pubs.acs.org/acschemicalbiology

Articles

Efficient Enzymatic Incorporation of Dehydroalanine Based on SAMDI-Assisted Identification of Optimized Tags for OspF/SpvC

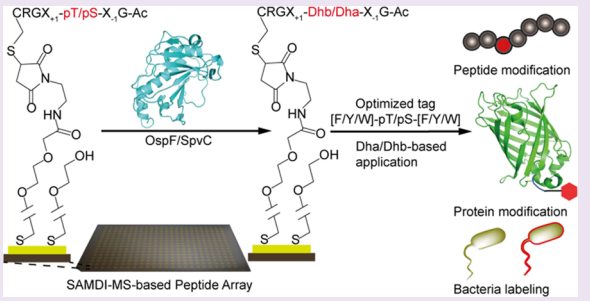
Anming Yang, Huanyu Tao, Lindsey C. Szymczak, Liang Lin, Junfeng Song, Yi Wang, Silei Bai, Justin Modica, Sheng-You Huang,* Milan Mrksich,* and Xinxin Feng*

Cite This: *ACS Chem. Biol.* 2022, 17, 414–425

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Site-specific modification of proteins has important applications in biological research and drug development. Reactive tags such as azide, alkyne, and tetrazine have been used extensively to achieve the abovementioned goal. However, bulky side-chain "ligation scars" are often left after the labeling and may hinder the biological application of such engineered protein products. Conjugation chemistry via dehydroalanine (Dha) may provide an opportunity for "traceless" ligation because the activated alkene moiety on Dha can then serve as an electrophile to react with radicalophile, thiol/amine nucleophile, and reactive phosphine probe to introduce a minimal linker in the protein post-translational modifications. In this report, we present a mild and highly efficient enzymatic approach to incorporate Dha with phosphothreonine/serine lyases, OspF and SpvC. These lyases originally catalyze an irreversible elimination reaction that converts a doubly phosphorylated substrate with phosphothreonine (pT) or phosphoserine (pS) to dehydrobutyrine (Dhb) or Dha. To generate a simple monophosphorylated tag for these lyases, we conducted a systematic approach to profile the substrate specificity of OspF and SpvC using peptide arrays and self-assembled monolayers for matrix-assisted laser desorption/ionization mass spectrometry. The optimized tag, [F/Y/W]-pT/pS-[F/Y/W] (where [F/Y/W] indicates an aromatic residue), results in a ∼10-fold enhancement of the overall peptide labeling efficiency via Dha chemistry and enables the first demonstration of protein labeling as well as live cell labeling with a minimal ligation linker via enzyme-mediated incorporation of Dha.



Downloaded via NORTHWESTERN UNIV on March 22, 2022 at 18:58:08 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Proteins represent the majority of nature's worker biomacromolecules. Post-expression site-specific modification of proteins diversify the structure of proteins, enabling their applications as biological tools as well as therapeutic molecules.[1,2] Modifications such as PPant attachment, lipoylation, and biotinylation are widely used; but these tags do not harbor additional chemical reactivity for next-step chemical modification. Instead, a "tag-and-graft" approach can be generalized as a site-selective protein-labeling method, where a functional group is introduced by modifying the preinstalled chemically reactive "tag" on the protein backbone (Figure 1a). In the past decades, the range of such reactive tags (e.g., azide, alkyne, tetrazine, and formylglycine) has been greatly expanded by chemical,[1] enzymatic[3,4] and genetic methods.[5−9] However, these reactive tags often leave attachment linkages larger than the amino acid residue itself (Figure 1a), and these side-chain "ligation scars" may hamper precise functionalization that is required in probing natural systems and creating therapeutic conjugates.[10,11]

The amino acid dehydroalanine (Dha) provides an opportunity to solve the abovementioned problem by finishing the conjugation with a minimal attachment linkage of a single atom (S, N, Se, or C) (Figure 1a).[12−14] The active alkene moieties on these residues can then serve as electrophiles in Michael addition reactions with thiol and amine nucleophiles,[13] with reactive phosphine probes,[15] or even radicalophiles, to install a full spectrum of authentic post-translational modifications (PTMs) via a "chemical mutagenesis" approach.[12,14,16−18] Therefore, Dha has been used to install multiple important PTMs, including glycosylation,[21] phosphorylation,[22] methylation, and acetylation.[23] In addition, a few recent reports have demonstrated the viability of diastereoselective Michael addition, taking advantage of the chiral protein environment[19] or a Rh-based catalyst.[20] Thus, it may be possible to make stereoselective modification using this Dha strategy with some necessary improvements in the future.

To date, the insertion of the Dha tag into proteins mostly proceeds chemically via installation of precursors,[13] as direct

Received: November 3, 2021
Accepted: January 24, 2022
Published: February 7, 2022

