doi.org/10.26434/chemrxiv.7150094.v1

Enzymatic Modification of N-Terminal Proline Residues Using Phenol Derivatives

Johnathan Maza, Daniel Bader, Lifeng Xiao, Alan Marmelstein, Daniel D. Brauer, Adel M. ElSohly, Matthew J. Smith, Matthew Francis

Submitted date: 29/09/2018 • Posted date: 01/10/2018

Citation information: Maza, Johnathan; Bader, Daniel; Xiao, Lifeng; Marmelstein, Alan; Brauer, Daniel D.; ElSohly, Adel M.; et al. (2018): Enzymatic Modification of N-Terminal Proline Residues Using Phenol Derivatives. ChemRxiv. Preprint.

A convenient enzymatic strategyis reported for the modification of proline residues in the N-terminal positions of proteins. Using a tyrosinase enzyme isolated from Agaricus bisporus(abTYR), phenols and catechols are oxidized to highly reactive o-quinone intermediates that then couple to N-terminal proline residues in high yield. Key advantages of this bioconjugation method include (1) the use of air-stable precursors that can be prepared on large scale if needed, (2) mild reaction conditions, including low temperatures, (3) the targeting of native functional groups that can be introduced readily on most proteins, and (4) the use of molecular oxygen as the sole oxidant. This coupling strategy was successfully demonstrated for the attachment of a variety of phenol-derivatized cargo molecules to a series of protein substrates, including self-assembled viral capsids, enzymes, and a chitin binding domain (CBD). The ability of the CBD to bind to the surfaces of yeast cells was found to be unperturbed by this modification reaction.

File list (2)

Enzymatic Modification of N-Terminal Proline Residues Using Phenol Derivatives

Johnathan C. Maza,[1] Daniel L. V. Bader,[1] Lifeng Xiao,[1] Alan M. Marmelstein,[1] Daniel D. Brauer,[1] Adel M. ElSohly,[1] Matthew J. Smith,[1] Matthew B. Francis[1,2,*]

[1]Department of Chemistry, University of California, Berkeley, California 94720, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratories, Berkeley, California 94720, USA. Email: mbfrancis@berkeley.edu

KEYWORDS: Bioconjugation, protein modification, N-terminus, proline, oxidative coupling, tyrosinase, chitin binding domain

ABSTRACT: A convenient enzymatic strategy is reported for the modification of proline residues in the N-terminal positions of proteins. Using a tyrosinase enzyme isolated from *Agaricus bisporus* (abTYR), phenols and catechols are oxidized to highly reactive *o*-quinone intermediates that then couple to N-terminal proline residues in high yield. Key advantages of this bioconjugation method include (1) the use of air-stable precursors that can be prepared on large scale if needed, (2) mild reaction conditions, including low temperatures, (3) the targeting of native functional groups that can be introduced readily on most proteins, and (4) the use of molecular oxygen as the sole oxidant. This coupling strategy was successfully demonstrated for the attachment of a variety of phenol-derivatized cargo molecules to a series of protein substrates, including self-assembled viral capsids, enzymes, and a chitin binding domain (CBD). The ability of the CBD to bind to the surfaces of yeast cells was found to be unperturbed by this modification reaction.

Introduction

The ability to construct site-specific protein bioconjugates has broad applicability in a range of disciplines. Fluorophores are commonly attached to proteins to study their localization in cells and biodistribution in living organisms,[1] and fluorescent labels can provide biophysical information about conformational changes.[2] Proteins are routinely immobilized on surfaces to probe mechanistic and binding features,[3,4] and increased efforts are exploring the utility of proteins as material components.[5] Most relevant to drug discovery is the synthesis of antibody-drug conjugates, where toxic payloads are covalently bound to malignant cell-type specific antibodies.[6] In all cases, the construction of a desired bioconjugate requires the specific folded conformation of the protein to be preserved to achieve proper function. Chemistries used to modify proteins must therefore proceed efficiently in aqueous conditions and under mild ranges of pH and temperature to minimize perturbations in protein structure.

A number of reactive strategies have been developed to meet these criteria. These reactions most commonly target lysine or cystine residues, which are the two most nucleophilic side chains on the surfaces of proteins.[7] However, it can be difficult to control the modification numbers and locations using lysine-specific reactions, and the reliance of many proteins on cysteine side chains for proper function and folding can complicate the targeting of this residue. More recently, non-canonical amino acid (ncAA) mutagenesis[8–10] has emerged as a powerful technique to introduce new reactive handles in defined locations, allowing for site-selective modification when the appropriate bioorthogonal chemistry[11] is employed.

As a complementary approach, our lab[12–15] and others[16–19] have focused on the development of mild reactions that target the N-terminus as a chemically distinct site in a given protein sequence. Compared to the primary amines of lysine side chains, protein N-termini offer lower pK_a values and adjacent side chain groups that can participate in the reactions.[12,20] Furthermore, a variety of protein N-terminal amino acids can be introduced directly during protein expression. Studies have shown that if the amino acid following the initial methionine encoded by the start codon is small, such as alanine or proline, the methionine is removed completely in bacterial and mammalian expression systems.[21,22]

One class of reactions with particular promise for site-specific protein labeling involves the oxidative coupling of *o*-quinoid intermediates to nucleophilic functional groups. In these strategies, *o*-aminophenols or *o*-catechols are first oxidized through the use of protein-compatible oxidants, such as sodium periodate or $K_3Fe(CN)_6$.[13,23,24] The resulting *o*-iminoquinone or *o*-quionone species react rapidly with *p*-aminophenylalanine ncAAs,[23] N-terminal prolines,[13] and reduced cysteine residues,[13] as shown in **Figure 1a**. These reactions have been applied to the efficient modification of viral capsids with nucleic acids,[25] polymers,[26] and proteins;[27] inorganic nanoparticles with polymers and proteins;[28] and electrochemically active surfaces with oligonucleotides capable of cell capture.[29,30] Taken together, these studies highlight the utility of this class of reactions for the preparation of highly complex bioconjugates for many applications. These reactions also bear resemblance to catechol-based interactions involved in the adhesion of mussels to surfaces, which have found substantial use as adhesives that can function in aqueous conditions.[31,32]

Despite their success as bioconjugation partners, the *o*-aminophenol and *o*-catechol substrates used in these oxidative coupling reactions have key limitations. While both reagents are readily activated in the presence of the mild oxidant $K_3Fe(CN)_6$, they also oxidize in the presence of air over time. This limits their shelf-life, while also adding challenges to the chemical preparation of these reagents. Indeed, when synthesizing aminophenol derivatives, our lab prefers to work with nitrophenols and then reduce them to *o*-aminophenols using sodium dithionite prior to reaction.[13,33] More recently, we reported *o*-methoxyphenols as alternative air-stable precursors to *o*-catechols.[24] In these cases, the *o*-methoxyphenol is oxidized to an *o*-quinone via a periodate-mediated reaction, in which the methoxy group is substituted with water. Further work on proteins has demonstrated that potentially deleterious side oxidations resulting from a large excess of periodate can be reduced if the oxidant is

quenched with mannose following *o*-methoxyphenol activation and prior to protein reaction.[24] Despite these workarounds, however, this chemistry would still benefit from additional routes to the key *o*-quinone intermediates. Ideally, these strategies should feature substrates that are commercially available, allow reagent storage, and minimize the need for large excesses of oxidants that must be quenched and removed following the bioconjugation reaction.

Previous work has demonstrated that enzymes are powerful tools for oxidizing phenol-containing compounds, playing myriad roles in biology, including neurotransmitter synthesis,[34,35] toxin removal,[36] and the production of melanin.[37,38] In melanosomes, tyrosinase catalyzes the 4-electron oxidation of tyrosine to dopaquinone, which undergoes an intramolecular 1,4-conjugate addition to produce dopachrome, **Figure 1b**.[37,38] This molecule is then further polymerized to produce various melanin compounds across a variety of organisms. Tyrosinase can also catalyze the two-electron oxidation of dopamine to the same dopaquinone intermediate, demonstrating the ability of the enzyme to make the same *o*-quinone intermediates from both catechol and phenolic substrates.[37,39]

Herein, we apply this biochemical concept to develop an enzymatic oxidative coupling strategy. Using a commercially available tyrosinase enzyme, the *o*-quinone intermediates required for oxidative coupling reactions can be accessed using simple catechol and phenol precursors. In addition to simplifying the substrate synthesis, this activation method uses only atmospheric oxygen as the sole reaction byproduct. Starting from commercially available NHS esters coupled to tyramine, this strategy is demonstrated through the attachment of a variety of phenolic derivatives to N-terminal proline residues in protein sequences. Finally,

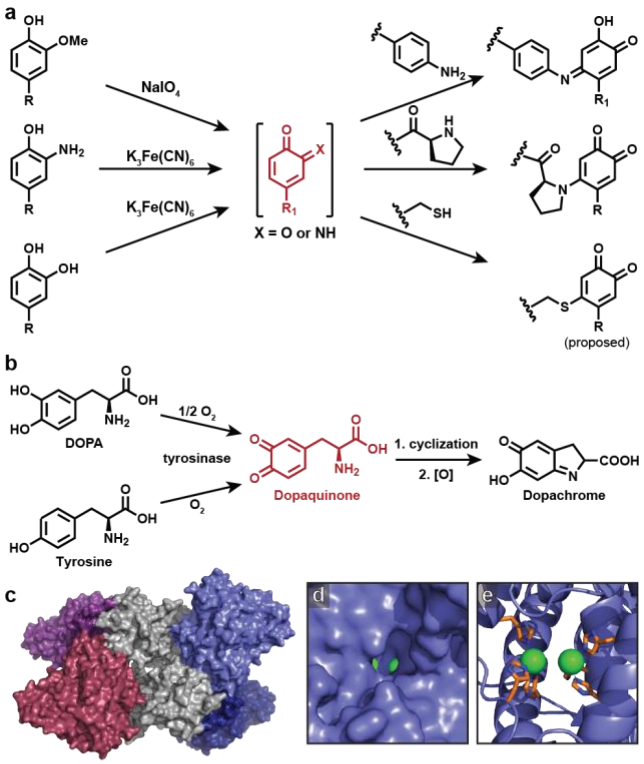


Figure 1. Chemoselective oxidative coupling reactions for bioconjugation and melanin biosynthesis. (a) Methoxyphenols, aminophenols, and catechols can be converted to *o*-quinoid intermediates *in situ* using periodate or the mild oxidant $K_3Fe(CN)_6$. These rapidly couple with anilines, N-terminal prolines, or thiols to form stable bioconjugation products. (b) A related pathway is catalyzed by tyrosinase as part of the melanin biosynthetic pathway. Both dopamine and tyrosine can serve as substrates for this reaction. (c) The tyrosinase isolated from *Agaricus bisporus* is shown (PDB ID: 2Y9W). The inactive domains are depicted in gray. (d,e) The active sites consist of Type-III binuclear copper complexes ligated by histidine residues. These ions can be accessed via a deep pocket in the protein structure.

we extend this technique towards the construction of an N-terminally labeled chitin-binding domain (CBD) for use as a potential targeting agent for yeast and pathogenic fungal organisms.

