

Title: Site-specific generation of protein-protein conjugates using native amino acids

Authors: Marco J. Lobba¹, Christof Fellmann^{2,3,4,*}, Alan M. Marmelstein¹, Johnathan C. Maza¹,
Elijah N. Kissman¹, Stephanie A. Robinson¹, Brett T. Staahl², Cole Urnes², Rachel J. Lew³,
Casey S. Mogilevsky¹, Jennifer A. Doudna^{1,2,3,5,6,*}, Matthew B. Francis^{1,7,*}

Affiliations:

¹Department of Chemistry, University of California, Berkeley, CA 94720, USA.

²Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA.

³Gladstone Institutes, San Francisco, CA 94158, USA.

⁴Department of Cellular and Molecular Pharmacology, School of Medicine, University of California, San Francisco, CA 94158, USA.

⁵Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA.

⁶Innovative Genomics Institute, University of California, Berkeley, CA 94720, USA.

⁷Materials Sciences Division, Lawrence Berkeley National Laboratories, Berkeley, CA 94720, USA.

*Correspondence should be addressed to C.F. (christof.fellmann@gladstone.ucsf.edu), J.A.D. (doudna@berkeley.edu), or M.B.F (mfrancis@berkeley.edu).

Abstract: Chimeric protein-protein conjugates provide platforms for immunotherapy, targeted drug delivery, and vaccine development. However, many desirable constructs cannot be produced through direct expression, and the targeted coupling of two proteins is chemically challenging. Here we present a new approach for the rapid and site-specific coupling of proteins using native amino acids. Tyrosinase oxidizes exposed tyrosine residues on polypeptides, generating *ortho*-quinones that react rapidly with strategically placed cysteine residues in other proteins. This approach was used to modify CRISPR-Cas9 and other substrates with small molecules, peptides and even intact proteins. The conjugation of cell penetrating peptides to CRISPR-Cas9 was shown to increase cellular genome editing efficiency by 20-fold relative to unmodified Cas9. This technology represents a new paradigm for biomolecular coupling, and paves the way to an unprecedented range of multifunctional bioconjugates.

One Sentence Summary: Tyrosinase oxidation allows for the site-specific coupling of whole proteins through native amino acids, yielding chimeric protein complexes.

Main Text: The covalent modification of proteins while preserving native function has long been a goal of chemical biology research (1–3). Unfortunately, it can be difficult to identify chemically distinct functional groups in protein sequences, typically leading to heterogeneous product mixtures with most modification strategies. Cysteine residues present a well-recognized exception, as their reduced forms are found infrequently on protein surfaces and they can be targeted chemoselectively using 1,4-conjugate addition reactions. While very useful for the installation of spectroscopic probes and polymer chains, this chemistry can be difficult to use for the installation of highly functionalized coupling partners, such as drug molecules, peptides, and additional proteins. Moreover, it is now also appreciated that cysteine-maleimide linkages are unstable in the presence of exogenous thiols, such as glutathione, which can limit their use for *in vivo* applications (4). As a result several recent advances in cysteine coupling chemistry have appeared based on maleimide alternatives with less reversibility, and a number of complementary approaches that target artificial amino acids and N-terminal sites have been developed (4).

Protein-protein coupling is particularly challenging because uniquely reactive chemical groups must be introduced at specific locations in both biomolecular substrates, and the chemical reactions to join them must proceed at micromolar concentrations while tolerating the significant steric interactions that develop. Currently, protein-protein coupling relies on native chemical ligation methods, the insertion of recognition domains for enzymatic labeling, or the use of non-canonical amino acids (1, 5–10). What has been lacking from the field is a technique that can achieve the rapid coupling of peptide and protein substrates with high site-selectivities, but with minimal levels of substrate engineering or functionalization. Here we describe an oxidative coupling strategy involving the reaction of N- or C-terminal tyrosine residues with cysteines that can be located at any desired position on a protein surface. The reaction features mild enzymatic activation conditions, excellent levels of site-selectivity, and the ability to couple peptides and full-size proteins together with high yields in under an hour. This new approach provides unprecedented access to the building of highly complex but still well-defined bioconjugates that can be used for a wide range of vaccine, therapeutic delivery, and biotechnology applications.

Tyrosinase from *Agaricus bisporus* (abTYR) is a copper-dependent enzyme that functions to convert tyrosine into melanin via an *o*-quinone intermediate (**Fig. 1A**) (11). We and others have shown that tyrosinase activation can occur selectively at tyrosine residues near the

N- or C-termini of proteins while leaving internal phenol groups untouched (12–16). We have further shown that the *o*-quinone intermediates can be intercepted by anilines and N-terminal proline residues with high efficiency (**Fig. 1A**)(16, 17), yielding complex bioconjugates at high dilution in under 1 h (2, 18). Based on this success, we became interested in the participation of cysteine thiolates in this chemistry. Although studies by other groups have examined *o*-quinone reactions with small molecule thiols(19), these pathways have not been explored for the site-selective coupling of large biomolecules(12–16).

We first tested the oxidative coupling between phenols and cysteine residues to modify genome-free MS2 bacteriophage capsids bearing N87C mutations in each of the subunit proteins. Each of these assemblies provides 180 cysteine residues facing the inside of a 27 nm sphere with pores that are large enough for peptide entry (**Fig. 1B**)(20–24). Modification was achieved by exposing 10 μ M (based on monomer) MS2 N87C to 200 μ M α -endorphin peptide (**Fig. 1C**) in the presence of 167 nM abTYR for 30 min at room temperature in 20 mM phosphate buffer, pH 6.5. Clean conversion of each capsid monomer to a single peptide conjugate was observed (**Fig. 1D**). This represents the installation of 180 peptide groups inside each capsid despite the high degree of steric crowding. Subsequent exposure to NEM showed no further addition (**Fig. 1E**, **supplemental fig. S1**), supporting the cysteine selectivity of the reaction. Similarly, no α -endorphin modification was observed for MS2 N87C that had been previously coupled to NEM, and there was no modification of wild type MS2 capsids lacking exposed cysteine residues (**supplemental fig S1**). Together, these results demonstrate an efficient and operationally simple approach to forming covalent bonds between native amino acid residues.

The structure of the bioconjugation product was determined using NMR for a small molecule analog. The product results from thiol addition to the 5 position of the *o*-quinone ring (**supplemental fig. S2**), and both NMR and mass spectral data indicated that the product exists in the reduced catechol form (**supplemental fig. S3**). The mass of this catechol product corresponds to those observed for protein addition products. We also note that cysteines on proteins avoid the tyrosinase inactivation that occurs using small-molecule thiol substrates (12, 25–27).

To test protein conjugation product stability, we subjected α -endorphin modified MS2 to buffers ranging from pH 5 to pH 8, potassium ferricyanide, 5 mM solutions of TCEP, and n-mercaptobutanol for up to 6 days at room temperature. None of the conditions led to greater than

5% loss of the attached peptides, illustrating the high stability of these conjugates (**supplemental table S1 and supplemental fig. S4**). Additional small molecule thiols were shown to add into the *ortho*-quinone ring in small amounts (**supplemental fig. S5**), presumably after aerobic oxidation to the *o*-quinone; however, these adducts also retained the stability of the linkage formed in the reaction.

Substrate compatibility tests using the tyrosinase activation/thiol addition protocol showed that small molecule phenols, including biotin, rhodamine, and 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA), all resulted in quantitative conversion of the starting protein to the expected products. Both super-folder GFP bearing the Y182C mutation (GFP Y182C, **Fig. 2A**) and the N87C MS2 coat protein (**supplemental fig. S6**) were used in these experiments. Some secondary adducts were observed with these small molecule substrates for sfGFP, which are believed to be the products of competing N-terminal modification (12). Peptides derived from IL-13 and the SV40 large T-antigen, which can assist with cellular uptake and localization (28, 29), were also coupled successfully, as were HIV Tat and polyarginine sequences bearing tyrosine groups (**Fig. 2B**). It is important to note that N-terminal acylation or an additional glycine was needed in these cases to prevent cyclization of the oxidized N-terminal tyrosine residues. The diversity of small molecules and peptides compatible with this coupling strategy demonstrates the utility of the reaction as a general tool for protein modification.

The potential of this chemistry to modify large proteins and enzymes was explored using CRISPR-Cas9 ribonucleoprotein (RNP) complexes. Cas9, an RNA-guided DNA endonuclease used widely for genome editing, contains two surface cysteines at positions 80 and 574, thus providing two native bioconjugation handles for peptide attachment. To generate a modified form of Cas9 bearing cell-penetrating peptides, coupling reactions were conducted using Cas9 bearing two copies of the SV40 large T-antigen nuclear localization sequence (NLS) peptide. Quantitative conversion was observed to the doubly modified protein in both the apo and RNA bound forms (**Fig. 3A**). Importantly, tyrosinase-modified Cas9 RNPs fully retained their target cleavage activity when compared to unmodified Cas9 RNPs (**Fig. 3B**). We also found that tyrosinase could be used to ligate Cas9 to green fluorescent protein (MYGGS-GFP, **Fig. 3C, D**) or supercharged cationic GFP (**supplemental fig. S7**).

The reaction was similarly successful for the coupling of Y182C sfGFP (5 μ M) to a HER2-binding scFv-GGSY construct (5 μ M) in pH 6.5 phosphate buffer at room temperature for 30 min (**Fig. 3E, F**). The scFv-GFP construct bound specifically to HER2 expressing cells, but not HER2 negative cells (**Fig. 4A**). No cellular fluorescence was observed with either GFP alone or in controls where GFP was not coupled to the scFv (**supplemental fig. S8, S9**). These and corresponding control experiments showed that only highly exposed tyrosines near the N- or C-terminus are capable of participating in the reaction (30). Furthermore, these results demonstrating the utility of tyrosinase mediated conjugation for site-specific protein-protein coupling while preserving native protein activity.

Genetic fusion of four consecutive SV40 NLSs to the N-terminus of Cas9 bearing two additional C-terminal NLSs has been shown to enable delivery of Cas9 RNPs to human neural progenitor cells (NPCs) in culture and *in vivo* (28). However, the production of such complexes is lengthy, requires dedicated cloning and protein purification, and conjugation of delivery molecules is limited to the native N- and C-terminus and peptides amenable to genetic fusion. Moreover, most commercially produced Cas9 RNPs demonstrate little-to-no cell penetrating capabilities. Thus, we examined whether standard Cas9 RNP complexes could be chemically modified with peptides or proteins for editing in human NPCs(28). Three different cell-penetrating peptides were coupled to the native cysteine residues of a Cas9RNP bearing two C-terminal NLSs. This version of Cas9 does not appreciably edit cells in the NPC assay on its own (**Fig. 4B**). After the reaction, ESI-TOF MS analysis confirmed that two copies of each tested peptide coupled quantitatively to Cas9 (**supplemental fig. S10**). These constructs, plus Cas9RNP with sfGFP, were then assayed for their editing efficiencies (31–34). Cas9 conjugates bearing +36 GFP proteins precipitated upon guide RNA addition, and thus could not be evaluated in this experiment. The Cas9RNP construct bearing two 2SV40 peptides (each peptide is composed of two consecutive SV40 NLSs) showed a 20-fold increase in editing efficiency (**Fig. 4B**) relative to the unmodified 2NLS-Cas9RNP. This indicates that the delivery of Cas9 RNP can be mediated by direct peptide coupling to the protein surface of readily available standard Cas9 protein. The addition of other peptides and sfGFP did not improve editing efficiency. In addition, experiments involving the original 4NLS Cas9 construct showed successful modification, but the added protein domains did not increase editing activity further (**supplemental fig. S11**).

