Supplemental Information

Tagging and Enriching Proteins Enables Cell-Specific Proteomics

Thomas S. Elliott, Ambra Bianco, Fiona M. Townsley, Stephen D. Fried, and Jason W. Chin
I. Supplemental Figures

Figure S1. Intact protein spectra prior to deconvolution

Figure S2. MS2 spectra of tryptic peptides

Figure S3. Fluorescence imaging of SORT-M

Figure S4. Quantification of SORT-E protein

Figure S5. Dilution of SORT-E input for enrichment threshold

Figure S6. SORT-E LFQ ratios and errors

Figure S7. SORT-E efficiently extracts low abundance proteins

Figure S8. Quantitation of SORT-E for TMT

Figure S9. SORT-E and codon abundance model

Figure S10. SORT-E and SASA model

Figure S11. SORT-M in *D. melanogaster* with different PylT variants

Figure S12. SORT-M and SORT-E in *D. melanogaster* comparison with *E. Coli*

Figure S13. SORT-E in *D. melanogaster* protein quantification

Table S1. Protein identities and LFQ of SORT-E from *E. coli*

Table S2. Protein identities and quantification of SORT-E via 9-Plex TMT

Table S3. Protein identities and LFQ of SORT-E from *D. Melanogaster*
Figure S1 (related to Figure 3). Raw mass spectra for intact T4 lysozyme (K-83-i)-His6 (top row), ubiquitin (K-6-i)-His6 (middle row), and ubiquitin (K-48-i)-His6 (bottom row), as pure proteins (i), TDB conjugates (ii), and cleaved dipyridiyl-diaza cycloaddition products (iii). Spectra correspond to the deconvolved masses shown in Figure 3.
Figure S2 (related to Figure 3). MS2 spectra of tryptic peptides derived from species iii of T4 lysozyme (K-83-1)-His6 (top), ubiquitin (K-6-1)-His6 (middle), and ubiquitin (K-48-1)-His6 (bottom). The spectra confirm the presence of the dipyridiyl-diaza cycloaddition product at the correct site in each of the proteins. In the case of T4 Lysozyme, possible over-reduction of the modified side chain is observed, giving rise to a fragment peak 2 Da higher in mass.
E. coli expressing PylRS and either PyltRNA\textsubscript{AGA}, PyltRNA\textsubscript{GCU}, PyltRNA\textsubscript{CAU}, or PyltRNA\textsubscript{UUU} were grown for 4 h at 37 °C – either with or without 1 (0.1 mM). Lysates were labelled with tetrazine-fluorophore conjugate, 6 (4 µM) for 4h, resolved by SDS-PAGE, and then analysed by Coomassie stain (bottom) or in-gel fluorescence (top). As based on the Coomassie image, all eight samples were identically loaded, but fluorescence labeling is strictly 1-dependent for any tRNA anticodon used.
Figure S4 (related to Figure 4). Quantity of eluted protein from enrichment with 2. A. SDS-PAGE gel analysis, lanes 1-7 loaded with serial dilutions of a lysate of known concentration from 42-0.42 µg of loaded protein (determined by Bradford assay). In lane 8 an aliquot of SORT-E proteins eluted from enrichment of GCU(Ser) lysate is loaded. Densitometry analysis of the lanes indicate that 1.3 µg of protein was loaded into lane 8. B. Standard curve generated from lane densitometry.
**Figure S5 (related to Figure 4).** *E. coli* expressing PylRS and PyltRNA\textsubscript{AGA} were grown for 4 h at 37 °C – either with or without 1 (0.1 mM). Next, lysates were either not diluted, diluted 10-fold, or diluted 100-fold with lysates from cultures that lacked 1. Lysates were labelled with 2, tagged proteins were captured with streptavidin beads, non-specifically bound proteins were washed away, and specifically captured proteins were eluted by cleavage of 2. Aliquots from the initial lysate, the final wash, and eluted fraction were analysed by SDS-PAGE. In the case of the 10-fold and 100-fold diluted samples, 10-fold and 100-fold more input was used, so that the total amount of labelled protein is comparable in all lanes, but the concentration of labelled proteins with respect to non-labelled is lower. At 10-fold dilution, the elution profile is similar. At 100-fold dilution, non-specific reactions of 2 with unlabeled proteins begin to match the level of specific capture.
Figure S6 (related to Figure 4). The subset of proteins that are identified in both the SORT-E samples and controls (blue circles in Figure 4 panels (B, D, F, G) from the main text are enriched, as judged by their LFQ values, in the SORT-E samples.
Figure S7 (related to Figure 5). (A, B) SORT-E enriches lower abundance proteins more efficiently than higher abundance proteins. The graphs show the log of the ratio of the relative abundance for proteins detected in SORT-E ((A) SORT-E (AGA, Ser); (B) SORT-E (CAU, Met)) to the relative abundance prior to enrichment. Each point corresponds to an average over all proteins in a given range of abundance, as defined in the PAX database (Wang et al., 2015; Wang et al., 2012). Error bars correspond to std. devs across biological triplicates. For SORT-E (AGA, Ser): $y = -0.132 \ln(x) + 0.622; R^2 = 0.86$. For SORT-E (CAU, Met): $y = -0.166 \ln(x) + 0.713; R^2 = 0.85$. 
Figure S8 (related to Figure 6). (A) SDS-PAGE gel analysis for the quantification of eluted protein from SORT-E AGA(Ser) and SORT-E CAU(Met) used for TMT analysis, lanes 1-6 loaded with an aliquot of the combined eluent, lanes 7-10 loaded with serial dilutions of a lysate of known concentration form 0.5-5 µg of loaded protein (determined by Bradford assay). Lane densitometry calculations reveal 109 µg, 116 µg and 142 µg of SORT-E AGA(Ser) protein yield in the total 250 µL eluents. And 102 µg, 98 µg and 106 µg of SORT-E CAU(Met) protein yield in the total 250 µL eluents.
Aliquots from the initial lysate, the final wash, and eluted fractions for each biological replicate of the SORT-E TMT experiment were analysed by SDS-PAGE.