ACS Publications

© 2022 American Chemical Society

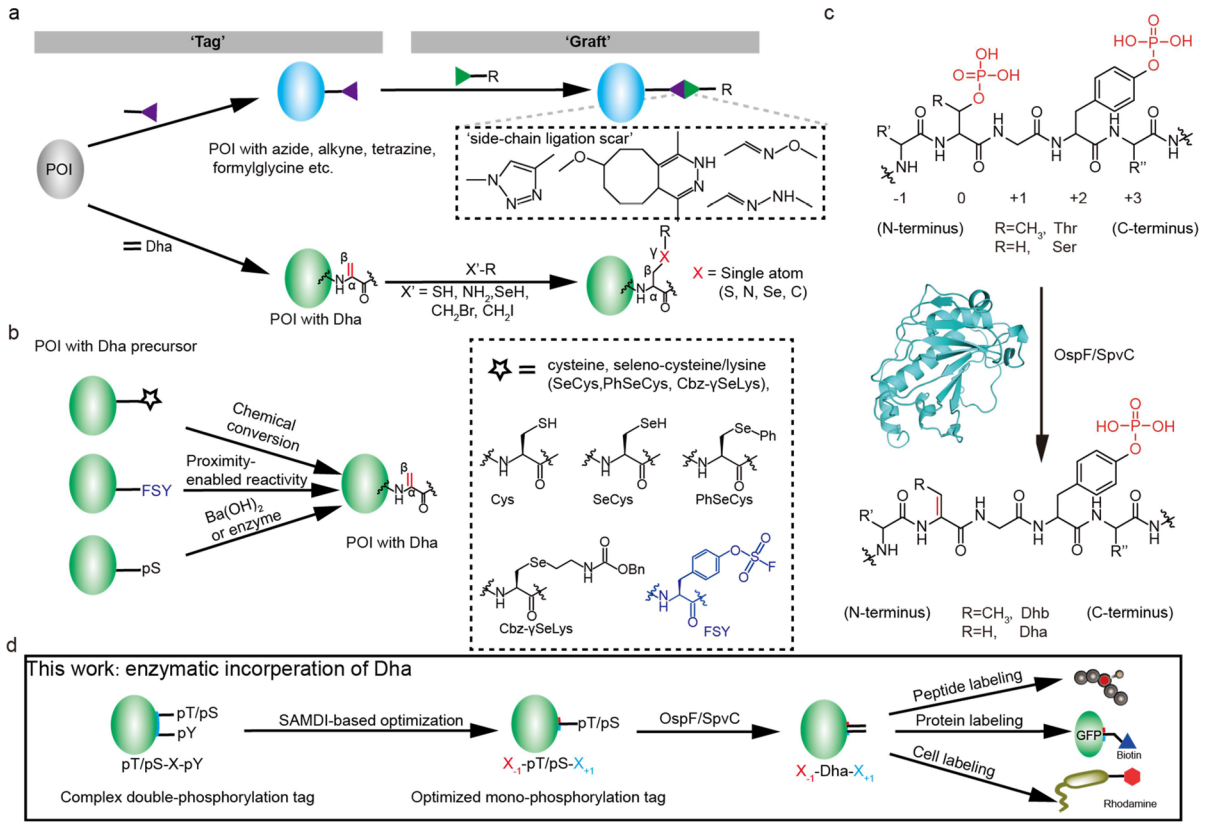


Figure 1. Formation and reactions of Dha and Dhb. (a) "Tag-and-graft" approach is generally used as a site-selective protein-labeling method for protein of interest (POI). (b) Current methods of introducing Dha. Cysteine, seleno-cysteine/lysine, and pS were converted to Dha by 2,5-dibromohexanediamide, H_2O_2, and Ba $(OH)_2$, respectively. (c) Lyases OspF and SpvC eliminate phosphate groups from phosphothreonine/serine residues in substrate proteins and convert them to the corresponding Dhb or Dha products. (d) Our proposed strategy to identify a monophosphorylated tag for OspF and SpvC for the efficient enzymatic incorporation of Dha or Dhb.

insertion of unnatural amino acids Dha is not yet possible due to the interference of its reactive side chain to the protein translation machinery.[8,9] Common chemical precursors of Dha include cysteine, seleno-cysteine/lysine (SeCys, PhSeCys, or Cbz-γSeLys), and phosphorylated serine (pS) (Figure 1b).[13,24] Recently, a proximity-enabled GECCO strategy also enables the site-specific conversion of fluorosulfate-L-tyrosine (FSY)/serine to Dha (Figure 1b).[25] Compared to direct chemical conversion, enzymatic protein labeling, mimicking nature's toolbox of enzymatic PTMs,[26,27] is highly efficient and robust under mild reaction conditions;[26,27] however, enzymatic incorporation of Dha has not been reported thus far, and the phosphothreonine/serine lyases OspF and SpvC are attractive candidates for such applications.

OspF and SpvC, two phosphothreonine/serine lyases secreted by *Shigella flexneri* and *Salmonella typhimurium*, respectively, can modify the host mitogen activated protein kinases (MAPKs, e.g., p38, JNK, and Erk1/2) through *αβ*-elimination reaction of the phosphate group on phosphothreonine (pT) or phosphoserine (pS) to yield Dhb or Dha, respectively.[28,29] The use of OspF- or SpvC-mediated generation of the Dha residue, however, has not been straightforward, in part because the native substrate sequences are constrained to the consensus pT/pS-X-pY motif (Figure 1c).[28,30,31] The pY on +2 position contributes significantly (3–10 fold difference in catalytic efficiency) to the substrate recognition of the lyases.[28] Such doubly phosphorylated recognition sequences create substantial difficulty in tag incorporation in target protein. Therefore, we sought to develop a monophosphorylated substrate for OspF and SpvC to provide a practical enzymatic route for site-specific labeling of proteins.

In this paper, we report our approach to identify a monophosphorylated tag for OspF/SpvC via substrate specificity profiling using self-assembled monolayers for matrix-assisted laser-desorption–ionization mass spectrometry (SAMDI-MS) and peptide arrays (Figure 1d). SAMDI-MS uses self-assembled monolayers of long-chain alkanethiolates on a gold-coated surface, to which peptide substrates are immobilized (Figure 2a). As it is a quantitative method compatible with high-throughput 384 and 1536 spot formats, the SAMDI-MS technique is well-suited for profiling enzyme activity and has been successfully used to characterize the activities of a broad range of enzymes, including phosphatases,[32] proteases,[33] deacetylases,[34] acetylases,[35] glycosyltransferases,[36] and cytochrome P411.[37] Here, we use SAMDI-MS to profile the activity of OspF and SpvC on peptide substrates that have a pS or pT residue flanked by all possible amino acids (except cysteine) at the ±1 positions. This approach revealed monophosphorylated tags that display comparable activity to native di-phosphorylated substrates in generating Dha and Dhb. The structure–activity relationship analysis as well as molecular docking studies indicate that hydrophobic and aromatic interactions are the two major determinants for the observed substrate specificity. We also successfully demonstrate that, with the optimized monophosphorylated tag for OspF, the efficiency of the Dha-based peptide labeling is enhanced 10-fold. Finally, we report the first examples of

