*), although a low level of background labeling was observed in experiments with the mutant strain (Fig. 2.3*E*). As expected, infections with *Yersinia* that lacked the NLL-MetRS resulted in HeLa cell rounding, but no evidence of labeled proteins in the cytoplasm (Fig. S2.8). These results indicate that Anl-labeling can be used to detect injected virulence factors inside host cells.

Identification of Virulence Factors

As a first step toward identification of injected virulence factors, we performed a directed MS search for Anl-labeled Yops secreted under LCR conditions. We were able to detect incorporation of Anl at Met positions distributed throughout the secreted proteins (Figs. S2.9 – S2.11). We next sought to enrich and identify injected T3SS substrates from HeLa cells by using a shotgun MS approach. HeLa cells were infected with *Y. enterocolitica* in Anl-supplemented medium, and selectively lysed with digitonin after the infection. Cell lysates were treated with probe **4** for affinity enrichment of Anl-labeled proteins. Biotinylation was detectable by western blot analysis (Fig. S2.12), and labeled proteins were affinity-enriched on streptavidin resin (Fig. S2.13). In-gel tryptic digestion was performed on these samples and the resulting peptide mixtures were analyzed on a nanoLC-LTQ-FT mass spectrometer (Figs. S2.14 and S2.15). Type III-specific virulence factors were determined by comparison of lysates of HeLa cells infected by the T3SS-Wt and T3SS-Mut strains (Fig. 2.3*F*). This analysis identified previously reported T3SS-specific substrates, including Yops D, E, H, M, N, P and Q, and LcrV.[17,22] We did not find YopT or YopO, perhaps because these proteins are associated with host membranes. YopT was previously observed in the insoluble fraction of Hela cell lysates,[23,24] and YpkA (the YopO counterpart in *Yersinia pseudotuberculosis*) has been reported to associate with the plasma membrane after

injection.[25] It is also possible that YopT and YopO are made or secreted at low levels under our conditions.

In addition to identifying proteins injected into HeLa cells, we investigated proteins that were secreted into the medium during infection (Fig. S2.16). This was done by precipitating the proteins from the infection medium and enriching Anl-labeled proteins. Comparison of T3SS substrates that were injected into HeLa cells with those secreted into medium during infection (Fig. 2.3*G*) showed that YopB, YscP and YscH/YopR were found only in the medium – not inside infected HeLa cells. This result supports a previously proposed mechanism of action of these Yops, in which secretion of YscP is followed by secretion of YopR into the extracellular medium, resulting in injection of YopN and other effector Yops into host cells.[19]

Nineteen *Yersinia* proteins were found both in lysates prepared from HeLa cells infected with the secretion-competent strain and in those prepared from cells infected with the secretion-mutant strain (Fig. 2.3*F*). These proteins, which are not T3SS substrates, fall into two classes: i). highly abundant bacterial proteins, and ii). proteins associated with the bacterial cell surface. The highly abundant proteins (defined as those ranked among the top 10% of bacterial proteins in terms of abundance according to the PaxDb database)[26] are listed in italic type in Fig. 2.3*F*, and include Tuf1 (ranked #1 of 3163 proteins in terms of abundance), TufA (ranked #4), GapA (#5), Eno (#6), GroL (#7) and DnaK (#10). We have also listed LivK in this group, although it falls at 10.8% in terms of abundance, as well as MetG (MetRS), because it is over-expressed in both *Yersinia* strains. It seems likely that these highly abundant proteins are found in the HeLa cell lysate as a result of a low level of adventitious bacterial cell lysis, although we cannot rule out other mechanisms of transfer. Previous proteomic studies of factors secreted by *Yersinia* found subsets of these proteins, including HtpG, OmpA, GroL and several elongation factors [27], as well as DnaK and Eno.[28]

The second class of proteins found in experiments conducted with the secretion-mutant strain includes the membrane-associated proteins A1js30, Sif15, A1jpb8, OmpA and Ail. *Yersinia* surface proteins invasin, Ail and YadA mediate binding to host cells,[29] and it has been shown that Ail mediates attachment and uptake of bacterially secreted outer membrane vesicles.[11] Outer membrane protein A (OmpA), also identified in our analysis, is known to be present in such vesicles and released by Gram-negative bacteria. [11,30] OmpA has been observed in monocyte cell lysates after infection with *Y. pesti*, [27] is known to bind scavenger receptors,[31] and is considered a potent *Yersinia* virulence factor.[30] Sif15, also known as systemic factor protein-a (Sfpa), is

involved in systemic infection of *Y. enterocolitica*, is induced at 37°C, and is necessary for colonization of mesenteric lymph nodes in a mouse Peyer's patch infection model.[32] We also found the putative exported protein A1jpb8 and the outer membrane porin A1js30 in lysates prepared from HeLa cells infected with the secretion-mutant strain. Transfer of these proteins to the HeLa cell lysate could occur via a variety of mechanisms, including regulated release of outer membrane vesicles. Further work will be required to establish the mode of transfer of each protein.

Identification of Virulence Factors Secreted by Internalized Bacterial Cells

Many pathogens, including *Yersinia* strains, invade host cells during infection. Internalized pathogens may secrete virulence factors that are distinct from those released by extracellular bacteria.[33] We used the gentamicin protection assay with pulsed Anl labeling to compare the type III secretion profiles of internalized and extracellular *Yersinia* cells. Two parallel infections were initiated with the T3SS-Wt strain. After 1 hour to allow internalization of the pathogen by HeLa cells, we added gentamicin to one of the samples, thereby inhibiting protein synthesis in the extracellular bacteria in this sample. Thereafter Anl was introduced into both samples for identical labeling times of 3 hours. Confocal microscopy verified selective labeling of *Yersinia* proteins inside infected HeLa cells (Fig. 2.4*A*). In the absence of gentamicin, both extracellular and intracellular *Yersinia* were labeled (Fig. S2.17, S2.18).

Comparison of Anl-labeled proteins in HeLa cell lysates by in-gel fluorescence detection revealed a distinct pattern of proteins secreted by internalized *Yersinia* (Fig. 2.4*B*); internalized cells appear to secrete a subset of virulence factors. MS analysis, after enrichment of injected proteins, also indicated that a subset of Yops is secreted by internalized *Yersinia* (Fig. S2.19); Yops M, P and Q were not detected, while YopD, YopN, LcrV and effectors YopE and YopH were injected by the intracellular subpopulation.

Yops are Injected in a Temporally Distinct Manner

The BONCAT method is ideally suited to the study of time-dependent cellular phenomena. To analyze the order in which Yops are injected into host cells, we used an Anl pulse-labeling strategy wherein Anl was added to the medium at specified times following infection (Fig. 2.4*C*). MS analysis showed that injection of YopD, which is part of the type III needle complex that inserts into the host-cell membrane, is followed by injection of effector Yops E and H. Identification of YopD as the earliest injected substrate is supported by previous reports

indicating its injection is required to establish translocation of other Yops. [34] YopD, E and H have previously been detected on the surface of bacteria prior to contact with host cells, potentially allowing rapid injection of these substrates to stop phagocytosis.[35] YopE disrupts the host cell cytoskeleton and can interfere with phagocytosis, and Yops E and H are thought to control injection of effector Yops.[36] Our finding that Yops N and M are injected after YopE is supported by the fact that impassable YopE-DHFR fusion substrates can be used to block injection of Yops N and M.[37] YopP was first detected 60-90 minutes after initiation of infection, in agreement with previous findings that its cytotoxic effect is not detected until 60 minutes post-infection, and its inhibition of NF-κB signaling in dendritic cells is detected starting at 90 minutes after infection.[38] Detection of YscM and YopO may indicate that Anl pulse-labeling may be particularly good for identification of low-abundance and transiently injected proteins that would otherwise be undetected. Taken together, these results demonstrate the capacity of the BONCAT method to elucidate the hierarchy of secretion of virulence factors.

Discussion

Identification of effector proteins that are secreted or injected by pathogenic bacteria offers new opportunities for understanding pathogenesis mechanisms and developing novel therapeutics. Here we show that cell-selective, non-canonical amino acid tagging enables a new approach to the labeling, enrichment and identification of virulence factors secreted by pathogenic bacteria. Cell-selective proteomic labeling was achieved by outfitting *Yersinia* cells with the *E. coli* NLL-MetRS, which charges *Yersinia* $tRNA^{Met}$ with the azide-functionalized non-canonical amino acid Anl. Treatment of Anl-labeled proteins with alkyne affinity reagents provided a selective chemical tagging method, and enabled enrichment of secreted virulence factors from abundant host proteins. *Yersinia* proteins isolated from HeLa cell lysates included 8 T3SS substrates and 19 proteins that were transferred via type III-independent mechanisms. Because some (or perhaps all) of the latter proteins may have been released via adventitious bacterial cell lysis, the overall selectivity of the method might be enhanced by further improvements in the removal of bacterially shed proteins and better host cell lysis techniques. For live cell applications, cyclooctyne-functionalized reagents can be used to tag the azide side chain of Anl residues in a copper-free manner.[39,40] Our chemical tagging strategy is compatible with routine MS sample preparation methods such as GeLC-MS, FASP [41,42] and MudPIT,[43] and is easily combined with SILAC, iTRAQ and multiple-reaction monitoring quantitative MS methods.[44,45] This approach can be complemented with candidate protein methods, such as expression of tagged substrates, to verify the secretion and identify the location of newly found substrates inside host cells.

Pulsed Anl-labeling was combined with a gentamicin protection assay to identify proteins injected into HeLa cells by internalized *Yersinia*. These results demonstrate that the method can be used in different compartments of the host and should be applicable to studies of functional redundancy, wherein multiple effectors carry out similar functions.[46] Pulsed Anl labeling was used to study the hierarchy of Yop injection, enabling direct elucidation of the injection order of T3SS substrates. As an alternative to pulsed Anl-labeling, spatial and temporal resolution may be achieved in future studies by putting the NLL-MetRS under control of specific promoters that are stage-specific or spatiotemporally regulated.[47] This approach may allow for Anl labeling at different stages of infection, in particular host cells, or sub-cellular compartments. Straightforward extensions of the technology will enable investigation of both pathogen and host proteins during infections in animals. The simplicity of the approach makes it suitable for the study of numerous host-microbe interactions.

Materials and Methods

Expression of the NLL-MetRS in *Y. enterocolitica*

DH10B strains were used for genetic manipulations. The *E. coli* NLL-MetRS with mutations L13N, Y360L, and H301L, was isolated from pJTN1[7] by Nhe1 digestion and inserted into pQE80 (Qiagen). Kanamycin resistance was used for selection because *Y. enterocolitica* is resistant to ampicillin[48] and nalidixic acid. The resulting plasmid, which carries NLL-MetRS under control of the endogenous *E. coli* MetG promoter is termed pAM1 and has been deposited in Addgene. The plasmid was transformed into electrocompetent *Y. enterocolitica* and transformants were grown at 26°C on agar plates or in Luria-Bertani (LB) medium, both containing 50 µg/ml of kanamycin.

Synthesis of Anl by Copper-Catalyzed Diazo Transfer

The synthesis of Anl was based on a previously published protocol; the starting material was Boc-lysine.[49] Briefly, 5.27 g (81.1 mmol) of sodium azide was treated with 2.7 ml (16 mmol) of distilled triflic anhydride in 13 ml of water for 2 hours. The triflic azide product was extracted with 10 ml methylene chloride and added dropwise to a flask containing Boc-Lys-OH (2 g, 8.1 mmol), K_2CO_3 (1.68 g, 12.2 mmol) and $CuSO_4$ (20 mg, 0.08 mmol) in 26 ml of water and 250 ml of methanol. After 20 hours at room temperature the product was extracted with ethyl acetate, redissolved in methylene chloride and purified by silica gel chromatography. After Boc deprotection with hydrochloric acid, the final product was purified by cation exchange chromatography.

Copper-Catalyzed Click Reaction

The copper-catalyzed azide-alkyne coupling reaction, and the synthesis of the requisite THPTA ligand were performed as described previously.[18] Labeling was carried out at room temperature for 2 hours, at a final concentration of 0.2-0.5 mg/ml of HeLa lysate proteins, 0.1 mM copper sulfate, 0.5 mM THPTA ligand, 5 mM sodium ascorbate, 5 mM aminoguanidine and 100 µM alkyne probe. EDTA-free protease inhibitor (Roche) was added to all reactions and lysates.

Secretion of T3SS Substrates under LCR Conditions

Y. enterocolitica W2273 were diluted 1:50 from an overnight LB culture into M9 medium at 26 °C with agitation at 250 rpm. At OD_{600}=0.5, protein secretion was initiated by a temperature shift to 37°C. Labeling with Anl was performed in M9 medium lacking calcium and containing 1 mM Anl. After 2.5 hours, bacteria were sedimented for 15 minutes at 15000-rcf at 4°C. The medium was passed through a 0.2 µm filter, and proteins were precipitated with chloroform/methanol.

HeLa Cell Infection and Anl Labeling of T3SS Substrates

HeLa cells (ATCC) were routinely cultured in DMEM supplemented with 10% fetal bovine serum, trypsinized (Gibco) and expanded every 72 hours. Before infection, cells were washed twice with phosphate buffered saline (PBS) and resuspended in Opti-MEM medium (Gibco). *Y. enterocolitica* were diluted 1:25 from overnight cultures in LB and incubated at 26°C with agitation at 250 rpm until OD_{600}=0.5 was reached. *Yersinia* were pre-incubated at 37°C for 3 hours prior to start of infection. The pre-incubation time was determined by tracking T3SS injection of Anl labeled proteins (Fig. 2.S6). Labeling was performed at a multiplicity of infection of 100, with 10^7 HeLa cells per condition, at 1 mM Anl and 50 µg/ml kanamycin. Infections were carried out for 3.5 hours. Infected HeLa cells were lysed with digitonin as described below for analysis of injected *Yersinia* proteins in HeLa cells (Fig. 2.3*F*).

Selective Lysis of HeLa Cells after Infection

After infection, HeLa cells were washed five times with PBS to remove surface bound proteins. Cells were incubated with 0.1% (w/v) digitonin in PBS for 20 minutes at room temperature with 100 rpm agitation. EDTA-free protease inhibitor (Roche) was added to the lysis buffer. Lysis was increased by pipetting the lysis solution over the cells. Bacterial cells were removed from the lysates by centrifugation at 15000-rcf for 15 minutes at 4°C and filtration through a 0.2 µm filter. Western blot analysis with an antibody for RpoA was used to confirm absence of *Yersinia* lysis.

Enrichment of Anl-Labeled Proteins

Probe **4** was appended to Anl-labeled proteins by the copper-catalyzed azide-alkyne cycloaddition as described above. Proteins were precipitated with acetone, dissolved in 250 µl of 4% SDS in PBS, and diluted to 0.1% SDS by addition of PBS supplemented with protease inhibitor (Roche). Proteins were incubated with 400 µl Streptavidin Plus Ultralink resin (Pierce) for 1.5 hours at room temperature. Affinity purification was performed according to a previously published protocol [50]. Elution fractions were combined with Amicon Ultra 0.5 centrifuge filters (Millipore). Enrichment can also be performed with Click-iT alkyne-agarose resin (Invitrogen).

Detection of Proteins in Gels and Western Blots

Bicinchoninic acid protein quantification (Pierce) was used to equalize the amounts of proteins analyzed under different conditions. After dye labeling via the copper-catalyzed click reaction described above, proteins were washed with methanol to remove unreacted dye, and

electrophoresed on a Novex 12% Bis-Tris polyacrylamide gel (Invitrogen). Colloidal blue dye (Invitrogen) was used for nonspecific protein detection. Antibodies were used at the following dilutions: YopD: 1:20000, YopE: 1:40000, YopH: 1:4000, RpoA: 1:40000, α-tubulin (Abcam), anti-rabbit IgG-AlexaFluor 488 conjugate (Cell Signal Technologies, secondary antibody): 1:1000. Fluorescence imaging of western blots and gels was performed with a Typhoon 9400 molecular imager (GE Healthcare).

Fluorescence Confocal Microscopy

Adherent HeLa cells were infected as described above and fixed with 3.7% formaldehyde in PBS before labeling with 10 µg/ml Alexa Fluor 633-WGA conjugate (Invitrogen) in PBS for 30 minutes. Cells were permeabilized with ice-cold methanol for 3 minutes. Labeling with alkyne-TAMRA (Invitrogen) was performed as described above. Fluorescence confocal images were obtained on a Zeiss LSM 510 microscope.

Comparison of Yops injected by extracellular and internalized *Y. enterocolitica*

Y. enterocolitica (T3SS Wt) were diluted 1:25 from overnight cultures in LB and incubated at 26°C with agitation until OD_{600}=0.5 was reached. Two parallel infections of 5 x 10^7 HeLa cells each were initiated at MOI of 100 in OptiMeM. After 1 hour of infection, gentamicin was added at 80 µg/ml to one sample of infected HeLa cells; the other sample did not contain the antibiotic. After 1 hour, the medium was changed to OptiMeM without gentamicin and Anl was added to both samples at 1 mM. The +gent condition was supplemented with 4 µg/ml gentamicin to maintain inhibition of protein synthesis by extracellular bacteria. After 3 hours of labeling, HeLa cells in both samples were lysed with 0.1% digitonin for enrichment and MS analysis or fixed with 3.7% formaldehyde for fluorescence confocal microscopy as described above.

Determination of the order of Yop injection

T3SS Wt *Y. enterocolitica* were diluted 1:25 from overnight cultures in LB and incubated at 26°C with agitation until OD_{600}=0.5. Cells were pelleted at 5000-rcf and washed with PBS. Infection of four parallel samples, corresponding to the four time-windows of interest (Fig. 2.4*C*), was initiated as described above, with no pre-incubation at 37°C. Anl was added to the infection medium at 1 mM for the indicated times (Fig. 2.4*C*) and HeLa cells were lysed with digitonin at the end of each interval. HeLa cell lysates were treated with probe **4** as described above for enrichment and identification of injected proteins by MS.

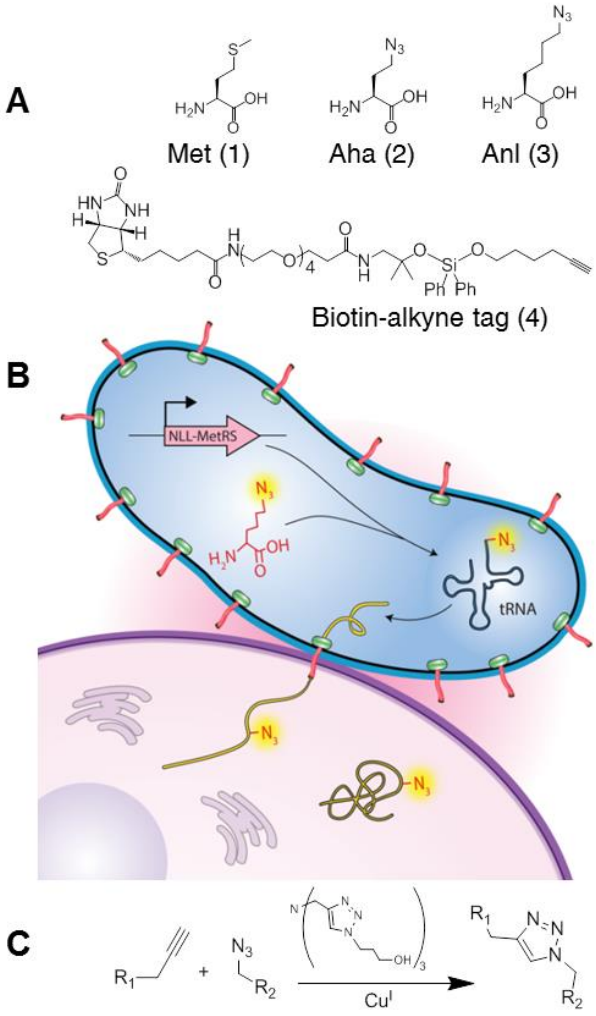


Figure 2.1. Incorporation of Anl into injected pathogen proteins and enrichment of labeled proteins. (*A*) Structures of amino acids Met (**1**) and Met analogues Aha (**2**) and Anl (**3**). Alkyne-functionalized biotin affinity probe (**4**) contains an acid-cleavable silane linker. The probe can be appended to the azide side chain of Anl. Cleavage with formic acid transfers a small mass tag to each modified Anl residue. (*B*) NLL-MetRS charges the $tRNA^{Met}$ with Anl; its expression in the pathogen allows cell-selective labeling of pathogen proteins during infection. Injected pathogen proteins are enriched after labeling with **4** and identified by tandem MS. (*C*) Copper-catalyzed azide-alkyne cycloaddition yields a triazole linkage.

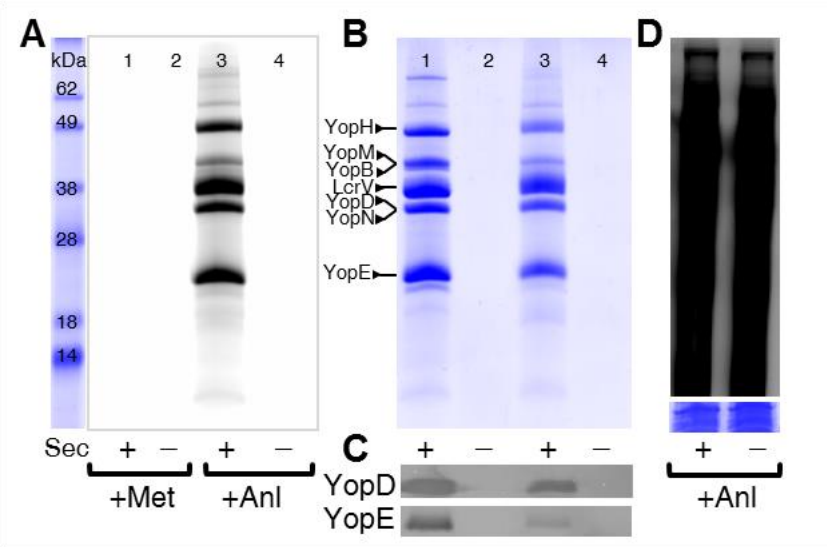


Figure 2.2. Labeling of secreted T3SS substrates. (*A*) Secretion competent (+) and secretion mutant (-) *Y. enterocolitica* strains harboring the NLL-MetRS were induced to secrete T3SS substrates under LCR conditions. Secreted proteins were labeled with alkyne-TAMRA and detected by in-gel fluorescence. (*B*) Detection of all secreted proteins in (*A*) by colloidal blue staining. (*C*) Western blot detection of YopD and YopE in samples from (*A*). (*D*) In-gel fluorescence detection of alkyne-TAMRA labeling of *Y. enterocolitica* lysates from conditions corresponding to lanes 3 and 4 in (*A*) shows proteome-wide incorporation of Anl; inset at bottom shows colloidal blue staining of the same samples.

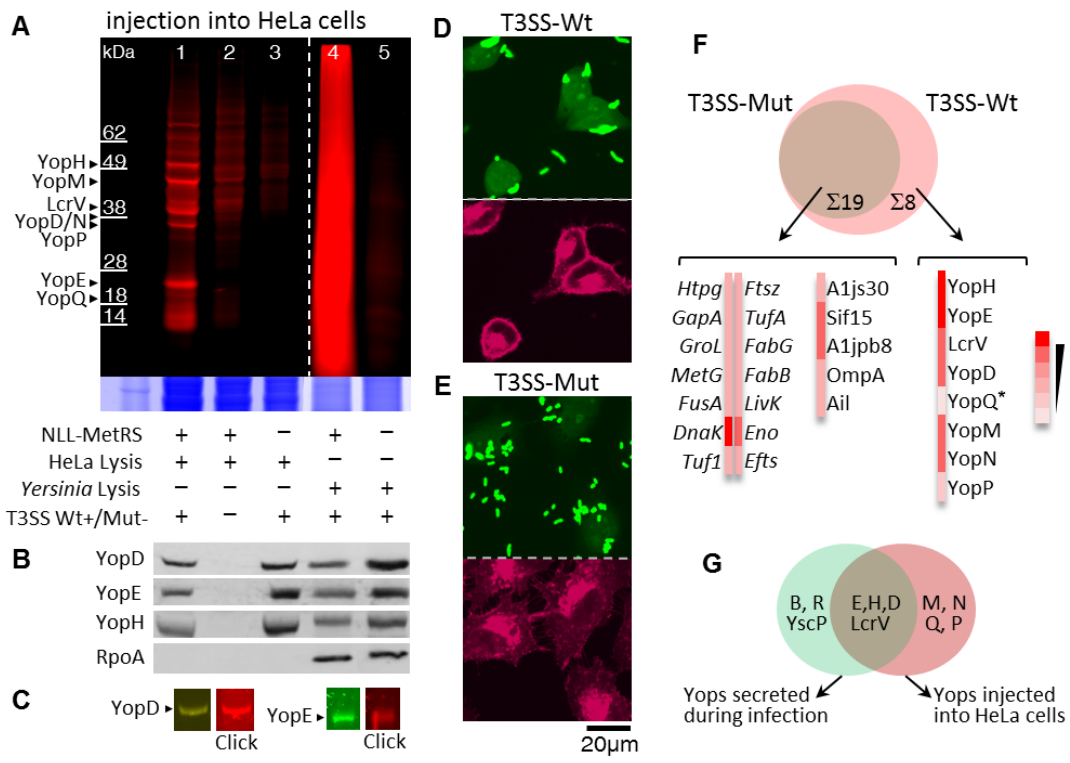


Figure 2.3. Detection and identification of injected *Y. enterocolitica* virulence factors in HeLa cells. (*A*) HeLa cells were infected with *Y. enterocolitica* in media containing 1 mM Anl. Proteins were labeled with alkyne-TAMRA and detected by in-gel fluorescence. Inset at bottom shows colloidal blue staining of the same gels. (*B*) Western blots of HeLa cell lysates with antibodies specific to YopD, YopE, YopH and bacterial RNA polymerase A (RpoA). Legend at the bottom of section *A* applies to section *B*. (*C*) Lysates of infected HeLa cells were treated with alkyne-TAMRA (click) and transferred to nitrocellulose membranes after SDS-PAGE. The same membranes were probed with antibodies specific to YopD and YopE. (*D*) Detection of injected *Yersinia* proteins in HeLa cells by fluorescence confocal microscopy. Anl-tagged proteins were labeled with alkyne-AlexaFluor 488 (green). An AlexaFluor 633-wheat germ agglutinin (WGA) conjugate was used to label membranes of HeLa cells (red). The same analysis was performed for infections with secretion mutant *Y. enterocolitica* in (*E*). (*F*) Shotgun proteomic identification of virulence factors injected into HeLa cells. Color code indicates the number of independent experiments in which each protein was detected. *YopQ was detected in only one infection. (*G*) Venn diagram showing Yops injected into HeLa cells or secreted into medium during infection.

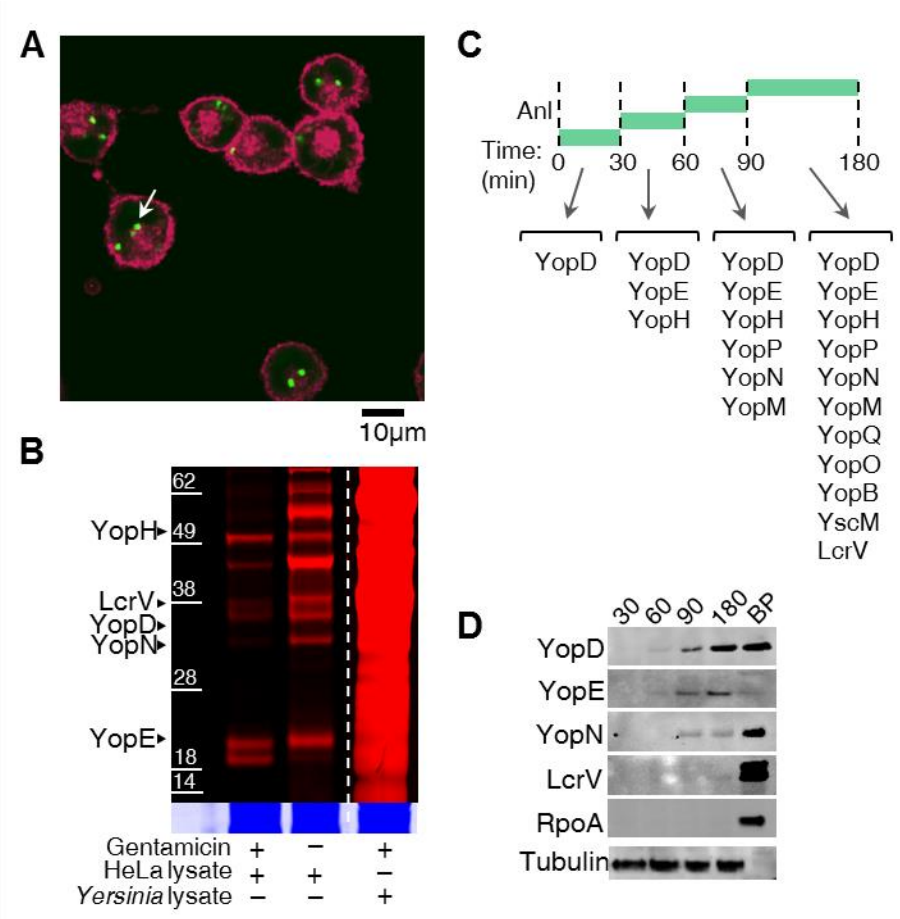
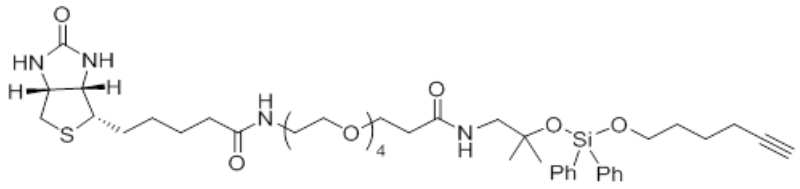


Figure 2.4. Labeling of proteins injected into HeLa cells by internalized *Y. enterocolitica* and identification of the order of Yop injection into HeLa cells (*A*) Confocal fluorescence microscopy shows Anl incorporation into the proteome of internalized *Y. enterocolitica.* HeLa cell membranes were labeled with AlexaFluor 633-WGA conjugate (red). Anl residues were labeled with alkyne-AlexaFluor 488 (green). Arrow indicates labeled *Y. enterocolitica* inside HeLa cell. (*B*) Infected HeLa cells were selectively lysed with digitonin and treated with alkyne-TAMRA to detect the proteins injected by internalized *Y. enterocolitica*. Inset at bottom shows colloidal blue staining of the same gel. In the presence of gentamicin only internalized *Yersinia* can inject proteins into HeLa cells. (*C*) The order of injection of Yops was determined by pulsed-Anl labeling and shotgun MS. Anl was added only during the indicated times for each infection and HeLa cells were lysed with digitonin at the end of each time interval. (D) Western blot detection of Yops in pulsed-Anl labeling experiments. RpoA serves as control for bacterial lysis; antibody for α-tubulin was used as loading control for HeLa lysates. BP: bacterial pellet.

Figure S2.1.

Structure of acid-cleavable biotin-alkyne tag.



Acid-cleavable biotin-alkyne affinity enrichment tag *N*-(2-(((hex-5-yn-1-yloxy)diphenylsilyl)oxy)-2-methylpropyl)-1-(5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-d]imidazol-4-yl)pentanamido)-3,6,9,12-tetraoxapentadecan-15-amide. Synthesis of the acid cleavable alkyne tag was based on our recent work on the analogous acid-sensitive dialkoxydiphenylsilane biotin-azide tag (*45*), with the exception that hex-5-yn-1-ol was used in place of 6-azidohexanol.^{1}H NMR (500 MHz, $CDCl_3$), δ (ppm): 7.72-7.50 (m, 4H), 7.40 (m, 6H), 7.06-6.95 (m, 1H), 6.66 (s, 1H), 6.50 (t, *J* = 5.65, 5.65 Hz, 1H), 5.80 (s, 1H), 4.53-4.37 (m, 1H), 4.34-4.19 (m, 1H), 3.73 (td, *J* = 19.58, 6.04, 6.04 Hz, 4H), 3.65-3.48 (m, 14H), 3.48-3.35 (m, 2H), 3.35-3.22 (m, 1H), 3.11 (dd, *J* = 11.73, 7.06 Hz, 1H), 2.85 (td, *J* = 16.69, 8.51, 8.51 Hz, 1H), 2.72 (t, *J* = 10.43, 10.43 Hz, 2H), 2.41 (t, *J* = 6.01, 6.01 Hz, 2H), 2.27-2.05 (m, 4H), 1.95 (dd, *J* = 3.09, 2.17 Hz, 1H), 1.81-1.52 (m, 8H), 1.50-1.33 (m, 2H), 1.23 (d, *J* = 22.65 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$), δ (ppm): 173.5, 171.2, 164.1, 134.8 (m, 6C), 130.2 (m, 2C), 127.9 (m, 4C), 84.3, 75.7, 70.4 (m, 8C), 68.6, 67.3, 62.6, 61.8, 60.2, 55.7, 50.4, 40.5, 39.1, 37.1, 35.9, 31.3, 28.3, 28.0, 27.6, 25.6, 24.8, 18.1; *m/z* calculated for $C_{43}H_{67}N_7O_9SSi$ $[M+H]^+$: 840.4163, MS found: 863.6 (M+Na). HRMS found: 863.4228.

Figure S2.1 continued.

NMR spectra of acid-cleavable biotin-alkyne tag (top, ^{1}H NMR spectrum; bottom, ^{13}C NMR spectrum)

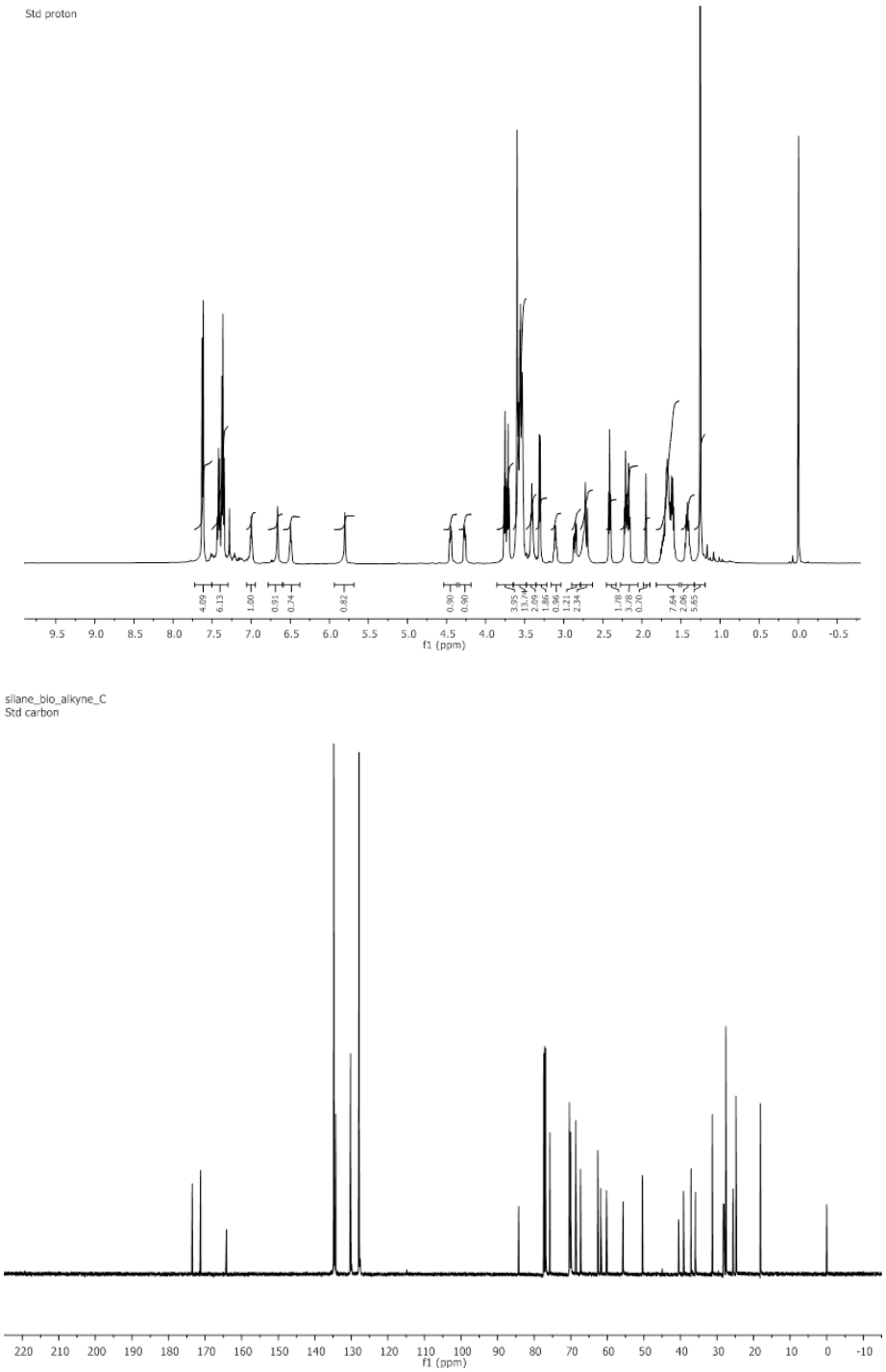
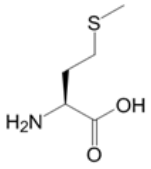


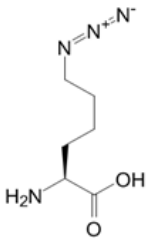
Figure S2.2.

Mass shifts at methionine positions as a result of Anl incorporation and addition of cleavable tag. Mass modifications associated with replacement of Met by Anl, with reaction of Anl with the biotin-alkyne affinity enrichment tag, and with cleavage of the tag by formic acid are represented. These three mass modifications are included as variable modifications for tandem mass spectrometry.

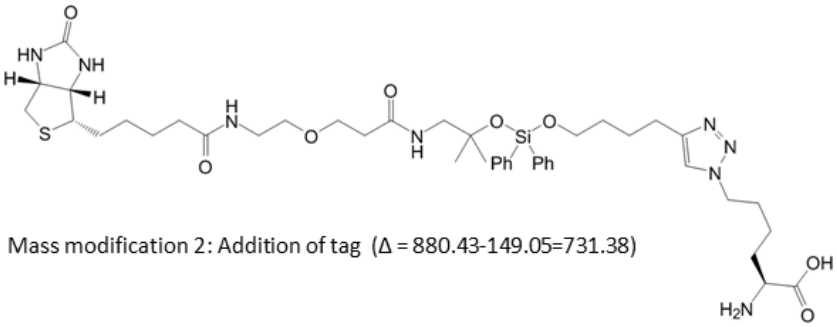
Met



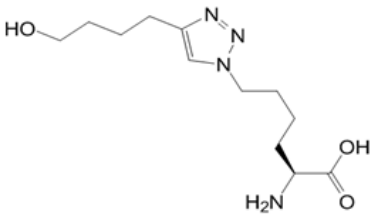
Anl



Mass modification1: Met to Anl (Δ=172.10-149.05=+23.05)



Mass modification 2: Addition of tag (Δ = 880.43-149.05=731.38)



Mass modification 3: Cleaved tagged Anl (Δ = 270.17-149.05=121.12)

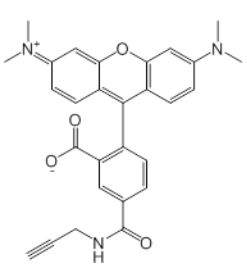
Figure S2.3.

Sequence of the NLL-MetRS expression construct. DNA sequence of the NheI-flanked expression cassette with NLL-MetRS under control of the *E. coli* MetG promoter and followed by its natural *E. coli* transcriptional termination signal. The MetRS sequence is underlined. Mutations L13N, Y260L, H301L are highlighted in green.

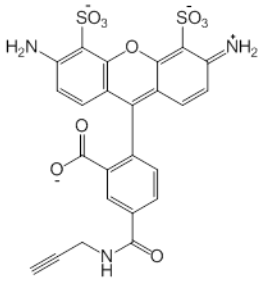
GCTAGCTCTAGAGACGTCCGGCCGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAGTCACTTACTTAACATTTTCCCATTTGGTACT
ATCTAACCCCTTTTCACTATTAAGAAGTAATGCCTACTATGACTCAAGTCGCGAAGAAAATTCTGGTGACGTGCGCAAACCCGTAC
GCTAACGGCTCAATCCACCTCGGCCATATGCTGGAGCACATCCAGGCTGATGTCTGGGTCCGTTACCAGCGAATGCGCGGCCACG
AGGTCAACTTCATCTGCGCCGACGATGCCCACGGTACACCGATCATGCTGAAAGCTCAGCAGCTTGGTATCACCCCGGAGCAGAT
GATTGGCGAAATGAGTCAGGAGCATCAGACTGATTTCGCAGGCTTTAACATCAGCTATGACAACTATCACTCGACGCACAGCGAA
GAGAACCGCCAGTTGTCAGAACTTATCTACTCTCGCCTGAAAGAAAACGGTTTTATTAAAAACCGCACCATCTCTCAGCTGTACGA
TCCGGAAAAAGGCATGTTCCTGCCGGACCGTTTTGTGAAAGGCACCTGCCCGAAATGTAAATCCCCGGATCAATACGGCGATAAC
TGCGAAGTCTGCGGCGCGACCTACAGCCCGACTGAACTGATCGAGCCGAAATCGGTGGTTTCTGGCGCTACGCCGGTAATGCGT
GATTCTGAACACTTCTTCTTTGATCTGCCCTCTTTCAGCGAAATGTTGCAGGCATGGACCCGCAGCGGTGCGTTGCAGGAGCAGGT
GGCAAATAAAATGCAGGAGTGGTTTGAATCTGGCCTGCAACAGTGGGATATCTCCCGCGACGCCCCTTACTTCGGTTTTGAAATTC
CGAACGCGCCGGGCAAATATTTCTACGTCTGGCTGGACGCACCGATTGGCCTGATGGGTTCTTTCAAGAATCTGTGCGACAAGCG
CGGCGACAGCGTAAGCTTCGATGAATACTGGAAGAAAGACTCCACCGCCGAGCTGTACCACTTCATCGGTAAAGATATTGTTTAC
TTCCTGAGCCTGTTCTGGCCTGCCATGCTGGAAGGCAGCAACTTCCGCAAGCCGTCCAACCTGTTTGTTCATGGCTATGTGACGGT
GAACGGCGCAAAGATGTCCAAGTCTCGCGGCACCTTTATTAAAGCCAGCACCTGGCTGAATCATTTTGACGCAGACAGCCTGCGT
TACTACTACACTGCGAAACTCTCTTCGCGCATTGATGATATCGATCTCAACCTGGAAGATTTCGTTCAGCGTGTGAATGCCGATATC
GTTAACAAAGTGGTTAACCTGGCCTCCCGTAATGCGGGCTTTATCAACAAGCGTTTTGACGGCGTGCTGGCAAGCGAACTGGCTG
ACCCGCAGTTGTACAAAACCTTCACTGATGCCGCTGAAGTGATTGGTGAAGCGTGGGAAAGCCGTGAATTTGGTAAAGCCGTGCG
CGAAATCATGGCGCTGGCTGATCTGGCTAACCGCTATGTCGATGAACAGGCTCCGTGGGTGGTGGCGAAACAGGAAGGCCGCGA
TGCCGACCTGCAGGCAATTTGCTCAATGGGCATCAACCTGTTCCGCGTGCTGATGACTTACCTGAAGCCGGTACTGCCGAAACTGA
CCGAGCGTGCAGAAGCATTCCTCAATACGGAACTGACCTGGGATGGTATCCAGCAACCGCTGCTGGGCCACAAAGTGAATCCGTT
CAAGGCGCTGTATAACCGCATCGATATGAGGCAGGTTGAAGCACTGGTGGAAGCCTCTAAATGAGAAGTAAAAGCCGCTGCCGC
GCCGGTAACTGGCCCGCTGGCAGATGATCCGATTCAGGAAACCATCACCTTTGACGACTTCGCTAAAGTTGACCTGCGCGTGGCG
CTGATTGAAAACGCAGAGTTTGTTGAAGGTTCTGACAAACTGCTGCGCCTGACGCTGGATCTCGGCGGTGAAAAACGCAATGTCT
TCTCCGGTATTCGTTCTGCTTACCCGGATCCGCAGGCACTGATTGGTCGTCACACCATTATGGTGGCTAACCTGGCACCACGTAAA
ATGCGCTTCGGTATCTCTGAAGGCATGGTGATGGCTGCCGGTCCTGGCGGGAAAGATATTTTCCTGCTAAGCCCGGATGCCGGTG
CTAAACCGGGTCATCAGGTGAAATAATCCCCCTTCAAGGCGCTGCATCGACAGCCTTTTGCTTTATAAATTCCTAAAGTTGTTTTCT
TGCGATTTTGTCTCTCTCTAACCCGCATAAATACTGGTAGCATCTGCATTCAACTGGATAAAATTACAGGGATGCAGAATGAGACA
CTTTATCTATCAGGACGAAAAATCACATAAATTCAGGGCAGTTGAGCAACAGGGAAACGAGTTGCATATCAGTTGGGGAAAAGTT
GGCACCAAAGGCAAAGCCAGATAAAAAGTTTTTCAGATGCTGCGGCAGCGGCAAAAGCGGAGCCCGACCTCGAGGGGGGGCCC
GGTACCCGGCCGGACGTCTCTAGAGCTAGC

Figure S2.4.

Alkyne functionalized fluorescent dyes for visualization of Anl labeled proteins. Alkyne functionalized tetramethylrhodamine (TAMRA) or Alexa Fluor 488 dyes are appended to Anl residues by copper catalyzed click reaction.



Alkyne-TAMRA



Alkyne-AlexaFluor 488

Figure S2.5.

HeLa cell morphology after labeling and infection by various *Y. enterocolitica* strains. HeLa cells were incubated with *Y. enterocolitica* at multiplicity of infection of 100 in the presence of 1 mM Met or 1 mM Anl. Infections with secretion wild type *Yersinia* resulted in a rounded morphology, whereas infections with secretion mutant *Yersinia* did not change HeLa cell morphology. Magnified versions are shown in the insets to the right of each image.

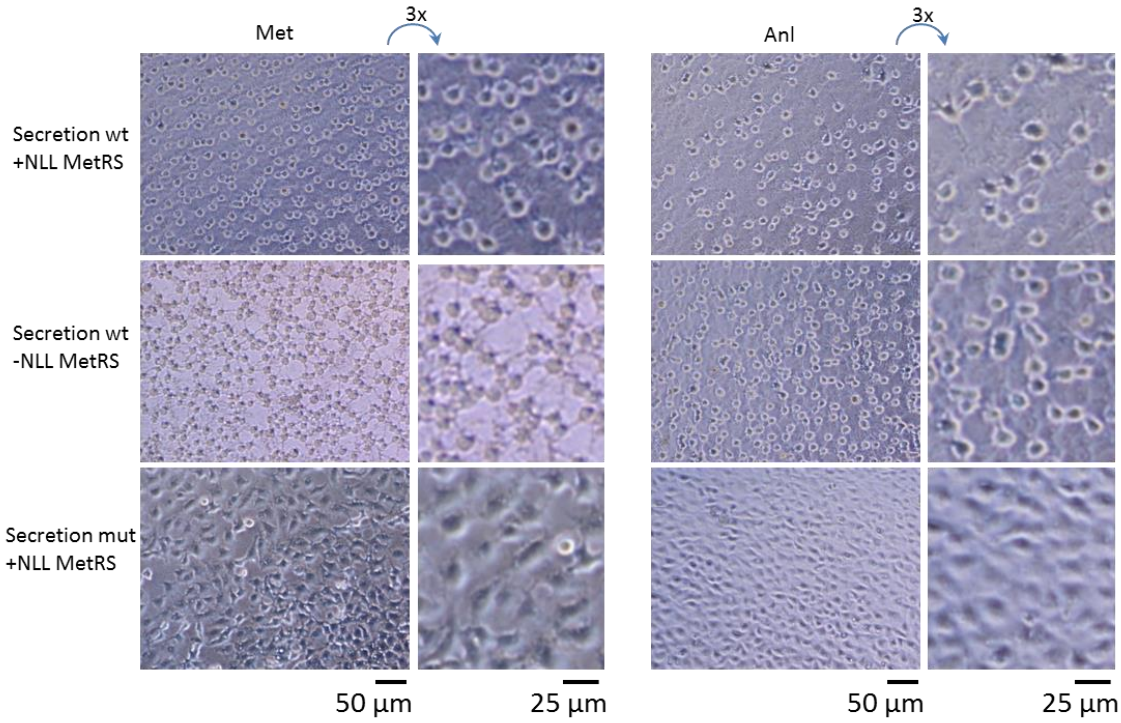


Figure S2.6.

Identifying infection conditions to increase efficiency of Yop injection into HeLa cells. Optimal conditions for Anl labeling of injected Yops were identified by investigating Anl concentrations of 0.5, 1 and 2 mM, as well as the timing of T3SS activation. *A.* Anl incorporation was detected by labeling with TAMRA alkyne dye and in-gel fluorescence detection. Activation of T3SS machinery was induced by pre-incubation of *Yersinia* at 37 °C for various times (0-60 min) before the start of infection. The labeling control (HeLa only) did not include any *Yersinia* cells. A pre-incubation step at 37 °C for 120 min was included to determine if longer pre-incubation would result in more labeling of injected proteins. Higher Anl concentrations consistently yielded higher levels of labeling with TAMRA dye, and longer pre-incubation at 37 °C resulted in more efficient injection. Secretion of Yops under LCR conditions (M) is included as a basis of comparison for the expected position of Yops *B.* Colloidal blue staining of the same gels shows equal protein loading across different conditions. Pre-incubation of *Yersinia* at 37 °C increases the efficiency of injection of labeled T3SS substrates. *C.* T3SS secretion competent *Y. enterocolitica* were pre-incubated at 37 °C for 1-5 hours before start of infection of HeLa cells. Infections were carried out in media containing 0.5, 1, 2 or 4 mM Anl. *D.* Colloidal blue dye staining of the same samples shows equal loading of HeLa lysates. H: HeLa lysates after infection with secretion WT *Yersinia*. M: Media secreted Yops under LCR conditions. Positions corresponding to different Yops are indicated to the right of the image.

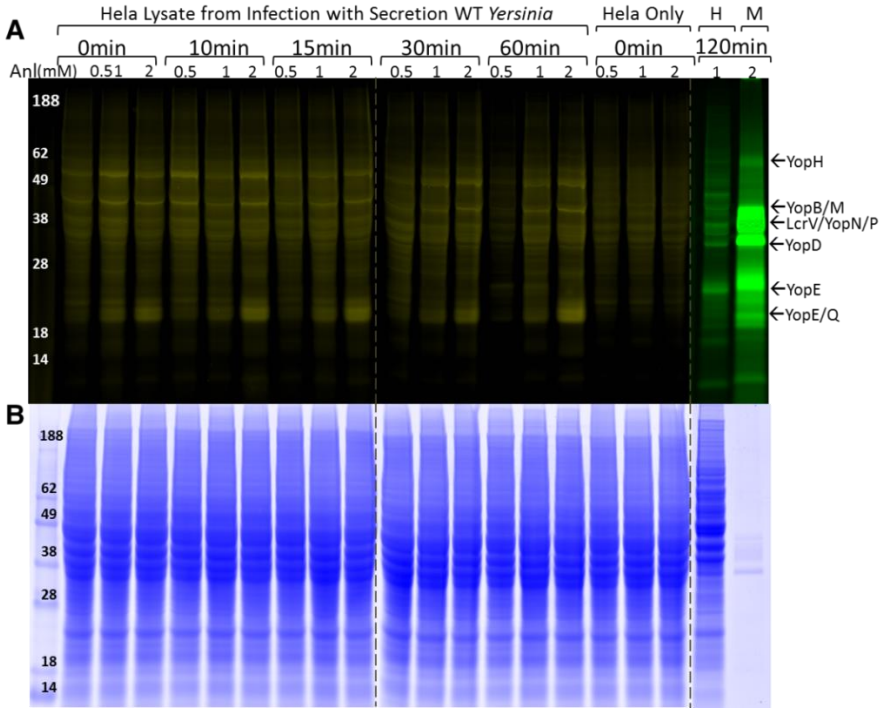


Figure S2.6 continued.

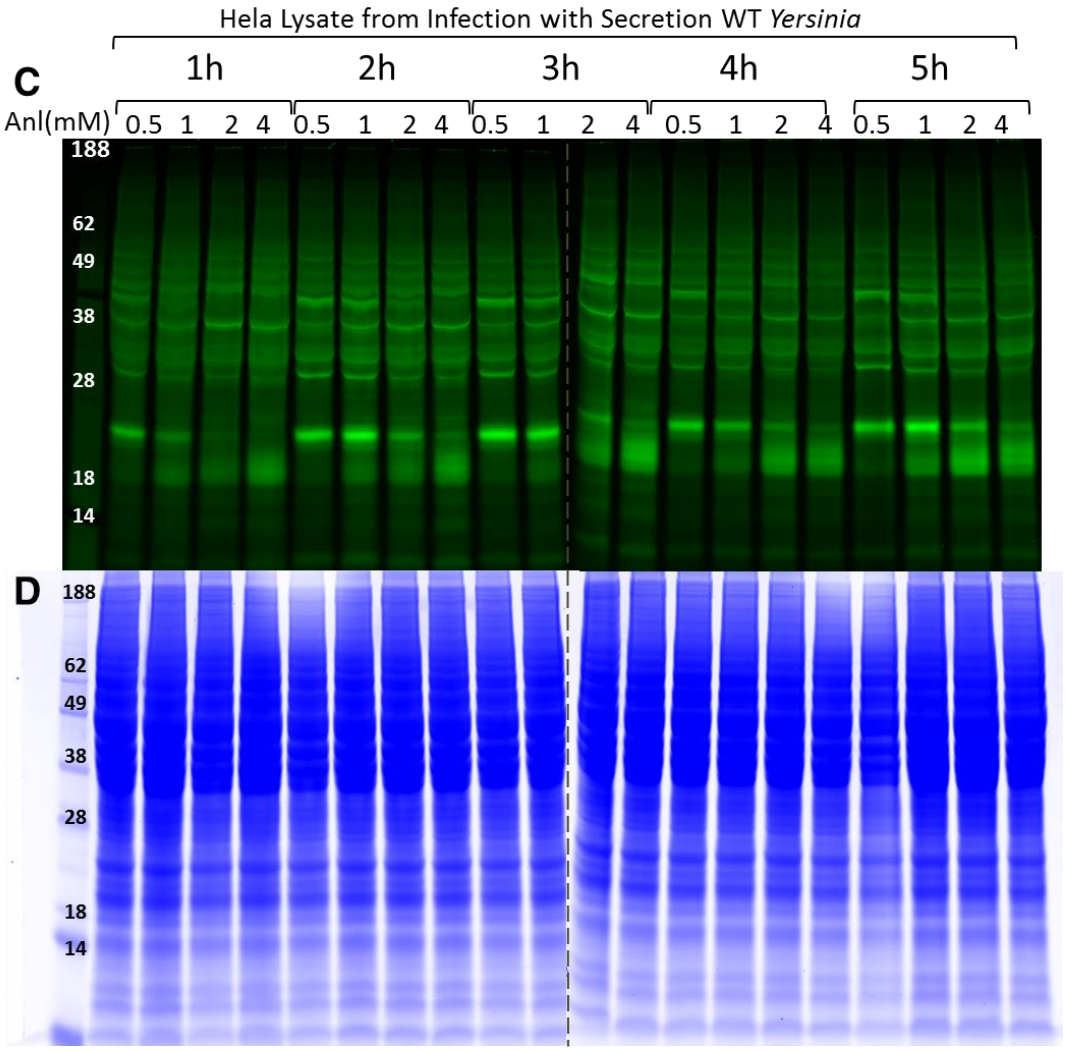


Figure S2.7.

Selective lysis of HeLa cells in the presence of intact *Y. enterocolitica.* Digitonin was used as a detergent to disrupt the HeLa cell membrane but not the bacterial envelope. Co-cultures of HeLa and *Yersinia* cells were lysed with various concentrations of digitonin in PBS. Briefly, 6-well plates containing 1.5 x 10^6 HeLa cells and *Yersinia* at MOI of 100 were lysed for 20 min at room temperature. After lysis of HeLa cells, *Yersinia* cells were removed by centrifugation at 15000 rcf and the remaining supernatant proteins were precipitated with chloroform/methanol. The proteins were resuspended in 1% SDS and electrophoresed on a 12% polyacrylamide gel and stained with colloidal blue. Efficient lysis of HeLa cells was observed with digitonin concentrations as low as 0.025% v/w. In the absence of HeLa cells and presence of *Yersinia*, no protein could be observed after incubation with digitonin, indicating lack of *Yersinia* lysis. Western blot analysis with antibody specific for *Yersinia* RpoA confirmed the lack of bacterial lysis. A small amount of non-specific binding was observed with this antibody under all conditions, including those with no bacteria. *Yersinia* lysis with 2% SDS served as a positive control for RpoA labeling. Western blot analysis with an antibody specific to α-tubulin was used as a loading control to show the extent of HeLa cell lysis. Western blots probed with α-tubulin antibody confirmed that 0.05% digitonin is sufficient for HeLa cell lysis. Digitonin concentrations are given as weight per volume of PBS.

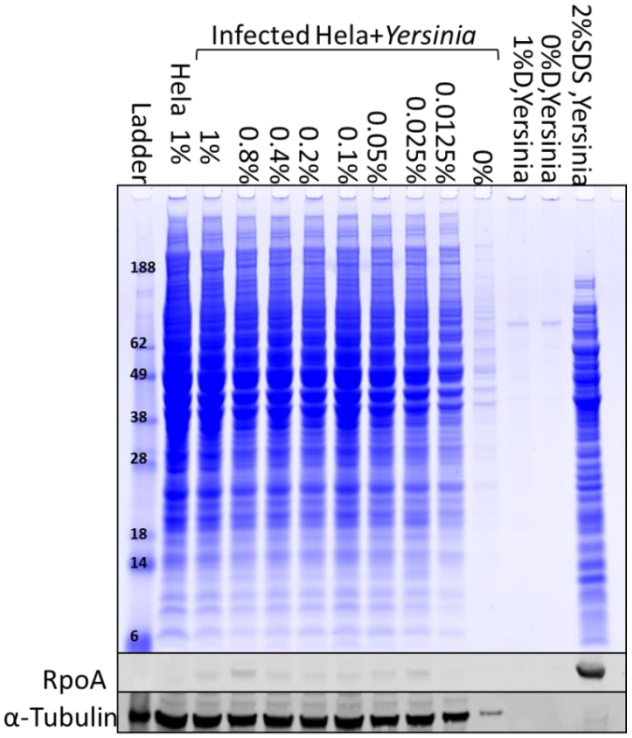
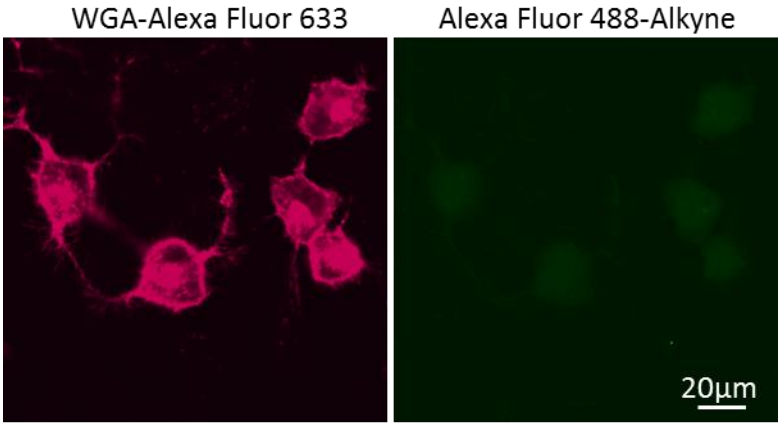


Figure S2.8.

Control conditions for detection of labeled proteins in absence of Anl incorporation. The images in this section are labeling controls for the infection studies shown in Fig. 2.3*D*, *E* of the main text. The images were obtained from infections in 1 mM Anl with secretion wild type *Y. enterocolitica* lacking the NLL-MetRS. Anl residues were labeled with alkyne-functionalized AlexaFluor 488; the resulting fluorescence is shown in green. AlexaFluor 633 conjugated to wheat germ agglutinin (WGA) labels membranes of HeLa cells; the corresponding fluorescence is shown in red. In the absence of NLL-MetRS, Anl is not incorporated into *Yersinia* proteins; therefore there is no labeling with alkyne-functionalized AlexaFluor 488.



The wild-type MetRS does not activate Anl to any appreciable extent; however the NLL-MetRS does activate Met. We have previously measured the kinetics of amino acid activation by NLL-MetRS; k_{cat}/K_m is 350±70 $M^{-1}s^{-1}$ for Met and 410±80 $M^{-1}s^{-1}$ for Anl. See Ref. 6 of the main text for measurement of the catalytic efficiency of MetRS variants. As shown in Figure 2.3*A* lane 5 of the main text, labeling in absence of NLL-MetRS is negligible. This is also evident in the images above as bacteria lacking the NLL-MetRS are not labeled, even in the presence of Anl. The Optimem medium contains about 100 µM Met; infections were performed in 1 mM Anl,

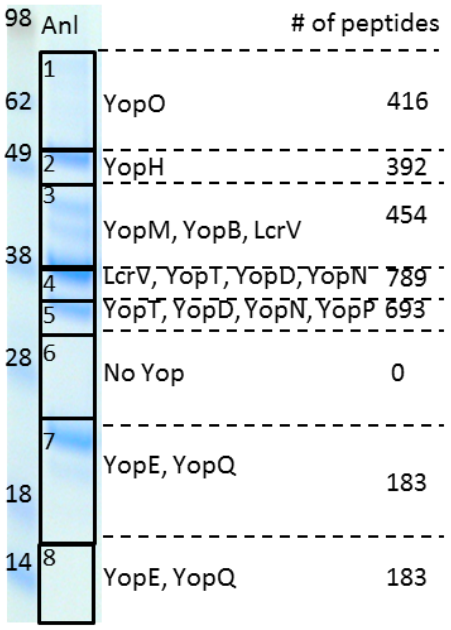
Figure S2.9.

Directed mass spectrometry search for gel-extracted peptides derived from Yops secreted under LCR conditions. A directed mass spectrometric search was used to determine the Met positions occupied by Anl in each Yop. Yops were secreted under LCR conditions in the presence of 1 mM Anl in 19-amino acid M9 medium supplemented with 25 µM Met and lacking calcium chloride. Secreted proteins were precipitated by chloroform/methanol treatment, resuspended in 1% SDS in PBS, electrophoresed on a 12% bis-tris acrylamide gel and stained with colloidal blue. The resulting bands were cut out, and in-gel tryptic digestion was performed for each gel fragment. To help identify peptides corresponding to each Yop, an m/z inclusion list for expected peptide ions was generated taking into account 2 missed cleavages by trypsin as well as fixed carbamidomethyl modification of cysteine and variable oxidation of Met. Replacement of Met by Anl was included as a variable modification. This inclusion list was used to perform directed LC-MS/MS experiments on an LTQ-FT Ultra mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source (Thermo Fisher Scientific) connected to an EASY-nLC II (Thermo Fisher Scientific).

A. Image of the cut out gel fragments with the total number of identified spectra shown for each fragment. The corresponding Yop positions are indicated next to the image. *B.* The total number of Met positions based on the protein sequence for each Yop and the molecular weight is listed. *C.* The number of identified spectra for each Yop is shown for each extracted gel section, the order of the gel slices correspond to decreasing molecular weight of protein.

Figure S2.9 continued.

a.



b.

Name of protein	MW	# Met
YopO	81280	18
YopH	50940	13
YopB	41916	15
YopM	41339	4
LcrV	37296	7
YopT	36082	8
YopD	33232	13
YopN	32925	5
YopP	32418	8
YopE	23016	9
YopQ	20829	4

c.