Histograms showing the coefficients of variation (CV) for protein abundances from three biological replicates. Protein abundances, as based on relative TMT ion counts, are very reproducible for all conditions studied. The average CV for proteins from unmodified *E. coli* lysates is 0.075. The average CV for proteins from SORT-E AGA(Ser) and CAU(Met) input proteins are 0.113 and 0.137, respectively. The average CV from SORT-E AGA(Ser) and CAU(Met) proteins are 0.141 and 0.161, respectively.
Figure S9 (related to Figure 6). (A, B) If stochastic recoding of sense codons were a rare event ($P_{\text{SORT}}$ is small), a protein that has more occurrences of the recodable codons has a higher likelihood of getting labeled, as based on Poisson statistics. (C, D) If stochastic recoding is relatively frequent, then a protein with more recodable codons will be labeled more times, but the likelihood of getting one label saturates quickly with
respect to the number of recodable codons. (E, F, G, H) The number of cognate codons in a given protein that can be decoded by the PyiT variant only weakly explains the variance in the SORT-E pull-down efficiencies across proteins, though the correlation is stronger for serine than for methionine. It is reasonable that \( P_{\text{SORT}} \) would be lower for serine because the intracellular tRNA\(_{\text{Ser}}\) concentration is ca. 4-fold greater than that of tRNA\(_{\text{Met}}\) (Dong et al., 1996). The graphs show the log of the ratio of the relative abundance for proteins detected in SORT-E to the relative abundance prior to enrichment plotted against codons counts (where “Sers” includes UCU, UCA, and UCC). The least-square regression lines have the following parameters: (E) slope = 0.018, \( R^2 = 0.019, p = 10^{-6} \). (F) slope = 0.015, \( R^2 = 0.012, p = 10^{-4} \). (G) slope = 0.022, \( R^2 = 0.023, p = 10^{-7} \). (H) slope = 0.042, \( R^2 = 0.076, p = 10^{-23} \). These results are more consistent with high-frequency recoding (model C) than low-frequency recoding (model A). However, this model does not take into account the efficiency with which tagged proteins may be captured.
Figure S10 (related to Figure 6). (A, B) Proteins that have incorporated \( \textbf{1} \) are labeled to a tetrazine diazobenzene biotin compound, enabling them to be captured on streptavidin-coated beads. The efficiency of labeling could be affected by the accessibility of the positions in the protein where \( \textbf{1} \) can be incorporated, especially if labeling chemistry were slow and inefficient. (C, D) The median solvent-accessible surface area (SASA) of a \( \text{SORT}- \text{decodable site in a given protein explains almost no variation in the SORT-E pull-down efficiencies across proteins. Normalized SASAs were calculated in PyMol for all } \text{E. coli} \text{ proteins for which an x-ray structure was available in the PDB, and for each protein, the median SASA across all residues that could be replaced with } \textbf{1} \text{ was taken. For (C), } R^2 = 0.002; \text{ for (D), } R^2 = 0.0066. \text{ These observations are consistent with the labeling chemistry being very efficient, and therefore not very hindered by protein structure.} \)
**Figure S11 (related to Figure 7).** (A) SDS-PAGE analysis of SORT-M labeling of ovary proteomes from *D. melanogaster*. Ovaries were harvested from 15 females expressing PylRS and PylT<sub>xxx</sub> fed normal food containing I (10 mM), or in the control lane from 15 females not expressing PylRS and PylT<sub>xxx</sub> but fed normal food containing I (10 mM). The protein lysates were labeled with tetrazine fluorophore conjugate 7 (4 µM) (Elliott et al., 2014). PylT<sub>xxx</sub> corresponds to variants of PylT bearing anti codon A = PylT<sub>UGC</sub> (Ala) or S = PylT<sub>GCU</sub> (Ser) or L = PylT<sub>CAG</sub> (Leu) or M = PylT<sub>CAU</sub> (Met) or combinations bearing two variants of PylT<sub>xxx</sub> for example AM = PylT<sub>UGC</sub> (Ala) and PylT<sub>CAU</sub> (Met).
(B) SDS-PAGE analysis of SORT-M labeling of ovary proteomes from *D. melanogaster*. Ovaries were harvested from 15 females expressing PylRS and PylT, fed either normal food containing 1 (10 mM) or normal food without 1. The protein lysates were labeled with tetrazine fluorophore conjugate 7 (4 µM) (Elliott et al., 2014). Here PylT corresponds to flies bearing combinations of two or three PylT variants for example LM = PylT_{CAG} (Leu) and PylT_{CAU} (Met); ASM = PylT_{UGC} (Ala) and PylT_{GCU} (Ser) and PylT_{CAU} (Met).

(C) Structure of tetrazine fluorophore conjugate, 7 (Elliott et al., 2014).
Figure S12 (related to Figure 7). (A) A representative SDS-PAGE analysis comparing the efficiency of SORT-M from *E. coli* cells to SORT-M from *D. melanogaster* ovaries. Equal quantities of protein lysate from *E. coli*, bearing PylRS and PylT<sub>AGA</sub> grown in LB supplemented with 1 (0.1 mM) or grown in LB without 1; and protein lysate from *D. melanogaster*, bearing PylRS and three variants of PylT (PylT<sub>UGC</sub>, PylT<sub>GCU</sub> and PylT<sub>CAU</sub>), fed either normal food supplemented with 1 (10 mM) or normal food without 1; were labeled with tetrazine fluorophore conjugate 6 (4 µM) (Lang et al., 2012). Densitometry calculations on fluorescence gels of three independent replicates show that *E. coli* proteome labeling is only 2.13±0.75 times more efficient than germline cell labeling. (B) SDS-PAGE analysis comparing the efficiency of SORT-E from *E. coli* cells to SORT-E from *D. melanogaster* ovaries. 3.5 mg of protein lysate from *E. coli*, bearing PylRS and PylT<sub>AGA</sub> grown in LB supplemented with 1 (0.1 mM) or grown in LB without 1; and 7 mg of ovary protein lysate from *D. melanogaster*, bearing PylRS and three variants of PylT (PylT<sub>UGC</sub>, PylT<sub>GCU</sub> and PylT<sub>CAU</sub>), fed either normal food supplemented with 1 (10 mM) or normal food without 1; were labeled with TDB probe 2 (20 µM) and subjected to the same SORT-E enrichment protocol. SDS-PAGE analysis of the input material, the final wash step, and the elution with Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub> show comparable protein enrichment between *E. coli* and fly, in both an amino acid and Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub> dependent manner. Indicating that germline cell labeling efficiency approaches that of *E. coli* cell labeling.
Figure S13 (related to Figure 7). (A) Venn diagram representing the number of proteins identified by mass spectrometry from the SORT-E elution (grey) and the no amino acid control (blue) of proteins from D. melanogaster ovaries. The majority of proteins identified in the no amino acid control were also identified in the corresponding SORT-E sample. (B) Volcano plot of those 10 proteins identified in both the SORT-E sample (+1) and the no amino acid control sample (-1). The volcano plot shows the ratio of the LFQ values for each protein in this subset (i.e., the enrichment factor), plotted against the p-value of the null hypothesis that there is no difference between the LFQ values. The area shaded in gray corresponds to the threshold of p-values <0.05, i.e. values with a greater than 95% probability of correctly rejecting the null hypothesis.
Table S1 (related to Figure 4). Protein identifications and LFQ from MS analysis of SORT-E AGA(Ser), GCU(Ser), CAU(Met) and UUU(Lys). LFQ values for each replicate were generated in MaxQuant.

Table S2 (related to Figure 6). Protein identifications and relative quantitation from 9-plex TMT analysis of SORT-E AGA(Ser) and CAU(Met). Reporter ions 126, 127N and 127C correspond to E. coli control replicates, ions 128N, 128C and 129N correspond to AGA(Ser) or CAU(Met) input replicates (Labeled NPD1, 2 or 3) and ions 129C, 130N and 130C correspond to SORT-E AGA(Ser) or SORT-E CAU(Met) replicates (Labeled PD1, 2 or 3).

Table S3 (related to Figure 7). Protein identifications and LFQ from MS analysis of SORT-E from D. melanogaster ASM. LFQ values for each replicate were generated in MaxQuant.
II. Supplemental Methods

Chemical syntheses - general methods

All chemicals and solvents were purchased from Sigma-Aldrich, Alfa Aesar or Fisher Scientific and used without further purification unless otherwise stated. Qualitative analysis by thin layer chromatography (TLC) was performed on aluminium sheets coated with silica (Merck TLC 60F-254). The spots were visualized under short wavelength ultra-violet lamp (254nm) or stained with basic, aqueous potassium permanganate, ethanolic ninhydrin or vanillin. Flash column chromatography was performed with specified solvent systems on silica gel 60 (mesh 230-400).

LC-MS analysis was performed on Agilent 1200 machine. The solvents used consisted of 0.2 % formic acid in water (buffer A) and 0.2 % formic acid in acetonitrile (buffer B). LC was performed using Phenomenex Jupiter C18 column (150 × 2 mm, 5µm) and monitored using variable wavelengths. Retention times (Rt) are recorded to a nearest 0.1 min and m/z ratio to nearest 0.01 mass units. The following programme was used for small molecule LC gradient: 0-1 min (A:B 10:90-10:90, 0.3 mL/min), 1-8 min (A:B 10:90-90:10, 0.3 mL/min), 8-10 min (A:B 90:10-90:10, 0.3 mL/min), 10-12 (A:B 90:10-10:90, 0.3 mL/min).