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414–425

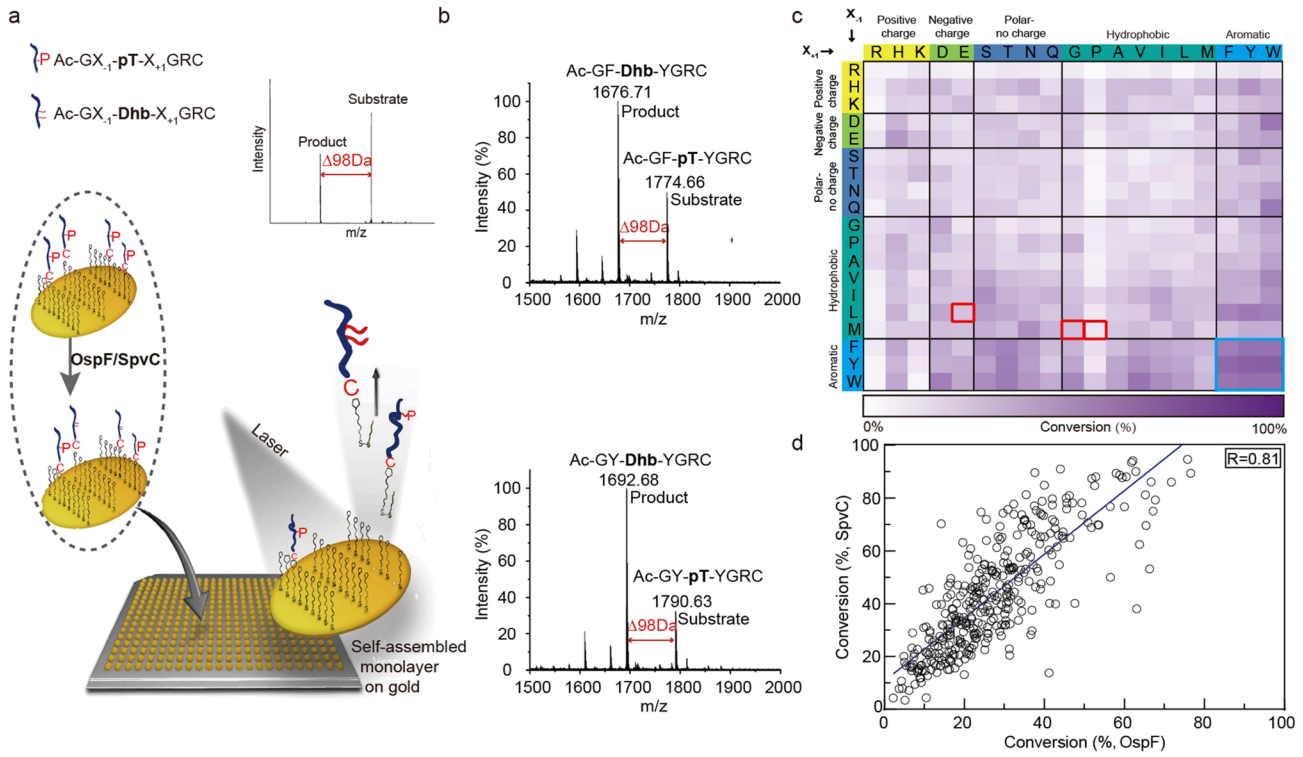


Figure 2. Substrate specificity profiling for OspF and SpvC. (a) Overview of SAMDI-based activity characterization of OspF and SpvC. Peptide solutions from a Ac-GX$_{-1}$-pT-X$_{+1}$GRC library were individually transferred to monolayer array plates where the peptide substrates were immobilized to maleimide-terminated monolayers. These peptide-functionalized spots were then treated with OspF/SpvC to give products. In SAMDI-MS, the monolayer is irradiated with a laser that dissociates the thiolate−gold bond, resulting in the release of ionized alkanethiolate−peptide species from the surface. (b) Examples of SAMDI spectra for enzyme activity quantification. A mass shift of −98 Da indicates the formation of elimination products. Lyase activity is determined from characteristic mass shifts and relative peak intensities in the MS spectrum. (c) Heatmap of OspF activity profiling with the Ac-GX$_{-1}$-pT-X$_{+1}$GRC library. The red boxes represent X$_{-1}$ and X$_{+1}$ of the three natural substrates. The cyan box indicates that OspF prefers aromatic amino acids for the X$_{-1}$ and X$_{+1}$ positions. (d) Correlation of conversion between OspF and SpvC. Each data point represents one peptide sequence from the Ac-GX$_{-1}$-pT-X$_{+1}$GRC library. Numeric data for these experiments can be found in Figure S6.

protein and live bacteria labeling via enzyme-mediated Dha incorporation.

RESULTS AND DISCUSSION

Substrate Specificity Profiling for OspF and SpvC. We first expressed and purified OspF and SpvC (Figure S1)[28,30] and confirmed their activities using the di-phosphorylated substrate (Erk2 analogue, P1 = CDHTGFL-pT-E-pY-VATR) and monophosphorylated substrate (P38 analogue, P2 = CDEM-pT-GYV) with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), where a mass shift of 98 Da confirmed the formation of elimination products (Figures S2 and S3). We also performed a kinetic characterization with the EnzChek assay[38] for native di-phosphorylated and monophosphorylated substrates (Table S2) and found kinetic parameters that were consistent with previous studies[28] (Figure S4).

We next designed the peptide array needed for profiling the substrate specificity of the two enzymes. It is known that tyrosine on the +2 position is highly conserved and critical for activity among the MAPK family homologues.[28,31] Also, the −1 and +1 positions of the peptide substrate are reported to be critical for enzymatic activity: for example, placement of alanine at the −1 position resulted in no detectable enzymatic activity, and substitutions at the +1 position had the greatest impact on activity for substitutions at the −3 to +3 positions.[31] Therefore, we reasoned that by profiling the specificity of ±1 positions flanking pT/pS, we could identify a more efficient monophosphorylated tag for enzymatic incorporation of Dhb or Dha by OspF/SpvC. We next prepared a peptide array having 361 peptides with the sequence Ac-GX$_{-1}$-pT-X$_{+1}$GRC, where X$_{+1}$ and X$_{-1}$ were one of 19 amino acids, with Cys being excluded to avoid interference with the immobilization reaction. A library with pT is selected, as the pT substrates give higher overall activities with OspF and thus better signal-to-noise ratios in the specificity study. The peptides (2 μL, 20 μM) were transferred onto array plates having 384 gold islands that were modified with a monolayer-presenting maleimide groups at a density of approximately 10% mixed with a tri(ethylene glycol)-terminated alkanethiolate.[39] The peptides were immobilized through reaction of the terminal cysteine groups with the maleimide groups. The tri(ethyleneglycol) groups prevent nonspecific adsorption of the enzymes.[39] We then applied the OspF/SpvC enzyme (0.1 μM, 2 μL) to each island and allowed the reaction to proceed for 2 h. The array plates were rinsed, treated with matrix, and analyzed by MALDI-TOF to obtain a mass spectrum for each spot (Figure 2a). The lyase activity was calculated by measuring the relative intensity (I) of each molecular ion peak (M^+) (Figures 2b and S5) from each spectrum using the following equation

$$\text{Conversion (\%)} = \frac{I(\text{Product})}{I(\text{Product}) + I(\text{Substrate})} \times 100\%$$