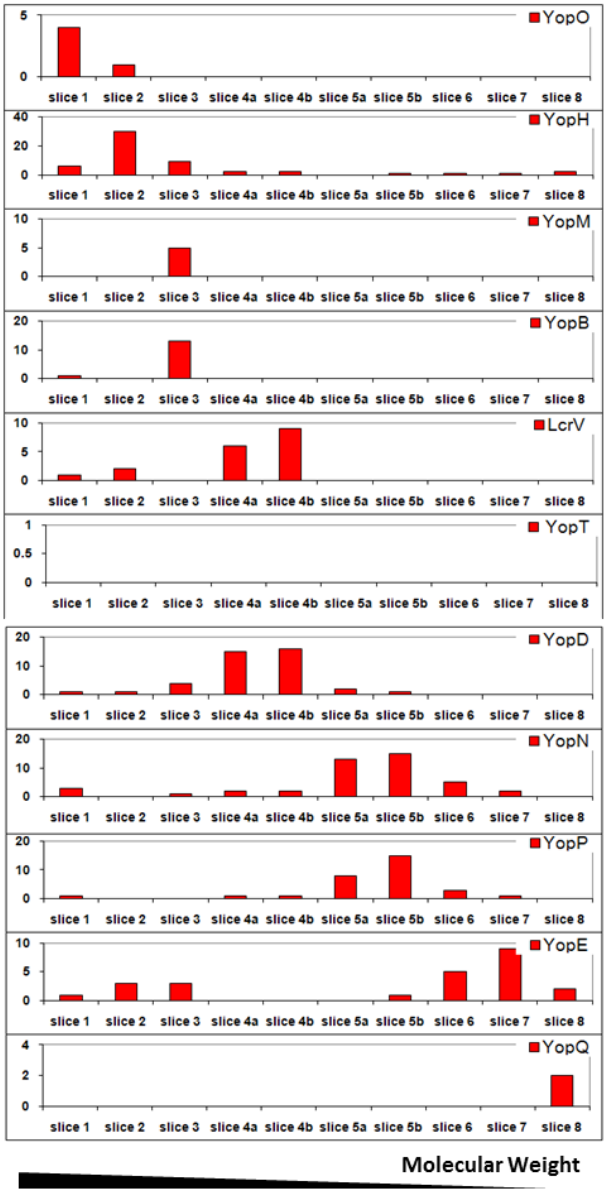


Figure S2.10.

Detection of residue-specific incorporation of Anl in Yops. Directed mass spectrometry was performed based on an expected peptide ion list (see Fig. S2.9 for details) from gel-extracted Yops from proteins secreted under LCR conditions An example of this analysis for YopH is illustrated here showing the sequence coverage obtained in presence or absence of Anl labeling

A. Highlighted sequences depict coverage by LC-MS/MS; modified Met residues are highlighted in green indicating oxidation of methionine in the "+Methionine" condition and incorporation of Anl in the "+Azidonorleucine" condition. Small differences were observed between the sequence coverages obtained for Met- and Anl-treated samples. *B.* The Met positions at which Anl was observed are shown for YopH as a function of residue position along the sequence. Black boxes are positions at which Anl was observed, and the frequencies at which Met and Anl were observed are shown for each position. *C.* A representative spectrum from an LC-MS/MS analysis on a precursor ion of the M^{+23}TQQVGLGDGIM$^{+16(ox)}$ADM^{+23}YTLTIR peptide is shown, with y-ions shown in blue and b-ions shown in red. A mass increment of 16 amu indicates that the Met at position 328 is oxidized. Anl is incorporated at positions 307 and 331, as indicated by mass shifts of 23 amu each. This mass shift corresponds to replacement of Met by Anl (Fig S2.2). A similar analysis was performed for each Yop.

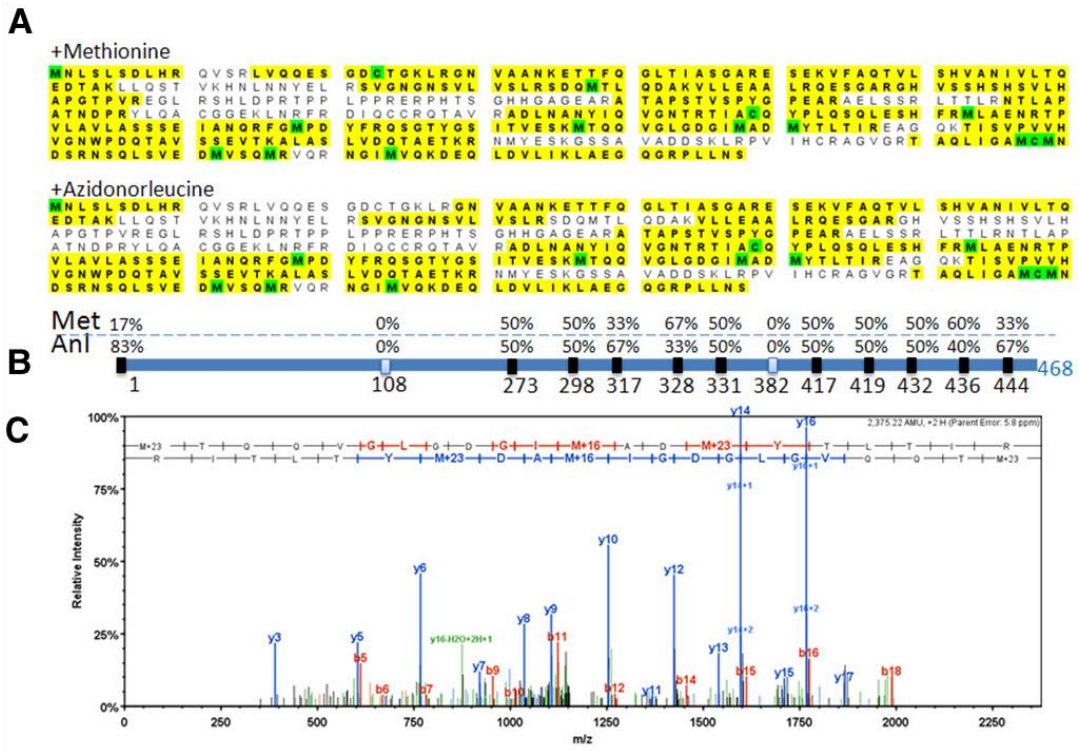
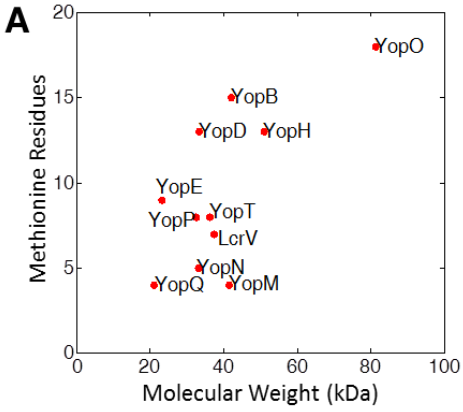
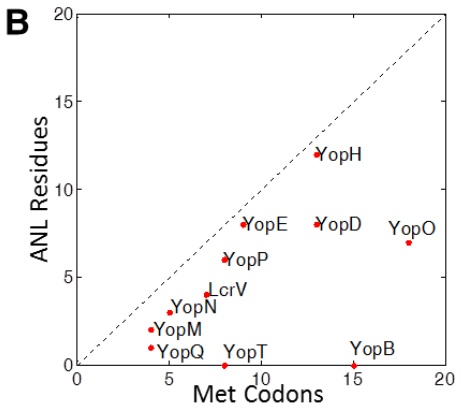


Figure S2.11.

Identifying the pattern of residue-specific replacement of methionine by azidonorleucine in Yops as a function of protein molecular weight and number of methionines. Directed mass spectrometry was used to determine the total number of Met positions at which Anl incorporation was observed. *A.* The number of Met positions as a function of molecular weight for each Yop is plotted to show the number of methionines contained in sequence of each Yop. *B.* The experimentally observed number of Anl residues (replacement of methionine) for each Yop as identified by MS. *C.* The sequence coverage observed by MS for each Yop. Sequence coverage is represented in yellow, and the position of each Met is represented as a circle. Black circles represent detection of Anl at a given Met position. These results were obtained from secretion of Yops under LCR conditions. The extent of labeling and observed sequence coverage may differ from those of Yops injected in HeLa cell infection.





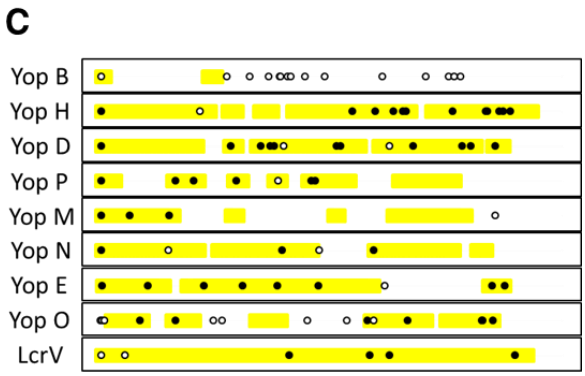


Figure S2.12.

Detection of biotin after click reaction of alkyne-biotin probe to proteins containing Anl. HeLa infections and secretion of Yops under LCR conditions were performed as described in the methods section of the main text. Proteins were treated with the biotin-alkyne affinity tag (structure **4** in the main text) and biotinylation was detected by western blot with a streptavidin Alexa fluor-633 conjugate antibody. Image below shows biotin labeling in injected HeLa cell lysates as well as proteins secreted under LCR conditions. Biotinylated *horseradish peroxidase (HRP)* was used as positive control for labeling.

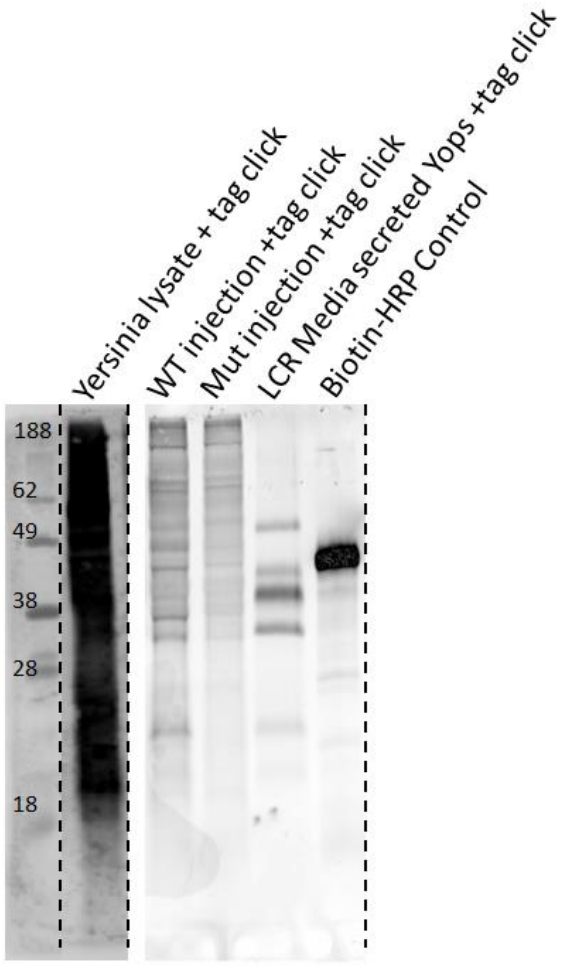


Figure S2.13.

Affinity enrichment of injected *Y. enterocolitica* proteins from infected HeLa cell lysates. *A.* Infected HeLa cell lysates were treated with probe **4** and affinity purified on a streptavidin column. Flow-through (FT) and elution (EL) fractions were stained with colloidal blue. *B.* Western blot detection of YopE and mammalian α-tubulin in the same fractions as those represented in *A.* YopE was enriched from HeLa infections with T3SS-Wt *Yersinia*. Lack of α-tubulin in the elution fractions confirmed efficient removal of HeLa cell proteins.

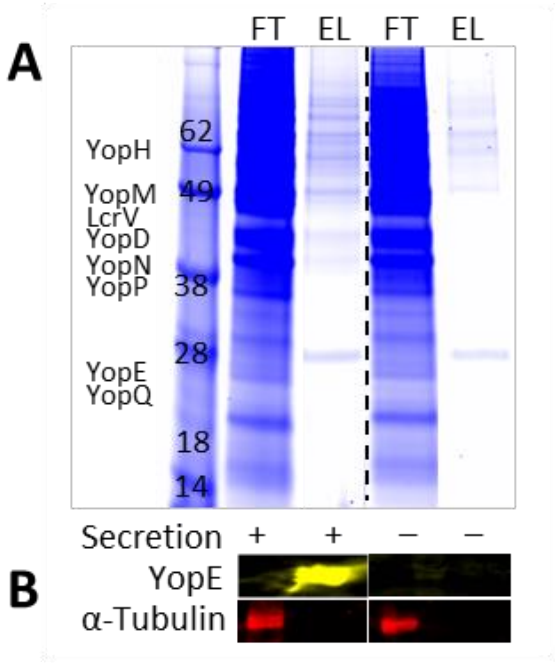


Figure S2.14.

Comparison of MS/MS spectra from Yops secreted under LCR conditions and Yops injected into HeLa cells. *A.* Tandem mass spectrum corresponding to a peptide from YopD from proteins secreted under LCR conditions. *B.* Tandem mass spectrum identified from infected HeLa cell lysates and corresponding to the same peptide sequence as in *A*. Both samples were gel extracted and desalted on C18 ziptip columns. A doubly charged m/z 701.86 amu peptide ion corresponding to the sequence "EKEVNASIAANEK" gave a precursor mass of 1401.71 (theoretical value 1401.7172). The y- and b-ions are color coded blue and red, respectively, along with the matching peptide sequence on top of each spectrum. Spectra were obtained on a nanoLc-Thermo Fisher Scientific LTQ-FT spectrometer with the peptide mass range set to 200-1700 amu.

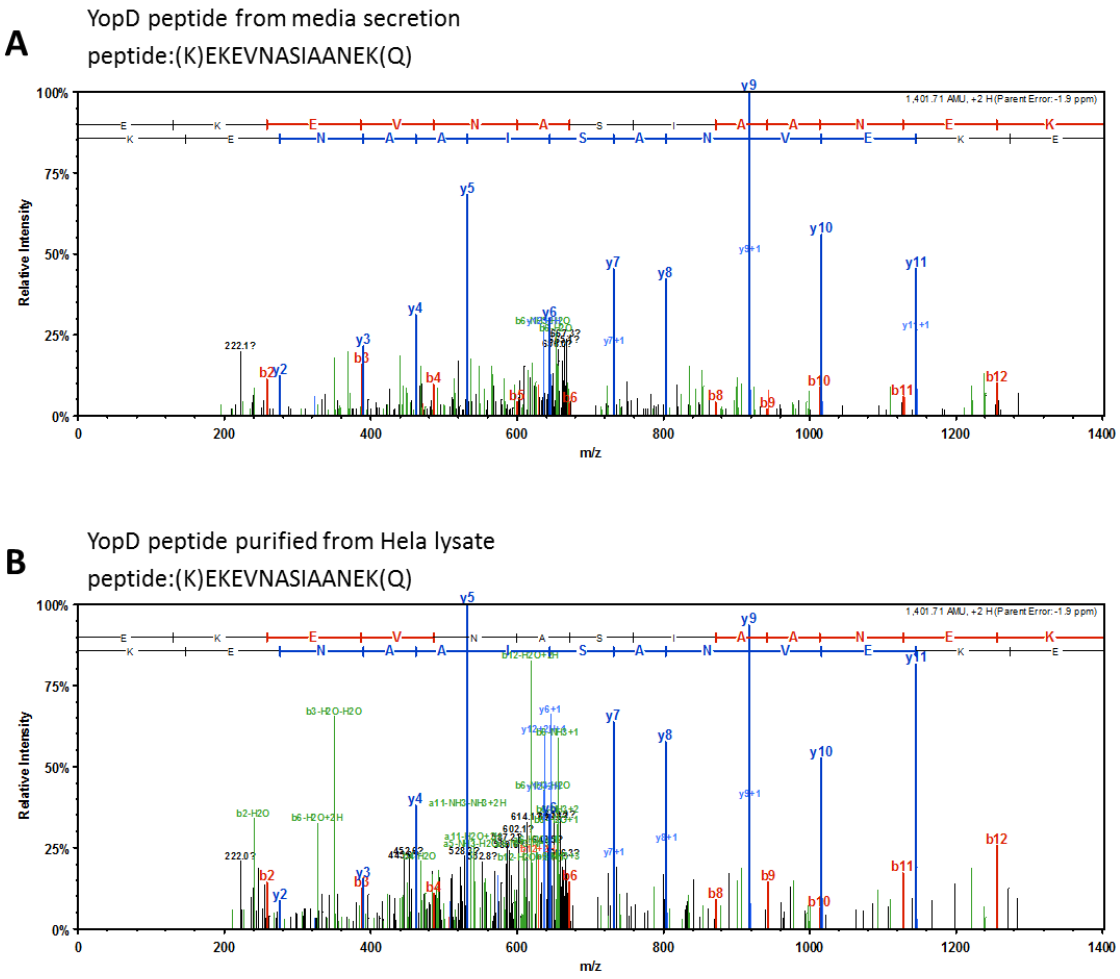


Figure S2.15.

Representative MS/MS spectra from Yops peptides identified after enrichment of inected proteins from HeLa cell lysates. The sequence of the peptide is shown on top of each spectrum. The y- and b-ions are color coded blue and red, respectively, along with the matching peptide sequence. A mass modification of M+121 amu at each Met position indicates the mass addition resulting from attachment and cleavage of the biotin-alkyne tag.

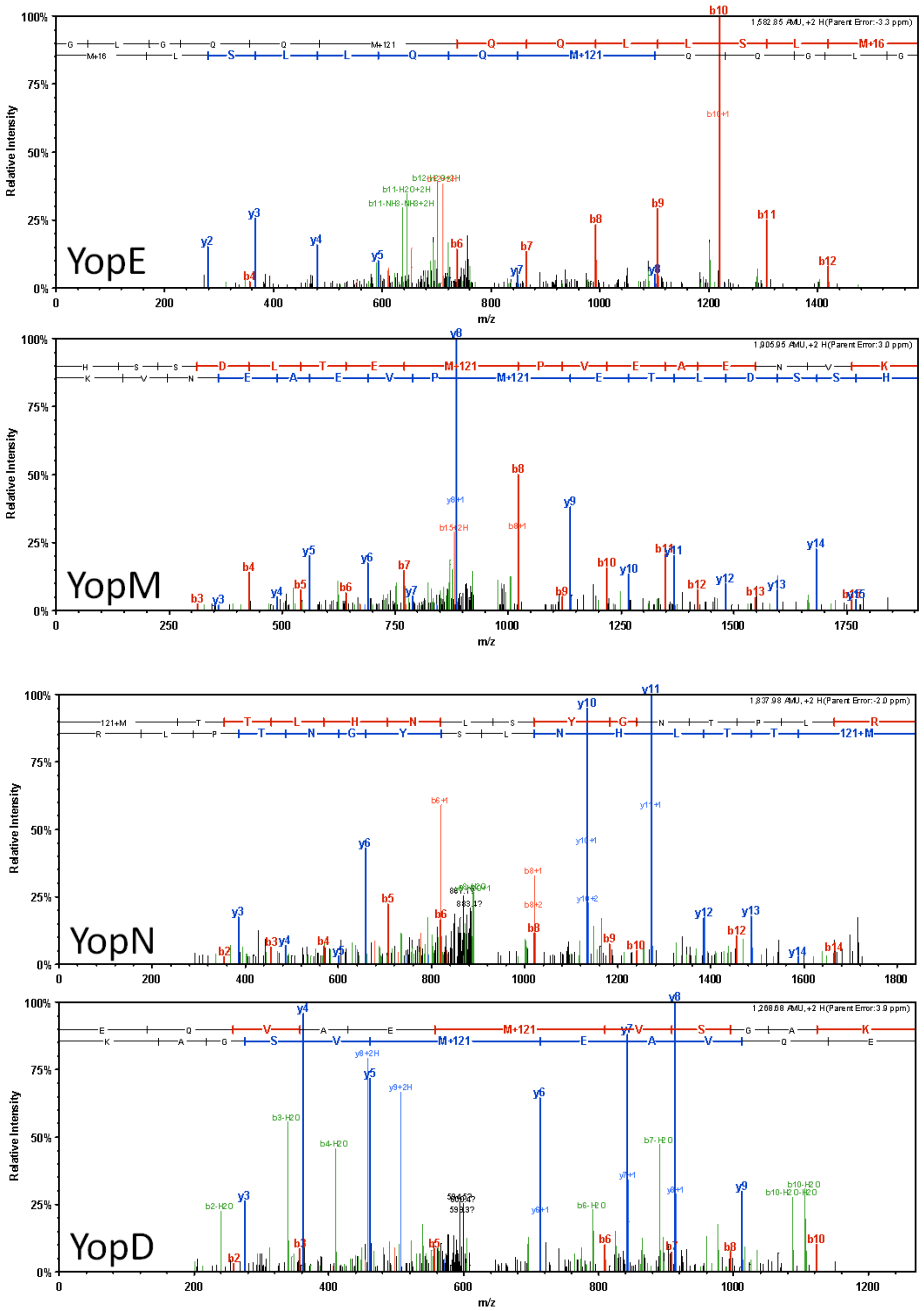


Figure S2.16.

Identification of secreted *Yersinia* proteins during infection. Infection medium was removed after the completion of HeLa infections, and proteins were precipitated with chloroform-methanol. Anl labeled proteins were tagged with probe **4**, enriched on a streptavidin column and analyzed by LC-MS. The secreted proteins identified here correspond to infections in Fig. 2.3*F* of main text. Three independent replicates were performed. The color-code next to the protein names indicates the number of identifications of each protein across experimental replicates. Venn diagram shows proteins detected in infections with T3SS-Wt *Yersinia* or the T3SS-Mut strain.

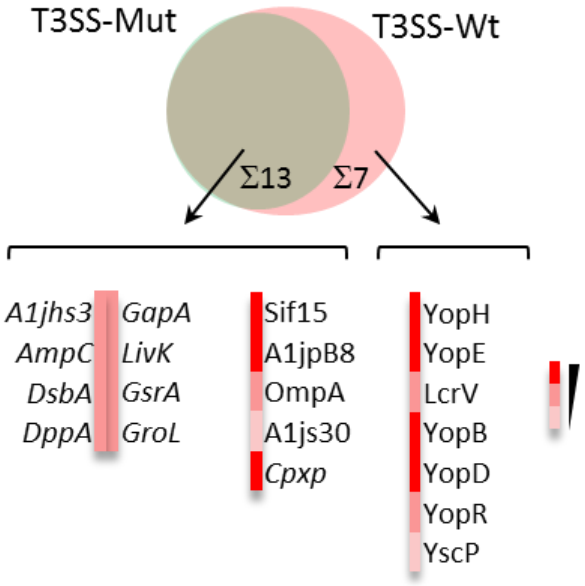


Figure S2.17.

Anl labeling of intracellular and extracellular bacteria in the absence of gentamicin. Infections corresponding to Fig. 2.4*A* of the main text were carried out using identical conditions; infections were carried out in the absence of gentamicin. Extracellular bacteria incorporated Anl into newly synthesized proteins and can be readily observed outside HeLa cells. In the image below, confocal fluorescence microscopy shows labeling of Anl residues with alkyne-AlexaFluor 488 (green). HeLa cell membranes were labeled with AlexaFluor633-WGA conjugate (red).

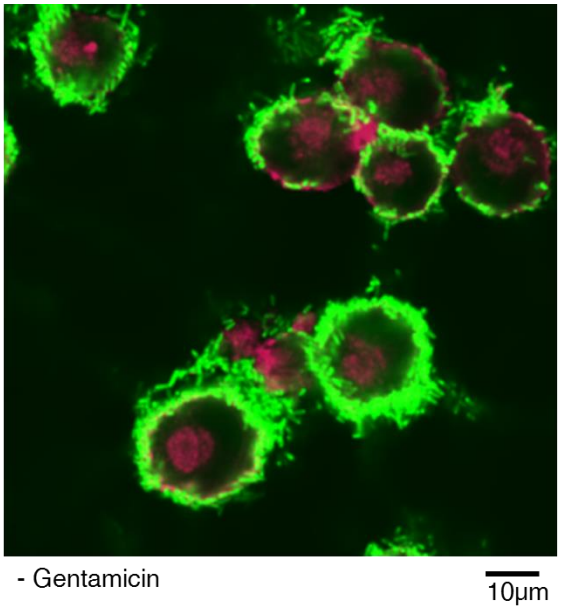
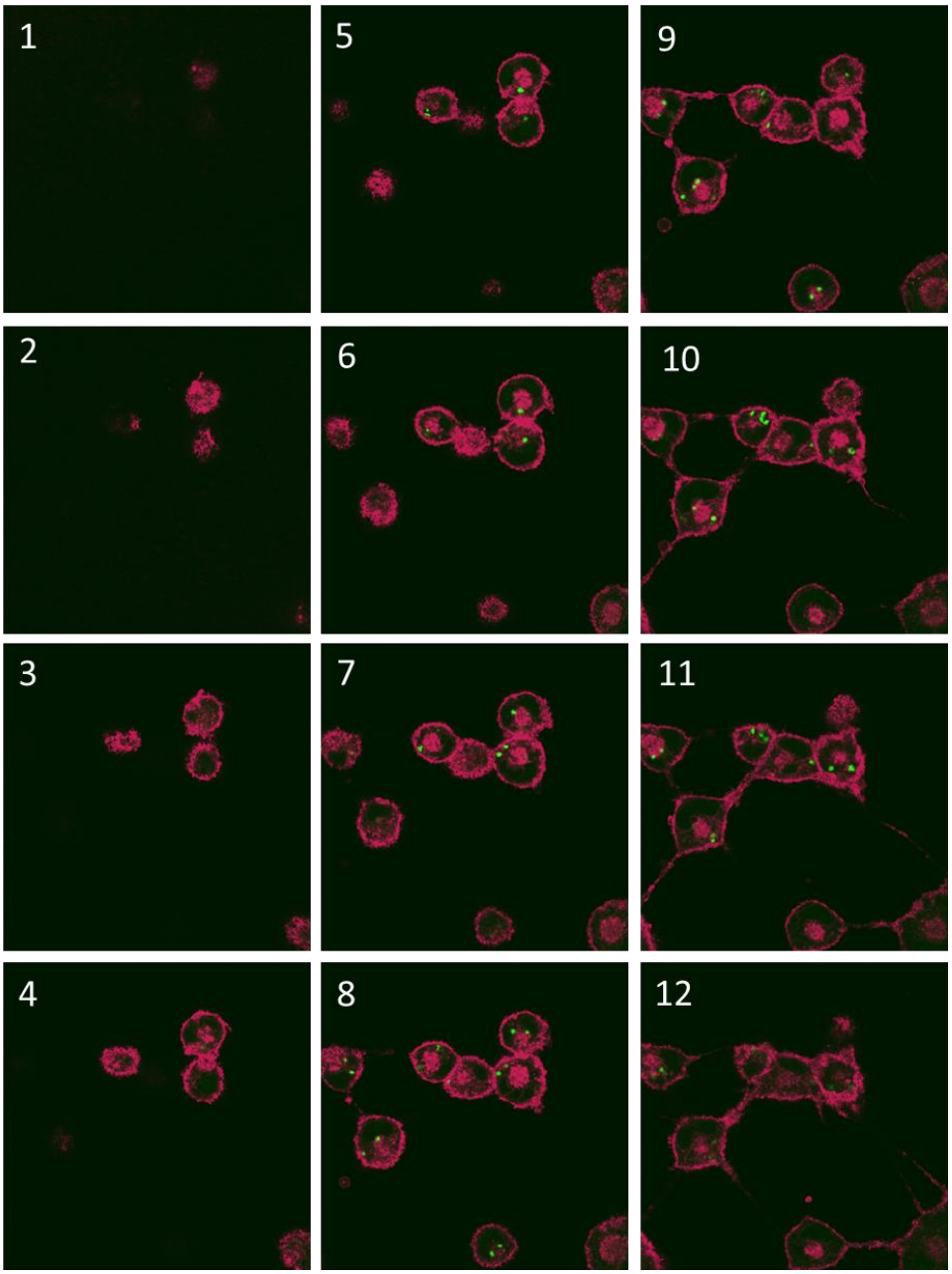


Figure S2.18.

Fluorescence confocal microscopy shows internalized *Yersinia* inside infected HeLa cells. Confocal planes from top to bottom of HeLa cell samples are numbered 1-12. AlexaFluor 633 conjugated to WGA was used to label the membranes of HeLa cells; the associated fluorescence is shown in red. Anl-labeled proteins were treated with alkyne-functionalized AlexaFluor 488; the associated fluorescence is shown in green. *A.* Gentamicin was added in this experiment to kill extracellular bacteria before Anl incorporation. Internalized *Yersinia* can be observed inside infected HeLa cells. See intracellular labeling video S1.



B. Gentamicin was not added in this experiment, and labeling was performed as described above. Both extracellular and intracellular bacteria have incorporated Anl into newly synthesized proteins and are labeled with alkyne-functionalized AlexaFluor 488 (green). See intracellular and extracellular labeling video S2.

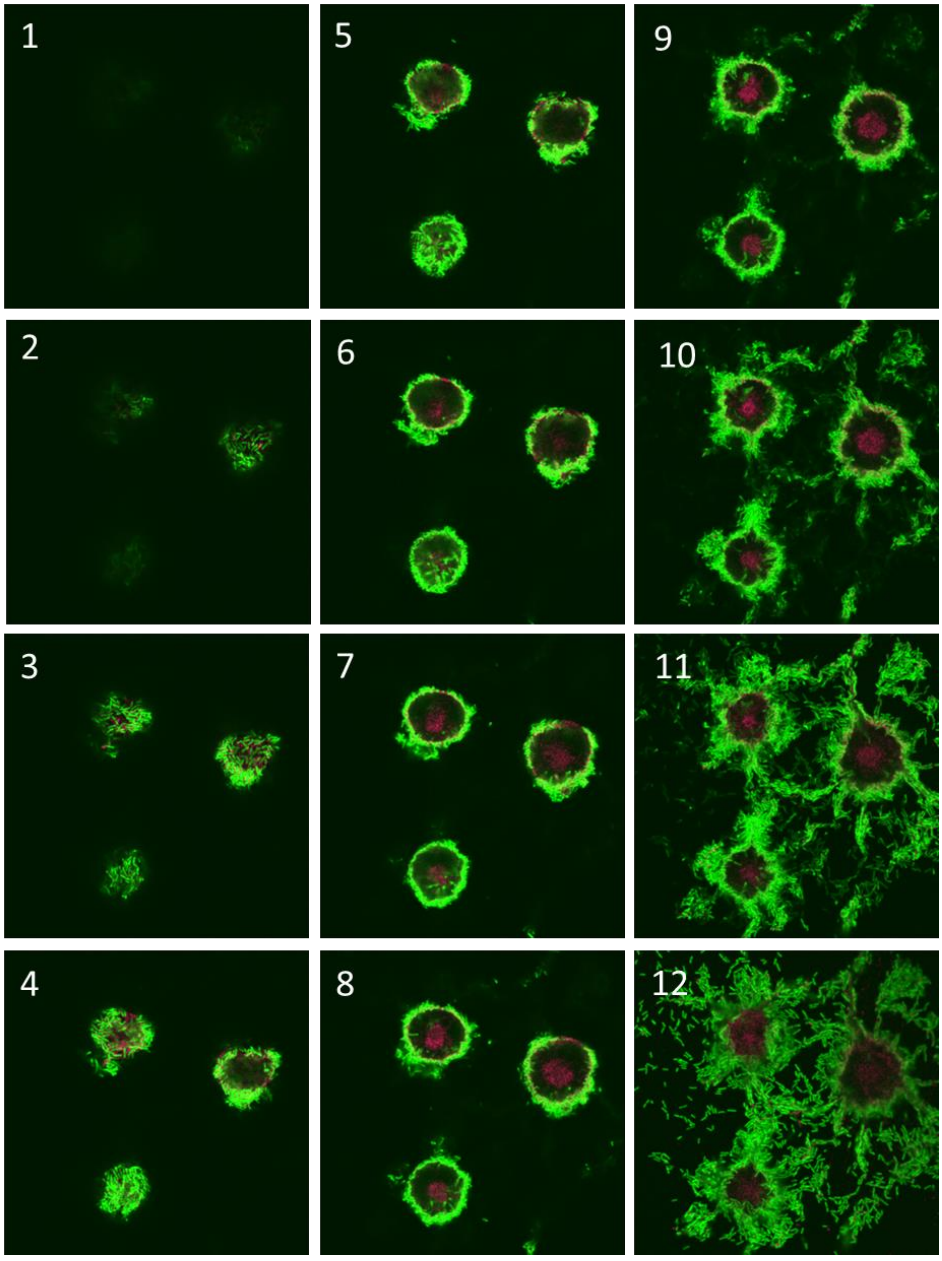


Figure S2.19.

Comparison of injected proteins in HeLa cells in presence or absence of gentamicin. Lysates of HeLa cells from infections corresponding to Fig 2.4*B* of the main text were analyzed by shotgun MS. Both infection mixtures were labeled with Anl for the same duration under identical conditions (except for the presence or absence of gentamicin). Anl labeled proteins were enriched using probe **4** before MS identification. Venn diagram shows the T3SS substrates identified under the two conditions, showing that intracellular *Yersinia* (+gent) injected a subset of Yops.

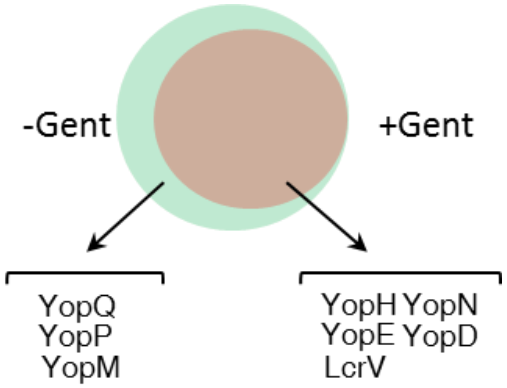


Table S2.1.

Proteins identified in infections with secretion wild type and secretion mutant *Y. enterocolitica*. Peptides were fractionated on a 15 cm reversed phase analytical column (75 μm ID) in-house packed with 3 μm C18 beads (ReproSil-Pur C18-AQ medium; Dr. Maisch GmbH) with a 60-minute gradient from 5% to 28% acetonitrile in 0.2% formic acid, over 50 minutes, followed by 10 minutes at 80 % acetonitrile in 0.2% formic acid, at a flow rate of 350 nl per minute.

The mass spectrometers were operated in data-dependent mode automatically switching between full-scan MS and tandem MS acquisition. Survey full scan mass spectra were acquired after accumulation of 500,000 ions, with a resolution of 60,000 (Orbitrap) and 50,000 (LTQ-FT) at 400 m/z. The top ten most intense ions from the survey scan were isolated and, after the accumulation of 5000 ions, fragmented in the linear ion trap by collision induced dissociation. Preview scan mode was enabled in the Orbitrap but not in the LTQ-FT. Precursor ion charge state screening was enabled and singly charged and unassigned charge states were rejected. The dynamic exclusion list was enabled with a relative mass window of 10 ppm and early expiration turned on.

T3SS substrates in HeLa infections were identified by a bottom up shotgun approach. MGF files were generated from Thermo RAW files using ReAdW4Mascot2 (v. 20090305a). The MGF files were searched using Mascot software (Matrix science, v2.2.06). The files were searched against the *Y. enterocolitica* proteome (5109 sequences, Uniprot) and a set of common contaminants (262 sequences). A decoy database consisting of the target protein sequences reversed was constructed to estimate the false discovery rate. The database search included mass modifications associated with Anl tagging (outlined in Fig. S2.2), as well as fixed carbamidomethyl modification of cysteine (+57.0214), variable oxidation of methionine (+15.9949) and N-terminal acetylation. A precursor tolerance of 10 ppm and fragment tolerance of 0.5 Da were used. Trypsin was the specified digestion enzyme and up to two missed cleavages were allowed. Scaffold software (Proteome software v3) was used to generate the list of identified proteins, using 99.9% protein probability, minimum of 2 peptides per protein and 95% peptide probability (Table S2.1) as thresholds for acceptance. Each identified protein was present in at least three independent infections. The false discovery rate was less than 1% in all experiments. MS analysis was performed on a nanoLC LTQ-FT mass spectrometer. MS data were analyzed by Mascot and Scaffold software packages.

All identified proteins had a false discovery rate of less than 1%. The criterion for inclusion of a protein found in the HeLa cell lysate was that it was observed in at least three different samples (of a total of six infections). The number of times a proteins was present was counted independently for secretion WT and secretion Mut infections; the larger of the two numbers is reported by the color code. For proteins secreted into the medium, the protein had to be present in two different samples (in a total of three infections) to be included in the Venn diagram. If a protein was observed in both the secretion WT and Mut infections in any of these samples, it was considered to be present in the joint distribution. YopQ was observed once in the secretion WT infection in HeLa cells and YscP was observed once in the secreted fraction isolated from infection with secretion WT *Yersinia*. YopE was observed once in the HeLa fraction of infections with secretion Mut *Yersinia*, but because it was present in every sample of injected proteins with secretion WT *Yersinia* it was retained in the secretion WT fraction in Fig. 2.3*F*. A1js30 was observed in three independent HeLa injections as reported in Fig. 2.3*F*. A1js30 was observed in only one of the samples of secreted proteins; although it did not meet the standard for inclusion in the secreted proteins, we included it to show that this protein may be present in the infection medium. Experiments are listed in reverse chronological order in the table below. Proteins identified in each experiment are sorted from highest to lowest based on the number of identified peptides and listed as such in the table. For time-series measurements using pulse-Anl labeling, three replicates were performed for early and intermediate time-points of 60 and 120 minutes respectively and identification was based on at least one observation of T3SS substrates across the samples. For the 210 minute sample we used the same list of proteins as that identified for Fig. 2.3*F* of the paper because these experiments used identical Anl labeling windows and infections. Variations in the numbers of identified proteins were observed as we improved our enrichment methods and instrument settings. Final enrichment methods and instrument settings are reported in the methods section.

Secretion Wt Injected Proteins 1	Accession Number
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
YopN OS=Yersinia enterocolitica GN=yopN PE=4 SV=1	O68333_YEREN
Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
Yop effector YopP OS=Yersinia enterocolitica GN=yopP PE=4 SV=1	O52162_YEREN
Chaperone protein htpG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=htpG PE=3 SV=1	HTPG_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Attachment invasion locus protein OS=Yersinia enterocolitica GN=ail PE=4 SV=1	D7PM19_YEREN
Yop effector YopM OS=Yersinia enterocolitica GN=yopM PE=4 SV=1	P74988_YEREN
Elongation factor G OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fusA PE=3 SV=1	A1JS54_YERE8
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
50S ribosomal protein L1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplA PE=3 SV=1	RL1_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8
Peptidoglycan-associated lipoprotein Pal OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pal PE=3 SV=1	A1JRK6_YERE8
Virulence-associated V antigen OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lcrV PE=4 SV=1	A1JU73_YERE8
Trigger factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tig PE=3 SV=1	TIG_YERE8
Outer membrane porin protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2463 PE=4 SV=1	A1JS30_YERE8
Adhesin yadA OS=Yersinia enterocolitica GN=yadA PE=1 SV=1	YADA1_YEREN
Putative outer membrane porin A protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ompA PE=3 SV=1	A1JMT3_YERE8

Cell division protein ftsZ OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ftsZ PE=3 SV=1	A1JJJ8_YERE8
Methionyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=metG PE=3 SV=1	SYM_YERE8
Secretion Wt Injected Proteins 2	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
Translocator protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopD PE=4 SV=1	A1JU70_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
Chaperone protein htpG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=htpG PE=3 SV=1	HTPG_YERE8
Elongation factor G OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fusA PE=3 SV=1	A1JS54_YERE8
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8
Clp ATPase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=clpB PE=3 SV=1	A1JK85_YERE8
Transaldolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tal PE=3 SV=1	TAL_YERE8
Methionyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=metG PE=3 SV=1	SYM_YERE8
Branched-chain amino acid-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=livK PE=3 SV=1	A1JIC6_YERE8
Ribosome-recycling factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=frr PE=3 SV=1	RRF_YERE8
Protein tolB OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tolB PE=3 SV=1	TOLB_YERE8
Putative outer membrane porin A protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ompA PE=3 SV=1	A1JMT3_YERE8
30S ribosomal protein S1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsA PE=4 SV=1	A1JMI7_YERE8

ATP-dependent protease ATPase subunit HslU OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hslU PE=3 SV=1	HSLU_YERE8
Trigger factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tig PE=3 SV=1	TIG_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
3-oxoacyl-[acyl-carrier protein] reductase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabG PE=3 SV=1	A1JN73_YERE8
Cell division protein ftsZ OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ftsZ PE=3 SV=1	A1JJJ8_YERE8
Phosphoglycerate kinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pgk PE=3 SV=1	PGK_YERE8
3-oxoacyl-[acyl-carrier-protein] synthase I OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabB PE=3 SV=1	A1JKM0_YERE8
Putative outer membrane porin F protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ompF PE=3 SV=1	A1JMN7_YERE8
DNA-directed RNA polymerase subunit alpha OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpoA PE=3 SV=1	RPOA_YERE8
Ribose-phosphate pyrophosphokinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=prsA PE=3 SV=1	A1JRV4_YERE8
ATP synthase subunit beta OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=atpD PE=3 SV=1	ATPB_YERE8
Lysyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lysS PE=3 SV=1	SYK_YERE8
Secretion Wt Injected Proteins 3	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
Chaperone protein htpG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=htpG PE=3 SV=1	HTPG_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Clp ATPase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=clpB PE=3 SV=1	A1JK85_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8

YopD OS=Yersinia enterocolitica GN=yopD PE=4 SV=1	C5IZG7_YEREN
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
Elongation factor G OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fusA PE=3 SV=1	A1JS54_YERE8
Ribosome-recycling factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=frr PE=3 SV=1	RRF_YERE8
Phosphoglycerate kinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pgk PE=3 SV=1	PGK_YERE8
Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
3-oxoacyl-[acyl-carrier-protein] synthase I OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabB PE=3 SV=1	A1JKM0_YERE8
Serine hydroxymethyltransferase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=glyA PE=3 SV=1	GLYA_YERE8
ATP-dependent protease ATPase subunit HslU OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hslU PE=3 SV=1	HSLU_YERE8
Trigger factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tig PE=3 SV=1	TIG_YERE8
Putative alkyl hydroperoxide reductase subunit c OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ahpC PE=4 SV=1	A1JNT6_YERE8
Formate acetyltransferase 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pflB PE=1 SV=1	A1JMG9_YERE8
Aldehyde-alcohol dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=adhE PE=4 SV=1	A1JQ89_YERE8
Fructose-bisphosphate aldolase class II OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fbaA PE=4 SV=1	A1JPQ8_YERE8
ATP synthase subunit alpha OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=atpA PE=3 SV=1	ATPA_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
3-oxoacyl-[acyl-carrier protein] reductase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabG PE=3 SV=1	A1JN73_YERE8
Transaldolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tal PE=3 SV=1	TAL_YERE8
Methionyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=metG PE=3 SV=1	SYM_YERE8
LcrV OS=Yersinia enterocolitica GN=lcrV PE=4 SV=1	O87495_YEREN
YopN OS=Yersinia enterocolitica GN=yopN PE=4 SV=1	O68333_YEREN
DNA protection during starvation protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dps PE=3 SV=1	DPS_YERE8

30S ribosomal protein S1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsA PE=4 SV=1	A1JMI7_YERE8
DNA-directed RNA polymerase subunit alpha OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpoA PE=3 SV=1	RPOA_YERE8
PTS system, mannose-specific IIAB component OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=manX PE=4 SV=1	A1JMB5_YERE8
Deoxyribose-phosphate aldolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=deoC PE=3 SV=1	DEOC_YERE8
50S ribosomal protein L10 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplJ PE=3 SV=1	RL10_YERE8
Attachment invasion locus protein OS=Yersinia enterocolitica GN=ail PE=4 SV=1	D7PM19_YEREN
Phosphoribosylaminoimidazole-succinocarboxamide synthase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=purC PE=3 SV=1	PUR7_YERE8
30S ribosomal protein S3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsC PE=3 SV=1	RS3_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Branched-chain amino acid-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=livK PE=3 SV=1	A1JIC6_YERE8
2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dapD PE=3 SV=1	DAPD_YERE8
30S ribosomal protein S5 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsE PE=3 SV=1	RS5_YERE8
Phosphate acetyltransferase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pta PE=4 SV=1	A1JLD1_YERE8
50S ribosomal protein L1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplA PE=3 SV=1	RL1_YERE8
50S ribosomal protein L4 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplD PE=3 SV=1	RL4_YERE8
50S ribosomal protein L9 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplI PE=3 SV=1	RL9_YERE8
Putative inner membrane protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2442 PE=4 SV=1	A1JRX6_YERE8
50S ribosomal protein L3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplC PE=3 SV=1	RL3_YERE8
Global stress requirement protein GsrA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gsrA PE=3 SV=1	A1JJQ8_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166	A1JPB8_YERE8

Secreted thiol:disulfide interchange protein DsbA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dsbA PE=3 SV=1	A1JHT2_YERE8
Pyruvate kinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pykF PE=3 SV=1	A1JPB0_YERE8
Virulence plasmid protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopQ PE=4 SV=1	A1JU64_YERE8
PTS sytem, enzyme I component OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ptsI PE=3 SV=1	A1JL94_YERE8
Probable alcohol dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2821 PE=3 SV=1	A1JU12_YERE8
Oligopeptidase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=opdA PE=3 SV=1	A1JSQ3_YERE8
ATP synthase subunit beta OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=atpD PE=3 SV=1	ATPB_YERE8
Yop effector YopP OS=Yersinia enterocolitica GN=yopP PE=4 SV=1	O52162_YEREN
Beta-lactamase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ampC PE=4 SV=1	A1JRW9_YERE8
DNA-directed RNA polymerase subunit beta' OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpoC PE=3 SV=1	RPOC_YERE8
Glutathione S-transferase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gst PE=4 SV=1	A1JNU4_YERE8
30S ribosomal protein S4 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsD PE=3 SV=1	RS4_YERE8
50S ribosomal protein L11 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplK PE=3 SV=1	RL11_YERE8
50S ribosomal protein L2 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplB PE=3 SV=1	RL2_YERE8
Uridylate kinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pyrH PE=3 SV=1	PYRH_YERE8
Adenylate kinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=adk PE=3 SV=1	KAD_YERE8
Acetate kinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ackA PE=3 SV=1	A1JLD2_YERE8
DNA-binding protein Hns OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hns PE=4 SV=1	A1JQA0_YERE8
Transketolase 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tktA PE=4 SV=1	A1JPR2_YERE8
GMP synthase [glutamine-hydrolyzing] OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=guaA PE=3 SV=1	GUAA_YERE8
Peptidyl-prolyl cis-trans isomerase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ppiB PE=3 SV=1	A1JNQ1_YERE8
50S ribosomal protein L15 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplO PE=3	RL15_YERE8
Protein tolB OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tolB PE=3 SV=1	TOLB_YERE8

Malate dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=mdh PE=3 SV=1	MDH_YERE8
30S ribosomal protein S7 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsG PE=3 SV=1	RS7_YERE8
Cell division protein ftsZ OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ftsZ PE=3 SV=1	A1JJJ8_YERE8
30S ribosomal protein S9 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsI PE=3 SV=1	RS9_YERE8
DnaK suppressor protein homologue OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dksA PE=4 SV=1	A1JJP2_YERE8
50S ribosomal protein L17 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplQ PE=3 SV=1	RL17_YERE8
Translation initiation factor IF-3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=infC PE=3 SV=1	IF3_YERE8
10 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groS PE=3 SV=1	CH10_YERE8
DNA-directed RNA polymerase subunit beta OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpoB PE=3 SV=1	RPOB_YERE8
6-phosphofructokinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pfkA PE=3 SV=1	K6PF_YERE8
Outer membrane protein C, porin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=meoA PE=3 SV=1	A1JLE1_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yrbC PE=3 SV=1	A1JRA6_YERE8
Putative D-ribose-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rbsB PE=3 SV=1	A1JHS3_YERE8
Cationic 19 kDa outer membrane protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ompH PE=3 SV=1	A1JP72_YERE8
Adhesin yadA OS=Yersinia enterocolitica GN=yadA PE=1 SV=1	YADA1_YEREN
UPF0234 protein YE3147 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3147 PE=3 SV=1	Y3147_YERE8
Putative uncharacterized protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE1738 PE=4 SV=1	A1JLU4_YERE8
Yop effector YopM OS=Yersinia enterocolitica GN=yopM PE=4 SV=1	P74988_YEREN
Cysteine synthase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cysK PE=3 SV=1	A1JL97_YERE8
Single-stranded DNA-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ssb PE=4 SV=1	A1JRR5_YERE8
30S ribosomal protein S8 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsH PE=3 SV=1	RS8_YERE8

UPF0304 protein YE1336 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE1336 PE=3 SV=1	Y1336_YERE8
Galactose-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=mglB PE=3 SV=1	A1JU06_YERE8
NADH:flavin oxidoreductase / NADH oxidase family protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2152 PE=4 SV=1	A1JP27_YERE8
cAMP-regulatory protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=crp PE=4 SV=1	A1JS98_YERE8
Uracil phosphoribosyltransferase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=upp PE=3 SV=1	UPP_YERE8
Putative stringent starvation protein A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=sspA PE=4 SV=1	A1JR91_YERE8
Putative lipoprotein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE1703 PE=4 SV=1	A1JME2_YERE8
PTS system, glucose-specific IIA component OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=crr PE=4 SV=1	A1JL89_YERE8
Ribose-phosphate pyrophosphokinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=prsA PE=3 SV=1	A1JRV4_YERE8
UPF0082 protein YE2395 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2395 PE=3 SV=1	Y2395_YERE8
Biotin carboxylase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=accC PE=4 SV=1	A1JRK3_YERE8
D-lactate dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ldhA PE=4 SV=1	A1JN94_YERE8
Thioredoxin reductase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=trxB PE=3 SV=1	A1JME7_YERE8
Putative lipoprotein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE1719 PE=3 SV=1	A1JLF7_YERE8
Heat-shock chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hslO PE=3 SV=1	A1JSE6_YERE8
Protease OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=degQ PE=3 SV=1	A1JR98_YERE8
Asparaginyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=asnS PE=3 SV=1	SYN_YERE8
Seryl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=serS PE=3 SV=1	SYS_YERE8
Transcription antitermination protein nusG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=nusG PE=3 SV=1	A1JIH5_YERE8
Phosphoheptose isomerase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gmhA PE=3 SV=1	GMHA_YERE8

Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3737 PE=3 SV=1	A1JR88_YERE8
Putative NAD(P)H-dependent FMN reductase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ssuE PE=3 SV=1	A1JT96_YERE8
50S ribosomal protein L16 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplP PE=3 SV=1	RL16_YERE8
Triosephosphate isomerase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tpiA PE=3 SV=1	TPIS_YERE8
ABC transporter protein, ATP-binding component OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE0584 PE=4 SV=1	A1JJB0_YERE8
Peptidoglycan-associated lipoprotein Pal OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pal PE=3 SV=1	A1JRK6_YERE8
Bifunctional purine biosynthesis protein purH OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=purH PE=3 SV=1	PUR9_YERE8
Secretion Wt Injected Proteins 4	Accession Number
Translocator protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopD PE=4 SV=1	A1JU70_YERE8
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
Type III secretion outer membrane protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopN PE=4	A1JU81_YERE8
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
Virulence-associated V antigen OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lcrV PE=4 SV=1	A1JU73_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Yop type III secretion system effector protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopM PE=4	A1JU68_YERE8
Secretion Wt Injected Proteins 5	Accession Number
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
Virulence-associated V antigen OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lcrV PE=4 SV=1	A1JU73_YERE8
Translocator protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopD PE=4 SV=1	A1JU70_YERE8

YopN OS=Yersinia enterocolitica GN=yopN PE=4 SV=1	O68333_YEREN
Yop effector YopM OS=Yersinia enterocolitica GN=yopM PE=4 SV=1	P74988_YEREN
Secretion Wt Injected Proteins 6	Accession Number
LcrV OS=Yersinia enterocolitica GN=lcrV PE=4 SV=1	O87495_YEREN
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
YopD OS=Yersinia enterocolitica GN=yopD PE=4 SV=1	C5IZG7_YEREN
YopN OS=Yersinia enterocolitica GN=yopN PE=4 SV=1	O68333_YEREN
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
Yop effector YopM OS=Yersinia enterocolitica GN=yopM PE=4 SV=1	P74988_YEREN
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Secretion Mut Injected Proteins 1	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Chaperone protein htpG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=htpG PE=3 SV=1	HTPG_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Clp ATPase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=clpB PE=3 SV=1	A1JK85_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8
50S ribosomal protein L1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplA PE=3 SV=1	RL1_YERE8
Branched-chain amino acid-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=livK PE=3 SV=1	A1JIC6_YERE8
Global stress requirement protein GsrA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gsrA PE=3 SV=1	A1JJQ8_YERE8
Elongation factor G OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fusA PE=3 SV=1	A1JS54_YERE8
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
3-oxoacyl-[acyl-carrier-protein] synthase I OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabB PE=3 SV=1	A1JKM0_YERE8
Ribosome-recycling factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=frr PE=3 SV=1	RRF_YERE8

Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
Trigger factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tig PE=3 SV=1	TIG_YERE8
Primosomal protein n' OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=priA PE=4 SV=1	A1JI12_YERE8
Methionyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=metG PE=3 SV=1	SYM_YERE8
ATP-dependent protease ATPase subunit HslU OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hslU PE=3 SV=1	HSLU_YERE8
Putative alkyl hydroperoxide reductase subunit c OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ahpC PE=4 SV=1	A1JNT6_YERE8
50S ribosomal protein L10 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplJ PE=3 SV=1	RL10_YERE8
3-oxoacyl-[acyl-carrier protein] reductase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabG PE=3 SV=1	A1JN73_YERE8
Putative D-ribose-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rbsB PE=3 SV=1	A1JHS3_YERE8
PTS sytem, enzyme I component OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ptsI PE=3 SV=1	A1JL94_YERE8
Aldehyde-alcohol dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=adhE PE=4 SV=1	A1JQ89_YERE8
Secretion Mut Injected Proteins 2	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Chaperone protein htpG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=htpG PE=3 SV=1	HTPG_YERE8
Elongation factor G OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fusA PE=3 SV=1	A1JS54_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Methionyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=metG PE=3 SV=1	SYM_YERE8
Trigger factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tig PE=3 SV=1	TIG_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN

Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8
3-oxoacyl-[acyl-carrier protein] reductase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabG PE=3 SV=1	A1JN73_YERE8
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
Branched-chain amino acid-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=livK PE=3 SV=1	A1JIC6_YERE8
3-oxoacyl-[acyl-carrier-protein] synthase I OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabB PE=3 SV=1	A1JKM0_YERE8
Secretion Mut Injected Proteins 3	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Clp ATPase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=clpB PE=3 SV=1	A1JK85_YERE8
Chaperone protein htpG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=htpG PE=3 SV=1	HTPG_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
Elongation factor G OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fusA PE=3 SV=1	A1JS54_YERE8
Putative alkyl hydroperoxide reductase subunit c OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ahpC PE=4 SV=1	A1JNT6_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Flagellar hook-associated protein 2 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fliD PE=4 SV=1	A1JSR2_YERE8
Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
Trigger factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tig PE=3 SV=1	TIG_YERE8
FliC OS=Yersinia enterocolitica GN=fliC PE=4 SV=1	C9EHF4_YEREN
30S ribosomal protein S1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsA PE=4 S	A1JMI7_YERE8
ATP-dependent protease ATPase subunit HslU OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hslU PE=3 SV=1	HSLU_YERE8

Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Aldehyde-alcohol dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=adhE PE=4 SV=1	A1JQ89_YERE8
Ribosome-recycling factor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=frr PE=3 SV=1	RRF_YERE8
3-oxoacyl-[acyl-carrier protein] reductase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabG PE=3 SV=1	A1JN73_YERE8
ATP synthase subunit beta OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=atpD PE=3 SV=1	ATPB_YERE8
ATP synthase subunit alpha OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=atpA PE=3 SV=1	ATPA_YERE8
30S ribosomal protein S3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsC PE=3 SV=1	RS3_YERE8
3-oxoacyl-[acyl-carrier-protein] synthase I OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabB PE=3 SV=1	A1JKM0_YERE8
DNA-directed RNA polymerase subunit alpha OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpoA PE=3 SV=1	RPOA_YERE8
Global stress requirement protein GsrA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gsrA PE=3 SV=1	A1JJQ8_YERE8
Protease OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=degQ PE=3 SV=1	A1JR98_YERE8
50S ribosomal protein L3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplC PE=3 SV=1	RL3_YERE8
Fructose-bisphosphate aldolase class II OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fbaA PE=4 SV=1	A1JPQ8_YERE8
Formate acetyltransferase 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pflB PE=1 SV=1	A1JMG9_YERE8
Methionyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=metG PE=3 SV=1	SYM_YERE8
Transaldolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tal PE=3 SV=1	TAL_YERE8
Adhesin, Ail protein OS=Yersinia enterocolitica (type O:3) GN=ail PE=4 SV=1	Q70AM3_YEREN
30S ribosomal protein S5 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsE PE=3 SV=1	RS5_YERE8
50S ribosomal protein L4 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplD PE=3 SV=1	RL4_YERE8
50S ribosomal protein L9 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplI PE=3 SV	RL9_YERE8
Putative inner membrane protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2442 PE=4 SV=1	A1JRX6_YERE8

Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
PTS sytem, enzyme I component OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ptsI PE=3 SV=1	A1JL94_YERE8
PTS system, mannose-specific IIAB component OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=manX PE=4 SV=1	A1JMB5_YERE8
DNA protection during starvation protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dps PE=3 SV=1	DPS_YERE8
Deoxyribose-phosphate aldolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=deoC PE=3 SV=1	DEOC_YERE8
50S ribosomal protein L1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplA PE=3 SV=1	RL1_YERE8
Protein recA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=recA PE=3 SV=1	RECA_YERE8
50S ribosomal protein L10 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplJ PE=3 SV=1	RL10_YERE8
50S ribosomal protein L17 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplQ PE=3 SV=1	RL17_YERE8
Branched-chain amino acid-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=livK PE=3 SV=1	A1JIC6_YERE8
50S ribosomal protein L5 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplE PE=3 SV=1	RL5_YERE8
30S ribosomal protein S2 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsB PE=3 SV=1	RS2_YERE8
Serine hydroxymethyltransferase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=glyA PE=3 SV=1	GLYA_YERE8
DNA-directed RNA polymerase subunit beta' OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpoC PE=3 SV=1	RPOC_YERE8
Malate dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=mdh PE=3 SV=1	MDH_YERE8
30S ribosomal protein S4 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsD PE=3 SV=1	RS4_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
Probable alcohol dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2821 PE=3 SV=1	A1JU12_YERE8
Translation initiation factor IF-3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=infC PE=3 SV=1	IF3_YERE8
Primosomal protein n' OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=priA PE=4	A1JI12_YERE8

SV=1	
Pyruvate kinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pykF PE=3 SV=1	A1JPB0_YERE8
Cell division protein ftsZ OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ftsZ PE=3 SV=1	A1JJJ8_YERE8
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
Beta-lactamase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ampC PE=4 SV=1	A1JRW9_YERE8
Dihydrolipoyl dehydrogenase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lpdA PE=3 SV=1	A1JJM0_YERE8
Asparaginyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=asnS PE=3 SV=1	SYN_YERE8
Uridylate kinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pyrH PE=3 SV=1	PYRH_YERE8
Putative membrane protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hflK PE=4 SV=1	A1JIR7_YERE8
Ribose-phosphate pyrophosphokinase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=prsA PE=3 SV=1	A1JRV4_YERE8
UPF0234 protein YE3147 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3147 PE=3 SV=1	Y3147_YERE8
50S ribosomal protein L2 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplB PE=3 SV=1	RL2_YERE8
Triosephosphate isomerase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tpiA PE=3 SV=1	TPIS_YERE8
50S ribosomal protein L11 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplK PE=3 SV=1	RL11_YERE8
Peptidyl-prolyl cis-trans isomerase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ppiB PE=3 SV=1	A1JNQ1_YERE8
Seryl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=serS PE=3 SV=1	SYS_YERE8
30S ribosomal protein S7 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsG PE=3 SV=1	RS7_YERE8
Outer membrane porin protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2463 PE=4 SV=1	A1JS30_YERE8
Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=aceF PE=3 SV=1	A1JJL9_YERE8
Flagellar hook-associated protein 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=flgK PE=3 SV=1	A1JSZ7_YERE8
Secretion Mut Injected Proteins 4	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8

Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3	A1JQH1_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
Outer membrane porin protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2463 PE=4 SV=1	A1JS30_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
Secretion Mut Injected Proteins 5	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3	A1JQH1_YERE8
Outer membrane porin protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2463 PE=4 SV=1	A1JS30_YERE8
Attachment invasion locus protein OS=Yersinia enterocolitica GN=ail PE=4 SV=1	D7PM19_YEREN
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
Secretion Mut Injected Proteins 6	Accession Number
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Attachment invasion locus protein OS=Yersinia enterocolitica GN=ail PE=4 SV=1	D7PM19_YEREN
Pulsed Anl labeling, 0-30 min	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Major outer membrane lipoprotein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081)	A1JPB4_YERE8

30S ribosomal protein S3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsC PE=3 SV=1	RS3_YERE8
50S ribosomal protein L1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplA PE=3 SV=1	RL1_YERE8
Putative uncharacterized protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3042 PE=4 SV=1	A1JNW8_YERE8
Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
Elongation factor Tu (Fragment) OS=Yersinia enterocolitica GN=tuf PE=4 SV=1	D1MWR3_YEREN
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
Flagellar basal-body rod protein flgB OS=Yersinia enterocolitica GN=flgB PE=3 SV=3	FLGB_YEREN
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE1132 PE=3 SV=1	A1JL02_YERE8
YopD OS=Yersinia enterocolitica GN=yopD PE=4 SV=1	D0FH99_YEREN
Preprotein translocase subunit secY OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=secY PE=3 SV=1	A1JS07_YERE8
Citrate lyase acyl carrier protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=citD PE=3 SV=1	CITD_YERE8
DNA-binding protein Hns OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hns PE=4 SV=1	A1JQA0_YERE8
Pseudouridine synthase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rluA PE=3 SV=1	A1JJF9_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3439 PE=3 SV=1	A1JPV0_YERE8
Pulsed Anl labeling, 30-60 min	Accession Number
Outer membrane virulence protein yopE OS=Yersinia enterocolitica GN=yopE PE=2 SV=1	YOPE_YEREN
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Tyrosine-protein phosphatase yopH OS=Yersinia enterocolitica GN=yopH PE=1 SV=1	YOPH_YEREN
Major outer membrane lipoprotein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lpp PE=3 SV=1	A1JPB4_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
30S ribosomal protein S3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsC PE=3 SV=1	RS3_YERE8
2,3-bisphosphoglycerate-dependent phosphoglycerate mutase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gpmA PE=3 SV=1	GPMA_YERE8
Elongation factor Tu (Fragment) OS=Yersinia enterocolitica GN=tuf PE=4 SV=1	D1MWR3_YEREN

Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
Pseudouridine synthase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rluA PE=3 SV=1	A1JJF9_YERE8
Putative uncharacterized protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3042 PE=4 SV=1	A1JNW8_YERE8
Primosomal protein n' OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=priA PE=4 SV=1	A1JI12_YERE8
Translocator protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopD PE=4 SV=1	A1JU70_YERE8
50S ribosomal protein L15 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplO PE=3 SV=1	RL15_YERE8
Pulsed Anl labeling, 60-90 min	Accession Number
Tyrosine-protein phosphatase yopH OS=Yersinia enterocolitica GN=yopH PE=1 SV=1	YOPH_YEREN
Outer membrane virulence protein yopE OS=Yersinia enterocolitica GN=yopE PE=2 SV=1	YOPE_YEREN
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
YopD OS=Yersinia enterocolitica GN=yopD PE=4 SV=1	D0FH99_YEREN
Yop effector YopP OS=Yersinia enterocolitica GN=yopP PE=4 SV=1	O52162_YEREN
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Elongation factor Tu 2 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf2 PE=3 SV=1	EFTU2_YERE8
30S ribosomal protein S3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsC PE=3 SV=1	RS3_YERE8
YopN OS=Yersinia enterocolitica GN=yopN PE=4 SV=1	O68333_YEREN
Type III secretion outer membrane protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopN PE=4 SV=1	A1JU81_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Elongation factor Tu (Fragment) OS=Yersinia enterocolitica GN=tuf PE=4 SV=1	D1MWR3_YEREN
50S ribosomal protein L1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplA PE=3 SV=1	RL1_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8
30S ribosomal protein S5 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsE PE=3 SV=1	RS5_YERE8
Pseudouridine synthase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rluA PE=3	A1JJF9_YERE8

Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
Flagellar basal-body rod protein flgB OS=Yersinia enterocolitica GN=flgB PE=3 SV=3	FLGB_YEREN
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE1132 PE=3 SV=1	A1JL02_YERE8
30S ribosomal protein S8 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsH PE=3 SV=1	RS8_YERE8
Major outer membrane lipoprotein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lpp PE=3 SV=1	A1JPB4_YERE8
Putative prophage integrase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE0993 PE=4 SV=1	A1JKH9_YERE8
PduX OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pduX PE=4 SV=1	A1JSN3_YERE8
Yop effector YopM OS=Yersinia enterocolitica GN=yopM PE=4 SV=1	P74988_YEREN
Putative methyl-accepting chemotaxis protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=aer PE=4 SV=1	A1JLQ5_YERE8
50S ribosomal protein L5 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplE PE=3 SV=1	RL5_YERE8
Arginine transport system permease protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=artM PE=3 SV=1	A1JM76_YERE8
Primosomal protein n' OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=priA PE=4 SV=1	A1JI12_YERE8
Pulsed Anl labeling, 90-180 min	Accession Number
Tyrosine-protein phosphatase yopH OS=Yersinia enterocolitica GN=yopH PE=1 SV=1	YOPH_YEREN
Outer membrane virulence protein yopE OS=Yersinia enterocolitica GN=yopE PE=2 SV=1	YOPE_YEREN
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Protein yopD OS=Yersinia enterocolitica GN=yopD PE=4 SV=1	YOPD_YEREN
Putative outer membrane virulence protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YEP0053 PE=3 SV=1	A1JUA9_YERE8
Translocator protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopD PE=4 SV=1	A1JU70_YERE8
Yop effector YopP OS=Yersinia enterocolitica GN=yopP PE=4 SV=1	O52162_YEREN
YopN OS=Yersinia enterocolitica GN=yopN PE=4 SV=1	O68333_YEREN
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8

Type III secretion system effector protein (Putative targeted effector protein) OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopJ PE=4 SV=1	A1JUC5_YERE8
Type III secretion outer membrane protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopN PE=4 SV=1	A1JU81_YERE8
Yop effector YopM OS=Yersinia enterocolitica GN=yopM PE=4 SV=1	P74988_YEREN
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
Attachment invasion locus protein (Fragment) OS=Yersinia enterocolitica (type O:3) GN=ail PE=4 SV=1	Q4ZJF6_YEREN
Protein yopQ OS=Yersinia enterocolitica GN=yopQ PE=4 SV=1	YOPQ_YEREN
Protein kinase YopO OS=Yersinia enterocolitica GN=yopO PE=4 SV=1	O85239_YEREN
Chaperone protein htpG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=htpG PE=3 SV=1	HTPG_YERE8
Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
Major outer membrane lipoprotein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lpp PE=3 SV=1	A1JPB4_YERE8
Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
2,3-bisphosphoglycerate-dependent phosphoglycerate mutase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gpmA PE=3 SV=1	GPMA_YERE8
ORF protein (Fragment) OS=Yersinia enterocolitica GN=ORF PE=4 SV=1	Q56881_YEREN
Protein yopB OS=Yersinia enterocolitica GN=yopB PE=4 SV=1	YOPB_YEREN
Virulence-associated V antigen OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lcrV PE=4 SV=1	A1JU73_YERE8
Regulatory protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yscM2 PE=4 SV=1	A1JUC6_YERE8
30S ribosomal protein S3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsC PE=3 SV=1	RS3_YERE8
Peptidoglycan-associated lipoprotein Pal OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pal PE=3 SV=1	A1JRK6_YERE8
Outer membrane porin protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2463 PE=4 SV=1	A1JS30_YERE8
50S ribosomal protein L1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplA PE=3 SV=1	RL1_YERE8
Global stress requirement protein GsrA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gsrA PE=3 SV=1	A1JJQ8_YERE8
Enolase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=eno PE=3 SV=1	ENO_YERE8
Putative alkyl hydroperoxide reductase subunit c OS=Yersinia enterocolitica serotype O:8 / biotype 1B	A1JNT6_YERE8

Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Yop proteins translocation protein M OS=Yersinia enterocolitica GN=yscM PE=2 SV=1	YSCM_YEREN
Pseudouridine synthase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rluA PE=3 SV=1	A1JJF9_YERE8
Protein tolB OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tolB PE=3 SV=1	TOLB_YERE8
Elongation factor Ts OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tsf PE=3 SV=1	EFTS_YERE8
Putative transcription accessory protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=teX PE=4 SV=1	A1JSG5_YERE8
DNA-directed RNA polymerase subunit beta' OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpoC PE=3 SV=1	RPOC_YERE8
Gluconate utilization system Gnt-I transcriptional repressor OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gntR PE=4 SV=1	A1JSK2_YERE8
LysR-family transcriptional regulatory protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3720 PE=4 SV=1	A1JR54_YERE8
Ni/Fe-hydrogenase 2 b-type cytochrome subunit OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=hybB PE=4 SV=1	A1JQK5_YERE8
PduX OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pduX PE=4 SV=1	A1JSN3_YERE8
Putative methyl-accepting chemotaxis protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=aer PE=4 SV=1	A1JLQ5_YERE8
Recombinase OS=Yersinia enterocolitica GN=ORF2 PE=4 SV=1	Q70W49_YEREN
Probable terminase, ATPase subunit OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gpP PE=4 SV=1	A1JNI4_YERE8
Protease OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=degQ PE=3 SV=1	A1JR98_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yacK PE=3 SV=1	A1JJN0_YERE8
Secretion Wt Injected Proteins Gentamicin	Accession Number
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
LcrV OS=Yersinia enterocolitica GN=lcrV PE=4 SV=1	O87495_YEREN
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Translocator protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yopD PE=4 SV=1	A1JU70_YERE8

Chaperone protein dnaK OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dnaK PE=2 SV=1	DNAK_YERE8
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Chaperone protein htpG OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=htpG PE=3 SV=1	HTPG_YERE8
YopN OS=Yersinia enterocolitica GN=yopN PE=4 SV=1	O68333_YEREN
Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
Elongation factor G OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fusA PE=3 SV=1	A1JS54_YERE8
Methionyl-tRNA synthetase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=metG PE=3 SV=1	SYM_YERE8
Secretion Wt Media Secreted Proteins 1	Accession Number
YopD OS=Yersinia enterocolitica GN=yopD PE=4 SV=1	C5IZG7_YEREN
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
Putative outer membrane porin A protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ompA PE=3 SV=1	A1JMT3_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
LcrV OS=Yersinia enterocolitica GN=lcrV PE=4 SV=1	O87495_YEREN
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2930 PE=3 SV=1	A1JRP4_YERE8
Protein tolB OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tolB PE=3 SV=1	TOLB_YERE8
Putative D-ribose-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rbsB PE=3 SV=1	A1JHS3_YERE8
Arginine-binding periplasmic protein 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=artI PE=3 SV=1	A1JM82_YERE8
Branched-chain amino acid-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=livK PE=3 SV=1	A1JIC6_YERE8
Secreted thiol:disulfide interchange protein DsbA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dsbA PE=3 SV=1	A1JHT2_YERE8
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3727 PE=3 SV=1	A1JR71_YERE8

Iron(III)-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yfuA PE=3 SV=1	A1JLH5_YERE8
YopB OS=Yersinia enterocolitica GN=yopB PE=4 SV=1	C5IZG6_YEREN
Putrescine-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=potF PE=3 SV=1	A1JM59_YERE8
Outer-membrane lipoprotein carrier protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=lolA PE=3 SV=1	LOLA_YERE8
Cationic 19 kDa outer membrane protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ompH PE=3 SV=1	A1JP72_YERE8
Peptidoglycan-associated lipoprotein Pal OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=pal PE=3 SV=1	A1JRK6_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2080 PE=3 SV=1	A1JN51_YERE8
Periplasmic dipeptide transport protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dppA PE=3 SV=1	A1JSV0_YERE8
Secretion Wt Media Secreted Proteins 2	Accession Number
YopD OS=Yersinia enterocolitica GN=yopD PE=4 SV=1	C5IZG7_YEREN
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
Beta-lactamase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ampC PE=4 SV=1	A1JRW9_YERE8
LcrV OS=Yersinia enterocolitica GN=lcrV PE=4 SV=1	O87495_YEREN
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
YopB OS=Yersinia enterocolitica GN=yopB PE=4 SV=1	C5IZG6_YEREN
Branched-chain amino acid-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=livK PE=3 SV=1	A1JIC6_YERE8
Putative type III secretion protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yscH PE=3 SV=1	A1JU99_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
YscP OS=Yersinia enterocolitica (type O:2.3) GN=yscP PE=4 SV=1	B6DX70_YEREN
Secreted thiol:disulfide interchange protein DsbA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dsbA PE=3 SV=1	A1JHT2_YERE8
Periplasmic dipeptide transport protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dppA PE=3 SV=1	A1JSV0_YERE8
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8

Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Maltose-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=malE PE=3 SV=1	A1JRV3_YERE8
Periplasmic oligopeptide-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=oppA PE=3 SV=1	A1JQ72_YERE8
Putative D-ribose-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rbsB PE=3 SV=1	A1JHS3_YERE8
Secretion Wt Media Secreted Proteins 3	Accession Number
YopD OS=Yersinia enterocolitica GN=yopD PE=4 SV=1	C5IZG7_YEREN
Yop effector YopH OS=Yersinia enterocolitica GN=yopH PE=4 SV=1	Q7BRY8_YEREN
Yop effector YopE OS=Yersinia enterocolitica GN=yopE PE=4 SV=1	Q7BRY7_YEREN
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
Beta-lactamase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ampC PE=4 SV=1	A1JRW9_YERE8
Global stress requirement protein GsrA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gsrA PE=3 SV=1	A1JJQ8_YERE8
YopB OS=Yersinia enterocolitica GN=yopB PE=4 SV=1	C5IZG6_YEREN
Putative outer membrane porin A protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ompA PE=3 SV=1	A1JMT3_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Putative type III secretion protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=yscH PE=3 SV=1	A1JU99_YERE8
Protease OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=degQ PE=3 SV=1	A1JR98_YERE8
50S ribosomal protein L9 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplI PE=3 SV=1	RL9_YERE8
3-oxoacyl-[acyl-carrier-protein] synthase I OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabB PE=3 SV=1	A1JKM0_YERE8
Outer membrane porin protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2463 PE=4 SV=1	A1JS30_YERE8
50S ribosomal protein L1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplA PE=3 SV=1	RL1_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
Secretion Mut Media Secreted Proteins 1	Accession Number

Branched-chain amino acid-binding protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=livK PE=3 SV=1	A1JIC6_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2166 PE=3 SV=1	A1JPB8_YERE8
Putative outer membrane porin A protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=ompA PE=3 SV=1	A1JMT3_YERE8
Protein tolB OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tolB PE=3 SV=1	TOLB_YERE8
Putative D-ribose-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rbsB PE=3 SV=1	A1JHS3_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE3039 PE=3 SV=1	A1JP43_YERE8
Secreted thiol:disulfide interchange protein DsbA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=dsbA PE=3 SV=1	A1JHT2_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=YE2930 PE=3 SV=1	A1JRP4_YERE8
Osmotically inducible protein Y OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=osmY PE=3 SV=1	A1JJ93_YERE8
Putrescine-binding periplasmic protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=potF PE=3 SV=1	A1JM59_YERE8
Secretion Mut Media Secreted Proteins 2	Accession Number
FliC OS=Yersinia enterocolitica GN=fliC PE=4 SV=1	C9EHF4_YEREN
Flagellar hook-associated protein 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=flgK PE=3 SV=1	A1JSZ7_YERE8
Flagellar hook-associated protein 2 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fliD PE=4 SV=1	A1JSR2_YERE8
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Global stress requirement protein GsrA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gsrA PE=3 SV=1	A1JJQ8_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
3-oxoacyl-[acyl-carrier protein] reductase OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabG PE=3 SV=1	A1JN73_YERE8
3-oxoacyl-[acyl-carrier-protein] synthase I OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=fabB PE=3 SV=1	A1JKM0_YERE8
50S ribosomal protein L9 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplI PE=3 SV=1	RL9_YERE8

Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8
Elongation factor Tu A (Fragment) OS=Yersinia enterocolitica GN=tufA PE=4 SV=1	B6RBB1_YEREN
50S ribosomal protein L10 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rplJ PE=3 SV=1	RL10_YERE8
FliC3 OS=Yersinia enterocolitica GN=fliC3 PE=4 SV=1	C9EHF3_YEREN
30S ribosomal protein S3 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=rpsC PE=3 SV=1	RS3_YERE8
Secretion Mut Media Secreted Proteins 3	Accession Number
60 kDa chaperonin OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=groL PE=3 SV=1	CH60_YERE8
Global stress requirement protein GsrA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gsrA PE=3 SV=1	A1JJQ8_YERE8
Putative exported protein OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=cpxP PE=3 SV=1	A1JHZ1_YERE8
Elongation factor Tu 1 OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=tuf1 PE=3 SV=1	EFTU1_YERE8
Glyceraldehyde 3-phosphate dehydrogenase A OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain 8081) GN=gapA PE=3 SV=1	A1JQH1_YERE8

References

(1) Cornelis, G. R. *Nat. Rev. Mol. Cell. Biol.* **2002**, *3*, 742.

(2) Cornelis, G. R. *Nat. Rev. Microbiol.* **2006**, *4*, 811.

(3) Schmidt, F.; Volker, U. *Proteomics* **2011**, *11*, 3203.

(4) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482.

(5) Dieterich, D. C.; Lee, J. J.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Nat. Protoc.* **2007**, *2*, 532.

(6) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A., 3rd; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15285.

(7) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Biol.* **2009**, *5*, 715.

(8) Grammel, M.; Zhang, M. M.; Hang, H. C. *Angew. Chem. Int. Ed.* **2010**, *49*, 5970.

(9) Deng, W.; de Hoog, C. L.; Yu, H. B.; Li, Y.; Croxen, M. A.; Thomas, N. A.; Puente, J. L.; Foster, L. J.; Finlay, B. B. *J. Biol. Chem.* **2010**, *285*, 6790.

(10) Rechavi, O.; Kalman, M.; Fang, Y.; Vernitsky, H.; Jacob-Hirsch, J.; Foster, L. J.; Kloog, Y.; Goldstein, I. *Nat. Methods* **2010**, *7*, 923.

(11) Kuehn, M. J.; Kesty, N. C. *Genes Dev.* **2005**, *19*, 2645.

(12) Cornelis, G. R. *Biol. Chem.* **2010**, *391*, 745.

(13) Lee, V. T.; Mazmanian, S. K.; Schneewind, O. *J. Bacteriol.* **2001**, *183*, 4970.

(14) DeBord, K. L.; Galanopoulos, N. S.; Schneewind, O. *J. Bacteriol.* **2003**, *185*, 3499.

(15) Snellings, N. J.; Popek, M.; Lindler, L. E. *Infect. Immun.* **2001**, *69*, 4627.

(16) Allaoui, A.; Woestyn, S.; Sluiters, C.; Cornelis, G. R. *J. Bacteriol.* **1994**, *176*, 4534.

(17) Riordan, K. E.; Schneewind, O. *Mol. Microbiol.* **2008**, *68*, 1485.

(18) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 9879.

(19) Blaylock, B.; Berube, B. J.; Schneewind, O. *Mol. Microbiol.* **2010**, *75*, 221.

(20) Aepfelbacher, M.; Trasak, C.; Ruckdeschel, K. *Thromb Haemost* **2007**, *98*, 521.

(21) Sorg, J. A.; Blaylock, B.; Schneewind, O. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16490.

(22) Anderson, D. M.; Ramamurthi, K. S.; Tam, C.; Schneewind, O. *J. Bacteriol.* **2002**, *184*, 1287.

(23) Hakansson, S.; Galyov, E. E.; Rosqvist, R.; Wolf-Watz, H. *Mol. Microbiol.* **1996**, *20*, 593.

(24) Schmidt, G. *Eur. J. Cell Biol.* **2011**, *90*, 955.

(25) Navarro, L.; Alto, N. M.; Dixon, J. E. *Curr. Opin. Microbiol.* **2005**, *8*, 21.

(26) Wang, M.; Weiss, M.; Simonovic, M.; Haertinger, G.; Schrimpf, S. P.; Hengartner, M. O.; von Mering, C. *Mol Cell Proteomics* **2012**, *11*, 492.

(27) Chromy, B. A.; Choi, M. W.; Murphy, G. A.; Gonzales, A. D.; Corzett, C. H.; Chang, B. C.; Fitch, J. P.; McCutchen-Maloney, S. L. *J. Bacteriol.* **2005**, *187*, 8172.

(28) Ponnusamy, D.; Hartson, S. D.; Clinkenbeard, K. D. *Vet. Microbiol.* **2011**, *150*, 146.

(29) Miller, V. L.; Bliska, J. B.; Falkow, S. *J. Bacteriol.* **1990**, *172*, 1062.

(30) Hellman, J.; Warren, H. S. *J. Endotoxin. Res.* **2001**, *7*, 69.

(31) Jeannin, P.; Bottazzi, B.; Sironi, M.; Doni, A.; Rusnati, M.; Presta, M.; Maina, V.; Magistrelli, G.; Haeuw, J. F.; Hoeffel, G.; Thieblemont, N.; Corvaia, N.; Garlanda, C.; Delneste, Y.; Mantovani, A. *Immunity* **2005**, *22*, 551.

(32) Mildiner-Earley, S.; Miller, V. L. *Infect. Immun.* **2006**, *74*, 4361.

(33) Steele-Mortimer, O.; Brumell, J. H.; Knodler, L. A.; Meresse, S.; Lopez, A.; Finlay, B. B. *Cell. Microbiol.* **2002**, *4*, 43.

(34) Francis, M. S.; Wolf-Watz, H. *Mol Microbiol* **1998**, *29*, 799.

(35) Akopyan, K.; Edgren, T.; Wang-Edgren, H.; Rosqvist, R.; Fahlgren, A.; Wolf-Watz, H.; Fallman, M. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1639.

(36) Wulff-Strobel, C. R.; Williams, A. W.; Straley, S. C. *Mol Microbiol* **2002**, *43*, 411.

(37) Sorg, J. A.; Miller, N. C.; Marketon, M. M.; Schneewind, O. *J Bacteriol* **2005**, *187*, 7090.

(38) Adkins, I.; Schulz, S.; Borgmann, S.; Autenrieth, I. B.; Grobner, S. *J. Med. Microbiol.* **2008**, *57*, 139.

(39) Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A. *Chembiochem* **2010**, *11*, 2092.

(40) Beatty, K. E.; Szychowski, J.; Fisk, J. D.; Tirrell, D. A. *Chembiochem* **2011**, *12*, 2137.

(41) Wisniewski, J. R.; Zougman, A.; Mann, M. *J. Proteome Res.* **2009**, *8*, 5674.

(42) Manza, L. L.; Stamer, S. L.; Ham, A. J.; Codreanu, S. G.; Liebler, D. C. *Proteomics* **2005**, *5*, 1742.

(43) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat. Biotechnol.* **2001**, *19*, 242.

(44) Mann, M. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 952.

(45) Howden, A. J.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nat Methods* **2013**, *10*, 343.

(46) Isberg, R. R.; O'Connor, T. J.; Heidtman, M. *Nat. Rev. Microbiol.* **2009**, *7*, 13.

(47) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. *ACS. Chem. Biol.* **2012**, *7*, 1326.

(48) Cornelis, G. *J. Gen. Microbiol.* **1975**, *91*, 391.

(49) Link, A. J.; Vink, M. K.; Tirrell, D. A. *Nat. Protoc.* **2007**, *2*, 1879.

(50) Szychowski, J.; Mahdavi, A.; Hodas, J. J.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *J. Am. Chem. Soc.* **2010**, *132*, 18351.

CHAPTER III

A Genetically Encoded AND Gate for Cell-Targeted Metabolic Labeling of Proteins

Abstract

We describe a genetic AND gate for cell-targeted metabolic labeling and proteomic analysis in complex cellular systems. The centerpiece of the AND gate is a bisected methionyl-tRNA synthetase (MetRS) that charges the Met surrogate azidonorleucine (Anl) to $tRNA^{Met}$. Cellular protein labeling occurs only upon activation of two different promoters that drive expression of the N- and C-terminal fragments of the bisected MetRS. Anl-labeled proteins can be tagged with fluorescent dyes or affinity reagents via either copper-catalyzed or strain-promoted azide-alkyne cycloaddition. Protein labeling is apparent within five minutes after addition of Anl to bacterial cells in which the AND gate has been activated. This method allows spatial and temporal control of proteomic labeling and identification of proteins made in specific cellular subpopulations. The approach is demonstrated by selective labeling of proteins in bacterial cells immobilized in the center of a laminar-flow microfluidic channel, where they are exposed to overlapping, opposed gradients of inducers of the N- and C-terminal MetRS fragments. The observed labeling profile is predicted accurately from the strengths of the individual input signals.

This chapter has been published as:

Mahdavi, A.; Segall-Shapiro, T. H.; Kou, S.; Jindal, G. A.; Hoff, K. G.; Liu, S.; Chitsaz, M.; Ismagilov, R. F.; Silberg, J. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2013**, *135*, 2979.

Introduction

The bio-orthogonal non-canonical amino acid tagging (BONCAT) method of proteomic analysis enables enrichment and identification of selected subsets of cellular proteins[1]. The initial demonstration of the BONCAT method used the methionine (Met) surrogate azidohomoalanine (Aha) as a metabolic label; addition of Aha to cells at selected time points provides temporal control of proteomic labeling. In contrast to Aha, which is incorporated into proteins in all cells through the action of the wild-type methionyl-tRNA synthetase (MetRS), the bulkier Met surrogate azidonorleucine (Anl) can be used to restrict labeling to cells that express a mutant MetRS (designated NLL-MetRS)[2]. Sensitivity to cell state (e.g., to oxidative stress) can be achieved by controlling expression of the mutant synthetase with carefully chosen promoters[3]. However, it is not yet possible to specify cell state more precisely in such experiments, and to control proteomic labeling by integrating multiple signals to activate the NLL-MetRS.

Here we introduce a new approach in which multiple promoters (P1 and P2) are used to regulate Anl incorporation into the E. coli proteome. The approach relies on construction of a split MetRS that generates a translational output only in cells in which the P1 and P2 inputs are simultaneously on. The split MetRS functions as an AND gate to control protein labeling (Figure 3.1).

Results and Discussion

We used protein fragment complementation to make a bisected MetRS[4], prompted by the work of Schimmel and coworkers who showed that the E. coli MetRS could be cut at six different sites to yield functional split variants[5]. To assess which of the peptide bonds in the MetRS backbone should be broken to create fragments that associate to form the most active enzymes, transposase mutagenesis was used to construct a library of vectors that express split MetRS[6] and active variants were selected via bacterial complementation. To avoid variants that require only a single fragment for activity, a gene fragment encoding MetRS residues 1-548 (designated MetRS henceforth) was used to construct the vector library (Figure 3.2a). The 1-548 variant is the smallest MetRS fragment that has been reported to exhibit near-native activity and stability[7]. E. coli strain CS50 was used for selecting functional variants on medium lacking Met[8]; growth was visible only when cells expressed an active MetRS (Figure 3.2b).

Selection by CS50-DE3 complementation yielded hundreds of colonies with varying complementation strengths (Figure S3.1). DNA sequencing of these clones identified MetRS variants with backbone cleavage sites lying within all four domains of the synthetase (Figure 3.2c, Table TS3.1). Variants with the greatest complementation strength were cut within the connective polypeptide and Rossmann domains. Mapping the locations of backbone cleavage onto the MetRS structure (Figure 3.2d) revealed that cleavage in these domains yields fragments that make large numbers of intermolecular residue-residue contacts in the assembled enzyme (Figure S3.2).

We have shown previously that two MetRS triple mutants, L13N/Y260/H301 (NLL) and L13P/Y260/H301 (PLL), charge Anl with good efficiency[9]. The corresponding mutations were introduced into the vectors encoding the bisected MetRS variants that exhibited the strongest complementation (those with cut sites at residues 48, 131, 183, 247, 272, 278, and 456) (Figure S3.3). *E. coli* CS50-DE3 cells transformed with each mutant vector were grown to mid-log phase in liquid culture, and labeling was initiated by addition of 1 mM Anl (**1**, Figure 3.2e). After 1 h, cells were lysed and total cellular proteins were treated with alkyne-tetramethylrhodamine (TAMRA) dye (**2**, Figure 3.2e) for labeling via copper-catalyzed azide-alkyne cycloaddition (**3**, Figure 3.2e)[10-12]. After SDS-PAGE, TAMRA-labeled proteins were detected by in-gel fluorescence imaging (Figure 3.2f). All fourteen of the NLL and PLL split MetRS variants enabled Anl incorporation into cellular proteins. On the basis of this result and the strong

complementation observed for wild-type MetRS(247) (Figure 3.2b), NLL-MetRS(247) was selected for further development.

To construct an AND gate for controlled labeling of cellular proteins with Anl, the coding sequences for the N- and C-terminal fragments of NLL-MetRS(247) were placed in a pair of vectors under control of P_{T5} and P_{BAD} promoters, respectively. As expected, NLL-MetRS(247) showed IPTG- and arabinose-dependent activity, with the highest levels of metabolic protein labeling observed in cultures supplemented with both inducers. To be useful as an AND gate, the system described here should be active only when both IPTG and arabinose are present – not in the presence of a single inducer. However, we observed some activity even in the absence of IPTG, presumably as a consequence of leaky expression of the N-terminal fragment (Figure S3.4). This activity was readily eliminated by catabolite repression (Figure S3.5)[13]; in glucose-supplemented media, Anl labeling was observed only upon addition of both IPTG and arabinose (Figure 3.3a-c).

To investigate how the output of the AND gate was controlled by variation in input levels, we used varying combinations of arabinose and IPTG concentrations to induce MetRS fragment expression and Anl incorporation. We found concentration-dependent expression of the N- and C-terminal fragments (Figure S3.6) and steep transitions between the on and off states (Figure 3.3d, and Figure S3.6). The time-dependence of Anl-labeling under control of the activated AND gate was monitored by detection of labeled proteins with an azadibenzocyclooctyne-TAMRA dye; labeling was apparent within 5 min after addition of Anl (Figure 3.3e). The rate of AND-gate activation upon induction by IPTG and arabinose was also measured (Figure S3.7). The rate of activation will depend on the particular promoters used to drive fragment expression.

Spatially controlled proteomic labeling of cells within complex multi-cellular systems is useful for study of a variety of important biological phenomena including biofilm formation and developmental patterning[14,15]. The MetRS AND gate provides a useful solution to such challenges because it can be activated by spatially regulated signals. As a step toward this goal, we sought to control Anl labeling in cells that are subjected to spatial gradients of activators. A laminar-flow microfluidic channel was used to expose surface-immobilized bacterial cells harboring the AND gate to opposing gradients of IPTG and arabinose. Anl incorporation should occur only in cells that encounter sufficient levels of both inducers (Figure 3.4a). The diffusion of IPTG and arabinose in the microfluidic channel was modeled (Figure S3.8) and diffusion analysis results were used to estimate inducer profiles across the width of the channel (Figure 3.4b).

Dye diffusion experiments enabled visualization of similar gradients during flow, and confirmed the modeling results (Figure S3.9). To estimate the strengths of the individual AND-gate inputs, E. coli DH10B cells were transformed with constructs that expressed GFP under control of the same PT5 and PBAD promoters that were used for expression of the N- and C-terminal fragments of the AND gate. GFP expression revealed the activation profile of each promoter as a function of position across the width of the channel (Figure 3.4c-e).

Cells harboring the AND gate were immobilized in the microfluidic channel, exposed to the inducer gradients and Anl for 4 h, then fixed and treated with an alkyne-functionalized AlexaFluor 488 dye. Fluorescence imaging showed that Anl incorporation was confined to cells in the center of the channel (Figure 3.4e).

To confirm that the observed labeling profile is predictable from the strengths of the input signals, we used the activation profile of each promoter as reported by GFP expression. As expected, either for a bimolecular binding equilibrium or for a Boolean AND operation, the scalar product of the inputs closely matched the output signal as measured by Anl incorporation (Figure 3.4e). This result suggests that the protein labeling profile can be adjusted, by tuning the input promoter strengths, to be more- or less-sharply dependent on the gradients of local activators or morphogens. Conversely, using endogenous promoters to drive the AND gate should provide labeling outputs that closely match spatial patterns of promoter activation.

Conclusions

This work introduces a new AND gate based on protein fragment complementation[4]. The AND gate can be used to restrict protein labeling to subsets of cells in which specified promoters or genes of interest are active. Because the AND gate is genetically encoded, it should be easily integrated with previously reported transcription-based logic gates (e.g., AND, OR, NOR, and XOR gates)[13,16,17] to create more complex genetic circuits with protein labeling as the output. As shown previously[1,18,19], azide-labeled proteins can be affinity-tagged and purified for analysis by mass spectrometry. Thus the AND gate provides a powerful new tool for targeted analysis of cellular protein synthesis.

Materials and Methods

Creating a Library of Vectors that Express Fragmented MetRS

A *metG* fragment encoding MetRS residues 1-548 was cloned into the NcoI and SalI sites of pPROEX1 (Life Technologies) to yield pPROEX-MetRS, a vector that uses a leaky Trc promoter for expression. To make this vector compatible with library construction [6], it was digested (XbaI and SpeI) and ligated to create pPROEX-MetRSΔNotI-XS1, which lacks a NotI site. This vector was used for library construction. A NotI-flanked artificial transposon encoding chloramphenicol resistance (M1-CamR, Finnzymes) was inserted into pPROEX-MetRSΔNotI-XS1 using HyperMu, a mutant MuA transposase with increased activity (Epicentre). Reaction mixtures (20 µL) containing HyperMu buffer, 300 ng of pPROEX-MetRSΔNotI-XS1, 100 ng M1-CamR, and 1 U of HyperMu MuA transposase were incubated at 37°C for 16 h. Reactions were terminated by adding 2 µL of HyperMu 10x Stop Solution, gently mixing, and incubating each reaction at 70°C for 10 min. Total DNA was purified (Zymo Research) and electroporated into *E. coli* MegaX DH10B. Colonies (~180,000) appearing after 24 h on selective medium (20 µg/mL chloramphenicol) were harvested from plates and pooled, and total plasmid DNA was purified using a Qiagen Miniprep Kit to obtain a library of vectors encoding MetRS genes with a transposon (M1-CamR) inserted at different locations. This library was digested with restriction enzymes (BspHI, ClaI, and NdeI) that cut at two sites within *metG* (base pairs 73-78 and 1603-8) and five sites within the pPROEX vector. MetRS genes (base pairs 73-1608) containing M1-CamR were separated from other DNA fragments by agarose gel electrophoresis, purified, and cloned back into pPROEX-MetRSΔNotI-XS1. These vectors containing size-selected MetRS genes were electroporated into *E. coli* MegaX DH10B, colonies (~118,000) appearing on LB-agar plates containing 20 µg/mL chloramphenicol were harvested, and total DNA was purified to obtain a size-selected library. The M1-CamR was removed from the size-selected library by NotI digestion and replaced with *f1-kan*^{R}, a DNA insert that terminates translation preceding the insert and initiates transcription and translation following the insert [6]. The resulting two-promoter library was electroporated into *E. coli* MegaX DH10B, plated onto LB-agar medium containing 25 µg/mL kanamycin and incubated for 24 h at 37°C. Colonies (~248,000) were harvested, and total DNA was purified to obtain the two-promoter library used for functional selections.

An *E. coli* Selection for Functional MetRS

The λDE3 Lysogenization Kit (Novagen) was used to generate DE3 lysogens of *E. coli* CS50 (Yale Stock Center), a strain that grows slowly in the absence of methionine because of a MetRS

mutation that increases the K_M for Met [8]. To identify a CS50-DE3 lysogen that constitutively expresses T7 RNA polymerase, cells were transformed with pET-Venus-Grx2[20] and their fluorescence was measured using a Tecan M1000 plate reader. An *E. coli* CS50-DE3 strain from this screen was electroporated with the final two-promoter library, plated on M9 minimal media containing all amino acids except methionine, and grown for 24 hours at 37°C. Vectors from one hundred colonies were picked and used to inoculate LB cultures containing 50 µg/mL ampicillin. DNA purified from each overnight culture was sequenced.

Evaluation of MetRS Complementation Strength

After sequencing, complementing vectors were transformed into *E. coli* CS50-DE3, plated onto LB-agar plates containing 50 µg/mL ampicillin, and incubated at 37°C for 24 h to obtain colonies. Overnight LB cultures from three separate colonies of each variant were diluted to OD_{600} = 2.0, 0.2, 0.02, and 0.002, and 10 µL of each was spotted onto +Met and –Met M9-agar plates. Prior to spotting, cells were harvested by centrifugation, washed three times with 25% glycerol, and resuspended in 25% glycerol to OD_{600} = 2. Growth was visually assessed after 24 h, when cells harboring pPROEX showed no visible growth and cells containing pPROEX-MetRSΔNotI-XS1 grew at all titers. Cell growth was scored on a scale of 1 to 4, based on the number of dilutions where growth was visible. MetRS-NLL did not complement *E. coli* CS50-DE3 growth after 24 h, and thus could not be used as a parent for laboratory evolution experiments.

Calculation of Residue-Residue Contacts in Each Split Variant

The number of intermolecular contacts made between fragments of all possible split MetRS variants was calculated as the number of residue pairs with atoms within a distance cutoff of 4.5 Å. Protein Data Bank crystal structure 1f4I was used for this analysis [21].

Introduction of Active Site Mutations into Selected MetRS Vectors

Site-directed mutagenesis (Stratagene) was used to mutate MetRS residues L13, Y260 and H301 to PLL (or NLL) in seven complementing vectors selected from the two-promoter library (pPROEX-MetRSΔNotI-XS1 derived), including variants 48, 131, 183, 247, 272, 278 and 456 (see Fig. S3.3). Split proteins with mutations are designated NLL-MetRS(cut site) and PLL-MetRS(cut site), respectively.

Vectors for Regulated Protein Expression

A fragment of the MetRS gene encoding NLL-MetRS residues 1-247 was cloned into pQE80L-Kan between the BamH1 and Sal1 restriction sites to create pQE80His6_N247. This vector drives expression of the N-terminal MetRS fragment with an N-terminal His_6-tag under control of an IPTG-inducible T5 promoter. A fragment of the MetRS gene encoding NLL-MetRS residues 247-548 was cloned into pBAD33 between the Sac1 and Kpn1 restriction sites to create pBAD33HA_C247. This vector drives expression of the C-terminal MetRS fragment with a C-terminal hemagglutinin tag (YPYDVPDYA) under control of an arabinose-inducible and glucose-repressible P_{BAD} promoter. Translational initiation on pBAD33-derived mRNAs was accomplished by introducing before the ATG start codon a DNA sequence (AGGAGGAATTCACC) that encodes a ribosome-binding site. Vectors for assessing promoter activities in gradients of inducers were generated by cloning GFP into pBAD33 and pQE80L-Kan to create pQE80L-GFP and pBAD33-GFP, respectively.

Synthesis of Azidonorleucine

Azidonorleucine synthesis was based on a previous protocol for synthesis of azidohomoalanine, using Boc-lysine as the starting material [22]. Briefly, 5.27 g (81.1 mmol) of sodium azide was treated with 2.7 mL (16 mmol) of distilled triflic anhydride in 13 mL of water for 2 h. The triflic azide product was extracted with 10 mL methylene chloride and added dropwise to a flask containing Boc-Lys-OH (2 g, 8.1 mmol), K_2CO_3 (1.68 g, 12.2 mmol) and $CuSO_4$ (20 mg, 0.08 mmol) in 26 mL of water and 250 mL of methanol. After 20 h at room temperature the product was extracted with ethyl acetate, redissolved in methylene chloride and purified by silica gel chromatography. After Boc deprotection with hydrochloric acid, the final product was purified by cation exchange chromatography.

Metabolic Incorporation of Anl into Cells Grown in Liquid Culture

To test metabolic incorporation of Anl into proteins in *E. coli* cells expressing split MetRS variants, overnight LB cultures were used to inoculate cultures containing M9 minimal media containing 19 amino acids (100 μM), 100 μM methionine, and 1% glycerol. Cells were grown in minimal media to an OD_{600} = 0.25, Anl was added at a final concentration of 1 mM, and cells were grown for 1 h to allow for metabolic incorporation of Anl into newly synthesized proteins. For experiments involving cells transformed with pairs of inducible vectors (*e.g.*, pQE80His6_N247 and pBAD33HA_C247), cells were grown to an OD_{600} = 0.25, arabinose and IPTG inducers were added at varying concentrations, cells were grown for 2 h, Anl was added at a final concentration of 1 mM, and cells were grown for 1 h to allow for metabolic incorporation

of Anl into newly synthesized proteins. For AND-gate expression, based on the two vector system (pQE80His6_N247, pBAD33HA_C247), 50 μg/ml of kanamycin and 100 μg/ml of chloramphenicol were used. Cells were harvested by centrifugation at 10,000 rcf for 5 minutes and lysed with 4% SDS (10 μL for cell pellets resulting from 1 mL of culture medium) prior to labeling of the lysate with alkyne-TAMRA dye (structure **2**, Fig. 3.2E).

Labeling Anl-tagged Proteins in Cell Lysates with an Alkyne-TAMRA Dye

Cells were lysed with 4% SDS in phosphate buffered saline (PBS). Ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor (Roche) was added to the lysates to reduce protease activity. The EDTA-free version of the protease inhibitor was required for copper-catalyzed cycloaddition reactions because EDTA chelates copper ions and interferes with copper-dependent reactions. PBS was added to dilute the SDS concentration to 1%, and cell lysates were centrifuged at 14,000 rcf for 10 min to remove cellular debris. Protein concentrations were measured by using a bicinchoninic protein quantification kit (BCA assay; Pierce, Rockford, IL). The same amount of protein was used for each condition; concentrations ranged from 0.1 to 0.4 mg/mL. Copper-catalyzed reactions were performed at room temperature for 2 h in 1.5 mL centrifuge tubes, according to the following protocol: (i) alkyne-TAMRA dye was added to protein solution to a final concentration of 100 μM; (ii) tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) and copper sulfate were pre-incubated together for 1 min and then added to the protein solution to final concentrations of 0.1 mM copper sulfate and 0.5 mM THPTA; (iii) aminoguanidine was added followed by sodium ascorbate (made fresh in water), and (iv) the solution was mixed once and protected from light with no further mixing. The final concentrations of aminoguanidine and sodium ascorbate were 5 mM. The THPTA ligand was synthesized according to methods published by Finn and coworkers [12]. Proteins were precipitated with chloroform/methanol, washed with methanol to remove unreacted dye and resuspended in protein loading buffer containing 2% SDS and 10% mM 2-mercaptoethanol. Proteins were electrophoresed using 12% Bis-Tris polyacrylamide gels (Invitrogen). TAMRA ($\lambda_{excitation}$ = 555 nm and $\lambda_{emission}$ = 580 nm) was excited at 532 nm and detected with a 580 band-pass 30 nm filter. In-gel fluorescence images were acquired on a Typhoon 9400 instrument (GE Healthcare).

Copper-free Protein Labeling with Dibenzocyclooctyne-TAMRA

Cells were lysed with reducing buffer containing 5 mM dithiothreitol, 4% SDS in PBS and incubated for 30 min at 50ºC. The SDS concentration was then reduced to 1% by addition of PBS. The lysates were subsequently incubated in the dark for 30 min with 100 mM

iodoacetamide. The reduction and alkylation steps were used to reduce background labeling due to reaction with free thiols. After reduction and alkylation, the proteins were reacted with 20 μM azadibenzocyclooctyne-TAMRA dye conjugate (Click Chemistry Tools, AZ) for 15 min at room temperature, precipitated with chloroform/methanol and washed with methanol to remove unreacted dye.

Western Blot Detection of Split-MetRS Fragment Expression

N-terminal fragments expressed from pQE80His6_N247 contained N-terminal His_6 tags. These fragments were detected by western blot analysis using an AlexaFluor 488 conjugated anti-penta-His antibody (Qiagen). AlexaFluor 488 ($\lambda_{excitation}$ = 495 nm and $\lambda_{emission}$ = 520 nm) was excited at 488 nm and detected with a 520 band-pass 40 nm filter. C-terminal fragments expressed from pBAD33HA_C247 contained C-terminal hemagglutinin (HA) tags. HA-tagged proteins were detected by western blot analysis with AlexaFluor 633 conjugated anti-HA antibody (Santa Cruz). AlexaFluor 633 ($\lambda_{excitation}$ = 632 nm and $\lambda_{emission}$ = 647 nm) was excited at 633 nm and detected with a 670 band pass 30 nm filter. All images were obtained on a Typhoon 9400 instrument (GE Healthcare).

Fabrication of the Microfluidic Laminar Flow Device

The microfluidic device was fabricated using rapid prototyping in polydimethylsiloxane (PDMS) [23]. The channel geometry is shown in Fig. S3.8. Inlet channel widths were 850 μm; the main channel width was 870 μm. The channel depth was 175 μm. The PDMS mold was plasma oxidized and attached to a 0.17 mm thick glass slide. The attached PDMS mold and glass slide were heat cured at 110°C for 10 min. To render the glass surface and channel surface cell-adhesive, the device was filled with 1 mg/mL Poly-D-lysine (Millipore, MA) in PBS and then dried for 12 h at 37°C [24].

Modeling Diffusion in the Microfluidic Channel

The channel geometry was constructed in AutoCad 2012 (Autodesk Student version) software and diffusion simulations were performed using COMSOL® Multiphysics v4.2a with the Laminar Flow module and Transport of Diluted Species module. A mesh with a maximum element size of 25 μm was used for this analysis. For these simulations, we used a flow rate of 1 μL/min and inlet concentrations of arabinose and IPTG of 5 mM and 1 mM, respectively. The diffusion coefficients of arabinose and IPTG were estimated using the Stokes-Einstein equation according to the diffusion coefficient of glucose in water at 37 ºC, which is 9.59×10^{-10} $m^2 \cdot s^{-1}$ [25].

The estimated values we used here were 1.02×10^{-9} $m^2\cdot s^{-1}$ for arabinose and 8.74×10^{-10} $m^2\cdot s^{-1}$ for IPTG. These values are in agreement with previously reported diffusion coefficients of arabinose [26] and IPTG [27] at 37 ^{0}C. See Fig. S3.8-S3.9 for further details.

Imaging Dye Diffusion in the Microfluidic Channel

Dye diffusion experiments were performed to measure dye profiles in the channel and compare the measured and predicted profiles. One inlet contained a solution of AlexaFluor488 dye at a concentration of 2.82 μM; the second inlet contained PBS. Fluorescence images were obtained with a Leica DMI6000 microscope (Leica, CT) with a 10× (0.40 N.A.) objective, which was coupled to a cooled-CCD camera with 12-bit, 1344×1024 resolution (Hamamatsu Photonics, UK) and a 0.6× coupler. The filter cube (Leica) had excitation range: BP 480/40; dichromatic mirror: 505; suppression filter: BP527/30. Images were obtained at the solid-liquid interface on the glass surface of the channel. Imaging the dye diffusion at the glass-liquid interface was done as a surrogate for the PDMS-liquid interface on which the cells are seeded, to reduce effects from bulk fluorescence of the dye.

Cell Seeding and Induction in the Microfluidic Channel

Overnight LB cultures were inoculated into M9 minimal media containing 19 amino acids (100 μM), 100 μM Met, and 1% glycerol, and incubated at 37^{0}C until they reached an OD_{600} = 0.4. The cells were seeded on the PDMS surface of the microfluidic channel for 2 h at 30°C with no fluid flow. The micro-channel was flushed with M9 minimal medium containing 0.5 mg/mL glucose. Cells were incubated in this medium at 37°C for 30 min prior to induction, with no flow. M9 minimal media containing 1 mM Anl and one of the two inducers (1 mM IPTG or 5 mM arabinose) were loaded into two different 250 μL gas-tight Hamilton syringes (Hamilton Company, MA), and the two converging laminar streams were pumped through the device. Flow rates were controlled using syringe pumps (Harvard Apparatus, MA); a flow rate of 1 μL/min was used for each stream. Inductions were performed for 4 h. After induction, inducer solutions were flushed from the channel with PBS, and cells were fixed with 3.7% formaldehyde at 37°C for 15 min and washed for 10 min with PBS. Cells expressing GFP were imaged in PBS, whereas cells expressing the split MetRS were permeabilized with methanol for 3 min, washed with PBS for 10 min, and treated with alkyne-AlexaFluor 488 dye prior to imaging (under the same conditions described above for labeling of lysates with alkyne-TAMRA dye). The dye was washed out of the channel for 1 h with PBS (2 μL/min) before fluorescence imaging.

Fluorescence Imaging of Cells in the Microfluidic Channel

Fluorescence and bright field images were obtained with a Leica DMI6000 microscope (Leica, CT). We used a 10× (0.40 N.A.) objective, which was coupled with a cooled CCD camera with 12-bit, 1344×1024 resolution (Hamamatsu Photonics, UK) with a 0.6× coupler. MetaMorph Imaging software (version 7.7) was used for image acquisition. The filter cube (Leica) used for GFP was L5 (Excitation range: BP 480/40; Dichromatic mirror: 505; Suppression filter: BP527/30), the same filter set was used for imaging alkyne-AlexaFluor 488 dye.

Quantification of Fluorescence in the Microfluidic Channel

Fluorescence images of the cells in the microfluidic channel were analyzed using ImageJ software. For quantification of fluorescence, the width of the channel was divided into 17 equally spaced boxes (as shown in Fig. 3.4*C*), starting 50 μm from each wall to avoid edge effects. Each box was 45 μm in height and 420 μm in length; a picture of one box located 50 μm from the edge is shown in Fig. 3.4*C*. The mean fluorescence intensity in each box was measured. This measurement was performed at three adjacent locations, with the upstream edge of the first box located 2 mm downstream from the y-junction of the channel, and adjacent boxes placed immediately downstream. Fluorescence measurements were used to calculate the mean and standard deviation at each position across the width of the channel. Fluorescence intensities are reported relative to the highest value.

Fluorescence Confocal Images of Labeled Proteins in Cells

Fluorescence confocal images were not obtained directly from the microfluidic channel due to the thickness of the glass slide. Overnight LB cultures were inoculated using the same method used for seeding in the microfluidic channel. Cells were seeded into a 6-well chamber slide (Lab-Tek, Thermo Fischer Scientific), and induction was performed according to the method used for the microfluidic channel. After induction, cells were fixed with 3.7% formaldehyde at 37°C for 15 min and washed three times with PBS. Cells were permeabilized with methanol for 3 min and washed three times with PBS. The copper-catalyzed dye-labeling reaction was performed (according to the conditions described for labeling of lysates with alkyne-TAMRA dye above) using alkyne-AlexaFluor 488 dye. The reaction volume was 200 μL per well. After dye-labeling for 2 h, the cells were washed with five sequential PBS washes of 300 μL each, over a period of 2 h. The chamber slide was removed from the glass slide, and the cells were covered with 10% glycerol in PBS under a coverslip. Imaging of AlexaFluor 488 fluorescence was performed on a

Zeiss LSM 510 microscope using excitation and emission wavelengths of 495 nm and 519 nm, respectively.

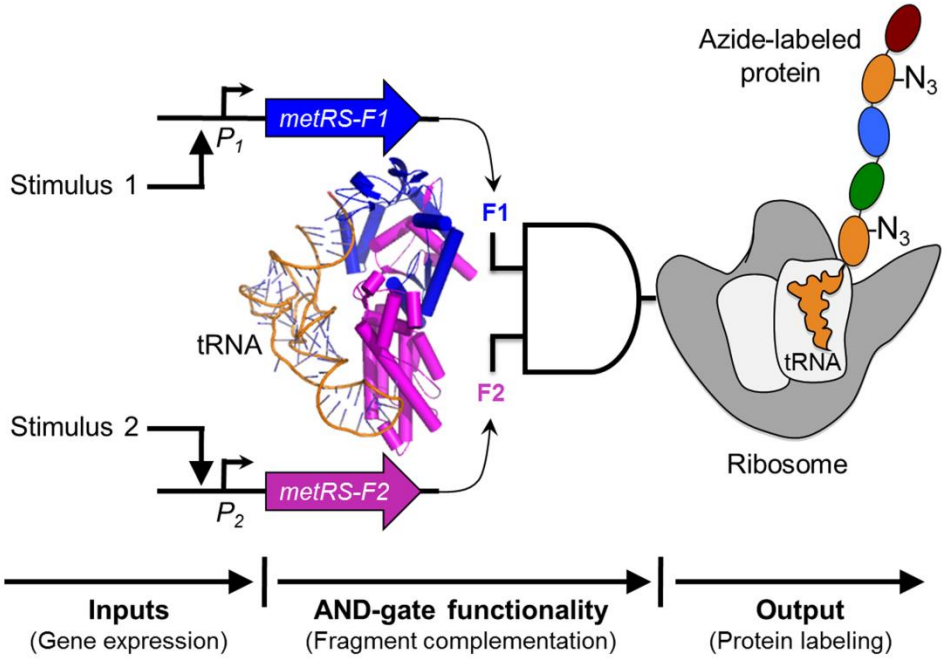


Figure 3.1. A genetic AND gate for proteomic labeling. Two promoters (P1 and P2) are used as inputs to drive expression of E. coli MetRS fragments F1 and F2. tRNAMet is charged with Anl only when the MetRS fragments are both expressed, and associate to form a functional enzyme. Ribosomes use Anl-charged tRNAMet during protein synthesis in competition with Met-charged tRNAMet. Newly synthesized proteins are labeled with azide groups, which can be selectively tagged by the azide-alkyne cycloaddition[10]. Diverse promoter inputs can be plugged into the AND gate to control proteomic labeling with Anl. The MetRS image was created using PDB 2CSX[28].

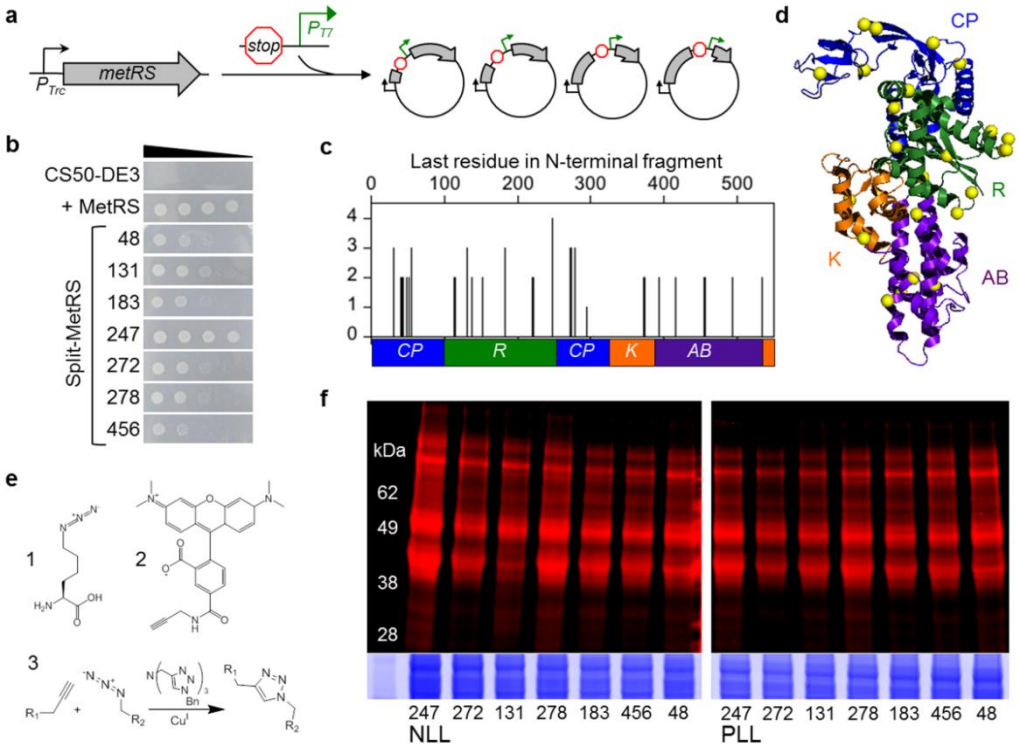


Figure 3.2. Identifying fragmented MetRS variants that metabolically label proteins with Anl. (a) DNA that contains a terminator, promoter (P_{T7}), and ribosomal binding site was randomly inserted into the MetRS gene to create a library of vectors that express two-piece MetRS variants[6]. (b) Growth of *E. coli* CS50-DE3 cells in the absence of Met is compared for cells that express a functional MetRS truncation comprised of residues 1-548 and split MetRS variants. Ten-fold serial dilutions of cells were spotted on M9-agar plates and growth was imaged after 24 h. (c) Complementation strengths of fragmented MetRS variants, which were named based on the last residue encoded by their N-terminal fragments. (d) Color coding of the connective polypeptide [CP, blue], Rossmann fold [R, green], KMSKS [K, orange], and anticodon-binding [AB, purple] domains maps the locations of backbone fission (yellow spheres) in functional variants onto the MetRS structure[29]. (e) Structure of Anl (1), which can be detected in proteins after reaction with alkyne-functionalized tetramethylrhodamine (TAMRA) dye (2). Copper-catalyzed azide-alkyne cycloaddition (3) results in a stable triazole linkage. (f) Cells expressing the split NLL-MetRS and PLL-MetRS from the vectors selected using bacterial complementation were grown in the presence of Anl, and cell lysates were treated with TAMRA-alkyne dye. In-gel fluorescence image shows TAMRA labeling, which indicates Anl incorporation into cellular proteins. Colloidal blue staining (bottom) was performed to compare protein loading across lanes.

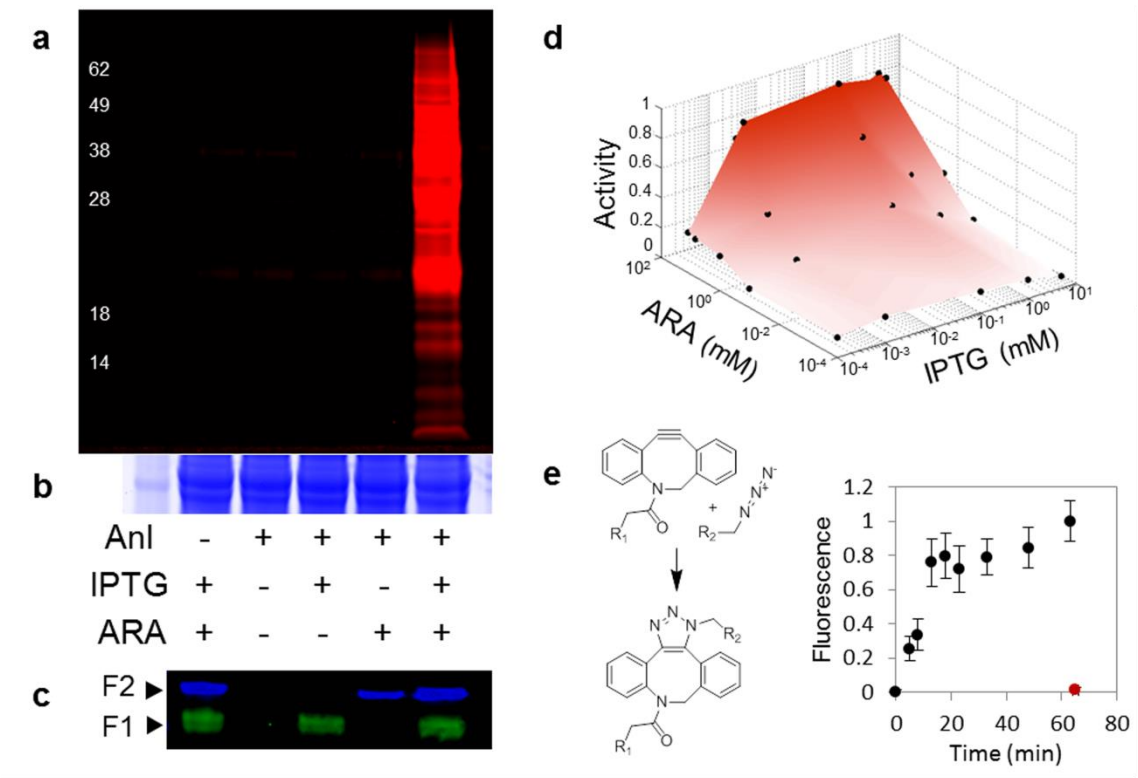


Figure 3.3. Genetic control of Anl incorporation into proteins in cells that express the N- and C-terminal fragments of NLL-MetRS(247) under control of IPTG- and arabinose-inducible promoters. (a) Anl labeling of *E. coli* proteins in the presence or absence of IPTG (1 mM), arabinose (5 mM), and Anl (1 mM). Cell lysates were treated with an alkyne-TAMRA dye, and in-gel TAMRA fluorescence was imaged to detect Anl incorporation. (b) Colloidal blue staining of the gel in (a) confirms similar total protein in each sample. (c) Western blot detection of the N-terminal (green) fragment containing a His_6 tag and the C-terminal (blue) fragment containing a hemagglutinin tag. (d) Effects of arabinose and IPTG levels on Anl incorporation. Cell lysates were treated with alkyne-TAMRA, and the in-gel fluorescence signal from TAMRA-labeled proteins was used to determine relative Anl incorporation levels. Dots on the surface show conditions under which Anl incorporation was measured. (e) Time-dependence of protein labeling upon addition of Anl to cells harboring the activated AND gate. IPTG (1 mM) and arabinose (5 mM) were added to activate the AND gate; Anl (1 mM) was added 3 h later (at 0 min on the time axis). Anl-labeled proteins were detected by treatment with azadibenzocyclooctyne-TAMRA dye (R1); labeling was quantified by in-gel fluorescence detection. When the AND gate was not activated but Anl was present, no measurable incorporation was observed even after 1 h, as shown by the red symbol.

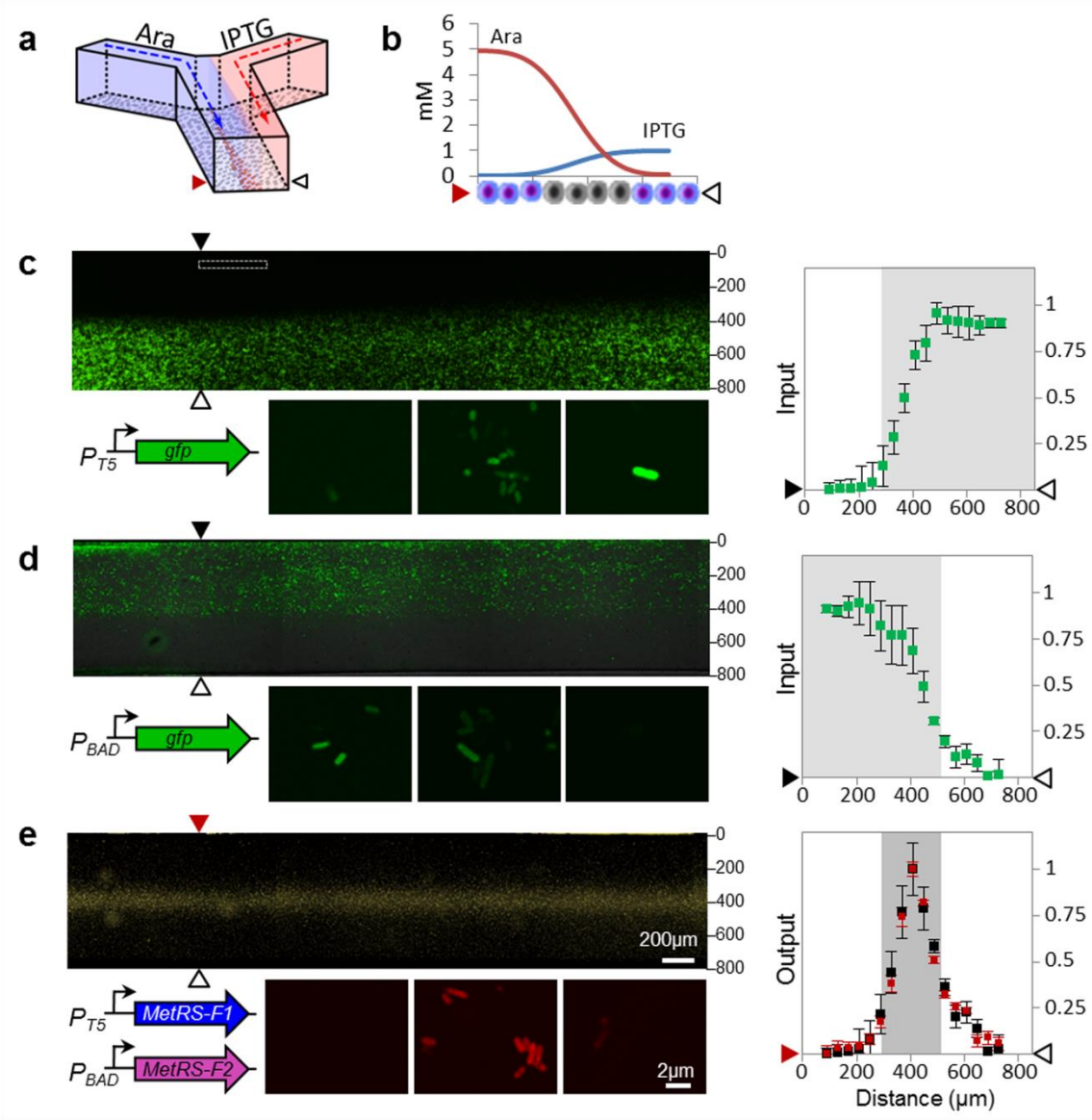


Figure 3.4. Spatial control of protein labeling. (a) A laminar flow microfluidic channel was used to expose surface-immobilized cells to gradients of arabinose and IPTG. In this device, the AND gate should be activated only in the center of the channel, where cells are exposed to both IPTG and arabinose. (b) Modeling of inducer diffusion was used to determine cross-channel concentration profiles of arabinose and IPTG at the locations shown by the arrows in (c, d, and e). Schematic shows that cells in the middle of the channel should incorporate Anl (gray); those on the periphery should not be labeled (purple). (c) Fluorescence image of the microfluidic channel (top) seeded with cells expressing GFP under IPTG control, and images of individual cells (bottom) that were incubated with varying combinations of inducers (left to right; 5 mM arabinose, 1 mM arabinose plus 0.1 mM IPTG, and 1 mM IPTG). GFP fluorescence was

quantified as a function of distance from the wall of the channel at positions indicted by the arrows, and is plotted to the right. Fluorescence values were normalized by setting the maximum fluorescence to 1. (d) Fluorescence detection of arabinose-regulated GFP expression under conditions identical to (c); the corresponding normalized fluorescence values are plotted to the right. (e) Detection of Anl incorporation by labeling with TAMRA-alkyne dye. Conditions are identical to those used for (c). The plot to the right shows the TAMRA fluorescence intensity (red) as a function of distance from the wall of the channel. TAMRA fluorescence values represent the experimental output of the AND gate, and were normalized by setting the maximum fluorescence to 1. Black dots represent the products of the GFP fluorescence signals from plots in (c) and (d), which were point-wise multiplied to provide a measure of the predicted output of the AND gate. The products were normalized by setting the maximum signal to 1 in order to allow direct comparison to the TAMRA fluorescence values.

Table TS3.1

Functional MetRS variants discovered using *E. coli* CS50-DE3 complementation. For each variant, we indicate: (i) the MetRS residues produced by the first open reading frame, ORF1, (ii), the relative orientations of the promoters preceding and within *metG*[6], (iii) if open reading frame following the internal promoter is in frame, (iv) the MetRS residues predicted to be produced by ORF2 following the internal promoter, (v) the relative complementation strength of each variant scored on a scale of 1-4, and (vi) the number of occurrences of each unique variant within the sequenced clones. In cases where ORF2 is out of frame, the MetRS residues present in the second fragment are unknown and the initial residue is designated alternative (alt) start codon. In the clone where the promoters are anti-parallel, ORF2 is not translated or transcribed.