Mass spectrometry analysis following LC was carried out in ESI mode on a 6130 Quadrupole spectrometer and recorded in both positive and negative ion modes. NMR analysis was carried out on a Bruker 400MHz instrument. All reported chemical shifts (δ) relative to TMS were referenced to the residual protons in deuterated solvents used: d1 – chloroform (1H δ = 7.26 ppm, 13C δ = 77.16 ppm), d6 – dimethylsulfoxide (1H δ = 2.49 ppm, 13C δ = 39.52 ppm), D2O (1H δ = 4.70). APT or two-dimentional experiments (COSY, HSQC) were always performed to provide additional information used for analysis where needed. Coupling constants are given in Hz and described as: singlet – s, doublet – d, triplet – t, quartet – q, broad singlet – br, multiplet – m, doublet of doublets – dd, etc. and combinations thereof.
Synthesis of tetrazine diazobenzene biotin 2

**Scheme S1.** Synthesis of 2. **Reagents and conditions:** i. (a) 5N HCl, NaNO₂, 0 °C, (b) NaHCO₃, acetone, 0 °C, 71% yield; ii. N-hydroxyiscuccinimide, DMF, EDCI, DMAP, RT, 48% yield; iii. 2-amino-N-(6-(6-(pyridin-2-yl)-1,2,4,5-tetrazin-3-yl)pyridin-3-yl)acetamide, Et₃N, DMF, RT; iv. TFA, CH₂Cl₂, RT, quantitative yield; v. NHS-PEG₄-Biotin (Thermo scientific), DMF, Et₃N, RT, 18% yield.

i. **(E)-4-((5-(2-((tert-butoxycarbonyl)amino)ethyl)-2-hydroxyphenyl)diazenyl)benzoic acid 3**

NaNO₂ (629 mg, 9.12 mmol, 2.5 eq) was added to a suspension of 4-aminobenzoate (500 mg, 3.65 mmol, 1 eq) in hydrochloric acid (5N, 10 mL) at 0 °C, in portions to maintain a constant 0 °C. The suspension was stirred vigorously and after approximately 1 h, at 0 °C, the solid had dissolved to give a yellow solution. The yellow solution was transferred to a dropping funnel and added dropwise to a suspension of N-boc-tyramine (865 mg, 3.65 mmol, 1 eq) and NaHCO₃ (4 g, 47.6 mmol, 13 eq) in saturated NaHCO₃ solution (40 mL) and acetone (5 mL) maintained at 0 °C. The mixture was stirred vigorously for 18 h at 0 °C and adjudged complete at this time. The reaction mixture was then diluted with HCl (1N, 100 mL) and the resultant precipitate filtered. The solid was washed with water (3× 30 mL) and then acetone (3× 30 mL) and dried under vacuum desiccation. 3 was obtained as a brown powder (1g, 71% yield). LRMS m/z (ES⁺) 386 [M+H]⁺, ¹H NMR analysis was in excellent agreement with the published values.

ii. **2,5-Dioxopyrrolidin-1-yl (E)-4-((5-(2-((tert-butoxycarbonyl)amino)ethyl)-2-hydroxyphenyl)diazenyl)benzoate 4**
Carboxylic acid 3 (50 mg, 0.13 mmol, 1eq) was dissolved in DMF (1 mL), to this solution was added N-hydroxysuccinimide (18.1 mg, 0.16 mmol, 1.2 eq), followed by EDCI (30 mg, 0.16 mmol, 1.2 eq) and DMAP (8 mg, 0.07 mmol, 0.5 eq). The reaction mixture was stirred at room temperature for 2 h and adjudged complete at this time by LCMS analysis. The reaction mixture was diluted with EtOAc (10 mL) and washed with HCl (1 M, 2× 10 mL) and then brine (10 mL). The organics were dried (Na₂SO₄) filtered and concentrated under vacuum to give a crude gum. The crude material was purified by silica gel column chromatography eluting EtOAc/hexane (40/60). The product fractions were combined and concentrated under vacuum to give 4 as a dark orange gum (30 mg, 48% yield). LRMS m/z (ES⁻) 481 [M-H]⁻

iii. tert-Butyl (E)-(4-hydroxy-3-(((2-oxo-2-((6-(6-(pyridin-2-yl)-1,2,4,5-tetrazin-3-yl)pyridin-3-yl)amino)ethyl)carbamoyl)phenyl)diazenyl)phenethyl)carbamate

Activated ester 4 (30 mg, 0.062 mmol, 1 eq) was dissolved in DMF (1 mL), to this solution was added aminotetrazine (32 mg, 0.093 mmol, 1.5 eq) followed by Et₃N (45 µL, 0.31 mmol, 5 eq). The reaction was stirred at room temperature for 1 h and was adjudged complete by LCMS analysis. The crude reaction mixture was diluted with EtOAc (5 mL) causing a precipitate to form. The solid was filtered and washed with HCl (1 M, 5 mL). The solid was retained, and the filtrate extracted with more EtOAc (2× 5 mL). The organic fraction was dried under vacuum and the resultant solid combined with the solid isolated from filtration. This gave the desired product 5 (41 mg) as a red/orange solid, that used without further purification. LRMS m/z (ES⁺) 676 [M+H]⁺, m/z (ES⁻) 674 [M-H]⁻.

iv. (E)-4-(((5-(2-Aminoethyl)-2-hydroxyphenyl)diazenyl)-N-(2-oxo-2-((6-(6-(pyridin-2-yl)-1,2,4,5-tetrazin-3-yl)pyridin-3-yl)amino)ethyl)benzamide 5a

Solid 5 (41 mg, 0.062 mmol, 1 eq) was suspended in CH₂Cl₂ (2 mL), to which was added TFA (2 mL) at room temperature, this immediately formed a red/orange solution. LCMS analysis after 5 minutes showed complete boc deprotection. The reaction was therefore concentrated to dryness by passing a stream of nitrogen over the reaction, giving the crude amine as a red/orange gum that was used directly without further purification. LRMS m/z (ES⁺) 576 [M+H]⁺.

v. Tetrazine-diazobenzene-biotin 2
Crude 5a (approximately 0.062 mmol, 1eq) was dissolved in DMF (1 mL) to this was added NHS-Peg4-Biotin (55 mg, 0.093 mmol, 1.5 eq, Thermo scientific EZ-Link NHS-PEG4-Biotin – 21363) and Et₃N (45 µL, 0.31 mmol, 5 eq). The reaction was monitored by LCMS analysis and was adjudged complete after 30 minutes. The reaction mixture was concentrated to dryness and the product purified by semi-preparative HPLC (10%-90% MeCN in H₂O over 35 minute gradient at 4 mL min⁻¹ using Phenomenex Luna, 5µ, C18, 100 Å column). The product fractions were combined and freeze dried to give 2 (12 mg, 18% yield) as a red/orange powder. LRMS m/z (ES⁺) 1049 [M+H]⁺.
Expression, purification and labeling of proteins site-specifically incorporating 1.