Results and Discussion

Screening the enzymatic oxidative coupling with small molecule substrates. Tyrosinase is conveniently available from commercial sources following isolation from the common button mushroom *Agaricus bisporus* (abTYR).[38,40] This variant is a ~120 kDa MW tetrameric protein containing multiple active subunits and inactive subunits. The enzyme has type III di-copper binding sites located in deep and spacious active sites, **Figure 1c**.[39,40] It is these large cavities that allow abTYR to act on a variety of phenols beyond tyrosine. Previous studies have shown abTYR is capable of binding and oxidizing acetaminophen, BPA, and even 17-β-estradiol,[41–43] indicating that the enzyme is promiscuous for a variety of phenol reagents with varying amounts of added complexity. This capacity for phenol oxidation has recently been exploited in the context of bioconjugation through the conversion of engineered, solvent accessible tyrosine residues to *o*-quinones to allow hetero-Diels-Alder reactions with cyclooctyne derivatives.[44] This chemistry was demonstrated for the attachment of cargo molecules to C-terminal tyrosine residues extending from properly engineered antibodies. A tyrosinase-based approach has also been described to convert tyrosine residues in peptides and proteins to *o*-catechols to allow reversible adduct formation with aryl boronic acids.[45]

To explore the potential of this enzyme in the context of the oxidative coupling chemistry, we first used abTYR to mediate the coupling of catechols to a proline engineered at the N-terminus of super-folder GFP (Pro-sfGFP, MW = 27575 Da), thereby obviating the need for large excesses of oxidants as required in previous oxidative coupling reactions. Briefly, samples containing 10 µM Pro-sfGFP and 10 equivalents of 4-methylcatechol were prepared in 20 mM pH 6.5 phosphate buffer. A solution of abTYR was added at final concentrations of 83 nM or 8 nM, and the reactions were incubated for 30 min at room temperature. As a comparison, an additional reaction was run with 0.5 mM $K_3Fe(CN)_6$, as previous studies have shown that this reagent is capable of oxidizing catechols to *o*-quinones for attachment to proline N-termini.[13] High conversion to the expected product was observed using abTYR at both high and low concentrations, **Figure 2, X=OH**. When incubated with Pro-sfGFP alone, no background oxidation of its native tyrosines was observed, suggesting that abTYR is not capable of binding and oxidizing the endogenous tyrosines within this timeframe. This result was in accordance with reports for other tyrosine-containing proteins. A survey of sfGFP substrates with other N-terminal amino acids confirmed the requirement for the proline residues, with conversions below 40% being observed with other N-terminal residues, **Supporting Information Figure S1**. This result was in line with previous findings for the ferricyanide-mediated version of the reaction.[13]

We next explored the ability of abTYR to mediate the attachment of simple phenols to N-terminal proline residues. Phenol derivatives are easy to prepare using commercially available materials and store readily for long periods of time, thus overcoming the inherent limitations of aminophenol and catechol reagents. Samples of Pro-sfGFP were exposed to 10 equivalents of *p*-cresol in the presence of abTYR under the conditions described above. After 30 min, 83 nM of abTYR successfully oxidized *p*-cresol to the *o*-quinone intermediate, completely converting the Pro-sfGFP starting material to the same product that was obtained with 4-methylcatechol, **Figure 2, X=H**. As anticipated, this reaction did not proceed using $K_3Fe(CN)_6$ as the oxidant. An increased dependence on abTYR concentration was observed compared to the catechol version of the reaction, but full conversion was still realized in 30 min using less than 1 mol% of enzyme relative to protein substrate.

In addition to the desired reaction product, reactions using both catechols and phenols produced a small amount (15% or less) of a double modification product. To date, we have not been able to obtain sufficient quantities of this species to characterize it. It may arise from the addition of particularly exposed lysine side chains to the quinone intermediates, or it could result from oligomerizations of the quinones

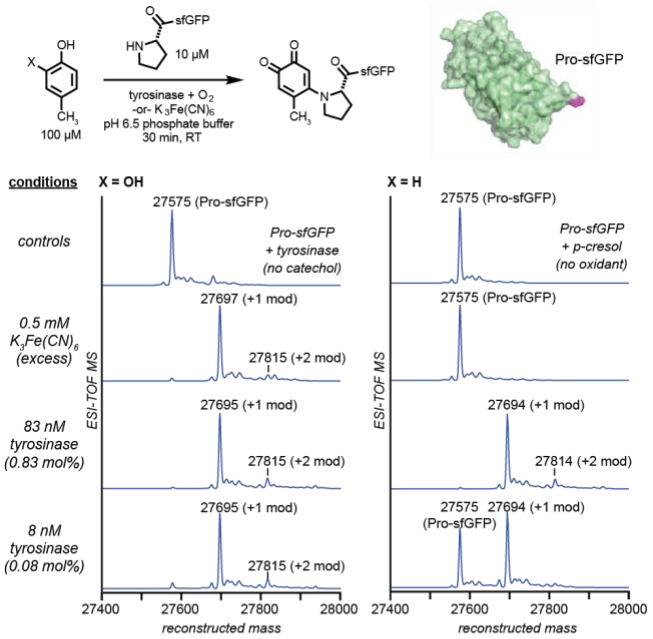


Figure 2. Site-selective oxidative coupling reactions using the tyrosinase from *Agaricus bisporus* (abTYR). Superfolder GFP with an N-terminal proline residue shown in pink (Pro-sfGFP, MW=27575 Da) was evaluated as an initial substrate. The expected modification product ($[M+H]^+$) has a mass of 27695 Da. In some cases a small amount (<15%) of double modification can be detected at 27815 Da. TCEP was added to the $K_3Fe(CN)_6$ sample upon workup, resulting in a mass increase of 2 Da due to reduction of the *o*-quinine to the catechol. Structure based on PDB ID: 2B3P.

through Diels-Alder or other pathways.[46] If problematic, this product can be minimized by lowering the reaction temperature to 4 ºC (*vide infra*).

Reaction optimization for site-selective biotinylation. The majority of bioconjugation reactions rely on commercially available small molecules, such as NHS-esters and maleimides, that are pre-activated for attachment to lysine and cysteine residues, respectively. To make use of this convenient set of reagents, we next developed a simple protocol to convert NHS-esters to N-terminal proline specific reagents using tyramine. Biotin was chosen as an initial target for methodology development because it is commonly used to tag proteins for fluorescence labeling,[47] affinity capture,[48] and surface immobilization.[49]

To prepare biotin-phenol **1a**, a 1.1 equivalent portion of the corresponding NHS-ester was added to tyramine in dry DMF, **Figure 3a**. The reaction was vortexed overnight at room temperature. To ensure hydrolysis of the remaining NHS ester groups, water was added before diluting the resulting phenol for use in protein coupling reactions. Portions of this solution were added to a final concentration of 100 µM to a 10 µM solution of Pro-sfGFP in 20 mM phosphate buffer at pH 6.5. A solution of abTYR was added at concentrations varying from 60-200 nM, and the reactions were incubated at room temperature for 30 min. At all concentrations, the enzyme successfully mediated the attachment of **1a** to the proline N-terminus of Pro-sfGFP, as indicated by ESI-TOF MS, **Supporting Information Figure S2**. Using 80 nM abTYR, ~90% conversion of the starting protein was observed, and complete conversion was observed at 200 nM.

Next, we screened the effects of altered pH on the abTYR mediated attachment of **1a** to Pro-sfGFP. A solution containing final concentrations of Pro-sfGFP at 10 µM and **1a** at 100 µM was prepared in 20 mM phosphate buffer at pH values ranging from 5 to 9. Upon addition of a final concentration of 160 nM enzyme and incubation at RT for 30 min, product formation was observed at all pH values, **Supporting Information Figure S2**. Yields were reduced at pH values less than 6.5, likely due to the increased protonation of the proline N-

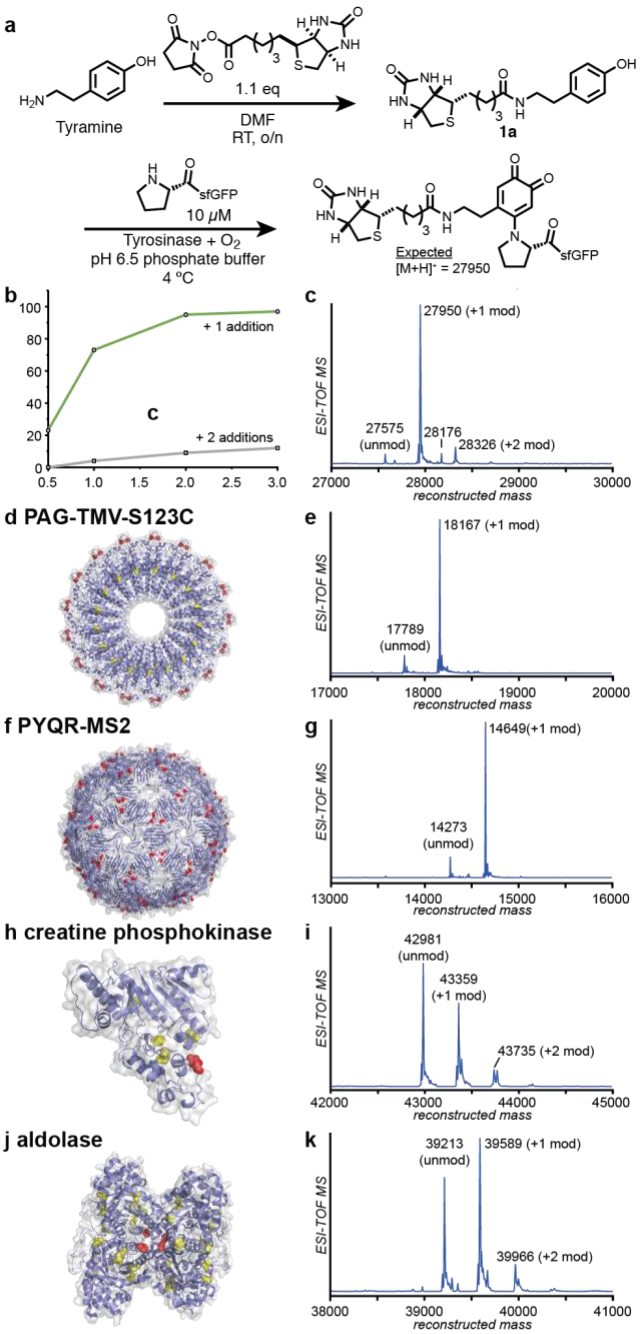


Figure 3. Site-selective biotinylation of proteins with N-terminal proline residues. (a) Using optimized conditions, biotin-phenol **1a** was attached to variety of proteins with proline residues at the N-terminus. (b) A time course experiment for Pro-sfGFP at 4 ºC showed that the reaction proceeds efficiently while minimizing over addition products. (c) High conversion to singly biotinylated Pro-sfGFP can be achieved after 2 h at 4 ºC. These conditions were used for the biotinylation of additional protein substrates with **1a**, including (d,e) TMV double disks with PAG N-terminal extensions, (f,g) MS2 viral capsids with PYQR N-terminal extensions, (h,i) commercially obtained creatine phosphokinase, and (j,k) comercially obtained aldolase. For the PAG-TMV, CPK, and aldolase, the free cysteine residues (yellow) were protected with Ellman's reagent prior to the oxidative coupling reaction. The proteins were reduced with TCEP to liberate the cysteines before analysis.

terminus. The enzyme abTYR was found to be capable of oxidizing **1a** in all buffers surveyed, **Supporting Information Figure S2.**

Finally, reactions on 10 µM Pro-sfGFP in 20 mM phosphate buffer at pH 6.5 were prepared as described above with varying

concentrations of phenol **1a**. After incubation with 160 nM abTYR for 30 min at room temp, product conversion was analyzed. Complete conversion was only achieved if 100 µM (10 equivalents) or greater of **1a** were used, likely because lower concentrations were too far below the K_M value of the enzyme (reported as 0.5 mM for L-tyrosine).[50] Based on these screens, the optimal conditions chosen were 10 µM protein and 100 µM of **1a** with 200 nM abTYR in 20 mM phosphate buffer at pH 6.5.