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414−425

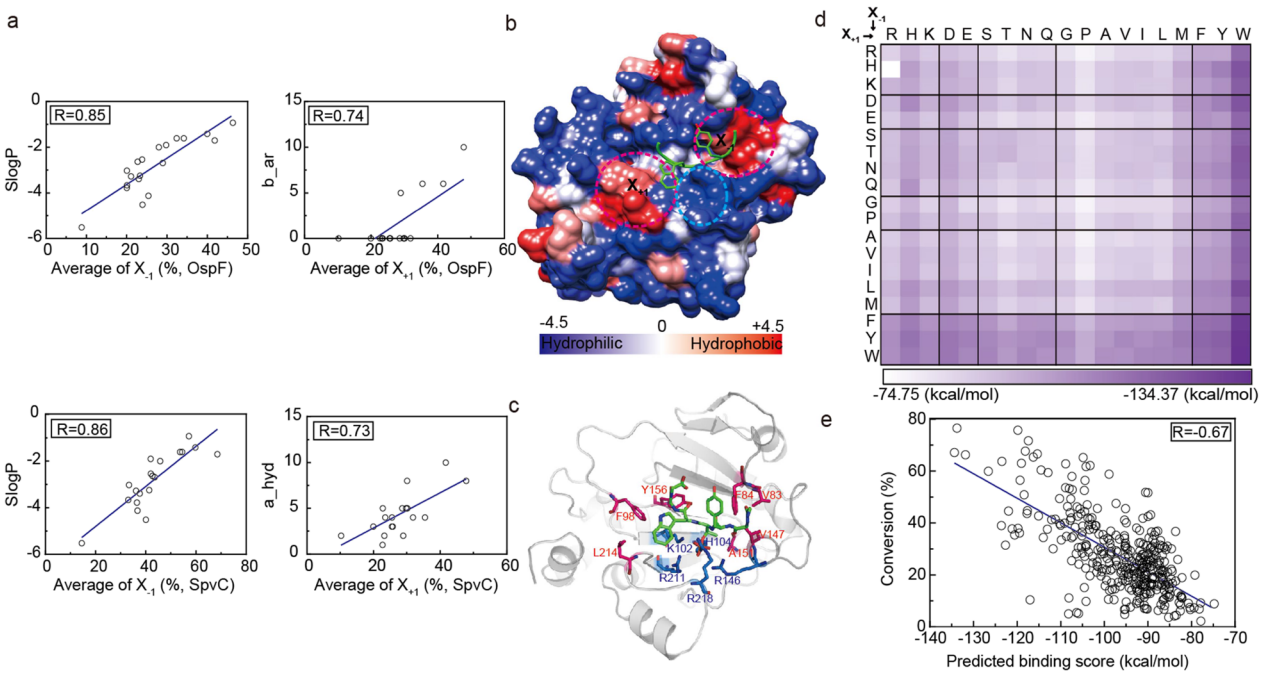


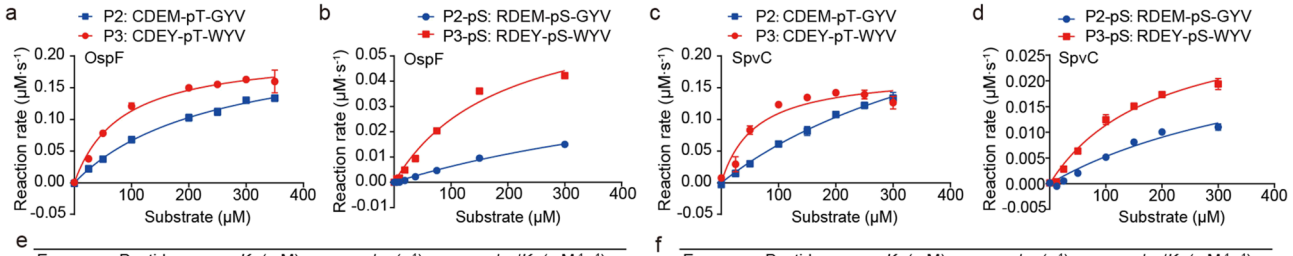
Figure 3. Quantitative and structural analysis of substrate specificity of OspF and SpvC. (a) Average conversion of each residue at X_{-1} or X_{+1} with a linear regression model against SlogP, b_ar and a_hyd descriptors. The average conversion of each residue at X_{-1} was calculated as the averaged conversion of all 19 residues at the X_{+1} position; similarly, the average conversion of residue at X_{+1} was calculated as the averaged conversion of all 19 residues at X_{-1} position. The 19 average conversion values for the X_{-1} or X_{+1} position were used to train a minimal linear regression model with each of the 365 molecular descriptors for the 19 amino acids, making slogP (hydrophobicity), b_ar (number of aromatic atoms), and a_hyd (number of hydrophobic atoms) the best predictors of the average conversion. (b) Modeled complex structure of OspF and a selected peptide GGY-pT-WGG by our HPEPDOCK program. The surface of OspF is pictured here, where hydrophilic residues are colored blue and hydrophobic residues are colored red according to the hydrophobicity scale of Kyte and Doolittle.[31] The blue circle and the two red circles represent the central hydrophilic region and the flanking hydrophobic regions, respectively. The PDB file (Complex PDB) for this modeled structure is provided in Supporting Information. (c) Residues involved in the central hydrophilic region (in blue) and the flanking hydrophobic regions (in red). (d) Heatmap of predicted binding energy scores between OspF and the $GGX_{-1}pTX_{+1}GG$ peptide library calculated by the ITScorePP scoring function. (e) Correlation between experimental activity profiling (Figure 2c) and the predicted binding energy scores (Figure 3d) for OspF. Each data point represents one peptide sequence from the $X_{-1}pTX_{+1}$ library.

As shown in Figures 2c and S6, yields for the lyase-mediated conversion of Ac-GX_{-1}-pT-X_{+1}GRC into Ac-GX_{-1}-Dhb-X_{+1}GRC ranged from 0% to nearly 90%, with OspF and SpvC exhibiting a similar pattern of substrate specificity on the array (Figure 2d). The average conversions for the native motif with M-pT-G (p38), L-pT-G (Erk1/2), and M-pT-P (JNK) were 23 and 33% for OspF and SpvC, respectively (boxed in red, Figures 2c and S6 for numeric data), indicating a suitable dynamic range for the lyase activity assay. Low conversion was observed for Pro (a conformationally constrained residue) at the X_{+1} position, which is consistent with required conformational flexibility observed for the pT-X-pY motif in native substrates.[30] In addition, inclusion of the basic residue Arg at the X_{-1} and X_{+1} positions also resulted in reduced activity, but this was not observed for the other two basic residues, His and Lys. When the 19 amino acids were grouped according to the hydrophobicity and charge of the side chain functional groups, we found that aromatic amino acids Phe, Tyr, and Trp were preferred at the X_{-1} and X_{+1} positions (Figure 2c, boxed in cyan). While one-side substitution with aromatic residues resulted in substantial increase of activity compared to the native motif, double-side substitution of [F/Y/W]-pT-[F/Y/W] (where [F/Y/W] indicates F, Y, or W residue) gave 2∼3 times higher conversions compared to the native motif (69% vs 23% for OspF and 80% vs 33% for SpvC; see Figure S7 for details of calculation).

Molecular Mechanisms of Substrate Specificity. To gain mechanistic insights into the observed specificity pattern, we applied a chemoinformatic approach to determine the factors contributing most to the activity of the lyases.[40,41] We calculated 365 molecular descriptors for each of the 19 amino acids. Each of these descriptors was used to train a linear least squares regression model that predicts the average conversion efficiency for each amino acid residue. Through this analysis, it was revealed that *SlogP* (log of the octanol/water partition coefficient) and number of aromatic/hydrophobic atoms were the most significant parameters that correlated with the average conversion efficiency for the X_{-1} and X_{+1} position analogues, respectively (Figure 3a).

The results of our experiments were also consistent with molecular modeling studies. Figure 3b shows the binding mode of one of the best substrate peptides, GGY-pT-WGG, to the OspF protein, which was predicted by the hierarchical protein−peptide docking program HPEPDOCK.[42−45] We see that the binding pocket accommodating the peptide consists of three distinct regions: one hydrophilic region in the middle and two hydrophobic regions on each side (Figure 3b).[46] The middle hydrophilic region is formed by five positively charged residues (K102, R146, R211, R218, and H104) responsible for the binding of pT at the X_0 position of the peptide (Figure 3c). One of the two hydrophobic regions, which is formed by four hydrophobic residues (V83, F84, V147, and A151) is located

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414−425



e

Enzyme	Peptide	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)
OspF	P2	0.25±0.02	2±0	8.8±0.7
	P3	0.081±0.006	2.06±0.07	25±2
	P2-pS	0.8±0.1	0.053±0.006	0.07±0.01
	P3-pS	0.189±0.004	0.072±0.001	0.38±0.01

f

Enzyme	Peptide	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)
SpvC	P2	0.51±0.02	3.7±0.2	7.3±0.5
	P3	0.059±0.005	1.77±0.04	30±3
	P2-pS	0.49±0.05	0.031±0.003	0.063±0.009
	P3-pS	0.189±0.008	0.033±0.002	0.17±0.01

Figure 4. Kinetic comparison of peptide substrates. P2 (CDEM-pT-GYV) and P2-pS (CDEM-pS-GYV) are peptide mimics of the best substrate protein p38. P3 (CDEY-pT-WYV) and P3-pS (CDEY-pS-WYV) are the corresponding optimized peptides with substitution of M/G to Y/W at the ±1 positions. (a) Kinetic data for the elimination of phosphate from peptides P2 and P3 using OspF. (b) Kinetic data for the elimination of phosphate from peptides P2-pS and P3-pS using OspF. (c) Kinetic data for the elimination of phosphate from P2 and P3 by SpvC. (d) Kinetic data for the elimination of the P2-pS and P3-pS substrates by SpvC. (e) Best-fit values for K_m and k_{cat} were used to determine the catalytic efficiency k_{cat}/K_m for each substrate of OspF. (f) Best-fit values for K_m and k_{cat} were used to determine the catalytic efficiency k_{cat}/K_m for each substrate of SpvC.

at the X_{-1} position (Figure 3c). The other hydrophobic region, which is mainly formed by three hydrophobic residues (F98, Y156, and, L214), is located at the X_{+1} position (Figure 3c). The presence of these two hydrophobic regions explains the observed preference of aromatic residues Phe/Tyr/Trp at both X_{-1} and X_{+1} positions, as they can form favorable hydrophobic interactions with the hydrophobic regions on the protein.