ORF1 (residues)	Promoter orientations	ORF2 in frame	ORF2 (residues)	Relative Activity	Number clones
1-31	parallel	no	alt-548	3	1
1-40	parallel	no	alt-548	2	1
1-43	parallel	yes	43-548	2	1
1-48	parallel	yes	48-548	2	1
1-52	parallel	no	alt-548	2	1
1-55	parallel	yes	55-548	3	1
1-114	parallel	no	alt-548	2	1
1-131	parallel	yes	131-548	3	1
1-138	parallel	no	alt-548	2	1
1-152	parallel	no	alt-548	2	1
1-183	parallel	no	alt-548	2	1
1-183	parallel	yes	183-548	3	2
1-221	parallel	no	alt-548	2	1
1-247	parallel	yes	247-548	4	46
1-272	parallel	yes	272-548	3	1
1-273	parallel	no	alt-548	1	2
1-274	parallel	yes	274-548	3	1
1-278	parallel	yes	278-548	3	2
1-294	parallel	yes	294-548	1	1
1-373	parallel	no	alt-548	2	1
1-393	parallel	yes	393-548	2	1
1-416	parallel	no	alt-548	2	1
1-454	parallel	no	alt-548	2	1
1-456	parallel	yes	456-548	2	1
1-494	parallel	yes	494-548	2	1
1-534	parallel	no	alt-548	2	1
1-538	anti-parallel	---	---	3	2

Figure S3.1.

Complementation of *E. coli* CS50-DE3 by selected MetRS variants. Growth after 24 hours at 37°C of *E. coli* CS50-DE3 transformed with: i) pPRO-EX, ii) pPROEX-MetRSΔNotI-XS1, and pPRO-EX vectors that express fragmented MetRS variants discovered in selections. After sequencing, each vector was transformed into CS50-DE3 to obtain individual colonies on LB-agar plates, and the complementation of three separate colonies was analyzed on minimal medium containing or lacking methionine. Serial dilutions (1x, 10x, 100x, and 1000x) of the resuspended cells (10 µL each) were spotted in triplicate onto two M9-agar plates. NLL-MetRS did not complement *E. coli* CS50 in the absence of methionine.

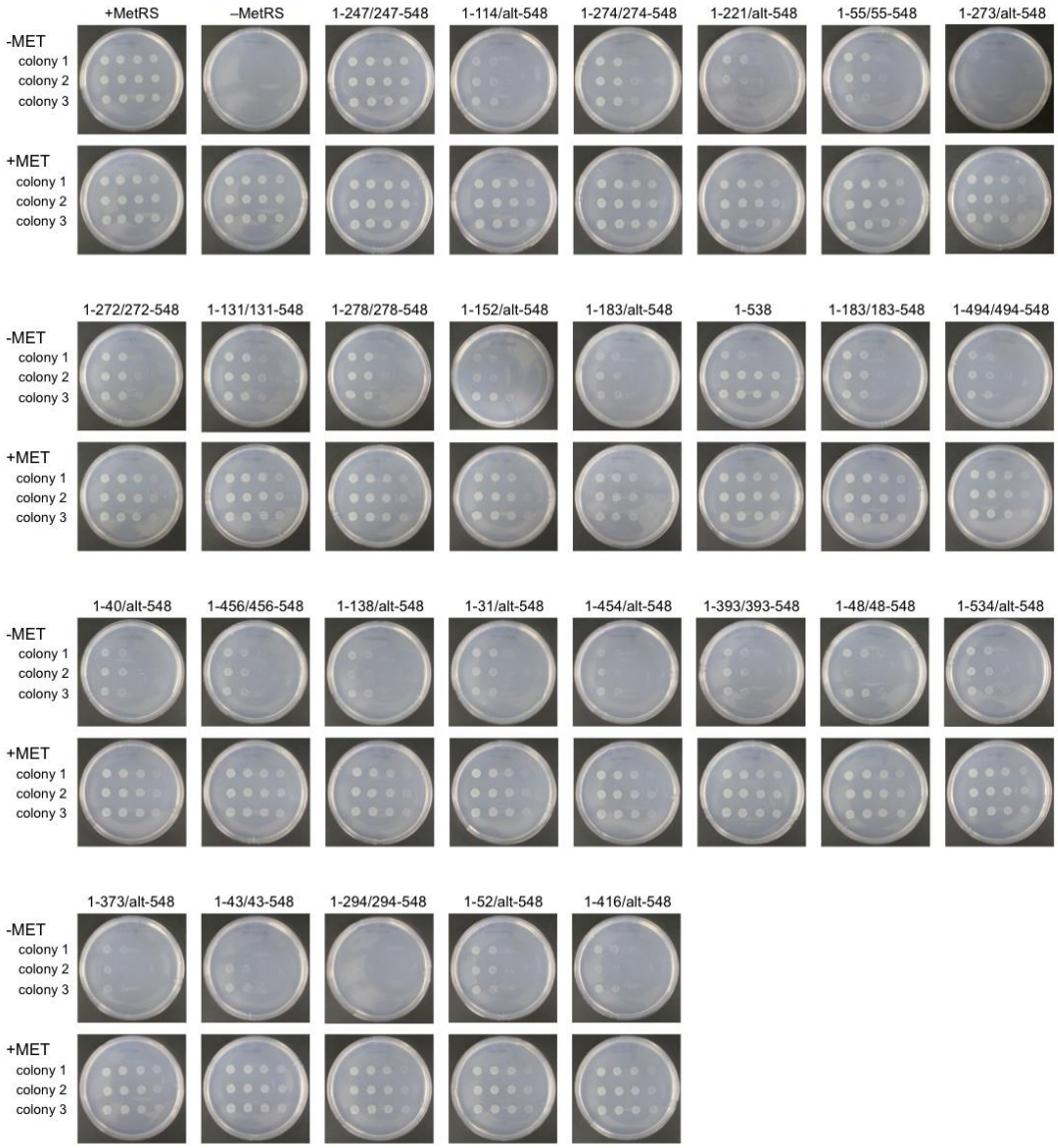


Figure S3.2.

Intermolecular contacts between fragments. The number of intermolecular interactions between fragments in each split MetRS variant was calculated by counting the number of residue–residue contacts across the inter-fragment interface. Residues were considered interacting if any of their atoms were within 4.5 Å of one another. Calculations are based on the MetRS crystal structure (Protein Data Bank code 1f4l) [21].

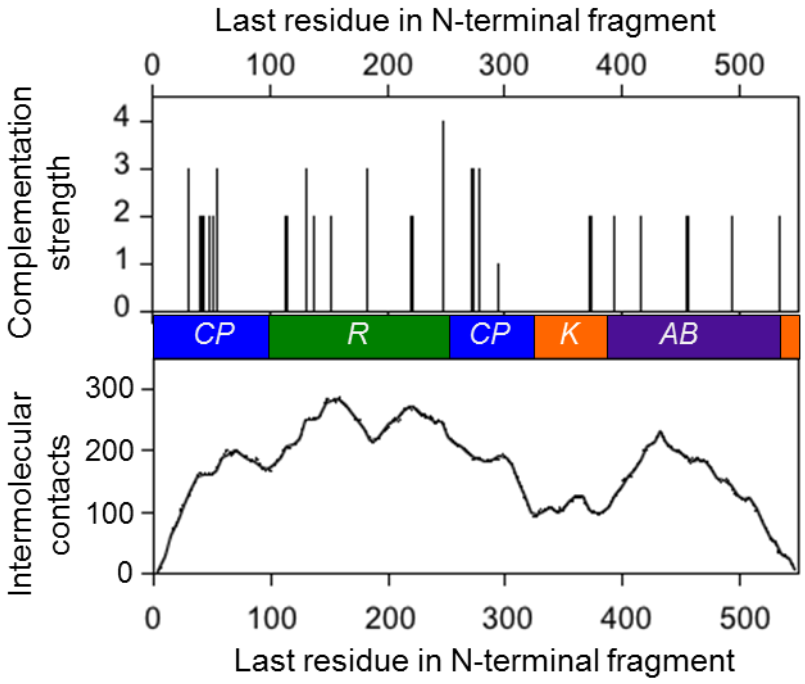


Figure S3.3.

Structural models for fragmented MetRS variants. PDB structure 2CSX[30] was used to generate structural models for the seven split MetRS variants that were screened for Anl labeling activity after mutation of residues L13, Y260 and H301 to NLL and PLL, respectively. These variants were chosen because they were among the most active fragmented MetRS in bacterial complementation measurements. The backbone cleavage sites were distributed through different domains of the protein in these variants. For each variant, residues in the N-terminal MetRS fragment are shown in blue, residues from the C-terminal fragment are shown in magenta, and bound tRNA is shown in orange. Split MetRS are named based on the last residue within the N-terminal fragment before peptide backbone cleavage, *e.g.*, MetRS(247). This residue is duplicated and found in both the N-terminal and C-terminal fragments in our constructs generated by transposase mutagenesis [6].

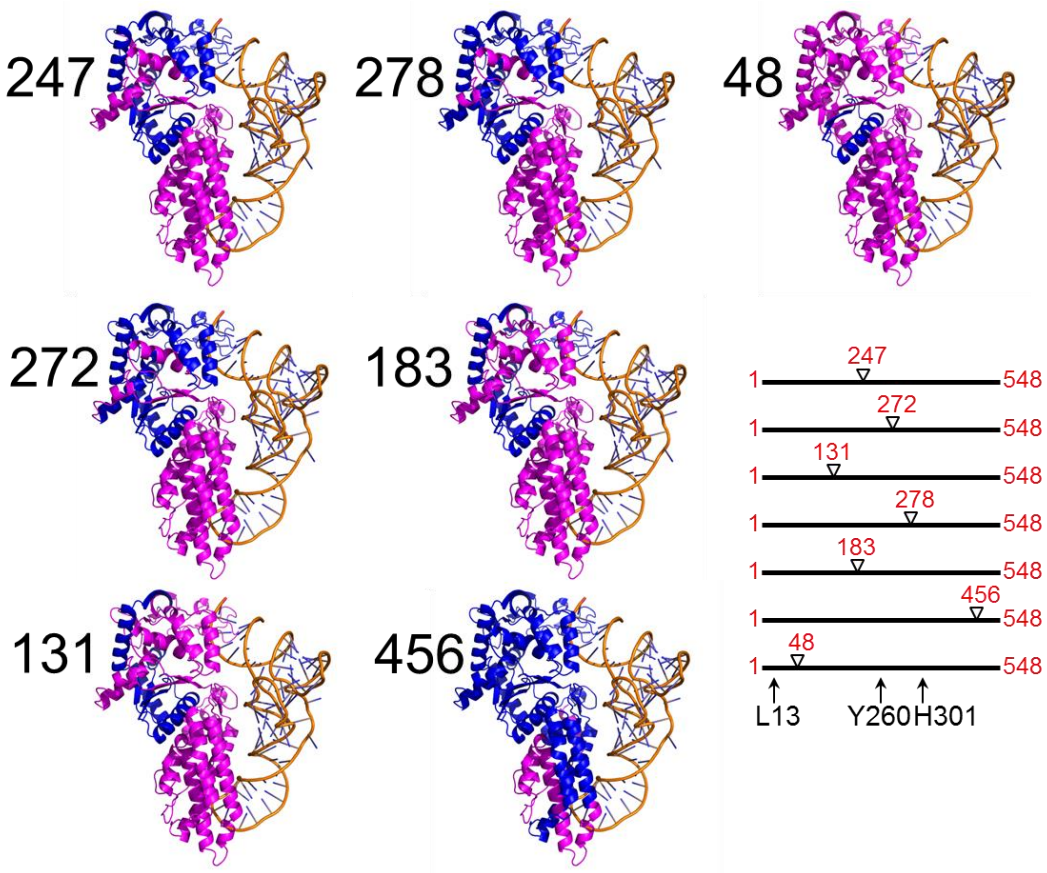


Figure S3.4.

Anl labeling in cells expressing NLL-MetRS(247) under control of regulated promoters. *E. coli* DH10B cells transformed with vectors encoding NLL-MetRS(247) [pQE80His6_N247 and pBAD33HA_C247] were grown in M9 minimal media containing 19 amino acids, 100 μM methionine, and 1% glycerol. At OD_{600} = 0.25, IPTG and arabinose were added to final concentrations of 1 mM and 5 mM, respectively. Cells were grown for 2 h after induction, Anl was added to a final concentration of 1 mM, and cells were grown for an additional 1 h to allow for Anl incorporation into newly synthesized proteins. Cells were lysed, and lysates were treated with a TAMRA-alkyne dye. Proteins from the cell lysate were separated by SDS-PAGE, and TAMRA-labeled proteins were detected by in-gel fluorescence imaging.

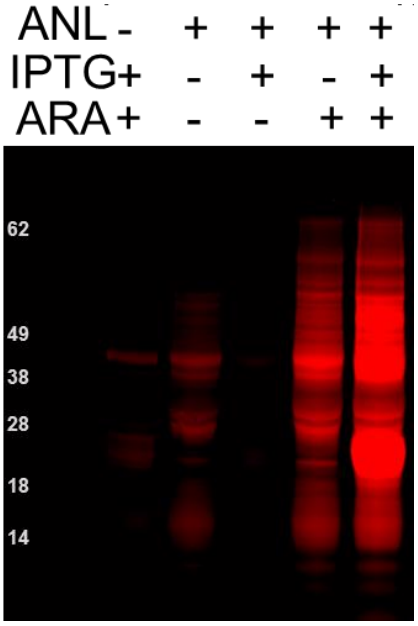


Figure S3.5.

Effect of glucose on NLL-MetRS(247) activity. *E. coli* DH10B cells transformed with vectors encoding NLL-MetRS(247) [pQE80His6_N247 and pBAD33HA_C247] were grown in M9 minimal media containing 19 amino acids, 100 μM methionine, 1% glycerol, and varying glucose concentrations (0, 0.1, 0.5, 1, 5 mg/ml). At OD_{600} = 0.25, IPTG and arabinose were added to final concentrations of 1 mM and 5 mM, respectively. Cells were grown for 2 h after induction, Anl was added to a final concentration of 1 mM, and cells were grown for 1 h to allow Anl incorporation into newly synthesized proteins. Cells were lysed and lysates were treated with a TAMRA-alkyne dye. Proteins from the cell lysate were separated by SDS-PAGE and Anl-labeled proteins were detected by in-gel fluorescence imaging. Colloidal blue staining of the same gel shows similar protein levels in each lane. Labeling observed under similar conditions with cells expressing intact NLL-MetRS under control of its endogenous MetG promoter in pQE80[2] is shown for comparison.

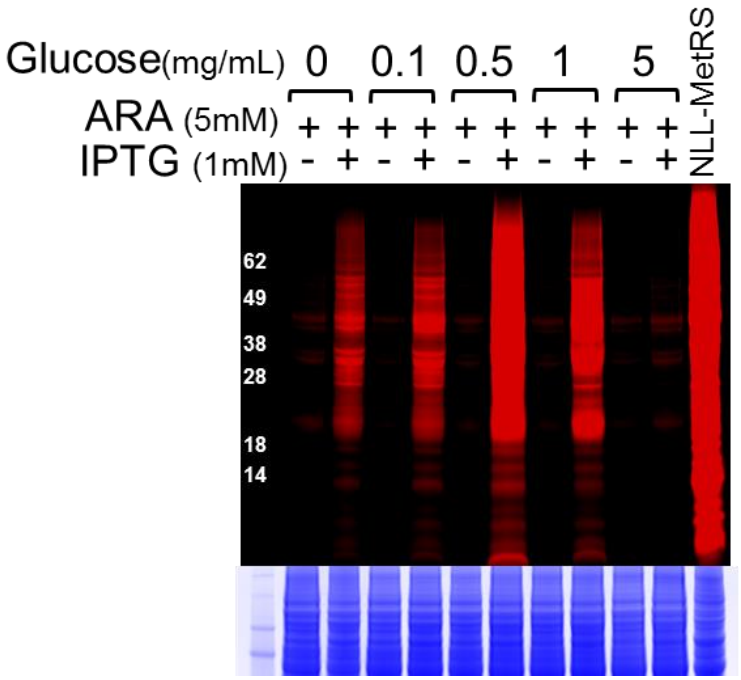


Figure S3.6.

Effect of varying IPTG and arabinose concentrations on NLL-MetRS(247) activity. *E. coli* DH10B cells transformed with vectors encoding NLL-MetRS(247) [pQE80His6_N247 and pBAD33HA_C247] were grown in M9 minimal media containing 19 amino acids, 100 μM methionine, 1% glycerol, and 0.5 mg/mL glucose. At OD_{600} = 0.25, 25 different concentrations of arabinose and IPTG were added to cultures. IPTG was added to final concentrations of 0 μM, 10 μM, 100 μM, 1 mM and 5 mM. Arabinose was added to final concentrations of 0 mM, 0.1 mM, 1 mM, 6.7 mM and 12 mM. Cells were grown for 2 h after induction, Anl was added to a final concentration of 1 mM, and cells were grown for 1 h to allow Anl incorporation into newly synthesized proteins. Cells were lysed and lysates were treated with a TAMRA-alkyne dye. Proteins from the cell lysate were separated by SDS-PAGE, and Anl-labeled proteins were detected by in-gel fluorescence imaging (top panel). The bottom gel shows a Western blot of the cell lysates using an AlexaFluor488-conjugated anti penta-His antibody that detects the N-terminal MetRS fragment (green) and an AlexaFluor647-conjugated anti hemagglutinin antibody that detects the C-terminal MetRS fragment (blue).

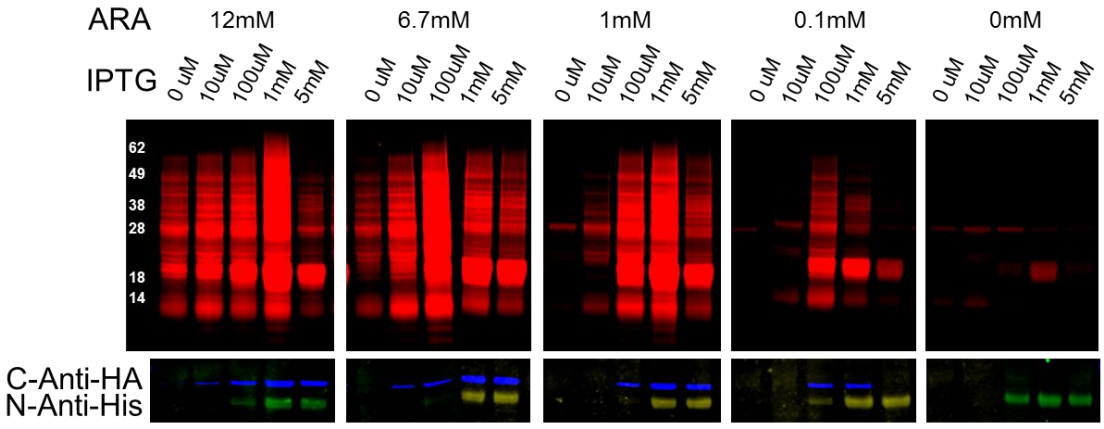
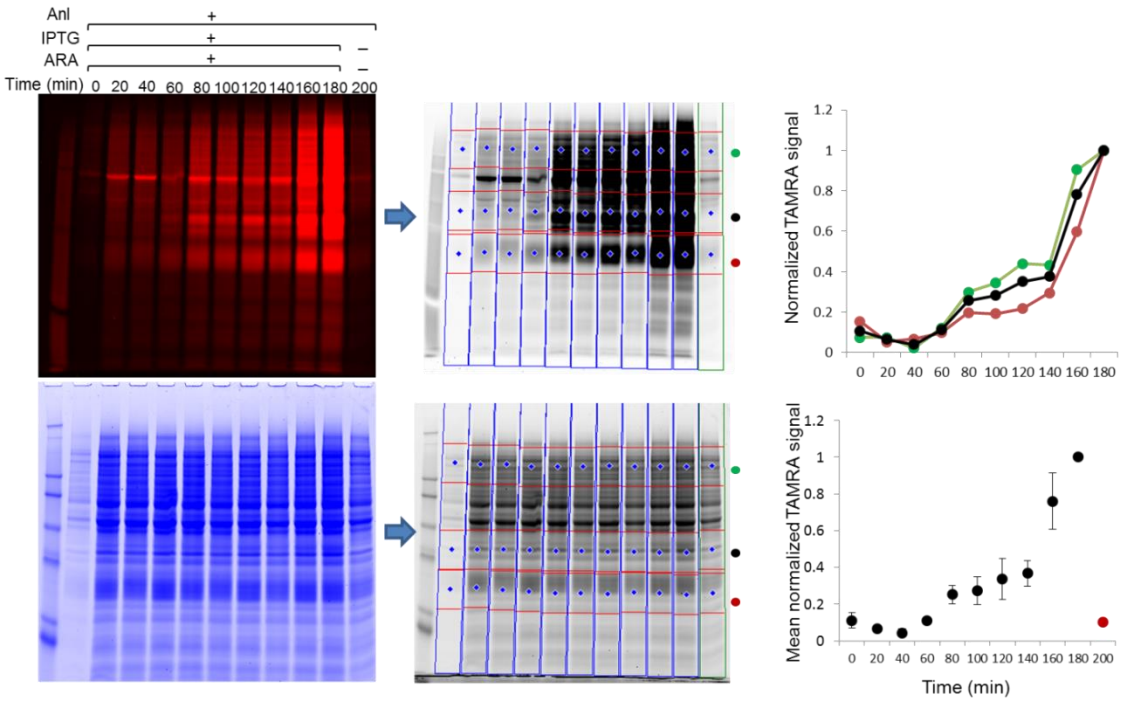


Figure S3.7.

Kinetics of AND-gate activation and protein labeling. (a) Kinetics of AND gate activation: *E. coli* DH10B cells transformed with vectors encoding NLL-MetRS(247) [pQE80His6_N247 and pBAD33HA_C247] were grown in M9 minimal medium containing 19 amino acids, 100 μM methionine, 1% glycerol, and 0.5 mg/mL glucose. The N- and C-terminal fragments were induced by addition of IPTG and arabinose at 1 mM and 5 mM, respectively. Anl was added at 1 mM at the time of induction. As a control for background labeling, cells were also incubated in 1 mM Anl without IPTG and arabinose for 200 min. After Anl incorporation for various times (0 – 180 min), cells were lysed and treated with azadibenzocyclooctyne-TAMRA dye (see supplementary methods). TAMRA labeling (top left) indicating Anl incorporation was detected by in-gel fluorescence imaging. Colloidal blue labeling of the same gel (bottom left) was used to compare protein loading across lanes. Quantification of Anl incorporation was done by measuring TAMRA fluorescence at three different bands in each lane and dividing these values by the band intensities from colloidal blue labeling, thereby normalizing for differences in protein amounts between different conditions. The level of Anl labeling in the absence of IPTG and arabinose was also measured after 200 min, and is represented by the red dot in the mean-normalized TAMRA fluorescence intensity plot.



(b) Kinetics of Anl incorporation under control of the activated AND gate: cells transformed with vectors encoding NLL-MetRS(247) [pQE80His6_N247 and pBAD33HA_C247] were grown in M9 minimal media containing 19 amino acids, 100 μM methionine, 1% glycerol, and 0.5 mg/mL glucose. The N- and C-terminal fragments were induced by addition of IPTG and arabinose at 1 mM and 5 mM, respectively, for 3 h. Anl was then added at 1 mM. Under these conditions, fragment expression is not the rate-limiting step for Anl incorporation. As a control for background labeling, cells were also incubated in 1 mM Anl without IPTG and arabinose for 65 min; the TAMRA signal for this control is shown by red symbol in the quantification results. Quantification was performed as described above.

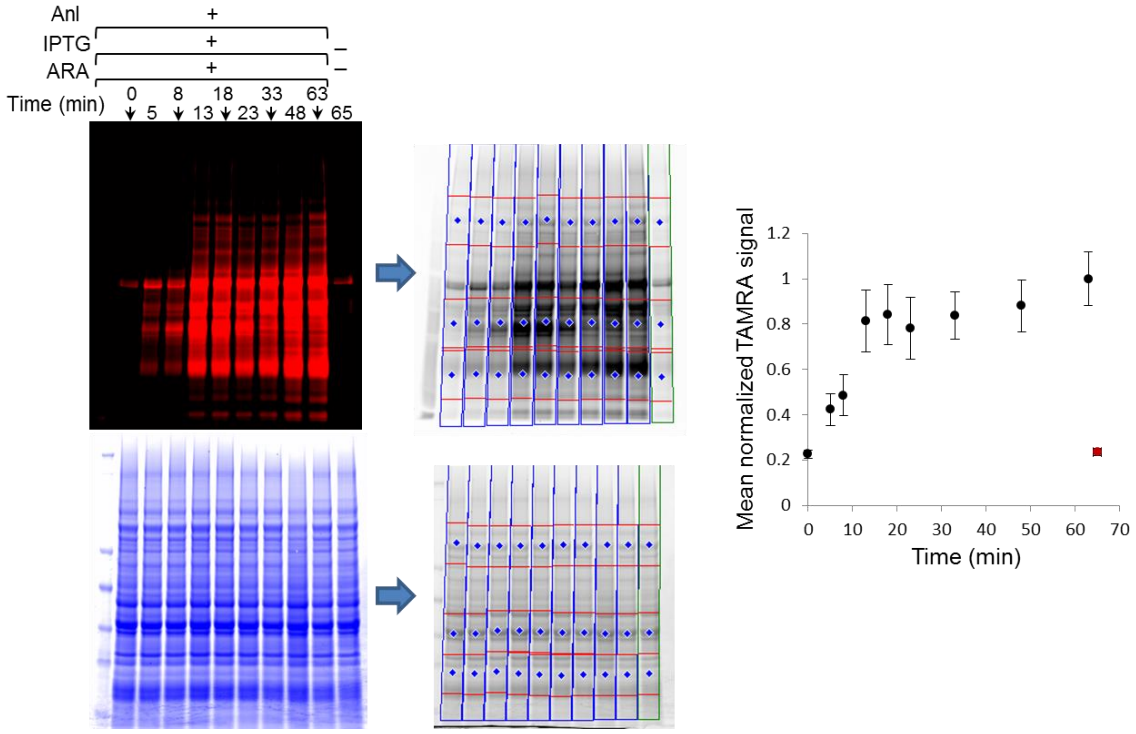
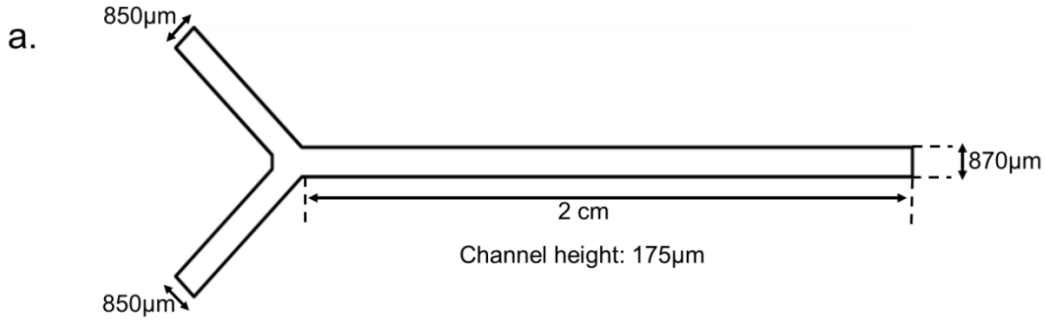


Figure S3.8.

Modeling the diffusion of IPTG and arabinose in the microfluidic channel. To determine the profiles of IPTG and arabinose in the microfluidic channel, the dimensions of the channel shown in (*A*) were entered into AutoCad software and imported into Comsol software to model diffusion at the liquid-liquid interface. A flow rate of 1 µL/min was used for each inducer.



(*B*) Model-predicted IPTG inducer concentration profile. Inlet concentration is 1 mM.

(*C*) Model-predicted arabinose inducer concentration profile. Inlet concentration is 5 mM.

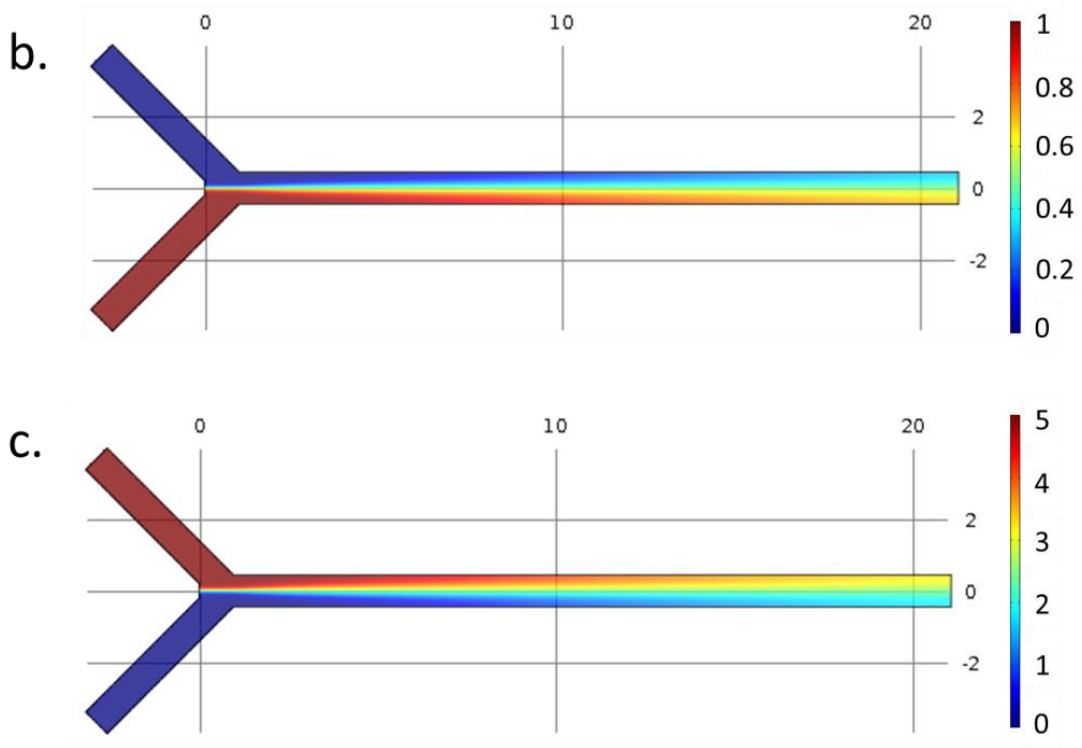


Figure S3.9.

Model prediction and experimental measurement of dye diffusion in the microfluidic channel. The diffusion of AlexaFluor488 dye in the microfluidic channel was computed under the conditions used for the experiments. Reported values for the diffusion coefficient of AlexaFluor488 dye range from 1.96 $\times 10^{-10}$ $m^2 \cdot s^{-1}$ to 4.62×10^{-10} $m^2 \cdot s^{-1}$ at room temperature [31-37]. We used 4.3×10^{-10} $m^2 \cdot s^{-1}$ because this value is commonly used[38-45] as the diffusion coefficient for this dye in water or aqueous buffer at room temperature. The diffusion profiles 0 mm, 6 mm and 12 mm downstream from the y-junction were predicted for AlexaFluor488. To verify these predictions, dye diffusion experiments were performed; the results closely matched the predicted concentration profiles. These results validated the prediction from simulations.

(*A*) To visualize diffusion in the laminar flow microfluidic channel, two liquid streams containing AlexaFluor488 dye (2.82 µM) and PBS were used. Fluorescence images were obtained from the glass-liquid interface inside the microfluidic channel. A series of images was obtained along the length of the channel and images were stitched together to provide the entire profile along the channel. Diffusion of dye at the interface of the two streams is visible along the length of the channel. (*B*) Diffusion of AlexaFluor488 dye in the microfluidic channel was modeled and the predicted profile of dye concentration at the liquid-surface interface along the entire length of the channel is plotted. AlexaFluor488 dye is depicted in red; PBS in blue.

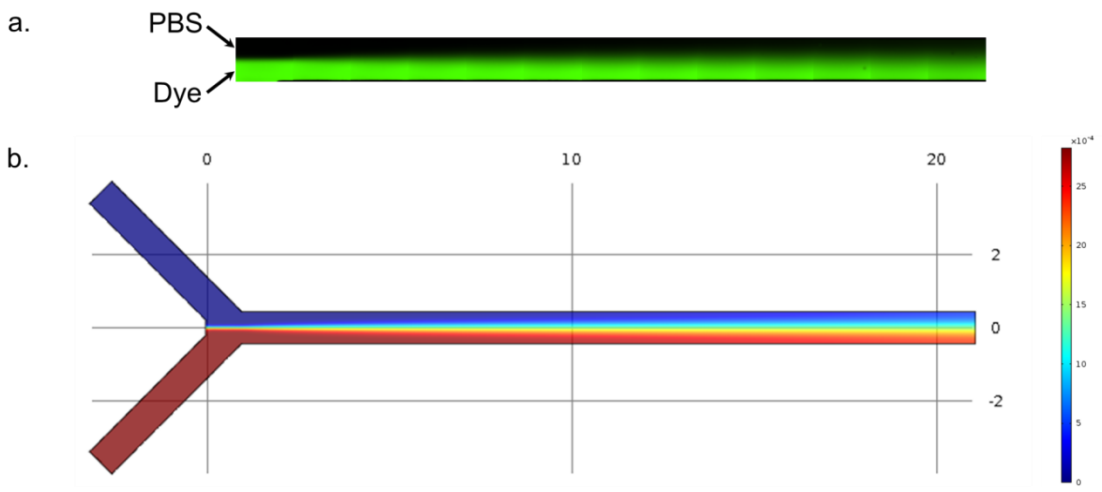
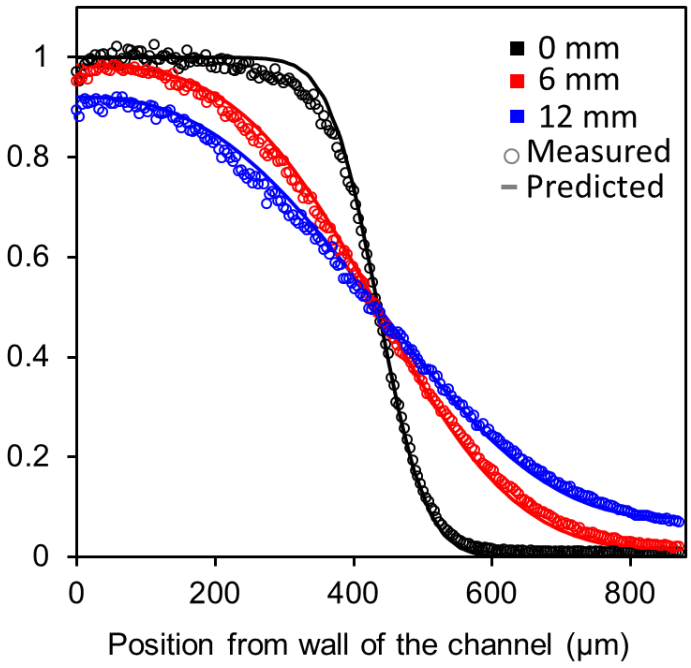


Figure S3.9 continued.

(**c**) Comparison of predicted and measured dye concentration profiles in the microfluidic channel. Predicted cross-channel dye profiles 0, 6, and 12 mm downstream from the y-junction are plotted as solid lines with color codes black, red and blue, respectively. Fluorescence quantification was also performed at these locations. Experimental data points are plotted as open circles, color coded as above, and show good agreement between the predicted and measured dye concentration profiles.



References

(1) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482.

(2) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Biol.* **2009**, *5*, 715.

(3) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. *ACS. Chem. Biol.* **2012**, *7*, 1326.

(4) Ullmann, A.; Jacob, F.; Monod, J. *J. Mol. Biol.* **1967**, *24*, 339.

(5) Burbaum, J. J.; Schimmel, P. *Biochemistry* **1991**, *30*, 319.

(6) Segall-Shapiro, T. H.; Nguyen, P. Q.; Dos Santos, E. D.; Subedi, S.; Judd, J.; Suh, J.; Silberg, J. J. *J. Mol. Biol.* **2011**, *406*, 135.

(7) Mellot, P.; Mechulam, Y.; Le Corre, D.; Blanquet, S.; Fayat, G. *J. Mol. Biol.* **1989**, *208*, 429.

(8) Somerville, C. R.; Ahmed, A. *J. Mol. Biol.* **1977**, *111*, 77.

(9) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A., 3rd; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15285.

(10) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596.

(11) Tornoe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057.

(12) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 9879.

(13) Wang, B.; Kitney, R. I.; Joly, N.; Buck, M. *Nat. Commun.* **2011**, *2*, 508.

(14) Lucchetta, E. M.; Lee, J. H.; Fu, L. A.; Patel, N. H.; Ismagilov, R. F. *Nature* **2005**, *434*, 1134.

(15) Berk, V.; Fong, J. C.; Dempsey, G. T.; Develioglu, O. N.; Zhuang, X.; Liphardt, J.; Yildiz, F. H.; Chu, S. *Science* **2012**, *337*, 236.

(16) Tamsir, A.; Tabor, J. J.; Voigt, C. A. *Nature* **2011**, *469*, 212.

(17) Moon, T. S.; Lou, C.; Tamsir, A.; Stanton, B. C.; Voigt, C. A. *Nature* **2012**, *491*, 249.

(18) Hodas, J. J.; Nehring, A.; Hoche, N.; Sweredoski, M. J.; Pielot, R.; Hess, S.; Tirrell, D. A.; Dieterich, D. C.; Schuman, E. M. *Proteomics* **2012**, *12*, 2464.

(19) Eichelbaum, K.; Winter, M.; Diaz, M. B.; Herzig, S.; Krijgsveld, J. *Nat. Biotechnol.* **2012**, *30*, 984.

(20) Hoff, K. G.; Culler, S. J.; Nguyen, P. Q.; McGuire, R. M.; Silberg, J. J.; Smolke, C. D. *Chem. Biol.* **2009**, *16*, 1299.

(21) Serre, L.; Verdon, G.; Choinowski, T.; Hervouet, N.; Risler, J. L.; Zelwer, C. *J. Mol. Biol.* **2001**, *306*, 863.

(22) Link, A. J.; Vink, M. K.; Tirrell, D. A. *Nat. Protoc.* **2007**, *2*, 1884.

(23) Anderson, J. R.; Chiu, D. T.; Jackman, R. J.; Cherniavskaya, O.; McDonald, J. C.; Wu, H.; Whitesides, S. H.; Whitesides, G. M. *Anal. Chem.* **2000**, *72*, 3158.

(24) Cowan, S. E.; Liepmann, D.; Keasling, J. D. *Biotechnol. Lett.* **2001**, *23*, 1235.

(25) Bashkatov, A. N.; Genina, E. A.; Sinichkin, Y. P.; Kochubey, V. I.; Lakodina, N. A.; Tuchin, V. V. *Biophys. J.* **2003**, *85*, 3310.

(26) Mogi, N.; Sugai, E.; Fuse, Y.; Funazukuri, T. *J. Chem. Eng. Data* **2007**, *52*, 40.

(27) Stewart, P. S. *J Bacteriol* **2003**, *185*, 1485.

(28) Nakanishi, K.; Ogiso, Y.; Nakama, T.; Fukai, S.; Nureki, O. *Nat. Struct. Mol. Biol.* **2005**, *12*, 931.

(29) Schmitt, E.; Tanrikulu, I. C.; Yoo, T. H.; Panvert, M.; Tirrell, D. A.; Mechulam, Y. *J. Mol. Biol.* **2009**, *394*, 843.

(30) Nakanishi, K.; Ogiso, Y.; Nakama, T.; Fukai, S.; Nureki, O. *Nat. Struct. Mol. Biol.* **2005**, *12*, 931.

(31) Pristinski, D.; Kozlovskaya, V.; Sukhishvili, S. A. *J. Chem. Phys.* **2005**, *122*, 14907.

(32) Doeven, M. K.; Folgering, J. H.; Krasnikov, V.; Geertsma, E. R.; van den Bogaart, G.; Poolman, B. *Biophys. J.* **2005**, *88*, 1134.

(33) Modos, K.; Galantai, R.; Bardos-Nagy, I.; Wachsmuth, M.; Toth, K.; Fidy, J.; Langowski, J. *Eur. Biophys. J.* **2004**, *33*, 59.

(34) Needleman, D. J.; Xu, Y.; Mitchison, T. J. *Biophys. J.* **2009**, *96*, 5050.

(35) Ramadurai, S.; Duurkens, R.; Krasnikov, V. V.; Poolman, B. *Biophys. J.* **2010**, *99*, 1482.

(36) Sanchez, S. A.; Brunet, J. E.; Jameson, D. M.; Lagos, R.; Monasterio, O. *Protein Sci.* **2004**, *13*, 81.

(37) Vendelin, M.; Birkedal, R. *Am. J. Physiol. Cell Physiol.* **2008**, *295*, 24.

(38) Dross, N.; Spriet, C.; Zwerger, M.; Muller, G.; Waldeck, W.; Langowski, J. *PLoS One* **2009**, *4*, 4.

(39) Nitsche, J. M.; Chang, H. C.; Weber, P. A.; Nicholson, B. J. *Biophys. J.* **2004**, *86*, 2058.

(40) Oh, D.; Zidovska, A.; Xu, Y.; Needleman, D. J. *Biophys. J.* **2011**, *101*, 1546.

(41) Petrášek, Z.; Schwille, P. *Biophys. J.* **2008**, *94*, 1437.

(42) Ries, J.; Ruckstuhl, T.; Verdes, D.; Schwille, P. *Biophys. J.* **2008**, *94*, 221.

(43) Yordanov, S.; Best, A.; Weisshart, K.; Koynov, K. *Rev. Sci. Instrum.* **2011**, *82*, 036105.

(44) Zareh, S. K.; DeSantis, M. C.; Kessler, J. M.; Li, J. L.; Wang, Y. M. *Biophys. J.* **2012**, *102*, 1685.

(45) Zengel, P.; Nguyen-Hoang, A.; Schildhammer, C.; Zantl, R.; Kahl, V.; Horn, E. *BMC Cell Biol.* **2011**, *12*, 21.

CHAPTER IV

Cell-Selective Identification of Newly Synthesized Proteins in Mammalian Cells Using an Engineered Aminocyl tRNA Synthetase

Abstract

Cell-selective identification of newly synthesized proteins will facilitate studies of biological processes in multicellular organisms. We have previously developed methods to identify newly made proteins in a cell-selective manner in bacterial cells; however, comparable strategies for proteomic analysis in mammalian cells are not available. Here we introduce a mutant mammalian methionyl tRNA synthetase (AnlRS) that incorporates the non-canonical amino acid azidonorleucine (Anl) into newly synthesized proteins. Controlled expression of this enzyme restricts protein labeling to cells of interest and facilitates identification by mass spectrometry. This approach does not require expression of orthogonal tRNAs or the removal of canonical amino acids. Successful Anl incorporation in a variety of mammalian cells demonstrates the potential utility of this enzyme for a wide range of proteomic studies.

Introduction

The incorporation of non-canonical amino acids into newly synthesized proteins has emerged as an important tool for study of proteome dynamics in cells.[1-3] This approach, named BONCAT,[4] utilizes the metabolic incorporation of non-canonical amino acids into newly made proteins to enable their enrichment and subsequent identification by mass spectrometry. A number of recent studies have used a combination of BONCAT and SILAC techniques[5] to quantitatively analyze proteins in mammalian cell lines.[6-8] A similar strategy was used to identify the secretome of different cell lines in serum containing media.[9] These studies highlight the importance of methods for analyzing proteome dynamics in mammalian cells. However, the inability to restrict labeling and thereby identify newly made proteins from subsets of mammalian cells has thus far limited the scope of problems that can be addressed using BONCAT. Here we introduce a new enzyme that allows cell-selective, time-resolved, identification of newly synthesized proteins in mammalian cells.

The BONCAT method was initially extended to achieve cell-selective protein labeling in bacterial cells[10] and we subsequently demonstrated an application of this approach to study host-pathogen interactions wherein the proteins of pathogenic microbes were selectively labeled and identified in the presence of host cells.[11] These studies utilized a mutant *E. coli* methionyl-tRNA synthetase (MetRS), which is a class I aminoacyl-tRNA synthetase (aaRS). The triple mutant (L13N/Y260L/H301L)[10] MetRS variant, designated NLL-MetRS, charges the non-canonical amino acid L-azidonorleucine (Anl) to $tRNA^{Met}$ and thereby enables incorporation of Anl at Met codons. Since the wild-type MetRS does not charge Anl, only cells that express the NLL-MetRS incorporate Anl. Controlled expression of NLL-MetRS using various promoters enabled the context-dependent and time-resolved labeling of newly synthesized proteins in bacteria.[12,13] Ideally, one could use an analogous approach for cell-selective protein labeling in mammalian cells (Scheme 4.1a). However, the *E.coli* NLL-MetRS enzyme only charges Anl to initiator tRNA ($tRNA^{fMet}$) and not the elongator mammalian tRNA ($tRNA^{Met}$). Therefore, this enzyme incorporates Anl site-selectively at N-terminal positions and not at internal Met residues in mammalian proteins. Based on this observation as well as the fact methionine aminopeptidase (MetAP) removes N-terminal Met residues from a wide range of proteins,[14] we sought to identify new enzyme variants to incorporate Anl into mammalian cell proteins. Newly synthesized proteins can be detected in fixed cells or cell lysates by 3+2 cycloaddition reactions between the azide side-chain of Anl and alkyne[15] or alkynyl probes[16,17] (Scheme 4.1b). Anl-labeled proteins can be tagged for affinity enrichment and identified by tandem mass spectrometry (Scheme 4.1c).

Results and Discussion

We investigated two complementary approaches towards cell-selective Anl incorporation in mammalian cells. The first approach aimed to identify mutant mammalian aaRSs that charge Anl to the endogenous $tRNA^{Met}$. The second approach involved co-expression of *E. coli* NLL-MetRS and *E. coli* elongator $tRNA^{Met}$ in mammalian cells. Reports by Rajbhandary and co-workers showing successful concomitant expression of an active *E. coli* glutaminyl-tRNA synthetase and a functional glutamine-inserting amber suppressor tRNA that is not aminoacylated *in vivo* by any mammalian aaRS suggested to us that the mammalian translation machinery could co-opt an *E. coli* elongator $tRNA^{Met}$ that is charged by the NLL-MetRS.[18,19] Additionally, the fact that NLL-MetRS exclusively charges Anl to the mammalian $tRNA^{fMet}$,[20] (and that the *E. coli* MetRS only charges Met to mammalian $tRNA^{fMet}$)[21] provided further evidence for the potential feasibility of this approach.

To identify mutant mammalian aaRSs that charge Anl to the endogenous $tRNA^{Met}$, we started with residues 13, 260 and 301 in the Met-binding pocket of *E. coli* MetRS[22] which can be mutated to expand the substrate specificity towards Anl.[23] Although there are no reported structures of a mammalian MetRS, we used sequence alignment to determine the corresponding positions in eukaryotic MetRSs (Figure 4.1a). Based on the similarity of eukaryotic MetRSs and the conservation of residues that align with *E. coli* MetRS residues 13/260/301, we hypothesized that site-directed mutagenesis of corresponding positions in the mouse methionyl tRNA synthetase (Mars) would potentially yield active variants that charge Anl. Therefore, we constructed a small mutant library in which residues L274, Y527, H562 of the Mars enzyme were mutated to SLL, CLL, PLL and NLL respectively. We also made an L274G mutant. These mutations correspond to *E. coli* MetRS variants that charge Anl in *E. coli* (Figure S4.1-S4.6). In addition we also overexpressed the wild-type Mars enzyme to assess the level of background incorporation. The various MetRS variants were expressed under control of a CMV promoter (Figure 4.1b).

To express the *E. coli* elongator tRNA, we followed the work of the Rajbhandary group and used 5' and 3' flanking sequences from the human initiator tRNA, which enable proper RNA editing (the *E. coli* version of these sequences are not recognized in mammalian cells)[18] (Figure 4.1c). Notably, efficient transcription by mammalian RNA polymerase III requires a purine at position 9 of the tRNA coding sequence. In this case, adenine was used. This adenine lies in the A box of the internal promoter of eukaryotic tRNA genes.[18] Transcription of the tRNA is terminated by a

stretch of thymidines. The 5' sequence of the tRNA is removed by RNase P, and the 3' end is removed by tRNase Z[24]. In mammalian cells, the 3' CCA tail is added by a nucleotidyltransferase;[24] therefore, we tested tRNA sequences with and without a CCA tail (Figure S4.7-S4.9). We measured the activity of these various constructs for Anl incorporation by transiently transfecting several widely used mammalian cell lines, including Chinese hamster ovary (CHO) cells, monkey kidney cells (COS7) and an immortalized human cervical cancer cell line (HeLa). Cells were incubated in serum containing media supplemented with 1.5 mM Anl for a total of 10 hours and thereafter lysed. Total cellular proteins were treated with alkyne-tetramethylrhodamine (TAMRA) dye (Figure S4.10) for labeling via copper-catalyzed azide-alkyne cycloaddition (reaction **1**, scheme 4.1b)[15,25,26]. After SDS-PAGE, TAMRA-labeled proteins were detected by in-gel fluorescence imaging (Figure 4.1d). As expected, Anl was not charged to any appreciable extent by the wild-type translation machinery of the cells. Anl incorporation was detected in cells carrying the *E. coli* NLL-MetRS. We did not observe a significant increase in Anl incorporation as a function of tRNA expression. The variability of incorporation across different cell lines suggests that the expression levels of the tRNA and the synthetase may need to be optimized for each cell type, thereby limiting the general applicability of this approach. Nonetheless, this strategy may prove to be useful for protein labeling in specific cases. For example, a similar strategy was used to incorporate several non-canonical amino acids using a genetically encoded pyrrolysyl-tRNA synthetase/tRNA pair in *Drosophila*.[27]

In contrast to our results with the *E. coli* NLL-MetRS/tRNA^{Met} (Figure 4.1D, *E. coli* MetRS), the L274G version of the mouse Mars enzyme (henceforth called AnlRS) resulted in robust Anl incorporation across all three cell lines while none of the other mutant Mars variants charged Anl (Figure 4.1D, mouse Mars). This result was encouraging because it demonstrates that a single mutation to the endogenous mammalian MetRS sequence is sufficient for Anl incorporation without requiring expression of exogenous tRNAs.

An important parameter for characterizing incorporation of non-canonical amino acids is the selectivity of a mutant aaRS towards a non-canonical amino acid versus its cognate substrate. To determine the selectivity of the AnlRS for Anl over Met, we aimed to measure enzyme activity in a manner that is most relevant to typical labeling experiments in mammalian cells, for example by measuring enzyme activity in the context of live cells. Therefore, we measured protein labeling in CHO cells by incubating the cells at different Anl concentrations in serum containing media. The DMEM component of media typically contains 150-200 μM of Met. After Anl incorporation, proteins from cell lysates were labeled with an alkyne-TAMRA dye (Figure 4.2a). We also

performed Met competition assays by incubating CHO cells with 1.5mM Anl at different Met concentrations (Figure 4.2b). We reasoned that because the same enzyme can charge both Anl and Met, it is important to know its selectivity as determined by the ratio of k_{cat}/k_M for Anl versus Met. Based on the enzyme kinetics of activation for both substrates, we obtained the following expression for product formation rate *V* (see figure S4.11 for derivation):

$$V = \frac{A(K^A/K^M)\,C}{M + A(K^A/K^M)}$$

Where K^A and K^M are k_{cat}/k_M for Anl and Met, respectively. Met and Anl concentrations (as depicted by M and A) are known, and C is a constant. Fitting this equation to the quantified TAMRA dye fluorescence measurements (which provide a direct measure of Anl incorporation in newly synthesized proteins) yielded a selectivity value of 0.25 for Anl versus Met (i.e., 1 Anl residue is incorporated for every 4 Met residues under these conditions). These results demonstrate the ability to tune the level of Anl incorporation by controlling extracellular Anl concentration as required for various biological studies. Previous work involving incorporation of the non-canonical amino acid azidohomoalanine (Aha), also a Met surrogate, that is incorporated non-cell-selectively by the endogenous MetRS, showed that an incorporation of approximately 6% of total Met codons was sufficient for proteomic identification.[7]

The ability to restrict AnlRS expression to cells of interest should enable the cell-selective labeling of proteins for targeted analysis of protein synthesis under specific conditions. To validate this concept, we developed a reporter construct expressing both AnlRS and mCherry under a CMV promoter (pMarsC). A T2A linker was used to express two disjointed proteins in the same open reading frame; we reasoned that fusing AnlRS and mCherry might alter the activity of the synthetase and confound protein labeling results (Figure 4.3a). As an additional control, we developed a similar expression construct without mCherry (pMars) (Figure S4.12-13).When transfected into CHO cells, both constructs labeled cellular proteins at comparable levels, indicating that AnlRS was functional in both contexts. A C-terminal Myc tag on the mCherry protein showed that AnlRS and mCherry are expressed as two separate proteins and not a fusion (Figures 4.3b and S4.14). We therefore used the mCherry signal to track expression of AnlRS in the pMarsC transfected cells. Cells expressing mCherry should also contain AnlRS and incorporate Anl. CHO cells transfected with pMarsC and incubated with Anl (1.5 mM, 6h) were fixed, permeabilized and reacted with alkyne-TAMRA through a copper-catalyzed cycloaddition reaction. We observed TAMRA fluorescence exclusively in the mCherry-expressing cells which indicated cell-selective incorporation of Anl (Figure 4.3c). In addition to promoter controlled

expression of AnlRS, it would be convenient to develop complementary methods for conditional and regulatable transgene expression. Towards this aim, we explored the use of site-specific Cre-Lox recombinase technology because a variety of well-established Cre reporter constructs, Cre-expressing organisms, and cell lines are readily available. We therefore developed a proof-of-concept Cre-Lox system for the conditional expression of AnlRS.

We first developed a stable CHO cell line containing an AnlRS-T2A-mCherry gene under a CMV promoter that is inactive because of the presence of a loxP-flanked stop sequence (Figure 4.4d). These cells were transiently transfected with a second (activating) construct containing a strong elongation factor (Ef1α) promoter that drives the expression of an eGFP-Cre fusion protein. We used the eGFP signal to track expression of Cre and followed the mCherry signal to monitor the successful recombination of the loxP sites. As expected, in the subset of cells that expressed Cre (as determined by the GFP signal), we observed mCherry expression, which is indicative of loxP recombination mediated AnlRS expression and TAMRA labeling which demonstrates selective Anl incorporation. These results show that cell-selective proteomic labeling can be achieved through regulatable transgene expression. Based on these results we anticipate that targeted insertion of LoxP-AnlRS reporter constructs into the ROSA26 locus of mouse embryonic stem (ES) cells[28] would potentially lead to the development of Cre-Lox mice in which tissue-selective proteomic studies can be performed. To validate that the labeled proteins can be identified by shotgun proteomics, we used an affinity chromatography column containing azadibenzocyclooctyne-functionalized beads to chemoselectively tag (using reaction **3**, scheme 4.1) and enrich Anl-labeled proteins from lysates of CHO cells. After on-resin trypsinization of purified proteins, we analyzed the resulting tryptic peptides by tandem mass spectrometry (Table S1). We identified a total of 820 proteins using a short (60 minute) run on a linear trap quadrupole (LTQ) Orbitrap mass spectrometer. We detected proteins from a variety of cellular components including nuclear, cytoskeletal, organelle-specific and membrane proteins (Figure 4.3e). Comparing these results to complementary methods of translation profiling, such as ribosome profiling,[29-32] may provide valuable information regarding similarities and differences between snapshots of translation and the distribution of newly synthesized proteins in cells. Such analysis may be useful to study mechanisms of the folding, maturation and processing of newly made proteins.

Conclusions

This work introduces a mutant mouse MetRS that metabolically incorporates the non-canonical amino acid Anl into mammalian cell proteins. This enzyme can be both genetically encoded and conditionally activated for time-resolved and cell-selective identification of newly synthesized proteins. Targeted expression of this enzyme provides a readily accessible approach for proteome and secretome analysis in subpopulations of mammalian cells.

Materials and Methods

Development of Mammalian Vectors for Expression of *E. Coli* NLL-MetRS/tRNA^{Met}

For PCR and cloning purposes, unless otherwise state, we used chemically competent *E. coli* MegaX DH10B (Zymo Research). Plasmid DNA was purified using a Miniprep kit (Qiagen), colony selection was performed on LB-agar plates with 100 µg/mL ampicillin. All plasmids were verified by sequencing (Laragen). The *E. Coli* NLL-MetRS enzyme was obtained from pAM1 vector (Addgene plasmid 51401) through PCR amplification using a Nhe1 forward primer and an Xho1 reverse primer and inserted into the multiple cloning site of the mammalian expression vector pCDNA3.1+ (Invitrogen). This vector contains a CMV promoter, a bovine growth hormone (BGH) transcriptional stop sequence, and a neomycin resistance gene for G418 selection. The resulting plasmid was named pMetRSNLL_G. For simultaneous expression of *E. Coli* tRNA^{Met}/NLL-MetRS, the mouse tRNA^{Met} sequence was synthesized (Integrated DNA Technologies) with 5' and 3' flanking sequences of the human tRNA^{fMet} as well as flanking BglII restriction enzyme cut sites, this construct was inserted into the BglII site in the backbone of the pMetRSNLL_G plasmid to produce the pMetRSNLLtRNA_G plasmid (sequence included in Figure S4.8). The pMetRSNLLtRNA_G plasmid is a pCDNA3.1+-based vector expressing the NLL-MetRS under CMV control as well as the *E. Coli* tRNA^{Met}. A second tRNA sequence lacking the C-terminal CCA tail was synthesized (Integrated DNA Technologies) and inserted into the BglII restriction site of MetRSNLL_G to make the pMetRSNLLtRNAdcca_G plasmid (sequence included in Figure S4.9).

Introduction of Mutations into the Endogenous Mars Sequence and Development of Associated Mammalian Expression Vectors

The endogenous sequence of mammalian Mars enzyme was obtained from a cDNA clone from American Type Culture Collection clone ID 6414029 (ATCC). Site-directed mutagenesis (Agilent) was used to introduce the NLL, CLL, PLL and SLL mutations at residues L274, Y527 and H562 respectively. The L274G mutation was also introduced. These sequences were PCR amplified with a Nhe1 forward primer and an Xho1 reverse primer and inserted into pCDNA3.1+ (Invitrogen), resulting in plasmids: pMarsWT_G, pMarsL274G_G/pMars_G, pMarsSLL _G, pMarsPLL_G, pMarsCLL_G, pMarsNLL_G (with sequences in Figures S4.1-S4.6 respectively).

Development of Mammalian Expression Vectors for Cell-Selective Proteomic Labeling and Cre-Lox Mediated Regulatable Expression Vector

For the pMarsC plasmid, the L274G Mars enzyme sequence was connected to an mCherry sequence through a T2A linker sequence by sewing PCR. Briefly, first the Mars enzyme sequence (obtained from pMarsL274G_G/pMars) was PCR amplified using a Nhe1 forward primer containing a Flag sequence tag and reverse primer containing a T2A sequence. The mCherry sequence was amplified using a matching T2A sequence in the forward primer and a reverse primers consisting of a C-terminal Myc-tag and stop codon. Sewing PCR was used to amplify the final product with the following sequence components: "Nhe1-FlagTag-Mars-T2A-Mcherry-MycTag-Xho1". This sequence was inserted between the Nhe1 and Xho1 cut sites of a pCDNA3.1+ vector containing a hygromycin resistance cassette (Invitrogen) to yield the pMarsC plasmid (sequence in Figure S4.12). For Cre-Lox mediated recombination, a LoxP flanked transcriptional stop sequence was inserted after the CMV promoter in the pMarsC plasmid. The LoxP flanked transcriptional stop sequence consisting of forward LoxP ('ATAACTTCGTATAGCATACATTATACGAAGTTAT') sequences flanking the transcriptional stop sequence was appended with Nhe1 restriction sites on both ends and synthesized (Integrated DNA Technologies). This insert was ligated into the Nhe1 cut site of pMarsC plasmid to yield the Cre-Lox plasmid pMarsC_lox_H (sequence in Figure S4.13). The correct orientation of the insert into the Nhe1 cut site was verified by sequencing.

Cell Culture and Media Compositions

Cells were passaged every 3 days on tissue-culture treated plates and incubated at 37°C and 5% CO_2. CHO-K1 cells were maintained in RMPI (Invitrogen) media with 10% fetal bovine serum, and supplemented with Pen/Strep, L-glutamine and non-essential amino acids (Invitrogen). HeLa and COS7 cells were cultured in DMEM (Invitrogen) media with 10% fetal bovine serum, and supplemented with Pen/Strep, L-glutamine and non-essential amino acids (Invitrogen).

Cell Transfection, Selection and Conditional Transgene Activation

Mammalian expression plasmids were amplified in *E. coli* MegaX DH10B and purified using endotoxin-free plasmid Maxi-kit (Qiagen). Lipfecatmine 2000 (Invitrogen) was used for all transfections according to manufacturer recommended procedures. For identification of MetRS variants that charge Anl and to study Anl incorporation, all cells were transiently transfected 30 hours prior to Anl labeling. For Cre-Lox transgene activation study, the pMarsC_lox_H was linearized using the BglII cut site in the backbone of the plasmid and transfected into CHO cells. After selection on hygromycin at 100 µg/ml for 10 days, surviving colonies were picked and expanded to yield a stable cell line. Cre-mediated recombination in these cells was achieved

through transient transfection with a plasmid expressing eGFP-Cre under control of a EF1α promoter (Addgene plasmid 11923).

Synthesis of Azidonorleucine

Azidonorleucine synthesis was based on a previous protocol for azidohomoalanine synthesis, using Boc-lysine as the starting material.[33] Briefly, 5.27 g (81.1 mmol) of sodium azide was treated with 2.7 mL (16 mmol) of distilled triflic anhydride in 13 mL of water for 2 h. The triflic azide product was extracted with 10 mL dichloromethane and added dropwise to a flask containing Boc-Lys-OH (2 g, 8.1 mmol), K_2CO_3 (1.68 g, 12.2 mmol) and $CuSO_4$ (20 mg, 0.08 mmol) in 26 mL of water and 250 mL of methanol. After 20 h at room temperature the product was extracted with ethyl acetate, redissolved in dichloromethane and purified by silica gel chromatography. After Boc deprotection with hydrochloric acid, the final product was purified by cation exchange chromatography.

Copper Catalyzed Reaction of Alkyne-TAMRA Dye with Anl-labeled Proteins in Cell Lysates and Detection by In-gel Fluorescence

Cells were lysed with 4% SDS in phosphate buffered saline (PBS). Ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor (Roche) was added to the lysates to reduce protease activity. PBS was added to dilute the SDS concentration to 1%, and cell lysates were centrifuged at 14,000 rcf for 10 min to remove cellular debris. Protein concentrations were measured by using a bicinchoninic protein quantification kit (BCA assay; Pierce). The same amount of protein was used for each condition; concentrations ranged from 0.1 to 0.4 mg/mL. Copper-catalyzed reactions was performed using Click-IT TAMRA protein analysis kit (Invitrogen). Proteins were precipitated with chloroform/methanol, washed with methanol to remove unreacted dye and resuspended in protein loading buffer containing 2% SDS and 10% mM 2-mercaptoethanol. Proteins were electrophoresed using 12% Bis-Tris polyacrylamide gels (Invitrogen). TAMRA ($\lambda_{excitation}$ = 555 nm and $\lambda_{emission}$ = 580 nm) was excited at 532 nm and detected with a 580 band-pass 30 nm filter. In-gel fluorescence images were acquired on a Typhoon 9400 instrument (GE Healthcare).

Detection of Proteins in Gels and Western Blots

Bicinchoninic acid protein quantification (Pierce) was used to equalize the amounts of proteins analyzed under different conditions. After dye labeling via the copper-catalyzed click reaction described above, proteins were washed with methanol to remove unreacted dye and then

electrophoresed on a Novex 12% Bis-Tris polyacrylamide gel (Invitrogen). Colloidal blue dye (Invitrogen) was used for nonspecific protein detection. For Western blots the proteins were transferred to a nitrocellulose membrane (GE Healthcare), and probed with a Myc-tag-Alexa Fluor 488 conjugate monoclonal antibody (Cell Signal Technologies) used at 1:1000 dilution in PBS with 0.2% v/w Tween20 (Sigma). Imaging of Western blots and gels was performed with a Typhoon 9400 molecular imager (GE Healthcare).