During the optimization studies we also determined the ideal storage conditions for abTYR. We found that the lyophilized powder obtained commercially could be stored at –20 ºC indefinitely. For use in reactions, stock solutions were prepared at 2 mg/mL (appx. 17 µM) in 50 mM phosphate buffer at pH 6.5. These solutions could be stored at –80 ºC for over 100 days before use; however, storage of these solutions at –20 ºC or higher temperatures led to decreases in activity over time. A summary of these studies appears in **Supporting Information Figure S3**.

Throughout these studies, the over modification product was observed to varying degrees depending on the conditions employed. To decrease this unwanted product, the reaction was run at reduced temperatures. A reaction was prepared using the optimized conditions described above and was placed in a 4 ºC cold room. At various times 20 µL portions of this solution were sampled and quenched with a final concentration of 1.9 mM of both TCEP and tropolone, which reduced the activated *o*-quinone[13] and inhibited the enzyme,[38,40] respectively. Following analysis with ESI-TOF MS, complete conversion of Pro-sfGFP to the desired product was achieved in 2 h, **Figure 3b and 3c**. Furthermore, the lower temperature reduced the unwanted secondary modification by ~10%. The reaction is also compatible with elevated temperatures, achieving full conversion of the Pro-sfGFP starting material in only 15 min at 37 °C, albeit with increased secondary modification observed (see **Supporting Information Figure S4**).

We also explored the efficiency of the abTYR-mediated biotinylation for other proline N-terminal proteins. Previous work from our group has shown that a K53R, K68R double mutant of the tobacco mosaic virus (TMV) capsid protein can be engineered to contain an N-terminal proline while retaining its ability to self-assemble into stable double disks.[26] This protein served as one substrate in our panel, also possessing a free cysteine group in position 123 of each monomer. Recent work from our lab and the Tullman-Ercek group has shown that the bacteriophage MS2 viral capsid can be engineered to contain proline residues at the N-termini of the coat proteins if included at the end of XXR extensions.[51,52] Finally, we examined creatine phosphokinase (CPK) and aldolase (ALD) as two commercially available enzymes that possess N-terminal proline residues.

These proteins were evaluated for site-selective biotinylation using the fully optimized reaction protocol for attachment of **1a** described above. First, the free thiols of TMV, CPK, and ALD were temporarily protected with Ellman's reagent, as we have described previously for oxidative coupling reactions.[13] Then, to 10 µM of each protein in 20 mM phosphate buffer at pH 6.5 was added phenol **1a** at a final concentration of 100 µM CPK and ALD or 200 µM for MS2 and TMV. To this was added a final concentration of 200 nM abTYR, and the reactions were run for 2 h at 4 ºC. Following a 1 min quench and reduction using final concentrations of 1.9 mM TCEP and tropolone, all proteins showed conversion to the expected modified product. The addition of TCEP also had the effect of reducing the Ellman's Reagent disulfides to yield the free thiols. The engineered proline N-termini on MS2 and TMV performed better than those on CPK and ALD, likely because the extended linkers helped to increase the solvent accessibility of the proline residues.

Application of abTYR-mediated oxidative coupling to a chitin-binding domain. Chitin-binding domains (CBD) are protein segments capable of selectively binding chitin on the surface of yeast and pathogenic fungal cells.[52] This binding ability makes these biomolecules attractive for the targeted delivery of antifungal agents. The abTYR oxidative coupling reaction could provide a useful route to access CBD bioconjugates by targeting a proline residues engineered at their N-termini, as this location is remotely disposed from the binding surface.[53] For use in these studies, a proline N-terminal variant of a CBD from the archaea *Pyrococcus furiosus* was generated.[52] This

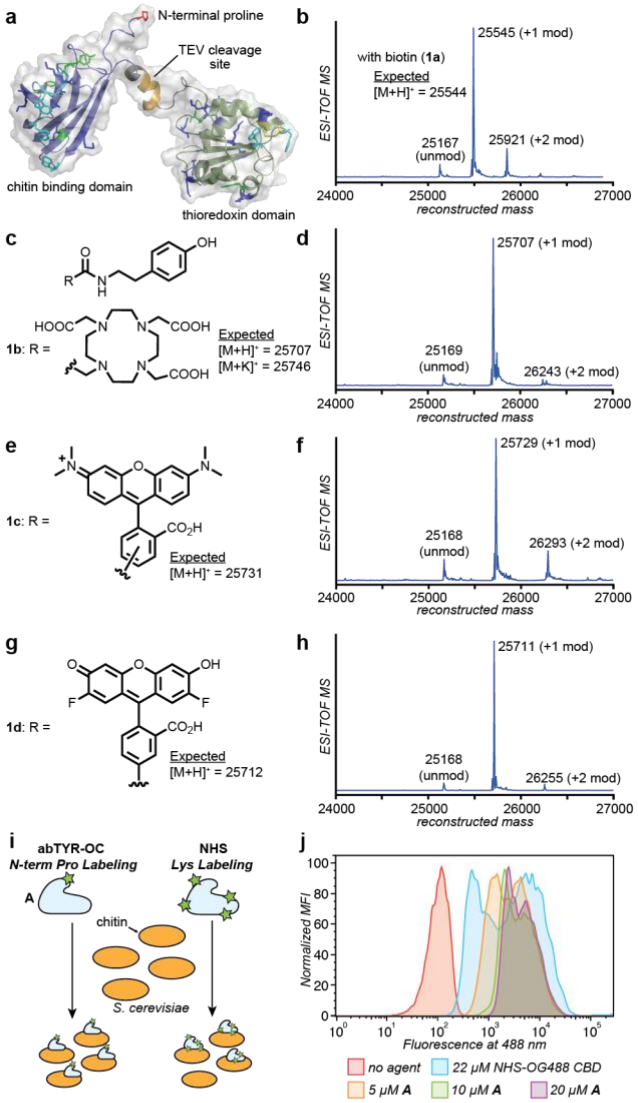


Figure 4. Tyrosinase-based labeling of a chitin binding domain (CBD). (a) A CBD-thioredoxin fusion protein was generated with an added N-terminal proline (red). Other amino acid residues shown include tryptophans (cyan), tyrosines (green), lysines (blue) and a cysteine disulfide (yellow). Structure is based on PBD IDs 2CWR and 3DBX. Using the procedure outlined in Figure 3a, the CBD construct was labeled with (b) biotin-phenol **1a**, (c,d) a DOTA-phenol, (e,f) a rhodamine-phenol, and (g,h) an Oregon Green 488-phenol. See Supporting Information for the specific conditions used in each case. (i) Yeast cells were treated with CBD labeled **1d** using tyrosinase (**A**) or Oregon Green NHS at RT. (j) Flow cytometry was used to quantify the degree of binding, reported as normalized mean fluorescent intensity values.

protein was expressed fused to a C-terminal Thioredoxin A domain to aid in solubility and to increase the production yields. Linking the two proteins was a TEV cleavage site, allowing for future removal of the solubilizing tag, if desired, **Figure 4a**. For the modification of this protein, we synthesized a variety of phenol derivatives from commercially available NHS-ester precursors using the method described above, which included fluorophores with different emission wavelengths and metal-chelating agents, **Figure 4**, compounds **1b-d**.

The panel of phenols was coupled to 10 µM solutions of Pro-CBD-TrxA using 200 nM tyrosinase in 20 mM phosphate buffer at pH 6.5 for 2 h at 4 °C. Biotin phenol **1a** was used at 100 µM, and bulkier substrates **1b-d** were used at 400 µM. To minimize higher order modifications, some derivatives were placed in a water-ice bath, which we found better maintains the colder temperatures as the exothermic coupling proceeds. All phenolic compounds coupled readily to the

proline N-terminus of the CBD construct, with little over modification observed for all compounds, **Figure 4b-h**.

To test the ability of the engineered Pro-CBD-Trx construct as a fungal binding agent, we used abTYR to attach Oregon Green 488-phenol **1d** to the proline N-terminus using abTYR, yielding bioconjugate **A** in **Figure 4g,h**. The labeled product was obtained in high yield (~88% single addition, <5% double addition). Conjugate **A** was then incubated with *S. cerevisiae* cells overnight at room temperature in the absence of light (to prevent photobleaching). Analysis via flow cytometry showed that the Pro-CBD-TRx construct with N-terminal OG488 groups retained its binding ability, **Figure 4i,j**. Furthermore, compared to a CBD that had been non-specifically labeled using Oregon Green 488-NHS (see **Supporting Information Figure S5**), conjugate **A** showed an overall increase in the fluorescently labeled population of yeast and greater homogeneity, **Figure 4j**. These results highlight the ease with which phenol compounds can be used to construct site-specific proline N-terminal bioconjugates that retain proper activity.

Conclusion

Herein we have presented a new bond forming strategy for the convenient site-selective labeling of bioconjugates. Using easily prepared and often commercially available NHS ester precursors, a variety of phenol derivatives of interest can be synthesized without the need for separate purification steps. These derivatives can be oxidized by a commercially available tyrosinase enzyme, after which they couple readily to proline residues introduced in the N-terminal positions of proteins. Compared to NHS-ester chemistry, this coupling achieves excellent regioselectivity, and reaction times are short even at 4 ºC. Given the ease with which N-terminal proline residues can be introduced during protein expression, this method will likely be compatible with a wide range of biomolecular substrates. Finally, its reliance on low-cost and readily storable reaction precursors offers excellent potential for its use in large scale bioconjugate preparation.

ASSOCIATED CONTENT

Supporting Information

Full experimental details, cloning procedures, and protein sequences are provided as Supporting Information. The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

* mbfrancis@berkeley.edu

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation (CHE 1808189), as well as the Chemical Biology Graduate Program at UC Berkeley (NIH T32-GM066698). The studies involving the chitin binding domain were supported by the BASF CARA program. J.C.M. was supported by a UC Berkeley Fellowship for Graduate Studies and the NSF Graduate Research Fellowship Program.