Here, we have shown that the abTYR activation of exposed phenols on proteins, peptides, and small molecules leads to efficient coupling with solvent exposed cysteines on target proteins. This simple but versatile approach allows for the targeted modification of proteins with a diverse array of molecules in potentially any desired location. It does not require additional modification steps or exposure to aggressive reaction conditions, and is thereby likely to conserve protein folding and function in most applications. In particular, the success of Cas9 delivery via peptide conjugation highlights a potential application for tyrosinase mediated coupling to facilitate the screening of candidate peptides for the cellular entry of large biomolecules without the need to clone, express, and purify individual genetic fusions for each combination. One of the most important aspects of this new bioconjugation approach is its ability to couple full-sized proteins together with high yields and minimal stoichiometric excesses as in the case of the scFv-GFP conjugate. It also offers the advantage of moving the attachment sites by changing the location of the introduced thiol component, which has been a longstanding feature of cysteine-based modification chemistry. The ability demonstrated herein to couple proteins together in such a straightforward way has significant implications for vaccine development, immunotherapy, and many additional biopharmaceutical and biomaterial applications.

References and Notes:

1. K. D. Brune, D. B. Leneghan, I. J. Brian, A. S. Ishizuka, M. F. Bachmann, S. J. Draper, S. Biswas, M. Howarth, Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Sci. Rep.* **6**, 19234 (2016).
2. A. C. Obermeyer, J. B. Jarman, M. B. Francis, N-Terminal Modification of Proteins with o -Aminophenols. *J. Am. Chem. Soc.* **136**, 9572–9579 (2014).
3. Q. Luo, Z. Dong, C. Hou, J. Liu, Protein-based supramolecular polymers: progress and prospect. *Chem. Commun.* **50**, 9997 (2014).
4. J. M. J. M. Ravasco, H. Faustino, A. Trindade, P. M. P. Gois, Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. - A Eur. J.* **25**, 43–59 (2019).
5. I. L. Aanei, T. Huynh, Y. Seo, M. B. Francis, Vascular Cell Adhesion Molecule-Targeted MS2 Viral Capsids for the Detection of Early-Stage Atherosclerotic Plaques. *Bioconjug. Chem.* **29**, 2526–2530 (2018).
6. S. C. Reddington, M. Howarth, Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher. *Curr. Opin. Chem. Biol.* **29**, 94–99 (2015).
7. C. S. Theile, M. D. Witte, A. E. M. Blom, L. Kundrat, H. L. Ploegh, C. P. Guimaraes, Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1800–1807 (2013).
8. T. P. Huang, K. T. Zhao, S. M. Miller, N. M. Gaudelli, B. L. Oakes, C. Fellmann, D. F. Savage, D. R. Liu, Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* **37** (2019), pp. 626–631.
9. X. Chen, J. L. Zaro, W.-C. Shen, Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).
10. H. L. Zhao, X. Q. Yao, C. Xue, Y. Wang, X. H. Xiong, Z. M. Liu, Increasing the homogeneity, stability and activity of human serum albumin and interferon- α 2b fusion protein by linker engineering. *Protein Expr. Purif.* **61**, 73–77 (2008).
11. G. Faccio, K. Kruus, M. Saloheimo, L. Thöny-Meyer, Bacterial tyrosinases and their applications. *Process Biochem.* **47**, 1749–1760 (2012).
12. J. C. Maza, D. L. V. Bader, L. Xiao, A. M. Marmelstein, D. D. Brauer, A. M. ElSohly, M. J. Smith, S. W. Krska, C. A. Parish, M. B. Francis, Enzymatic Modification of N-Terminal Proline Residues Using Phenol Derivatives. *J. Am. Chem. Soc.* **141**, 3885–3892 (2019).
13. E. Montanari, A. Gennari, M. Pelliccia, L. Manzi, R. Donno, N. J. Oldham, A. MacDonald, N. Tirelli, Tyrosinase-Mediated Bioconjugation. A Versatile Approach to Chimeric Macromolecules. *Bioconjug. Chem.* **29**, 2550–2560 (2018).
14. R. Lantto, E. Puolanne, K. Kruus, J. Buchert, K. Autio, Tyrosinase-Aided Protein Cross-Linking: Effects on Gel Formation of Chicken Breast Myofibrils and Texture and Water-Holding of Chicken Breast Meat Homogenate Gels. *J. Agric. Food Chem.* **55**, 1248–1255 (2007).
15. M. Fairhead, L. Thöny-Meyer, Bacterial tyrosinases: old enzymes with new relevance to

- biotechnology. *N. Biotechnol.* **29**, 183–191 (2012).
16. E. J. Land, C. A. Ramsden, P. A. Riley, Quinone Chemistry and Melanogenesis. *Methods Enzymol.* **378**, 88–109 (2004).
 17. D. G. Anderson, S. V. S. Mariappan, G. R. Buettner, J. A. Doorn, Oxidation of 3,4-Dihydroxyphenylacetaldehyde, a Toxic Dopaminergic Metabolite, to a Semiquinone Radical and an ortho -Quinone. *J. Biol. Chem.* **286**, 26978–26986 (2011).
 18. C. R. Behrens, J. M. Hooker, A. C. Obermeyer, D. W. Romanini, E. M. Katz, M. B. Francis, Rapid chemoselective bioconjugation through oxidative coupling of anilines and aminophenols. *J. Am. Chem. Soc.* **133**, 16398–16401 (2011).
 19. T. H. Huang, T. Kuwana, A. Warsinke, Analysis of thiols with tyrosinase-modified carbon paste electrodes based on blocking of substrate recycling. *Biosens. Bioelectron.* **17**, 1107–1113 (2002).
 20. F. A. Galaway, P. G. Stockley, MS2 viruslike particles: A robust, semisynthetic targeted drug delivery platform. *Mol. Pharm.* **10**, 59–68 (2013).
 21. I. L. Aanei, A. M. Elsohly, M. E. Farkas, C. Netirojjanakul, M. Regan, S. Taylor Murphy, J. P. O’Neil, Y. Seo, M. B. Francis, Biodistribution of antibody-MS2 viral capsid conjugates in breast cancer models. *Mol. Pharm.* **13**, 3764–3772 (2016).
 22. M. E. Farkas, I. L. Aanei, C. R. Behrens, G. J. Tong, S. T. Murphy, J. P. O’Neil, M. B. Francis, PET Imaging and Biodistribution of Chemically Modified Bacteriophage MS2. *Mol. Pharm.* **10**, 69–76 (2013).
 23. Z. M. Carrico, D. W. Romanini, R. A. Mehl, M. B. Francis, Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* **0**, 1205 (2008).
 24. S. L. Capehart, M. P. Coyle, J. E. Glasgow, M. B. Francis, Controlled integration of gold nanoparticles and organic fluorophores using synthetically modified MS2 viral capsids. *J. Am. Chem. Soc.* **135**, 3011–3016 (2013).
 25. G. Agrup, C. Hansson, H. Rorsman, E. Rosengren, The effect of cysteine on oxidation of tyrosine dopa, and cysteinyl dopas. *Arch. Dermatol. Res.* **272**, 103–115 (1981).
 26. A. M. Elsohly, M. B. Francis, Development of Oxidative Coupling Strategies for Site-Selective Protein Modification. *Acc. Chem. Res.* **48**, 1971–1978 (2015).
 27. V. S. Murty, T. M. Penning, Polycyclic aromatic hydrocarbon (PAH) ortho-quinone conjugate chemistry: Kinetics of thiol addition to PAH ortho-quinones and structures of thioether adducts of naphthalene-1,2-dione. *Chem. Biol. Interact.* **84**, 169–188 (1992).
 28. B. T. Staahl, M. Benekareddy, C. Coulon-Bainier, A. A. Banfal, S. N. Floor, J. K. Sabo, C. Urnes, G. A. Munares, A. Ghosh, J. A. Doudna, Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.* **35**, 431–434 (2017).
 29. H. Pandya, D. M. Gibo, S. Garg, S. Kridel, W. Debinski, An interleukin 13 receptor α 2-specific peptide homes to human Glioblastoma multiforme xenografts. *Neuro. Oncol.* **14**, 6–18 (2012).

30. A. M. Marmelstein, M. J. Lobba, C. S. Mogilevsky, J. C. Maza, D. D. Brauer, M. B. Francis, Tyrosinase-Mediated Oxidative Coupling of Peptide and Protein Tyrosine Tags.
31. J. K. Allen, D. J. Brock, H. M. Kondow-McConaghy, J.-P. Pellois, J. K. Allen, D. J. Brock, H. M. Kondow-McConaghy, J.-P. Pellois, Efficient Delivery of Macromolecules into Human Cells by Improving the Endosomal Escape Activity of Cell-Penetrating Peptides: Lessons Learned from dfTAT and its Analogs. *Biomolecules*. **8**, 50 (2018).
32. J. J. Cronican, D. B. Thompson, K. T. Beier, B. R. McNaughton, C. L. Cepko, D. R. Liu, Potent Delivery of Functional Proteins into Mammalian Cells in Vitro and in Vivo Using a Supercharged Protein. *ACS Chem. Biol.* **5**, 747–752 (2010).
33. P. Lönn, A. D. Kacsinta, X.-S. Cui, A. S. Hamil, M. Kaulich, K. Gogoi, S. F. Dowdy, Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.* **6**, 32301 (2016).
34. S. Futaki, T. Suzuki, W. Ohashi, T. Yagami, S. Tanaka, K. Ueda, Y. Sugiura, Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J. Biol. Chem.* **276**, 5836–40 (2001).
35. M. FLING, N. H. HOROWITZ, S. F. HEINEMANN, The isolation and properties of crystalline tyrosinase from *Neurospora*. *J. Biol. Chem.* **238**, 2045–53 (1963).
36. C. Anders, M. Jinek, in *Methods in enzymology* (Elsevier Inc., ed. 1, 2014), vol. 546, pp. 1–20.
37. G. J. Tong, S. C. Hsiao, Z. M. Carrico, M. B. Francis, Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* **131**, 11174–11178 (2009).
38. C. Engler, R. Kandzia, S. Marillonnet, A one pot, one step, precision cloning method with high throughput capability. *PLoS One*. **3**, e3647 (2008).

Acknowledgments: The authors wish to thank Jamie Gleason for the kind gift of the plasmid containing GFP Y182C. The authors wish to thank Dr. Jeffrey Pelton for help with 900 MHz NMR data collection and analysis. **Funding:** MBF was supported by the NSF (CHE-1808189). Funds for the 900 MHz NMR spectrometer were provided by the NIH through grant GM68933. CF was supported by a US NIH Pathway to Independence Award (K99GM118909, R00GM118909) from the National Institute of General Medical Sciences (NIGMS). JAD receives funding from the William M. Keck Foundation, a Collaborative MS Research Center Award from the National Multiple Sclerosis Society, the Centers for Excellence in Genomic Science of the National Institutes of Health under award number RM1HG009490, the Somatic Cell Genome Editing Program of the Common Fund of the National Institutes of Health under award number U01AI142817-02, and the National Science Foundation under award number 1817593. JAD is an Investigator of the Howard Hughes Medical Institute and is a Paul Allen Distinguished Investigator. **Author Contributions:** MJL developed the project idea, conducted coupling reactions, ran NMR experiments, and prepared the manuscript and figures. CF ran cell culture assays with the help of RJL, assisted with writing, assisted with flow cytometry data analysis, and aided in experimental design. AMM synthesized and purified the small molecule product, assisted in the structural characterization, provided Trastuzumab scFV samples, and ran

scFv-GFP cell coupling studies. ENK ran MS2 couplings, stability studies, and NEM capping experiments. JCM contributed to experimental design, provided tyrosinase stocks, and assisted in data analysis. CM assisted in reagent preparation and conducted stability studies. SAR assisted in Cas9 activity assays and the initial Cas9 coupling reactions. BTS provided initial samples of 4NLS Cas9 protein for NPC assays and established the NPC line and assay protocols. CJU ran NPC assays. JAD provided funding support, assisted in experimental design, and contributed to the text and figures. MBF provided funding support, assisted in data analysis, assisted with experimental design, and contributed to the text and figures. **Competing Interests:** J.A.D. is an Investigator of the Howard Hughes Medical Institute and executive director of the Innovative Genomics Institute at the University of California, Berkeley and the University of California, San Francisco. J.A.D. is a co-founder of Editas Medicine, Intellia Therapeutics, and Caribou Biosciences and a scientific adviser to Caribou, Intellia, eFFECTOR Therapeutics and Driver. The Regents of the University of California have patents pending for CRISPR technologies and the use of tyrosinase in the abovementioned manner on which the authors are inventors. **Data and materials availability:** All data is available in the main text or the supplementary materials.