Copper Catalyzed Reaction of Alkyne-TAMRA Dye in Adherent Cells and Fluorescence Confocal Microscopy

Copper catalyzed azide-alkyne cycloaddition reaction and synthesis of requisite THPTA ligand were performed as described previously[26,33]. Adherent CHO cells seeded onto glass bottom tissue culture plates (MatTek) were incubated in fresh CHO media, as described above, supplemented with Anl at 1.5mM for 6 hours. Thereafter the cells were washed twice with PBS, fixed with 3.7% formaldehyde in PBS for 15 min at room temperature, followed by permeabilization with 4°C cold methanol for 10 min, and washed three times with PBS at room temperature. Labeling with alkyne-TAMRA (Invitrogen) was performed at room temperature in pH 7.4 PBS for 2 hours, using a final concentration of 0.1mM copper sulfate, 0.5mM THPTA ligand, 5mM sodium ascorbate, 5mM aminoguanidine and 10 µM of alkyne-TAMRA. To remove unreacted dye and reaction components the cells were washed five times at 30 min intervals with PBS. Cell nuclei were stained with 300nM DAPI in PBS for 30 minutes at room temperature and washed three times with PBS before imaging. Fluorescence confocal images were obtained on a Zeiss LSM 510 microscope.

Affinity Enrichment of Anl-labeled Proteins using Azadibenzocyclooctyne Functionalized Resin

Cells were lysed with 4% SDS in PBS supplemented with (EDTA)-free protease inhibitor (Roche) and 100 mM chloroacetamide. Alkylation step was used to reduce background reaction with free thiols. Cell lysate was sonicated using a tip sonicator to reduce viscosity and thereafter centrifuged at 14,000 rcf for 15 min to remove cellular debris. PBS was added to reduce SDS concentration to 1% and 1.5ml of lysate at 1mg/ml was reacted with 200µl of azadibenzocyclooctyne resin which is 50% slurry by volume (Click Chemistry Tools) for 20 min at room temperature on a rotating table. Direct conjugation of proteins to resin allows for stringent washing conditions without loss of labeled proteins and is advantageous because of lack of any streptavidin. Resin was washed 10 times with 1% SDS in PBS, proteins were reduced

using 1.25 µL of 500mM TCEP in water for 10 min, and alkylated using 5 µL of 500mM iodoactamide (fresh and dissolved in water) and incubate for 15 minutes at room temperature in the dark. Thereafter the resin was washed extensively first with 1% SDS in PBS for 20 times at room temperature, and an additional 20 times with PBS to remove residual SDS, and finally washed three times with Tris-HCl pH 8.5.

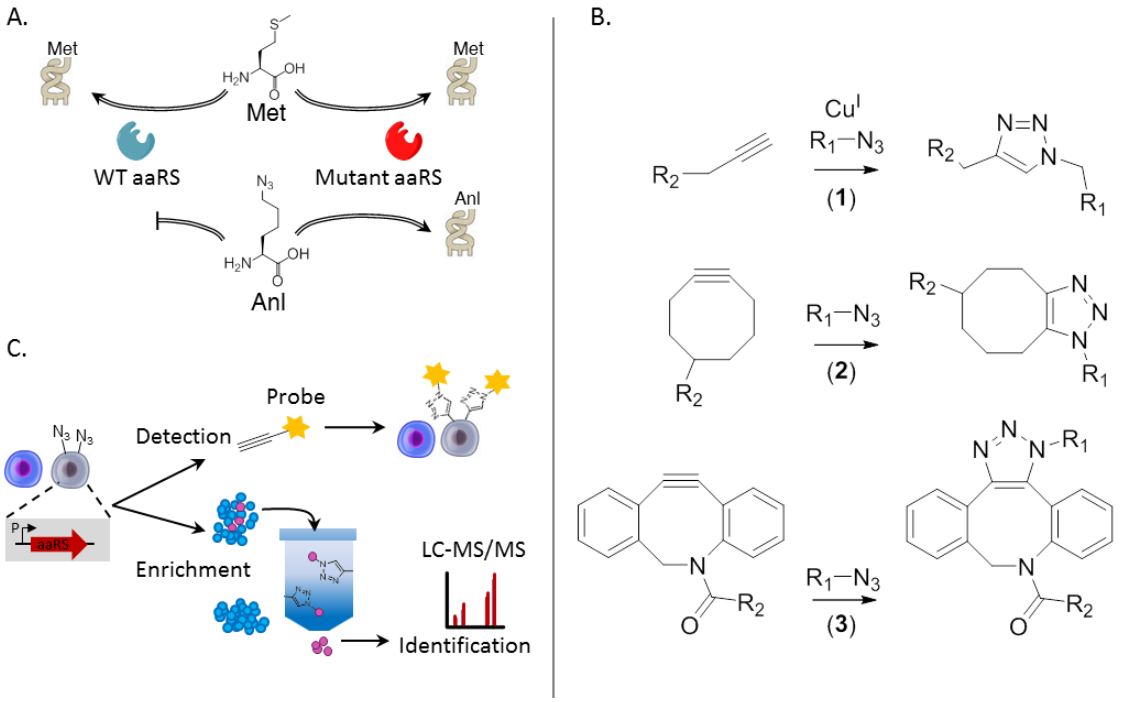
Trypsinization of Affinity Purified Proteins and Sample Preparation for Mass Spectrometry

After affinity enrichment, protein-bound resin was resuspended in 200µl Tris-HCl pH 8.5. On resin digestion was carried out with addition of 10µL of 0.1µg/µL of Lysyl endopeptidase(Wako) in 100mM Tris-HCl pH 8.5. Sample was incubated for 4 hrs at room temperature in the dark. Thereafter 20µL of 0.5µg/µL trypsin(Wako) in water was added and the sample was incubated in the dark overnight at room temperature. The eluent was centrifuged at 14000 rcf for 20 min using a 10kDa molecular weight cutoff spin filter (Pierce) to remove undigested proteins as well as trypsin which remains in the filter, the flow-through containing tryptic peptides was retained. This solution was acidified to 0.2% CF_3COOH. Solution containing peptides was desalted based on a previously published desalting protocol by Mann can coworkers[34] using a 3ml MILI-SPE C18-SD extraction disk cartridge (3M) as follows: 1) Cartridge was first washed with 1ml of CH_3OH and centrifuged at 1500 rcf for 1min. 2) Cartridge was washed once with 0.5ml of 0.1% CF_3COOH, 70% CH_3CN in water and centrifuged at 1500 rcf for 1min. 3) The cartridge was washed with 0.1% CF_3COOH in water and centrifuged at 1500 rcf for 1 min. 4) Peptide sample was loaded to the cartridge and passed through three times, each time the cartridge was centrifuged at 150 rcf for 3min. 5) The cartridge was washed twice with 0.5 ml of 0.1% CF_3COOH in water and centrifuged at 150 rcf for 3 min. 6) To elute the desalted peptides, cartridge was washed with 0.5ml of CH_3CN in water and centrifuged at 150 rcf for 3 min, the eluent containing desalted peptides was retained. The desalted peptides were lyophilized and stored at 4°C before analysis by mass spectrometry.

Mass Spectrometry

Analyses were performed with a hybrid LTQ-Orbitrap (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source connected to an EASY-nLC II instrument (Thermo Fisher Scientific). Fractionation of peptides was performed using a 15 cm reversed phase analytical column (75 µm ID) with 3 µm C18 beads (ReproSil-Pur C18-AQ) with a 60-minute gradient from 5% to 28% acetonitrile in 0.2% formic acid, over 50 minutes, followed by 10 minutes at 80 %

acetonitrile in 0.2% formic acid. The mass spectrometer was operated in data-dependent mode. Survey full scan mass spectra were acquired with a resolution of 60,000 at 400 m/z. The top ten most intense ions from the survey scan were isolated and, after the accumulation of 5000 ions, fragmented in the linear ion trap by collision induced dissociation. Precursor ion charge state screening was enabled and singly charged and unassigned charge states were rejected. The dynamic exclusion list was enabled containing trypsin peptide sequences, with a relative mass window of 10 ppm.



Scheme 4.1. Targeted expression of genetically encoded mutant aminoacyl tRNA synthetase (aaRS) for cell-selective BONCAT in mammalian cells. (a) The endogenous mammalian aaRS charges Met but not Anl. A mutant aaRS is used to charge Anl to $tRNA^{Met}$. (b) The azide side chain of Anl in proteins (R_1) is chemoselectively tagged through azide-alkyne cycloaddition reactions. Reaction 1 uses Cu^I as a catalyst, and reactions 2-3 are copper-free. (c) The controlled expression of mutant aaRS in cells of interest restricts labeling to those cells and enables the detection and identification of proteins in a cell-selective manner.

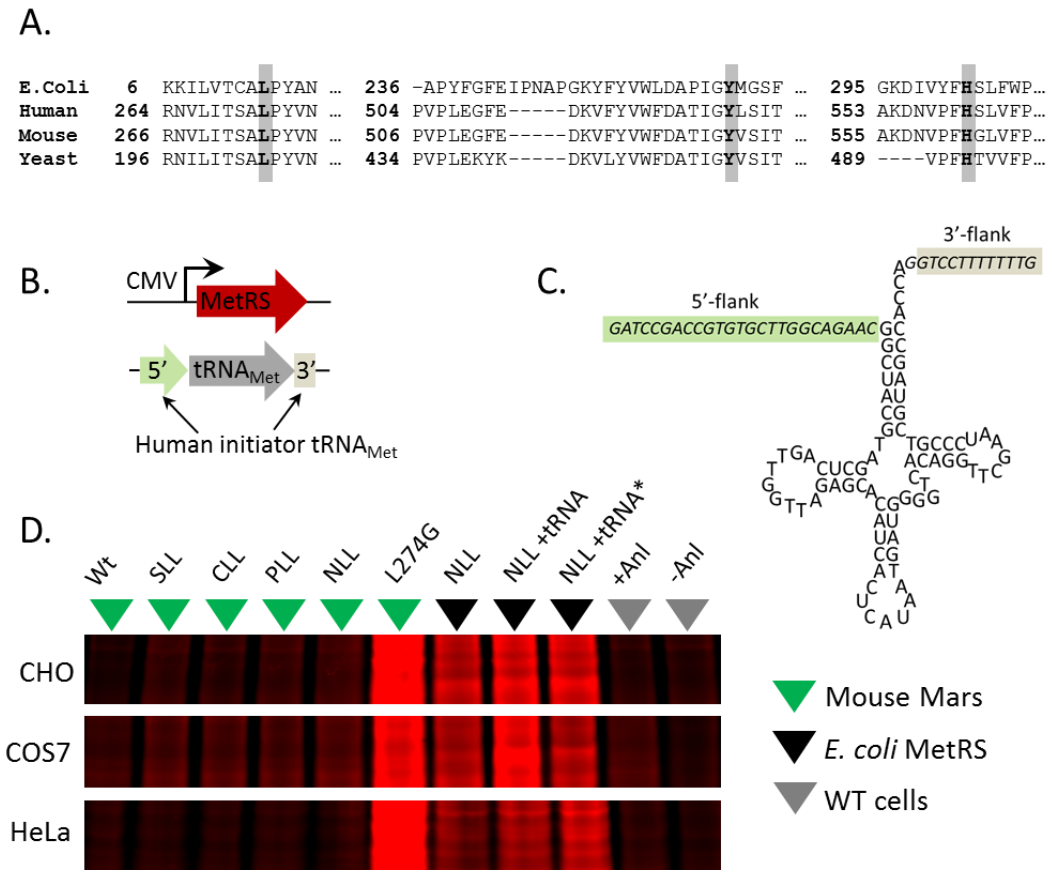


Figure 4.1. Identification of MetRS variants for the metabolic incorporation of Anl in mammalian cells. (a) Sequence alignment of various MetRS enzymes with positions 13/260/301 of the *E. coli* MetRS highlighted. (b) Constructs for the expression of MetRS and *E. Coli* $tRNA^{Met}$ under the CMV promoter and an internal promoter of eukaryotic tRNAs, respectively (c) *E. coli* $tRNA^{Met}$ coding sequence containing 5’ and 3’ flanking sequences from human initiator $tRNA^{Met}$. (d) In-gel fluorescence image shows TAMRA labeling, which indicates Anl incorporation in cellular proteins. tRNA* designates the presence of a 3’ CCA tail in the tRNA coding sequence. The wild type mouse MetRS enzyme is indicated by WT and the ±Anl conditions correspond to empty vector labeling controls.

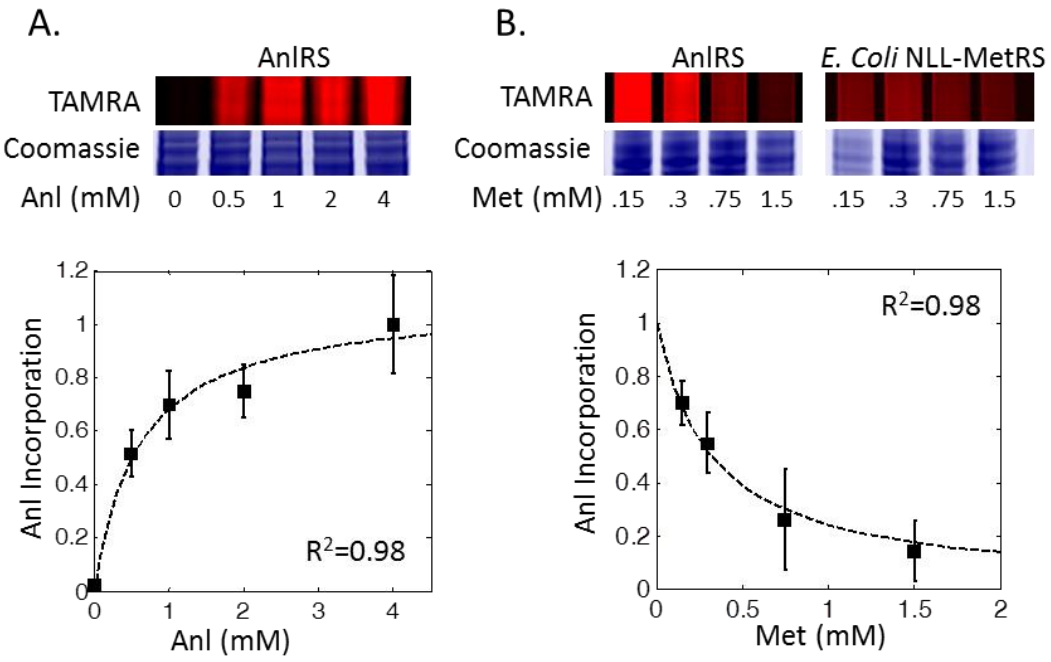


Figure 4.2. Measuring selectivity of AnlRS. In-gel TAMRA fluorescence and coomassie staining show Anl incorporation and total protein content, respectively. Fluorescence values were quantified and normalized to total protein levels. (a) Protein labeling in CHO media using the AnlRS at different Anl concentrations. (b) Protein labeling using both the AnlRS enzyme and the *E. coli* NLL-MetRS at increasing Met concentrations and 1.5 mM Anl. Plots on the bottom show Anl incorporation based on quantitated TAMRA fluorescence values. Error bars indicate standard deviation, and the dotted line represents line of best fit based on the equation above, and was used to calculate selectivity.

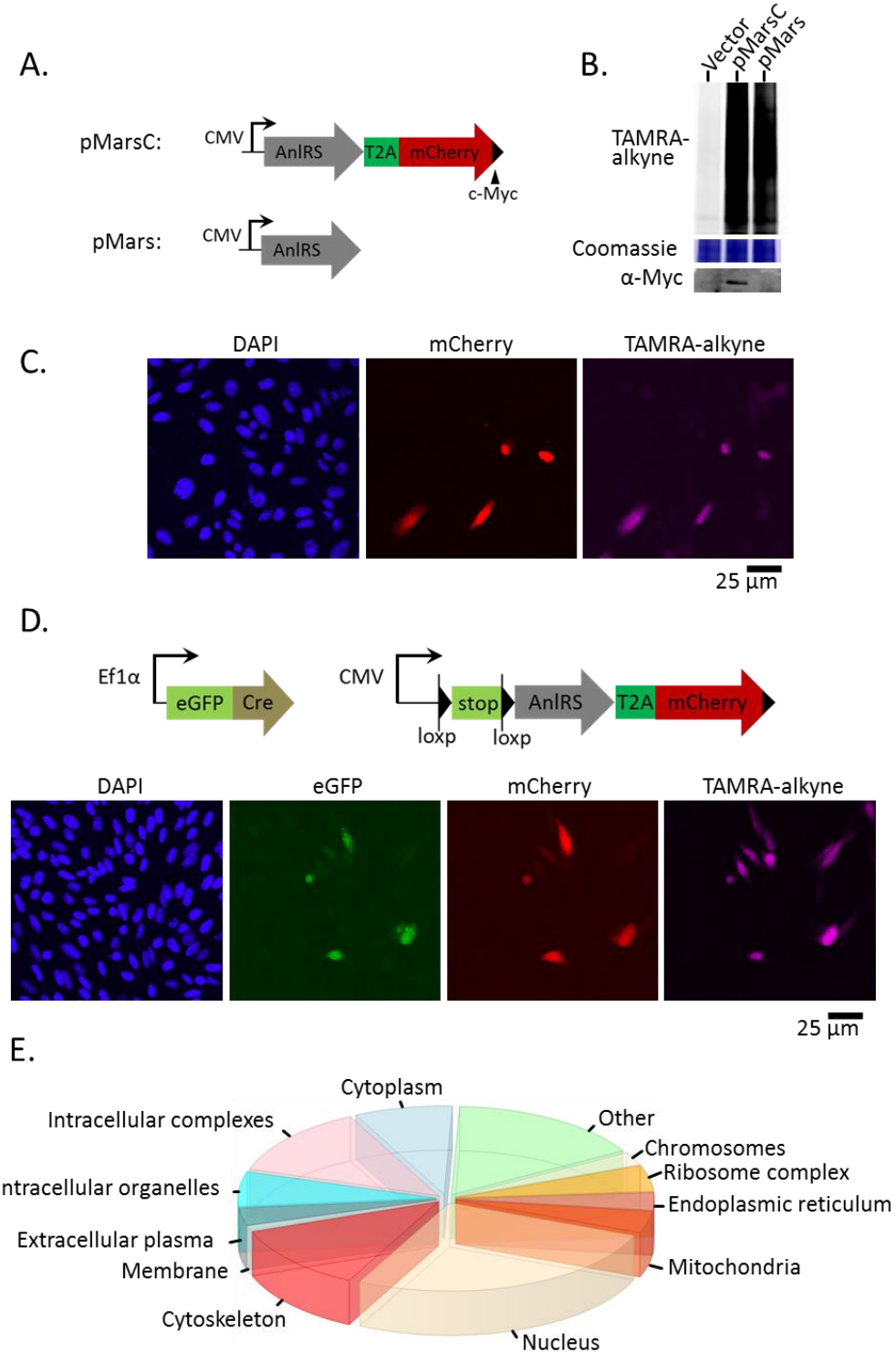
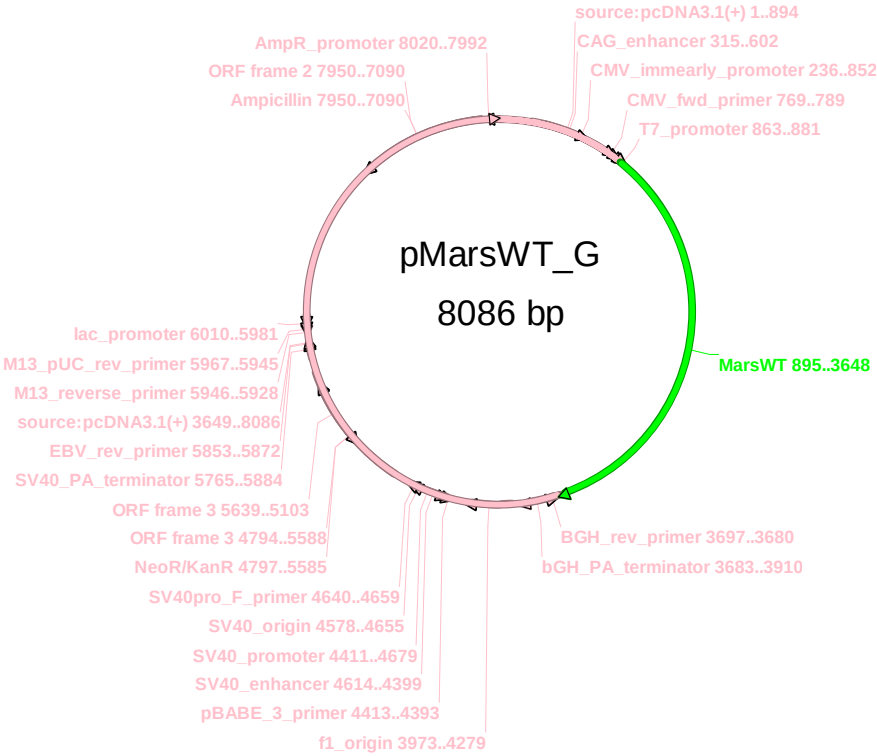


Figure 4.3. Cell-selective proteomic labeling and identification. (a) Design of pMarsC and pMars constructs for the expression of AnlRS and a C-terminal Myc-tagged mCherry protein to track synthetase expression. (b) Metabolic incorporation of Anl by pMars- and pMarsC-

transfected CHO cells. The in-gel fluorescence image on top shows TAMRA labeling, which indicates Anl incorporation. A western blot at the bottom using anti-Myc antibody shows the detection of mCherry at approximately 25 kDa (c) Fluorescence confocal microscopy image of CHO cells containing the pMarsC construct. DAPI shows the position of cell nuclei. (d) Fluorescence confocal microscopy image of CHO cells with a LoxP-containing AnlRS-T2A-mCherry sequence transiently transfected with an eGFP-Cre containing vector. The GFP signal corresponds to cells expressing Cre; the mCherry signal corresponds to cells in which AnlRS is expressed. (e) Identification of Anl-labeled proteins by tandem mass spectrometry and annotation of proteins into different cellular components using STRAP software[35]. Proteins in each group are listed in table S1.

Figure S4.1.

The pMarsWT_G vector for expression of wild-type Mars enzyme under CMV promoter control. Restriction enzymes are highlighted in yellow and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence.



GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCA
TCACCATATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGGCCGCGAGGG
CCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTCCTTACCCGG
CCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTGCCGATATTT
TTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTGGAATGGGAGGCAACAGAACTGC
AGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATACTTGGCCCA
CTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGCTGGGGACAC
AGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTGCAAGACCCAGCTTACCTCCCTG

AGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGAACCGTGTCAGCGAGCTGCAGAG
ACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGAAACAGCCACAGCCTCAGCCCCC
GCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTGAGGAGGACA
TCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAGCAGCATCCA
GTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCCTCCCCTATGTCAACAATGTCCCCCA
CCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTCGCCAGTGGA
ATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGACCAAGGCCATGGAGGAGGGCCTA
ACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCTACCGCTGGTTCGGCATATCGTT
CGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAGGACATCTTCCAGAGGTTGCTGA
CCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTCCTGGCTGAC
CGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGACAGGTGTGG
CAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTGTGGTGAGGT
CCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGGAAGACAGTG
CCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGCTTCGAGATGGCCTCAAGCCACG
ATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGGTATTTTACG
TCTGGTTTGATGCTACTATTGGCTACGTGTCCATCACAGCCAACTACACAGACCAATGGGAGAAATGGTGG
AAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATGTTCCCTTCCATGGCTTGGTCTT
TCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCACATCATTGCTACAGAGTACCTGA
ACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAGGATACAGGA
ATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGGGCCAGGACAGTGCCTTCTCCTG
GACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGCAACTTCATCAACAGAGCTGGCA
TGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAACCCCTGATGACAGACGCCTGGTG
GCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGATGCCTTGCG
CAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAATGAGCCCTGGAAACGGATTAAAG
GTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGTGAACATGGCTGCCTTGCTGTCT
GTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCCACCTGCAGC
CTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCACCGAATTGGCACAGTCAGTCCTT
TGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAGGCTAAAGGC
TCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGCACATACAAACGCTGACGGATGA
GGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCAGACAAGAACCAGGTTGCTGCAG
AGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAAACTCCTAAA
GGCAAGAAGAAAAAGTGATAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCAT
TAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTA
GTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTC
TTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGAT
TTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTG
TCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
CAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCC
GCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGA
AGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTC
GGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCG

TGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGT
TGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACC
TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACC
TGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGA
TCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCA
TGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGC
CGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTAC
CCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAA
TGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTG
GGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT
ATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA
TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTC
CGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
CTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT
CTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

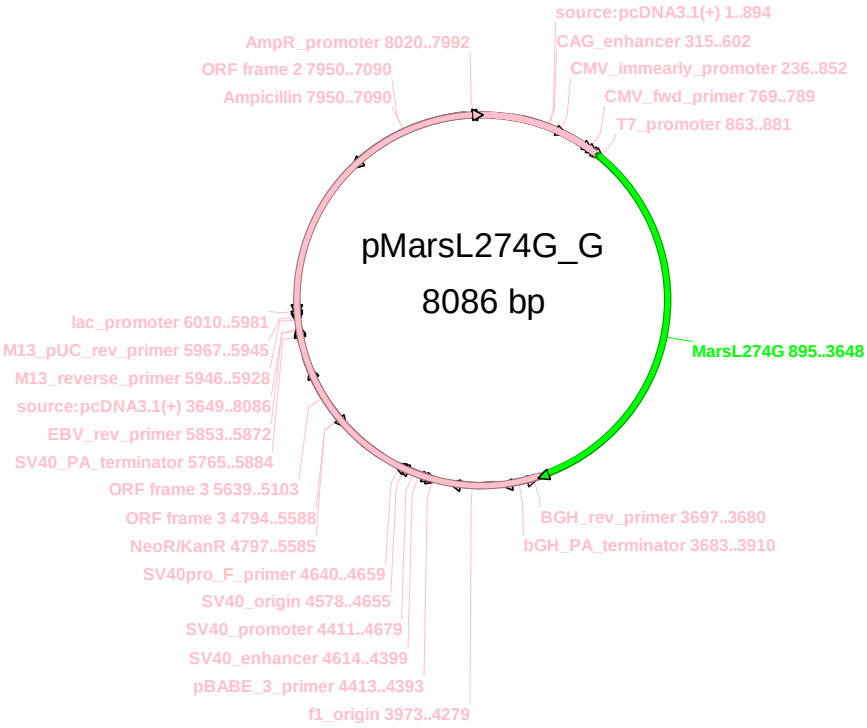
Protein Sequence

MAHHHHHHMRLFVSEGSPGSLPVLAAAARARGRAELLISTVGPEECVVPFLTRPKVPVLQLDSGNYLFSASAICRYFF
LLCGWEQDDLTNQWLEWEATELQPVLSAALHCLVVQGKKGEDILGPLRRVLTHIDHSLSRQNCPFLAGDTESLADIVL

WGALYPLLQDPAYLPEELGALQSWFQTLSTQEPCQRAAETVLKQQGVLALRLYLQKQPQPQPPPPEGRTVSNELEEEE
LATLSEEDIVTAVAAWEKGLESLPPLKLQQHPVLPVPGERNVLITSALPYVNNVPHLGNIIGCVLSADVFARYCRLRQ
WNTLYLCGTDEYGTATETKAMEEGLTPREICDKYHAIHADIYRWFGISFDTFGRTTTPQQTKITQDIFQRLLTRGFVL
RDTVEQLRCERCARFLADRFVEGVCPFCGYEEARGDQCDRCGKLINAIELKKPQCKICRSCPVVRSSQHLFLDLPKLE
KRLEDWLGKTVPGSDWTPNARFIIRSWLRDGLKPRCITRDLKWGTPVPLEGFEDKVFYVWFDATIGYVSITANYTDQW
EKWWKNPEQVDLYQFMAKDNVPFHGLVFPCSVLGAEDNYTLVKHIIATEYLNYEDGKFSKSRGIGVFGDMAKDTGIPA
DIWRFYLLYIRPEGQDSAFSWTDLLIKNNSELLNNLGNFINRAGMFVSKFFGGCVPEMALTPDDRRLVAHVSWELQHY
HQLLEKVRIRDALRSILTISRHGNQYIQVNEPWKRIKGGEMDRQRAGTVTGMAVNMAALLSVMLQPYMPTVSSTIQTQ
LQLPPAACRILATSFICTLPAGHRIGTVSPLFQKLENDQIENLRQRFGGGQAKGSPKPAAVEAVTAAGSQHIQTLTDE
VTKQGNVVRELKAQKADKNQVAAEVAKLLDLKKQLALAEGKPIETPKGKKKK-

Figure S4.2.

The pMarsL274G_G vector for expression of L274G mutant Mars enzyme under CMV promoter control. Restriction enzymes are highlighted in yellow and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence.



GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCA
TCACCATATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGGCCGCGAGGG
CCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTCCTTACCCGG
CCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTGCCGATATTT
TTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTGGAATGGGAGGCAACAGAACTGC
AGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATACTTGGCCCA
CTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGCTGGGGACAC
AGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTGCAAGACCCAGCTTACCTCCCTG

AGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGAACCGTGTCAGCGAGCTGCAGAG
ACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGAAACAGCCACAGCCTCAGCCCCC
GCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTGAGGAGGACA
TCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAGCAGCATCCA
GTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCGGACCCTATGTCAACAATGTCCCCCA
CCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTCGCCAGTGGA
ATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGACCAAGGCCATGGAGGAGGGCCTA
ACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCTACCGCTGGTTCGGCATATCGTT
CGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAGGACATCTTCCAGAGGTTGCTGA
CCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTCCTGGCTGAC
CGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGACAGGTGTGG
CAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTGTGGTGAGGT
CCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGGAAGACAGTG
CCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGCTTCGAGATGGCCTCAAGCCACG
ATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGGTATTTTACG
TCTGGTTTGATGCTACTATTGGCTACGTGTCCATCACAGCCAACTACACAGACCAATGGGAGAAATGGTGG
AAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATGTTCCCTTCCATGGCTTGGTCTT
TCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCACATCATTGCTACAGAGTACCTGA
ACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAGGATACAGGA
ATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGGGCCAGGACAGTGCCTTCTCCTG
GACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGCAACTTCATCAACAGAGCTGGCA
TGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAACCCCTGATGACAGACGCCTGGTG
GCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGATGCCTTGCG
CAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAATGAGCCCTGGAAACGGATTAAAG
GTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGTGAACATGGCTGCCTTGCTGTCT
GTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCCACCTGCAGC
CTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCACCGAATTGGCACAGTCAGTCCTT
TGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAGGCTAAAGGC
TCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGCACATACAAACGCTGACGGATGA
GGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCAGACAAGAACCAGGTTGCTGCAG
AGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAAACTCCTAAA
GGCAAGAAGAAAAAGTGATAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCAT
TAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTA
GTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTC
TTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGAT
TTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTG
TCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
CAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCC
GCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGA
AGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTC
GGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCG

TGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGT
TGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACC
TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACC
TGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGA
TCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCA
TGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGC
CGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTAC
CCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAA
TGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTG
GGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT
ATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA
TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTC
CGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
CTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT
CTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

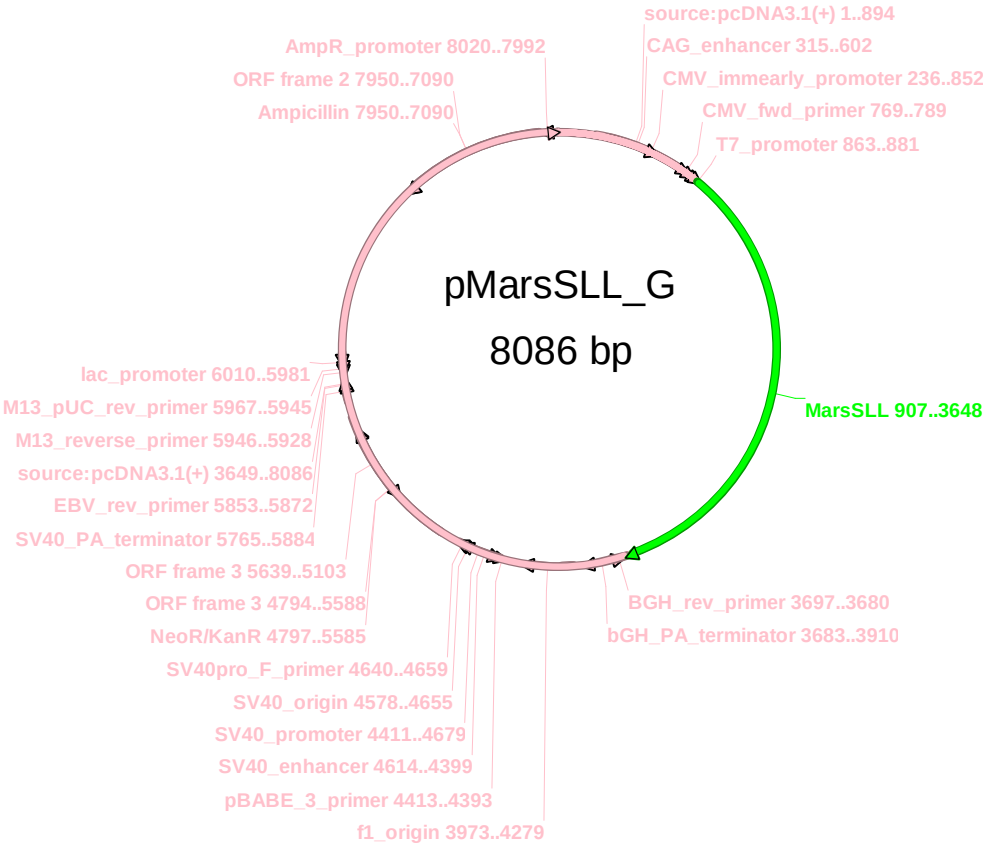
Protein Sequence

MAHHHHHHMRLFVSEGSPGSLPVLAAAARARGRAELLISTVGPEECVVPFLTRPKVPVLQLDSGNYLFSAS
AICRYFFLLCGWEQDDLTNQWLEWEATELQPVLSAALHCLVVQGKKGEDILGPLRRVLTHIDHSLSRQNCP

FLAGDTESLADIVLWGALYPLLQDPAYLPEELGALQSWFQTLSTQEPCQRAAETVLKQQGVLALRLYLQKQ
PQPQPPPPEGRTVSNELEEEELATLSEEDIVTAVAAWEKGLESLPPLKLQQHPVLPVPGERNVLITSAGPY
VNNVPHLGNIIGCVLSADVFARYCRLRQWNTLYLCGTDEYGTATETKAMEEGLTPREICDKYHAIHADIYR
WFGISFDTFGRTTTPQQTKITQDIFQRLLTRGFVLRDTVEQLRCERCARFLADRFVEGVCPFCGYEEARGD
QCDRCGKLINAIELKKPQCKICRSCPVVRSSQHLFLDLPKLEKRLEDWLGKTVPGSDWTPNARFIIRSWLR
DGLKPRCITRDLKWGTPVPLEGFEDKVFYVWFDATIGYVSITANYTDQWEKWWKNPEQVDLYQFMAKDNVP
FHGLVFPCSVLGAEDNYTLVKHIIATEYLNYEDGKFSKSRGIGVFGDMAKDTGIPADIWRFYLLYIRPEGQ
DSAFSWTDLLIKNNSELLNNLGNFINRAGMFVSKFFGGCVPEMALTPDDRRLVAHVSWELQHYHQLLEKVR
IRDALRSILTISRHGNQYIQVNEPWKRIKGGEMDRQRAGTVTGMAVNMAALLSVMLQPYMPTVSSTIQTQL
QLPPAACRILATSFICTLPAGHRIGTVSPLFQKLENDQIENLRQRFGGGQAKGSPKPAAVEAVTAAGSQHI
QTLTDEVTKQGNVVRELKAQKADKNQVAAEVAKLLDLKKQLALAEGKPIETPKGKKKK-

Figure S4.3.

The pMarsSLL_G vector for expression of SLL Mars enzyme under CMV promoter control. Restriction enzymes are highlighted in yellow and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence.



GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCA
TCACCATATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGGCCGCGAGGG
CCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTCCTTACCCGG
CCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTGCCGATATTT
TTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTGGAATGGGAGGCAACAGAACTGC
AGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATACTTGGCCCA

CTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGCTGGGGACAC
AGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTGCAAGACCCAGCTTACCTCCCTG
AGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGAACCGTGTCAGCGAGCTGCAGAG
ACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGAAACAGCCACAGCCTCAGCCCCC
GCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTGAGGAGGACA
TCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAGCAGCATCCA
GTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCTCCCCCTATGTCAACAATGTCCCCCA
CCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTCGCCAGTGGA
ATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGACCAAGGCCATGGAGGAGGGCCTA
ACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCTACCGCTGGTTCGGCATATCGTT
CGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAGGACATCTTCCAGAGGTTGCTGA
CCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTCCTGGCTGAC
CGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGACAGGTGTGG
CAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTGTGGTGAGGT
CCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGGAAGACAGTG
CCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGCTTCGAGATGGCCTCAAGCCACG
ATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGGTATTTTACG
TCTGGTTTGATGCTACTATTGGCCTGGTGTCCATCACAGCCAACTACACAGACCAATGGGAGAAATGGTGG
AAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATGTTCCCTTCCTGGGCTTGGTCTT
TCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCACATCATTGCTACAGAGTACCTGA
ACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAGGATACAGGA
ATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGGGCCAGGACAGTGCCTTCTCCTG
GACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGCAACTTCATCAACAGAGCTGGCA
TGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAACCCCTGATGACAGACGCCTGGTG
GCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGATGCCTTGCG
CAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAATGAGCCCTGGAAACGGATTAAAG
GTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGTGAACATGGCTGCCTTGCTGTCT
GTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCCACCTGCAGC
CTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCACCGAATTGGCACAGTCAGTCCTT
TGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAGGCTAAAGGC
TCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGCACATACAAACGCTGACGGATGA
GGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCAGACAAGAACCAGGTTGCTGCAG
AGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAAACTCCTAAA
GGCAAGAAGAAAAAGTGATAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCAT
TAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTA
GTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTC
TTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGAT
TTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTG
TCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
CAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCC
GCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGA
AGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTC

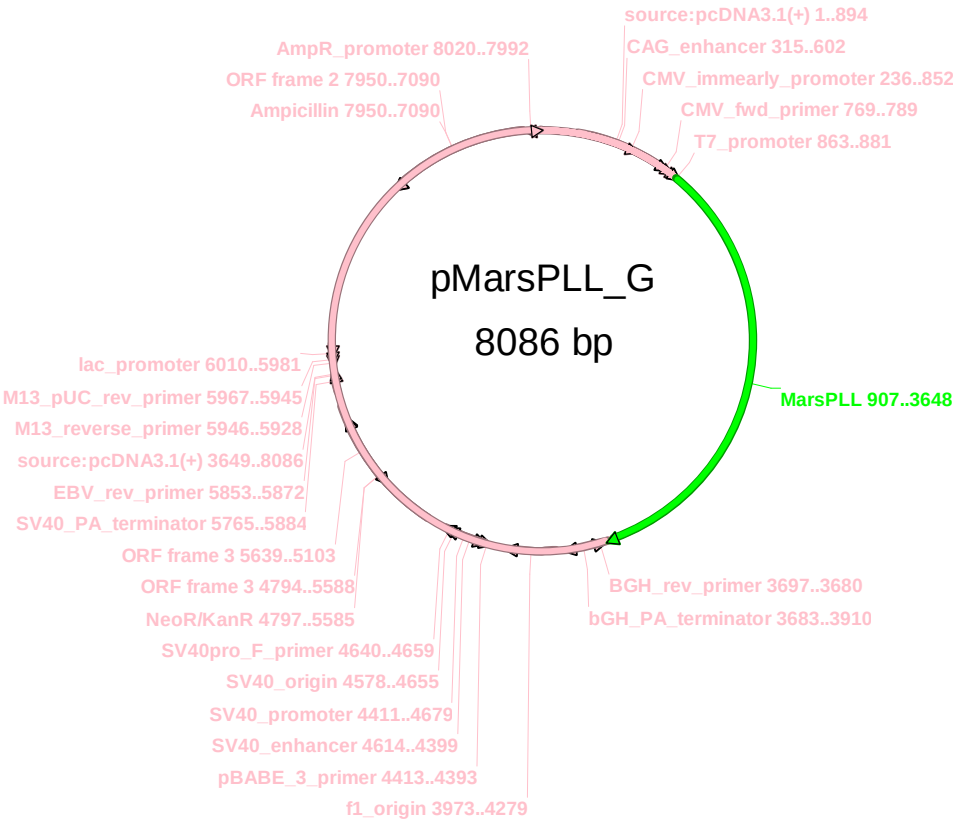
GGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCG
TGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGT
TGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACC
TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACC
TGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGA
TCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCA
TGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGC
CGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTAC
CCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAA
TGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTG
GGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT
ATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA
TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTC
CGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
CTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT
CTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

MAHHHHHHMRLFVSEGSPGSLPVLAAAARARGRAELLISTVGPEECVVPFLTRPKVPVLQLDSGNYLFSASAICRYFF
LLCGWEQDDLTNQWLEWEATELQPVLSAALHCLVVQGKKGEDILGPLRRVLTHIDHSLSRQNCPFLAGDTESLADIVL

WGALYPLLQDPAYLPEELGALQSWFQTLSTQEPCQRAAETVLKQQGVLALRLYLQKQPQPQPPPPEGRTVSNELEEEE
LATLSEEDIVTAVAAWEKGLESLPPLKLQQHPVLPVPGERNVLITSASPYVNNVPHLGNIIGCVLSADVFARYCRLRQ
WNTLYLCGTDEYGTATETKAMEEGLTPREICDKYHAIHADIYRWFGISFDTFGRTTTPQQTKITQDIFQRLLTRGFVL
RDTVEQLRCERCARFLADRFVEGVCPFCGYEEARGDQCDRCGKLINAIELKKPQCKICRSCPVVRSSQHLFLDLPKLE
KRLEDWLGKTVPGSDWTPNARFIIRSWLRDGLKPRCITRDLKWGTPVPLEGFEDKVFYVWFDATIGLVSITANYTDQW
EKWWKNPEQVDLYQFMAKDNVPFLGLVFPCSVLGAEDNYTLVKHIIATEYLNYEDGKFSKSRGIGVFGDMAKDTGIPA
DIWRFYLLYIRPEGQDSAFSWTDLLIKNNSELLNNLGNFINRAGMFVSKFFGGCVPEMALTPDDRRLVAHVSWELQHY
HQLLEKVRIRDALRSILTISRHGNQYIQVNEPWKRIKGGEMDRQRAGTVTGMAVNMAALLSVMLQPYMPTVSSTIQTQ
LQLPPAACRILATSFICTLPAGHRIGTVSPLFQKLENDQIENLRQRFGGGQAKGSPKPAAVEAVTAAGSQHIQTLTDE
VTKQGNVVRELKAQKADKNQVAAEVAKLLDLKKQLALAEGKPIETPKGKKKK--

Figure S4.4.

The pMarsPLL_G vector for expression of SLL Mars enzyme under CMV promoter control. Restriction enzymes are highlighted in yellow and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence.



GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCA
TCACCATATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGGCCGCGAGGG
CCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTCCTTACCCGG
CCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTGCCGATATTT
TTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTGGAATGGGAGGCAACAGAACTGC
AGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATACTTGGCCCA

CTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGCTGGGGACAC
AGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTGCAAGACCCAGCTTACCTCCCTG
AGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGAACCGTGTCAGCGAGCTGCAGAG
ACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGAAACAGCCACAGCCTCAGCCCCC
GCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTGAGGAGGACA
TCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAGCAGCATCCA
GTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCCCACCCTATGTCAACAATGTCCCCCA
CCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTCGCCAGTGGA
ATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGACCAAGGCCATGGAGGAGGGCCTA
ACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCTACCGCTGGTTCGGCATATCGTT
CGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAGGACATCTTCCAGAGGTTGCTGA
CCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTCCTGGCTGAC
CGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGACAGGTGTGG
CAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTGTGGTGAGGT
CCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGGAAGACAGTG
CCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGCTTCGAGATGGCCTCAAGCCACG
ATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGGTATTTTACG
TCTGGTTTGATGCTACTATTGGCCTGGTGTCCATCACAGCCAACTACACAGACCAATGGGAGAAATGGTGG
AAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATGTTCCCTTCCTGGGCTTGGTCTT
TCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCACATCATTGCTACAGAGTACCTGA
ACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAGGATACAGGA
ATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGGGCCAGGACAGTGCCTTCTCCTG
GACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGCAACTTCATCAACAGAGCTGGCA
TGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAACCCCTGATGACAGACGCCTGGTG
GCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGATGCCTTGCG
CAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAATGAGCCCTGGAAACGGATTAAAG
GTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGTGAACATGGCTGCCTTGCTGTCT
GTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCCACCTGCAGC
CTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCACCGAATTGGCACAGTCAGTCCTT
TGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAGGCTAAAGGC
TCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGCACATACAAACGCTGACGGATGA
GGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCAGACAAGAACCAGGTTGCTGCAG
AGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAAACTCCTAAA
GGCAAGAAGAAAAAGTGATAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCAT
TAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTA
GTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTC
TTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGAT
TTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTG
TCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
CAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCC
GCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGA
AGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTC

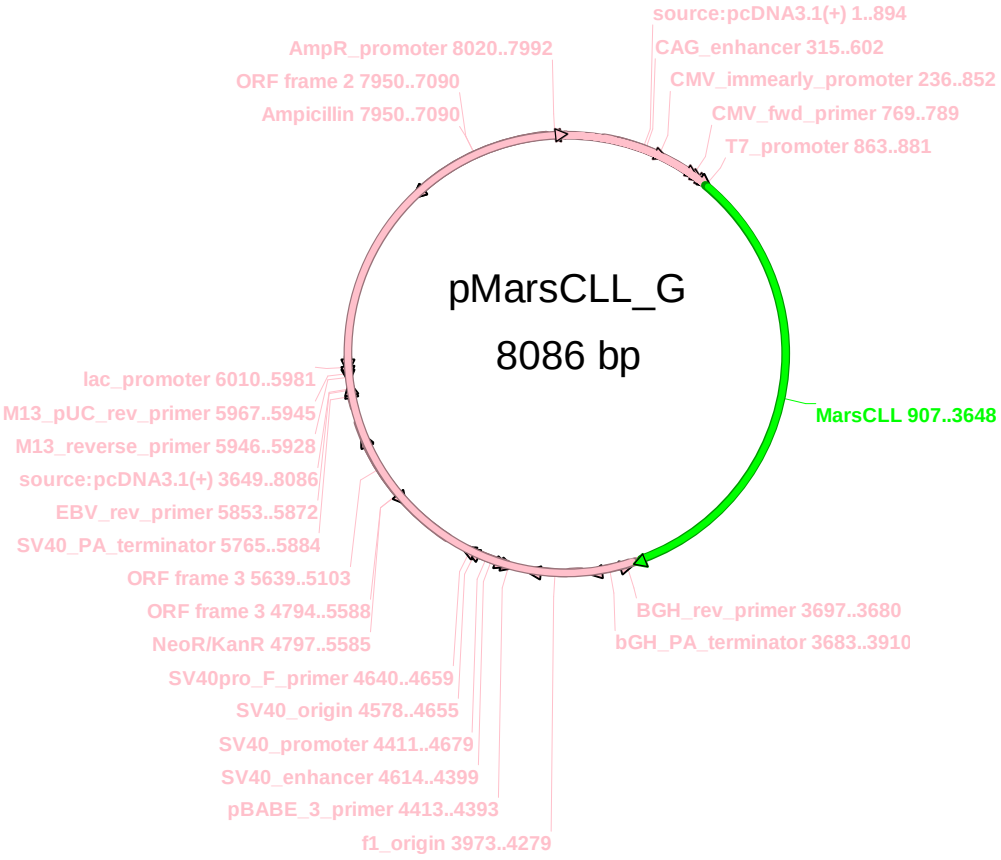
GGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCG
TGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGT
TGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACC
TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACC
TGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGA
TCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCA
TGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGC
CGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTAC
CCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAA
TGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTG
GGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT
ATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA
TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTC
CGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
CTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT
CTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

MAHHHHHHMRLFVSEGSPGSLPVLAAAARARGRAELLISTVGPEECVVPFLTRPKVPVLQLDSGNYLFSASAICRYFF
LLCGWEQDDLTNQWLEWEATELQPVLSAALHCLVVQGKKGEDILGPLRRVLTHIDHSLSRQNCPFLAGDTESLADIVL

WGALYPLLQDPAYLPEELGALQSWFQTLSTQEPCQRAAETVLKQQGVLALRLYLQKQPQPQPPPPEGRTVSNELEEEE
LATLSEEDIVTAVAAWEKGLESLPPLKLQQHPVLPVPGERNVLITSAPPYVNNVPHLGNIIGCVLSADVFARYCRLRQ
WNTLYLCGTDEYGTATETKAMEEGLTPREICDKYHAIHADIYRWFGISFDTFGRTTTPQQTKITQDIFQRLLTRGFVL
RDTVEQLRCERCARFLADRFVEGVCPFCGYEEARGDQCDRCGKLINAIELKKPQCKICRSCPVVRSSQHLFLDLPKLE
KRLEDWLGKTVPGSDWTPNARFIIRSWLRDGLKPRCITRDLKWGTPVPLEGFEDKVFYVWFDATIGLVSITANYTDQW
EKWWKNPEQVDLYQFMAKDNVPFLGLVFPCSVLGAEDNYTLVKHIIATEYLNYEDGKFSKSRGIGVFGDMAKDTGIPA
DIWRFYLLYIRPEGQDSAFSWTDLLIKNNSELLNNLGNFINRAGMFVSKFFGGCVPEMALTPDDRRLVAHVSWELQHY
HQLLEKVRIRDALRSILTISRHGNQYIQVNEPWKRIKGGEMDRQRAGTVTGMAVNMAALLSVMLQPYMPTVSSTIQTQ
LQLPPAACRILATSFICTLPAGHRIGTVSPLFQKLENDQIENLRQRFGGGQAKGSPKPAAVEAVTAAGSQHIQTLTDE
VTKQGNVVRELKAQKADKNQVAAEVAKLLDLKKQLALAEGKPIETPKGKKKK--

Figure S4.5.

The pMarsCLL_G vector for expression of SLL Mars enzyme under CMV promoter control. Restriction enzymes are highlighted in yellow and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence.



GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCA
TCACCATATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGGCCGCGAGGG
CCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTCCTTACCCGG
CCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTGCCGATATTT
TTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTGGAATGGGAGGCAACAGAACTGC

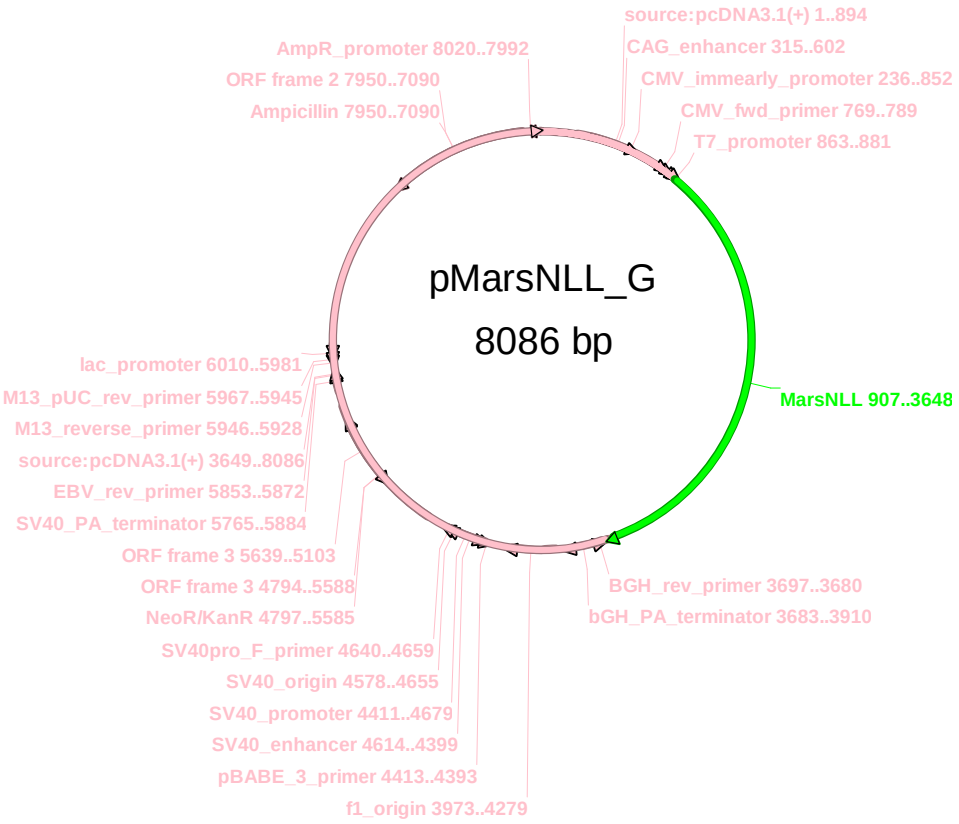
AGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATACTTGGCCCA
CTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGCTGGGGACAC
AGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTGCAAGACCCAGCTTACCTCCCTG
AGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGAACCGTGTCAGCGAGCTGCAGAG
ACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGAAACAGCCACAGCCTCAGCCCCC
GCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTGAGGAGGACA
TCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAGCAGCATCCA
GTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCTGTCCCTATGTCAACAATGTCCCCCA
CCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTCGCCAGTGGA
ATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGACCAAGGCCATGGAGGAGGGCCTA
ACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCTACCGCTGGTTCGGCATATCGTT
CGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAGGACATCTTCCAGAGGTTGCTGA
CCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTCCTGGCTGAC
CGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGACAGGTGTGG
CAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTGTGGTGAGGT
CCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGGAAGACAGTG
CCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGCTTCGAGATGGCCTCAAGCCACG
ATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGGTATTTTACG
TCTGGTTTGATGCTACTATTGGCCTGGTGTCCATCACAGCCAACTACACAGACCAATGGGAGAAATGGTGG
AAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATGTTCCCTTCCTGGGCTTGGTCTT
TCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCACATCATTGCTACAGAGTACCTGA
ACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAGGATACAGGA
ATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGGGCCAGGACAGTGCCTTCTCCTG
GACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGCAACTTCATCAACAGAGCTGGCA
TGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAACCCCTGATGACAGACGCCTGGTG
GCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGATGCCTTGCG
CAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAATGAGCCCTGGAAACGGATTAAAG
GTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGTGAACATGGCTGCCTTGCTGTCT
GTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCCACCTGCAGC
CTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCACCGAATTGGCACAGTCAGTCCTT
TGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAGGCTAAAGGC
TCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGCACATACAAACGCTGACGGATGA
GGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCAGACAAGAACCAGGTTGCTGCAG
AGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAAACTCCTAAA
GGCAAGAAGAAAAAGTGATAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCAT
TAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTA
GTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTC
TTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGAT
TTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTG
TCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
CAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCC
GCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGA

AGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTC
GGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCG
TGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGT
TGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACC
TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACC
TGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGA
TCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCA
TGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGC
CGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTAC
CCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAA
TGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTG
GGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT
ATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA
TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTC
CGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
CTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT
CTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

MAHHHHHHMRLFVSEGSPGSLPVLAAAARARGRAELLISTVGPEECVVPFLTRPKVPVLQLDSGNYLFSASAICRYFF
LLCGWEQDDLTNQWLEWEATELQPVLSAALHCLVVQGKKGEDILGPLRRVLTHIDHSLSRQNCPFLAGDTESLADIVL
WGALYPLLQDPAYLPEELGALQSWFQTLSTQEPCQRAAETVLKQQGVLALRLYLQKQPQPQPPPPEGRTVSNELEEEE
LATLSEEDIVTAVAAWEKGLESLPPLKLQQHPVLPVPGERNVLITSACPYVNNVPHLGNIIGCVLSADVFARYCRLRQ
WNTLYLCGTDEYGTATETKAMEEGLTPREICDKYHAIHADIYRWFGISFDTFGRTTTPQQTKITQDIFQRLLTRGFVL
RDTVEQLRCERCARFLADRFVEGVCPFCGYEEARGDQCDRCGKLINAIELKKPQCKICRSCPVVRSSQHLFLDLPKLE
KRLEDWLGKTVPGSDWTPNARFIIRSWLRDGLKPRCITRDLKWGTPVPLEGFEDKVFYVWFDATIGLVSITANYTDQW
EKWWKNPEQVDLYQFMAKDNVPFLGLVFPCSVLGAEDNYTLVKHIIATEYLNYEDGKFSKSRGIGVFGDMAKDTGIPA
DIWRFYLLYIRPEGQDSAFSWTDLLIKNNSELLNNLGNFINRAGMFVSKFFGGCVPEMALTPDDRRLVAHVSWELQHY
HQLLEKVRIRDALRSILTISRHGNQYIQVNEPWKRIKGGEMDRQRAGTVTGMAVNMAALLSVMLQPYMPTVSSTIQTQ
LQLPPAACRILATSFICTLPAGHRIGTVSPLFQKLENDQIENLRQRFGGGQAKGSPKPAAVEAVTAAGSQHIQTLTDE
VTKQGNVVRELKAQKADKNQVAAEVAKLLDLKKQLALAEGKPIETPKGKKKK--

Figure S4.6.

The pMarsNLL_G vector for expression of SLL Mars enzyme under CMV promoter control. Restriction enzymes are highlighted in yellow and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence.



GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCA
TCACCATATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGGCCGCGAGGG
CCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTCCTTACCCGG
CCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTGCCGATATTT
TTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTGGAATGGGAGGCAACAGAACTGC
AGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATACTTGGCCCA

CTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGCTGGGGACAC
AGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTGCAAGACCCAGCTTACCTCCCTG
AGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGAACCGTGTCAGCGAGCTGCAGAG
ACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGAAACAGCCACAGCCTCAGCCCCC
GCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTGAGGAGGACA
TCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAGCAGCATCCA
GTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCAACCCCTATGTCAACAATGTCCCCCA
CCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTCGCCAGTGGA
ATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGACCAAGGCCATGGAGGAGGGCCTA
ACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCTACCGCTGGTTCGGCATATCGTT
CGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAGGACATCTTCCAGAGGTTGCTGA
CCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTCCTGGCTGAC
CGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGACAGGTGTGG
CAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTGTGGTGAGGT
CCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGGAAGACAGTG
CCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGCTTCGAGATGGCCTCAAGCCACG
ATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGGTATTTTACG
TCTGGTTTGATGCTACTATTGGCCTGGTGTCCATCACAGCCAACTACACAGACCAATGGGAGAAATGGTGG
AAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATGTTCCCTTCCTGGGCTTGGTCTT
TCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCACATCATTGCTACAGAGTACCTGA
ACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAGGATACAGGA
ATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGGGCCAGGACAGTGCCTTCTCCTG
GACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGCAACTTCATCAACAGAGCTGGCA
TGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAACCCCTGATGACAGACGCCTGGTG
GCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGATGCCTTGCG
CAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAATGAGCCCTGGAAACGGATTAAAG
GTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGTGAACATGGCTGCCTTGCTGTCT
GTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCCACCTGCAGC
CTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCACCGAATTGGCACAGTCAGTCCTT
TGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAGGCTAAAGGC
TCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGCACATACAAACGCTGACGGATGA
GGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCAGACAAGAACCAGGTTGCTGCAG
AGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAAACTCCTAAA
GGCAAGAAGAAAAAGTGATAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCAT
TAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTA
GTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTC
TTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGAT
TTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTG
TCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
CAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCC
GCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGA
AGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTC

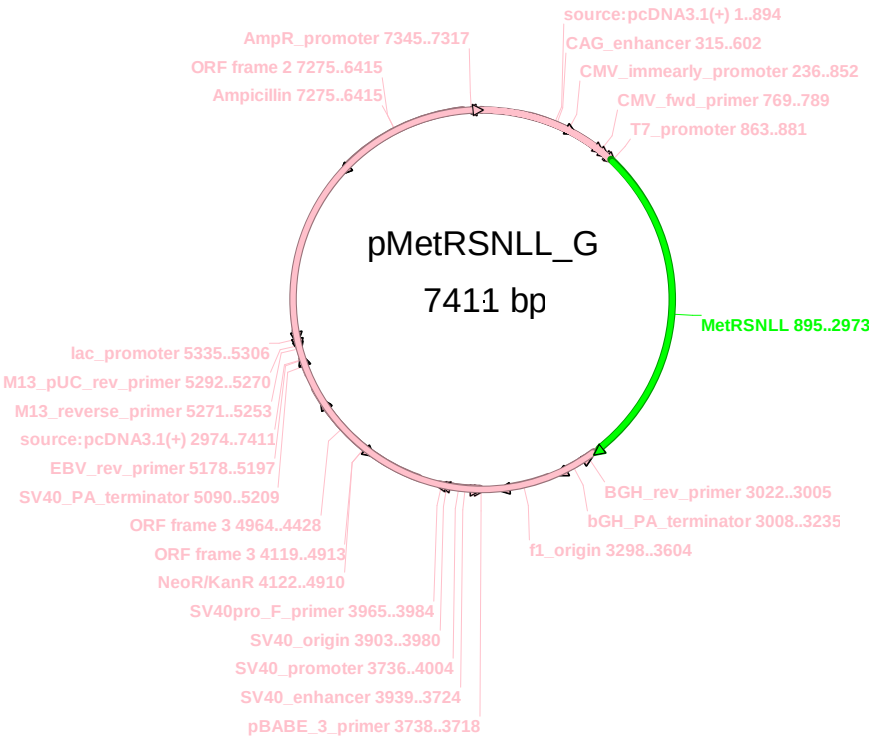
GGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCG
TGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGT
TGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACC
TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACC
TGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGA
TCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCA
TGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGC
CGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTAC
CCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAA
TGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTG
GGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT
ATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA
TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTT
TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTC
CGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
CTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT
CTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

MAHHHHHHMRLFVSEGSPGSLPVLAAAARARGRAELLISTVGPEECVVPFLTRPKVPVLQLDSGNYLFSASAICRYFF
LLCGWEQDDLTNQWLEWEATELQPVLSAALHCLVVQGKKGEDILGPLRRVLTHIDHSLSRQNCPFLAGDTESLADIVL

WGALYPLLQDPAYLPEELGALQSWFQTLSTQEPCQRAAETVLKQQGVLALRLYLQKQPQPQPPPPEGRTVSNELEEEE
LATLSEEDIVTAVAAWEKGLESLPPLKLQQHPVLPVPGERNVLITSANPYVNNVPHLGNIIGCVLSADVFARYCRLRQ
WNTLYLCGTDEYGTATETKAMEEGLTPREICDKYHAIHADIYRWFGISFDTFGRTTTPQQTKITQDIFQRLLTRGFVL
RDTVEQLRCERCARFLADRFVEGVCPFCGYEEARGDQCDRCGKLINAIELKKPQCKICRSCPVVRSSQHLFLDLPKLE
KRLEDWLGKTVPGSDWTPNARFIIRSWLRDGLKPRCITRDLKWGTPVPLEGFEDKVFYVWFDATIGLVSITANYTDQW
EKWWKNPEQVDLYQFMAKDNVPFLGLVFPCSVLGAEDNYTLVKHIIATEYLNYEDGKFSKSRGIGVFGDMAKDTGIPA
DIWRFYLLYIRPEGQDSAFSWTDLLIKNNSELLNNLGNFINRAGMFVSKFFGGCVPEMALTPDDRRLVAHVSWELQHY
HQLLEKVRIRDALRSILTISRHGNQYIQVNEPWKRIKGGEMDRQRAGTVTGMAVNMAALLSVMLQPYMPTVSSTIQTQ
LQLPPAACRILATSFICTLPAGHRIGTVSPLFQKLENDQIENLRQRFGGGQAKGSPKPAAVEAVTAAGSQHIQTLTDE
VTKQGNVVRELKAQKADKNQVAAEVAKLLDLKKQLALAEGKPIETPKGKKKK--

Figure S4.7.