ABBREVIATIONS

abTYR = tyrosinase from *Agaricus bisporus*

sfGFP = superfolder green fluorescent protein

CBD = chitin binding domain

Trx = thioredoxin

CPK = creatine phosphokinase

ALD = aldolase

TMV = tobacco mosaic virus

OG488 = Oregon Green 488 chromophore

REFERENCES

1. O'Hare, H. M., Johnsson, K. & Gautier, A. Chemical probes shed light on protein function. *Curr Opin Struc Biol* **17,** 488–494 (2007).

2. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527,** 110 (2015).

3. Hutsell, S. Q., Kimple, R. J., Siderovski, D. P., Willard, F. S. & Kimple, A. J. Surface Plasmon Resonance. *Methods Mol Biology Clifton N J* **627,** 75–90 (2010).

4. Malmqvist, M. & Karlsson, R. Biomolecular interaction analysis: affinity biosensor technologies for functional analysis of proteins. *Curr Opin Chem Biol* **1,** 378–383 (1997).

5. Witus, L. S. & Francis, M. B. Using Synthetically Modified Proteins to Make New Materials. *Accounts Chem Res* **44,** 774–783 (2011).

6. Alley, S. C., Okeley, N. M. & Senter, P. D. Antibody–drug conjugates: targeted drug delivery for cancer. *Curr Opin Chem Biol* **14,** 529–537 (2010).

7. Baslé, E., Joubert, N. & Pucheault, M. Protein Chemical Modification on Endogenous Amino Acids. *Chem Biol* **17,** 213–227 (2010).

8. Young, T. S. & Schultz, P. G. Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J Biol Chem* **285,** 11039–11044 (2010).

9. Ngo, J. T. & Tirrell, D. A. Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis. *Accounts Chem Res* **44,** 677–685 (2011).

10. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550,** 53 (2017).

11. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angewandte Chemie Int Ed* **48,** 6974–6998 (2009).

12. MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nature Chemical Biology* **11,** 326–331 (2015).

13. Obermeyer, A. C., Jarman, J. B. & Francis, M. B.

N-Terminal Modification of Proteins with *o*-Aminophenols. *Journal of the American Chemical Society* **136,** 9572–9 (2014).

14. Gilmore, J. M., Scheck, R. A., Esser-Kahn, A. P., Joshi, N. S. & Francis, M. B. N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angewandte Chemie Int Ed* **45,** 5307–5311 (2006).

15. Witus, L. S. *et al.* Site-Specific Protein Transamination UsingN-Methylpyridinium-4-carboxaldehyde. *J Am Chem Soc* **135,** 17223–17229 (2013).

16. Theile, C. S. *et al.* Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat Protoc* **8,** nprot.2013.102 (2013).

17. Mahrus, S. *et al.* Global Sequencing of Proteolytic Cleavage Sites in Apoptosis by Specific Labeling of Protein N Termini. *Cell* **134,** 866–876 (2008).

18. Li, X., Zhang, L., Hall, S. E. & Tam, J. P. A new ligation method for N-terminal tryptophan-containing peptides using the Pictet–Spengler reaction. *Tetrahedron Lett* **41,** 4069–4073 (2000).

19. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science* **266,** 776–779 (1994).

20. Sereda, T. J., Mant, C. T., Quinn, A. M. & Hodges, R. S. Effect of the α-amino group on peptide retention behaviour in reversed-phase chromatography Determination of the pKa values of the α-amino group of 19 different N-terminal amino acid residues. *J Chromatogr A* **646,** 17–30 (1993).

21. Hirel, Schmitter, M., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proceedings of the National Academy of Sciences* **86,** 8247–8251 (1989).

22. Xiao, Q., Zhang, F., Nacev, B. A., Liu, J. O. & Pei, D. Protein N-Terminal Processing: Substrate Specificity of Escherichia coli and Human Methionine Aminopeptidases. *Biochemistry-us* **49,** 5588–99 (2010).

23. Obermeyer, A. C., Jarman, J. B., Netirojjanakul, C., El Muslemany, K. & Francis, M. B. Mild Bioconjugation Through the Oxidative Coupling of ortho-Aminophenols and Anilines with Ferricyanide. *Angewandte Chemie Int Ed* **53,** 1057–1061 (2014).

24. ElSohly, A. M. *et al.* ortho-Methoxyphenols as Convenient Oxidative Bioconjugation Reagents with Application to Site-Selective Heterobifunctional Crosslinkers. *J Am Chem Soc* **139,** 3767–3773 (2017).

25. Wang, D. *et al.* Hierarchical Assembly of Plasmonic Nanostructures Using Virus Capsid Scaffolds on DNA Origami Templates. *Acs Nano* **8,** 7896–7904 (2014).

26. Finbloom, J. A. *et al.* Stable Disk Assemblies of a Tobacco Mosaic Virus Mutant as Nanoscale Scaffolds for Applications in Drug Delivery. *Bioconjugate Chem* (2016). doi:10.1021/acs.bioconjchem.6b00424

27. ElSohly, A. M. *et al.* Synthetically Modified Viral Capsids as Versatile Carriers for Use in Antibody-Based Cell Targeting. *Bioconjugate Chem* **26,** 1590–1596 (2015).

28. Capehart, S. L., ElSohly, A. M., Obermeyer, A. C. & Francis, M. B. Bioconjugation of Gold Nanoparticles through the Oxidative Coupling ofortho-Aminophenols and Anilines. *Bioconjugate Chem* **25,** 1888–1892 (2014).

29. Furst, A. L., Smith, M. J. & Francis, M. B. Direct Electrochemical Bioconjugation on Metal Surfaces. *J Am Chem Soc* (2017). doi:10.1021/jacs.7b06385

30. Furst, A. L., Hoepker, A. C. & Francis, M. B. Quantifying Hormone Disruptors with an Engineered Bacterial Biosensor. *Acs Central Sci* **3,** 110–116 (2017).

31. Lee, H., Dellatore, S. M., Miller, W. M. & Messersmith, P. B. Mussel-Inspired Surface Chemistry for Multifunctional Coatings. *Science* **318,** 426–430 (2007).

32. Zhou, J., Wang, C., Wang, P., Messersmith, P. B. & Duan, H. Multifunctional Magnetic Nanochains: Exploiting Self-Polymerization and Versatile Reactivity of Mussel-Inspired Polydopamine. *Chem Mater* **27,** 3071–3076

33. ElSohly, A. M. & Francis, M. B. Development of Oxidative Coupling Strategies for Site-Selective Protein Modification. *Accounts Chem Res* **48,** 1971–1978

34. Grobe, N. *et al.* Mammalian Cytochrome P450 Enzymes Catalyze the Phenol-coupling Step in Endogenous Morphine Biosynthesis. *J Biol Chem* **284,** 24425–24431 (2009).

35. Abell, C. W. Progress in Brain Research. **65,** 139–152 (1986).

36. Basheer, L. & Kerem, Z. Interactions between CYP3A4 and Dietary Polyphenols. *Oxid Med Cell Longev* **2015,** 854015 (2015).

37. Land, E. J., Ramsden, C. A. & Riley, P. A. Tyrosinase Autoactivation and the Chemistry ofortho-Quinone Amines†. *Accounts Chem Res* **36,** 300–308 (2003).

38. Chang, T.-S. An Updated Review of Tyrosinase Inhibitors. *International Journal of Molecular Sciences* **10,** 2440–2475 (2009).

39. Goldfeder, M., Kanteev, M., Isaschar-Ovdat, S., Adir, N. & Fishman, A. Determination of tyrosinase substrate-

binding modes reveals mechanistic differences between type-3 copper proteins. *Nat Commun* **5,** (2014).

40. Ismaya, W. T. *et al.* Crystal structure of Agaricus bisporus mushroom tyrosinase: identity of the tetramer subunits and interaction with tropolone. *Biochemistry* **50,** 5477–86 (2011).

41. Valero, E., Varón, R. & García-Carmona, F. Tyrosinase-Mediated Oxidation of Acetaminophen to 4-Acetamido-o- Benzoquinone. *Biological Chemistry* **383,** 1931–1939 (2005).

42. Xu, D.-Y. & Yang, Z. Cross-linked tyrosinase aggregates for elimination of phenolic compounds from wastewater. *Chemosphere* **92,** 391–398 (2013).

43. Pezzella, A., Lista, L., Napolitano, A. & d'Ischia, M. Tyrosinase-Catalyzed Oxidation of 17β-Estradiol: Structure Elucidation of the Products Formed beyond Catechol Estrogen Quinones. *Chem Res Toxicol* **18,** 1413–1419 (2005).

44. Bruins, J. *et al.* Inducible, Site-specific Protein Labelling by Tyrosine Oxidation–Strain-Promoted (4+2) Cycloaddition. *Bioconjugate Chem* (2017). doi:10.1021/acs.bioconjchem.7b00046

45. Montanari, E. *et al.* Tyrosinase-assisted bioconjugation. A versatile approach to chimeric macromolecules. *Bioconjugate Chem* **29,** 2550–2560 (2018).

46. Yang, J., Saggiomo, V., Velders, A. H., Stuart, M. A. & Kamperman, M. Reaction Pathways in Catechol/Primary Amine Mixtures: A Window on Crosslinking Chemistry. *PLOS ONE* **11,** e0166490 (2016).

47. Jacobsen, M. T., Fairhead, M., Fogelstrand, P. & Howarth, M. Amine Landscaping to Maximize Protein-Dye Fluorescence and Ultrastable Protein-Ligand Interaction. *Cell Chem Biol* **24,** 1040-1047.e4 (2017).

48. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nature Biotechnology* **30,** 1143–1148 (2012).

49. Fung, E. *et al.* Protein Arrays. **278,** 085–100 (2004).

50. Espín, J., Jolivet, S. & Wichers, H. J. Kinetic Study of the Oxidation of γ-l-Glutaminyl-4-hydroxybenzene Catalyzed by Mushroom (Agaricus bisporus) Tyrosinase. *J Agr Food Chem* **47,** 3495–3502 (1999).

51. Hartman, E. C. *et al.* Quantitative characterization of all single amino acid variants of a viral capsid-based drug delivery vehicle. *Nat Commun* **9,** 1385 (2018).

52. Brauer, D.D. *et al.* Systematic Engineering of a Protein Nanocage for High-yield, Site-specific Modification. (*Preprint, under review*). DOI: 10.26434/chemrxiv.7150097

53. Nakamura, T. *et al.* Tertiary Structure and Carbohydrate Recognition by the Chitin-Binding Domain of a Hyperthermophilic Chitinase from Pyrococcus furiosus. *J Mol Biol* **381,** 670–680 (2008).

Enzymatic Modification of N-terminal Proline Residues Using Phenol Derivatives

Johnathan C. Maza,[1] *Daniel L.V. Bader,*[1] *Lifeng Xiao,*[1] *Alan M. Marmelstein,*[1] *Daniel Brauer,*[1] *Adel ElSohly,*[1] *Matthew J. Smith,*[1] *and Matthew B. Francis* [1,2,*]

[1]Department of Chemistry, University of California, Berkeley, California 94720, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratories, Berkeley, California 94720,USA.