Expression of T4lysozyme K83-1

Electro-competent E. coli DH10B cells were co-transformed with pBK-MbPylRS and pBAD-T4lysK83TAG-His6 PylT (Nguyen et al., 2011). Transformed cells were recovered in S.O.B. (1 mL, supplemented with 0.2% glucose) for 1 h at 37 ºC and used to inoculate LB containing 50 µg/mL kanamycin and 50 µg/mL tetracycline (LB-KT). The cells were incubated with shaking overnight at 37 ºC, 250 r.p.m. 1 mL of overnight culture was used to inoculate 100 mL of LB-KT½, which was then incubated (37 ºC, 250 r.p.m). At O.D.₆₀₀ ~0.3, the culture was divided equally and supplemented with either 1 (1 mM) or H₂O (500 µL) and incubated further (37 ºC, 250 r.p.m). At O.D.₆₀₀ ~0.6 protein expression was induced by the addition of arabinose (0.2%). After 4 h, the cells were harvested by centrifugation (4000 r.p.m, 20 min) and the pellet frozen until further use.

Expression of Ubiquitin K6-1 and Ubiquitin K48-1

Chemically competent E. coli BL21(DE3) cells (Merck Biosciences) were co-transformed with pBK-MbPylRS and pCDF-pylT-UbTAG6-His6 or pCDF-pylT-UbTAG48-His6 (Madrzak et al., 2015; Virdee et al., 2010). Transformed cells were recovered in S.O.B. (1 mL, supplemented with 0.2% glucose) for 1 h at 37 ºC and used to inoculate LB containing 50 µg/mL kanamycin and 25 µg/mL spectinomycin (LB-KS). The cells were incubated with shaking overnight at 37 ºC, 250 r.p.m. 1 mL of overnight culture was used to inoculate 100 mL of LB-KS½, which was then incubated (37 ºC, 250 r.p.m). At O.D.₆₀₀ ~0.3, the culture was divided equally and supplemented with either 1 (1 mM) or H₂O (500 µL) and incubated further (37 ºC, 250 r.p.m). At O.D.₆₀₀ ~0.6 protein expression was induced by the addition of IPTG (0.5 mM). After 4 h, the cells were harvested by centrifugation (4000 r.p.m, 20 min) and the pellet frozen until further use.
Purification of T4lysozyme K83-1, Ubiquitin K6-1 and Ubiquitin K48-1 from E. coli

The frozen bacterial pellets were thawed on ice and resuspended in 2.5 mL lysis buffer (Bugbuster®, Novagen®, 50 µg/mL DNAse 1, Roche inhibitor cocktail and 20 mM imidazole). Cells were incubated (4 °C, 30 minutes) then clarified by centrifugation (16000 g, 4 °C, 30 minutes). The clarified lysates were transferred to fresh tubes and 100 µL Ni-NTA slurry added. The mixtures was incubated with agitation (4 °C, 1 h) and then collected by centrifugation (1000 g, 4 °C, 5 min). The beads were resuspended three times in 500 µL wash buffer (10 mM Tris-HCL, 40 mM imidazole, 200 mM NaCl, pH 8) and collected by centrifugation (1000 g, 4 °C, 5 min). Finally, the beads were resuspended in 100 µL elution buffer (10 mM Tris-HCL, 300 mM imidazole, 200 mM NaCl, pH 8), pelleted by centrifugation (1000 g, 4 °C, 5 min) and the supernatant collected into fresh tubes. The elution was repeated three times with 100 µL of elution buffer. The purified proteins were analysed by 4-12% SDS-PAGE and LC-MS and a Bradford assay was performed to determine protein concentration (using BSA standard, for the standard curve).

Intact Protein Mass Spectrometry

ESI-MS was carried out using an Agilent 1200 LC-MS system with a 6130 Quadrupole spectrometer. The solvent system consisted of 0.2 % (v/v) formic acid in H2O as buffer A, and 0.2 % (v/v) formic acid in acetonitrile (MeCN) as buffer B. Protein UV absorbance was monitored at 214 and 280 nm. Protein MS spectra were acquired in positive ionisation mode, scanning between 400-2000 m/z. Collected spectra were averaged over the entire total ion current (TIC). Intact protein masses were calculated via spectral deconvolution using Agilent’s LC/MSD Chemstation software with built-in deconvolution tool. The default deconvolution parameters were used (masses between 500-50000 Da, with a maximum allowable charge of +50, a minimum of 5 peaks in a peak set, a noise cut off of 1000 counts and an abundance cut off at 10% for selected peaks).
In vitro labeling of T4lysozyme K83-1, Ubiquitin K6-1 and Ubiquitin K48-1 with Probe 2 and subsequent reduction with Na₂S₂O₄

Purified proteins were diluted to a concentration of 20 µM in 8M GdmCl and used at a final reaction volume of 50 µL. 2 (10 molar equivalents, 5 µL from a 2 mM stock solution in DMSO) was added and the reactants mixed by aspirating several times, the mixture was then incubated at room temperature. The reaction progress was monitored by LC-MS (injecting 5 µL/100 pmol of protein) and showed complete conversion to the ligated products T4lysozyme K83-ii, Ubiquitin K6-ii and Ubiquitin K48-ii (Figure 2) after 4 hours.

Test reduction with Na₂S₂O₄

The crude reaction mixtures were then treated with Na₂S₂O₄ (25 mM, 4 µL of a 250 mM stock), mixed by aspirating several times and incubated at room temperature. The reaction progress was monitored by LC-MS (injecting 5 µL/100 pmol of protein) and showed complete reduction to the desired products T4lysozyme K83-iii, Ubiquitin K6-iii and Ubiquitin K48-iii (Figure 2) within 30 minutes. Note: The reaction turned from yellow to colourless with seconds of adding the Na₂S₂O₄.

Proteomic incorporation of 1 via SORT in E. coli and chemoselective labeling of lysates with tetrazine probes

Proteomic incorporation of 1 via SORT

Electrocompetent E. coli DH10B cells (50 µL) were doubly transformed with pBAD_wtT4L_MbPylTXXX plasmid (Elliott et al., 2014) (2 µL, necessary for expression of PyltRNAXXX) and pBK-MbPylS plasmid (Elliott et al., 2014) (2 µL necessary for expression of PylRS). Transformed cells were recovered in 1 ml S.O.B. (supplemented with 0.2% glucose) for 1 h at 37 °C. 100 µL of the recovery was used to inoculate 50 ml LB-KT (50 µg/ml kanamycin and 25 µg/ml tetracycline) and the cultures incubated overnight (37 °C, 250 r.p.m.). Overnight cultures were diluted to OD₆₀₀ 0.3 with LB-KT½ and divided into two 50 ml culture. One culture was supplemented with 1 (0.1 mM final concentration) and other with H₂O (50 µL). Cultures were then incubated (37 °C, 250 r.p.m.). At OD₆₀₀ ~ 0.6 the cultures were incubated for a further 4 h. Cells were
harvested by centrifugation (4,000 r.p.m., 4 °C, 20 min) and then washed three times by resuspending in 1 mL of ice cold PBS, and collecting by centrifugation (4,000 r.p.m., 4 °C, 20 min) between each wash. The final bacterial pellets were immediately frozen for storage.

**SORT-M: Chemoselective labeling of proteomes tagged with 1 by SORT with tetrazine-fluorophore conjugates.**

Frozen bacterial pellets were resuspended in 5 ml Urea buffer (8M urea, 25 mM Tris-HCl, pH 8) and lysed using a microtip sonicator (amplitude 15, 90 s total sonication time, Soniprep 150 plus). The lysate was clarified by centrifugation (4 °C, 14,000 r.p.m., 30 min) and the supernatant transferred to a fresh tube. A Bradford assay was performed to determine protein concentration (using BSA standards for the standard curve). To 50 µL of cleared cell lysate was added tetrazine-fluorophore conjugate 6 (4 µM final concentration, from a 200 µM stock in DMSO, Figure S3) (Lang et al., 2012). The reactions were mixed by aspirating several times and the samples then incubated in the dark (room temperature, 4 h). After this time a 5 µL aliquot was diluted into 10 µL of pure water; 5 µL of 4× LDS sample buffer supplemented (6 mM BCN-OH and 20 mM DTT) was added and mixed by vortexing gently. Samples were incubated for 10 min before boiling at 90 °C for 10 min. Samples were analysed by 4–12% SDS-PAGE and fluorescent images were acquired using a Typhoon Trio phosphoimager (GE Life Sciences).