Supplementary Materials:

Materials and Methods

Figures S1-S11

Tables S1

References (2, 12, 28, 30, 32, 35–38)

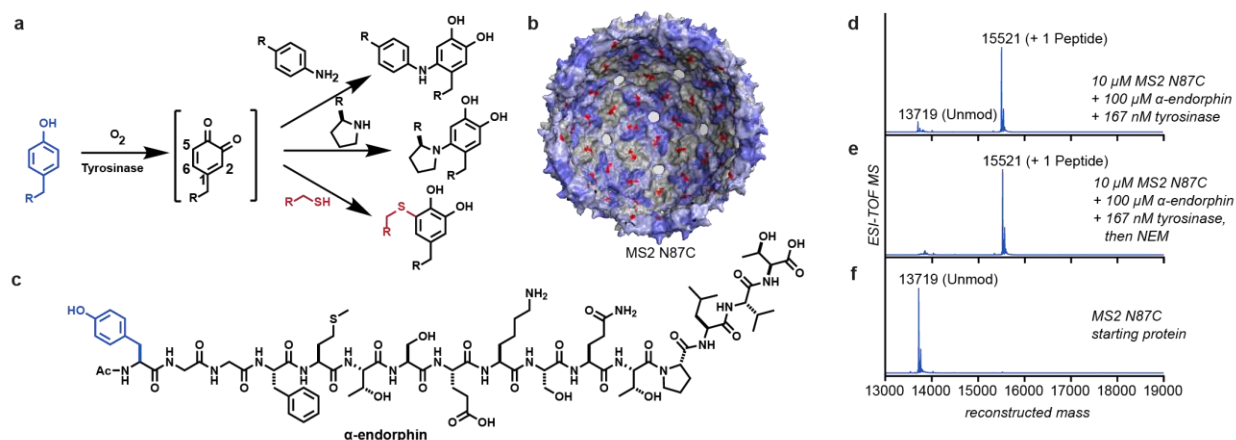


Fig. 1. Tyrosinase-mediated oxidative coupling reactions. (a) Phenols are oxidized by tyrosinase to yield *o*-quinone intermediates that couple with a variety of nucleophiles on biomolecules. In this work, the addition of cysteine thiolates is explored. (b) An MS2 viral capsid is shown, with 180 interior cysteine residues (N87C) indicated in red. PDB ID: 2MS2. Pores in the capsid shell allow entry to peptides and small molecules. (c) The structure of α -endorphin is shown, with the targeted tyrosine residue in blue. (d-f) Coupling reactions were screened using ESI-TOF MS. Reaction conditions: pH 6.5 phosphate buffer, RT, 30 min. Expected mass values: MS2 N87C $[M+H]^+ = 13719$; MS2 N87C–N-ethylmaleimide (NEM) adduct $[M+H]^+ = 13844$; MS2 N87C–endorphin quinone product $[M+H]^+ = 15521$.

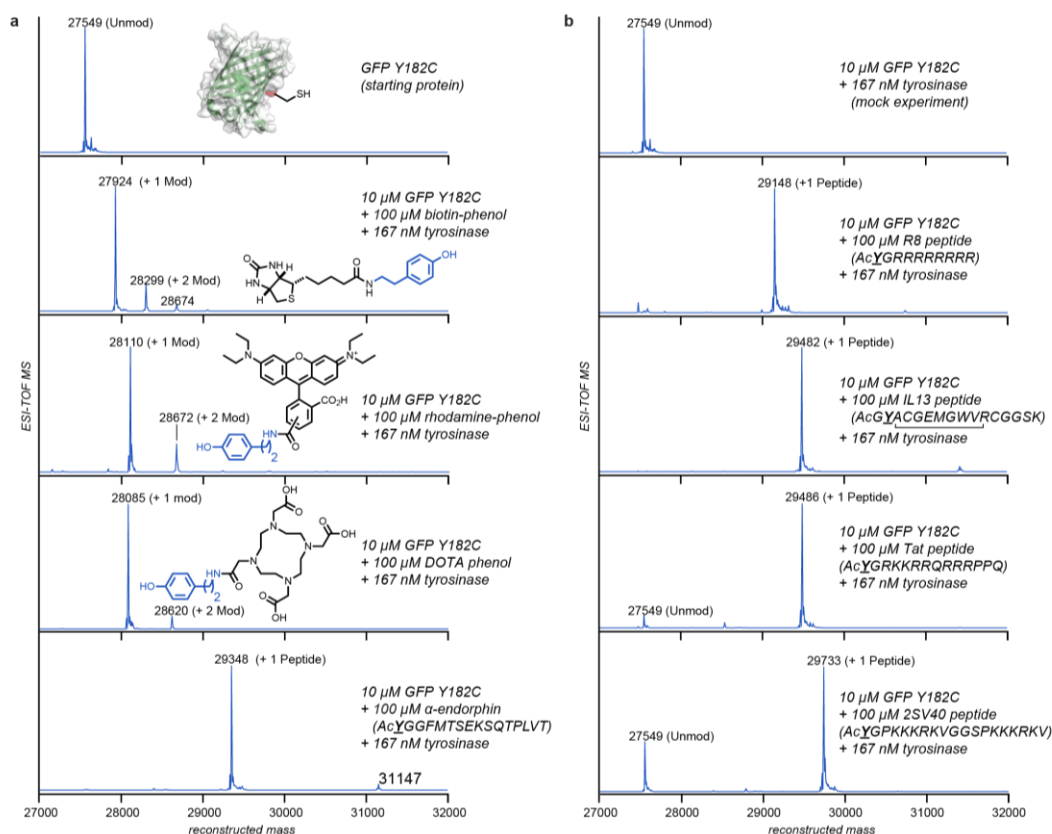


Fig. 2. Substrate scope for thiol-directed enzymatic oxidative coupling reactions (a,b). A superfolder GFP thiol mutant (Y182C) was used as the protein component. All reactions were characterized using ESI-TOF MS.

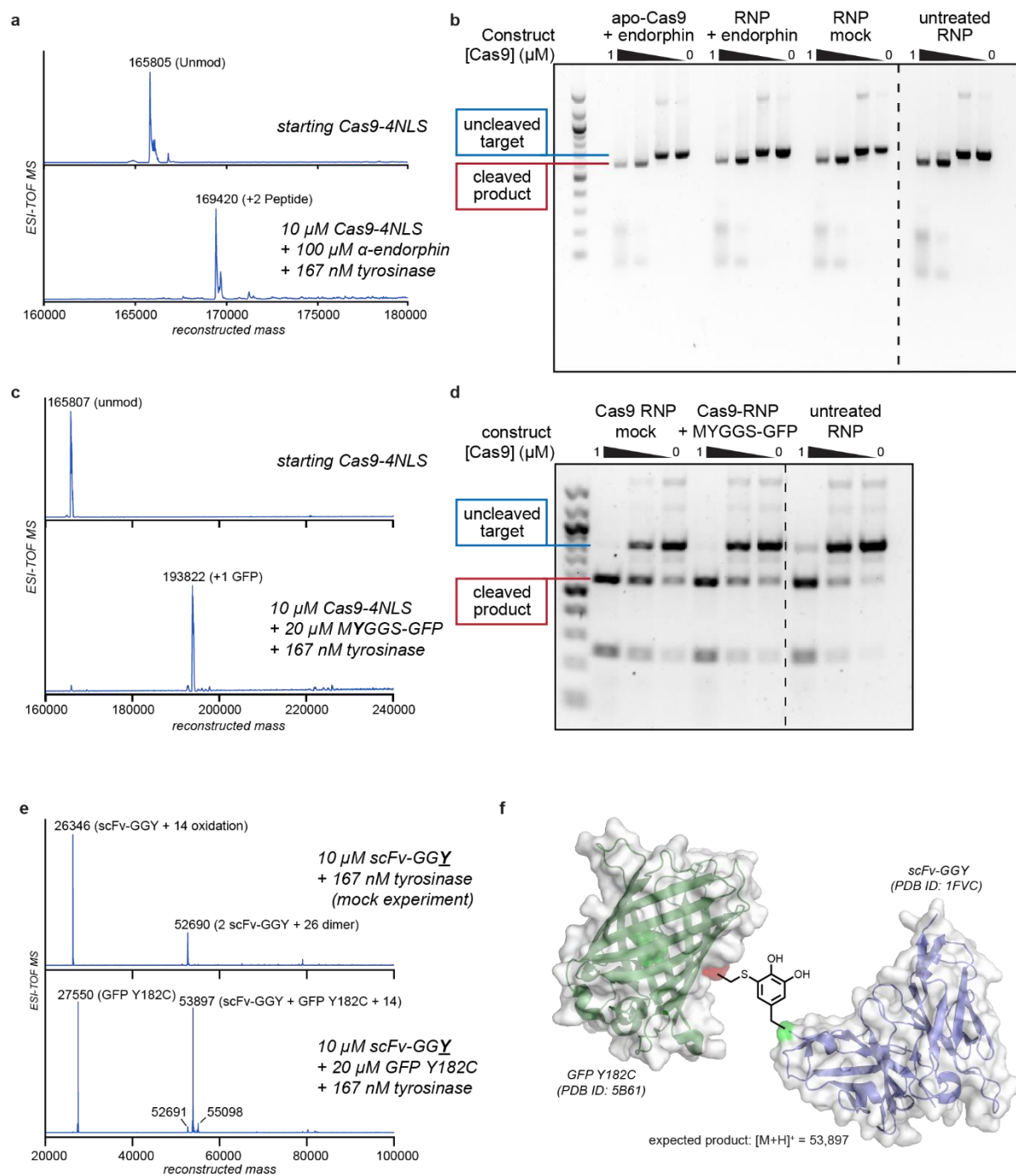


Fig. 3. Modification of CRISPR-Cas9 with two C-terminal and four N-terminal nuclear localization sequences (Cas9-4NLS), which has two surface thiols (cys 80 and cys 574) in the native sequence. (a) ESI-TOF MS data indicated complete coupling at both sites with the α -endorphin peptide. In contrast, cas9 that was pretreated with NEM showed no peptide coupling (see Supporting Information). (b) The site-specific DNA cleaving ability of the conjugate was unchanged relative to an untreated control. This was true when the pre-assembled ribonucleoprotein (RNP) was modified directly, or when the RNA-free protein (apo-cas9) was used for the modification reaction. In the latter case, the single guide RNA strand was

before the DNA cleavage experiment. (c) The coupling reaction was successful even when the coupling partner was superfolder GFP bearing an exposed tyrosine residue near the N-terminus (MYGGS-GFP). (d) This conjugate retained site-specific DNA cleaving ability despite the high degree of added steric bulk. (e) The coupling reaction was similarly successful for the coupling of GFP Y182C to a HER2-binding scFv-GGY construct. The observed product mass is consistent with the linkage depicted in (f).