*Corresponding author. Email: mbfrancis@berkeley.edu

Supporting Information

Table of Contents

Materials and Methods

General Methods and Instrumentation

All reagents were obtained from commercial sources and used without any further purification. EZ-link NHS-biotin, NHS-TAMRA, and NHS-5'-Oregon Green 488 were purchased from Thermo Fisher (Waltham, MA). DOTA-NHS was purchased from Macrocyclics (Plano, Tx). Tyrosinase isolated from *Agaricus bisporus* (abTYR, both 25 kU [SKU = T3824-25KU] and 50 kU [SKU = T3824-50KU] were used in these studies), aldolase isolated from rabbit muscle, and creatine phosphokinase isolated from rabbit muscle were purchased from Sigma-Aldrich. Spin concentrators with 10 and 100 kDa molecular weight cutoffs (MWCO) and sterile spin filters with 0.22 µm pores were purchased from Millipore (Billerica, MA). Doubly distilled water (ddH_2O) was obtained from a Millipore purification system.

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 MΩ·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 × 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 reconsctructed 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) 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 out as preweighed dry stocks in Eppendorf tubes. These were stored at -20 ºC. Stocks of the enzyme were prepared by dissolving a dry portion of enzyme to yield a 2 mg/mL solution in 50 mM phosphate buffer at pH 6.5 (a ~16.7 µM concnetration of enzyme). The stock solutions were stored at -80 ºC, which was shown to be the optimal storage conditions from an abTYR stability study, shown below. 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.

Disulfide exchange with Ellman's reagent. Free cysteines on Pro-TMV, aldolase, and creatine phosphokinase were protected from potential modification by disulfide formation with Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid), DTNB). To a solution of each protein (50 µL of a 100 µM solution) was added DTNB (1 µL of a 20 mM solution in 100 mM phosphate buffer, pH 7.2 with 1 mM EDTA). The reaction mixture was incubated at rt for 15 min, upon which the solution turned from clear to bright yellow, indicating the protection of reactive cysteine residues. Excess DTNB was removed by repeated (3-6 times) centrifugal filtration against a 100 kDa MWCO membrane. After reaction, the Ellman's reagent protecting group was removed via addition of TCEP (final concentration of 1.9 mM) and incubation at RT for 5 min.

*abTYR-mediated oxidative coupling of biotin-phenol (**1a**) to proline N-termini.* To a 10 µM protein solution in 20 mM phosphate buffer (pH = 6.5) was added the phenol coupling partner to a final concentration of 100 µM. A 2 µM 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 then allowed to proceed in a 4 ºC cold room for 2 h. Reactions were quenched through the addition of TCEP and tropolone (final concentrations of 1.9 mM for both), 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. The resulting samples were analyzed using ESI-TOF MS.

*abTYR-mediated oxidative coupling of bulkier phenols (compounds **1b-c**) to proline N-termini.* To a 10 µM solution of protein in 20 mM phosphate buffer (pH = 6.5) was added the phenol coupling partner to a final concentration of 400 µM. A 2 µM solution of abTYR in 50 mM phosphate buffer (pH = 6.5) was added to reach a final concentration of 200 nM. The reaction was then allowed to proceed on a wet ice bath in a 4 ºC cold room for 2 h. Reactions were quenched with 1.9 mM of TCEP and tropolone and allowed to sit at RT for 1 min before spin concentration against 10 kDa MWCO filters with 10 mM bistTris at pH 7.2. The resulting samples were analyzed using ESI-TOF MS.

*N-terminal labeling of Pro-CBD-TrxA with OG-488-phenol (**A**).* To *a* 10 µM solution of protein was added phenol coupling partner **1d** (400 µM) in 20 mM phosphate buffer (pH = 6.5). To this was then added abTYR to a final concentration of 200 nM. After 2 h, the reaction was loaded onto an Illustra NAP-5 column (GE Healthcare) and eluted with 20 mM phosphate buffer

(pH = 7.0). The absorbance values of the fractions at 280 and 496 nm were used to determine which fractions contained fluorescently labeled protein (see protocol below). ESI-TOF MS data showed that the N-terminally OG-488 labeled CBD was obtained with 92% conversion (see **Figure 4h**).

Oregon Green-488-NHS ester labeling of Pro-CBD-TrxA. Oregon Green 488 (50 mM stock solution in DMF, 5 mM final concentration, 44 eq.) was added to Pro-CBD-TrxA solution (final concentration of 115 μM) in 100 mM HEPES buffer, containing 200 mM NaCl at pH 8.5. The reaction mixture was vortexed briefly and incubated at room temperature for 2 h. Next, excess Oregon Green 488 was removed by passing the reaction mixture through a NAP-5 column equilibrated with 100 mM HEPES buffer, with 200 mM NaCl (pH = 7.0). The Pro-CBD-TrxA dye conjugate was further concentrated by centrifugal filtration with a 500 μL Amicon Ultra-4 spin concentrator (EMD Millipore, Burlington, MA) with a molecular weight cutoff off (MWCO) of 10 kDa against 100 mM sodium phosphate buffer at pH 8.5. The protein dye conjugate was obtained at a final concentration of 55.8 μM. This was calculated with the following equation (from the ThermoFisher, Protein Labelling Kit), where 0.12 represents the correction factor accounting for 280 nm dye absorption, and 41,490 $cm^{-1}M^{-1}$ is the molar extinction coefficient of P-CBD-TrxA.

$$\text{protein concentration (M)} = ([A_{280} - (A_{496} \times 0.12)] \times \text{dilution factor})/41490$$

The conversion efficiency was determined by ESI-TOF-LC/MS (see **Supplementary Information Figure S4**).

Binding of yeast cells using OG-488-labeled CBD-TrxA. Saccharomyces cerevisiae BY4743 were cultured in a medium of yeast extract, peptone, and dextrose (YPD media) at 30 ºC for 20 h (OD600= 2.23) prior to use. A 1 mL portion of the culture medium was transfered to a clean Eppendorf tube. The cells were pelleted by centrifugation (Eppendorf Spin, 3200 rpm, 1 min), resuspended with Dulbecco's phosphate buffered saline (DPBS, pH=7.4), and pelleted again by centrifugation using the same speed. After washing with DPBS for a total of three times, the cells were resuspended in 1 mL of DPBS buffer. The resulting yeast sample was portioned into Eppendorf tubes (10 μL for each), and then treated with either NHS-OG488 labeled P-CBD-TrxA at 22 μM, OG488-phenol/abTYR labeled Pro-CBD-TrxA (**A**) at 5 μM, 10 μM, or 20 μM, or no agent. The samples were incubated at room temperature for 1 d in the absence of light. After incubation, the cells were washed with DPBS three times and then resuspended in PBS buffer. The cells were analyzed on a LSR Fortessa X20 flow cytometer equipped with a 488 nm laser (Flow Cytometry Facility at UC Berkeley). For each sample 20,000 cells were counted.

Tyrosinase (abTYR) Activity Assay. The following procedure was modified from Yoon *et al.*[1] 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 μL of 100 mM phosphate buffer (pH 6.5), 92.6 μL of 2.7 mM L-Tyrosine, and 695.4 μL of Milli-Q water. A 12.0 μL portion of the abTYR aliquot was added with thorough mixing to start the reaction (final concentrations: 20 mM phosphate, 250 μM 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 second intervals around each timepoint between 0:16 and 9:45. The maximum slope typically occurred between 5 and 7 minutes, and was converted to activity units (U or μmol substrate / min) using the molar extinction coefficient for dopachrome of 3600 $M^{-1}cm^{-1}$.[2] 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.

Tyrosinase (abTYR) Stability Study. All solutions were pre-chilled and kept on ice. A solid sample of tyrosinase was dissolved in enough 100 mM phosphate buffer (pH 6.5) to yield a 2.0 mg/mL solution. The resulting solution was divided into four equal aliquots, which were each diluted 1:1 with appropriate buffers to yield the solutions in **SI Table S1**. A 12.0 μL aliquots were stored in separate Eppendorf tubes to avoid subjecting the samples to repeated freeze / thaw cycles. A series of 200 μL portions of samples **E** through **H** were stored in single tubes which were purposely thawed and re-frozen for each activity check to test the effect on stability. The aliquots were stored at either -20 °C or -80 °C. Since buffer composition used here did not affect enzyme stability or activity, the activity of all samples at the different storage temperatures or freeze/thaw conditions were averaged, see **Supporting Information Figure S7**.

Expression of N-terminal Pro-sfGFP. Expression was carried out as previously reported.[3] Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of panel of sfGFP variants. The N-terminal proline residue of the Pro-sfGFP gene in a pBAD vector was mutated to various amino acid residues using the following primers and standard QuickChange protocols:

Gly mutant forward primer: *5'-ggagaaaggtcatgggtcgtaaaggcgaagagcTGTTCACTGG-3'*

Gly mutant reverse primer: *5'-gctcttcgcctttacgacccatgacctttctcctctTTAAGAGC-3'*

Ala mutant forward primer: *5'-gaggagaaaggtcatggcgcgtaaaggcgaagagc-3'*

Ala mutant reverse primer: *5'-gctcttcgcctttacgcgccatgacctttctcctc-3'*

Ser mutant forward primer: *5'-ggagaaaggtcatgtctcgtaaaggcgaagagCTGTTCACTGG-3'*

Ser mutant reverse primer: *5'-gctcttcgcctttacgagacatgacctttctcctctTTAAGAGC-3'*

Met mutant foward primer: *5'-GCTCTTAAAGAggagaaaggtcatgcgtaaaggcgaagagc-3'*

Met mutant reverse primer: *5'-CCAGTGAACAgctcttcgcctttacgcatgacctttctcc-3'*

Following sequence verification, mutants were transformed into DH10B *E. coli* cells for expression. Expression cultures at 100 mL in 2XYT broth were prepared using 5 mL of confluent culture. These were grown at 37 ºC until cultures reached an OD_{600} value between 0.6-0.8, upon which protein expression was induced with a final concentration of 0.5% arabinose. Cells were then shaken overnight at 30 ºC and pelleted. After freezing at -80 ºC, cell pellets were resuspended in 4 mL of an equilibration buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole at pH = 7.4) and 4 mL of BugBuster (Thermo) and then lysed via sonication for 20 min at 70% amplitude. Cells were loaded onto 1 mL of preequilibrated Ni-NTA spin columns (Thermo) and allowed to mix in an end-over-end rotator for 30 min at 4 ºC. After mixing, 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 one resin bed volume 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 ºC until use.

Expression of N-terminal Pro-TMV. Expression and purification was carried out as previously reported.[4] Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Expression of N-terminal Pro-MS2. Expression was carried out as previously reported.[5] Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7 using 10k MWCO spin filters prior to reaction.