**SORT-E: Chemoselective labeling of proteomes tagged with 1 by tetrazine-diazobenzene-biotin conjugate 2, enrichment and elution of SORT-E proteins**

Cleared bacterial protein lysates were prepared in the same way as described above. For capture, enrichment and elution of tagged proteins, 500 µL of cleared cell lysate (8 mg mL⁻¹ protein lysate) was typically used. Thus, to 500 µL of cleared cell lysate DTT (1 mM, final concentration), was added and mixed with gentle vortexing, the mixture was incubated at room temperature for 45 minutes. After this time iodoacetamide (5.5 mM, final concentration) was added, mixed with gentle vortexing, then incubated for 30 minutes at room temperature. 2 (5 µL, 20 µM final concentration, from a 2 mM stock in
DMSO) was then added and the reactions were mixed by gentle vortexing and the samples then incubated in the dark at room temperature overnight with end-over-end rotation (Hula mixer). 10 µL aliquots were taken at this stage as input sample for subsequent SDS-PAGE analysis. The samples were then diluted with PBS to a final volume of 5 mL and high capacity streptavidin beads (150 µL of settled resin, pre-equilibrated in PBS, Thermo scientific Streptavidin Agarose resin) added. The beads were incubated with end-over-end rotation (Hula mixer) for 1.5 hours, and then collected by gravity filtration through a column (Poly-Prep® chromatography column Bio-Rad 731-1550). The beads were resuspended in urea buffer (500 µL, 8M urea, 25 mM tris, pH 8) and transferred to a smaller spin column (Mini Bio-spin® chromatography column Bio-Rad 731-1550). The beads were collected by mild centrifugation (1000 rpm, 5 second pulse) and washed 2× urea buffer (500 µL) then 2× 1% SDS in PBS (500 µL) with 15 minute incubation times between each wash. The final wash acts as a control for the Na$_2$S$_2$O$_4$ specific elution, the beads were resuspended in 1% SDS in PBS (150 µL) and incubated at room temperature for 30 minutes with end-over-end rotation (Hula mixer). The beads collected by mild centrifugation (1000 rpm, 5 second pulse) and the supernatant kept for subsequent analysis by SDS-PAGE. Specifically bound proteins were then eluted by resuspending the beads in 1% SDS in PBS supplemented with 25 mM Na$_2$S$_2$O$_4$ (150 µL) and incubating at room temperature for 30 minutes with end-over-end rotation (Hula mixer). The supernatant was collected by mild centrifugation (1000 rpm, 5 second pulse) and analysed by SDS-PAGE.

**SDS-PAGE analysis and LC-MS/MS of enriched proteins tagged via SORT**

5 µL aliquots of each lysate input sample was diluted into 10 µL of pure water; 15 µL of each final wash sample and 15 µL of each elution sample were transferred to fresh tubes. 5 µL of 4× LDS sample buffer (supplemented with 80 mM DTT) was added and mixed by vortexing gently. Samples were incubated for 10 min before boiling at 90 °C for 10 min. Samples were analyzed by 4–12% SDS-PAGE, 200 V constant, 40 minutes. Gels were stained with Sypro®-Ruby using the standard manufacturer’s protocol and imaged with a Typhoon Trio phosphoimager (GE Life Sciences) excitation – 410 nm, emission filter at 580 nm.
For LC-MS/MS each pull-down was performed on 3 biological replicates. Eluted proteins were loaded onto 4–12% SDS-PAGE precast gels, as described above and each lane excised and cut into approximately 24 slices. The excised protein gel pieces were placed in a well of a 96-well microtiter plate and destained with 50% v/v acetonitrile and 50 mM ammonium bicarbonate, reduced with 10 mM DTT and alkylated with 55 mM iodoacetamide. After alkylation, proteins were digested with 6 ng/µL trypsin (Promega, UK) overnight at 37 °C. The resulting peptides were extracted in 2% v/v formic acid, 2% v/v acetonitrile. The digest was analyzed by nano-scale capillary LC-MS/MS using an Ultimate U3000 HPLC (Thermo Scientific Dionex, San Jose, USA) to deliver a flow of ~300 nL/min. A C18 Acclaim PepMap100 5 µm, 100 µm × 20 mm nanoViper (Thermo Scientific Dionex, San Jose, USA) trapped the peptides before separation on a C18 Acclaim PepMap100 3 µm, 75 µm × 250 mm nanoViper (ThermoScientific Dionex, San Jose, USA). Peptides were eluted with a gradient of acetonitrile. The analytical column outlet was directly interfaced by means of a modified nano-flow electrospray ionization source, with a hybrid dual-pressure linear ion trap mass spectrometer (Orbitrap Velos, Thermo Scientific, San Jose, USA). Data-dependent analysis was carried out, using a resolution of 30,000 for the full MS spectrum, followed by ten MS/MS spectra in the linear ion trap. MS spectra were collected over a m/z range of 300–2,000. MS/MS scans were collected using a threshold energy of 35 for collision-induced dissociation. LC-MS/MS data were then searched against a protein database (UniProt KB) using the Mascot search engine program (Matrix Science, UK). Database search parameters were set with a precursor tolerance of 5 p.p.m. and a fragment ion mass tolerance of 0.8 Da. Two missed enzyme cleavages were allowed and variable modifications for oxidized methionine, carbamidomethyl cysteine, pyroglutamic acid, phosphorylated serine, threonine and tyrosine were included. Proteomic data were worked up in the MaxLFQ software suite (1) to identify proteins from peptide sequences and calculate label-free quantifications (LFQ).

**SORT-E: Combined with a 9-Plex TMT analysis**

*Production of control protein lysates*
E. coli DH10B cells were inoculated into three LB cultures (50mL) and incubated overnight (37 °C, 250 r.p.m.). Overnight cultures were diluted to OD<sub>600</sub> 0.3 with LB (250mL) and incubated (37 °C, 250 r.p.m.). At OD<sub>600</sub> ~ 0.6 the cultures were incubated for a further 4 h. Cells were then harvested by centrifugation (4,000 r.p.m., 4 °C, 20 min) and frozen until further use.

Frozen bacterial pellets were resuspended in 50 mL Urea buffer (8M urea, 25 mM tris, pH 8) and lysed using a microtip sonicator (amplitude 15, 90 s total sonication time, Soniprep 150 plus). The lysate was cleared by centrifugation (4 °C, 14,000 r.p.m., 30 min) and the supernatant aspirated to a fresh tube, a Bradford assay was performed to determine protein concentration (using BSA standards, for standard curve).

Production of SORT-E input material
For SORT-E AGA(Ser) and SORT-E CAU(Met) three biological replicates of cleared bacterial protein lysates were prepared in the same way as described above for SORT-E, but scaled up by a factor of 10.