We also substituted the X_{-1} and/or X_{+1} positions with each of the 19 amino acids and calculated the binding energy scores between OspF and these peptides. We found that the calculated binding energy scores showed a good correlation with the experimental data. That is, peptides with hydrophobic residues at X_{-1} and X_{+1} positions generally give lower binding energy scores (Figure 3d, see Figure S8 for numeric data). Overall, the Pearson correlation coefficient between the experimental data and the calculated binding scores is 0.67 (Figure 3e). This result suggests the reliability of our molecular modeling and proposed binding mechanism.

Optimization of a Monophosphorylated Tag for OspF/SpvC. We next asked whether the optimal residues at the ±1 positions could be used to improve the activity of the monophosphorylated peptide sequence from the best native substrate protein p38, CDEM-pT-GYV (P2). We synthesized a peptide that substituted ±1 positions with Y/W to obtain CDEY-pT-WYV (P3). Similarly, we prepared the pS version of both P2 and P3, RDEM-pS-GYV (P2-pS) and RDEY-pS-WYV (P3-pS), respectively.

We then characterized the conversion of these four peptides by OspF and SpvC and obtained kinetic parameters by fitting to the Michaelis–Menten model (Figure 4). In all four cases, the activity of the modified substrates (P3 and P3-pS) was higher than the original substrate (P2 and P2-pS), with the increased activity between the pS substrate catalyzed by OspF being the most significant. Quantitatively (Figure 4), we found that substitution of M/G to Y/W at the ±1 positions resulted in a 2.6–8.6-fold decrease in K_m values for different pT/pS peptide substrates catalyzed by OspF or SpvC, with the maximum change of 8.6-fold observed for the pT peptides catalyzed by SpvC (P2: 0.51 mM; P3: 0.059 mM). The effect of substitution on k_{cat} was smaller than that of K_m, with a maximum increase of 1.4-fold for the pS peptides catalyzed by OspF (P2-pS: 0.053 s^{-1}; P3-pS: 0.072 s^{-1}). As an overall result, the catalytic efficiency k_{cat}/K_m thus achieved a maximum increase of 5.4-fold for the pS peptide when catalyzed by OspF (P2-pS: 0.07 mM^{-1} s^{-1}; P3-pS: 0.38 mM^{-1} s^{-1}). These data indicate that the substitution of M/G to Y/W significantly increases the binding affinity between the peptide and the lyases and serves as the major contribution to the observed increase in catalytic efficiency for the modified monophosphorylated peptides. Molecular docking studies of these protein–peptide interactions also provided an insight into the enhanced lyase-substrate binding: the binding energy of P3-OspF is 56.3 kcal/mol lower than that of P2-OspF (−185.29 kcal/mol vs −128.96 kcal/mol). In short, the optimization of ±1 positions resulted in a 2.8–5.8-fold overall improvement of catalytic efficiency. This, therefore, enables the simplification of the di-phosphorylated X-pT/pS-X-pY motif to a monophosphorylated [F/Y/W]-pT/pS-[F/Y/W] sequence that shows comparable activity with the best di-phosphorylated substrate reported to date.[28] This also provides the possibility of downstream applications of OspF/SpvC-mediated Dhb/Dha insertion, as we demonstrate below.

Enhancement of Peptide Labeling Efficiency with the Optimized Monophosphorylated Tag. The usefulness of the OspF/SpvC-mediated transformations stems from the versatility of the product residues, Dha/Dhb, which can be used as sites for covalent modification of peptides or proteins through nucleophilic addition (thia-Michael addition and aza-Michael addition) as well as radical addition (Figure 1b).[13] As a proof-of-concept demonstration of the application of this lyase-mediated α,β-unsaturated carbonyl insertion, we first investigated the reactivity of Dha and Dhb in the optimized peptide (R/CDEY-Dha/Dhb-WYV) toward four thiols [3-mercaptopropane-1,2-diol, ethyl-2-mercaptoacetate, *N*-acetylcysteamine, and 2-(dimethylamino) ethanethiol]. The Dha/Dhb-containing peptide R/CDEY-Dha/Dhb-WYV was first generated by treating phosphopeptide (R/CDEY-pS/pT-WYV) with OspF for 30 min, and subsequent Michael addition reactions were carried out with 20 mM thiol at room temperature for 2 h (Figure 5a). Indeed, Dha showed good reactivity as a thiol-Michael acceptor with conversion of 75–92% with the four thiols (Figure 5b), as characterized by MALDI-TOF MS and MS/MS (Figures 5c,d and S9–11). However, the reaction of Dhb with thiol compounds did not