Procedure for expression and purification of pCBD-TEV-TrxA protein. The Chitin-Binding Domain (CBD) gene from *Pyrococcus furiosus*[6] and Thioredoxin A (TrxA) were purchased as gene blocks from International DNA Technologies, Inc. (IDT) as follows:

```
CBD from Pyrococcus furiosus:   AGG TCT CAC ATG ACG ACT CCA GTC CCA GTA AGC GGC
TCT CTG GAA GTC AAA GTC AAC GAT TGG GGG TCT GGG GCC GAA TAT GAC GTG ACA TTA AAT
TTA GAC GGT CAG TAC GAC TGG ACA GTA AAG GTG AAA CTT GCT CCG GGG GCA ACC GTG GGC
AGC TTC TGG AGC GCA AAT AAA CAG GAA GGC AAT GGG TAT GTT ATC TTT ACT CCG GTA TCG
TGG AAC AAA GGA CCA ACC GCG ACT TTC GGT TTC ATC GTC AAC GGA CCA CAG GGT GAC AAA
GTC GAG GAG ATC ACC TTA GAG ATT AAC GGC CAG GTT ATT GGA GGC TCA GGA GGC TCG TAA
GTG AGA CCT


Thioredoxin A:  AGG TCT CAC ATG ATG TCG GAT AAA ATT ATC CAT TTA ACG GAT GAT AGT
TTT GAT ACT GAC GTG TTG AAG GCC GAC GGG GCC ATC TTG GTA GAC TTC TGG GCA GAG TGG
TGC GGT CCG TGT AAG ATG ATT GCA CCC ATT TTA GAT GAA ATT GCT GAT GAG TAC CAA GGG
AAG TTG ACA GTT GCT AAA TTG AAT ATC GAC CAG AAT CCA GGG ACC GCA CCC AAA TAC GGC
ATT CGT GGA ATC CCT ACA TTG TTA CTT TTC AAG AAT GGG GAG GTC GCT GCC ACT AAG GTT
GGC GCT CTT AGC AAG GGG CAA TTA AAA GAA TTT TTG GAC GCA AAC TTG GCG TAA GTG AGA
CCT
```

The Thioredoxin A gene block was first extended and amplified using sequential 5' primers to add a 5'-TEV cleavage site, a GGSGGS flexible linker, and a Golden Gate (BsaI) cutsite:

Forward primer 1: *5'-TTCAGGGTGGAGGCTCAGGAGGCTCGATGTCGGATAAAAT-3'*

Forward primer 2: *5'- AGGTCTCACTCGGAAAACCTGTATTTTCAGGGTGGAGGCTAGGA-3'*

In both of these steps, the 3' end was amplified using a single primer to add a His6 tag for purification and a BsaI cut site for vector insertion:

Reverse primer: *5'-AGGTCTCACTTAATGATGATGATGATGATGCGCCAAGTTTCGTCCAAAAATTCTT-3'*

The Chitin-Binding Domain gene block was extended and amplified, affixing 5' N-terminal proline base pairs (CCT) and Golden Gate (BsaI) cut sites at the 3' and 5' ends:

Forward primer: *5'-AGGTCTCACATGCCTACGACTCCAGTCCCAGTAAGCG -3'*

Reverse primer: *5'-AGGTCTCTCGAGCCTCCTGAGCCTCCAATAAC-3'*

After PCR purification, both DNA inserts were ligated into a pET vector by the standard Golden Gate assembly procedure.[ref] The resulting plasmid was transformed into BL21 star *E. coli* cells via heat shock and streaked onto LB-agar plates

with 1 mM kanamycin. Plasmids were isolated from 5 mL overnight cultures by Qiagen Spin MiniPrep Kit. The plasmid was sequenced to confirm the following:

ATGCCAACGACTCCAGTCCCAGTAAGCGGCTCTCTGGAAGTCAAAGTCAACGATTGGGGGTCTGGGGCCGAATAT–
GACGTGACATTAAATTTAGACGGTCAGTACGACTGGACAGTAAAGGTGAAACTTGCTCCGGGGGCAACCGTGG–
GCAGCTTCTGGAGCGCAAATAAACAGGAAGGCAATGGGTATGTTATCTTTACTCCGGTATCGTGGAACAAAGGAC–
CAACCGCGACTTTCGGTTTCATCGTCAACGGACCACAGGGTGACAAAGTCGAGGAGATCACCTTAGAGATTAACGGC–
CAGGTTATTGGAGGCTCAGGAGGCTCGgaaaacctgtattttcagggtGGAGGCTCAGGAGGCTCGATGTCGGATA–
AAATTATCCATTTAACGGATGATAGTTTTGATACTGACGTGTTGAAGGCCGACGGGGCCATCTTGGTAGACTTCTG–
GGCAGAGTGGTGCGGTCCGTGTAAGATGATTGCACCCATTTTAGATGAAATTGCTGATGAGTACCAAGGGAAGTTG–
ACAGTTGCTAAATTGAATATCGACCAGAATCCAGGGACCGCACCCAAATACGGCATTCGTGGAATCCCTACATTGT–
TACTTTTCAAGAATGGGGAGGTCGCTGCCACTAAGGTTGGCGCTCTTAGCAAGGGGCAATTAAAAGAATTTTTGGAC–
GCAAACTTGGCGcatcatcatcatcatcat

Which corresponds to the following amino acid sequence:

(M)PTTPVPVSGSLEVKVNDWGSGAEYDVTLNLDGQYDWTVKVKLAPGATVGSFWSANKQEGNGYVI-
FTPVSWNKGPTATFGFIVNGPQGDKVEEITLEINGQVIGGSGGSENLYFQGGGSGGSMSDKIIHLTDDSF-
DTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKN-
GEVAATKVGALSKGQLKEFLDANLAHHHHHH

A 5 mL overnight culture was grown in Luria-Burtani (LB) media containing 1 mM kanamycin at 37 °C. A 1 mL portion was then added to 1 L of LB media and grown at 37 °C to an OD600 of 0.6. Expression was induced with a final concentration of 1 mM IPTG. After 20 h the cells were collected by centrifugation at 5000*x*g for 20 min at 4 °C. The cells were re-suspended in 24 mL of 50 mM HEPES buffer (pH = 8.5) containing 500 mM NaCl, 1 mM protease inhibitor cocktail (composed of leupeptin, aprotinin, and phenylmethylsulfonylfluoride) and sonicated for 15 min on ice at 50% amplitude. Lysed cells were centrifuged at 8000*x*g at 4 °C for 30 min. Next, the supernatant was loaded on 3 mL of Ni-NTA resin and shaken for 1 h at 4 °C. Resin was washed with 50 mM HEPES buffer (pH = 8.5) containing 500 mM NaCl and 25 mM imidazole (3 x 8 mL), and the protein was eluted with 50 mM HEPES buffer (pH = 8.5) containing 500 mM NaCl and 250 mM imidazole. The fractions containing pCBD-TEV-TrxA were collected and concentrated using Amicon Ultra 10 kD MWCO centrifugal concentrators (MilliporeSigma). Purified protein was spin concentrated into 20 mM phosphate buffer at pH 7.0 using 10k MWCO spin filters prior to reaction.

Synthetic Procedures

*Synthesis biotin-phenol (**1a**).* To a 40 mM solution of tyramine (1 eq, Sigma Aldrich) in dry DMF was added E-Z link biotin-NHS (1.1 eq, Thermo) to a final concentration of 44 mM. The reaction was vortexed overnight and analyzed by LC-MS to confirm amide bond formation. To hydrolyze any remaining NHS ester, portions of this reaction solution were diluted with water to a final concentration of 2 mM and allowed to stand for 48 h before use in subsequent labeling reactions. ESI-Q-TRAP MS: *m/z* calc'd for $C_{18}H_{26}N_3O_3S$ $[M+H]^+$ 364.16, observed 364.7.

*Synthesis DOTA-phenol (**1b**).* To a 4 mM solution of tyramine (1 eq, Sigma Aldrich) in dry DMF was added DOTA-NHS (1.1 eq, Macrocyclics) to a final concentration of 4.4 mM and triethylamine (10 eq, Sigma Aldrich) to a final concentration of 40 mM. The reaction was vortexed overnight and anlyzed by LC-MS to confirm amide bond formation. No further hydrolysis was necessary. Portions of this reaction solution were diluted with DMF to a final concentration of 2 mM and used for subsequent labeling reactions. ESI-Q-TRAP MS: *m/z* calc'd for $C_{24}H_{38}N_5O_8$ $[M+H]^+$ 524.26, observed 524.

*Synthesis rhodamine-phenol (**1c**).* To a 40 mM of tyramine (1 eq, Sigma Aldrich) in dry DMF was added of NHS -Rhodamine (5/6-carboxy-tetramethyl-rhodamine succinimidyl ester, 1.1 eq, Thermo) to a final concentartion of 44 mM. The reaction was vortexed overnight and anlyzed by LC-MS to confirm amide bond formation. No further hydrolysis was necessary. Portions of this reaction solution were diluted with DMF to a final concentration of 2 mM and used for subsequent labeling reactions. ESI-Q-TRAP MS: *m/z* calc'd for $C_{33}H_{32}N_3O_5^+$ $[M]^+$ 550.23, observed 550.

*Synthesis Oregon Green-488-phenol (**1d**).* To a 4 mM solution of tyramine (1 eq, Sigma Aldrich) in dry DMF was added Oregon Green-488-NHS (Oregon Green™ 488 Carboxylic Acid, Succinimidyl Ester, 5-isomer, 1.1 eq, Thermo) to a final concentration of 4.4 mM. The reaction was vortexed overnight and anlyzed by LC-MS to confirm amide bond formation. No further hydrolysis was necessary. Portions of this reaction solution were diluted with DMF to a final concentration of 2 mM and used for subsequent labeling reactions. ESI-Q-TRAP MS: *m/z* calc'd for $C_{29}H_{26}F_2NO_7$ $[M+H]^+$ 532.11, observed 532.