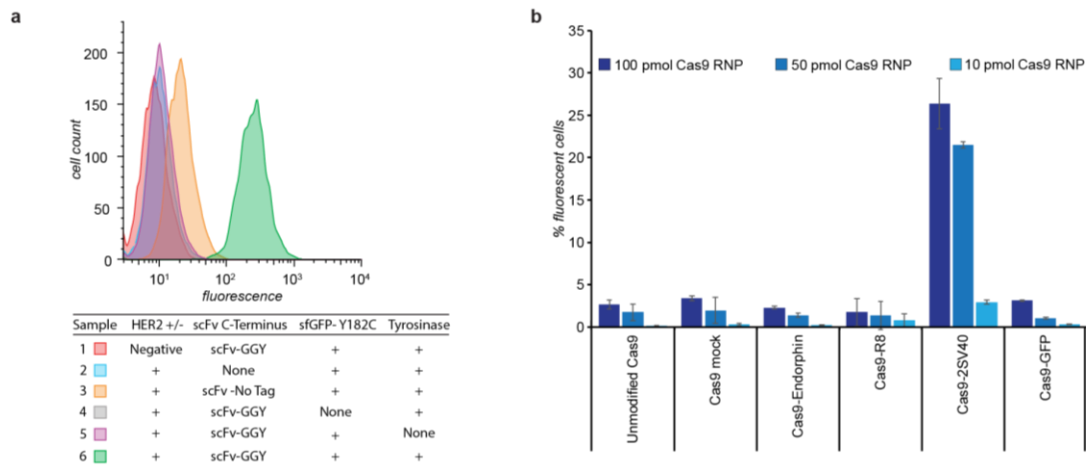


Fig. 4. Cell culture activity assays. a) Flow cytometry analysis of SK-BR-3 (HER2+) and MDA-MB-468 (HER2-) breast cancer cells treated with the Trastuzumab scFv-sfGFP construct or negative control showing cell specific binding. Gating and statistics shown in Supplemental Figure S9. b) Evaluation of modified CRISPR-Cas9 ribonucleoproteins (RNP) for gene editing in neural progenitor cells. The assay measured the increase in fluorescent protein expression following successful editing at a tdTomato locus in the cell genome. Cas9 modified with two copies of the SV40 nuclear localization sequence (Cas9-2SV40) showed a 20-fold increase in editing efficiency compared to unmodified Cas9. Data were collected at 72 h post-treatment.

Supplementary Materials for

Site-specific generation of protein-protein conjugates using native amino acids

Marco J. Lobba¹, Christof Fellmann^{2,3,4,*}, Alan M. Marmelstein¹, Johnathan C. Maza¹, Elijah N. Kissman¹, Stephanie A. Robinson¹, Brett T. Staahl², Cole Urnes², Rachel J. Lew³, Casey S. Mogilevsky¹, Jennifer A. Doudna^{1,2,3,5,6,*}, Matthew B. Francis^{1,7,*}

Correspondence should be addressed to C.F. (christof.fellmann@gladstone.ucsf.edu), J.A.D. (doudna@berkeley.edu), or M.B.F (mfrancis@berkeley.edu).

This PDF file includes:

Materials and Methods
Figs. S1 to S11
Table S1

Materials and Methods

General Methods and Instrumentation

All reagents were obtained from commercial sources and used without any further purification. Tyrosinase isolated from *Agaricus bisporus* (abTYR, both 25 kU [SKU = T3824-25KU] and 50 kU [SKU = T3824-50KU] were used in these studies). DOTA-NHS was purchased from Macrocyclics (Plano, Tx). Tyrosine containing peptides were purchased from GeneScript (Piscataway, NJ) and used without further purification. Spin concentrators with 10 and 100 kDa molecular weight cutoffs (MWCO) and sterile spin filters with 0.22 μ m pores were purchased from Millipore (Billerica, MA). Milli-Q was obtained from a Millipore purification system. Small molecule phenol substrates were synthesized as previously reported(12).

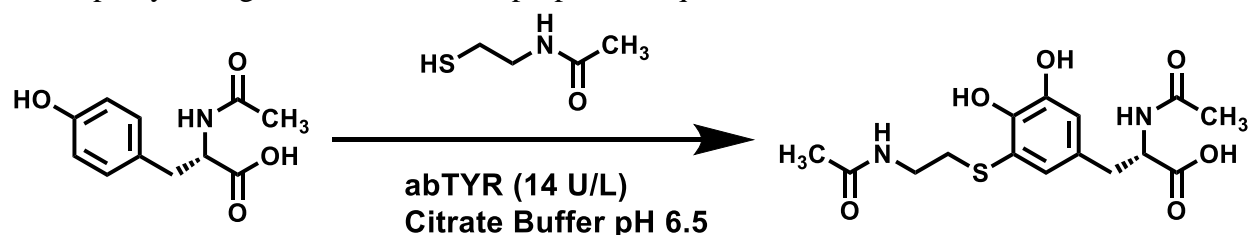
Liquid chromatography mass spectrometry (LC-MS) analysis. Acetonitrile (Optima grade, 99.9%, Fisher, Waltham, MA), formic acid (1 mL ampules, 99+%, Pierce, Rockford, IL), and water purified to a resistivity of 18.2 M Ω ·cm (at 25 °C) using a Milli-Q Gradient ultrapure water purification system (Millipore, Billerica, MA) were used to prepare mobile phase solvents for LC-MS. Electrospray ionization mass spectrometry (ESI-MS) of protein bioconjugates was performed using an Agilent 1260 series liquid chromatograph outfitted with an Agilent 6224 time-of-flight (TOF) LC-MS system (Santa Clara, CA). The LC was equipped with a Proswift RP-4H (monolithic phenyl, 1.0 mm \times 50 mm, Dionex) analytical column. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). For each sample, approximately 15 to 30 picomoles of analyte were injected onto the column. Following sample injection, a 5-100% B elution gradient was run at a flow rate of 0.30 mL/min over 8 min. Data was collected and analyzed by deconvolution of the entire elution profile in order to provide reconstructed mass spectra that are representative of the entire sample using Agilent MassHunter Qualitative Analysis B.05.00. Percent modification was determined through integration of MS peaks using opensource Chartograph software (www.chartograph.com). The integration of the completely unmodified protein peak served as an internal standard in determining the percent modification.

UV-VIS measurements. A NanoDrop 1000 (Thermo Fisher Scientific) was used to quantify the samples in this work based on absorbance values at 280 nm (or 488 nm for sfGFP).

Experimental Procedures

Preparation and storage of abTYR stock solutions. Tyrosinase isolated from common button mushroom (*Agaricus bisporus*) was purchased as a lyophilized dark-red powder from Sigma-Aldrich (either 25 kU or 50 kU). Upon receipt, the powder was portioned as preweighed dry samples in Eppendorf tubes. These were stored at -20 °C until use. Stock solutions were prepared by dissolving a dry portion of enzyme to yield a concentration of 2 mg/mL in 50 mM phosphate buffer at pH 6.5 (corresponding to ~16.7 μ M enzyme). Portions of this solution (10 μ L each) were transferred to 250 μ L Eppendorf tubes, which were then capped and stored at -80 °C until use. These storage conditions were shown to be the optimal for abTYR in a previous study (2). Stock solutions were used as single-use aliquots and diluted to desired working concentrations in 50 mM phosphate buffer at pH 6.5 prior to use.

Preparation and structural characterization of tyrosine-thiol conjugates. Stocks of 100 mM N-acetyltyrosine and acylcysteamine as well as 2 μ M abTYR were prepared in 100 mM citrate pH 6.5. The base of a 15 mL 3k MWCO spin concentrator was pre-loaded with 5 mL of 100 mM citrate and 10 mM acylcysteamine prepared by dilution of stock solutions. In the top section of the filter 1 mL of 10 mM N-acetyltyrosine was combined with 1 mL of 2 μ M Tyrosinase in 100 mM citrate at pH 6.5. The solution was incubated at RT for 30 min before centrifugation at 3000 rpm for 25 min. After this step, an additional 1 mL of 10 mM N-acetyl-L-tyrosine in 100 mM citrate at pH 6.5 was added and allowed to sit for a further 20 min on ice before centrifugation as above. This process was repeated an additional three times for a total added volume of 10 mM n-acetyltyrosine of 4 mL. After the final centrifugation the system was spun an additional 25 min at 4000 rpm, yielding a total of 11 mL of purple-red liquid.



Citrate buff. (pH 6.5)			N-Ac-Tyr			N-Ac-Cysteamine			A. bisporus Tyrosinase			H2O	Total Vol
i (mM)	f (mM)	vol (μ L)	i (mM)	f (mM)	vol (μ L)	i (mM)	f (mM)	vol (μ L)	i (U/L)	f (U/L)	vol (μ L)	(μ L)	(μ L)
100	20	3000.0	100.0	5.00	750.0	100.0	5.0	750.0	982	14.0	213.8	10286	15000

A secondary protocol was also used which generated lower yields of product. Here, to a 10.286 mL aliquot of Milli-Q H₂O in a falcon tube was added stock solutions of sodium citrate buffer (pH 6.5), N-acetyl-L-tyrosine, and N-acetylcysteamine with initial (stock solution) concentrations (i) and final (reaction mixture) concentrations (f) given in the table above. The abTYR stock solution was added last. After gentle mixing, the resulting reaction solution was allowed to stand at room temperature for 2 hours and then was stored at 4 °C for one week while awaiting purification. After passing the rxn solution through a 0.22 μ m Millipore polyethersulfone (PES) syringe filter, the crude reaction mixture was subject to analytical and preparatory HPLC.

Analysis and purification of the crude small molecule oxidative coupling product were performed on a Waters UPLC-MS system with the following modules: Waters 2545 Binary Gradient Module pumps, a Waters 2767 Sample Manager for sample injection and fraction collection, a Waters System Fluidics Organizer, a Waters 2489 UV/Visible Detector, and a Waters Acquity Ultra Performance LC mass spectrometer with SQ detector. Waters Mass Lynx software V4.1 (SCN919) was used to operate the instrument. For gradient separations, solvent A was: Milli-Q H₂O with 0.1% TFA, and solvent B was: HPLC-grade acetonitrile with 0.1% TFA. This instrument is owned and maintained by the LBNL/UCB Catalysis Group Instrumentation Facility and is supported by DOE Basic Energy Sciences program KC0302010.

For analytical HPLC, a Waters X-Bridge C18, 3.5 μ m, 4.6 x 100 mm column was used. 20 μ L injections of the filtered reaction mixture were subject to a linear gradient of 5 to 20% solvent B in solvent A over 15 minutes at a flow rate of 1.00 mL/min.

For preparatory HPLC, a Waters X-Bridge C18, 10 μ m OBD, 4.6 x 250 mm column was used. 1000 μ L injections of the filtered reaction mixture were subject to a linear gradient of 5 to 20% solvent B in solvent A over 15 minutes at a flow rate of 30.0 mL/min.

The product isolate then underwent evaporative removal of solvent. This yielded several flakes of red-brown powder in quantities too low to be effectively weighed. The dried product was resuspended in D₂O and submitted to NMR analysis on a Bruker 900 MHz system.

abTYR-mediated oxidative coupling of phenols to Y182C sfGFP. To a 10 μ M protein solution in 20 mM phosphate buffer (pH = 6.5) was added the phenol coupling partner to a final concentration of 100 μ M. A 2 μ M solution of abTYR in 50 mM phosphate buffer (pH=6.5) was added to reach a final enzyme concentration of 200 nM. The reaction was left unstirred for 30 min at RT. The reactions were subsequently quenched through the addition of tropolone (final concentrations of 1.9 mM), and allowed to stand at RT for 1 min before spin concentration against 10 kDa MWCO filters with 10 mM bisTris at pH 7.2 (three cycles at 13,000 rpm for 3 min) for small molecule removal. The resulting samples were analyzed using ESI-TOF MS.