Production of E. coli SORT-E material
For SORT-E AGA(Ser), 3.5 mL (25 mg protein) of the above cleared protein lysate were used for each replicate. For SORT-E CAU(Met), 7 mL (50 mg protein) of the above cleared protein lysates were used for each replicate. Thus, to each lysate DTT (1 mM) was added and mixed with gentle vortexing, the mixture was incubated at room temperature for 45 minutes. After this time iodoacetamide (5.5 mM) was added, mixed with gentle vortexing, then incubated for 30 minutes at room temperature. 2 (5 µL, 20 µM final concentration, from a 2 mM stock in DMSO) was then added and the reactions were mixed by gentle vortexing and the samples then incubated in the dark at room temperature overnight with end-over-end rotation (Hula mixer). 10 µL aliquots were taken at this stage as input sample for subsequent SDS-PAGE analysis. The labelled protein samples were precipitated by the addition of 10 volumes of ice-cold methanol and incubated at -25 °C overnight. The proteins were pelleted by centrifugation (1000g, 30min) and washed three times with further volumes (10 mL) of ice-cold methanol. Washed protein pellets were re-suspended in 3.5 mL or 7 mL of Urea buffer (8M urea, 25
mM tris, pH 8) respectively. High capacity streptavidin beads (50 µL of settled resin, Thermo scientific Streptavidin Agarose resin) were added to each sample and incubated with end-over-end rotation (Hula mixer) for 1.5 hours. Beads were then collected by gravity filtration through a column (Poly-Prep® chromatography column Bio-Rad 731-1550). The beads were resuspended in urea buffer (500 µL, 8M urea, 25 mM tris, pH 8) and transferred to a smaller spin column (Mini Bio-spin® chromatography column Bio-Rad 731-1550). The beads were collected by mild centrifugation (1000 rpm, 5 second pulse) and washed 2× urea buffer (500 µL) then 2× 1% SDS in PBS (500 µL) with 15 minute incubation times between each wash. The final wash acts as a control for the Na₂S₂O₄ specific elution, the beads were resuspended in 1% SDS in PBS (50 µL) and incubated at room temperature for 30 minutes with end-over-end rotation (Hula mixer). The beads were collected by mild centrifugation (1000 rpm, 5 second pulse) and the supernatant kept for subsequent analysis by SDS-PAGE. Specifically bound proteins were then eluted by resuspending the beads in 1% SDS in PBS supplemented with 50 mM Na₂S₂O₄ (50 µL) and incubating at room temperature for 30 minutes with end-over-end rotation (Hula mixer). The supernatant was collected by mild centrifugation (1000 rpm, 5 second pulse), this elution step was repeated a further four times and the eluents combined to give a total of 250 µL SORT-E protein fraction for each sample. A 15 µL aliquot was taken for SDS-PAGE analysis. SORT-E protein eluents were then 10-fold diluted with PBS and then concentrated using centrifugal spin filters (3 kDa cut-off, Merk-Millpore Amicon Ultra-0.5 mL) to a final volume of approximately 100 µL. Protein concentration was determined by a BCA colorimetric assay (ThermoFisher – 23225).

Preparation of 9-Plex TMT analysis

50 µg of each protein sample replicate for each test condition (3× control proteins, 3× SORT-E input and 3× SORT-E) were trypsinised and the resulting peptides labelled with unique isobaric mass tags as per the manufacturers instructions (TMT10plex™ Isobaric Mass Tag Labeling Kit; ThermoFisher – 90113). An aliquot of the TMT labelled pool was evaporated to dryness, resuspended in 5% formic acid and then desalted using SepPak cartridges according to the manufacturers instructions (Waters, Milford,
Massachusetts, USA). Eluate from the SepPak cartridge was again evaporated to dryness and resuspended in 1% formic acid prior to analysis by nano-LC MSMS using an Orbitrap Fusion Tribrid Mass Spectrometer.

**Nano-LC Mass Spectrometry**

The sample was fractionated using an Ultimate 3000 nanoHPLC system in line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). In brief, peptides in 1% (vol/vol) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific). After washing with 0.5% (vol/vol) acetonitrile 0.1% (vol/vol) formic acid peptides were resolved on a 250 mm × 75 µm Acclaim PepMap C18 reverse phase analytical column (Thermo Scientific) over a 150 min organic gradient, using 7 gradient segments (1-6% solvent B over 1min., 6-15% B over 58min., 15-32%B over 58min., 32-40%B over 5min., 40-90%B over 1min., held at 90%B for 6min and then reduced to 1%B over 1min.) with a flow rate of 300 nl min⁻¹. Solvent A was 0.1% formic acid and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid. Peptides were ionized by nano-electrospray ionization at 2.0kV using a stainless steel emitter with an internal diameter of 30 µm (Thermo Scientific) and a capillary temperature of 275°C.

All spectra were acquired using an Orbitrap Fusion Tribrid mass spectrometer controlled by Xcalibur 2.0 software (Thermo Scientific) and operated in data-dependent acquisition mode using an SPS-MS3 workflow. FTMS1 spectra were collected at a resolution of 120 000, with an automatic gain control (AGC) target of 200 000 and a max injection time of 50ms. The TopN most intense ions were selected for MS/MS. Precursors were filtered according to charge state (to include charge states 2-7) and with monoisotopic precursor selection. Previously interrogated precursors were excluded using a dynamic window (40s +/-10ppm). The MS2 precursors were isolated with a quadrupole mass filter set to a width of 1.2m/z. ITMS2 spectra were collected with an AGC target of 5000, max injection time of 120ms and CID collision energy of 35%.

For FTMS3 analysis, the Orbitrap was operated at 60 000 resolution with an AGC target of 50 000 and a max injection time of 120ms. Precursors were fragmented by high energy collision dissociation (HCD) at a normalised collision energy of 55% to ensure
maximal TMT reporter ion yield. Synchronous Precursor Selection (SPS) was enabled to include up to 5 MS2 fragment ions in the FTMS3 scan.

**Analysis of Proteomic Data**

Shotgun proteomic data were worked up in the MaxLFQ software suite (1) to identify proteins from peptide sequences and calculate label-free quantifications (LFQ). Identified proteins and LFQs were exported to Excel, where proteins were counted and compared across different experiments. The number of proteins in common between SORT-E (AGA, Ser), SORT-E (GCU, Ser), SORT-E (CAU, Met) and SORT-E (UUU, Lys) were calculated in Excel, and Venn-Diagrams were constructed using eulerAPE (2). For further analyses, identified proteins were paired with their molecular weights (MW), relative abundances, and codon counts (the number of occurrences of a given codon within the protein’s open reading frame). Codon counts and molecular weights were extracted on a protein-by-protein basis by parsing through the annotated genebank file for the *E. coli* str. K-12 substr. MG1655 whole genome (accession code NC_000913), matching gene names (including synonyms) against our proteomic data set, and extracting (or calculating) the relevant data. These operations were automated through a python script called CodonCounter.py (provided as supplemental program 1) equipped with the BioPython package. Protein abundances were taken from an integrated meta-dataset of *E. coli* protein abundances (accession code 511145), compiled by PaxDB (http://pax-db.org/). As the abundances were indexed under locus tags, CodonCounter.py also cross-referenced locus to gene names in order to assist the compilation of these data. Sorting and t-test calculations were carried out in Excel, and plots were prepared in Veusz (http://home.gna.org/veusz/).

**Analysis of Tandem Mass Tag (TMT) Proteomic Data**

Protein mixtures obtained from one of three replicates for one of five test conditions (unmodified *E. coli* lysate, lysates from *E. coli* equipped with components to perform SORT-E (AGA, Ser) or SORT-E (CAU, Met) (i.e., the SORT-E inputs), and the
corresponding SORT-E enrichments were reacted with one of nine tandem mass tag labels, and submitted to LC-MS/MS in sets of nine.