Supporting Information References

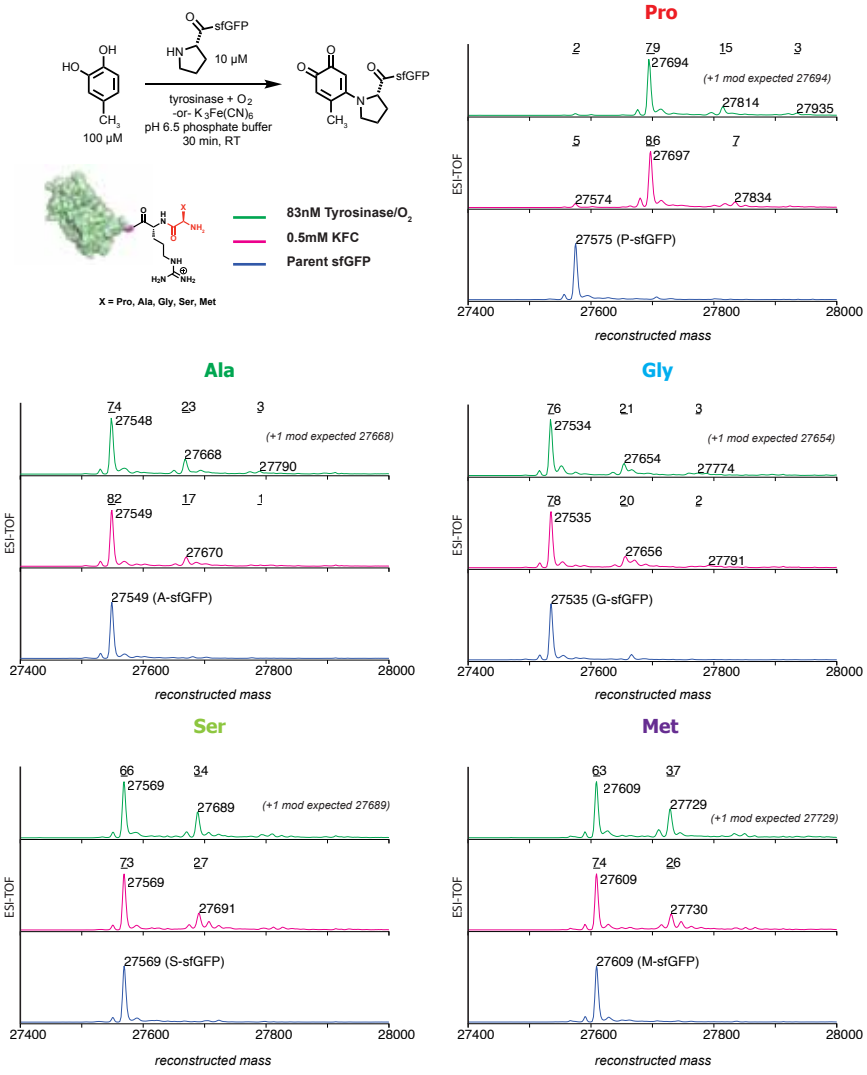
1. Baek, Y. et al. Tyrosinase inhibitory effects of 1,3-diphenylpropanes from Broussonetia kazinoki. Bioorgan Med Chem

17, 35–41 (2009).

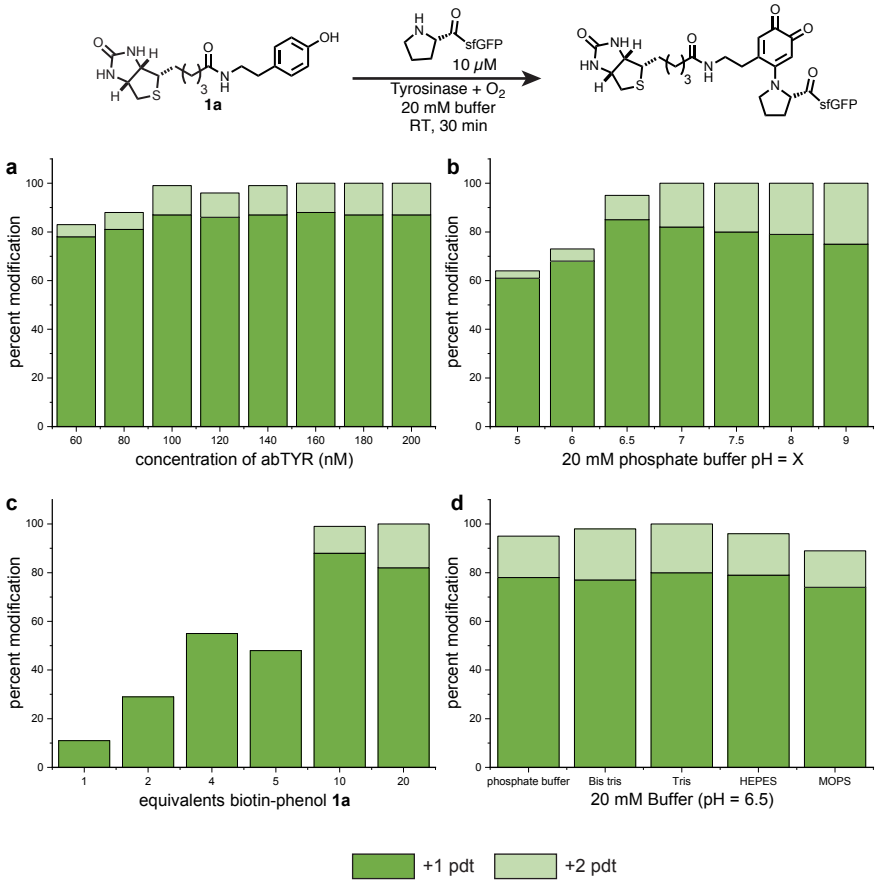
2. FLING, M., HOROWITZ, N. & HEINEMANN, S. The isolation and properties of crystalline tyrosinase from Neurospora. J Biological Chem 238, 2045–53 (1963).
3. Obermeyer, A. C., Jarman, J. B. & Francis, M. B. N Terminal Modification of Proteins with o Aminophenols. Journal of the American Chemical Society 136, 9572–9 (2014).
4. Finbloom, J. A. et al. Stable Disk Assemblies of a Tobacco Mosaic Virus Mutant as Nanoscale Scaffolds for Applications in Drug Delivery. Bioconjugate Chem (2016). doi:10.1021/acs.bioconjchem.6b00424
5. Brauer, D.D. et al. Systematic Engineering of a Protein Nanocage for High-yield, Site-specific Modi-fication. (Preprint, under review). DOI: 10.26434/chemrxiv.7150097
6. Nakamura, T. et al. Tertiary Structure and Carbohydrate Recognition by the Chitin-Binding Domain of a Hyperthermophilic Chitinase from Pyrococcus furiosus. J Mol Biol 381, 670–680 (2008).

series	storage temp (ºC)	freeze/thaw*	buffer	glycerol
A	-20	N	50 mM phosphate, pH 6.5	15%
B	-20	N	50 mM phosphate, pH 6.5	25%
C	-20	N	50 mM phosphate, pH 6.5	50%
D	-20	N	50 mM phosphate, pH 6.5, 150 mM NaCl	25%
E	-80	N	50 mM phosphate, pH 6.5	15%
F	-80	N	50 mM phosphate, pH 6.5	25%
G	-80	N	50 mM phosphate, pH 6.5	50%
H	-80	N	50 mM phosphate, pH 6.5, 150 mM NaCl	25%
E-F/T*	-80	Y	50 mM phosphate, pH 6.5	15%
F-F/T*	-80	Y	50 mM phosphate, pH 6.5	25%
G-F/T*	-80	Y	50 mM phosphate, pH 6.5	50%
H-F/T	-80	Y	50 mM phosphate, pH 6.5, 150 mM NaCl	25%

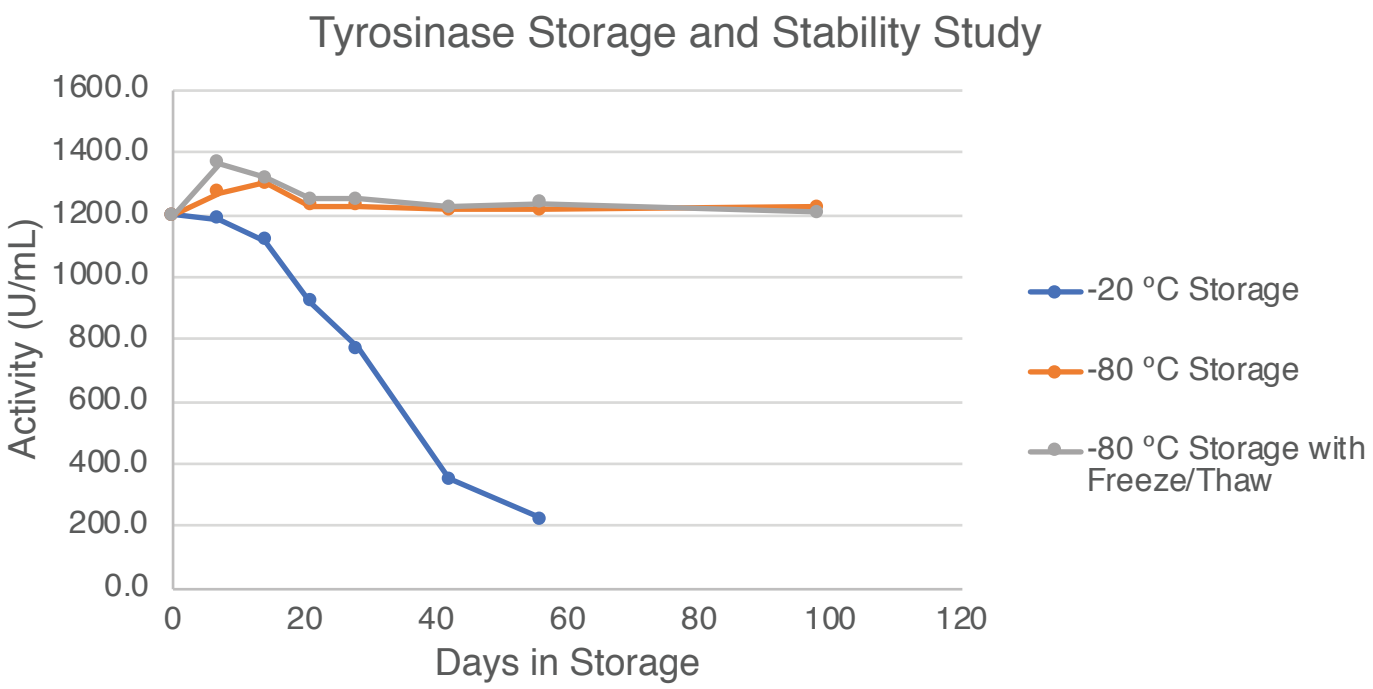
SI Table S1. Sample conditions for tyrosinase (abTYR) stability studies.



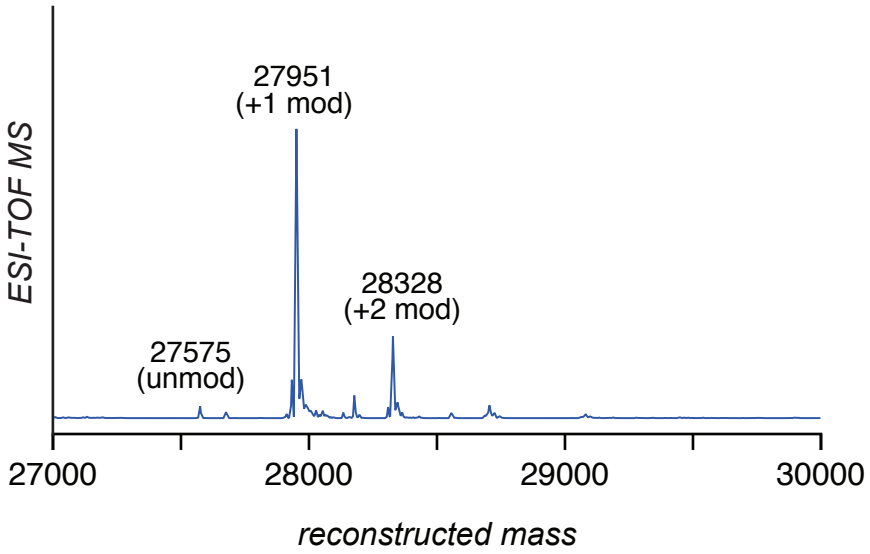
Supporting Information Figure S1. Modification of a panel of sfGFP substrates with various N-terminal residues. All reactions contained 10 µM protein mixed with a final concentration of 100 µM 4-methylcatechol, and were dissolved in 20 mM phosphate buffer at pH 6.5. To initiate each reaction, abTYR or $K_3Fe(CN)_6$ was added at final concentrations of 83 nM or 0.5 mM, respectively. After 30 min at room temp, reactions were analyzed by ESI-TOF MS. Complete conversion of the starting material was observed only when the N-terminal residue was proline. Otherwise, reaction conversions were typically 40% or less.