Cysteine modification of N87C MS2. To a 10 μ M solution of protein was added the phenol- or tyrosine-containing coupling partner (100 μ M) in 20 mM phosphate buffer (pH = 6.5). To this was then added abTYR to a final concentration of 200 nM. After 30 min the reaction was quenched with tropolone to a final concentration of 2 mM and the product was analyzed by ESI-MS (see **Figure 1**).

Stability study of cysteine modification products. To assess the stability of the oxidative coupling products, a solution of endorphin modified MS2 N87C (**Pdt**) was incubated in a variety of conditions. To assess the impact of pH on the linkage stability, a 10 μ M solution of **Pdt** with 1 mM tropolone (to prevent further oxidation via tyrosinase) was incubated in 50 mM phosphate buffer at pH 5.0, 7.0 and 8.0 at RT for 0, 3, or 6 days. At each time point a 10 μ L sample of the protein was diluted 1:1 with 50 mM pH 7.0 bisTris buffer and spin concentrated against a 10 kDa MWCO filter. The resulting samples were analyzed using ESI-TOF MS and the percent labeling was determined by peak integration using the opensource Chartograph software (www.chartograph.com). To assess the impact of nucleophiles on the linkage stability, a 10 μ M solution of **Pdt** with 1 mM tropolone was incubated in 50 mM phosphate buffer at pH 7.4 at RT for 0, 3, and 6 days in the presence of 10 mM concentrations of various nucleophiles. At each time point, the proteins were spin concentrated against 10 kDa MWCO filters with 10 mM bisTris at pH 7.2 for three cycles at 13,000 rpm for 3 min each. The resulting samples were analyzed using ESI-TOF MS and the percent labeling was determined via peak integration using the opensource Chartograph software. Note that starting modifications differ as each condition was treated as a biological replicate.

Quantitative tyrosinase (abTYR) activity assay. The following procedure was previously reported in Maza et. al (12). In brief, a 1.0 mg/mL aliquot of abTYR was removed from storage at either -20 °C or -80 °C and thawed on ice. To a clean cuvette was added 200 μ L of 100 mM phosphate buffer (pH 6.5), 93 μ L of 2.7 mM L-tyrosine, and 695 μ L of Milli-Q water. A 12 μ L portion of the abTYR aliquot was added with thorough mixing to start the reaction (final concentrations: 20 mM phosphate, 250 μ M L-tyrosine, and 0.012 mg/mL abTYR). The reaction was immediately transferred to a UV-Vis spectrometer and readings of the absorption at 475 nm

were recorded at 1 s intervals for at least 10 min. For data processing, the absorbance at 475 nm was plotted against time in minutes using Microsoft Excel. The instantaneous rate (Abs/min) at a given time was approximated by calculating the linear slope over 30 s intervals around each timepoint between 0:16 and 9:45. The maximum slope typically occurred between 5 and 7 min, and was converted to activity units (U or μmol substrate / min) using the molar extinction coefficient for dopachrome of $3600 \text{ M}^{-1}\text{cm}^{-1}$ (35). Dividing the U by the volume of the tyrosinase stock added to the reaction gave the U/mL of the stock. Freshly prepared 1 mg/mL tyrosinase typically displayed an activity of 1200 U/mL.

Expression of sfGFP. Expression was carried out as previously reported (2). Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of spCas9. Expression was carried out as previously reported (36). Purified protein was spin concentrated into 20 mM phosphate buffer with 200 mM trehalose and 300 mM NaCl at pH 7 using 100k MWCO spin filters prior to reaction. Cas9 sequences for 2NLS and 4NLS were identical to those used previously (28).

Expression of N87C MS2. Expression and purification was carried out as previously reported (37). Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

N87C MS2 amino acid sequence:

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQN
RKYTIKVEVPKVATQTVGGVELPVAAWRSYLCMELTIPIFATNSDCELIVKAMQGLL
KDGNIPIPSAIAANSKIY*

Expression of Trastuzumab scFv. Expression was carried out according to established protocols (36). Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of Y182C sfGFP and N-Terminal MYGGS_{H6}-sfGFP. Following sequence verification by sequencing, mutants were transformed into Rosetta DE3 *E. coli* cells for expression. Expression cultures at 1000 mL in TB broth were prepared using 10 mL of confluent culture. These were grown at 37 °C until cultures reached an OD₆₀₀ value between 0.6-0.8, upon which protein expression was induced with a final concentration of 0.1 g/L IPTG.

Cells were then shaken overnight at 37 °C and pelleted. After freezing at -80 °C, cell pellets were resuspended in 15 mL of an equilibration buffer (20 mM sodium phosphate, 2 M NaCl, 20 mM imidazole at pH = 7.4) and then lysed via sonication for 30 min at 60% amplitude. Cell lysate was spun down at 14,000 rpm for 30 min, then the supernatant decanted before loading on to a HisTrap FF Crude column. After binding, the bound protein was washed with 4 portions of two resin bed volumes of wash buffer (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole at pH = 7.4) and subsequently eluted with four resin bed volumes of elution buffer (20 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole at pH = 7.4). The purified sfGFP variant was then

spun concentrated into 20 mM phosphate buffer at pH 7.2 using 10k MWCO filters. Purified sfGFP samples were flash frozen and stored at -20 °C until use.

Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

MYGGS_{H6}-sfGFP amino acid sequence:

MYGGSHHHHHRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLT
LKFICTTGKLPVPWPTLVTTLTYGVCQFARYPDHMKQHDFFKSAMPEGYVQERTISFKD
DGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNHNVYITADKQKN
GIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRD
HMLLEFVTAAGITHGMDELYK*

Y182C sfGFP amino acid sequence:

MRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLT
LKFICTTGKLPVPWPTLVTTLTYGVCQFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGT
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNHNVYITADKQKN
GIKANFKIRHNVEDGSVQLADHCCQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRD
HMLLEFVTAAGITHGMDELYKHHHHH*

Expression of +36 MYGGS sfGFP. The +36 GFP gene(37) with addition of MYGGS-H₆ at the N-terminus was purchased as a gene block from Integrated DNA Technologies (IDT) as follows:

MYGGS-H6	+36	GFP:
TGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGT	TAACTTTAAGA	
AGGAGATATACCATGTACGGAGGCTCACACCATCACC	ATCACCATGGT	
GGAGCTTCAAAGGGTGAGCGTCTTTTTCGCGGCAA	AGTCCCTATCCTGG	
TTGAGTTGAAGGGGGATGTCAACGGACATAAATTCT	CAGTGCGCGGTA	
AAGGGAAAGGCGATGCAACCCGTGGCAA	ACTGACTTTGAAATTTATTT	
GCACAACTGGCAAATTACCTGTTCCCTGGCCCACCT	AGTAACGACACT	
GACTTATGGAGTACAGTGTTCAGTCGTTACCCCAA	ACACATGAAACGT	
CATGACTTCTTTAAAAGCGCAATGCCCAAGGGGT	ACGTGCAAGAACGT	
ACCATCTCTTTTAAAGAAAGACGGGAAGTACAAA	ACTCGTGCGGAAGTT	
AAGTTTGAAGGCCGTACGCTTGTCAACCGCATTA	AAGCTTAAAGGTCGC	
GACTTTAAAGAAAAGGGGAACATTCTGGGGCACA	AGTTGCGCTACAAT	
TTCAACAGCCACAAAGTGTATATCACAGCGGATA	AAGCGCAAGAATGGA	
ATTAAAGCTAAGTTCAAGATTCGTCATAACGTAA	AAGGACGGCAGCGTG	
CAGCTGGCGGATCATTACCAACAGAATACACCTAT	CGGACGTGGGCCT	
GTATTATTGCCTCGCAACCATTACCTGTCAACAC	CGCTCAAACTTTCCA	
AAGATCCGAAGGAGAAACGTGATCACATGGTACT	TTTGGAGTTTGTTA	
CCGCAGCTGGTATTAAGCACGGGCGCGATGAACG	TTACAAATAACTCG	
AGCACCACCACCACCACCTG		

Which corresponds to the following amino acid sequence after insertion into vector:

MYGGSHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGGK
GDATRGKLT
LKFICTTGKLPVPWPTLVTTLTYGVCQFSRYPKHMKRHDFFK
SAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTL
VNRIKLKGRDFKEKGN
ILGHKLRYNFNHSHKVYITADKRKNGIKAKFKIRHN
VKDGSVQLADHYQQN

TPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDER
YK*

A pET28b vector that was cleaved by XbaI and XhoI and column purified was then combined with the +36GFP gene block and cloned using Gibson 2x Master Mix from New England BioLabs (NEB) using standard procedure (38).

Following sequence verification, plasmids were transformed into Rosetta DE3 *E. coli* cells via heat shock and streaked onto LB-agar plates with 1 mM kanamycin. Expression and purification were carried out according to previous literature (38).

Cas9-tyrosinase coupling conditions. Cas9 coupling occurred under the following conditions: 20 mM Tris HCl, 300 mM KCl, 200 mM trehalose pH 7.0 (Buffer A), 4 °C, 1 h, 10 μ M Cas9. All samples were quenched with the tropolone solution as above, then solvent exchanged into Buffer A three times using a 100,000 kDa MWCO spin concentrator. For peptide coupling, peptides were added in five-fold excess, yielding 50 μ M peptide concentrations. In protein-protein coupling we found that a 1:1 ratio of Cas9:target protein yielded near-quantitative conversion to the singly modified Cas9 after filtration, as characterized using ESI-TOF MS. A 1:5 ratio of Cas9:target yielded full conversion to the doubly modified product. It is of particular interest that the trehalose buffer is tolerated as previous oxidative strategies were incompatible with saccharides in solution (35). Additionally, we have found that the internal tyrosine residues within both of these proteins are not affected by this reaction, as revealed by a lack of detectable oxidation for either Cas9 or a sfGFP that lacks the N-terminal tyrosine residue by ESI-TOF MS. Further supporting this is the lack of dimer formation between Cas9 molecules or peptide substrates in these reactions. Thus, it appears that only highly exposed tyrosines, ideally near the N- or C-termini are capable of participating in reactions (30).

Cas9 in vitro cleavage assay. Two gene fragments were produced via PCR for use as target sequences (Cas9 recognition sites shown underlined in red):

Cas9 cleavage sequence 1: 5’-

TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTA
CGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCCGTGCCC
TGGCCCAACCCTCGTGACCACCCTGACCTACGGCGTGCAAGTGCTTCAGCCGCTA
CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGC
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGA
AGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACG
GCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
CTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCA
ACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGAT
CACTCTCGGCATGGACGA-3’

Cas9 Cleavage sequence 2: 5’-

TGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTC

GAGGTTAAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTTCCTT
 TGAAAAACACGATGATAATATGGCCACAACCATGGTGAGCAAGGGCGAGGA
 GCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC
 GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGC
AAGCTGACCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGC
 CCACCCTCGTGACCACCCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCC
 GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACG
 TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
 CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGG
 CATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAA
 CTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATC
 AAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCG
 CCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
 CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
 AAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGATCACTC
 TCGGCATGGACGA-3'

This was done using a plasmid containing EGFP and the following PCR primers:

PCR forward primer 1: 5'-TAAACGGCCACAAGTTCAGC-3'

PCR forward primer 2: 5'-TGTATGGGATCTGATCTGGGG-3'

PCR reverse primer: 5'-TCGTCCATGCCGAGAGTGATCC-3'

Cas9 was combined with sgRNA (targeting sequence 5'-CAGGGTCAGCTTGCCGTAGG-3') produced by *in vitro* transcription as previously published to produce the active RNP(30). A 0.4 pmol sample of cleavage sequence 1 or 2 was then exposed to varying concentrations of Cas9 RNP in 20 mM mM Tris-HCl, pH 7.5, 75 mM KCl, 5 mM MgCl₂, 1 mM dithiothreitol (DTT), 5% glycerol buffered solution and incubated at 37 °C for 30 min, followed by denaturation at 95 °C for 2 min. Samples were then loaded onto a 1% agarose TAE gel with SYBR-safe and visualized on a BioRad EZDoc Gel imager.