The raw data files were processed and quantified using Proteome Discoverer software v1.4 (Thermo Scientific). Peptide precursor mass tolerance was set at 10ppm, and MS/MS tolerance was set at 0.6Da. Search criteria included oxidation of methionine (+15.9949 Da) as a variable modification and carbamidomethylation of cysteine (+57.0214 Da) and the addition of the TMT mass tag (+229.163 Da) to peptide N-termini and lysine as fixed modifications. Searches were performed with full tryptic digestion and a maximum of 1 missed cleavage was allowed. The reverse database search option was enabled and all peptide data was filtered to satisfy a false discovery rate (FDR) of 5%. Proteome Discoverer searched peptide sequences against the UniProt E. coli database using the SEQUEST algorithm to assign peptides to proteins. Relative peptide abundances were calculated based on TMT ion counts. The protein abundance was taken as the median abundance across all peptides identified to that protein, and these data were exported to an Excel spreadsheet. Statistics for protein abundances were generated by calculating the mean, standard deviation, and coefficient of variation (see Figure S7 D-H) of these medians across the three biological replicates.

The sum of the TMT ion counts across all proteins identified in the SORT-E pull-down experiments was about one-fourth the analogous sums for the non-pull-down experiments, despite the fact that an identical amount of protein was loaded for all experiments. We believe this is because SORT-E elution conditions solubilize a streptavidin contaminant from beads (which appears in gels as a dense low molecular-weight band in elution lanes (e.g., elution lanes in Figures 3, S5)), which lowers the fraction of peptides arising from E. coli proteins. Before further analysis, we therefore normalized the TMT ion counts from every mass channel, making the sum of all TMT ion counts in each channel the same.

Where we refer to a protein’s SORT-E pull down efficiency, we mean the average of that protein’s TMT ion counts from the three mass channels derived from the enrich samples divided by the average of that protein’s TMT ion counts from the three mass channels derived from the input samples prior to enrichment.
Mathematical manipulations, sorting, and t-test calculations were carried out in Excel, and plots were prepared in Veusz.

**Fly lines and culture conditions**

All flies were grown at 25°C on standard Iberian medium. Flies were fed 1 by mixing dried yeast with a solution of 1 (10mM) to form a paste. This paste was added as a supplement to the normal Iberian fly food for a minimum of 24h and the yeast was changed daily.

Double and triple sense codon lines were created by recombination using the original lines FT58 (A = PylT\textsubscript{UGC}, Ala), FT60 (S = PylT\textsubscript{GCU}, Ser), FT62 (L = PylT\textsubscript{CAG}, Leu) and FT63 (M = PylT\textsubscript{CAU}, Met)(Elliott et al., 2014). Trans-heterozygous virgins were collected for each pair-wise combination of sense codon (AS – FT58/FT60, AL – FT58/FT62, AM – FT58/FT63, SL – FT60/FT62, SM – FT60/FT63, and LM – FT62/FT63) and crossed to males of the third chromosome balancer stock \textit{w}; ; TM3/TM6. Potential recombinant males were identified based on eye colour, and individuals were backcrossed to virgins of \textit{w}; ; TM3/TM6 to make a balanced stock. Recombinant lines were then screened by crossing to nos-\textit{vp16-GAL4} virgins (Bloomington 4937) to create FT58-60/nos-\textit{vp16-GAL4} (AS), FT58-62/nos-\textit{vp16-GAL4} (AL) etc. and compared with the original single sense codon lines FT58/nos-\textit{vp16-GAL4}, FT60/nos-\textit{vp16-GAL4}, FT62/nos-\textit{vp16-GAL4} and FT63/nos-\textit{vp16-GAL4}. The females were fed 10 mM 1 for 24 - 48h and then the ovaries were extracted from 15 females of the indicated genotype and labeled with 4 \(\mu\text{M} 7\) (Figure S11) for 2h as described (Elliott et al., 2014).

Triple sense codon lines were generated in a similar manner. In this case the double sense codon lines were crossed to a different single sense codon line to generate trans-heterozygotes with three different sense codons. In this case not all combinations produced viable trans-heterozygotes and in some cases trans-heterozygotes were viable but gave no potential recombinant males. Successful combinations were FT58-63 (AM)/FT60 (S), FT60-63 (SM)/FT62 (L) and FT58-60 (AS)/FT63 (M) which generated the lines FT58-60-63 (ASM) and FT60-62-63 (SLM). These lines were screened by
crossing to nos-vp16-GAL4 virgins and compared with the double sense codon lines as described above (Elliott et al., 2014).

**SORT-E from D. melanogaster**

Ovaries were dissected from 250 females flies of FT58-60-63/nos-vp16-Gal4 (ASM) which had been fed normal food either supplemented with 1 (10 mM) or without 1. The ovaries were homogenized into 8 M urea 15 mM tris (250 µL) and the resultant protein lysate clarified by filtration. Bradford assay was used to determine the protein concentration. For SORT-E and subsequent mass spectrometry, typically 7 mg of fly ovary protein was used and labeled with 2 (20 µM) in an identical procedure as described in the protocol for *E. coli* above.
### III. Supplemental program

**Python scripts**

`#CodonCounter.py`

`#Call with one argument, a file with a list of proteins`

`#Requires Biopython package, an annotated genbank file, and the PaxDB file if abundance data is desired`

```python
from Bio import SeqIO
import sys
from Bio.Seq import Seq
from Bio.Alphabet import IUPAC
from Bio import SeqUtils
import numpy as np

proteinList = open( sys.argv[1] , 'r')
proteinListOut_fn = sys.argv[1].split('.')[0] + '_out.txt'
proteinListOut = open( proteinListOut_fn , 'w')

#Will hold tuples of form (TCT,ATG,AGC,AGT,AAA,AAG)
codons = []
K12 = SeqIO.read("Ecoli K12 genome.gb", "genbank")
abundanceData = np.genfromtxt('511145-E.coli_whole_organism-integrated_dataset.txt',skiprows=9, dtype=None)
abundanceRange = range( abundanceData.shape[0] )
abundanceLocusTags = [
abundances = []
for i in abundanceRange:
    abundanceLocusTags.append( abundanceData[i][1].split('.')[1] )
    abundances.append( abundanceData[i][2] )

for entry in proteinList:
    found = 0
    protein = entry.split()[0]
    for feature in K12.features:
        if (feature.type=="CDS"):
            gene = feature.qualifiers['gene'][0]
            synonyms = [x.strip() for x in feature.qualifiers['gene_synonym'][0].split(';')] 
            synonyms.append(gene)
            if protein in synonyms:
                found = 1
                start = feature.location.start.position
                end = feature.location.end.position
                sense = feature.strand
                locustag = feature.qualifiers['locus_tag'][0]
                if sense == 1:
                    geneseq = K12.seq[start:end]
                else:
                    geneseq = K12.seq[start:end].reverse_complement()
                if locustag in abundanceLocusTags:
                    abundances[abundanceLocusTags.index(locustag)] += 1
                proteinListOut.write("%s	%s	%s
" % (geneseq,locustag,abundanceLocusTags.index(locustag)))
```

abundanceIndex = abundanceLocusTags.index(locustag)

else:
    abundance = abundances[abundanceIndex]

if found == 0:
    proteinListOut.write('**' + entry)

proteinList.close()
proteinListOut.close()

**SASA calculator**

#GetSASAscores.py
#This script is fed a list with this structure
# gene-name, PDB-id, chain-id
# gene-name, PDB-id, chain-id
# ...
#For each protein, it calculates the SASA score for each site, the tRNA decoding efficiency for each site
#and provides final scores