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414–425

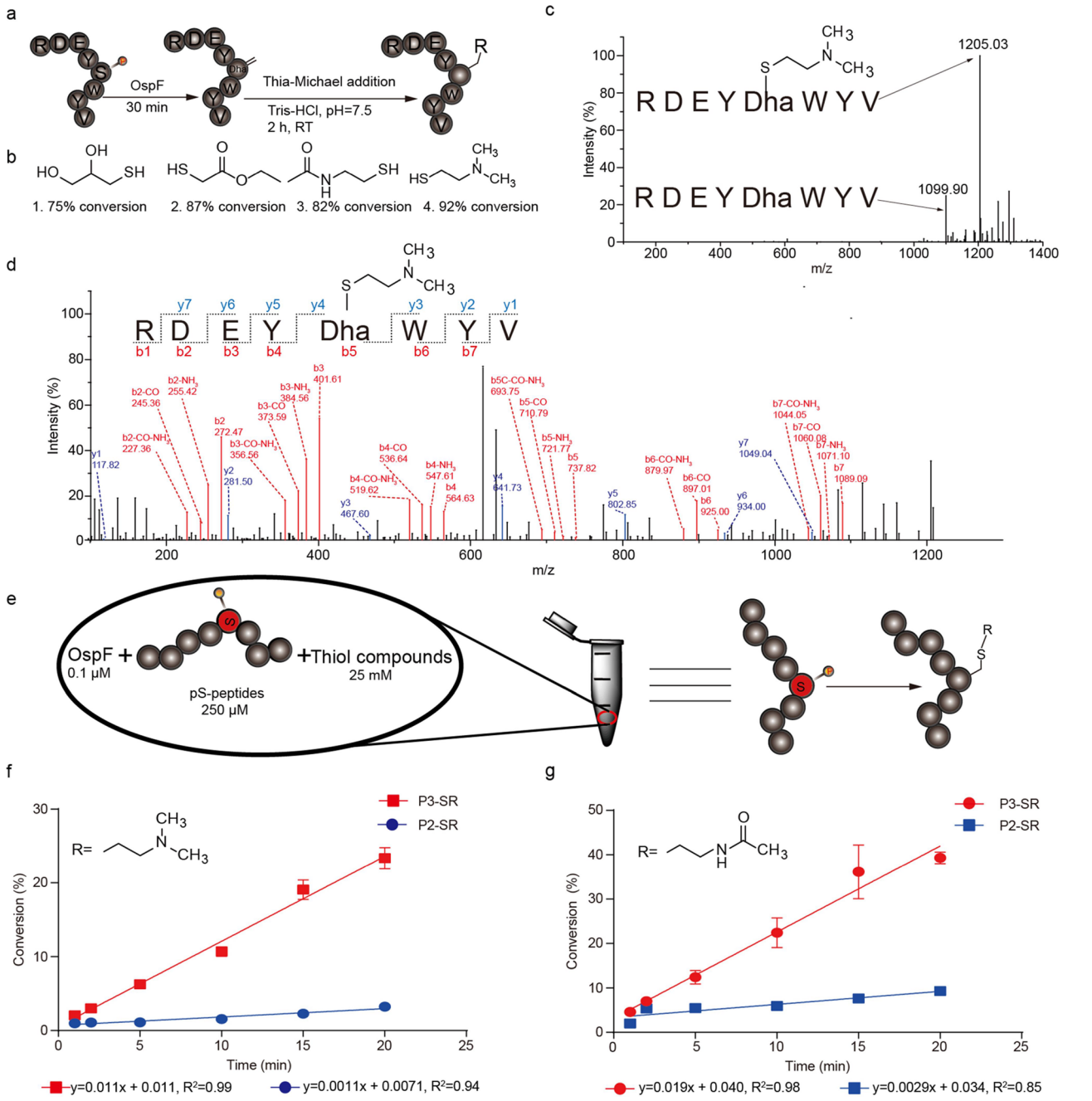


Figure 5. Site-specific peptide labeling via OspF-mediated Dha insertion. (a) Scheme for stepwise OspF-mediated Dha insertion and thia-Michael addition reactions to enable site-specific peptide labeling. (b) Chemical structures for the four thiols and the corresponding yields of the reactions. (c) Product of 2-(dimethylamino) ethanethiol and P3-Dha was detected with MALDI-TOF. (d) Tandem MS analysis of reaction products confirming successful labeling of Dha with 2-(dimethylamino) ethanethiol. Fragment ions "b" are shown in red, and fragment ions "y" are shown in blue. (e) Scheme of the one-pot reaction to compare the labeling efficiency of the optimized monophosphorylated tag (P3-pS) and the unoptimized monophosphorylated tag (P2-pS). (f) Labeling rate of optimized monophosphorylated tag (P3-pS) is 10× fold higher than the unoptimized monophosphorylated tag (P2-pS) with 2-(dimethylamino) ethanethiol. (g) Labeling rate of the optimized monophosphorylated tag (P3-pS) is 6.5× fold higher than the unoptimized monophosphorylated tag (P2-pS) with *N*-acetylcysteamine.

result in any detectable product formation in thia-Michael addition, due presumably to the attenuated electrophilicity of Dhb as a result of steric hindrance and hyperconjugation of the additional methyl group (Figure S12).[15,47,48] To increase the reactivity of Dhb, a recently developed Dhb-reactive phosphine probe may be coupled with the pT-based monophosphorylated tag for efficient labeling of protein targets.[15] In addition, ATP–Cu(II) catalyst and Mn(I) complexes were also reported to efficiently catalyze Michael reaction for structures bearing a Dhb motif.[49,50]

Having confirmed the reactivity of Dha, we then compared the labeling efficiency of the optimized monophosphorylated sequence RDEY-pS-WYV (P3-pS) and the unoptimized sequence RDEM-pS-GYV (P2-pS) in a "one-pot" reaction with OspF and the thiol compounds (Figure 5e). These reaction conditions are more relevant to intracellular protein labeling, where the target substrate, labeling enzyme, and