Cell Culture Procedures

Cell staining with Trastuzumab scFv-sfGFP conjugate. SK-BR-3 and MDA-MB-468 cells were obtained in separate T-25 flasks from the Berkeley Cell Culture Facility, transferred to a 37 °C, 5% CO₂ incubator and grown to 90% confluency. Growth medium was removed, and cells were detached by incubation with 6 mL of 0.13% Trypsin in DPBS at 37 °C for 10 min. After visually confirming detachment, each cell sample was diluted to 15 mL with Binding Buffer (DPBS with 10% FBS and 1% NaN₃). Cells were added to 15 mL Falcon tubes and subject to two cycles of centrifugation (300 g, 5 min, room temperature), removing the supernatant after each cycle and diluting with 14 mL of Binding Buffer before the second cycle. Cells were re-suspended in 0.8 mL of Cell Binding Buffer and counted with a BioRad TC20 automated cell counter. Live cell counts were 99 to 100% as indicated by trypan blue staining. Cell suspension density was adjusted to 1.9 x10⁶ cells/mL with additional Binding Buffer. 100 µL aliquots were distributed to Eppendorf tubes and incubated on ice. 40 µL of the 0.05 µg/µL scFv-sfGFP construct or appropriate negative control in PBS was added to each aliquot, and these samples were allowed to sit on ice, in the dark for 1 hour. Samples were then subject to 3 cycles of centrifugation (300 g, 3

min), re-suspending cells in 1000 μ L of Binding Buffer before each cycle and discarding supernatant after each cycle. After centrifugation, cell pellets were re-suspended in 500 μ L of DPBS with 0.5% paraformaldehyde and kept on ice in the dark for 1-3 hours until flow cytometry analysis.

Flow Cytometry Analysis for cell staining. Cell samples were diluted 2-3x in PBS and analyzed on to a BD LSR II flow cytometer (BD Biosciences). Fluorescence excitation at the 488 nm laser line and FITC filter set were used. Data was acquired using BD FACSDiva software (version 6.2, 2000) 10,000 events were recorded for each sample. Data was processed using FlowJo software (version 10.0).

Preparation of Cas9 for cell culture trials. Cas9 RNP was prepared based on previously described methods (28). To prepare the Cas9 RNP complexes, modified Cas9 protein was incubated with sgRNA at a 1:1.2 molar ration. Briefly, we added sgRNA to Buffer#1 (25mM NaPi, 150mM NaCl, 200mM trehalose, 1mM MgCl₂), then added the Cas9 to the sgRNA, slowly with swirling, and incubated at 37°C for 10 min to form RNP complexes. RNP complexes were filtered prior to use through a 0.22 μ m Costar 8160 Filter pre-wet with 200 μ l Buffer#1. If needed, the RNP sample was concentrated with a 0.5ml Millipore Ultra 100 kDa cutoff filter, part # UFC510096, until the desired volume was obtained.

Neural progenitor cell (NPCs) line assay. NPCs were treated based on a previously described protocol (32). Briefly, NPCs were isolated from cortices from embryonic day 13.5 Ai9-tdTomato homozygous mouse embryos. Cells were cultured as neurospheres in NPC medium: DMEM/F12 with glutamine, Na-Pyruvate, 10 mM HEPES, non-essential amino acid, penicillin and streptomycin, 2-mercaptoethanol, B-27 without vitamin A, N2 supplement, bFGF and EGF, both 20 ng/ml as final concentration. NPCs were passaged using MACS Neural Dissociation Kit (Papain) cat# 130-092-628 following manufacturer's protocol. bFGF and EGF were refreshed every other day and passaged every 6 days. The NPC line was authenticated by immunocytochemistry marker protein staining. They were tested for mycoplasma using Hoechst stain with visual analysis and were negative.

Cas9 RNP direct delivery. Direct delivery RNP in NPC experiments was added to media and incubated with cells for 24 h, then cells were washed twice with 200 U/ml heparin in DMEM media and allowed to grow for 24 additional hours before analysis by flow cytometry. Total quantity of Cas9 per well of a 96-well plate was titrated from 100 pmol to 10 pmol.

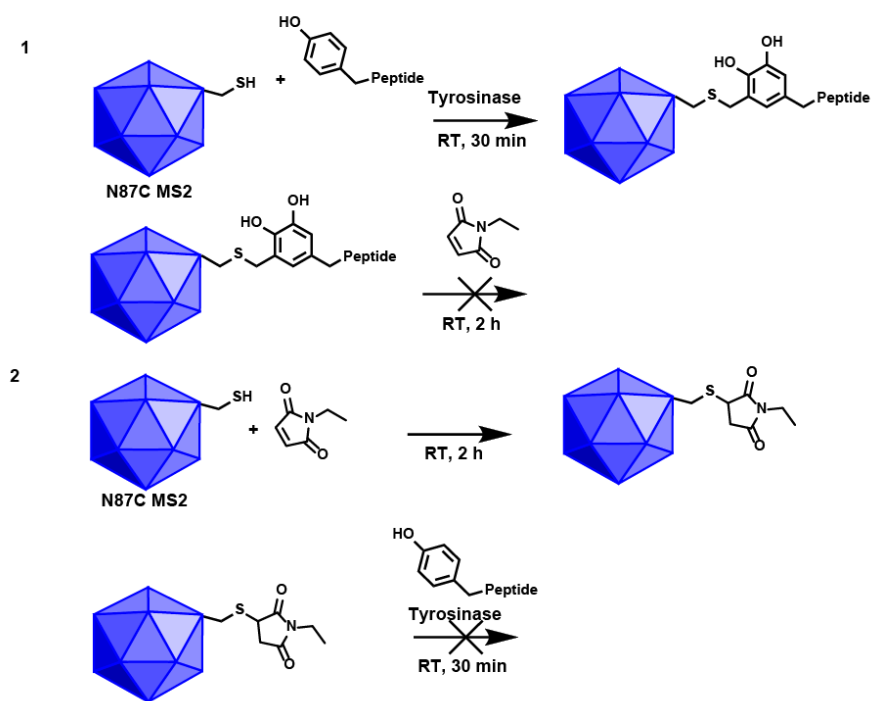


Fig. S1.

Flowchart representing Maleimide capping experiments. In a) we see that MS2 modified with endorphin does not react with additional maleimide coupling partner when exposed to coupling conditions in subsequent steps. Similarly, in b) we see that the maleimide modified MS2 is untouched by the tyrosinase reaction. Mass spectral data support these hypotheses, as c) shows NEM capped MS2, and d) shows NEM capped MS2 exposed to tyrosinase coupling conditions is not further modified by endorphin. Finally, e) shows WT MS2 is unmodified by tyrosinase conditions.

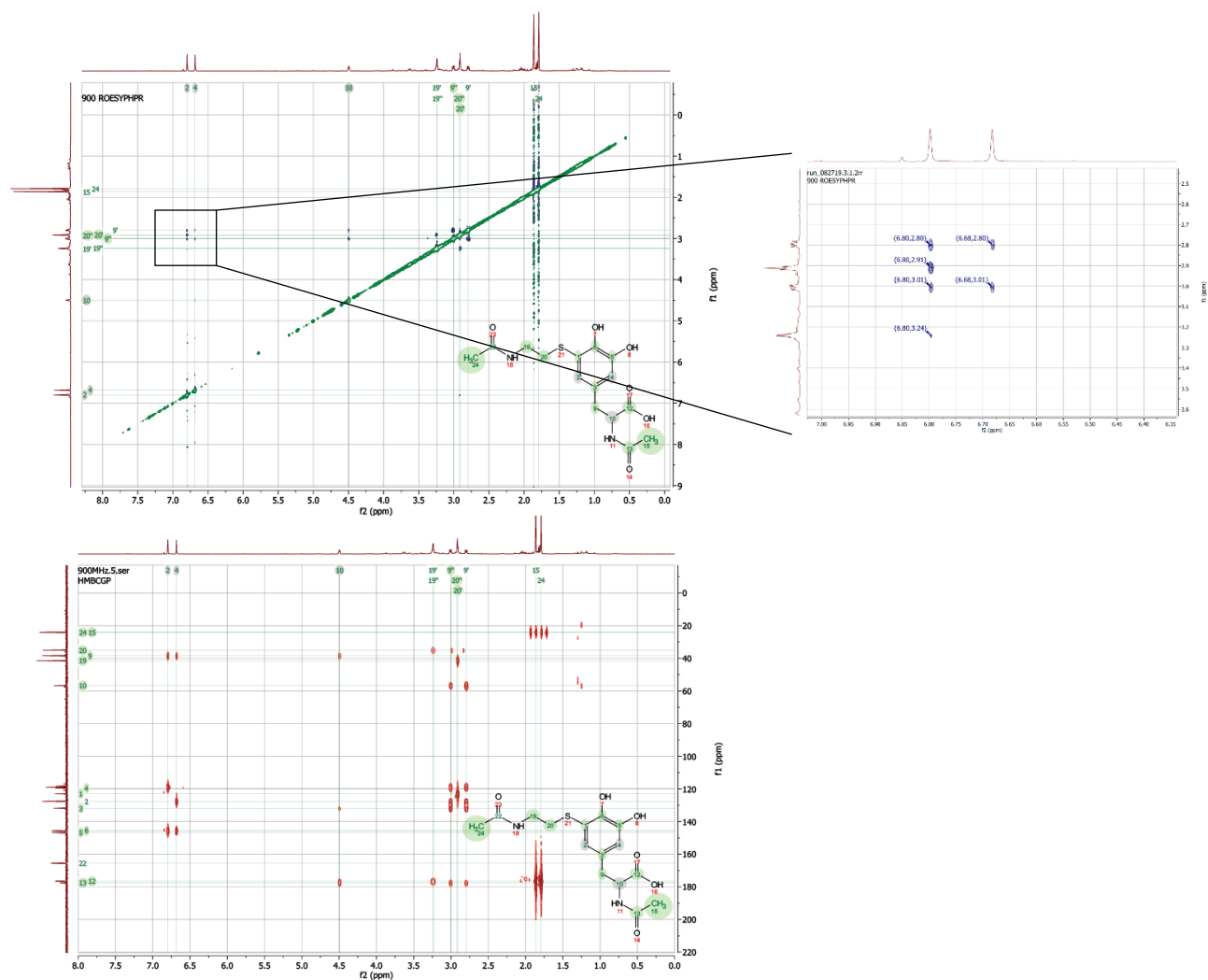


Fig. S2.

ROESY and HMBC data for N-acetyltyrosine coupled to N-acetylcysteamine. ROESY data shows clear correlation between only one proton (2) and the corresponding signal from the alkyl chain in cysteamine (20) while also coupling to the tyrosine alkyl chain (9). HMBC corroborates proton assignments.