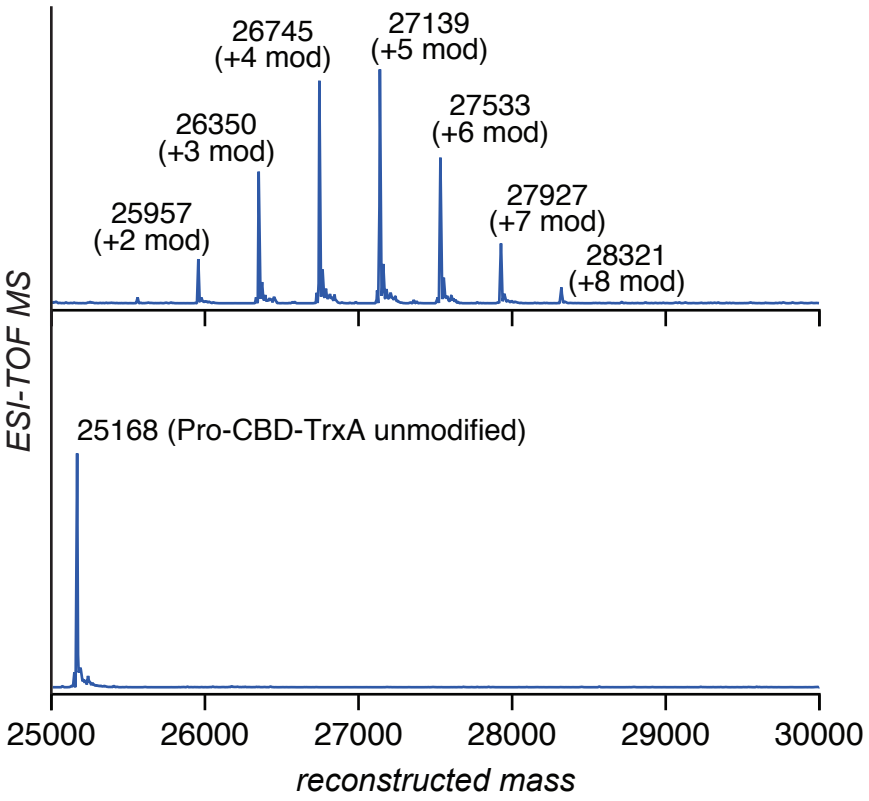
Supporting Information Figure S2. Reaction optimization for site-selective protein biotinylation on proline N-termini. (a) To explore the effects of enzyme concentration on the reaction, a reaction mix containing 10 μM Pro-sfGFP in 20 mM phosphate buffer at pH 6.5 was prepared with **1a** at a final concentration of 100 μM. The final concentration of abTYR was varied from 60 nM to 200 nM, and the reactions were allowed to proceed at room temp for 30 min prior to analysis. (b) To explore the effects of pH on the reaction, a reaction mix containing 10 μM Pro-sfGFP in 20 mM phosphate buffer at various pH ranges was prepared with **1a** at a final concentration of 100 μM. The enzyme abTYR was added at 160 nM, and reactions were allowed to proceed at room temp for 30 min prior to analysis. (c) To explore the effect that differing equivalencies of **1a** had on the reaction, a reaction mix containing 10 μM Pro-sfGFP in 20 mM phosphate buffer at pH 6.5 was prepared with **1a** at final concentrations ranging from 10 μM to 200 μM. The enzyme abTYR was added at 160 nM, and reactions were allowed to proceed at room temp for 30 min prior to analysis. (d) To explore the effects of various buffers on the reaction, a reaction mix containing 10 μM Pro-sfGFP in 20 mM of various buffers at pH 6.5 was prepared with **1a** at a final concentration of 100 μM. The enzyme abTYR was added at 200 nM, and reactions were allowed to proceed at room temp for 30 min prior to analysis.



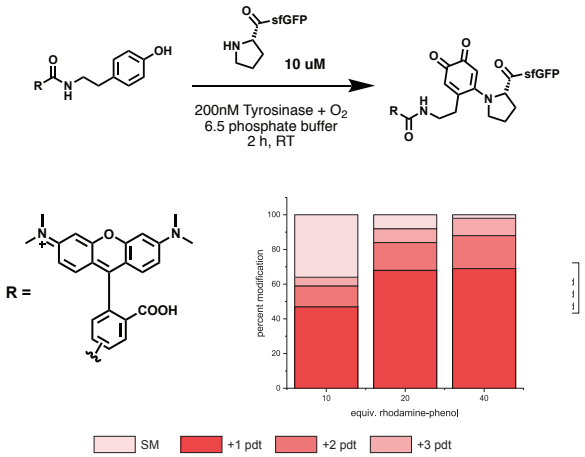
Supporting Information Figure S3. Stability study to determine optimal storage conditions for tyrosinase (abTYR). Analysis of enzyme activity over multiple weeks shows that -80 ºC is optimal for the storage of enzyme aliquots.



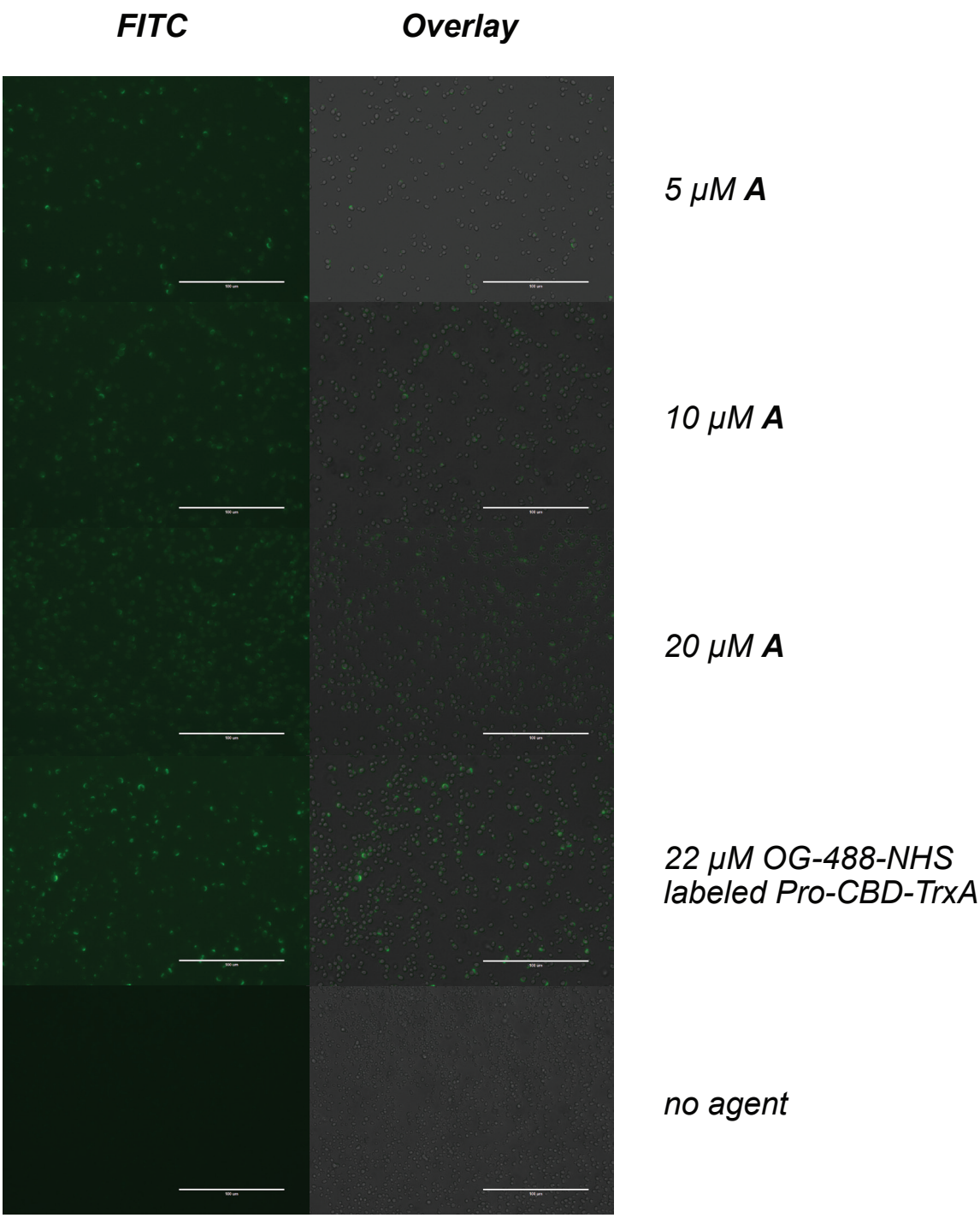
Supporting Information Figure S4. Site-selective protein biotinylation at 37 ºC. To 10 µM of Pro-sfGFP was added **1a** to a final concentration of 100 µM in 20 mM phosphate buffer at pH 6.5. To this was added 200 nM abTYR and the reaction was placed in a 37 ºC bath. ESI-TOF MS analysis at 15 min showed that the reaction had reached completion.



Supporting Information Figure S5. Modification of Pro-CBD-TrxA using an Oregon-Green 488-NHS ester. After reaction with the NHS-ester and purification wtih a NAP-5 column, Pro-CBD-TrxA is fully converted to a heterogenously labeled OG488 product, as indicated by the appearance of expected MW peaks in the ESI-TOF MS trace.



Supporting Information Figure S6. Reaction optimization for bulkier phenol substrates. Rhodamine-phenol 1c was synthesized using tyramine and 1.1 equivalents of an NHS-TAMRA in DMF. Reaction conditions were optimized for the attachment of this substrate to the N-terminus of Pro-sfGFP. Reaction conversion was quantified using ESI-TOF MS. Complete conversion was only acheived when using 40 equivalents of **1c**.



Supporting Information Figure S7. Fluorescent labeling of yeast using OG-488 labeled CBD. Pro-CBD-TrxA was N-terminally modified using abTYR and an OG-488-phenol (**1d**), yielding compound **A**. As a comparison, OG-488-NHS was used to label the surface lysines on Pro-CBD-TrxA non-specifically. After incubation, *S. cerevisiae* cells were imaged using fluorescence microscopy. Fluorescence images (488 nm channel) appear on the left, and flurescence merged with bright field images appear on the right. Scale bars represent 100 μm.