The pMetRSNLL_G vector for expression of the *E. Coli* NLL-MetRS under CMV promoter control. Restriction enzymes are highlighted in yellow, start codon in red, and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence.



GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCA
TCACCATACTCAAGTCGCGAAGAAAATTCTGGTGACGTGCGCAAACCCGTACGCTAACGGCTCAATCCACC
TCGGCCATATGCTGGAGCACATCCAGGCTGATGTCTGGGTCCGTTACCAGCGAATGCGCGGCCACGAGGTC
AACTTCATCTGCGCCGACGATGCCCACGGTACACCGATCATGCTGAAAGCTCAGCAGCTTGGTATCACCCC
GGAGCAGATGATTGGCGAAATGAGTCAGGAGCATCAGACTGATTTCGCAGGCTTTAACATCAGCTATGACA
ACTATCACTCGACGCACAGCGAAGAGAACCGCCAGTTGTCAGAACTTATCTACTCTCGCCTGAAAGAAAAC
GGTTTTATTAAAAACCGCACCATCTCTCAGCTGTACGATCCGGAAAAAGGCATGTTCCTGCCGGACCGTTT

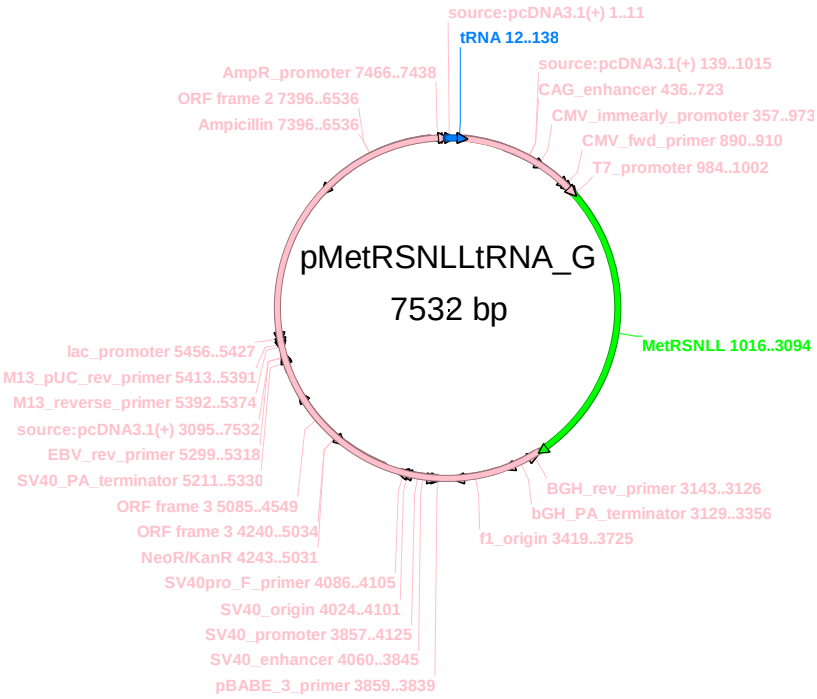
TGTGAAAGGCACCTGCCCGAAATGTAAATCCCCGGATCAATACGGCGATAACTGCGAAGTCTGCGGCGCGA
CCTACAGCCCGACTGAACTGATCGAGCCGAAATCGGTGGTTTCTGGCGCTACGCCGGTAATGCGTGATTCT
GAACACTTCTTCTTTGATCTGCCCTCTTTCAGCGAAATGTTGCAGGCATGGACCCGCAGCGGTGCGTTGCA
GGAGCAGGTGGCAAATAAAATGCAGGAGTGGTTTGAATCTGGCCTGCAACAGTGGGATATCTCCCGCGACG
CCCCTTACTTCGGTTTTGAAATTCCGAACGCGCCGGGCAAATATTTCTACGTCTGGCTGGACGCACCGATT
GGCCTGATGGGTTCTTTCAAGAATCTGTGCGACAAGCGCGGCGACAGCGTAAGCTTCGATGAATACTGGAA
GAAAGACTCCACCGCCGAGCTGTACCACTTCATCGGTAAAGATATTGTTTACTTCCTGAGCCTGTTCTGGC
CTGCCATGCTGGAAGGCAGCAACTTCCGCAAGCCGTCCAACCTGTTTGTTCATGGCTATGTGACGGTGAAC
GGCGCAAAGATGTCCAAGTCTCGCGGCACCTTTATTAAAGCCAGCACCTGGCTGAATCATTTTGACGCAGA
CAGCCTGCGTTACTACTACACTGCGAAACTCTCTTCGCGCATTGATGATATCGATCTCAACCTGGAAGATT
TCGTTCAGCGTGTGAATGCCGATATCGTTAACAAAGTGGTTAACCTGGCCTCCCGTAATGCGGGCTTTATC
AACAAGCGTTTTGACGGCGTGCTGGCAAGCGAACTGGCTGACCCGCAGTTGTACAAAACCTTCACTGATGC
CGCTGAAGTGATTGGTGAAGCGTGGGAAAGCCGTGAATTTGGTAAAGCCGTGCGCGAAATCATGGCGCTGG
CTGATCTGGCTAACCGCTATGTCGATGAACAGGCTCCGTGGGTGGTGGCGAAACAGGAAGGCCGCGATGCC
GACCTGCAGGCAATTTGCTCAATGGGCATCAACCTGTTCCGCGTGCTGATGACTTACCTGAAGCCGGTACT
GCCGAAACTGACCGAGCGTGCAGAAGCATTCCTCAATACGGAACTGACCTGGGATGGTATCCAGCAACCGC
TGCTGGGCCACAAAGTGAATCCGTTCAAGGCGCTGTATAACCGCATCGATATGAGGCAGGTTGAAGCACTG
GTGGAAGCCTCTAAAGAAGAAGTAAAAGCCGCTGCCGCGCCGGTAACTGGCCCGCTGGCAGATGATCCGAT
TCAGGAAACCATCACCTTTGACGACTTCGCTAAAGTTGACCTGCGCGTGGCGCTGATTGAAAACGCAGAGT
TTGTTGAAGGTTCTGACAAACTGCTGCGCCTGACGCTGGATCTCGGCGGTGAAAAACGCAATGTCTTCTCC
GGTATTCGTTCTGCTTACCCGGATCCGCAGGCACTGATTGGTCGTCACACCATTATGGTGGCTAACCTGGC
ACCACGTAAAATGCGCTTCGGTATCTCTGAAGGCATGGTGATGGCTGCCGGTCCTGGCGGGAAAGATATTT
TCCTGCTAAGCCCGGATGCCGGTGCTAAACCGGGTCATCAGGTGAAATAAGAGCTCCTCGAGTCTAGAGGG
CCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCC
CGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGC
ATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTC
TAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGA
CCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCC
GGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGA
CCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTT
TGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCG
GTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACA
AAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGC
AGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAG
CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATC
CCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGA
GGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTT
GCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGC
ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTG
GGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTT
TTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCC
ACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGG
CGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATG
CAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAG
CGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGC
GCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCG
ATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGT
GTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGC

TGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTG
ACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAG
ATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATG
ATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGG
TTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTT
TGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCA
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCAT
AAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTT
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGT
ATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATC
AGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAA
AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCT
GACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGC
GTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCT
TTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTT
CGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCG
TCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACA
AACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAG
ATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATG
AGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTAT
ATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT
TTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGC
CCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGC
CGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGG
AAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTG
TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCC
CATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGT
TATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTG
ACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTC
AATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGC
GAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCT
TCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG
GTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACA
TTTCCCCGAAAAGTGCCACCTGACGTC

MAHHHHHHTQVAKKILVTCANPYANGSIHLGHMLEHIQADVWVRYQRMRGHEVNFICADDAHGTPIMLKAQ
QLGITPEQMIGEMSQEHQTDFAGFNISYDNYHSTHSEENRQLSELIYSRLKENGFIKNRTISQLYDPEKGM
FLPDRFVKGTCPKCKSPDQYGDNCEVCGATYSPTELIEPKSVVSGATPVMRDSEHFFFDLPSFSEMLQAWT
RSGALQEQVANKMQEWFESGLQQWDISRDAPYFGFEIPNAPGKYFYVWLDAPIGLMGSFKNLCDKRGDSVS
FDEYWKKDSTAELYHFIGKDIVYFLSLFWPAMLEGSNFRKPSNLFVHGYVTVNGAKMSKSRGTFIKASTWL
NHFDADSLRYYYTAKLSSRIDDIDLNLEDFVQRVNADIVNKVVNLASRNAGFINKRFDGVLASELADPQLY
KTFTDAAEVIGEAWESREFGKAVREIMALADLANRYVDEQAPWVVAKQEGRDADLQAICSMGINLFRVLMT
YLKPVLPKLTERAEAFLNTELTWDGIQQPLLGHKVNPFKALYNRIDMRQVEALVEASKEEVKAAAAPVTGP
LADDPIQETITFDDFAKVDLRVALIENAEFVEGSDKLLRLTLDLGGEKRNVFSGIRSAYPDPQALIGRHTI
MVANLAPRKMRFGISEGMVMAAGPGGKDIFLLSPDAGAKPGHQVK-

Figure S4.8.

The pMetRSNLLtRNA_G vector for expression of the *E. Coli* NLL-MetRS under CMV promoter control. Restriction enzymes are highlighted in yellow, start codon in red, and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence. This plasmid also contains the tRNA expression cassette as outlined in Figure 4.1b of the main text. The tRNA cassette was inserted into the BglII restriction site in the backbone of the plasmid and is lighted in pink.



GACGGATCGGGAGATCTGATCCGACCGTGTGCTTGGCAGAACGGCTACGTAGCTCAGTTGGTTAGAGCACA
TCACTCATAATGATGGGGTCACAGGTTCGAATCCCGTCGTAGCCACCAGGTCCTTTTTTTGAGATCTCCCG
ATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCT
TGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAA
TTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTT
GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAG
TTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTC
AATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTAC
GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGAC
GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTG
ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG
GACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT
CTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTC
ACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCATCACCATACTCAAGTCGCGAA
GAAAATTCTGGTGACGTGCGCAAACCCGTACGCTAACGGCTCAATCCACCTCGGCCATATGCTGGAGCACA
TCCAGGCTGATGTCTGGGTCCGTTACCAGCGAATGCGCGGCCACGAGGTCAACTTCATCTGCGCCGACGAT
GCCCACGGTACACCGATCATGCTGAAAGCTCAGCAGCTTGGTATCACCCCGGAGCAGATGATTGGCGAAAT

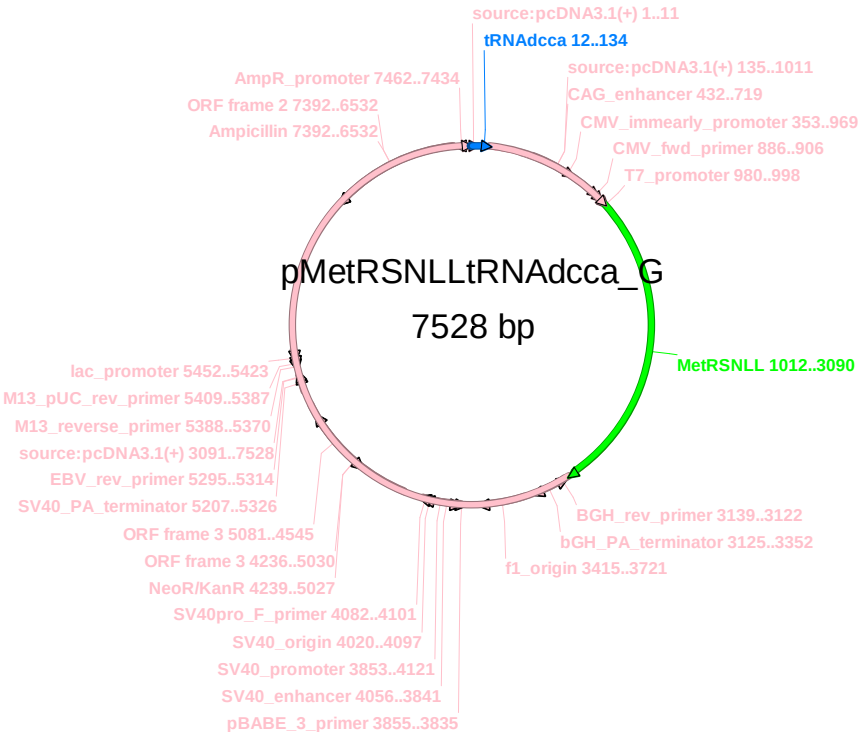
GAGTCAGGAGCATCAGACTGATTTCGCAGGCTTTAACATCAGCTATGACAACTATCACTCGACGCACAGCG
AAGAGAACCGCCAGTTGTCAGAACTTATCTACTCTCGCCTGAAAGAAAACGGTTTTATTAAAAACCGCACC
ATCTCTCAGCTGTACGATCCGGAAAAAGGCATGTTCCTGCCGGACCGTTTTGTGAAAGGCACCTGCCCGAA
ATGTAAATCCCCGGATCAATACGGCGATAACTGCGAAGTCTGCGGCGCGACCTACAGCCCGACTGAACTGA
TCGAGCCGAAATCGGTGGTTTCTGGCGCTACGCCGGTAATGCGTGATTCTGAACACTTCTTCTTTGATCTG
CCCTCTTTCAGCGAAATGTTGCAGGCATGGACCCGCAGCGGTGCGTTGCAGGAGCAGGTGGCAAATAAAAT
GCAGGAGTGGTTTGAATCTGGCCTGCAACAGTGGGATATCTCCCGCGACGCCCCTTACTTCGGTTTTGAAA
TTCCGAACGCGCCGGGCAAATATTTCTACGTCTGGCTGGACGCACCGATTGGCCTGATGGGTTCTTTCAAG
AATCTGTGCGACAAGCGCGGCGACAGCGTAAGCTTCGATGAATACTGGAAGAAAGACTCCACCGCCGAGCT
GTACCACTTCATCGGTAAAGATATTGTTTACTTCCTGAGCCTGTTCTGGCCTGCCATGCTGGAAGGCAGCA
ACTTCCGCAAGCCGTCCAACCTGTTTGTTCATGGCTATGTGACGGTGAACGGCGCAAAGATGTCCAAGTCT
CGCGGCACCTTTATTAAAGCCAGCACCTGGCTGAATCATTTTGACGCAGACAGCCTGCGTTACTACTACAC
TGCGAAACTCTCTTCGCGCATTGATGATATCGATCTCAACCTGGAAGATTTCGTTCAGCGTGTGAATGCCG
ATATCGTTAACAAAGTGGTTAACCTGGCCTCCCGTAATGCGGGCTTTATCAACAAGCGTTTTGACGGCGTG
CTGGCAAGCGAACTGGCTGACCCGCAGTTGTACAAAACCTTCACTGATGCCGCTGAAGTGATTGGTGAAGC
GTGGGAAAGCCGTGAATTTGGTAAAGCCGTGCGCGAAATCATGGCGCTGGCTGATCTGGCTAACCGCTATG
TCGATGAACAGGCTCCGTGGGTGGTGGCGAAACAGGAAGGCCGCGATGCCGACCTGCAGGCAATTTGCTCA
ATGGGCATCAACCTGTTCCGCGTGCTGATGACTTACCTGAAGCCGGTACTGCCGAAACTGACCGAGCGTGC
AGAAGCATTCCTCAATACGGAACTGACCTGGGATGGTATCCAGCAACCGCTGCTGGGCCACAAAGTGAATC
CGTTCAAGGCGCTGTATAACCGCATCGATATGAGGCAGGTTGAAGCACTGGTGGAAGCCTCTAAAGAAGAA
GTAAAAGCCGCTGCCGCGCCGGTAACTGGCCCGCTGGCAGATGATCCGATTCAGGAAACCATCACCTTTGA
CGACTTCGCTAAAGTTGACCTGCGCGTGGCGCTGATTGAAAACGCAGAGTTTGTTGAAGGTTCTGACAAAC
TGCTGCGCCTGACGCTGGATCTCGGCGGTGAAAAACGCAATGTCTTCTCCGGTATTCGTTCTGCTTACCCG
GATCCGCAGGCACTGATTGGTCGTCACACCATTATGGTGGCTAACCTGGCACCACGTAAAATGCGCTTCGG
TATCTCTGAAGGCATGGTGATGGCTGCCGGTCCTGGCGGGAAAGATATTTTCCTGCTAAGCCCGGATGCCG
GTGCTAAACCGGGTCATCAGGTGAAATAAGAGCTCCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCA
GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATT
CTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGG
GATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCC
CTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCC
TAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTA
AATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGG
TGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCT
TTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAA
GGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATT
CTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCAT
GCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGT
TCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTC
TGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTT
GTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTG
CACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTG
CTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCG
GTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCA
GCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCT
CCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGC
TTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAA
GCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAG

GCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCA
TGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGAC
ATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTA
CGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGAC
TCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCT
TCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTC
ATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCAT
CACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTAT
CTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTG
TGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGC
CTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAT
ACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGA
ACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGA
CGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCT
CGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGG
CGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTG
CACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTA
CAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTG
AAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTT
TTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGG
GGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTC
ACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGA
CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCT
GACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAG
TGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGC
CAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATG
GCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT
TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAG
CACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAG
TCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCC
ACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATG
TTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTC

MAHHHHHHTQVAKKILVTCANPYANGSIHLGHMLEHIQADVWVRYQRMRGHEVNFICADDAHGTPIMLKAQ
QLGITPEQMIGEMSQEHQTDFAGFNISYDNYHSTHSEENRQLSELIYSRLKENGFIKNRTISQLYDPEKGM
FLPDRFVKGTCPKCKSPDQYGDNCEVCGATYSPTELIEPKSVVSGATPVMRDSEHFFFDLPSFSEMLQAWT
RSGALQEQVANKMQEWFESGLQQWDISRDAPYFGFEIPNAPGKYFYVWLDAPIGLMGSFKNLCDKRGDSVS
FDEYWKKDSTAELYHFIGKDIVYFLSLFWPAMLEGSNFRKPSNLFVHGYVTVNGAKMSKSRGTFIKASTWL
NHFDADSLRYYYTAKLSSRIDDIDLNLEDFVQRVNADIVNKVVNLASRNAGFINKRFDGVLASELADPQLY
KTFTDAAEVIGEAWESREFGKAVREIMALADLANRYVDEQAPWVVAKQEGRDADLQAICSMGINLFRVLMT
YLKPVLPKLTERAEAFLNTELTWDGIQQPLLGHKVNPFKALYNRIDMRQVEALVEASKEEVKAAAAPVTGP
LADDPIQETITFDDFAKVDLRVALIENAEFVEGSDKLLRLTLDLGGEKRNVFSGIRSAYPDPQALIGRHTI
MVANLAPRKMRFGISEGMVMAAGPGGKDIFLLSPDAGAKPGHQVK-

Figure S4.9.

The pMetRSNLLtRNAdcca_G vector for expression of the *E. Coli* NLL-MetRS under CMV promoter control. Restriction enzymes are highlighted in yellow, start codon in red, and the enzyme coding sequence is highlighted in green. The expressed protein sequence is included after the plasmid sequence. This plasmid also contains the tRNA expression cassette as outlined in Figure 4.1b of the main text, wherein the CCA tail of the tRNA was removed in this sequence. The tRNA cassette was inserted into the BglII restriction site in the backbone of the plasmid and is lighted in pink.



GACGGATCGGGAGATCTGATCCGACCGTGTGCTTGGCAGAACGGCTACGTAGCTCAGTTGGTTAGAGCACA
TCACTCATAATGATGGGGTCACAGGTTCGAATCCCGTCGTAGCCAGTCCTTTTTTTGAGATCTCCCGATCC
CCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTG
TGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGC
ATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACA
TTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC
GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATA
ATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTA
AACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTA
AATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTA
TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTC
ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACT
TTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTAT
ATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTA
TAGGGAGACCCAAGCTGGCTAGCGCCACCATGGCCCACCATCACCATCACCATACTCAAGTCGCGAAGAAA

ATTCTGGTGACGTGCGCAAACCCGTACGCTAACGGCTCAATCCACCTCGGCCATATGCTGGAGCACATCCA
GGCTGATGTCTGGGTCCGTTACCAGCGAATGCGCGGCCACGAGGTCAACTTCATCTGCGCCGACGATGCCC
ACGGTACACCGATCATGCTGAAAGCTCAGCAGCTTGGTATCACCCCGGAGCAGATGATTGGCGAAATGAGT
CAGGAGCATCAGACTGATTTCGCAGGCTTTAACATCAGCTATGACAACTATCACTCGACGCACAGCGAAGA
GAACCGCCAGTTGTCAGAACTTATCTACTCTCGCCTGAAAGAAAACGGTTTTATTAAAAACCGCACCATCT
CTCAGCTGTACGATCCGGAAAAAGGCATGTTCCTGCCGGACCGTTTTGTGAAAGGCACCTGCCCGAAATGT
AAATCCCCGGATCAATACGGCGATAACTGCGAAGTCTGCGGCGCGACCTACAGCCCGACTGAACTGATCGA
GCCGAAATCGGTGGTTTCTGGCGCTACGCCGGTAATGCGTGATTCTGAACACTTCTTCTTTGATCTGCCCT
CTTTCAGCGAAATGTTGCAGGCATGGACCCGCAGCGGTGCGTTGCAGGAGCAGGTGGCAAATAAAATGCAG
GAGTGGTTTGAATCTGGCCTGCAACAGTGGGATATCTCCCGCGACGCCCCTTACTTCGGTTTTGAAATTCC
GAACGCGCCGGGCAAATATTTCTACGTCTGGCTGGACGCACCGATTGGCCTGATGGGTTCTTTCAAGAATC
TGTGCGACAAGCGCGGCGACAGCGTAAGCTTCGATGAATACTGGAAGAAAGACTCCACCGCCGAGCTGTAC
CACTTCATCGGTAAAGATATTGTTTACTTCCTGAGCCTGTTCTGGCCTGCCATGCTGGAAGGCAGCAACTT
CCGCAAGCCGTCCAACCTGTTTGTTCATGGCTATGTGACGGTGAACGGCGCAAAGATGTCCAAGTCTCGCG
GCACCTTTATTAAAGCCAGCACCTGGCTGAATCATTTTGACGCAGACAGCCTGCGTTACTACTACACTGCG
AAACTCTCTTCGCGCATTGATGATATCGATCTCAACCTGGAAGATTTCGTTCAGCGTGTGAATGCCGATAT
CGTTAACAAAGTGGTTAACCTGGCCTCCCGTAATGCGGGCTTTATCAACAAGCGTTTTGACGGCGTGCTGG
CAAGCGAACTGGCTGACCCGCAGTTGTACAAAACCTTCACTGATGCCGCTGAAGTGATTGGTGAAGCGTGG
GAAAGCCGTGAATTTGGTAAAGCCGTGCGCGAAATCATGGCGCTGGCTGATCTGGCTAACCGCTATGTCGA
TGAACAGGCTCCGTGGGTGGTGGCGAAACAGGAAGGCCGCGATGCCGACCTGCAGGCAATTTGCTCAATGG
GCATCAACCTGTTCCGCGTGCTGATGACTTACCTGAAGCCGGTACTGCCGAAACTGACCGAGCGTGCAGAA
GCATTCCTCAATACGGAACTGACCTGGGATGGTATCCAGCAACCGCTGCTGGGCCACAAAGTGAATCCGTT
CAAGGCGCTGTATAACCGCATCGATATGAGGCAGGTTGAAGCACTGGTGGAAGCCTCTAAAGAAGAAGTAA
AAGCCGCTGCCGCGCCGGTAACTGGCCCGCTGGCAGATGATCCGATTCAGGAAACCATCACCTTTGACGAC
TTCGCTAAAGTTGACCTGCGCGTGGCGCTGATTGAAAACGCAGAGTTTGTTGAAGGTTCTGACAAACTGCT
GCGCCTGACGCTGGATCTCGGCGGTGAAAAACGCAATGTCTTCTCCGGTATTCGTTCTGCTTACCCGGATC
CGCAGGCACTGATTGGTCGTCACACCATTATGGTGGCTAACCTGGCACCACGTAAAATGCGCTTCGGTATC
TCTGAAGGCATGGTGATGGCTGCCGGTCCTGGCGGGAAAGATATTTTCCTGCTAAGCCCGGATGCCGGTGC
TAAACCGGGTCATCAGGTGAAATAAGAGCTCCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCT
CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGT
GCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTAT
TCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG
CGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGT
AGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC
GCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATC
GGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGAT
GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAA
TAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGA
TTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGT
GGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCAT
CTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCA
TCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCG
CCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAG
CTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTAT
ATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACG
CAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCT
GATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGC
CCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTG
TGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTG

TCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGA
TCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCG
GTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTC
AAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGT
GGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAG
CGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGT
ATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTG
GGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTA
TGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGC
TGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACA
AATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTA
TCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA
ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAA
TGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGC
TCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCG
TAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT
CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTG
CGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCT
TTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACG
AACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACAC
GACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGC
CAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTT
GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTC
TGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCT
AGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGT
TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACT
CCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGT
CCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGT
TAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTT
CATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGC
TCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACT
GCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACAT
AGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCT
GTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCG
TTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGA
ATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACG
TC

MAHHHHHHTQVAKKILVTCANPYANGSIHLGHMLEHIQADVWVRYQRMRGHEVNFICADDAHGTPIMLKAQ
QLGITPEQMIGEMSQEHQTDFAGFNISYDNYHSTHSEENRQLSELIYSRLKENGFIKNRTISQLYDPEKGM
FLPDRFVKGTCPKCKSPDQYGDNCEVCGATYSPTELIEPKSVVSGATPVMRDSEHFFFDLPSFSEMLQAWT
RSGALQEQVANKMQEWFESGLQQWDISRDAPYFGFEIPNAPGKYFYVWLDAPIGLMGSFKNLCDKRGDSVS
FDEYWKKDSTAELYHFIGKDIVYFLSLFWPAMLEGSNFRKPSNLFVHGYVTVNGAKMSKSRGTFIKASTWL
NHFDADSLRYYYTAKLSSRIDDIDLNLEDFVQRVNADIVNKVVNLASRNAGFINKRFDGVLASELADPQLY

KTFTDAAEVIGEAWESREFGKAVREIMALADLANRYVDEQAPWVVAKQEGRDADLQAICSMGINLFRVLMT
YLKPVLPKLTERAEAFLNTELTWDGIQQPLLGHKVNPFKALYNRIDMRQVEALVEASKEEVKAAAAPVTGP
LADDPIQETITFDDFAKVDLRVALIENAEFVEGSDKLLRLTLDLGGEKRNVFSGIRSAYPDPQALIGRHTI
MVANLAPRKMRFGISEGMVMAAGPGGKDIFLLSPDAGAKPGHQVK-

Figure S4.10.

The chemical structure of alkyne-tetramethylrhodamine(TAMRA) dye.

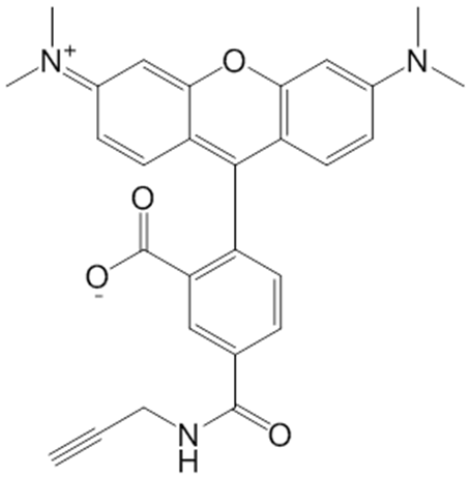


Figure S4.11.

Derivation of equation for selectivity of MetRS for Anl versus Met. In-gel fluorescence quantification is used as a measure of Anl incorporation by the L274GMars enzyme. At a fixed level of Met, changes in Anl concentration will directly translate into different incorporation levels of Anl that can be measured as a function of time.

Notation:

E:	Concentration of L274GMars
A:	Anl concentration
ES_A:	Enzyme substrate complex for Anl
ES_M:	Enzyme substrate complex for Met
PA:	Polypeptide containing an Anl residue
PM:	Polypeptide containing a Met residue
K_1^A:	K forward for Anl complex with L274GMars
K_{-1}^A:	K reverse for Anl complex with L274GMars
K_2^A:	rate of product formation based on concentration of L274GMars and Anl
K_1^M:	K forward for Met complex with L274GMars
K_{-1}^M:	K reverse for Met complex with L274GMars
K_2^M:	rate of product formation based on concentration of L274GMars and Met

Anl: $$E + A \underset{K_{-1}^A}{\overset{K_1^A}{\leftrightarrow}} ES_A \overset{K_2^A}{\rightarrow} E + P_A \qquad (1)$$

Met: $$E + M \underset{K_{-1}^M}{\overset{K_1^M}{\leftrightarrow}} ES_M \overset{K_2^M}{\rightarrow} E + P_M \qquad (2)$$

At steady state: $$ES_A' = 0 = E.A.K_1^A - ES_A(K_{-1}^A + K_2^A) \qquad (3)$$

Therefore:

$$ES_A = \frac{E.A}{K_A} \qquad (4)$$

where: $K_A = \frac{K_{-1}^A + K_2^A}{K_1^A}$

and

$$ES_M = \frac{E.M}{K_M} \qquad (5)$$

where: $K_M = \frac{K_{-1}^M + K_2^M}{K_1^M}$

$$E_0 = E + ES_A + ES_M \quad \rightarrow \quad E = \frac{E_0.K_A.K_M}{K_A.K_M + A.K_M + M.K_A} \qquad (6)$$

using (4) and (6): $$V_A = K_2^A ES_A = \frac{A}{K_A}\left(\frac{E_0.K_A.K_M}{K_A K_M + A.K_M + M.K_A}\right).K_2^A$$

using (5) and (6):

$$V_M = K_2^M ES_M = \frac{M}{K_M}\left(\frac{E_0.K_A.K_M}{K_A K_M + A.K_M + M.K_A}\right).K_2^M$$

Taking the ratio: $\frac{V_A}{V_M}=\frac{A}{M}\frac{K_2^A.K_M}{K_2^M.K_A}$

-or-

$$\frac{V_A}{V_M}=\frac{A}{M}\frac{\left(\frac{k_{cat}}{k_M}\right)^{Anl}}{\left(\frac{k_{cat}}{k_M}\right)^{Met}} \quad (7)$$

let, $K^A = \left(\frac{k_{cat}}{k_M}\right)^{Anl}$ and $K^M = \left(\frac{k_{cat}}{k_M}\right)^{Met}$ in (7)

$$\frac{V_A}{V_M}=\frac{A}{M}\frac{K^A}{K^M} \quad (8)$$

For a fixed time period "i", the Anl concentration is a fixed amount and the total protein synthesis rate is also a constant amount C: $C = V_{Ai} + V_{Mi}$

Rearranging (8) at any point "i" gives:

$$K^A = \frac{V_{Ai}}{V_{mi}}\frac{M}{A}K^M$$

Substitute using C:

$$K^A = \frac{V_{Ai}}{C - V_{Ai}}\frac{M}{A}K^M$$

-or-

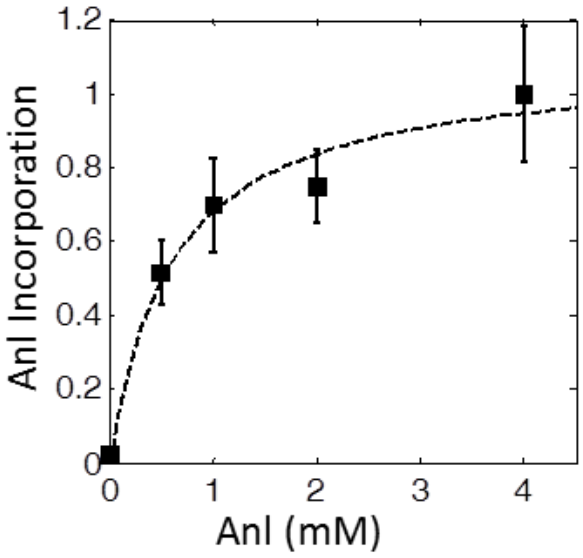
$$V_{Ai} = \frac{K^A C}{\frac{M}{A}K^M + K^A} \quad (9)$$

Using (9) and dividing through by K_M yields:

$$V_{Ai} = \frac{A\left(\frac{K^A}{K^M}\right)C}{M + A\left(\frac{K^A}{K^M}\right)} \quad (10)$$

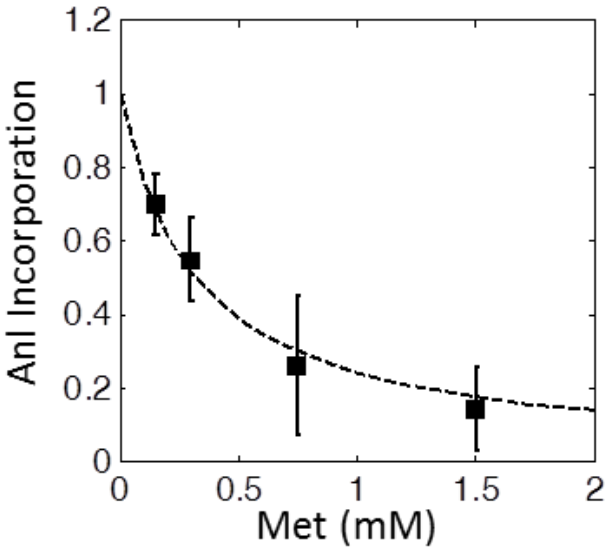
where, $\left(\frac{K_A}{K_M}\right)$ is selectivity, c is a constant and Met and Anl concentrations are known. Equation 10 is used for a least-squares fit to kinetics of Anl incorporation at different Anl or Met concentrations to determine selectivity of L274G enzyme.

The following Matlab functions was used to fit the data from Fig 4.3A in the main text, where the resulting parameter fits corresponding to selectivity and C in equation (10) were 0.2508 and 1.089 respectively.




```
------------------------------
function fitflouresenceR
function m = FF(x , xdata)
for i=1:size(xdata)
m(i)=((xdata(i)*x(1)*x(2))/((0.15+xdata(i)*x(1))));
end
m=m';
end

A=[0; 0.5; 1;  2; 4];
fl=[0.022871044;    0.516731636;    0.699062794; 0.752107911;   1]
err=[0.002135075;   0.086032748;    0.127708537;    0.098922244;    0.183627695];
errorbar(A,fl,err,'ro');

% initial guess for paramters
x0=[0.5 1] % This is the selectivity paratmer
% second 1 this is constant C

[x,resnorm, residual,~,exitflag,output]=lsqcurvefit(@FF,x0,A,fl);
hold on
counter=0; for i=0:0.1:4.5, counter=counter+1; II(counter)=i;
F(counter)=(i*0.2508*1.089)/(0.15+i*0.2508);end
plot(II,F)
x

SStotal = (length(fl)-1) * var(fl);
SSresid=sum(residual.^2);
rsq = 1 - SSresid/SStotal

end
------------------------------------------------
```

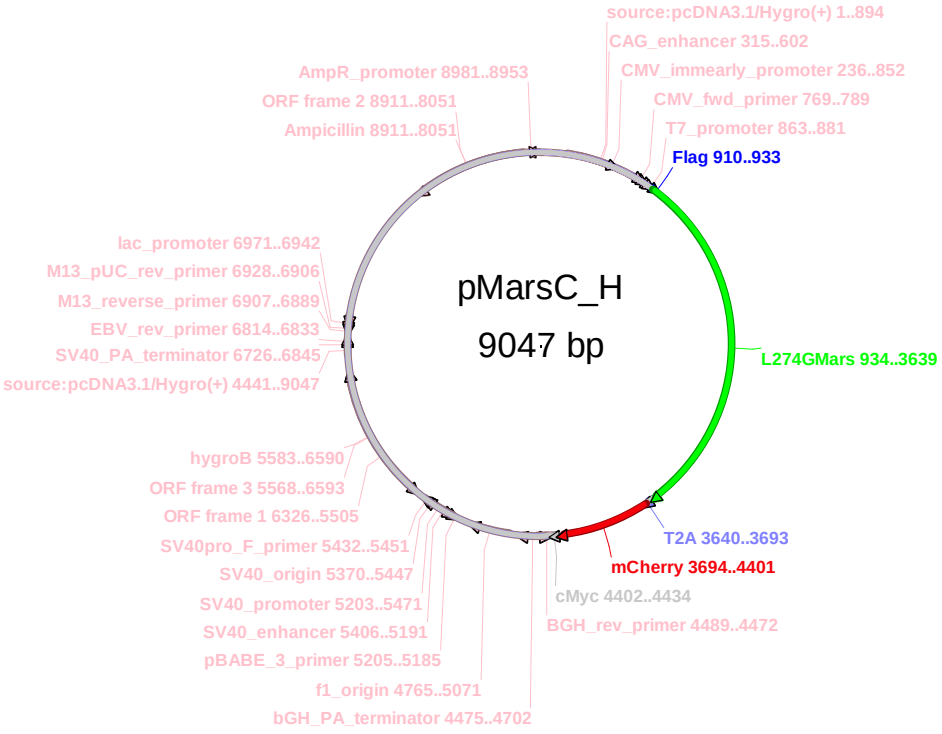
The following Matlab functions was used to fit the data from Fig 4.3B in the main text. The data was first normalized so that at the Met concentration of 0.15mM, the Anl incorporation corresponds to that of Fig 4.3A. Also, by definition of equation (10), at Met concentration of 0mM, the value of constant C should be 1. Based on these results, we plotted the data as well as the line of best fit. The selectivity obtained for this fit is 0.2131.




```
------------------------------
function fitflouresencetst
function m = FF(x , xdata)
for i=1:size(xdata)
m(i)=((1.5*x(1)*1)/((xdata(i)+1.5*x(1))))
end
m=m';
end

M=[0.15;     0.3;     0.75;    1.5];
fl=[0.7;     0.548851071;     0.263135683;     0.14119813];
err=[0.086722315;    0.112658654;     0.19100414; 0.11272512];
errorbar(M,fl,err,'ro');

% initial guess for parameters
x0=[0.1 1] % This is the selectivity parameter
% second 1 this is constant C

[x,resnorm, residual,~,exitflag,output]=lsqcurvefit(@FF,x0,M,fl);
x
residual
hold on
counter=0; for i=0:0.1:2, counter=counter+1; II(counter)=i;
F(counter)=(1.5*x(1)*x(2))/(i+1.5*x(1));end
plot(II,F)

SStotal = (length(fl)-1) * var(fl);
SSresid=sum(residual.^2);
rsq = 1 - SSresid/SStotal
end
---------------------------------------------
```

Figure S4.12.

The pMarC vector for expression of L274G Mars and mCherry proteins. The insertion cassette into the Nhe1/Xho1 restriction sites of the pcDNA3.1 plasmids is color coded corresponding to highlighted sequences. The pMars plasmid is the same construct as pMarsL274G, and below is the sequence for pMarsC which contains a T2A-Mcherry sequence appended to the C-terminal of L274G Mars enzyme.



Nhe1Flag-tagMars*T2AMcherryMyc-tagXho1

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGATTACAAGGATGA
TGATGATAAGATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGGCCGCGA
GGGCCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTCCTTACC
CGGCCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTGCCGATA
TTTTTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTGGAATGGGAGGCAACAGAAC
TGCAGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATACTTGGC

CCACTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGCTGGGGA
CACAGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTGCAAGACCCAGCTTACCTCC
CTGAGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGAACCGTGTCAGCGAGCTGCA
GAGACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGAAACAGCCACAGCCTCAGCC
CCCGCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTGAGGAGG
ACATCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAGCAGCAT
CCAGTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCGGACCCTATGTCAACAATGTCCC
CCACCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTCGCCAGT
GGAATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGACCAAGGCCATGGAGGAGGGC
CTAACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCTACCGCTGGTTCGGCATATC
GTTCGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAGGACATCTTCCAGAGGTTGC
TGACCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTCCTGGCT
GACCGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGACAGGTG
TGGCAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTGTGGTGA
GGTCCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGGAAGACA
GTGCCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGCTTCGAGATGGCCTCAAGCC
ACGATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGGTATTTT
ACGTCTGGTTTGATGCTACTATTGGCTACGTGTCCATCACAGCCAACTACACAGACCAATGGGAGAAATGG
TGGAAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATGTTCCCTTCCATGGCTTGGT
CTTTCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCACATCATTGCTACAGAGTACC
TGAACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAGGATACA
GGAATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGGGCCAGGACAGTGCCTTCTC
CTGGACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGCAACTTCATCAACAGAGCTG
GCATGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAACCCCTGATGACAGACGCCTG
GTGGCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGATGCCTT
GCGCAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAATGAGCCCTGGAAACGGATTA
AAGGTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGTGAACATGGCTGCCTTGCTG
TCTGTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCCACCTGC
AGCCTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCACCGAATTGGCACAGTCAGTC
CTTTGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAGGCTAAA
GGCTCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGCACATACAAACGCTGACGGA
TGAGGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCAGACAAGAACCAGGTTGCTG
CAGAGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAAACTCCT
AAAGGCAAGAAGAAAAAGG**AAGGGAGAGGAAGCCTCCTAACATGCGGTGACGTGGAGGAGAACCCAGGACC**
AATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGG
AGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAG
ACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCAT
GTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGG
GCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTG
CAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCA
GAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGA
TCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAG
AAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTA
CACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGG
AACAAAAACTTATTTCTGAAGAAGATCTGTAACTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCC
TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGG
TGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTA
TTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAT
GCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTG
TAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAG
CGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAAT
CGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGA
TGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTA
ATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGG
ATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTG
TGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCA
TCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGC

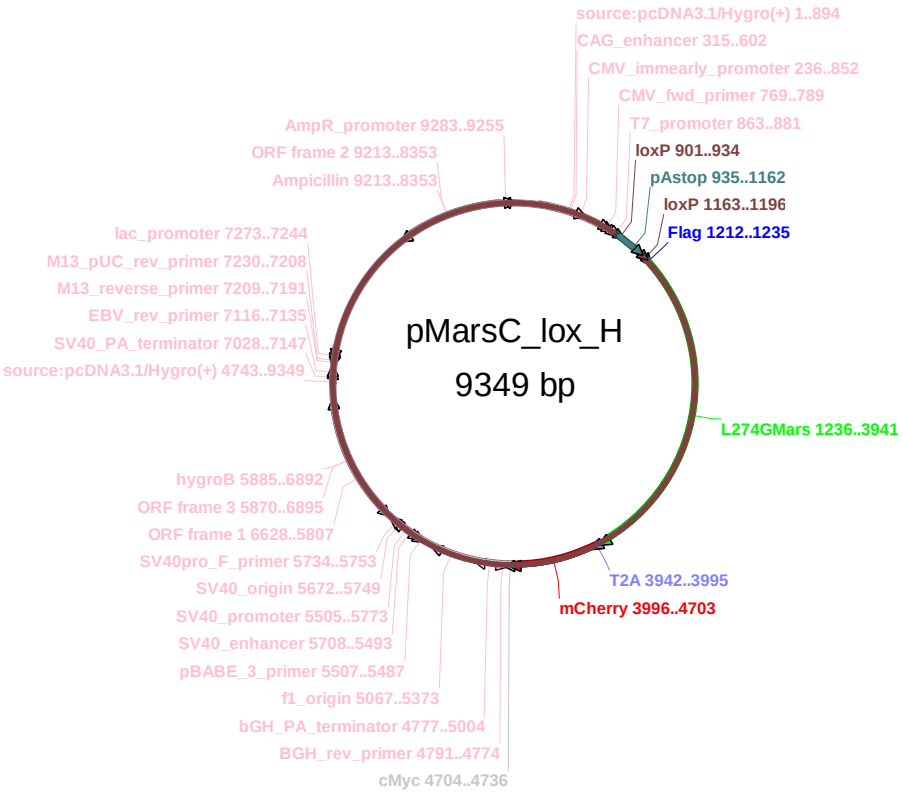
ATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCC
GCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGA
GCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTA
TATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCAC
GCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTC
TGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTG
CCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCT
GTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCT
GTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTG
ATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCC
GGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCT
CAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGG
TGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATA
GCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGG
TATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCT
GGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCT
ATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATG
CTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCAC
AAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTT
ATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGA
AATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTA
ATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACG
GTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACC
GTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGC
TCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGC
TTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCAC
GAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA
CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAG
AGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAG
CCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTT
TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACC
TAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAG
TTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGA
GACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG
TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAG
TTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCT
TCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAG
CTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCAC
TGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCA
TTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACA
TAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGC
TGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGC
GTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
GTC

MDYKDDDDKMRLFVSEGSPGSLPVLAAAARARGRAELLISTVGPEECVVPFLTRPKVPVLQLDSGNYLFSA
SAICRYFFLLCGWEQDDLTNQWLEWEATELQPVLSAALHCLVVQGKKGEDILGPLRRVLTHIDHSLSRQNC
PFLAGDTESLADIVLWGALYPLLQDPAYLPEELGALQSWFQTLSTQEPCQRAAETVLKQQGVLALRLYLQK
QPQPQPPPPEGRTVSNELEEEELATLSEEDIVTAVAAWEKGLESLPPLKLQQHPVLPVPGERNVLITSAGP
YVNNVPHLGNIIGCVLSADVFARYCRLRQWNTLYLCGTDEYGTATETKAMEEGLTPREICDKYHAIHADIY

RWFGISFDTFGRTTTPQQTKITQDIFQRLLTRGFVLRDTVEQLRCERCARFLADRFVEGVCPFCGYEEARG
DQCDRCGKLINAIELKKPQCKICRSCPVVRSSQHLFLDLPKLEKRLEDWLGKTVPGSDWTPNARFIIRSWL
RDGLKPRCITRDLKWGTPVPLEGFEDKVFYVWFDATIGYVSITANYTDQWEKWWKNPEQVDLYQFMAKDNV
PFHGLVFPCSVLGAEDNYTLVKHIIATEYLNYEDGKFSKSRGIGVFGDMAKDTGIPADIWRFYLLYIRPEG
QDSAFSWTDLLIKNNSELLNNLGNFINRAGMFVSKFFGGCVPEMALTPDDRRLVAHVSWELQHYHQLLEKV
RIRDALRSILTISRHGNQYIQVNEPWKRIKGGEMDRQRAGTVTGMAVNMAALLSVMLQPYMPTVSSTIQTQ
LQLPPAACRILATSFICTLPAGHRIGTVSPLFQKLENDQIENLRQRFGGGQAKGSPKPAAVEAVTAAGSQH
IQTLTDEVTKQGNVVRELKAQKADKNQVAAEVAKLLDLKKQLALAEGKPIETPKGKKKKEGRGSLLTCGDV
EENPGPMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI
LSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPS
DGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDIT
SHNEDYTIVEQYERAEGRHSTGGMDELYKEQKLISEEDL-

Figure S4.13.

The pMarlox vector for expression of L274G Mars and mCherry proteins. The insertion cassette into the Nhe1/Xho1 restriction sites of the pcDNA3.1 plasmids is color coded corresponding to highlighted sequences. The loxP sequences and transcriptional stop sequence are inserted in the Nhe1 cut site of this plasmid, thereby introducing a second Nhe1 cut site.



Nhe1loxpstoploxpSDFlagMars*T2AMcherryMycXho1

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCATAACTTCGTATAGCATACATTA
TACGAAGTTATCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGA

CCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGG
TGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCA
TGCTGGGGATGCGGTGGGCTCTATGGATAACTTCGTATAGCATACATTATACGAAGTTATGCTAGCCGCCA
CCATGCATTACAAGGATGATGATGATAAGATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCC
GTGCTGGCTGCGGCCGCGAGGGCCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTG
TGTGGTACCATTCCTTACCCGGCCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTG
CTAGTGCAATCTGCCGATATTTTTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTG
GAATGGGAGGCAACAGAACTGCAGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAA
AGGGGAAGATATACTTGGCCCACTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACT
GTCCTTTCCTGGCTGGGGACACAGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTG
CAAGACCCAGCTTACCTCCCTGAGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGA
ACCGTGTCAGCGAGCTGCAGAGACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGA
AACAGCCACAGCCTCAGCCCCCGCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTG
GCTACCTTGTCTGAGGAGGACATCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCC
GCTAAAGCTCCAGCAGCATCCAGTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCGGAC
CCTATGTCAACAATGTCCCCCACCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGG
TATTGTCGCCTTCGCCAGTGGAATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGAC
CAAGGCCATGGAGGAGGGCCTAACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCT
ACCGCTGGTTCGGCATATCGTTCGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAG
GACATCTTCCAGAGGTTGCTGACCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCG
GTGTGCACGTTTCCTGGCTGACCGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAG
GTGACCAGTGTGACAGGTGTGGCAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGC
CGCTCCTGCCCTGTGGTGAGGTCCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGA
GGACTGGTTGGGGAAGACAGTGCCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGC
TTCGAGATGGCCTCAAGCCACGATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGT
TTTGAGGACAAGGTATTTTACGTCTGGTTTGATGCTACTATTGGCTACGTGTCCATCACAGCCAACTACAC
AGACCAATGGGAGAAATGGTGGAAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATG
TTCCCTTCCATGGCTTGGTCTTTCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCAC
ATCATTGCTACAGAGTACCTGAACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGG
AGACATGGCCAAGGATACAGGAATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGG
GCCAGGACAGTGCCTTCTCCTGGACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGC
AACTTCATCAACAGAGCTGGCATGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAAC
CCCTGATGACAGACGCCTGGTGGCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGG
TTCGGATCCGGGATGCCTTGCGCAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAAT
GAGCCCTGGAAACGGATTAAAGGTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGT
GAACATGGCTGCCTTGCTGTCTGTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCC
AGCTGCAGCTCCCACCTGCAGCCTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCAC
CGAATTGGCACAGTCAGTCCTTTGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTT
TGGAGGGGGTCAGGCTAAAGGCTCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGC
ACATACAAACGCTGACGGATGAGGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCA
GACAAGAACCAGGTTGCTGCAGAGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGG
GAAACCCATTGAAACTCCTAAAGGCAAGAAGAAAAAGGGAAGGGAGAGGAAGCCTCCTAACATGCGGTGACG
TGGAGGAGAACCCAGGACCAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATG
CGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCG
CCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACA
TCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTG
AAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGT
GACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCT
CCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGAC

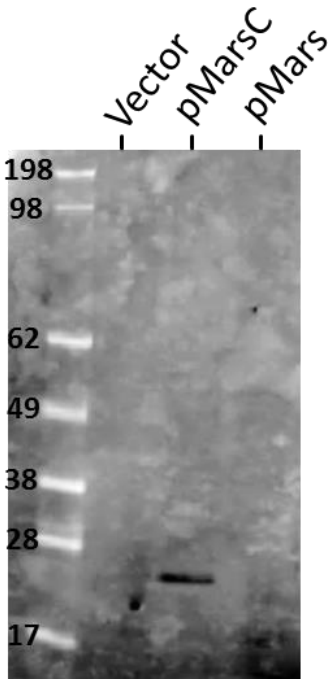
GGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAA
GACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCA
CCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGC
ATGGACGAGCTGTACAAGGAACAAAAACTTATTTCTGAAGAAGATCTGTAA**CTCGAG**TCTAGAGGGCCCGT
TTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGT
CTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAA
TAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGG
GGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCT
ACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTT
TCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACG
TTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTA
TTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAAT
TTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCA
GAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGC
AGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCC
CCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCG
AGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAA
AAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGAT
TGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCAC
AACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTC
AAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGAC
GGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAG
TGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATG
CGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGC
ACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAG
CCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCC
TGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGC
GGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACC
GCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAG
TTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTC
GATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCT
CCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACA
AATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCC
AAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTC
ATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGT
GTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAG
TCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG
GCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTC
ACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGC
CAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA
GCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTC
CCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTC
CCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTG
AGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA

CCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATT
ATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATG
AGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGT
TCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAG
TGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAA
GGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCT
AGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACG
CTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGT
TGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGG
TGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATAC
GGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAA
CTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAA
GGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTAT
TGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCC
CCGAAAAGTGCCACCTGACGTC

MDYKDDDDKMRLFVSEGSPGSLPVLAAAARARGRAELLISTVGPEECVVPFLTRPKVPVLQLDSGNYLFSA
SAICRYFFLLCGWEQDDLTNQWLEWEATELQPVLSAALHCLVVQGKKGEDILGPLRRVLTHIDHSLSRQNC
PFLAGDTESLADIVLWGALYPLLQDPAYLPEELGALQSWFQTLSTQEPCQRAAETVLKQQGVLALRLYLQK
QPQPQPPPPEGRTVSNELEEEELATLSEEDIVTAVAAWEKGLESLPPLKLQQHPVLPVPGERNVLITSAGP
YVNNVPHLGNIIGCVLSADVFARYCRLRQWNTLYLCGTDEYGTATETKAMEEGLTPREICDKYHAIHADIY
RWFGISFDTFGRTTTPQQTKITQDIFQRLLTRGFVLRDTVEQLRCERCARFLADRFVEGVCPFCGYEEARG
DQCDRCGKLINAIELKKPQCKICRSCPVVRSSQHLFLDLPKLEKRLEDWLGKTVPGSDWTPNARFIIRSWL
RDGLKPRCITRDLKWGTPVPLEGFEDKVFYVWFDATIGYVSITANYTDQWEKWWKNPEQVDLYQFMAKDNV
PFHGLVFPCSVLGAEDNYTLVKHIIATEYLNYEDGKFSKSRGIGVFGDMAKDTGIPADIWRFYLLYIRPEG
QDSAFSWTDLLIKNNSELLNNLGNFINRAGMFVSKFFGGCVPEMALTPDDRRLVAHVSWELQHYHQLLEKV
RIRDALRSILTISRHGNQYIQVNEPWKRIKGGEMDRQRAGTVTGMAVNMAALLSVMLQPYMPTVSSTIQTQ
LQLPPAACRILATSFICTLPAGHRIGTVSPLFQKLENDQIENLRQRFGGGQAKGSPKPAAVEAVTAAGSQH
IQTLTDEVTKQGNVVRELKAQKADKNQVAAEVAKLLDLKKQLALAEGKPIETPKGKKKKEGRGSLLTCGDV
EENPGPMVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDI
LSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPS
DGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDIT
SHNEDYTIVEQYERAEGRHSTGGMDELYKEQKLISEEDL-

Figure S4.14.

Western blot for detection of Myc-tagged mCherry protein. Western blot using a Myc-tag-Alexa Fluor 488 conjugate monoclonal antibody was used to probe for presence of mCherry in lystes of CHO cells transfected with pMarsC and pMars vectors. The pMarsC lane shows presence of a protein band at approximately 25 kDa corresponding to mCherry and no other bands at higher molecular weights indicating that mCherry is not fused to the AnlRS enzyme. The pMars vector lacks the mCherry sequence and as anticipated we do not observe a protein band corresponding to mCherry protein in the cell lysates. This image corresponds to the western blot image in Figure 4.3b of the main text.



References

(1) Hinz, F. I.; Dieterich, D. C.; Tirrell, D. A.; Schuman, E. M. *ACS Chem. Neurosci.* **2012**, *3*, 40.

(2) Ngo, J. T.; Tirrell, D. A. *Acc Chem Res* **2011**, *44*, 677.

(3) Yuet, K. P.; Tirrell, D. A. *Ann. Biomed. Eng.* **2014**, *42*, 299.

(4) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482.

(5) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol Cell Proteomics* **2002**, *1*, 376.

(6) Lu, Y. Y.; Sweredoski, M. J.; Huss, D.; Lansford, R.; Hess, S.; Tirrell, D. A. *ACS. Chem. Biol.* **2014**, *9*, 334.

(7) Bagert, J. D.; Xie, Y. J.; Sweredoski, M. J.; Qi, Y.; Hess, S.; Schuman, E. M.; Tirrell, D. A. *Mol Cell Proteomics* **2014**, *13*, 1352.

(8) Howden, A. J.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nat Methods* **2013**, *10*, 343.

(9) Eichelbaum, K.; Krijgsveld, J. *Methods Mol Biol* **2014**, *1174*, 101.

(10) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Biol.* **2009**, *5*, 715.

(11) Mahdavi, A.; Szychowski, J.; Ngo, J. T.; Sweredoski, M. J.; Graham, R. L.; Hess, S.; Schneewind, O.; Mazmanian, S. K.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 433.

(12) Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. *ACS. Chem. Biol.* **2012**, *7*, 1326.

(13) Mahdavi, A.; Segall-Shapiro, T. H.; Kou, S.; Jindal, G. A.; Hoff, K. G.; Liu, S.; Chitsaz, M.; Ismagilov, R. F.; Silberg, J. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2013**, *135*, 2979.

(14) Tsunasawa, S.; Stewart, J. W.; Sherman, F. *J Biol Chem* **1985**, *260*, 5382.

(15) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596.

(16) Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A. *Chembiochem* **2010**, *11*, 2092.

(17) Liang, Y.; Mackey, J. L.; Lopez, S. A.; Liu, F.; Houk, K. N. *J. Am. Chem. Soc.* **2012**, *134*, 17904.

(18) Drabkin, H. J.; Park, H. J.; RajBhandary, U. L. *Mol. Cell. Biol.* **1996**, *16*, 907.

(19) Kohrer, C.; Sullivan, E. L.; RajBhandary, U. L. *Nucleic acids research* **2004**, *32*, 6200.

(20) Ngo, J. T.; Schuman, E. M.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 4992.

(21) Meinnel, T.; Mechulam, Y.; Fayat, G.; Blanquet, S. *Nucleic acids research* **1992**, *20*, 4741.

(22) Link, A. J.; Vink, M. K.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10180.

(23) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A., 3rd; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15285.

(24) Lodish, H. F. *Molecular cell biology*; 4th ed.; W.H. Freeman: New York, 2000.

(25) Tornoe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057.

(26) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 9879.

(27) Elliott, T. S.; Townsley, F. M.; Bianco, A.; Ernst, R. J.; Sachdeva, A.; Elsasser, S. J.; Davis, L.; Lang, K.; Pisa, R.; Greiss, S.; Lilley, K. S.; Chin, J. W. *Nat Biotechnol* **2014**, *32*, 465.

(28) Friedrich, G.; Soriano, P. *Genes Dev.* **1991**, *5*, 1513.

(29) Li, G. W.; Burkhardt, D.; Gross, C.; Weissman, J. S. *Cell* **2014**, *157*, 624.

(30) Wolff, S.; Weissman, J. S.; Dillin, A. *Cell* **2014**, *157*, 52.

(31) Becker, A. H.; Oh, E.; Weissman, J. S.; Kramer, G.; Bukau, B. *Nat Protoc* **2013**, *8*, 2212.

(32) Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R.; Weissman, J. S. *Science* **2009**, *324*, 218.

(33) Link, A. J.; Vink, M. K.; Tirrell, D. A. *Nat. Protoc.* **2007**, *2*, 1884.

(34) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat Methods* **2009**, *6*, 359.

(35) Bhatia, V. N.; Perlman, D. H.; Costello, C. E.; McComb, M. E. *Anal Chem* **2009**, *81*, 9819.

CHAPTER V

Lineage-Specific Proteomic Analysis of Embryonic Stem Cell Differentiation by Non-canonical Amino Acid Tagging

Abstract

Advances in mass spectrometric instrumentation and methodologies have made proteomic analysis a valuable tool for biological discovery. A growing number of studies are using proteomics to investigate the mechanisms of pluripotency, differentiation, and lineage commitment of stem cells. These processes often involve the interactions of multiple cell types and occur in mixed populations. Therefore methods for proteomic analysis of specific cell types in heterogeneous cultures are needed. Here we introduce an approach, based on the cell-targeted metabolic incorporation of non-canonical amino acids, to enable lineage-specific proteomic analysis of stem cells in culture. We use this method to selectively label newly made proteins in subpopulations of mouse embryonic stem cells (ESCs) that are differentiating towards mesodermal and cardiac lineages that are tracked by brachyury and α-MHC promoters, respectively. This technique for lineage-specific proteomic labeling enables unbiased and comprehensive analysis of proteome changes during stem cell differentiation and cell-fate commitment.

Introduction

The field of stem cell research is quickly progressing towards understanding mechanisms of stem cell fate commitment and development of new therapies. Combined with recent advances in genome-editing[1], stem cell research provides exciting opportunities for cell-based therapies and new models of human disease[2]. Transcriptomic and epigenetic studies have been used to uncover transcriptional profiles and chromatin states of stem cells, and proteomic analysis is being used to study posttranslational modifications of these cells[3]. Such efforts have generated dense and unbiased datasets that provide representative snapshots of the proteome of stem cells during pluripotency and differentiation.[4]

Advances in proteomics platforms have been complemented with progress in chemical approaches that enable the selective and temporal labeling and identification of proteins. Through the SILAC technique[5], and more recently through the BONCAT[6] method, amino acid labeling approaches are quickly becoming widely applicable for studies of proteome changes in cells. Whereas SILAC enables identification and quantitation of newly made proteins through incorporation of isotope-labeled amino acids, BONCAT utilizes incorporation of non-canonical amino acids into newly made proteins so that they can be chemically tagged, enriched and identified by tandem mass spectrometry. SILAC has previously been used to study pluripotency[7,8] and differentiation[9,10] of stem cells. A combined BONCAT-SILAC approach was recently used to quantitate protein dynamics of primary cells[11] and to identify secreted proteins.[12,13] BONCAT was used to successfully identify low abundance secreted proteins of mammalian cells from an abundance of other proteins in serum containing media. In these studies BONCAT was performed using the methionine surrogate non-canonical amino acid L-azidohomoalanine (Aha) which incorporates into newly synthesized proteins at met codons. However, Aha is metabolically incorporated into all cells and therefore labels the entire population of cells without cell-type specificity. Ideally one would like to have spatial (cell-targeted) and temporal (stage-specific) control of proteome labeling for study of stem cell biology. To solve this problem, here we utilize a mutant (L274G) mouse methionyl tRNA synthetase enzyme (MMetRS) that charges the non-canonical amino acid L-azidonorleucine (Anl) in place of methionine. Only cells that express the MMetRS enzyme can incorporate Anl into newly made proteins, and therefore lineage-specific and temporally-defined proteomic labeling is achieved through controlled expression of this enzyme. We put the expression of MMetRS under the control of lineage-specific promoters that are activated during early stages of ESC differentiation and phenotype commitment. We focus on early mesodermal and cardiac differentiation of mouse ESCs because these processes can be

recapitulated in differentiating embryoid bodies (EBs), easily tracked using well-established markers[14], and provide readily accessible systems for proof-of-concept studies[15].

Results and Discussion

We aimed to perform linage-specific proteomic labeling at two transition points towards cardiac differentiation, first at the transition between Oct4+ pluripotent cells to mesoderm lineage, and then during the further differentiation of mesoderm cells towards cardiac progenitors.

For mesoderm lineage tracking, we used the promoter sequence of the Brachyury (T) gene to control MMetRS expression. Brachyury is a conserved transcription factor required for formation of mesoderm and notochord.[16] Previous promoter mapping experiments have demonstrated that the -500 bp 5' of the T gene, relative to ATG start codon, confers anterior mesoderm expression.[17] The -430:-280 5' sequence is required for T expression during early gastrulation, and the -280:-190 for late gastrulation. The sequence up to -350 is sufficient for expression in primitive streak, sequences up to -585 may function as enhancers whereas regulatory elements required for T expression in the notochord are not present until -2400 upstream of the gene[18]. We therefore used the limited promoter sequence of -645:-1 which supports transgene expression in early mesoderm cells excluding any that differentiate towards notochord. We made a reporter construct using this promoter sequence controlling the MMetRS expression and a constitutively active G418 resistance cassette for selection. This construct was electroporated into the G4 mouse ESC line to develop the BryMMetRS ES cell line expressing the MMetRS under the control of T promoter.