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414–425

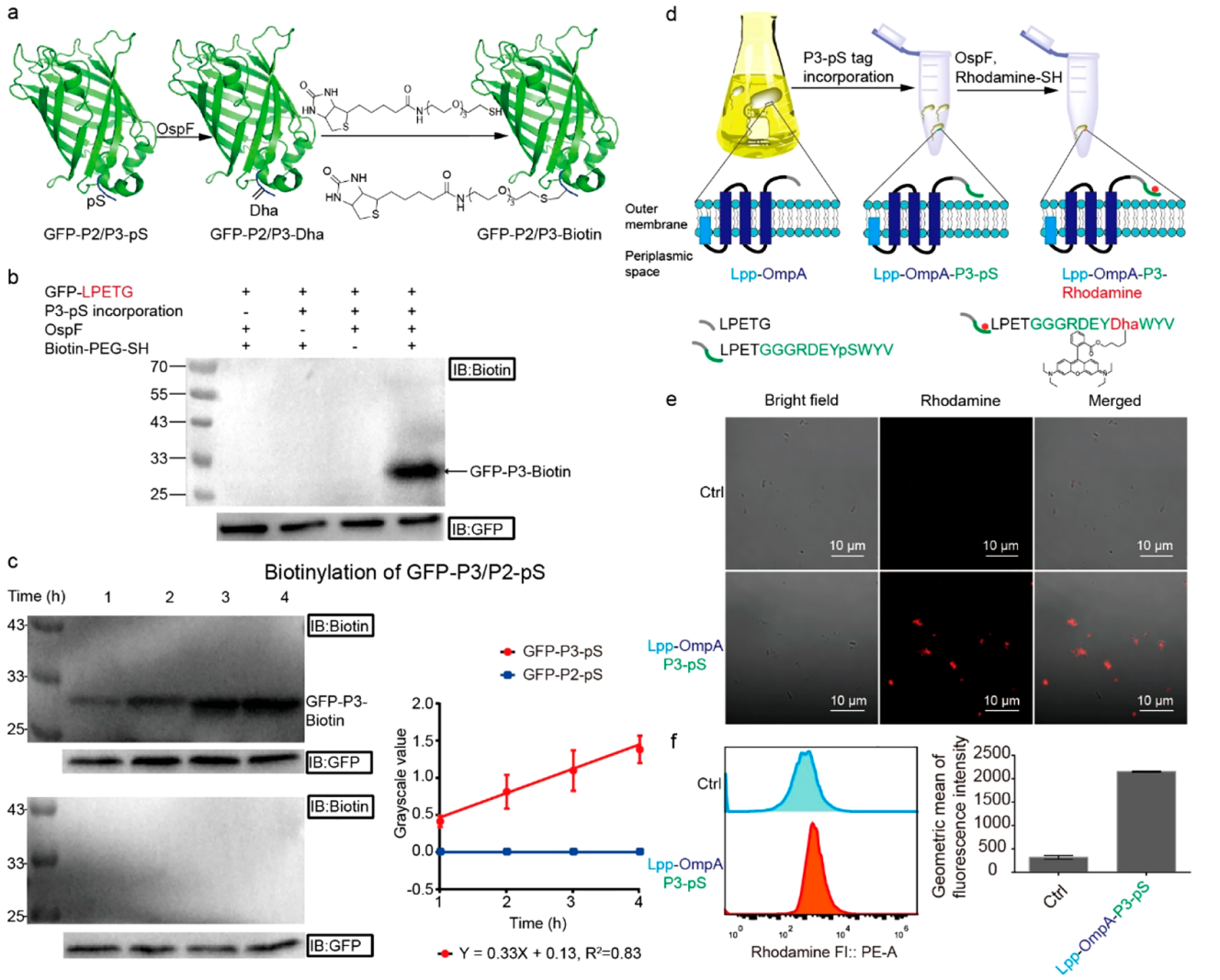


Figure 6. Protein and live cell labeling via OspF-mediated enzymatic incorporation of Dha. (a) Scheme showing the production of GFP-P2/P3-Dha and their reactions with biotin-PEG-SH. (b) Western blot detects biotinylated GFP (GFP-P3-Biotin) only in the positive sample with all four components (GFP-LPETG, StrA, OspF, and biotin-PEG-SH). (c) Labeling efficiency of biotinylated GFP-GGGRDEY-pS-WYV (GFP-P3-pS) is significantly higher than that of GFP-GGGRDEM-pS-GYV (GFP-P2-pS). (d) Scheme of outer membrane protein Lpp-OmpA modification of live bacterial cell surfaces via OspF-mediated Dha insertion. (e) Confocal microscopy of rhodamine-labeled *E. coli* showing that bacteria can be labeled via the surface-displayed Lpp-OmpA-GGGRDEY-pS-WYV. (f) Flow cytometry of rhodamine-labeled *E. coli*. The labeling efficiency of bacteria with Lpp-OmpA-GGGRDEY-pS-WYV is significantly higher than the control group (no overexpressed OmpA).

modification group are all present. The results in Figure 5f,g show that the optimized P3-pS displayed 10-fold and 6.5-fold enhancement of conversion over the unoptimized P2-pS, with 2-(dimethylamino) ethanethiol and N-(2-sulfanylethyl) acetamide, respectively. These data demonstrate that optimization of the residues at the ±1 positions of the monophosphorylated substrate was effective at identifying a monophosphorylated substrate for modification by lyases and subsequent use in peptide/protein-labeling applications.

OspF-Mediated Labeling of GFP Protein and *E. coli* Cell with the Optimized Monophosphorylated Tag. We next evaluated the OspF-mediated Dha insertion and labeling method in bacterial cells. We first evaluated whether the optimized monophosphorylated tag could be used to introduce a biotin group on the model protein GFP (Figure 6a). For such *in cellulo* applications, the Dha precursor can be readily introduced via efficient genetic encoding of phosphoserine.[51,52] Chin and co-workers evolved an orthogonal aminoacyl-tRNA synthetase/$tRNA_{CUA}$ pair that directs the efficient incorporation of pS into recombinant proteins in *E. coli*. This approach enables quantitative decoding of the amber stop codon as pS. They also discovered a phosphothreonyl-tRNA synthetase/$tRNA_{CUA}$ pair and created an entirely biosynthetic route to incorporate phosphothreonine in proteins.[53] In addition, phosphorylation of Ser or Thr on target peptide can be made possible with a choice of kinase with suitable substrate specificity (Table S2[54]), with regard to the ±2 and ±3 sequence of the target peptide.

Here, as a quick demonstration, we prepared C-terminus-tagged GFPs through ligation of GFP-LPETG and the monophosphorylated peptide (P4: GGGRDEM-pS-GYV or P5: GGGRDEY-pS-WYV) using the transpeptidase SrtA.[55] The resulting proteins, GFP-GGGRDEY-pS-WYV (GFP-P3-pS) and GFP-GGGRDEM-pS-GYV (GFP-P2-pS), were used to add a biotin group with a poly(ethylene glycol) linker (Figure 6b). We treated GFP-P3-pS (30 μM) with OspF (5

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414−425

μM) for 4 h at room temperature and a thiol-PEG-biotin (50 mM) to generate a biotinylated GFP that was visualized by streptavidin-HRP using Western blot. We next compared the rate of labeling between the two GFPs with different tags (GFP-P2-pS: GGGRDEM-pS-GYV GFP-P3-pS-GGGRDEY-pS-WYV). The results in Figure 6c show that within the time frame of the experiment, GFP with the optimized tag, GFP-P3-pS, was clearly labeled, but no detectable label was found in the GFP with the unoptimized tag, GFP-P2-pS.

Finally, we evaluated whether the optimized monophosphate tag could enable OspF-based labeling of live bacteria. In this study, we employed an *E. coli* strain that overexpresses a surface-exposed transmembrane protein Lpp-OmpA-LPETG.[56,57] We then installed the OspF recognition motif Y-pS-W on the surface of *E. coli* by using SrtA and P5 (GGGRDEY-pS-WYV) to modify Lpp-OmpA-LPETG on the C-terminus. The resulting surface-displayed Lpp-OmpA-GGGRDEY-pS-WYV then served as the site for OspF/Dha-mediated functionalization of the bacterial surfaces with thiol-PEG-rhodamine (Figure 6d). As shown in Figure 6e, *E. coli* cells treated with 5 μM OspF and 1 mM thiol-PEG-rhodamine for 0.5 h showed fluorescence intensity significantly higher than the control lacking overexpressed OmpA, indicating successful labeling of the cell surface under these conditions. (Figure 6f). These examples demonstrate the general utility for OspF as an enzymatic approach for orthogonal polypeptide modifications on proteins and whole cells.

CONCLUSIONS

This work describes the development of a monophosphorylated peptide tag for the installation of thiol-reactive Dha residues on proteins and whole cells using the bacterial lyases OspF and SpvC. In this study, we employed SAMDI-MS to determine the relationship between the catalytic efficiency of OspF/SpvC and the preferred amino acid residues at the ±1 positions of the peptide substrates. We found that aromatic and hydrophobic residues at these positions are preferred and that the use of these preferences guided the development of a mono-phosphorylated tag, [F/Y/W]-pT/pS-[F/Y/W], that displays a 2.8∼5.8-fold increase in overall catalytic efficiency compared to the native mono-phosphorylated substrates. This optimized tag also led to a ∼10-fold enhancement of the overall reaction efficiency and enabled the first demonstration of protein and bacteria labeling via OspF-mediated enzymatic incorporation of alkene functional groups.

These results demonstrate that this approach, coupled with genetic incorporation of pS/pT[51,52], and Dha/Dhb conjugation chemistry,[13,15−18] can serve as a valuable tool for site-specific in situ labeling of proteins and whole cells. The simplification from the native di-phosphorylated tag to monophosphorylated tag has further lowered technical barriers for such applications. Additionally, the difference in reactivity of Dha and Dhb can be exploited for applications in which dual-site protein labeling is desired.[58−60] Further, directed evolution of OspF may yield mutants that have even higher catalytic efficiency and specificity, or variants that selectively recognize different mono-phosphorylated sequences on substrates in vivo, a set of tools that would be highly useful in live cell phosphoproteomics studies.