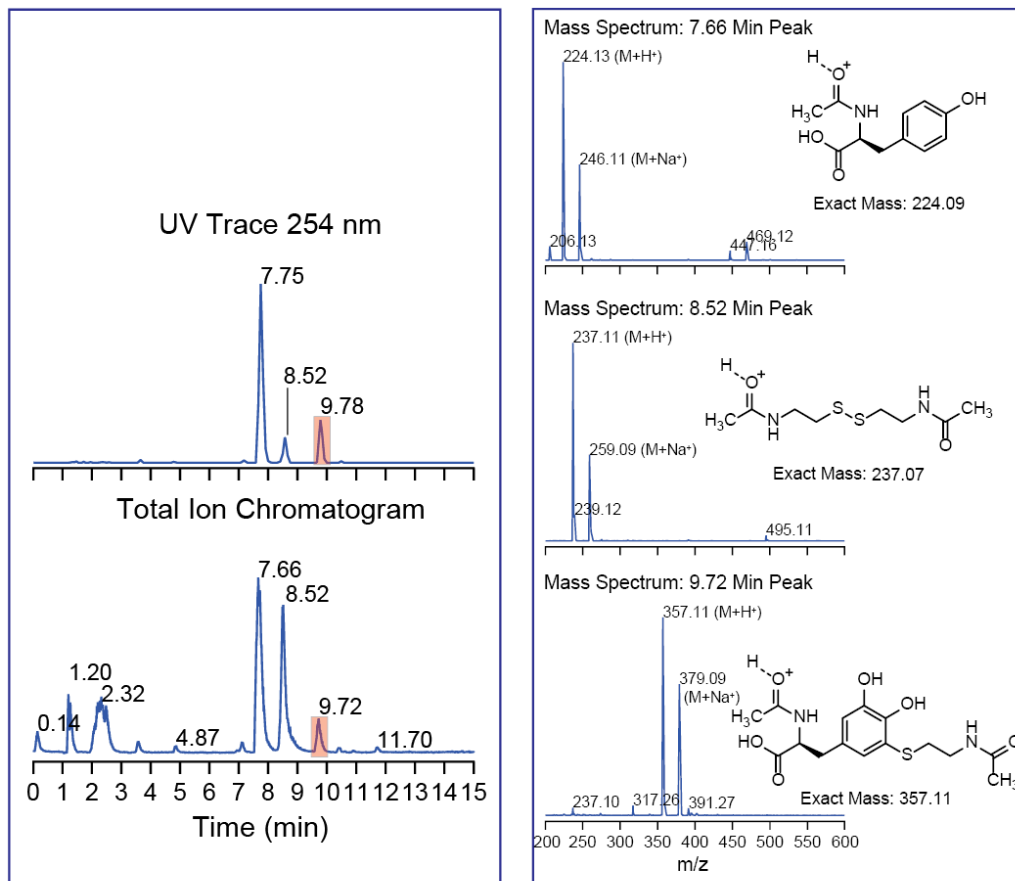


Fig. S3.

Analytical HPLC chromatogram of the small molecule thiol addition reaction sample. Red boxes indicate single adduct peak.

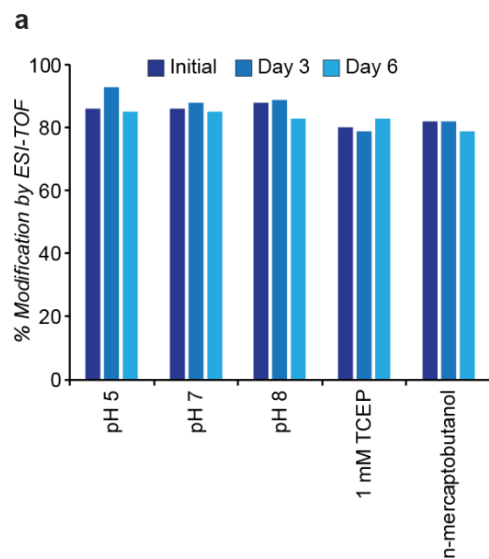


Fig. S4.

Addition of small molecule thiols to tyrosinase modified substrates. a) unmodified MS2 N87C b) N87C exposed to tyrosinase and n-mercaptobutanol yields no addition product c) MS2 N87C combined with endorphin and tyrosinase yields the +1 peptide modification d) MS2 N87C with endorphin exposed to n-mercaptobutanol yields some +1 peptide +1 n-mercaptobutanol product.

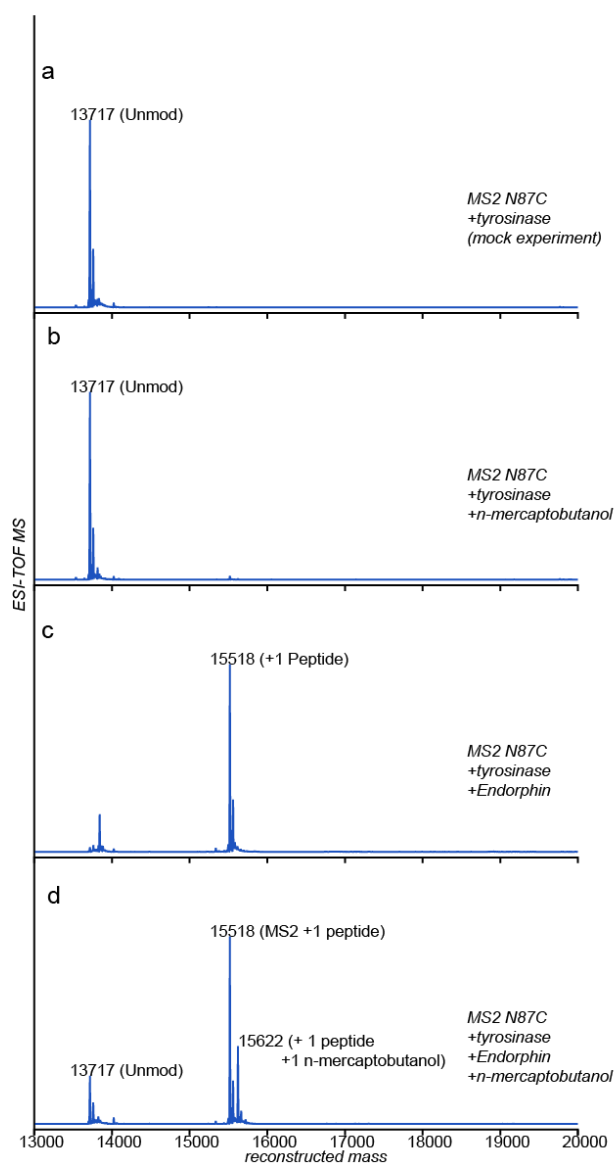


Fig. S5.

Addition of small molecule thiols to tyrosinase modified substrates. a) unmodified MS2 N87C b) N87C exposed to tyrosinase and n-mercaptobutanol yields no addition product c) MS2 N87C combined with endorphin and tyrosinase yields the +1 peptide modification d) MS2 N87C with endorphin exposed to n-mercaptobutanol yields some +1 peptide +1 n-mercaptobutanol product.

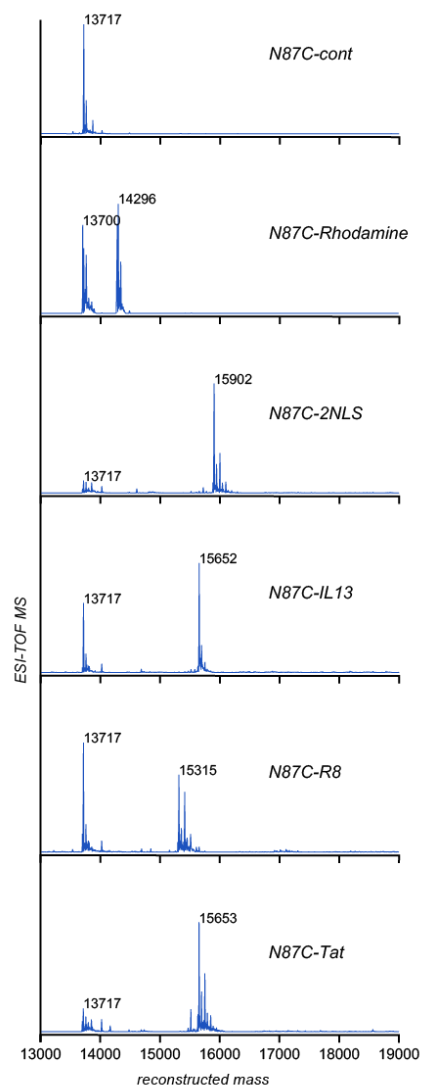


Fig. S6.

Addition of small molecule phenols and tyrosine containing peptides to MS2 N87C. 10 μ M of protein was combined with 100 μ M of phenol coupling partner and 167 nM abTyr in pH 6.5 phosphate buffer for 30 min at RT then quenched with 2 mM final concentration of tropolone and analyzed via ESI-TOF

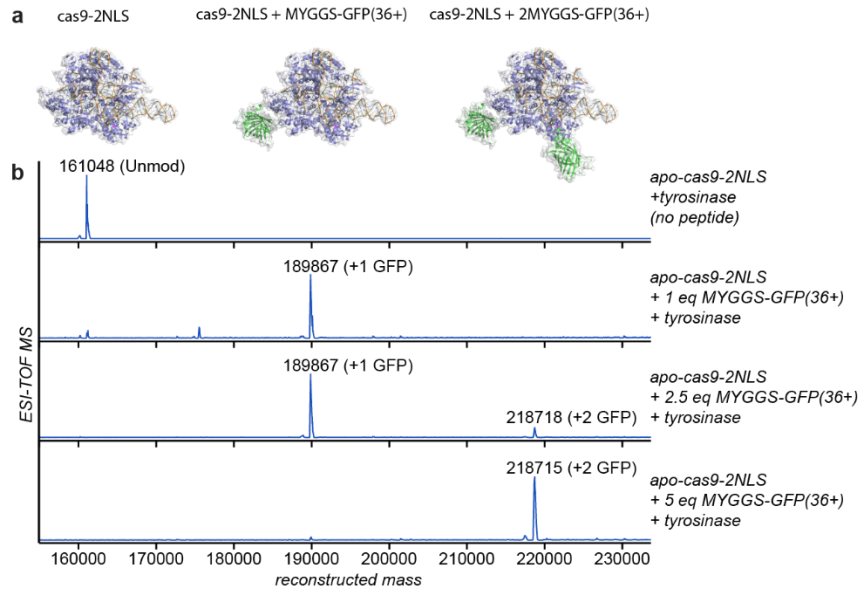


Fig. S7

Coupling of MYGGS +36 GFP to Cas9. In a) we show structural representations of the coupled proteins using PDBID 5F9R and 6B9C for Cas9 and GFP respectively. b) presents ESI-MS data indicating the quantitative conversion to singly or doubly modified Cas9 when 1 to 5 equivalents of GFP are added to tyrosinase reaction conditions.