Cardiac-specific promoters were used to control MMetRS expression for proteomic labeling during the transition from mesodermal and cardiac progenitors. Early cardiac differentiation is accompanied by upregulation of α-myosin heavy chain (α-MHC) protein, the transcription factor Nkx2.5, and downregulation of vascular endothelial growth factor receptor Flk1.[19] We used the α-MHC promoter to control MMetRS expression. The α-MHC gene encodes a cardiac muscle-specific protein involved in active force generation. Previous work has shown that the α-MHC cardiac specific enhancer can be mapped to the -344:-156 upstream region of the gene and containing binding sites for MEF2, BF-2, SRF, ARF, MCBF and GATA-4 and possibly other transcription factors.[20] The zinc finger transcription factor GATA-4, for example, is important to mesoderm induction and required for cardiac differentiation.[21] In addition to this minimum

required 5'UTR sequence, additional upstream sequences maybe needed to obtain cardiac specificity. A recent study used a 5.5k bp fragment of the α-MHC promoter to control eGFP expression and track cardiac differentiation[22]. We therefore put this 5'UTR sequence upsteam of the MMetRS transgene for cardiac lineage-specific proteomic labeling and developed the αMHCMMetRS ES cell line in a similar manner to the BryMMetRS cell line described above (Figure 5.1a). We used these two cell lines to probe for protein labeling during EB differentiation.

Cell-targeted Anl labeled proteins can be chemoselectively tagged using 3+2 cycloaddition reaction between the azide side-chain of Anl terminal alkyne[23] or strained alkyne probes.[24,25] Cells are fixed, permeabilized and reacted with probes for *in situ* detection of newly synthesized proteins using copper (I) catalyzed alkyne-azide cycloaddition (CuAAC) reaction with fluorescent probes. For proteomic analysis, total cellular lysates are reacted using copper-free cycloaddition with a strained cyclooctyne tag containing a dithionite-cleavable biotin moiety for binding to streptavidin and enrichment by affinity chromatography (Figure 5.1b). This probe reacts with Anl residues and enables selective enrichment of proteins from targeted cells and reduces sample complexity for mass spectrometric identification.

To probe for lineage-specific incorporation of Anl we used an EB differentiation protocol. Hanging-drop EBs were initiated from αMHCMMetRS and BryMMetRS cell lines; after 48 hours in suspension, the EBs were plated onto gelatinized tissue culture plates to allow for attachment and differentiation. Differentiation was enhanced by removal of leukemia inhibitor factor (LIF) ligand which signals through the Jak/Stat3 pathway to maintain mouse ESCs in an undifferentiated state[26]. Attached EBs were cultured for a total of 15 days and during this time ESCs differentiate towards a variety of lineages resulting in a mixed population of many different cell types. Previous reports show that Brachyury expression increases up to day 4 of differentiation and gradually decreases towards day 6 of differentiation in EBs[16]. αMHC mRNA levels are detectable at day 5 of differentiation and continue to increase up to day 16[27]; protein expression is observed starting on day 6[28] and the number of cells expressing αMHC in the EBs increases up to day 18[22]. We therefore pulse-labeled the EBs with Anl for a total of 10 hours on days 4 and 14 of differentiation to capture the transition points to mesoderm and cardiac lineage. On day 14 of differentiation we observed localized pulsating cell clusters associated with cardiac progenitors within differentiating EBs. Cells were fixed and reacted with alkyne-functionalized AlexaFluor 488 dye using CuAAC reaction. Clusters of αMHC+ cells were observable on day 14 of differentiation (as determined by immunofluorescence detection with αMHC antibody), but not on days 0 and 4 (Figure 5.2a). AlexaFluor 488 labeling was restricted to αMHC+ cells showing

selective lineage-specific Anl incorporation. As expected, we did not observe any Anl incorporation in the parental G4 cell line which lacks the MMetRS construct (Figure 5.2b). Oct4 labeling of these cells showed presence of pluripotent Oct4+ cells on day 0, a marked reduction of Oct4+ cells on day 4 and a lack of pluripotent cells by day 14 of differentiation. These results were in agreement with the timeline of differentiation of mouse ESCs.[27]

Labeling of BryMMetRS cells with a Bry-specific antibody showed localized clusters of Bry+ cells on day 4 of differentiation. Co-localized labeling of the same cell clusters with AlexaFluor 488 dye showed selective incorporation of Anl into mesoderm progenitors in this cell line (Figure 5.2c). In comparing the pattern of labeling across the different cell lines, both the Bry and Oct4 antibody labeling was nuclear and the αMHC labeling was cytoplasmic as would be expected for transcriptional factors and structural proteins, respectively. AlexaFluor 488 labeling was observed throughout the cells because Anl is metabolically incorporated into the proteome, whereas αMHC labeling is cytoplasmic (Figure 5.2d) and Bry labeling is nuclear (Figure 5.2e) in these cells.

To determine the timeline of activation of labeling in the BryMMetRS cell line during hanging drop differentiation we plated day 2 hanging drop EBs of these cells onto gelatin coated plates and incubated the cells with Anl on each successive day of differentiation and reacted the cell lysates with Alkyne-TAMRA dye to probe for Anl incorporation into the proteome. We observed a marked increase in labeling of proteins on day 3 of differentiation, gradually increasing towards day 4 and starting to decrease on day 6 of differentiation (Figure 5.3). These results were in agreement with the timeline of Brachyury promoter activity in differentiating EBs and demonstrate promoter activated labeling of cellular proteomes during differentiation.[27]

Taken together these results demonstrate that promoter-controlled expression of MMetRS can be used to obtain temporal and lineage-specific proteome labeling of differentiating mouse ESCs. The use of an EB differentiation protocol for this analysis highlights the selectivity of this protein labeling approach. Although we did not optimize differentiation towards specific lineages, the efficiency of differentiation towards cardiac and mesoderm cells can be increased by addition of exogenous factors such as retinoic acid and Bone morphogenic protein-4 (BMP4), respectively.

Materials and Methods

Embryonic Stem Cell Culture

Unless otherwise state, all cell culture reagents were obtained from Gibco (Invitrogen). Low passage G4 mouse ESCs were maintained at 37° C with 5% CO_2 in mESC medium containing high-glucose DMEM supplemented with 15vol% ES-qualified fetal bovine serum, 2mM L-glutamine, 100 μM β-mercaptoethanol, 1 mM sodium pyruvate, 0.1 mM nonessential amino acids, 50 μg/mL penicillin/streptomycin, and 500 pM mouse leukemia inhibitory factor. Media was changed every day, and cells were passaged every 3 days or at 50% confluence with 0.05% trypsin-EDTA and plated on 0.1% gel-coated tissue culture flasks.

Differentiation Using Hanging Drop Embryoid Bodies (EBs)

Hanging drop EBs were initiated with 600 cells in 24 μL of mouse ESC media lacking LIF. Hanging drop EBs were incubated for 48 hours in humidified chamber maintained at 37° C with 5% CO_2 and thereafter plated onto 0.1% gel-coated tissue culture plates to allow for surface attachment and differentiation of cells.

Reporter Constructs for Development of αMHCMMetRS and BryMMetRS Cell Lines

For PCR and cloning purposes, unless otherwise state, we used chemically competent *E. coli* MegaX DH10B (Zymo Research). Plasmid DNA was purified using a Miniprep kit (Qiagen), colony selection was performed on LB-agar plates with 100 μg/mL ampicillin. All plasmids were verified by sequencing (Laragen). The endogenous sequence of mammalian MetRS enzyme was obtained from a cDNA clone from American Type Culture Collection clone ID 6414029 (ATCC). Site-directed mutagenesis (Agilent) was used to introduce the L274G mutation into the MetRS sequence.

The Brachyury promoter was obtained from Addgene plasmid 21222 and the αMHC promoter was obtained from Addgene plasmid 21229. A fusion PCR was used to connect the Bry promoter sequence to the MMetRS gene and the resulting construct was inserted between the BglII and Xho1 restriction sites of pCDNA3.1+ (Invitrogen) vector containing a gentamycin resistance cassette. This plasmid was used to construct a second plasmid in which the Bry promoter was removed and the αMHC promoter was inserted 5' to the MMetRS. Sequences of these plasmids are included in Figure S5.1 and S5.2.

Electroporation of Mouse ESCs and Development of ES Cell Lines

Plasmids were amplified in *E. coli* MegaX DH10B and purified using endotoxin-free plasmid Maxi-kit (Qiagen). Linearized plasmids for Bry and αMHC promoter controlling MMetRS

expression were electroporated into the G4 mouse ES cell line. After electroporation, approximately 300,000 cells per plated on 0.1% gel-coated 150mm petri dishes to allow for individual colony growth. G418 was added to the media at 100μM final concentration, and the media was changed every day for a total of 8 days for selection. Approximately 60 individual colonies were picked from each electroporation and expanded to make hanging drop EBs. EB differentiation was followed for 4-16 days to identify colonies in which Anl incorporation matched expression pattern of Brachyury or αMHC proteins. The expression pattern of these proteins was determined by antibody labeling of cells and confocal fluorescence microscopy imaging of differentiating EBs. Successful clones were expanded to make stocks of αMHCMMetRS and BryMMetRS cell lines.

Synthesis of Azidonorleucine and Metabolic Labeling of Cells

Azidonorleucine synthesis was based on a previous protocol for azidohomoalanine synthesis, using Boc-lysine as the starting material.[29] Briefly, 5.27 g (81.1 mmol) of sodium azide was treated with 2.7 mL (16 mmol) of distilled triflic anhydride in 13 mL of water for 2 h. The triflic azide product was extracted with 10 mL dichloromethane and added dropwise to a flask containing Boc-Lys-OH (2 g, 8.1 mmol), K_2CO_3 (1.68 g, 12.2 mmol) and $CuSO_4$ (20 mg, 0.08 mmol) in 26 mL of water and 250 mL of methanol. After 20 h at room temperature the product was extracted with ethyl acetate, redissolved in dichloromethane and purified by silica gel chromatography. After Boc deprotection with hydrochloric acid, the final product was purified by cation exchange chromatography. Anl was diluted from a stock of 100mM in deionized water, and it was added directly to the cell culture medium to a final concentration of 2mM. Cells were labeled for a total of 10 hours with Anl. After Anl pulse, cells were either fixed and permeabilized for imaging by fluorescence microscopy, or lysed with 4% SDS for analysis of proteins in cell lysates as described below.

Labeling of Cell Lysate Proteins with Alkyne-TAMRA Dye and Detection by In-gel Fluorescence Imaging

Cells were lysed as described above, and lysates were supplemented with Ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor (Roche) to reduce protease activity. Lysates were diluted with PBS to a final SDS concentration of 1% and centrifuged at 14,000 rcf for 10 min to remove cellular debris. Concentration of protein in lysates was measured using a bicinchoninic protein quantification kit (BCA assay; Pierce). The same amount of protein was used for each time-point of cell differentiation. CuAAC reaction of on lysates was

performed with the Click-IT TAMRA protein analysis kit (Invitrogen), wherein Anl residues were tagged with an alkyne-functionalized tetramethylrhodamine (TAMRA) dye. Proteins were precipitated with chloroform-methanol protein precipitation method and unreacted dye labeling components were removed by two additional washes with methanol. Proteins were electrophoresed using 12% Bis-Tris polyacrylamide gels (Invitrogen). TAMRA dye ($\lambda_{excitation}$ = 555 nm and $\lambda_{emission}$ = 580 nm) was excited at 532 nm and detected with a 580 band-pass 30 nm filter. In-gel fluorescence images were acquired on a Typhoon 9400 instrument (GE Healthcare). Non-specific detection of total cellular proteins in each lane of the gel was performed by labeling the gel with coomassie dye.

Processing of EBs and Labeling of Anl-containing Cellular Proteins by Copper Catalyzed Alkyne-Azide Cycloaddition (CuAAC)

For fluorescence confocal microscopy imaging the EBs were plated onto glass bottom 12 well plates (MatTek). After Anl labeling, cells were washed once with PBS, fixed with 3.7% formaldehyde for 15 minutes at room temperature and permeabilized for 10 minutes with methanol at 4°C. Thereafter the cells were incubated for 2 hours in PBS containing 0.25 vol% Triton X-100. CuAAC reaction and synthesis of requisite THPTA ligand were performed as described previously[29,30]. Briefly, the reaction was performed at room temperature in pH 7.4 PBS for 2 hours using a final concentration of 0.1mM copper sulfate, 0.5mM THPTA ligand, 5mM sodium ascorbate, 5mM aminoguanidine and 4 μM of alkyne functionalized AlexaFluor-488 dye. Cells were washed extensively with PBS with 0.25vol% Triton X-100 to remove the unreacted dye components and further processed as described below for antibody labeling and fluorescence confocal microscopy.

Antibody Labeling and Imaging of Differentiating EBs by Fluorescence Confocal Microscopy

Cells were fixed, permeablized and reacted with an alkyne-functionalized dye as described above. Thereafter, samples were incubated in 5 vol % FBS in PBS for 24 hours to block nonspecific binding. We used an anti-Brachyury goat IgG antibody (R&D Systems), an anti-αMHC mouse IgG antibody (Abcam) and an Oct4/POU5F1 mouse IgG antibody (BD Biosciences), all at 1:200 dilutions. These primary antibodies were diluted in 5 vol % FBS in PBS containing 0.25vol% Triton X-100. Cells were incubated with the primary antibody solution for 24 hours at 4°C. We used a donkey anti-goat AlexaFluor-594 conjugate and an anti-mouse AlexaFluor-594 conjugate (both from Cell Signal Technologies), at 1:500 and 1:1000 dilutions, respectively. Secondary

antibodies were diluted in PBS with 0.25vol% Triton X-100. Cells were incubated with secondary antibody solutions for 10 hours at room temperature and thereafter washed with PBS. Cell nuclei were stained with 300nM DAPI in PBS for 30 minutes at room temperature and washed three times with PBS before imaging. Fluorescence confocal images were obtained on a Zeiss LSM 510 microscope.

Tryptic Digestion and Sample Preparation for Mass Spectrometry

Proteins were resuspended in 200µl Tris-HCl pH 8.5 and digested with 10µL of 0.1µg/µL of Lysyl endopeptidase(Wako) in 100mM Tris-HCl pH 8.5. Sample was incubated for 4 hrs at room temperature in the dark. Thereafter 20µL of 0.5µg/µL trypsin(Wako) in water was added and the sample was incubated in the dark overnight at room temperature. This solution was acidified to 0.2% CF_3COOH and desalted based on a previously published desalting protocol by Mann can coworkers[31] using a 3ml MILI-SPE C18-SD extraction disk cartridge (3M) as follows: 1) Cartridge was first washed with 1ml of CH_3OH and centrifuged at 1500 rcf for 1min. 2) Cartridge was washed once with 0.5ml of 0.1% CF_3COOH, 70% CH_3CN in water and centrifuged at 1500 rcf for 1min. 3) The cartridge was washed with 0.1% CF_3COOH in water and centrifuged at 1500 rcf for 1 min. 4) Peptide sample was loaded to the cartridge and passed through three times, each time the cartridge was centrifuged at 150 rcf for 3min. 5) The cartridge was washed twice with 0.5 ml of 0.1% CF_3COOH in water and centrifuged at 150 rcf for 3 min. 6) To elute the desalted peptides, cartridge was washed with 0.5ml of CH_3CN in water and centrifuged at 150 rcf for 3 min, the eluent containing desalted peptides was retained. The desalted peptides were lyophilized and stored at 4°C before analysis by mass spectrometry.

Mass Spectrometry

Proteomic analyses was performed with a hybrid LTQ-Orbitrap (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source connected to an EASY-nLC II instrument (Thermo Fisher Scientific). Fractionation of peptides was performed using a 15 cm reversed phase analytical column (75 µm ID) with 3 µm C18 beads (ReproSil-Pur C18-AQ) with a 60-minute gradient from 5% to 28% acetonitrile in 0.2% formic acid, over 50 minutes, followed by 10 minutes at 80 % acetonitrile in 0.2% formic acid. The mass spectrometer was operated in data-dependent mode. Survey full scan mass spectra were acquired with a resolution of 60,000 at 400 m/z. The top ten most intense ions from the survey scan were isolated and, after the accumulation of 5000 ions, fragmented in the linear ion trap by collision induced dissociation. Precursor ion charge state screening was enabled and singly charged and unassigned charge states were

rejected. The dynamic exclusion list was enabled containing trypsin peptide sequences, with a relative mass window of 10 ppm.

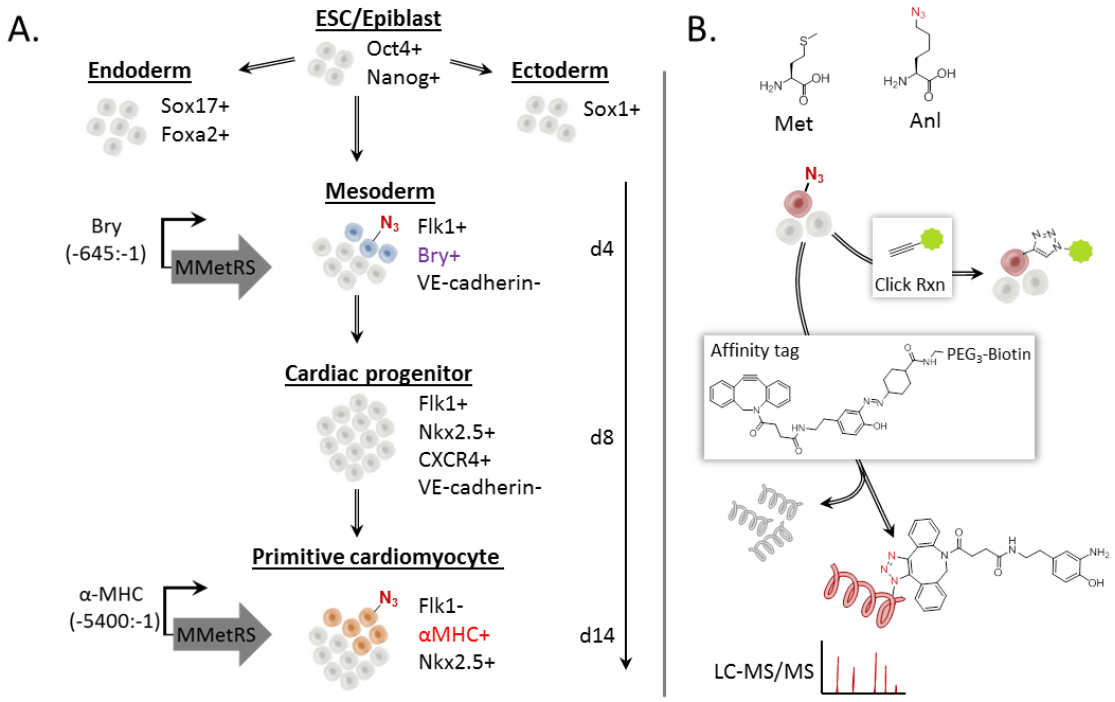


Figure 5.1. Lineage-specific labeling, detection and identification of proteins during mouse embryonic stem cell differentiation. (A) Selective expression of MMetRS under the control of brachyury and αMHC promoters in ES cells. The timeline of activity of each promoter as well as typical markers associated with each cell type are listed for each stage of differentiation. (B) Selective incorporation of methionine surrogate Anl enables detection and identification of newly synthesized proteins in targeted cell populations. Anl labeled proteins are detected by azide-alkyne cycloaddition with fluorescent dyes or enriched and identified by tandem MS. Inset shows structure of a cleavable biotin cyclooctyne tag that is used for affinity enrichment of Anl labeled proteins.

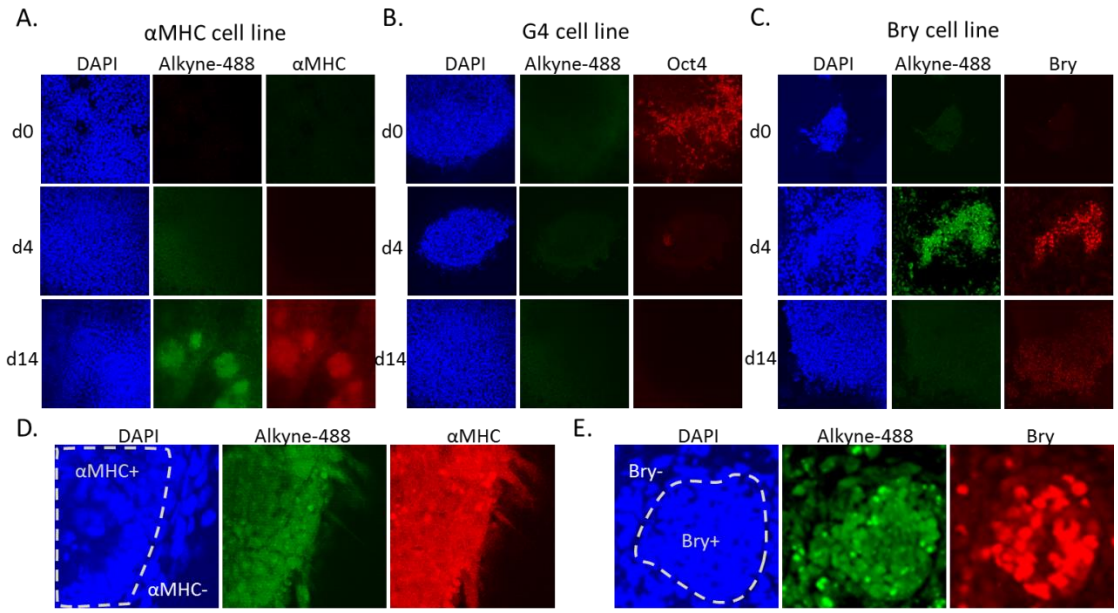


Figure 5.2. Lineage-specific proteome labeling of subpopulations of cells in differentiating embryoid bodies (A) EBs from the αMHCMMetRS cell line were differentiated for 0,4, and 14 days as indicated on the left of each image set. DAPI stains cell nuclei, Alkyne-488 indicates CuAAC labeling of Anl residues using alkyne-functionalized Alexa488 dye. Cardiac progenitor cells are labeled using an αMHC-specific antibody. (B) Differentiating EBs from the G4 cell line. Undifferentiated MESCs are labeled using an antibody for Oct4 protein. (C) Differentiating EBs from the BryMMetRS cell line. Mesoderm progenitors are labeled using a Bry-specific antibody. (D) Day 14 differentiating EBs from the αMHCMMetRS cell line showing cytoplasmic labeling of Anl residues and location of αMHC protein within cells. Arrows indicate interface of αMHC+ cells and differentiated cells of other lineages at the edge of an αMHC+ cell cluster. (E) Day 4 differentiating EBs from the BryMMetRS cell line showing Anl incorporation throughout the cells and labeling of the Bry protein which is localized to the nuclei of mesoderm progenitors.

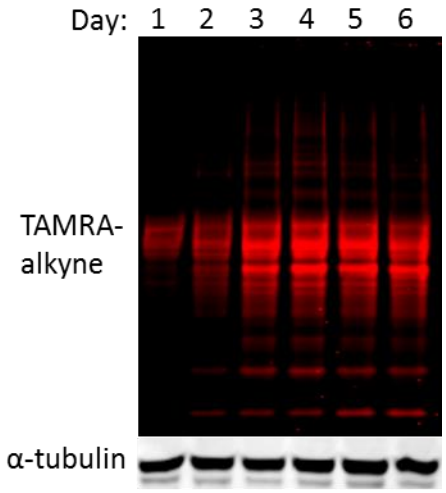
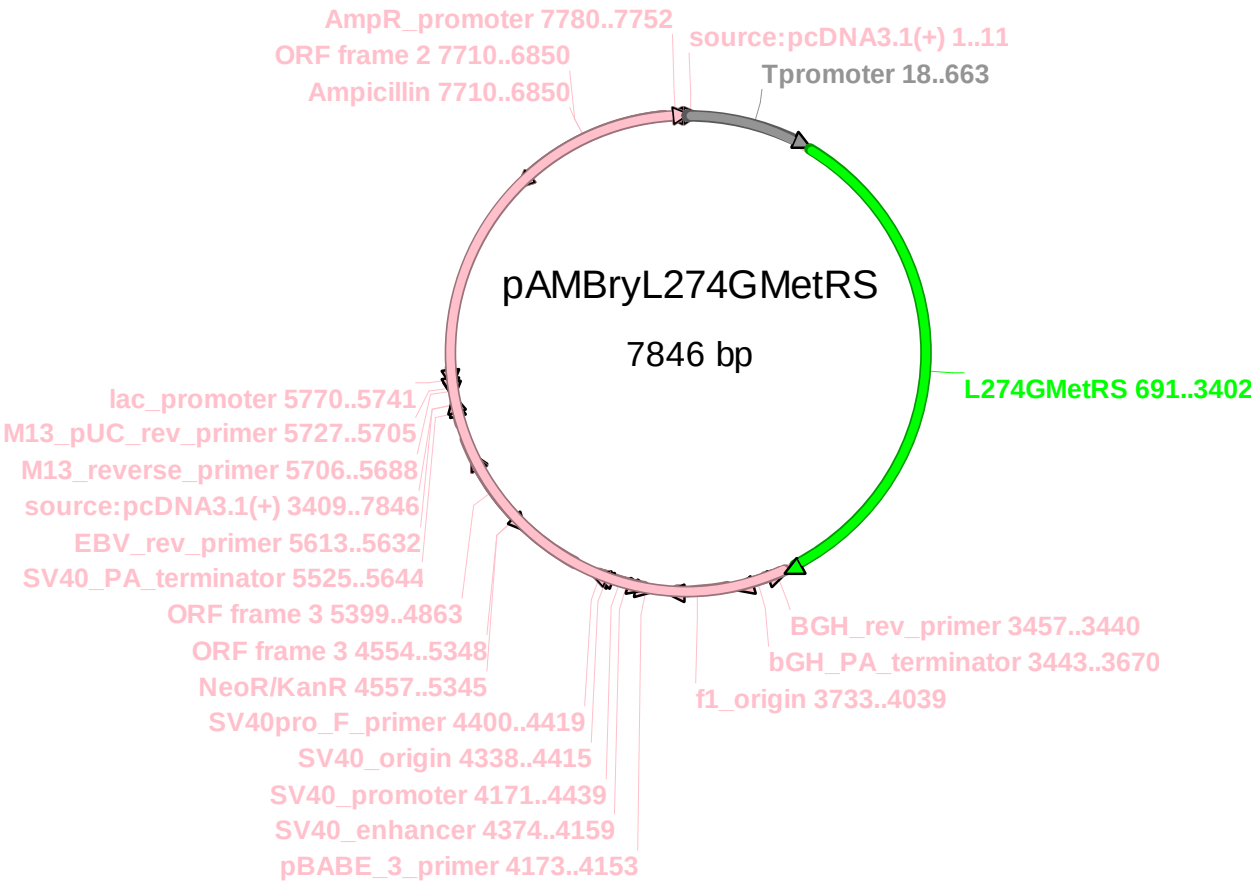


Figure 5.3. Anl incorporation in the BryMMetRS cell line during differentiation. In-gel fluorescence image of TAMRA-alkyne tagged cell lysates from the BryMMetRS cell line. Day 2 EBs were labeled with Anl on each day of differentiation indicated on the top of the image. α-tubulin serves as protein loading control.

Figure S5.1.

The pAMBryL274GMetRS vector for expression of L274G mouse MetRS under the control of Brachyury promoter.



-Tpromoter-FlagTag-L274MetRS-

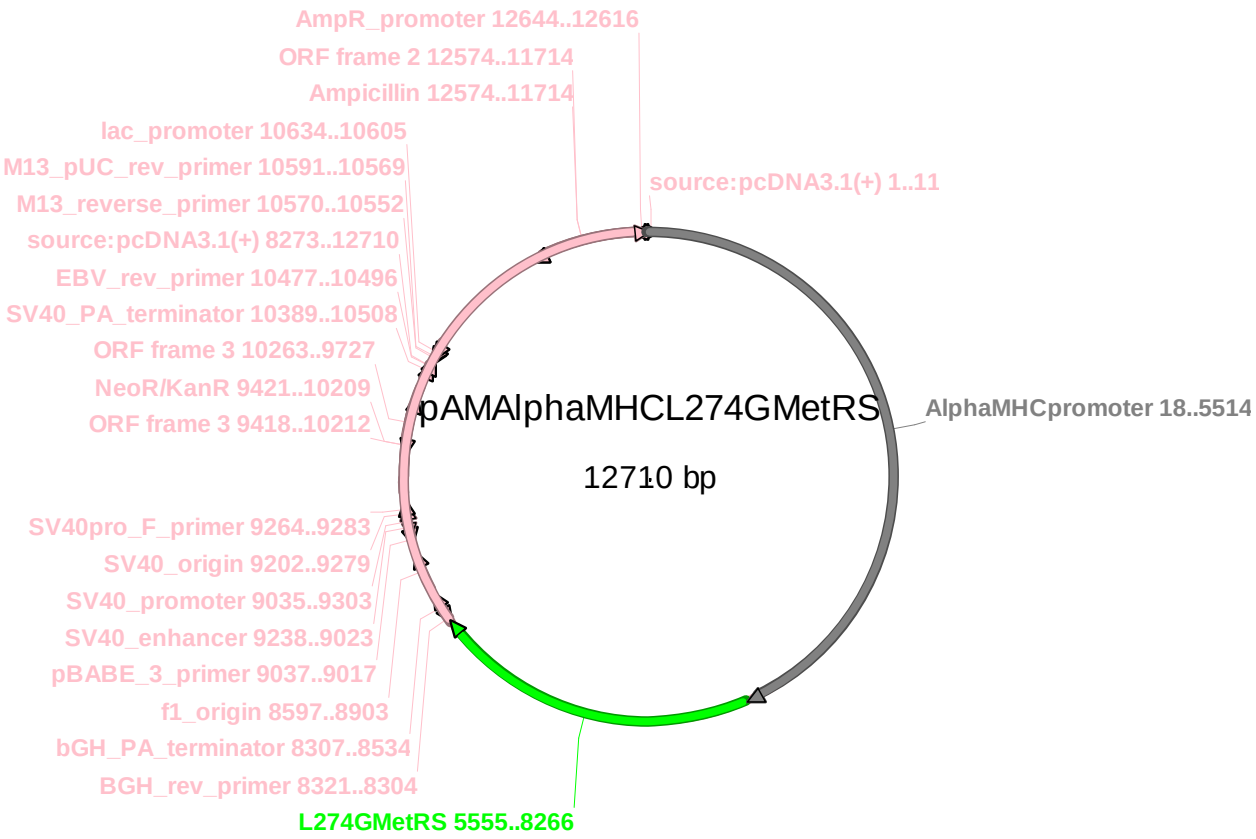
GACGGATCGGGAGATCTGGGCAAAGTCGCAGGCGCCGGTGTGCGCTTGGACAGCGCGTGGGAGTGGAGAGT
TTAGCAGTGGCTCTAGGAGCCAGGGTCCTGGGTGGCTCCAGCCCGGCTTCTCGCCCTCCCTCCCCCAGGGT
CCGCCCCGCCGCTTTGATGGAGGTGCAAACATTTGGGGGAGGGCGGGGGTGTCGGGACTGCGCCCGACGCT
TTCCTTACAGGAAGCGCGCGCTGGAGCCCATTGTTGGCCCCCAGCCTCCGGGCCCGCCCGGCCAGTCTGAC
ACGGCCGCGCACCGCCAATGGGCAGCTGCTCGGTACTTCAAAGGGTGTCCCGCCCAATCCGCCGCACCCCC
TGCCGAGGCCACCACGGCTGTATTTATGGGGAGGGGACCCATTTTTCTCTTCCCCAGAGACTTACTCTTGT
GCCGCCTTGCGGGAGTTCAAGTGGAGCCACGGCTCCCCAGGCCCTCTCCCCCATCCCCGCCCCCTTCCCCC
CTCATCCCGATCTCGGTGCTCCTTTGGCGAATGTGCAGGGACCCAGGTGTAATCTTTGGGCTCCGCAGAGT
GACCCTTTTTCTTGGAAAAGCGGTGGCGAGAGAAGTGAAGGTGGCTGTTGGGTAGGGAGTCAAGACTCCTG
GAAGGTGGAGAGGGTGGCGGGAGGATGCATTACAAGGATGATGATGATAAGATGAGACTGTTCGTGAGCGA
GGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGGCCGCGAGGGCCCGGGGTCGGGCGGAGCTGCTCATCA
GCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTCCTTACCCGGCCTAAGGTCCCTGTCTTGCAGCTGGAT
AGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTGCCGATATTTTTTTCTGTTATGTGGCTGGGAACAAGA
TGATCTCACCAACCAGTGGCTGGAATGGGAGGCAACAGAACTGCAGCCAGTTCTGTCTGCTGCCCTACACT

GTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATACTTGGCCCACTTCGGAGAGTCCTGACTCACATTGAT
CACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGCTGGGGACACAGAATCTCTAGCTGACATTGTTTTGTG
GGGAGCACTGTATCCTTTACTGCAAGACCCAGCTTACCTCCCTGAGGAGTTGGGTGCCCTGCAAAGTTGGT
TCCAGACACTGAGTACCCAGGAACCGTGTCAGCGAGCTGCAGAGACGGTGCTAAAACAGCAGGGTGTCCTG
GCACTTCGTCTGTACCTCCAGAAACAGCCACAGCCTCAGCCCCCGCCTCCTGAGGGGAGAACTGTCAGCAA
CGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTGAGGAGGACATCGTTACAGCTGTTGCCGCGTGGGAGA
AGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAGCAGCATCCAGTGTTGCCTGTGCCTGGAGAGAGGAAT
GTTCTCATCACCAGTGCCCTCCCCTATGTCAACAATGTCCCCCACCTTGGAAACATCATTGGCTGTGTGCT
CAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTCGCCAGTGGAATACCCTCTATCTGTGTGGTACAGATG
AGTATGGTACTGCGACAGAGACCAAGGCCATGGAGGAGGGCCTAACCCCACGGGAAATCTGTGACAAGTAC
CATGCCATCCATGCTGACATCTACCGCTGGTTCGGCATATCGTTCGATACTTTCGGGCGCACTACCACTCC
TCAGCAGACCAAAATCACCCAGGACATCTTCCAGAGGTTGCTGACCCGGGGGTTTGTGCTGCGAGATACTG
TGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTCCTGGCTGACCGCTTTGTGGAGGGTGTGTGTCCCTTC
TGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGACAGGTGTGGCAAGCTCATCAATGCCATTGAGCTCAA
GAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTGTGGTGAGGTCCTCACAGCACCTGTTTCTAGACTTGC
CTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGGAAGACAGTGCCTGGCAGTGACTGGACACCCAATGCC
AGGTTCATTATACGTTCCTGGCTTCGAGATGGCCTCAAGCCACGATGCATCACCAGAGACCTCAAATGGGG
AACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGGTATTTTACGTCTGGTTTGATGCTACTATTGGCTACG
TGTCCATCACAGCCAACTACACAGACCAATGGGAGAAATGGTGGAAGAACCCAGAACAAGTGGACCTTTAC
CAGTTCATGGCCAAAGACAATGTTCCCTTCCATGGCTTGGTCTTTCCGTGTTCAGTCCTAGGAGCTGAGGA
CAACTACACCCTGGTCAAGCACATCATTGCTACAGAGTACCTGAACTATGAGGATGGGAAATTCTCTAAGA
GCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAGGATACAGGAATCCCTGCTGACATCTGGCGATTCTAT
CTGCTATACATTCGGCCTGAGGGCCAGGACAGTGCCTTCTCCTGGACAGACTTGTTGATTAAAAACAATTC
TGAGCTGCTCAACAACCTGGGCAACTTCATCAACAGAGCTGGCATGTTTGTTTCTAAGTTTTTTGGCGGTT
GTGTGCCTGAGATGGCGCTAACCCCTGATGACAGACGCCTGGTGGCCCATGTCTCTTGGGAACTCCAGCAC
TATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGATGCCTTGCGCAGTATCCTCACCATATCTCGCCATGG
CAACCAATACATTCAAGTGAATGAGCCCTGGAAACGGATTAAAGGTGGTGAGATGGACAGGCAGCGGGCAG
GCACAGTGACAGGCATGGCAGTGAACATGGCTGCCTTGCTGTCTGTCATGCTGCAGCCATACATGCCCACA
GTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCCACCTGCAGCCTGCCGCATCCTTGCCACAAGCTTCAT
TTGTACCTTGCCAGCAGGCCACCGAATTGGCACAGTCAGTCCTTTGTTCCAAAAACTGGAAAATGACCAGA
TTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAGGCTAAAGGCTCCCCCAAGCCAGCAGCTGTGGAGGCA
GTTACAGCAGCAGGCTCGCAGCACATACAAACGCTGACGGATGAGGTGACCAAGCAGGGCAACGTCGTCCG
GGAACTGAAAGCACAGAAGGCAGACAAGAACCAGGTTGCTGCAGAGGTGGCTAAACTCTTGGATCTAAAGA
AACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAAACTCCTAAAGGCAAGAAGAAAAAGTGATAACTCGAG
TCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTG
CCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAA
TTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAG
GATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAG
CTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGC
GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCC
ACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACG
GCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTT
TTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAAC
CCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCT
GATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGG
CTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAG
GCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACT
CCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTAT
TTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCC

TAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGA
TCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGG
CTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCC
CGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCG
TGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCT
GCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCA
TGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACAT
CGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCA
GGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGA
CCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGC
CGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGG
CGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATC
GCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGC
CATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCC
GGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGC
TTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTA
GTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTG
GCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGC
CGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCAC
TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGC
GGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCG
AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
CGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG
ATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC
TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTG
TAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGG
TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGA
TTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGA
AGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATC
CGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGAT
CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATT
TTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAAT
CTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
TCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTA
CCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA
CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATT
GTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGC
ATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGG
CCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGC
TTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTG
CCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTT
CTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC
AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCA
TTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTT
CCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

Figure S5.2.

The pAMAlphaMHCL274GMetRS vector for expression of L274G mouse MetRS under the control of α-MHC promoter.



-AlphaMHCpromoter-FlagTag-L274MetRS-

GACGGATCGGGCTCGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCTGCAAGGTCACACAA
GGGTCTCCACCCACCAGGTGCCCTAGTCTCAATTTCAGTTTCCATGCCTTGTTCTCACAATGCTGGCCTCC
CCAGAGCTAATTTGGACTTTGTTTTTATTTCAAAAGGGCCTGAATGAGGAGTAGATCTTGTGCTACCCAGC
TCTAAGGGTGCCCGTGAAGCCCTCAGACCTGGAGCCTTTGCAACAGCCCTTTAGGTGGAAGCAGAATAAAG
CAATTTTCCTTAAAGCCAAAATCCTGCCTCTAGACTCTTCTTCTCTGACCTCGGTCCCTGGGCTCTAGGGT
GGGGAGGTGGGGCTTGGAAGAAGAAGGTGGGGAAGTGGCAAAAGCCGATCCCTAGGGCCCTGTGAAGTTCG
GAGCCTTCCCTGTACAGCACTGGCTCATAGATCCTCCTCCAGCCAAACATAGCAAGAAGTGATACCTCCTT
TGTGACTTCCCCAGGCCCAGTACCTGTCAGGTTGAAACAGGATTTAGAGAAGCCTCTGAACTCACCTGAAC
TCTGAAGCTCATCCACCAAGCAAGCACCTAGGTGCCACTGCTAGTTAGTATCCTACGCTGATAATATGCAG
AGCTGGGCCACAGAAGTCCTGGGGTGTAGGAACTGACCAGTGACTTTTCAGTCGGCAAAGGTATGACCCCC
TCAGCAGATGTAGTAATGTCCCCTTAGATCCCATCCCAGGCAGGTCTCTAAGAGGACATGGGATGAGAGAT
GTAGTCATGTGGCATTCCAAACACAGCTATCCACAGTGTCCCTTGCCCCTTCCACTTAGCCAGGAGGACAG
TAACCTTAGCCTATCTTTCTTCCTCCCCATCCTCCCAGGACACACCCCCTGGTCTGCAGTATTCATTTCTT
CCTTCACGTCCCCTCTGTGACTTCCATTTGCAAGGCTTTTGACCTCTGCAGCTGCTGGAAGATAGAGTTTG
GCCCTAGGTGTGGCAAGCCATCTCAAGAGAAAGCAGACAACAGGGGGACCAGATTTTGGAAGGATCAGGAA
CTAAATCACTGGCGGGCCTGGGGGTAGAAAAAAGAGTGAGTGAGTCCGCTCCAGCTAAGCCAAGCTAGTCC
CCGAGATACTCTGCCACAGCTGGGCTGCTCGGGGTAGCTTTAGGAATGTGGGTCTGAAAGACAATGGGATT
GGAAGACATCTCTTTGAGTCTCCCCTCAACCCCACCTACAGACACACTCGTGTGTGGCCAGACTCCTGTTC
AACAGCCCTCTGTGTTCTGACCACTGAGCTAGGCAACCAGAGCATGGGCCCTGTGCTGAGGATGAAGAGTT

GGTTACCAATAGCAAAAACAGCAGGGGAGGGAGAACAGAGAACGAAATAAGGAAGGAAGAAGGAAAGGCCA
GTCAATCAGATGCAGTCAGAAGAGATGGGAAGCCAACACACAGCTTGAGCAGAGGAAACAGAAAAGGGAGA
GATTCTGGGCATAAGGAGGCCACAGAAAGAAGAGCCCAGGCCCCCCAAGTCTCCTCTTTATACCCTCATCC
CGTCTCCCAATTAAGCCCACTCTTCTTCCTAGATCAGACCTGAGCTGCAGCGAAGAGACCCGTAGGGAGGA
TCACACTGGATGAAGGAGATGTGTGGAGAAGTCCAGGGCAACCTAAGAGCCAGAGCCTAAAAGAGCAAGAG
ATAAAGGTGCTTCAAAGGTGGCCAGGCTGTGCACACAGAGGGTCGAGGACTGGTGGTAGAGCCTCAAGATA
AGGATGATGCTCAGAATGGGCGGGGGGGGGGGATTCTGGGGGGGGGGAGAGAGAAGGTGAGAAGGAGCCTGGA
ACAGAGAATCTGGAAGCGCTGGAAACGATACCATAAAGGGAAGAACCCAGGCTACCTTTAGATGTAAATCA
TGAAAGACAGGGAGAAGGGAAGCTGGAGAGAGTAGAAGGACCCCGGGGCAAGACATGGAAGCAAGGACAAG
CCAGGTTGAGCGCTCCGTGAAATCAGCCTGCTGAAGGCAGAGCCCTGGTATGAGCACCAGAACAGCAGAGG
CTAGGGTTAATGTCGAGACAGGGAACAGAAGGTAGACACAGGAACAGACAGAGACGGGGGAGCCAGGTAAC
AAAGGAATGGTCCTTCTCACCTGTGGCCAGAGCGTCCATCTGTGTCCACATACTCTAGAATGTTCATCAGA
CTGCAGGGCTGGCTTGGGAGGCAGCTGGAAAGAGTATGTGAGAGCCAGGGGAGACAAGGGGGCCTAGGAAA
GGAAGAAGAGGGCAAACCAGGCCACACAAGAGGGCAGAGCCCAGAACTGAGTTAACTCCTTCCTTGTTGCA
TCTTCCATAGGAGGCAGTGGGAACTCTGTGACCACCATCCCCCATGAGCCCCCACTACCCATACCAAGTTT
GGCCTGAGTGGCATTCTAGGTTCCCTGAGGACAGAGCCTGGCCTTTGTCTCTTGGACCTGACCCAAGCTGA
CCCAATGTTCTCAGTACCTTATCATGCCCTCAAGAGCTTGAGAACCAGGCAGTGACATATTAGGCCATGGG
CTAACCCTGGAGCTTGCACACAGGAGCCTCAAGTGACCTCCAGGGACACAGCTGCAGACAGGTGGCCTTTA
TCCCCAAAGAGCAACCATTTGGCATAGGTGGCTGCAAATGGGAATGCAAGGTTGAATCAGGTCCCTTCAAG
AATACTGCATGCAAGACCTAAGACCCCTGGAGAGAGGGGTATGCTCCTGCCCCCACCCACCATAAGGGGAG
TGAACTATCCTAGGGGGCTGGCGACCTTGGGGAGACACCACATTACTGAGAGTGCTGAGCCCAGAAAAACT
GACCGCCCTGTGTCCTGCCCACCTCCACACTCTAGAGCTATATTGAGAGGTGACAGTAGATAGGGTGGGAG
CTGGTAGCAGGGAGAGTGTTCCTGGGTGTGAGGGTGTAGGGGAAAGCCAGAGCAGGGGAGTCTGGCTTTGT
CTCCTGAACACAATGTCTACTTAGTTATAACAGGCATGACCTGCTAAAGACCCAACATCTACGACCTCTGA
AAAGACAGCAGCCCTGGAGGACAGGGGTTGTCTCTGAGCCTTGGGTGCTTGATGGTGCCACAAAGGAGGGC
ATGAGTGTGAGTATAAGGCCCCAGGAGCGTTAGAGAAGGGCACTTGGGAAGGGGTCAGTCTGCAGAGCCCC
TATCCATGGAATCTGGAGCCTGGGGCCAACTGGTGTAAATCTCTGGGCCTGCCAGGCATTCAAAGCAGCAC
CTGCATCCTCTGGCAGCCTGGGGAGGCGGAAGGGAGCAACCCCCCACTTATACCCTTTCTCCCTCAGCCCC
AGGATTAACACCTCTGGCCTTCCCCCTTCCCACCTCCCATCAGGAGTGGAGGGTTGCAGAGGGAGGGTAAA
AACCTACATGTCCAAACATCATGGTGCACGATATATGGATCAGTATGTGTAGAGGCAAGAAAGGAAATCTG
CAGGCTTAACTGGGTTAATGTGTAAAGTCTGTGTGCATGTGTGTGTGTCTGACTGAAAACGGGCATGGCTG
TGCAGCTGTTCAGTTCTGTGCGTGAGGTTACCAGACTGCAGGTTTGTGTGTAAATTGCCCAAGGCAAAGTG
GGTGAATCCCTTCCATGGTTTAAAGAGATTGGATGATGGCCTGCATCTCAAGGACCATGGAAAATAGAATG
GACACTCTATATGTGTCTCTAAGCTAAGGTAGCAAGGTCTTTGGAGGACACCTGTCTAGAGATGTGGGCAA
CAGAGACTACAGACAGTATCTGTACAGAGTAAGGAGAGAGAGGAGGGGGTGTAGAATTCTCTTACTATCAA
AGGGAAACTGAGTCGTGCACCTGCAAAGTGGATGCTCTCCCTAGACATCATGACTTTGTCTCTGGGGAGCC
AGCACTGTGGAACTTCAGGTCTGAGAGAGTAGGAGGCTCCCCTCAGCCTGAAGCTATGCAGATAGCCAGGG
TTGAAAGGGGGAAGGGAGAGCCTGGGATGGGAGCTTGTGTGTTGGAGGCAGGGGACAGATATTAAGCCTGG
AAGAGAAGGTGACCCTTACCCAGTTGTTCAACTCACCCTTCAGATTAAAAATAACTGAGGTAAGGGCCTGG
GTAGGGGAGGTGGTGTGAGACGCTCCTGTCTCTCCTCTATCTGCCCATCGGCCCTTTGGGGAGGAGGAATG
TGCCCAAGGACTAAAAAAAGGCCATGGAGCCAGAGGGGCGAGGGCAACAGACCTTTCATGGGCAAACCTTG
GGGCCCTGCTGTCCTCCTGTCACCTCCAGAGCCAAGGGATCAAAGGAGGAGGAGCCAGGACAGGAGGGAAG
TGGGAGGGAGGGTCCCAGCAGAGGACTCCAAATTTAGGCAGCAGGCATATGGGATGGGATATAAAGGGGCT
GGAGCACTGAGAGCTGTCAGAGATTTCTCCAACCCAGGTAAGAGGGAGTTTCGGGTGGGGGCTCTTCACCC
ACACCAGACCTCTCCCCACCTAGAAGGAAACTGCCTTTCCTGGAAGTGGGGTTCAGGCCGGTCAGAGATCT
GACAGGGTGGCCTTCCACCAGCCTGGGAAGTTCTCAGTGGCAGGAGGTTTCCACAAGAAACACTGGATGCC
CCTTCCCTTACGCTGTCTTCTCCATCTTCCTCCTGGGGATGCTCCTCCCCGTCTTGGTTTATCTTGGCTCT
TCGTCTTCAGCAAGATTTGCCCTGTGCTGTCCACTCCATCTTTCTCTACTGTCTCCGTGCCTTGCCTTGCC
TTCTTGCGTGTCCTTCCTTTCCACCCATTTCTCACTTCACCTTTTCTCCCCTTCTCATTTGTATTCATCCT
TCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCTCCCTTCCTTCCTTCCTTCCTTCC
TTCCTTCCTTCCTTCCTTCCTGTGTCAGAGTGCTGAGAATCACACCTGGGGTTCCCACCCTTATGTAAACA
ATCTTCCAGTGAGCCACAGCTTCAGTGCTGCTGGGTGCTCTCTTACCTTCCTCACCCCCTGGCTTGTCCTG
TTCCATCCTGGTCAGGATCTCTAGATTGGTCTCCCAGCCTCTGCTACTCCTCTTCCTGCCTGTTCCTCTCT
CTGTCCAGCTGCGCCACTGTGGTGCCTCGTTCCAGCTGTGGTCCACATTCTTCAGGATTCTCTGAAAAGTT
AACCAGGTGAGAATGTTTCCCCTGTAGACAGCAGATCACGATTCTCCCGGAAGTCAGGCTTCCAGCCCTCT
CTTTCTCTGCCCAGCTGCCCGGCACTCTTAGCAAACCTCAGGCACCCTTACCCCACATAGACCTCTGACAG
AGAAGCAGGCACTTTACATGGAGTCCTGGTGGGAGAGCCATAGGCTACGGTGTAAAAGAGGCAGGGAAGTG

GTGGTGTAGGAAAGTCAGGACTTCACATAGAAGCCTAGCCCACACCAGAAATGACAGACAGATCCCTCCTA
TCTCCCCCATAAGAGTTTGAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGGGAGGATGCCATTACAA
GGATGATGATGATAAGATGAGACTGTTCGTGAGCGAGGGTTCCCCGGGGAGCCTGCCCGTGCTGGCTGCGG
CCGCGAGGGCCCGGGGTCGGGCGGAGCTGCTCATCAGCACCGTAGGCCCCGAAGAGTGTGTGGTACCATTC
CTTACCCGGCCTAAGGTCCCTGTCTTGCAGCTGGATAGTGGCAACTACCTCTTCTCTGCTAGTGCAATCTG
CCGATATTTTTTTCTGTTATGTGGCTGGGAACAAGATGATCTCACCAACCAGTGGCTGGAATGGGAGGCAA
CAGAACTGCAGCCAGTTCTGTCTGCTGCCCTACACTGTCTAGTGGTTCAAGGCAAGAAAGGGGAAGATATA
CTTGGCCCACTTCGGAGAGTCCTGACTCACATTGATCACAGCTTGAGTCGTCAAAACTGTCCTTTCCTGGC
TGGGGACACAGAATCTCTAGCTGACATTGTTTTGTGGGGAGCACTGTATCCTTTACTGCAAGACCCAGCTT
ACCTCCCTGAGGAGTTGGGTGCCCTGCAAAGTTGGTTCCAGACACTGAGTACCCAGGAACCGTGTCAGCGA
GCTGCAGAGACGGTGCTAAAACAGCAGGGTGTCCTGGCACTTCGTCTGTACCTCCAGAAACAGCCACAGCC
TCAGCCCCCGCCTCCTGAGGGGAGAACTGTCAGCAACGAGCTGGAGGAAGAGGAACTGGCTACCTTGTCTG
AGGAGGACATCGTTACAGCTGTTGCCGCGTGGGAGAAGGGTCTGGAAAGCCTGCCTCCGCTAAAGCTCCAG
CAGCATCCAGTGTTGCCTGTGCCTGGAGAGAGGAATGTTCTCATCACCAGTGCCCTCCCCTATGTCAACAA
TGTCCCCCACCTTGGAAACATCATTGGCTGTGTGCTCAGTGCTGATGTCTTTGCAAGGTATTGTCGCCTTC
GCCAGTGGAATACCCTCTATCTGTGTGGTACAGATGAGTATGGTACTGCGACAGAGACCAAGGCCATGGAG
GAGGGCCTAACCCCACGGGAAATCTGTGACAAGTACCATGCCATCCATGCTGACATCTACCGCTGGTTCGG
CATATCGTTCGATACTTTCGGGCGCACTACCACTCCTCAGCAGACCAAAATCACCCAGGACATCTTCCAGA
GGTTGCTGACCCGGGGGTTTGTGCTGCGAGATACTGTGGAGCAGCTTCGGTGTGAGCGGTGTGCACGTTTC
CTGGCTGACCGCTTTGTGGAGGGTGTGTGTCCCTTCTGTGGCTATGAAGAGGCCCGAGGTGACCAGTGTGA
CAGGTGTGGCAAGCTCATCAATGCCATTGAGCTCAAGAAACCACAGTGCAAAATCTGCCGCTCCTGCCCTG
TGGTGAGGTCCTCACAGCACCTGTTTCTAGACTTGCCTAAGTTGGAAAAGCGTCTGGAGGACTGGTTGGGG
AAGACAGTGCCTGGCAGTGACTGGACACCCAATGCCAGGTTCATTATACGTTCCTGGCTTCGAGATGGCCT
CAAGCCACGATGCATCACCAGAGACCTCAAATGGGGAACGCCTGTGCCCTTGGAAGGTTTTGAGGACAAGG
TATTTTACGTCTGGTTTGATGCTACTATTGGCTACGTGTCCATCACAGCCAACTACACAGACCAATGGGAG
AAATGGTGGAAGAACCCAGAACAAGTGGACCTTTACCAGTTCATGGCCAAAGACAATGTTCCCTTCCATGG
CTTGGTCTTTCCGTGTTCAGTCCTAGGAGCTGAGGACAACTACACCCTGGTCAAGCACATCATTGCTACAG
AGTACCTGAACTATGAGGATGGGAAATTCTCTAAGAGCCGGGGCATAGGAGTGTTTGGAGACATGGCCAAG
GATACAGGAATCCCTGCTGACATCTGGCGATTCTATCTGCTATACATTCGGCCTGAGGGCCAGGACAGTGC
CTTCTCCTGGACAGACTTGTTGATTAAAAACAATTCTGAGCTGCTCAACAACCTGGGCAACTTCATCAACA
GAGCTGGCATGTTTGTTTCTAAGTTTTTTGGCGGTTGTGTGCCTGAGATGGCGCTAACCCCTGATGACAGA
CGCCTGGTGGCCCATGTCTCTTGGGAACTCCAGCACTATCACCAGCTGTTGGAAAAGGTTCGGATCCGGGA
TGCCTTGCGCAGTATCCTCACCATATCTCGCCATGGCAACCAATACATTCAAGTGAATGAGCCCTGGAAAC
GGATTAAAGGTGGTGAGATGGACAGGCAGCGGGCAGGCACAGTGACAGGCATGGCAGTGAACATGGCTGCC
TTGCTGTCTGTCATGCTGCAGCCATACATGCCCACAGTCAGCTCTACCATCCAGACCCAGCTGCAGCTCCC
ACCTGCAGCCTGCCGCATCCTTGCCACAAGCTTCATTTGTACCTTGCCAGCAGGCCACCGAATTGGCACAG
TCAGTCCTTTGTTCCAAAAACTGGAAAATGACCAGATTGAAAATTTGAGGCAGCGCTTTGGAGGGGGTCAG
GCTAAAGGCTCCCCCAAGCCAGCAGCTGTGGAGGCAGTTACAGCAGCAGGCTCGCAGCACATACAAACGCT
GACGGATGAGGTGACCAAGCAGGGCAACGTCGTCCGGGAACTGAAAGCACAGAAGGCAGACAAGAACCAGG
TTGCTGCAGAGGTGGCTAAACTCTTGGATCTAAAGAAACAGTTGGCTTTGGCTGAGGGGAAACCCATTGAA
ACTCCTAAAGGCAAGAAGAAAAAGTGATAACTCCAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTC
GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTG
CCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATT
CTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGC
GGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTA
GCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCG
CCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATG
GTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAAT
AGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGAT
TTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTG
GAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATC
TCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCAT
CTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGC
CCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGC
TATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATA
TCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGC

AGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTG
ATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCC
CTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGT
GCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGT
CATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGAT
CCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGG
TCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCA
AGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTG
GAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGC
GTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTA
TCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGG
GGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTAT
GAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCT
GGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAA
ATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAT
CATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA
TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAAT
GAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAG
CTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGT
TATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGT
AAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT
TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGA
ACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACG
ACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAG
TTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCC
AGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTG
TTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCT
GACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTT
ACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTC
CCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC
CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTC
ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG
CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATT
CTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATA
GCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTG
TTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGT
TTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAA
TACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATA
TTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGT
C

Figure S5.3.

Immunofluorescence detection of αMHC and Brachyury in undifferentiated mouse ESC lines. Each cell line, indicated on the left of images was labeled with either αMHC antibody or Brachyury antibody for detection of background levels of these proteins in undifferentiated cells. DAPI dye stains cell nuclei in these confocal fluorescence images.

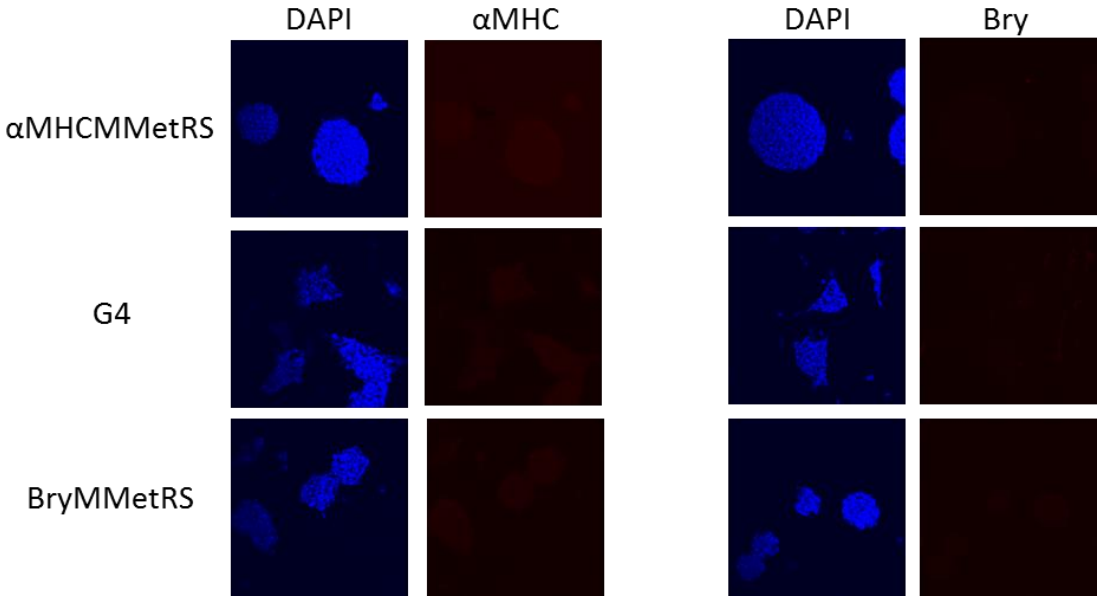
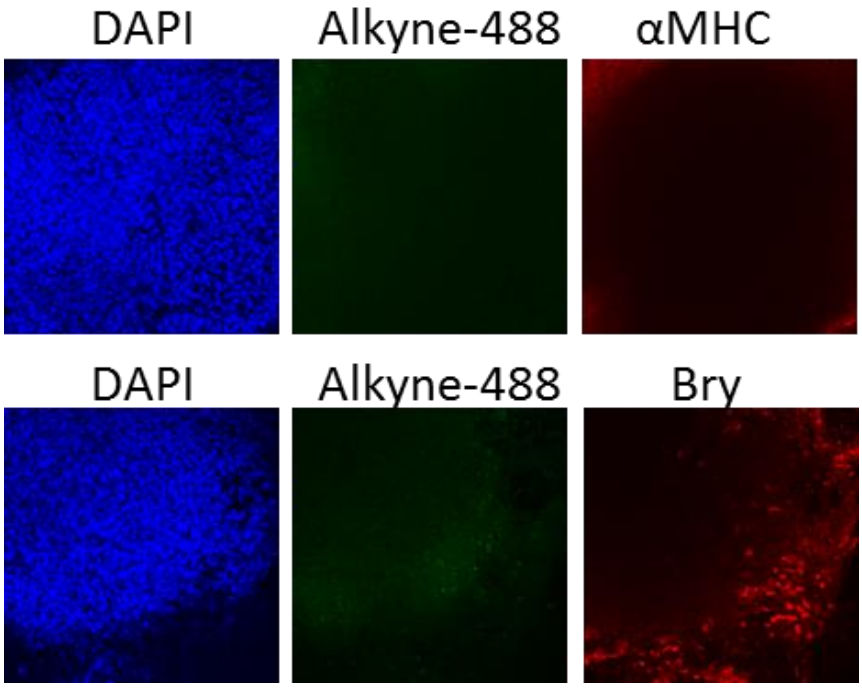


Figure S5.4.

Confocal fluorescence imaging of EBs from αMHCMMetRS cell line. Differentiating αMHCMMetRS cells on day 4 of differentiation were fixed, permeablized, and labeled with alkyne functionalized Alexa 488 dye and labeled with either the αMHC or the Brachyury antibodies. DAPI dye was used to stain cell nuclei.



References

(1) Wang, H.; Yang, H.; Shivalila, C. S.; Dawlaty, M. M.; Cheng, A. W.; Zhang, F.; Jaenisch, R. *Cell* **2013**, *153*, 910.

(2) Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. *Science* **2013**, *339*, 823.

(3) Hansson, J.; Krijgsveld, J. *Curr. Opin. Genet. Dev.* **2013**, *23*, 540.

(4) Buszczak, M.; Signer, R. A.; Morrison, S. J. *Cell* **2014**, *159*, 242.

(5) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol Cell Proteomics* **2002**, *1*, 376.

(6) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482.

(7) Reiland, S.; Salekdeh, G. H.; Krijgsveld, J. *Expert review of proteomics* **2011**, *8*, 29.

(8) Hansson, J.; Rafiee, M. R.; Reiland, S.; Polo, J. M.; Gehring, J.; Okawa, S.; Huber, W.; Hochedlinger, K.; Krijgsveld, J. *Cell reports* **2012**, *2*, 1579.

(9) van Hoof, D.; Krijgsveld, J.; Mummery, C. *Cold Spring Harbor perspectives in biology* **2012**, *4*.

(10) Klimmeck, D.; Hansson, J.; Raffel, S.; Vakhrushev, S. Y.; Trumpp, A.; Krijgsveld, J. *Mol Cell Proteomics* **2012**, *11*, 286.

(11) Howden, A. J.; Geoghegan, V.; Katsch, K.; Efstathiou, G.; Bhushan, B.; Boutureira, O.; Thomas, B.; Trudgian, D. C.; Kessler, B. M.; Dieterich, D. C.; Davis, B. G.; Acuto, O. *Nat Methods* **2013**, *10*, 343.

(12) Eichelbaum, K.; Winter, M.; Berriel Diaz, M.; Herzig, S.; Krijgsveld, J. *Nat Biotechnol* **2012**, *30*, 984.

(13) Eichelbaum, K.; Krijgsveld, J. *Methods Mol Biol* **2014**, *1174*, 101.

(14) Niebruegge, S.; Bauwens, C. L.; Peerani, R.; Thavandiran, N.; Masse, S.; Sevaptisidis, E.; Nanthakumar, K.; Woodhouse, K.; Husain, M.; Kumacheva, E.; Zandstra, P. W. *Biotechnol. Bioeng.* **2009**, *102*, 493.

(15) Discher, D. E.; Mooney, D. J.; Zandstra, P. W. *Science* **2009**, *324*, 1673.

(16) Evans, A. L.; Faial, T.; Gilchrist, M. J.; Down, T.; Vallier, L.; Pedersen, R. A.; Wardle, F. C.; Smith, J. C. *PLoS One* **2012**, *7*, e33346.

(17) Clements, D.; Taylor, H. C.; Herrmann, B. G.; Stott, D. *Mech. Dev.* **1996**, *56*, 139.

(18) Yamaguchi, H.; Niimi, T.; Kitagawa, Y.; Miki, K. *Dev. Growth Differ.* **1999**, *41*, 253.

(19) Narazaki, G.; Uosaki, H.; Teranishi, M.; Okita, K.; Kim, B.; Matsuoka, S.; Yamanaka, S.; Yamashita, J. K. *Circulation* **2008**, *118*, 498.