#This function writes up a SASA-calculating program that is shelled out to pymol as a subprocess
#It needs to know the PDB-id and chain-id of the protein you want
def writeProgram(PDB, chain, dPos):
    programText = ""
    import pymol
    from pymol import cmd
    from pymol import stored
    pymol.finish_launching()

    cmd.fetch('%%s')
    stored.resids = %s
    stored.resnames = []
    for i in stored.resids:
#check if residue exists
        stored.test = []
        cmd.iterate('resi '+str(i)+' and chain %s and name CA',
                        'stored.test.append(resn)')
        if stored.test == []:  
            stored.resnames.append('NONE')
        else:
            cmd.iterate('resi '+str(i)+' and chain %s and name CA',
                        'stored.resnames.append(resn)')
        stored.test = []
cmd.set('dot_solvent', 1)
cmd.set('dot_density', 2)

stored.sasa_per_residue = []
for i in stored.resids:
    # check if residue exists
    stored.test = []
    cmd.iterate('resi '+str(i)+' and chain %s and name CA',
                'stored.test.append(resn)')
    if stored.test == []:
        stored.sasa_per_residue.append(0.8)
    else:
        stored.sasa_per_residue.append(cmd.get_area('resi '+str(i)+'
and chain %s'))

stored.sa_per_residue = []
for i in stored.resnames:
    if i == 'ALA':
        stored.sa_per_residue.append(106)
    elif i == 'CYS':
        stored.sa_per_residue.append(135)
    elif i == 'ASP':
        stored.sa_per_residue.append(163)
    elif i == 'GLU':
        stored.sa_per_residue.append(194)
    elif i == 'PHE':
        stored.sa_per_residue.append(197)
    elif i == 'GLY':
        stored.sa_per_residue.append(84)
    elif i == 'HIS':
        stored.sa_per_residue.append(184)
    elif i == 'ILE':
        stored.sa_per_residue.append(169)
    elif i == 'LYS':
        stored.sa_per_residue.append(205)
    elif i == 'LEU':
        stored.sa_per_residue.append(164)
    elif i == 'MET':
        stored.sa_per_residue.append(188)
    elif i == 'ASN':
        stored.sa_per_residue.append(157)
    elif i == 'PRO':
        stored.sa_per_residue.append(136)
    elif i == 'GLN':
        stored.sa_per_residue.append(198)
    elif i == 'ARG':
        stored.sa_per_residue.append(248)
    elif i == 'SER':
        stored.sa_per_residue.append(130)
    elif i == 'THR':
        stored.sa_per_residue.append(142)
    elif i == 'VAL':
        stored.sa_per_residue.append(142)
    elif i == 'TRP':
        stored.sa_per_residue.append(227)
    elif i == 'TYR':
        stored.sa_per_residue.append(222)
elif i == 'NONE':
    stored.sa_per_residue.append(1)
else:
    stored.sa_per_residue.append(180)

stored.sasa_relative = [x/y for x,y in zip(stored.sasa_per_residue, stored.sa_per_residue)]
print stored.sasa_relative
return programText

import sys
import os
import re
import subprocess
from Bio import SeqIO
from Bio.Seq import Seq
from Bio.Alphabet import IUPAC
from Bio import SeqUtils
import numpy as np

#call the program with a csv, as above
listFn = sys.argv[1]
list = open( listFn , 'r')

outListFn = listFn.split('.')[0]+'_out.txt'
outList = open( outListFn , 'w')
sasaListFn = listFn.split('.')[0]+'_sasa.txt'
sasaList = open( sasaListFn , 'w')

K12 = SeqIO.read("Ecoli K12 genome.gb", "genbank")

for entry in list:
geneName = entry.split()[0]
try:
    PDBid = entry.split()[1]
except IndexError:
PDBid = '.'
try:
    chainid = entry.split()[2]
except IndexError:
    pass
for feature in K12.features:
    if (feature.type=="CDS"):
        if (feature.qualifiers['gene'][0] == geneName):
            start = feature.location.start.position
            end = feature.location.end.position
            sense = feature.strand
            locustag = feature.qualifiers['locus_tag'][0]
            if sense == 1:
                geneseq = K12.seq[start:end]
            else:
                geneseq = K12.seq[start:end].reverse_complement()
codons = [geneseq[i:i+3] for i in range(0,len(geneseq),3)]
    Mets = []
Sers = []
for j in range(1, len(codons)): #skip the initial codon because that uses initiator tRNA
    if codons[j] == 'ATG':
        Mets.append(j+1)
    if codons[j] == 'TCT' or codons[j] == 'TCA' or codons[j] == 'TCC':
        Sers.append(j+1)
decodablePositions = str(Mets + Sers)
break
#Calculate SASA at all the decodable positions in a subprocess that uses the pymol interpreter
if PDBid == '.':
    outList.write(geneName + 'n')
else:
    SASAfn = geneName+'.py'
    SASAscript = open(SASAfn, 'w')
    SASAscript.write(writeProgram(PDBid, chainid, decodablePositions))
    SASAscript.close()
    SASAoutput = subprocess.check_output('pymol -rqc %s' % SASAfn, shell=True)
    begin = SASAoutput.find('[')
    end = SASAoutput.find(']')
    SASAoutput1 = SASAoutput[begin:end]
    SASAs = [round(float(re.sub(r'^-9.\n', '', x)), 2) for x in SASAoutput1.split(',')]
    #makes sure all non-numerical characters are removed before trying to cast as float
    #Determine the Met Score
    if len(Mets) == 0:
        MetMedian = 0
        MetMax = 0
    else:
        MetMedian = np.median(SASAs[0:len(Mets)])
        MetMax = np.max(SASAs[0:len(Mets)])
    #Determine the Ser Score
    if len(Sers) == 0:
        SerMedian = 0
        SerMax = 0
    else:
        SerMedian = np.median(SASAs[len(Mets):len(Mets)+len(Sers)])
        SerMax = np.max(SASAs[len(Mets):len(Mets)+len(Sers)])
    outList.write(geneName + 't %s t %s t %s t %s
\n%s' % (MetMedian, MetMax, SerMedian, SerMax))
sasaList.write(geneName + 'n' + str(SASAs) + 'n')
list.close()
outList.close()
sasaList.close()
IV. References

Dong, H., Nilsson, L., and Kurland, C.G. (1996). Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. Journal of molecular biology 260, 649-663.

Elliott, T.S., Townsley, F.M., Bianco, A., Ernst, R.J., Sachdeva, A., Elsasser, S.J., Davis, L., Lang, K., Pisa, R., Greiss, S., et al. (2014). Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. Nature biotechnology 32, 465-472.

Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A., and Chin, J.W. (2012). Genetically encoded norbornene directs site-specific cellular protein labeling via a rapid bioorthogonal reaction. Nature chemistry 4, 298-304.

Madrzak, J., Fiedler, M., Johnson, C.M., Ewan, R., Knebel, A., Bienz, M., and Chin, J.W. (2015). Ubiquitination of the Dishevelled DIX domain blocks its head-to-tail polymerization. Nature communications 6, 6718.

Nguyen, D.P., Elliott, T., Holt, M., Muir, T.W., and Chin, J.W. (2011). Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bioorthogonal cyanobenzothiazole condensation. Journal of the American Chemical Society 133, 11418-11421.

Virdee, S., Ye, Y., Nguyen, D.P., Komander, D., and Chin, J.W. (2010). Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. Nature chemical biology 6, 750-757.

Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics 15, 3163-3168.

Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S.P., Hengartner, M.O., and von Mering, C. (2012). PaxDb, a database of protein abundance averages across all three domains of life. Molecular & cellular proteomics : MCP 11, 492-500.