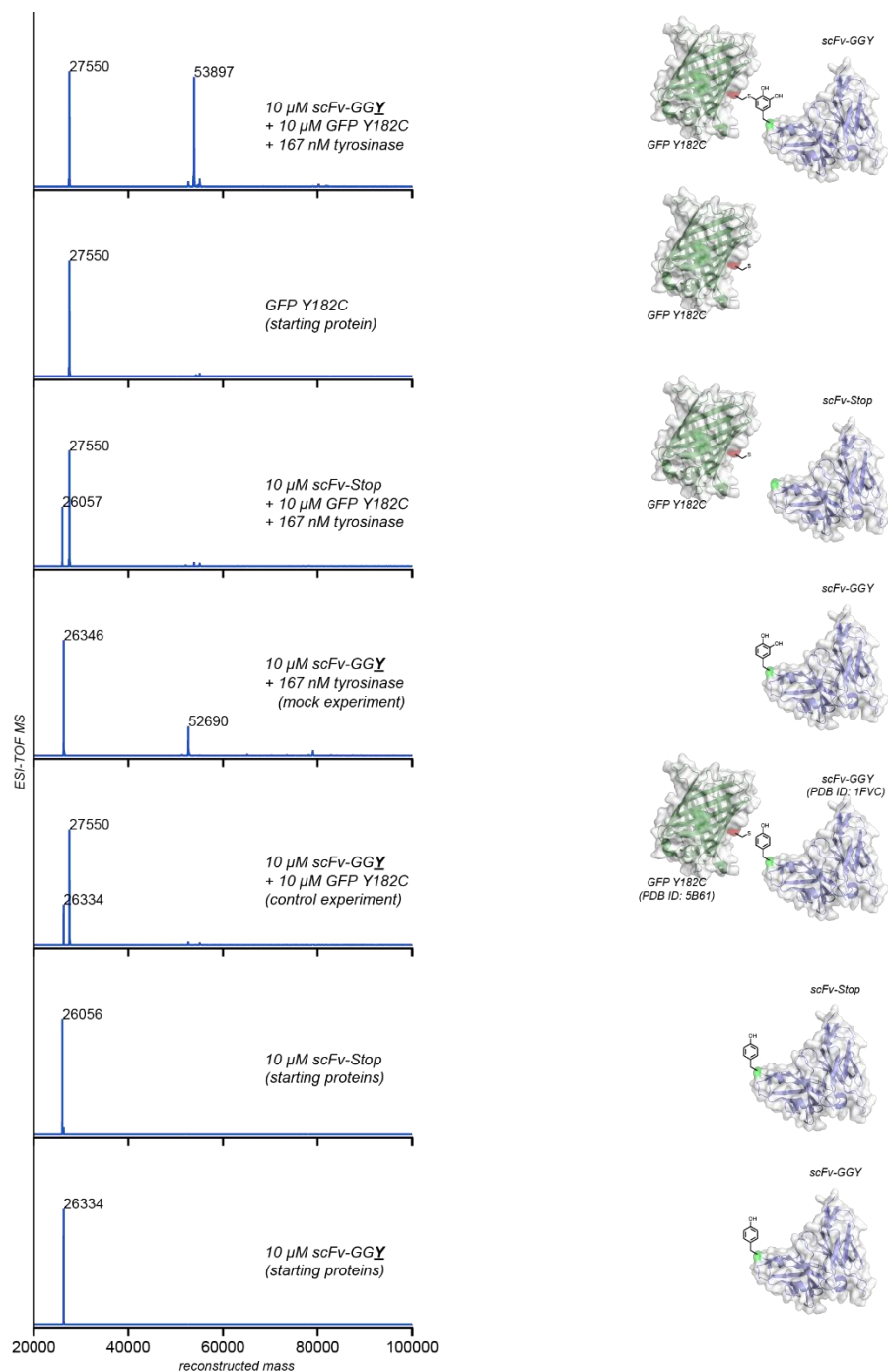
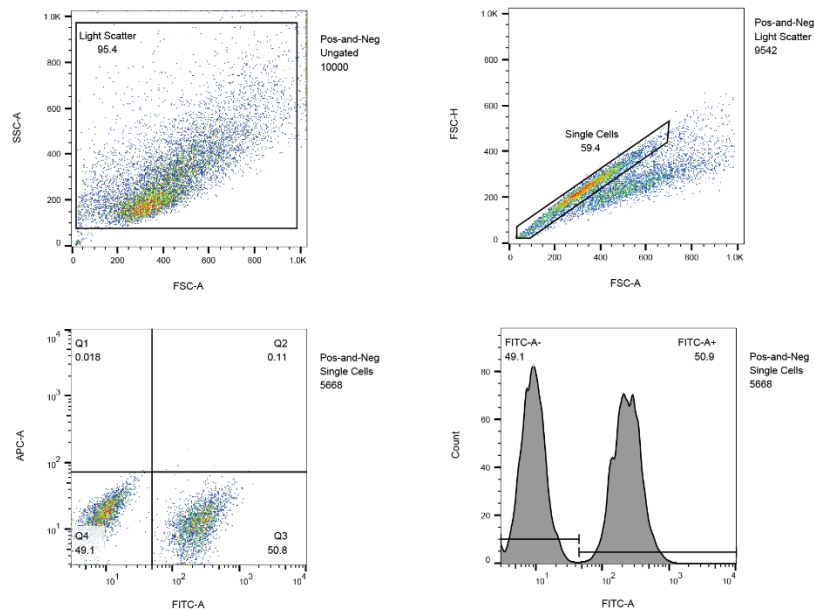


Fig. S8.

Coupling of Trastuzumab-derived scFV (Tras) and GFP Y182C. Mass spectrum from ESI-TOF showing coupling between the scFv and GFP Y182C. A 5 μM solution of scFv was combined with 5 μM of GFP-Y182C to yield near quantitative conversion to the GFP-scFv dimer.



Gate Depth	Name	Statistic (%)	#Cells	Gate Depth	Name	Statistic (%)	#Cells
	Pos-and-Neg		10000		Sample 4		10000
>	Pos-and-Neg/Light Scatter	95.4	9542	>	Sample 4/Light Scatter	93.3	9332
>>	Pos-and-Neg/Light Scatter/Single Cells	59.4	5668	>>	Sample 4/Light Scatter/Single Cells	67.2	6270
>>>	Pos-and-Neg/Light Scatter/Single Cells/FITC-A+	50.9	2886	>>>	Sample 4/Light Scatter/Single Cells/FITC-A+	0.21	13
>>>	Pos-and-Neg/Light Scatter/Single Cells/FITC-A-	49.1	2782	>>>	Sample 4/Light Scatter/Single Cells/FITC-A-	99.8	6257
	Sample 1		10000		Sample 5		10000
>	Sample 1/Light Scatter	98	9796	>	Sample 5/Light Scatter	93.9	9387
>>	Sample 1/Light Scatter/Single Cells	80.2	7852	>>	Sample 5/Light Scatter/Single Cells	67.3	6314
>>>	Sample 1/Light Scatter/Single Cells/FITC-A+	0.23	18	>>>	Sample 5/Light Scatter/Single Cells/FITC-A+	0.33	21
>>>	Sample 1/Light Scatter/Single Cells/FITC-A-	99.8	7834	>>>	Sample 5/Light Scatter/Single Cells/FITC-A-	99.7	6293
	Sample 2		10000		Sample 6		10000
>	Sample 2/Light Scatter	90.4	9041	>	Sample 6/Light Scatter	91.4	9138
>>	Sample 2/Light Scatter/Single Cells	62.7	5665	>>	Sample 6/Light Scatter/Single Cells	62.5	5707
>>>	Sample 2/Light Scatter/Single Cells/FITC-A+	0.19	11	>>>	Sample 6/Light Scatter/Single Cells/FITC-A+	99.2	5662
>>>	Sample 2/Light Scatter/Single Cells/FITC-A-	99.8	5654	>>>	Sample 6/Light Scatter/Single Cells/FITC-A-	0.79	45
	Sample 3		10000				
>	Sample 3/Light Scatter	94.5	9446				
>>	Sample 3/Light Scatter/Single Cells	67.9	6416				
>>>	Sample 3/Light Scatter/Single Cells/FITC-A+	5.77	370				
>>>	Sample 3/Light Scatter/Single Cells/FITC-A-	94.2	6046				

Fig. S9.

Labeling of HER2+ cells using Trastuzumab scFv-sfGFP conjugate. Representative plots showing gating scheme on a mixture of 1:1 untreated MDA-MB-468 (HER2-) cells and SK-BR-3 (HER2+) cells treated with Trastuzumab scFv-sfGFP conjugate (as in Sample 6). The gating scheme was applied uniformly to all samples and statistics are tabulated. The fluorescence intensity of single cells was measured using the 488 laser line and FITC filter set.

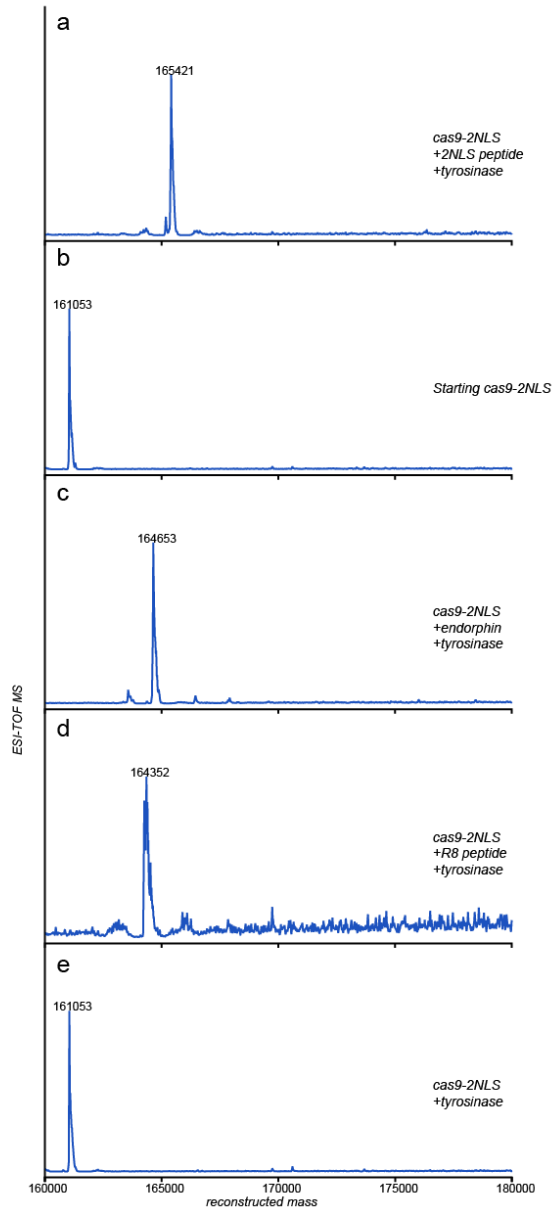


Fig. S10.

Cas9 peptide modifications. Cas9 at 10 μ M combined with 100 μ M of the respective peptides for 1 h at 4 $^{\circ}$ C in Buffer 1. a) Cas9 coupled to 2SV40 peptide b) control sample of Cas9-2NLS c) Cas9-2NLS coupled to endorphin d) Cas9-2NLS coupled to R8 peptide e) Cas9 exposed to tyrosinase conditions exhibits no oxidation.

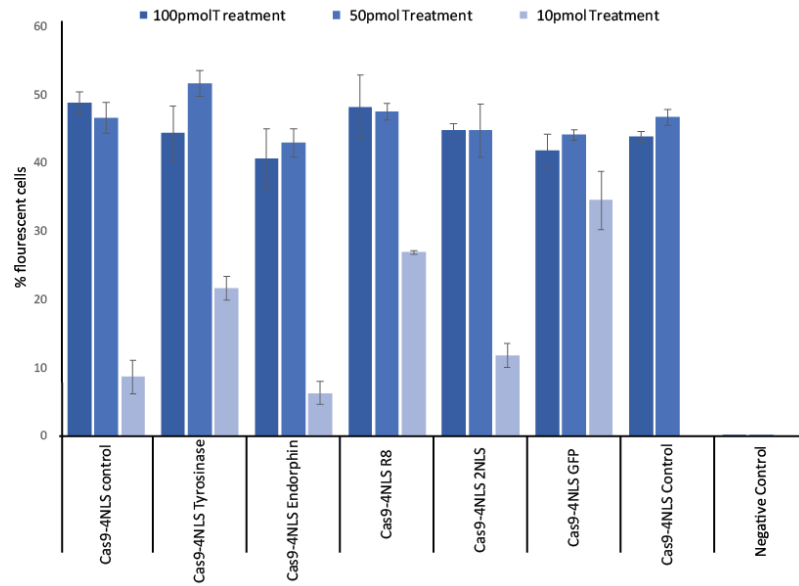


Fig. S11.

Cas9-4NLS editing in neural progenitor cells. All treatment conditions showed similar editing despite the addition of diverse peptides.

Table S1.

Cysteine-Tyrosine coupling stability			
Reaction Condition	Day 1	Day 3	Day 6
pH 5	86%	93%	85%
pH 7	86%	88%	85%
pH 8	88%	89%	83%
TCEP + Tropolone	80%	79%	83%
N-mercapto-butanol	82%	82%	79%

Modification rates for MS2-N87C coupled to α -endorphin over 6-day time course.