(20) Molkentin, J. D.; Jobe, S. M.; Markham, B. E. *J. Mol. Cell. Cardiol.* **1996**, *28*, 1211.

(21) Zhao, R.; Watt, A. J.; Battle, M. A.; Li, J.; Bondow, B. J.; Duncan, S. A. *Dev. Biol.* **2008**, *317*, 614.

(22) Takahashi, T.; Lord, B.; Schulze, P. C.; Fryer, R. M.; Sarang, S. S.; Gullans, S. R.; Lee, R. T. *Circulation* **2003**, *107*, 1912.

(23) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596.

(24) Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A. *Chembiochem* **2010**, *11*, 2092.

(25) Liang, Y.; Mackey, J. L.; Lopez, S. A.; Liu, F.; Houk, K. N. *J. Am. Chem. Soc.* **2012**, *134*, 17904.

(26) Mahdavi, A.; Davey, R. E.; Bhola, P.; Yin, T.; Zandstra, P. W. *PLoS Comput Biol* **2007**, *3*, e130.

(27) Koike, M.; Sakaki, S.; Amano, Y.; Kurosawa, H. *J. Biosci. Bioeng.* **2007**, *104*, 294.

(28) Doss, M. X.; Winkler, J.; Chen, S.; Hippler-Altenburg, R.; Sotiriadou, I.; Halbach, M.; Pfannkuche, K.; Liang, H.; Schulz, H.; Hummel, O.; Hubner, N.; Rottscheidt, R.; Hescheler, J.; Sachinidis, A. *Genome biology* **2007**, *8*, R56.

(29) Link, A. J.; Vink, M. K.; Tirrell, D. A. *Nat. Protoc.* **2007**, *2*, 1884.

(30) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 9879.

(31) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat Methods* **2009**, *6*, 359.

CONCLUDING REMARKS

The technologies and methods introduced in this thesis provide versatile tools for the comprehensive and unbiased detection and identification of newly synthesized proteins in complex multicellular systems. Time-resolved, genetically encoded and spatially defined non-canonical amino acid incorporation enables the identification of proteins involved in cell-cell interactions and the proteins made during specific cell states.

Future advances in mass spectrometers and chemical biology techniques will enhance our ability to probe the inner workings of the cell. For example, the cell-selective labeling of newly synthesized proteins will continue to provide a unique approach towards identifying proteins made by specific subpopulations of genomically homogenous cells. Without labeling, proteins of these cells are indistinguishable in mass spectrometric analyses. Even if the instrumentation advances obviate the need for affinity enrichment, the labeling that is afforded by non-canonical amino acid incorporation is indispensable for accurate identification of proteins made by specific cell types. For example, analyzing proteins that are secreted by subpopulations of cells during hematopoietic stem cell differentiation may provide insight into methods to expand and control the differentiation of these cells. Applying the same approaches to neurobiology will provide a detailed map of the brain proteome and the dynamics of protein synthesis during learning and memory formation. These applications demonstrate exciting possibilities and future directions.

The labeling strategies introduced here can be readily combined with various quantitative proteomics methods and analytical techniques to probe posttranslational modifications. These combined approaches will increase the scope of problems that can be investigated by mass spectrometry. Improvements in bioorthogonal chemistries will help to increasingly move towards *in vivo* labeling techniques that enable the tagging and analysis of proteins in their native states. In addition to the intended use of these tools for basic research, it is anticipated that these technologies will help develop diagnostic methods and new therapies for a variety of diseases including brain disorders, age-related illnesses and cancer. In the distant future, it is hoped that these technologies will help decrease the gap between humans and intelligent machines, perhaps by revealing more about the human brain and providing molecular interfaces for machines.

APPENDIX A

Cleavable Biotin Probes for Labeling of Biomolecules via Azide-Alkyne Cycloaddition

Abstract

The azide−alkyne cycloaddition provides a powerful tool for bio-orthogonal labeling of proteins, nucleic acids, glycans, and lipids. In some labeling experiments, e.g., in proteomic studies involving affinity purification and mass spectrometry, it is convenient to use cleavable probes that allow release of labeled biomolecules under mild conditions. Five cleavable biotin probes are described for use in labeling of proteins and other biomolecules via azide−alkyne cycloaddition. Subsequent to conjugation with metabolically labeled protein, these probes are subject to cleavage with either 50 mM $Na_2S_2O_4$, 2% $HOCH_2CH_2SH$, 10% HCO_2H, 95% CF_3CO_2H, or irradiation at 365 nm. Most strikingly, a probe constructed around a dialkoxydiphenylsilane (DADPS) linker was found to be cleaved efficiently when treated with 10% HCO_2H for 0.5 h. A model green fluorescent protein was used to demonstrate that the DADPS probe undergoes highly selective conjugation and leaves a small (143 Da) mass tag on the labeled protein after cleavage. These features make the DADPS probe especially attractive for use in biomolecular labeling and proteomic studies.

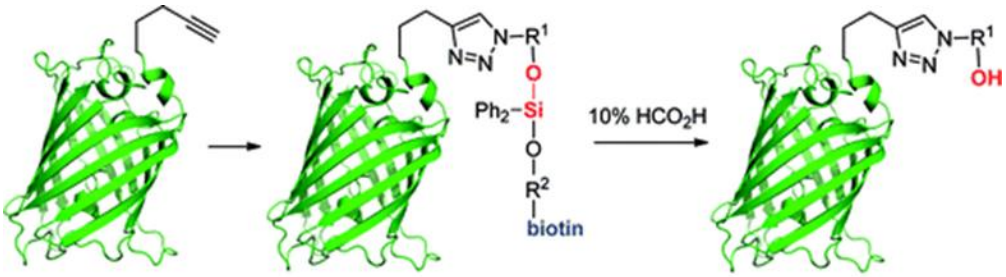


Figure AA. DADPS probe for tagging of biomolecules via azide−alkyne cycloaddition.

This work was reported in:

Szychowski, J.; Mahdavi, A.; Hodas, J. J.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *J. Am. Chem. Soc.* **2010**, *132*, 18351.

APPENDIX B

A Magnetic Cell-Based Sensor

Abstract

Cell-based sensing represents a new paradigm for performing direct and accurate detection of cell- or tissue-specific responses by incorporating living cells or tissues as an integral part of a sensor. Here we report a new magnetic cell-based sensing platform by combining magnetic sensors implemented in the complementary metal-oxide-semiconductor (CMOS) integrated microelectronics process with cardiac progenitor cells that are differentiated directly on-chip. We show that the pulsatile movements of on-chip cardiac progenitor cells can be monitored in a real-time manner. Our work provides a new low-cost approach to enable high-throughput screening systems as used in drug development and hand-held devices for point-of-care (PoC) biomedical diagnostic applications.

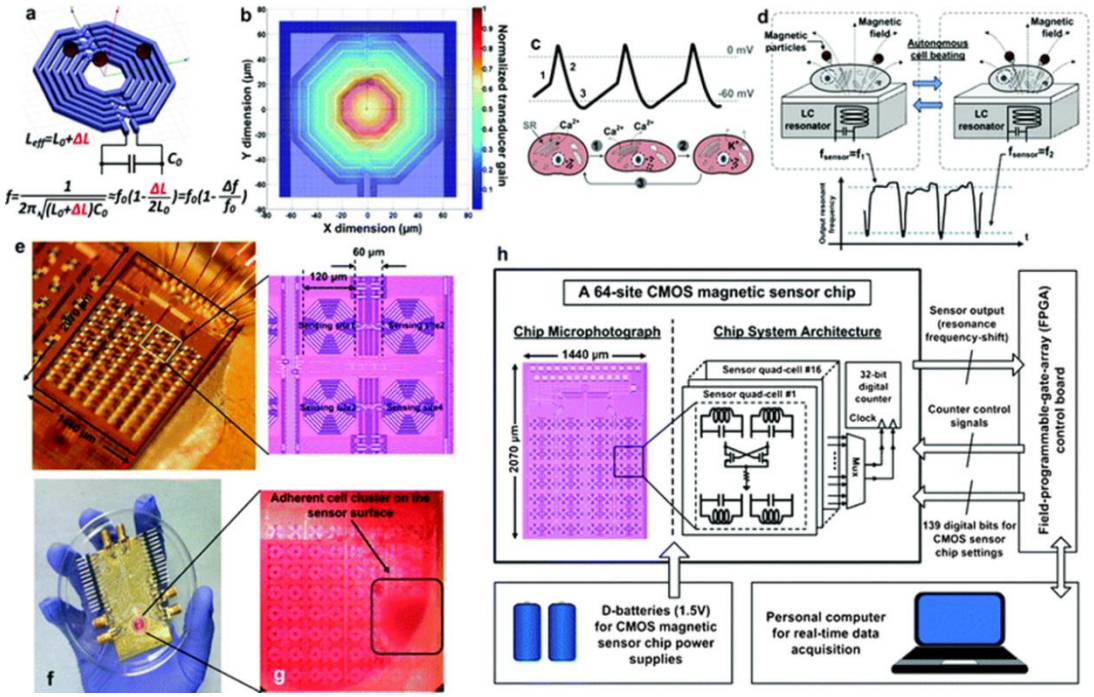


Figure AB. Magnetic cell-based sensor. (a) Magnetic particles result in an increased effective inductance L_{eff} in the LC resonator, which causes a downshift in the resonant frequency. This frequency shift can be detected by the CMOS circuits and serves as the readout for the sensing unit. (b) The normalized transducer gain (sensitivity to magnetic particles) of the CMOS magnetic sensor is plotted with respect to the sensor inductor geometry. (c) A spontaneous cardiac cell beating. Inset at top shows the cell potential changes (in milli-volts) during the periodic beating motion. SR = sarcoplasmic reticulum. (d) Autonomous beating of the cardiac progenitor cells leads to displacements of the magnetic particles and is detected by the CMOS magnetic sensor as periodic shifts in resonant frequency. (e) The sensor chip contains 64 (8 × 8) independent sensing sites as a sensor array. A zoom-in view shows the individual sensing site. (f) The sensor module fits in a petri dish. The module includes a PDMS reservoir to hold the cardiac progenitor cells and the medium. (g) An image of the PDMS sample reservoir shows a cell cluster on the CMOS sensor surface. (h) Diagram for the CMOS magnetic sensor chip microphotograph, chip system architecture, and the measurement setup.

This work was reported in:

Wang, H.; Mahdavi, A.; Tirrell, D. A.; Hajimiri, A. *Lab on a chip* **2012**, *12*, 4465.

APPENDIX C

Colony-Forming Cells in the Adult Mouse Pancreas are Expandable in Matrigel and Form Endocrine/Acinar Colonies in Laminin Hydrogel

Abstract

The study of hematopoietic colony-forming units using semisolid culture media has greatly advanced the knowledge of hematopoiesis. Here we report that similar methods can be used to study pancreatic colony-forming units. We have developed two pancreatic colony assays that enable quantitative and functional analyses of progenitor-like cells isolated from dissociated adult (2-4 mo old) murine pancreas. We find that a methylcellulose-based semisolid medium containing Matrigel allows growth of duct-like "Ring/Dense" colonies from a rare (~1%) population of total pancreatic single cells. With the addition of roof plate-specific spondin 1, a wingless-int agonist, Ring/Dense colony-forming cells can be expanded more than 100,000-fold when serially dissociated and replated in the presence of Matrigel. When cells grown in Matrigel are then transferred to a Matrigel-free semisolid medium with a unique laminin-based hydrogel, some cells grow and differentiate into another type of colony, which we name "Endocrine/Acinar." These Endocrine/Acinar colonies are comprised mostly of endocrine- and acinar-like cells, as ascertained by RNA expression analysis, immunohistochemistry, and electron microscopy. Most Endocrine/Acinar colonies contain beta-like cells that secrete insulin/C-peptide in response to D-glucose and theophylline. These results demonstrate robust self-renewal and differentiation of adult Ring/Dense colony-forming units in vitro and suggest an approach to producing beta-like cells for cell replacement of type 1 diabetes. The methods described, which include microfluidic expression analysis of single cells and colonies, should also advance study of pancreas development and pancreatic progenitor cells.

The artificial extracellular matrix proteins developed for this work are reported in Appendix G.

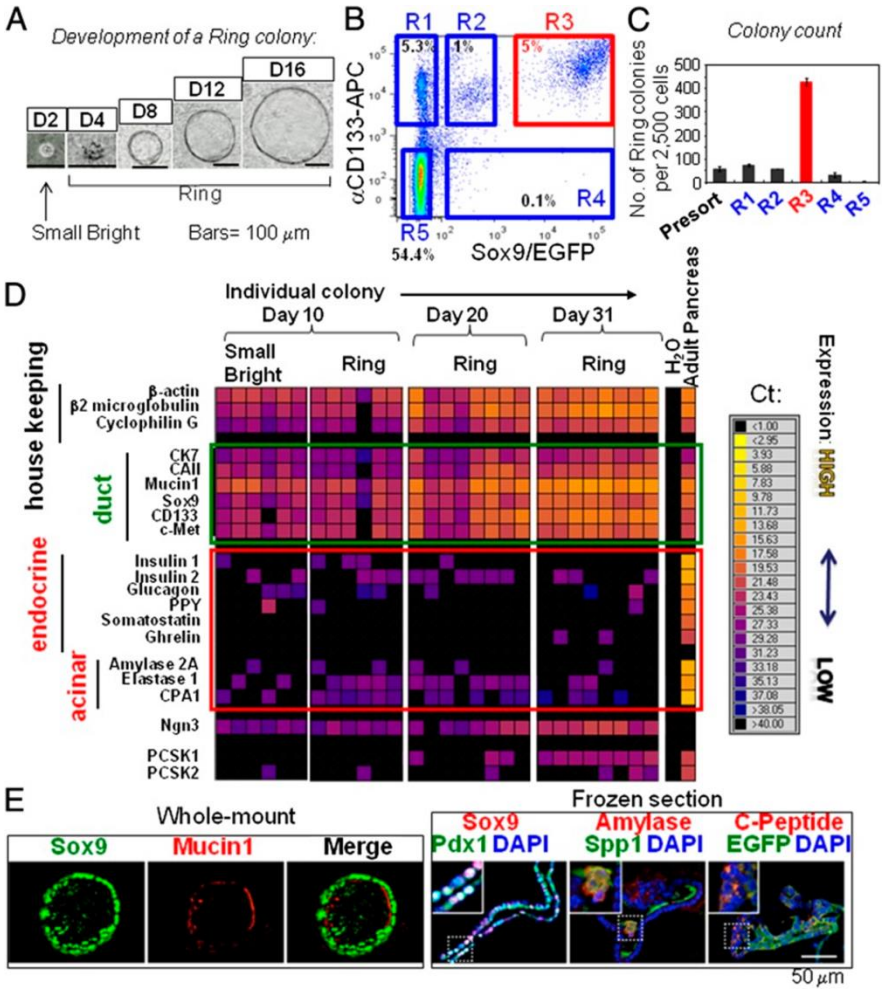


Figure AC. "Ring" colonies are formed in Matrigel-containing culture from CD133$^+$Sox9/EGFP$^+$ cells isolated from dissociated adult murine pancreata. (*A*) A Ring colony starts as a "Small Bright" colony and grows into a Ring colony. (*B*) Flow cytometry analysis of CD133 and Sox9/EGFP expression of total dissociated adult pancreata. Regions (R) drawn indicate sorting windows. (*C*) PCFUs–Ring are most enriched in the CD133$^+$Sox9/EGFP$^+$ R3 window. (*D*) Single-colony microfluidic qRT-PCR analysis demonstrates that Small Bright and Ring colonies express high levels of ductal markers and low but detectable levels of endocrine and acinar cell markers. Each column is from a single colony. (*E*) Whole-mount and frozen section immunostaining demonstrated protein expression of ductal (Sox9, Mucin1, and Spp1), acinar (Amylase), or endocrine (C-Peptide) markers.

This work was reported in:

Jin, L.; Feng, T.; Shih, H. P.; Zerda, R.; Luo, A.; Hsu, J.; Mahdavi, A.; Sander, M.; Tirrell, D. A.; Riggs, A. D.; Ku, H. T. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 3907.

APPENDIX D

Synthesis of Bioactive Protein Hydrogels by Genetically Encoded SpyTag-SpyCatcher Chemistry

Abstract

Protein-based hydrogels have emerged as promising alternatives to synthetic hydrogels for biomedical applications, owing to the precise control of structure and function enabled by protein engineering. Nevertheless, strategies for assembling 3D molecular networks that carry the biological information encoded in full-length proteins remain underdeveloped. Here we present a robust protein gelation strategy based on a pair of genetically encoded reactive partners, SpyTag and SpyCatcher, that spontaneously form covalent isopeptide linkages under physiological conditions. The resulting "network of Spies" may be designed to include cell-adhesion ligands, matrix metalloproteinase-1 cleavage sites, and full-length globular proteins [mCherry and leukemia inhibitory factor (LIF)]. The LIF network was used to encapsulate mouse embryonic stem cells; the encapsulated cells remained pluripotent in the absence of added LIF. These results illustrate a versatile strategy for the creation of information-rich biomaterials.

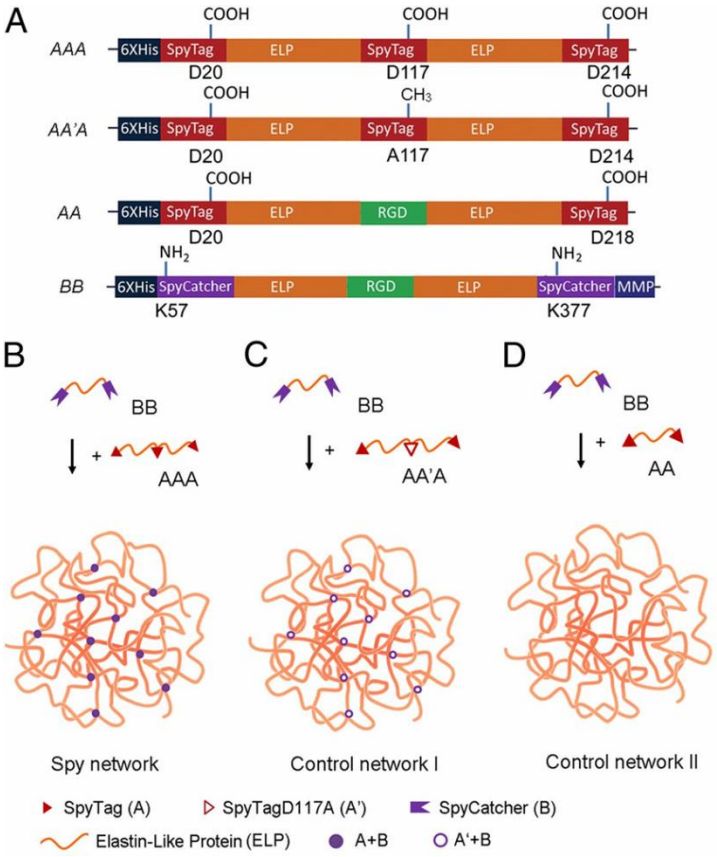


Figure AD. Spy network hydrogels. (*A*) Genetic constructs for the four protein precursors (AAA, AA′A, AA, and BB). The reactive groups are Asp-20, Asp-117, Asp-214, Lys-57, and Lys-337. AA′A is an Asp-117–to–Ala mutant of AAA, which enables only noncovalent molecular recognition between SpyTag and SpyCatcher at the A′ site. Construct BB contains an internal integrin-binding RGD sequence to facilitate cell adhesion and an MMP-1 cleavage site at the C-terminus to allow matrix remodeling by encapsulated cells. (*B–D*) Schematic illustration of the products formed by mixing protein precursors. AAA+BB leads to the formation of a covalently cross-linked gel, whereas AA′A+BB and AA+BB cannot form covalent molecular networks.

This work was reported in:

Sun, F.; Zhang, W. B.; Mahdavi, A.; Arnold, F. H.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 11269.

APPENDIX E

Colony-Forming Progenitor Cells in the Postnatal Mouse Liver and Pancreas Give Rise to Morphologically Distinct Insulin-Expressing Colonies in 3D Cultures

Abstract

In our previous studies, colony-forming progenitor cells isolated from murine embryonic stem cell-derived cultures were differentiated into morphologically distinct insulin-expressing colonies. These colonies were small and not light-reflective when observed by phase-contrast microscopy (therefore termed "Dark" colonies). A single progenitor cell capable of giving rise to a Dark colony was termed a Dark colony-forming unit (CFU-Dark). The goal of the current study was to test whether endogenous pancreas, and its developmentally related liver, harbored CFU-Dark. Here we show that dissociated single cells from liver and pancreas of one-week-old mice give rise to Dark colonies in methylcellulose-based semisolid culture media containing either Matrigel or laminin hydrogel (an artificial extracellular matrix protein). CFU-Dark comprise approximately 0.1% and 0.03% of the postnatal hepatic and pancreatic cells, respectively. Adult liver also contains CFU-Dark, but at a much lower frequency (~0.003%). Microfluidic qRT-PCR, immunostaining, and electron microscopy analyses of individually handpicked colonies reveal the expression of insulin in many, but not all, Dark colonies. Most pancreatic insulin-positive Dark colonies also express glucagon, whereas liver colonies do not. Liver CFU-Dark require Matrigel, but not laminin hydrogel, to become insulin-positive. In contrast, laminin hydrogel is sufficient to support the development of pancreatic Dark colonies that express insulin. Postnatal liver CFU-Dark display a cell surface marker $CD133^{+}CD49(flow)CD107b(low)$ phenotype, while pancreatic CFU-Dark are $CD133^{-}$. Together, these results demonstrate that specific progenitor cells in the postnatal liver and pancreas are capable of developing into insulin-expressing colonies, but they differ in frequency, marker expression, and matrix protein requirements for growth.

The artificial extracellular matrix proteins developed for this work are reported in Appendix G.

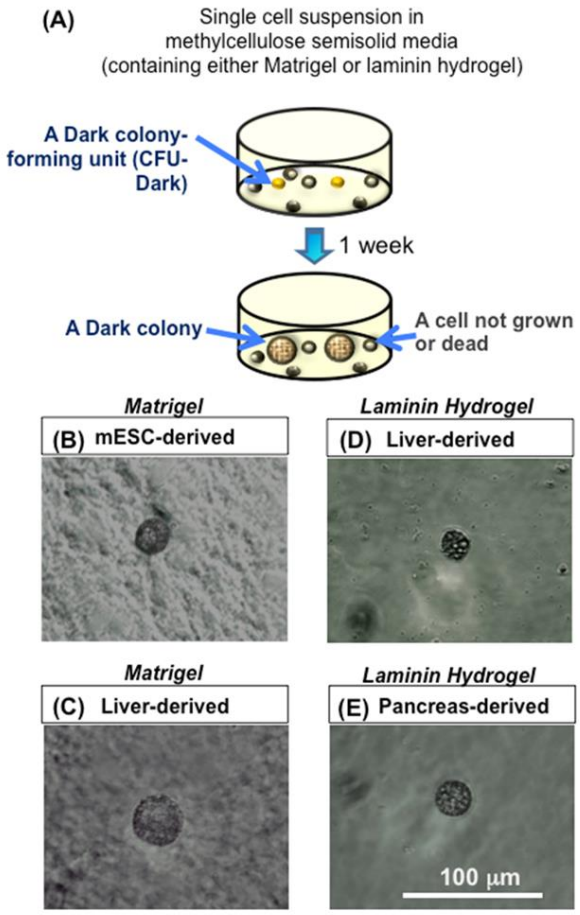


Figure AE. "Dark" colonies are formed from dissociated one-week-old murine liver or pancreas in our 3D colony assays. (A) Schematic of an *in vitro* methylcellulose-based colony assay for progenitor cells. Two different colony assays were used in this report; one contained Matrigel and the other contained laminin hydrogel. (B-E) Morphologically distinct colonies are identified when observed under a phase-contrast light microscope one week after plating. (B) A photomicrograph of a Dark colony developed from dissociated murine ES cell-derived day-16 cultures, confirming our prior finding. A Dark colony grown in Matrigel (C) or laminin hydrogel (D) from dissociated postnatal liver. (E) A Dark colony grown in laminin hydrogel from dissociated postnatal pancreas. mESC, mouse embryonic stem cells.

This work was reported in:

Jin, L.; Feng, T.; Chai, J.; Ghazalli, N.; Gao, D.; Zerda, R.; Li, Z.; Hsu, J.; Mahdavi, A.; Tirrell, D. A.; Riggs, A. D.; Ku, H. T. *The review of diabetic studies : RDS* **2014**, *11*, 35.

APPENDIX F

Cells with Surface Expression of $CD133^{high}CD71^{low}$ are Enriched for Colony-Forming Progenitors in the Adult Murine Pancreas

Abstract

Adult pancreatic stem and progenitor cells are potential sources of insulin-producing beta-like cells for cell replacement therapy of type 1 diabetes. In previous studies, we identified colony-forming progenitor cells in the murine adult pancreas that are capable of robust self-renewal and multi-lineage differentiation *in vitro*. However, one roadblock to effective study of these progenitors is that they are a minor population in the pancreas (up to 1% in 2-4 month-old mice). We report here that adult pancreatic progenitors express specific cell surface markers that allow enrichment by fluorescence-activated cell sorting. We find that $CD133^{high}CD71^{low}$ cells, but not other cell subpopulations in the adult pancreas, are the most enriched (up to 30-fold) for colony-forming progenitors. In semi-solid culture medium containing Matrigel, single $CD133^{high}CD71^{low}$ cells self-renewed and gave rise to duct-like cells. When cultured in laminin hydrogel, $CD133^{high}CD71^{low}$ cells gave rise to colonies composed of $insulin^{+}glucagon^{+}$ double-hormonal cells and acinar-like cells. Colonies with $insulin^{+}glucagon^{+}$ cells secreted C-peptide in response to D-glucose *in vitro*, and when placed under the renal capsules of diabetic mice for five weeks they gave rise to insulin or glucagon single-positive cells. Both $CD133^{high}CD71^{low}$ and $CD133^{+}CD71^{-}$ cells expressed ductal genes yet only $CD133^{high}CD71^{low}$ cells contained progenitors, suggesting functional heterogeneity of adult ducts. We conclude that adult pancreatic progenitors are distinguishable by cell surface markers and can self-renew and differentiate into multiple lineages of cells. This study establishes that adult pancreatic progenitors can be purified using live-cell sorting, which will enable in-depth understanding of these cells.

The artificial extracellular matrix proteins developed for this work are reported in Appendix G.

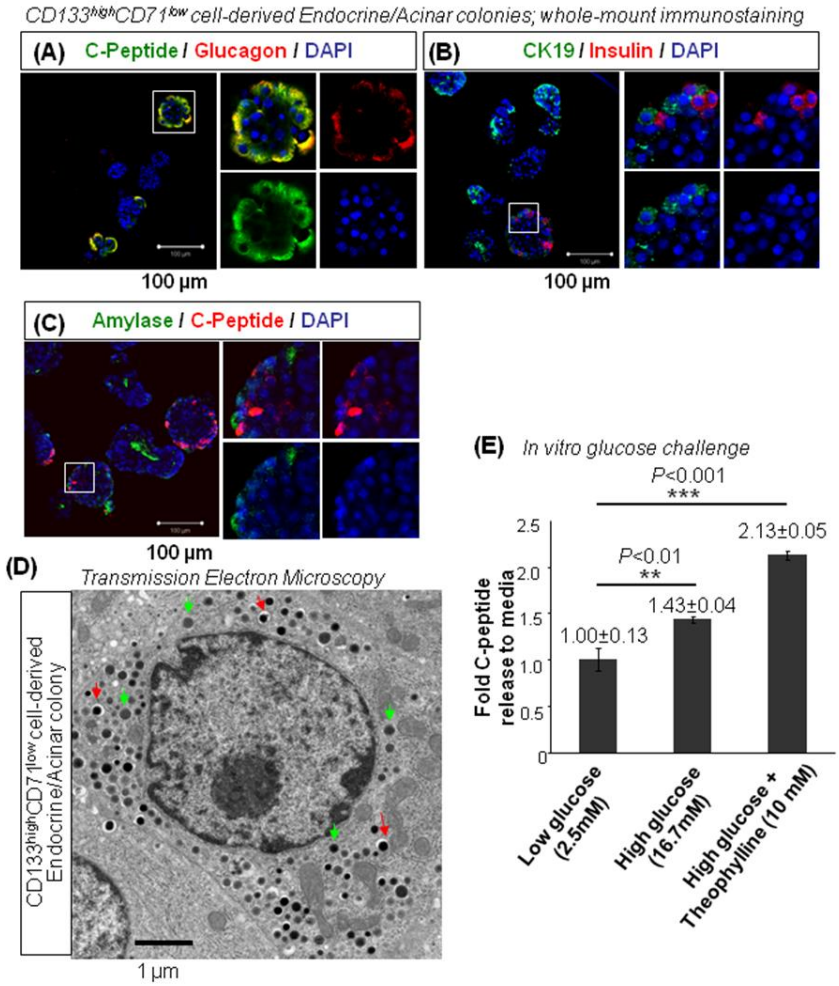


Figure AF. Some cells in Endocrine/Acinar colonies derived from the CD133^{high}CD71^{low} population are insulin^{+}glucagon^{+} and respond to D-glucose *in vitro*. Whole-mount immunostaining of Endocrine/Acinar colonies showed C-peptide^{+} or Insulin^{+} cells co-expressed Glucagon (A) but not CK19 (B) or Amylase (C). C-peptide is a surrogate marker for *de novo* synthesized insulin. (D) Transmission electron microscopy of a cell in an Endocrine/Acinar colony that contained two types of granules. Red and green arrows indicate insulin-like and non-insulin granules, respectively. (E) Endocrine/Acinar colonies were pooled and subjected to sequential *in vitro* incubation with designated levels of D-glucose and/or theophylline, a cAMP potentiator.

This work was submitted as follows:

Jin, L.; Gao, D.; Feng, T.; Rawson, J.; Luo, A.; Chai, J.; Wedeken, L.; Ghazalli, N.; Hsu, J.; LeBon, J.; Walker, S.; Tremblay, J.R.; Mahdavi, A.; Tirrell, D. A.; Riggs, A. D.; Ku, H. T. *Stem Cells*, submitted.

APPENDIX G

Artificial Extracellular Matrix Proteins with Laminin-Derived Sequences Enhance Survival and Differentiation of Pancreatic Progenitor Cells in Culture

Abstract

Information-rich and well-defined materials that recapitulate the essential functions of the extracellular matrix (ECM) are of interest as substrates for studies of cellular behavior and as scaffolds for use in regenerative medicine. For example, matrix-derived peptide sequences can be used to enhance survival of and insulin release from pancreatic beta cells encapsulated in poly (ethylene glycol) (PEG) hydrogels. The application of such strategies to progenitor populations has not been explored. Here we report the use of artificial ECM (aECM) proteins based on elastin and laminin to enhance survival and differentiation of beta cell progenitors in culture.

This manuscript is included because it provides information about the protein-based materials used in appendices C, E and F.

Well-defined materials that recapitulate the essential functions of the extracellular matrix (ECM) are of interest as substrates for studies of cellular behavior and as scaffolds for use in regenerative medicine.[1-3] For example, Anseth and coworkers have shown that matrix-derived peptide sequences can be used to enhance survival of and insulin release from pancreatic beta cells encapsulated in poly(ethylene glycol) (PEG) hydrogels.[4] To our knowledge, the application of such strategies to progenitor populations has not been explored. Here we report the use of artificial ECM (aECM) proteins based on elastin and laminin to enhance survival and differentiation of beta cell progenitors in culture.

A variety of extracellular matrix proteins, including laminin, collagen type IV, fibronectin, heparin sulfate proteoglycans and nidogen/entactin, are present in the microenvironments of pancreatic cells.[5] In the endocrine pancreas, insulin-expressing beta cells use VEGF-A to recruit endothelial cells, which form capillaries with a vascular basement membrane adjacent to the beta cells.[6] Among the vascular basement membrane proteins, laminins were found to induce insulin gene expression and increase beta cell proliferation both in a MIN6 insulinoma cell line and in VEGF-A$^{-/-}$ islets.[6] This effect was dependent on β_1-integrin receptor binding as determined by a blocking antibody assay.[6] Both the survival and glucose-stimulated insulin secretion of isolated adult rat islets may be enhanced by culturing the cells on laminin.[7,8]

These reports prompted us to explore the response of pancreatic cells to elastin-based aECM proteins that include three copies of the laminin-derived IKVAV sequence (designated 19mer) (Figure G1a). The elastin-based sequences dominate the physical properties of the protein and facilitate its purification by thermal cycling through the lower critical solution temperature (LCST). The 19 amino acid residues that include the IKVAV sequence were obtained from the α-chain of laminin. To investigate the specificity of the cell response to the IKVAV sequence, we designed two additional proteins (Figure G1a). The first includes a scrambled version of the IKVAV sequence (designated 19scr). As a control for effects of differential cell adhesion, we designed a second protein that contains a fibronectin-derived RGD peptide sequence (designated RGD). The aECM proteins were expressed in *E. coli* strain BL21 (DE3) pLysS by using an inducible pET28 expression vector (Supplementary Figure GS.1 and GS.2). After three rounds of thermal cycling, we obtained each aECM protein in pure form, as shown by gel electrophoresis (Figure G1c). The molecular weight of each protein was verified by MALDI-TOF mass spectrometry (Figure G1d). The LCST was approximately 27°C in phosphate buffered saline at a protein concentration of 10 mg/ml (Figure G1e).

To test the effects of aECM proteins, we chose to study postnatal pancreas—a rich source of progenitor cells[9,10] in addition to differentiated cells. Dissociated cells were harvested from pancreas and plated in 3-dimensional cultures containing semi-solid media that permit the survival, proliferation and/or differentiation of single cells.[11] We compared the effects of these materials to those of Matrigel™ which is a commercial source of crude ECM components.[12-15]

Cells were grown in the presence of either Matrigel (5%), 19mer or 19scr (100 μg/ml each) proteins for 7 days, and gene expression of various marker genes was analyzed by quantitative (Q) RT-PCR. The level of gene expression was first analyzed relative to an internal control, the housekeeping gene β-actin. Subsequently, gene expression levels from cells supported by aECM proteins were further normalized to those supported by Matrigel. Compared to Matrigel, both the 19mer and 19scr proteins enhanced expression of endocrine (glucagon$^+$ alpha cells and insulin$^+$ beta cells), neuronal (β3-tubulin), endothelial (CD31) and mesenchymal (vimentin) cell markers (Figure G2a and b). Compared to 19scr, the 19mer protein enhanced the levels of insulins 1 and 2 (Figure G1a), suggesting a specific effect of the IKVAV sequence to support the survival of single beta cells or the differentiation of their progenitors.

We next investigated the dose-dependent response of dissociated pancreatic cells to the 19mer protein. There was a positive correlation between the concentration of 19mer protein and expression of amylase 2A (an acinar cell marker), CD31, glucagon, insulins 1 and 2, and vimentin (Figure G2c). In contrast, the level of ductal cell marker, CK7, was lower in cells supported by the 19mer protein compared to Matrigel. Taken together, these findings indicate that in short-term cultures and in contrast to Matrigel, the 19mer protein enhances survival or differentiation away from ductal cells, and that the IKVAV sequence specifically increases insulin expression.

We next investigated the long-term effect of the 19mer protein by using 25 day cultures. QRT-PCR analysis showed increased levels of insulins 1 and 2 and glucagon above those characteristic of freshly isolated adult pancreatic cells (Figure G2d). Consistent with the short-term culture results and compared to Matrigel, the ductal cell marker CK7 was down-regulated, The expression of the delta cell marker somatostatin (SST) was retained and the exocrine marker amylase 2a was minimally up-regulated by the 19mer protein (Figure G2e). Taken together, these results indicate that, in long-term culture, the 19mer protein preferentially supports the survival of alpha and beta cells and to lesser extent acinar cells but in contrast to Matrigel, the 19mer protein does not support ductal cells.

To investigate the effects of aECM proteins specifically on progenitor cells, we analyzed single cell-derived colonies, whose lineage composition and number would indicate their lineage potential and survival, respectively. Single cells harvested from postnatal pancreas were plated in our clonogenic assay in the presence of various aECM proteins. After 7 days in culture, we observed light-reflective colonies with diameters ranging from 20 to 60 μm (Figure G3a). In the control culture containing Matrigel, cystic colonies typical of ductal cell morphology[12-15] were observed. Such ductal colonies have been extensively characterized in prior studies, including ours; they are formed in cultures initiated with purified adult human [12,13] or mouse [14,15] ductal cells in the presence of Matrigel.

To analyze lineage composition, individual 8-day-old colonies that showed morphologies consistent with those in Figure G3a were handpicked and subjected to microfluidic QRT-PCR analysis (Figure G3b). All values were normalized to β-actin and compared to those obtained from freshly isolated postnatal pancreas cells. We found that among the single colonies supported by the 19mer protein, 2 out of 6 expressed insulin 1 or 2, and none expressed ductal (CK7, mucin1, and HNF1b) or acinar (amylase 2a) cell markers. In contrast, 19scr- and RGD-supported colonies did not express insulins 1 or 2, but 4 out of 6 colonies from each group expressed at least one of the three ductal markers. These results demonstrate that the 19mer protein specifically supports differentiation of beta cell progenitors. Alternatively, the 19mer protein may divert lineage commitment to beta cells from a bi-potential progenitor that could otherwise differentiate toward both ductal and beta cell lineages. We did not detect expression of glucagon in any of the 6 colonies supported by the 19mer protein, which may suggest a lower prevalence of alpha- than beta-cell progenitors in postnatal pancreas.

To examine protein expression and to ensure glucagon-expressing colonies could be observed, we used whole-mount immunostaining and imaged colonies by confocal fluorescence microscopy. C-peptide was used as a surrogate marker for *de novo* synthesized insulin to distinguish it from exogenous recombinant insulin,[16] which was present in our culture media. We found colonies positively stained for either glucagon or C-peptide (Figure G3c). The staining was specific; incubation with isotype control antibody did not yield a positive signal. The staining patterns of glucagon or C-peptide were cytoplasmic, consistent with the expected locations of these proteins as secretory hormones. Together, these results indicate that the 19mer protein supports differentiation of both alpha and beta cell progenitors. Comparison of colony-forming frequency among cultures containing various aECM proteins revealed that the 19mer protein best supports the overall survival of progenitor cells (Figure G3d).

In summary, we have developed an elastin-based aECM protein bearing a laminin-derived IKVAV sequence that supports survival and differentiation of beta cell progenitors in culture. The observed effects may be due to better survival or selection for beta versus ductal cell lineages. Appropriately designed aECM proteins may eventually enable expansion and differentiation of stem and progenitor cells to be used in regenerative medicine.

Materials and Methods

Expression of aECM proteins

Cloning was performed in *E. coli* strain DH10B (Invitrogen, Carlsbad, CA). Bacteria transformation was accomplished with Z-competent *E. coli* transformation kits (Zymo Research, Irvine, CA). Restriction enzymes were obtained from New England Biolabs (Ipswich, MA). Ligations were performed with T4 DNA ligase (Roche, Indianapolis, IN). Plasmids were isolated with QIAprep kits (Qiagen, Valenica, CA). Final plasmid sequences were verified by restriction digestion and sequencing (Laragen, Culver City, CA). DNA sequences encoding various aECM proteins were ligated into pET28 (Novagen, Madison, WI) and induced under control of a bacteriophage T7 promoter. Expressions were performed in *E. coli* strain BL21 (DE3) pLysS (Novagen). Overnight cultures were used to inoculate Terrific Broth (TB) medium supplemented with 50 μg/ml kanamycin and 100 μg/ml chloramphenicol (Sigma, St. Louis, MO). Cells were grown to optical density at 600 nm (OD_{600}) of 5 in a 10 L BioFlow 3000 fermenter (New Brunswick Scientific, Edison, NJ) with oxygen and pH control as previously described[17,18]. Induction was initiated at OD_{600} = 5 by addition of 2.5 mM isopropyl-1-β-D-thiogalactopyranoside (IPTG). After 3h the OD_{600} was 10-15.

Purification of aECM proteins

Bacteria cells were pelleted by centrifugation (10,000 g, 15 min, 4°C) and lysed in TEN buffer (10 mM Tris-HCL, 1 mM EDTA, 0.1 M NaCl, pH 8) supplemented with 50 μg/ml phenylmethylsulfonyl flouride (Sigma) and 10 μg/ml each of ribonuclease A and deoxyribonuclease 1 (Sigma). Cells were lysed by a freeze-thaw cycle followed by sonication at 4°C. Lysates were centrifuged (35,000 g, 2 h, 4°C), and the soluble fraction was adjusted to pH 9 in 1 M NaCl. Each thermal cycle consisted of a temperature shift to 37°C, followed by centrifugation (35,000g, 2 h), resolubilization of the pellet in pH 9 water at 4°C, and centrifugation (35,000g, 2 h). After three thermal cycles, aECM proteins were dialyzed in water at 4°C and the product lyophilized. A solution of 0.25 mg/ml of protein was subjected to electrophoresis on a Novex 12% Bis-Tris polyacrylamide gel (Invitrogen), labeled with colloidal blue dye (Invitrogen) and imaged on a Typhoon 9400 molecular imager (GE Healthcare).

MALDI-TOF mass spectrometry

Matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) was performed on a Voyager DE-Pro MALDI TOF-MS (Applied Biosystems, Beverly, MA). Protein solutions at 30

mg/ml were added to matrix solution consisting of 10 mg/ml sinapinic acid in 0.07% trifluoroacetic acid and 30% acetonitrile. A matrix to protein ratio of 20:1 was used.

LCST measurement

Lower critical solution temperature measurements were performed at a protein concentration of 10 mg/ml in phosphate buffered saline (pH 7.4). Absorbance at 300 nm was measured on a DU7400 diode array UV-visible spectrophotomer (Beckman Coulter, Indianapolis, IN).

Pancreas dissociation

Pancreata were dissected from 8-day-old C57BL/6 mice, and dissociated with 4 mg/ml collagenase B (Roche, Indianapolis, IN) and 1000 U/ml deoxyribonuclease I (Calbiochem, Gibbstown, NJ) at 37°C for 20 min. Cells were washed twice with Dulbecco's PBS (Mediatech, Manassas, VA) containing 0.1% bovine serum albumin (Sigma) and filtered through a 40 μm mesh to yield single cell suspensions. All mice were maintained under specific pathogen-free conditions, and animal experiments were conducted according to the Institutional Animal Care and Use Committee at the City of Hope.

Mouse cell culture

The culture assay was reported previously[11]. Briefly, cells were dissociated as described above and resuspended in cold culture medium containing DMEM/F12 (Mediatech), 1% 1500 centipoise methylcellulose (Sinetsu Chemical, Tokyo, Japan), 50% conditioned medium from pancreatic-like cells differentiated from embryonic stem cells, 5% fetal bovine serum (Tissue Culture Biologicals, Seal Beach, CA), 10 mM nicotinamide (Sigma), 0.1 nM exendin-4 and 10 ng/ml human recombinant activin-β (R&D Systems, Minneapolis, MN). Unless specified otherwise, aECM proteins were added at 100 μg/ml. Matrigel (BD Biosciences, Bedford, MA) was added at 5% V/V. Cells were placed in 24-well Ultra-Low attachment plates (Corning Incorporated, Corning, NY) at $1x10^4$ cells/well and incubated at 37°C with 5% CO_2 air. Quadruplicate wells were routinely plated and analyzed.

RNA isolation, RT-PCR and quantitative PCR analysis

Total RNA was extracted by using RNeasy Micro Kits (Qiagen). Reverse transcription was performed using QuantiTect Reverse Transcription Kits (Qiagen). The cDNAs of interest were amplified with Taqman probes (Applied Biosystems) and reaction buffer PerfeCTa FastMix (Quanta Biosciences, Gaithersburg, MD) on an ABI 7900HT Fast Real-time PCR System

(Applied Biosystems). All samples were tested in duplicate and β-actin was used as an internal control to calculate relative (ΔC_T) gene expression among samples in the same PCR run. Values in Figure G2 were further normalized to that obtained from freshly isolated postnatal pancreatic cells (panel d) or from Matrigel-treated cells (rest of panels).

Microfluidic QRT-PCR analysis on single handpicked colonies

Individual colonies were handpicked and analyzed using the BioMark™ 48.48 Dynamic Array System (Fluidigm, South San Francisco, CA) according to the manufacturer's protocol. Threshold cycle (C_T) was determined from fluorescence intensity by the BioMark PCR Analysis software. ΔC_T was calculated relative to β-actin, similar to the QRT-PCR analysis.

Colony counts

Colonies were scored visually on an inverted optical microscope as described previously.

Whole-mount immunofluorescent staining and confocal microscopy

Colonies were handpicked, pooled, placed in round-bottom 96-well plates, and fixed with 4% paraformaldehyde (Pierce, Rockford, IL) in PBS for 30 min. After washing twice with PBS, cells were permeabilized with 0.1% Triton-X100 (Pierce) for 10 min and then washed again with PBS. Non-specific binding sites were first blocked by incubation overnight with Protein Block Serum-Free solution (Dako, Carpinteria, CA) at 4°C, followed by staining with primary antibody (rabbit anti-human Glucagon or C-peptide serum [Dako]) at 1:75 dilution in blocking solution for 24 h at 4°C. Rabbit IgG was used as an isotype control. After washing, Cy5-conjugated goat anti-rabbit antibody (Jackson Immunology Research Labs, West Grove, PA) (1:100) was added and incubated at room temperature for 4 h. After washing, colonies were resuspended in 50 μg/ml DPPI in PBS for 30 min prior to imaging on a glass-bottom petri dish (MatTek, Ashland, MA). Confocal imaging was performed on a LSM510 Meta NLO microscope (Zeiss) with 2-photon capability. DAPI was excited at 700 nm for 2-photon excitation.

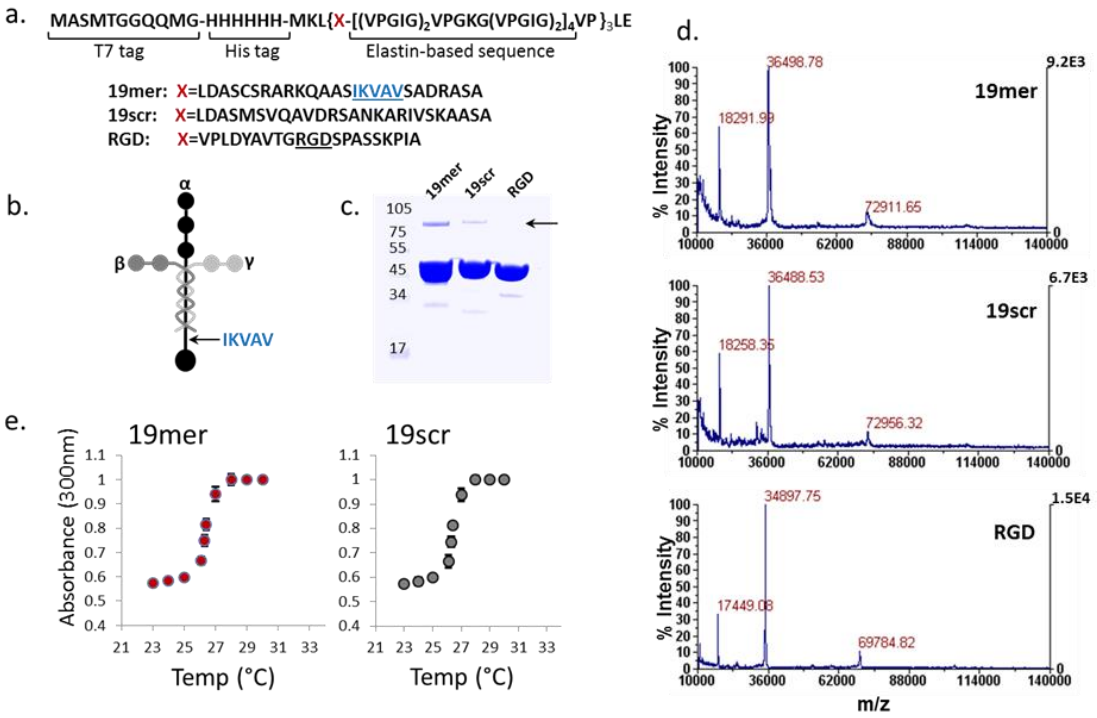


Figure G1. Artificial extracellular matrix (aECM) proteins. (A) Amino acid sequences of the aECM proteins generated in this report. (B) Schematic representation of laminin protein comprised of three chains (αβγ), showing the location of the IKVAV sequence. (C) Polyacrylamide gel stained with Colloidal Blue demonstrates the purity of aECM proteins. Arrow indicates protein dimers. (D) MALDI-TOF mass spectra of aECM proteins. Signal at approximately 18 kDa is assigned to the doubly charged protein and the peak at approximately 73 kDa is assigned to the protein dimer. (E) Absorbance at 300 nm was measured as a function of temperature to determine the transition temperature of the aECM proteins.

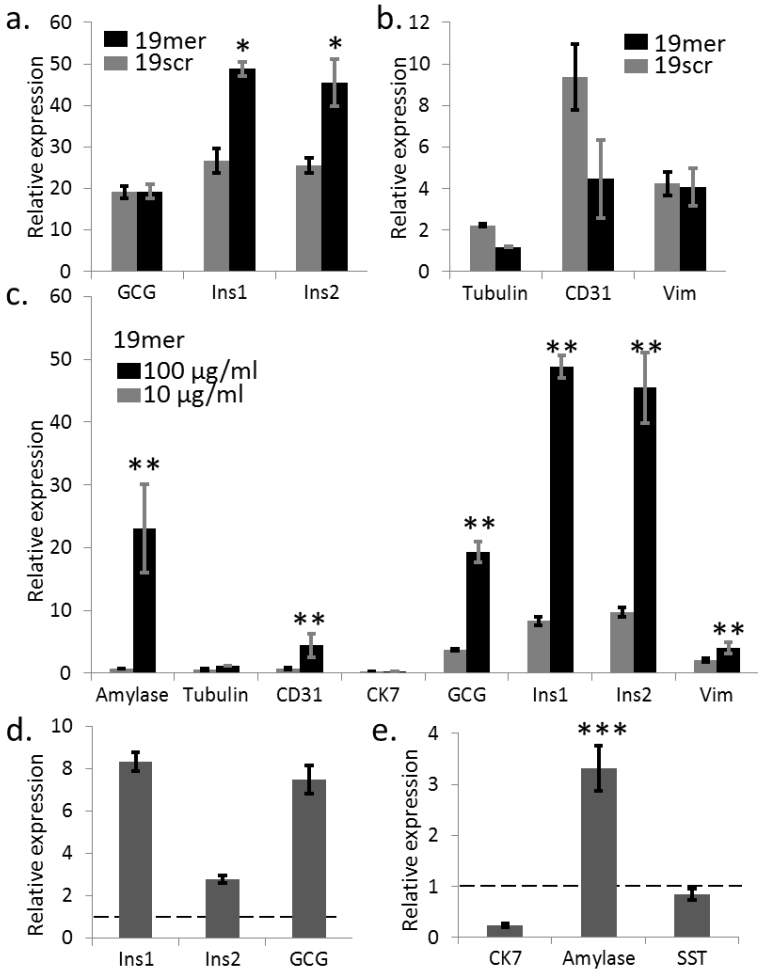


Figure G2. Opposing effects of aECM proteins versus Matrigel in pancreatic cell culture. Artificial ECM protein with laminin-derived sequences (19mer) preferentially supported the survival of pancreatic endocrine and acinar cells, but not ductal cells. Single cells procured from postnatal pancreas were plated into semi-solid media containing Matrigel, 19mer or 19scr. Seven (A-C) or 25 (D-E) days post-plating, total cells were collected and gene expression analyzed by QRT-PCR. Gene expression levels were normalized to those supported by Matrigel (A-C and E) or to freshly isolated day-8 pancreatic cells (D). Cells in panels d and e were grown in the presence of 100 µg/ml 19mer. Error bars represent standard deviation. *This number is different from 19scr at $P<0.05$. **This number is different from 10 µg/ml 19mer at $P<0.05$. ***This number is different from Matrigel control at $P<0.05$.

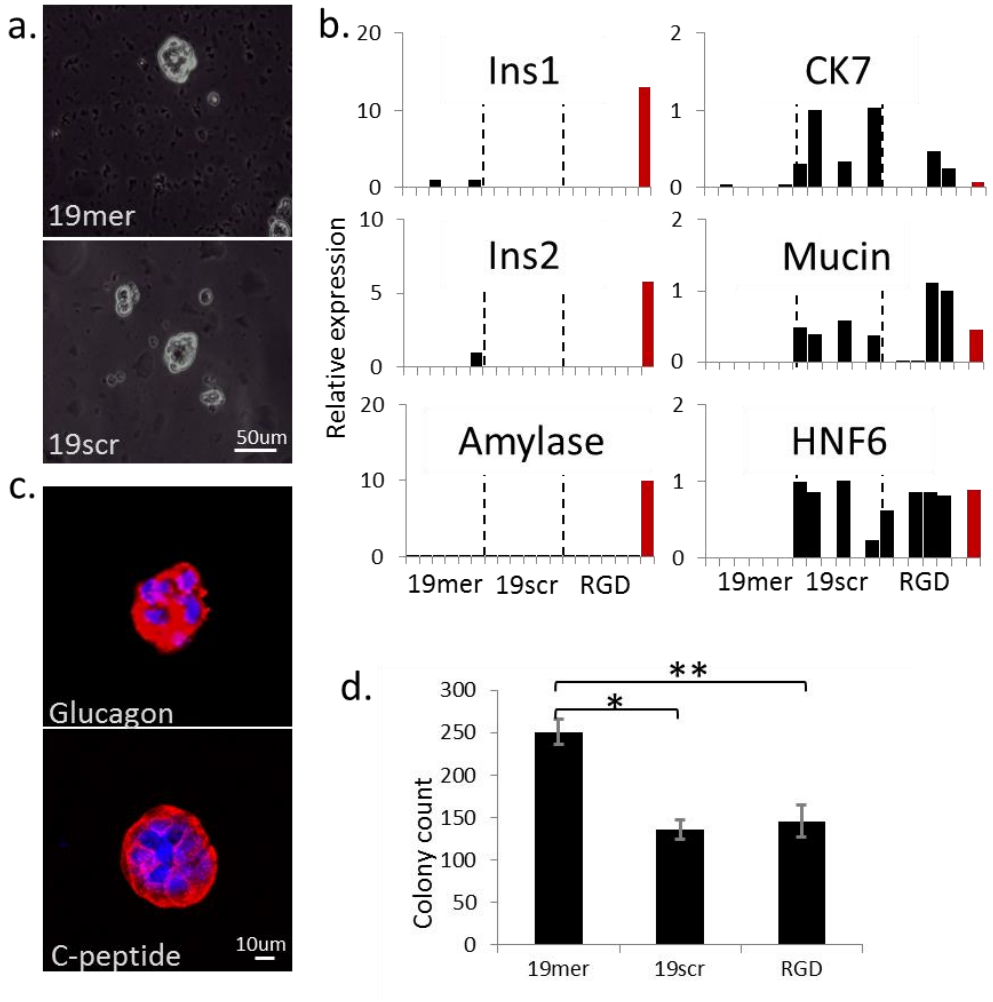


Figure G3. Pancreatic progenitor cells survive and differentiate in aECM proteins. Differentiation of beta-cell progenitor cells is best supported by aECM protein with laminin-derived sequences (19mer). (A) Photomicrographs of single cell-derived colonies (7-day-old) grown in the presence of designated aECM proteins. (B) Microfluidic QRT-PCR analysis of individually handpicked 8-day-old colonies grown on designated aECM proteins. Expression is relative to the internal control β-actin. Each black bar represents expression level of 1 colony (n=6 from each group). Red bars represent the expression levels from positive control postnatal pancreatic cells. (C) Whole-mount immunofluorescent staining of handpicked 8-day-old colonies grown in 19mer. Nuclei are stained with DAPI in blue. (D) Survival of pancreatic progenitor cells is enhanced by 19mer, as indicated by the higher number of single cell-derived colonies. Data represent the mean and standard deviation of the number of 8-day-old colonies in quadruplicate wells.*$P=1.27x10^{-4}$; **$P=2.31x10^{-5}$.

Figure GS.1.

Scheme of an inducible expression system used for the production of aECM proteins in bacteria. The *E-coli* expression strain, BL21 (DE3) pLysS, contains a gene for bacteriaphage T7 polymerase under the control of the UV5p promoter. In the absence of the inducer IPTG, the endogenous T7 polymerase is inhibited by the T7 lysozyme expressed by an episomal plasmid, pLysS, in the bacteria. After addition, IPTG binds to UV5p promoter and activates production of T7 polymerase, which subsequently triggers the expression of aECM gene, such as the 19mer shown, located in a pET28 plasmid.

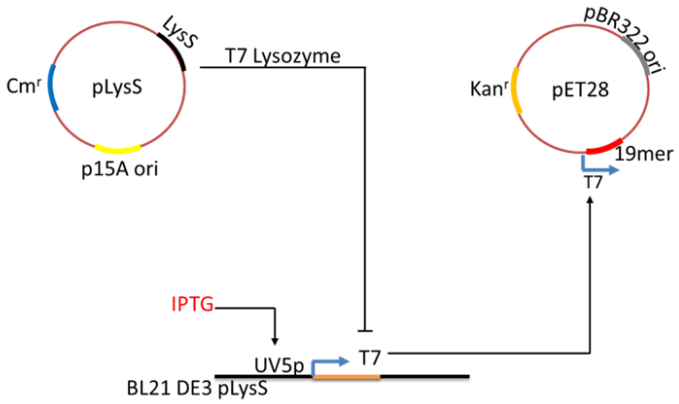
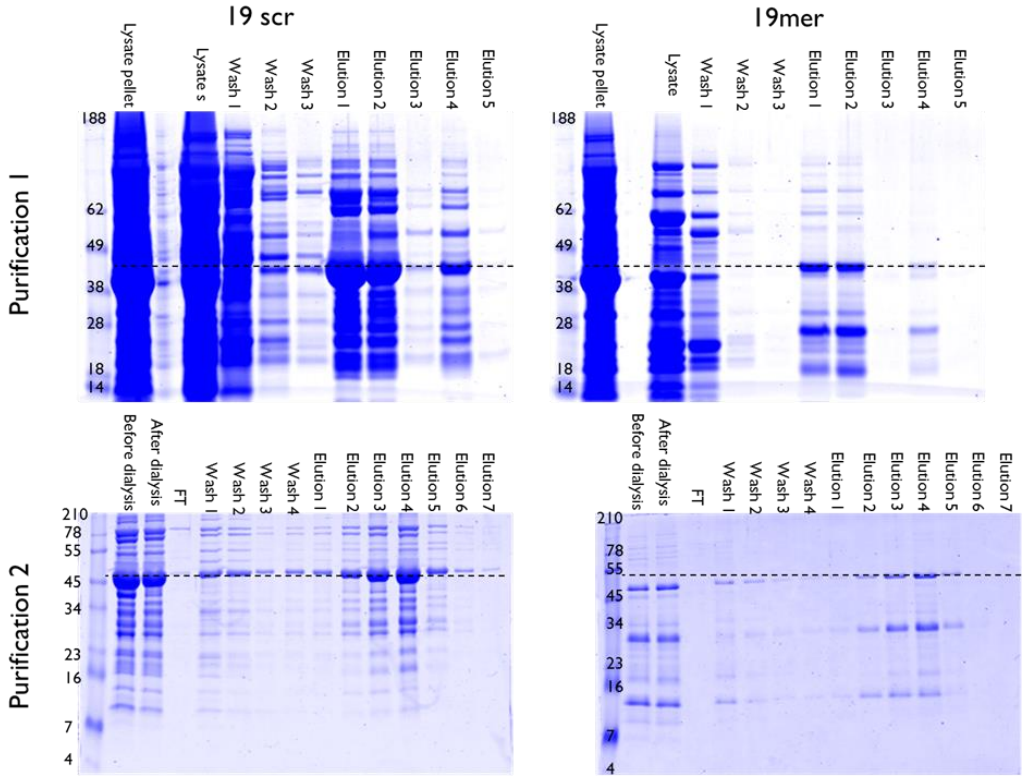


Figure GS.2.

Efficient purification of aECM proteins using cold-hot temperature cycles. In purification method 1 (upper panel), the nickel-nitrilotriacetic acid (Ni-NTA) resin affinity chromatography was used to enrich 19scr and 19mer proteins, because of their N-terminal polyhistidine-tag that can bind to the Ni-NTA resin. In purification method 2 (lower panel), temperature cycling based on the lower critical solution temperature (LCST) of the 19scr and 19mer proteins was also effective in purifying these proteins. Dotted lines show the position of the 19scr and 19mer proteins on polyacryamide gels stained with Coomassie Blue. Since temperature cycling is more cost-effective, we routinely purify aECM proteins based on this method.



References

(1) Alberti, K.; Davey, R. E.; Onishi, K.; George, S.; Salchert, K.; Seib, F. P.; Bornhauser, M.; Pompe, T.; Nagy, A.; Werner, C.; Zandstra, P. W. *Nat Methods* **2008**, *5*, 645.

(2) Khademhosseini, A.; Langer, R.; Borenstein, J.; Vacanti, J. P. *Proc Natl Acad Sci U S A* **2006**, *103*, 2480.

(3) Wheeldon, I.; Farhadi, A.; Bick, A. G.; Jabbari, E.; Khademhosseini, A. *Nanotechnology* **2011**, *22*, 212001.

(4) Weber, L. M.; Hayda, K. N.; Haskins, K.; Anseth, K. S. *Biomaterials* **2007**, *28*, 3004.

(5) Kragl, M.; Lammert, E. *Adv Exp Med Biol* **2010**, *654*, 217.

(6) Nikolova, G.; Jabs, N.; Konstantinova, I.; Domogatskaya, A.; Tryggvason, K.; Sorokin, L.; Fassler, R.; Gu, G.; Gerber, H. P.; Ferrara, N.; Melton, D. A.; Lammert, E. *Dev Cell* **2006**, *10*, 397.

(7) Pinkse, G. G.; Bouwman, W. P.; Jiawan-Lalai, R.; Terpstra, O. T.; Bruijn, J. A.; de Heer, E. *Diabetes* **2006**, *55*, 312.

(8) Parnaud, G.; Hammar, E.; Rouiller, D. G.; Armanet, M.; Halban, P. A.; Bosco, D. *Diabetes* **2006**, *55*, 1413.

(9) Bonner-Weir, S. *Endocrinology* **2000**, *141*, 1926.

(10) Herbach, N.; Bergmayr, M.; Goke, B.; Wolf, E.; Wanke, R. *PLoS One* **2011**, *6*, e22814.

(11) Winkler, M.; Trieu, N.; Feng, T.; Jin, L.; Walker, S.; Singh, L.; Ku, H. T. *J Vis Exp* **2011**, e3148.

(12) Bonner-Weir, S.; Taneja, M.; Weir, G. C.; Tatarkiewicz, K.; Song, K. H.; Sharma, A.; O'Neil, J. J. *Proc Natl Acad Sci U S A* **2000**, *97*, 7999.

(13) Gao, R.; Ustinov, J.; Pulkkinen, M. A.; Lundin, K.; Korsgren, O.; Otonkoski, T. *Diabetes* **2003**, *52*, 2007.

(14) Schreiber, F. S.; Deramaudt, T. B.; Brunner, T. B.; Boretti, M. I.; Gooch, K. J.; Stoffers, D. A.; Bernhard, E. J.; Rustgi, A. K. *Gastroenterology* **2004**, *127*, 250.

(15) Jin, L.; Feng, T.; Shih, H. P.; Dubois, C. L.; Winkler, M.; Zerda, R.; Schug, J.; Trieu, N.; Al-Abdulla, I.; Kandeel, F.; Kaestner, K.; Sander, M.; Ku, H. T. *Manuscript Submitted* **2012**.

(16) Rajagopal, J.; Anderson, W. J.; Kume, S.; Martinez, O. I.; Melton, D. A. *Science* **2003**, *299*, 363.

(17) Nowatzki, P. J.; Tirrell, D. A. *Biomaterials* **2004**, *25*, 1261.

(18) Heilshorn, S. C.; DiZio, K. A.; Welsh, E. R.; Tirrell, D. A. *Biomaterials* **2003**, *24*, 4245.