METHODS

General. Unless otherwise stated, all chemical reagents are of analytical grade, obtained from commercial suppliers and used without further purification.

Materials. *Genes and Cells.* The *Shigella flexneri ospf* gene was synthesized at Kingsray Biotechnology Co., Ltd. The *spvc*, *srtA*, and *ompA* genes were amplified from *Salmonella typhimurium*, *Staphylococcus aureus*, and *E. coli*, respectively. All competent cells, including shuffle T7, *E. coli* DE3, *E. coli* DH5α, and *E. coli* DH10B, were purchased from Thermo Fisher.

Chemicals, Enzymes, and Other Reagents. 2-amino-6-mercapto-7-methylpurine riboside (MESG) was purchased from Cayman Co., Ltd., and purine nucleoside phosphorylase (PNP) was purchased from Thermo Fisher. Potassium chloride (KCl), sodium chloride (NaCl), Tris−HCl, glycine, nickel(II) sulfate ($NiSO_4$), and sodium phosphate were purchased from Hefei Bomei Biotechnology Co., Ltd. Thiol compounds were purchased from Shanghai Macklin Biochemical Co., Ltd. The ClonExpress II One Step Cloning Kit was purchased from Novizan Biotechnology Co., Ltd. Restriction endonucleases and protease inhibitor cocktails were purchased from Thermo Fisher. Streptavidin was purchased from Sangon Biotech (Shanghai) Co., Ltd. The GFP antibody and secondary antibodies were purchased from Abmart Pharmaceutical Technology (Shanghai) Co., Ltd. 96-well filter plates (cat. no. AWFP-F20000, Arctic White LLC) and all fluorenylmethyloxycarbonyl (Fmoc) amino acids and Rink-amide resins were purchased from AnaSpec, Inc. All solvents, N,N-dimethylformamide (DMF), dichloromethane, trifuoroacetic acid (TFA), and piperidine were purchased from Thermo Fisher Scientific. Other chemical reagents used in peptide synthesis were purchased from Sigma-Aldrich unless otherwise noted.

Gene Cloning, Expression, and Purification. *Cloning.* All protein plasmids used in this study were generated by homologous recombination with standard protocols. Briefly, the target gene amplified with forward and backward primers, purified, and ligated into the corresponding vectors with recombinase. Information of vectors, restriction sites, and primers is given in Table S1.

Protein Expression. Plasmids were transformed to the corresponding competent cells (shuffle T7-K12 and *E. coli* DE3 for protein expression; *E. coli* DH5α and *E. coli* DH10B for plasmid amplification). Transformants were plated on an LB agar plate supplemented with ampicillin (100 μg/mL) or kanamycin (50 μg/mL) and incubated overnight at 30 °C. A single colony was inoculated into 20 mL of LB supplemented with ampicillin (100 μg/mL) or kanamycin (50 μg/mL) and agitated at 220 rpm at 30 °C. On the following day, overnight cell culture was diluted in 1000 mL LB supplemented with ampicillin (100 μg/mL) or kanamycin (50 μg/mL) and agitated at 220 rpm at 30 °C until the OD_{600} reached 0.6−0.8. Protein expression was induced with 1 mM IPTG at 18 °C for 20 h. Cells were collected by centrifugation at 8000 g for 40 min at 4 °C and stored at −80 °C.

Protein Purification. Cell pellets were resuspended in 25 mL lysis buffer (50 mM Tris−HCl, pH 8.0, 250 mM NaCl, lysozyme 1 mg/mL, DNase 0.1 mg/mL, and 1× protease inhibitor cocktails). The cell suspension was sonicated with Sonic Dismembrator (Ningbo Scientz Biotechnology Co.,LTD, with 40% output, 40 min, 2 s off, 1 s on) in an ice-water bath, followed by centrifugation to remove the debris (18,000 g, 60 min, 4 °C). The supernatant was collected and filtered through a 0.45 μm filter. The filtrate was loaded onto a Poly-Prep Chromatography Column, washed with lysis buffer, and eluted by gradually increasing the eluent buffer (50 mM Tris−HCl, pH 8.0, 500 mM NaCl, 250 mM imidazole) for 60 min. The eluate was concentrated and desalted against a buffer with no imidazole. Bacterial lysates and eluted fractions were characterized using SDS-PAGE stained with Coomassie brilliant blue. The GST tag was removed from the target protein by cleavage with HRV 3C at a concentration of 10 μg per mg target protein for an overnight at 4 °C. The digested protein mixture was purified again with a nickel column, and the eluates were collected and characterized with SDS-PAGE. The final protein product was concentrated and desalted into protein

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414−425

storage buffer (50 mM Tris−HCl, pH 7.4, and 150 mM NaCl, 25%v/v glycerol) at a concentration of 3.2 mg/mL, and stored at −80 °C.

SAMDI-Based Substrate Specificity Profiling. *Solid-Phase Synthesis of Peptide Arrays.* Solid-phase peptide synthesis (SPPS) was performed as described previously.[61] Briefly, SPPS was performed on Rink-amide resin (10 mg) housed in 96-well filter plates. N-terminal Fmoc protecting groups were deprotected with 20% piperidine in DMF at room temperature for 20 min. The resin was filtered and rinsed five times with DMF on a multiscreen vacuum manifold. Amino acids were coupled to the resin with PyBop and *N*-methylmorpholine (NMM) in DMF at a 4:4:8 molar excess for 30 min, twice. The deprotection and coupling were repeated for each residue, following deprotection of the final residue. Peptides were cleaved and deprotected in 95% TFA, 2.5% triethylsilane, and 2.5% water for 16 h. The cleavage solution was evaporated with N_2 gas flow. The peptides were resuspended in 0.1% TFA in water, lyophilized, and resuspended again in 0.1% TFA in water to a final concentration of 200 μM and stored at −80 °C.

Preparation of SAMDI Plates. Steel array plates evaporated with 384 gold spots (with a diameter of 3.0 mm) were soaked at 4 °C in a solution of 1 mM total disulfide with 0.8 mM tri(ethylene glycol)-terminated C11-alkane disulfide and 0.2 mM C11-alkane disulfide with one terminal tri(ethylene glycol) and one terminal maleimide in ethanol for 2 days to allow assembly of the monolayer (10% maleimide coverage). Peptides were diluted to a final concentration of 20 μM with 100 mM Tris−HCl buffer (pH 7.5) in a 384-well plate with 5 μL of TCEP beads (Thermo Scientific) in each well. Using a TECAN robotic liquid handler, 2 μL of peptide solution from each well was transferred on to the corresponding gold spot on the monolayer-presenting plate and incubated at 37 °C in a humidified chamber for 1 h to allow for peptide immobilization.

Profiling Conversion Efficiency of OspF/SpvC. OspF/SpvC (0.1 μM) was diluted in Tris−HCl buffer (100 mM Tris−HCl, pH 7.5, 50 mM NaCl), and 2 μL was applied to each gold spot on the peptide array plate with a Multidrop Combi (Thermo Scientific). The reactions were incubated at 37 °C in a humidified chamber for 2 h. After the reaction was complete, the plate was rinsed with water and ethanol, treated with a 1 μL matrix (10 mg/mL THAP, 5 mg/mL ammonium citrate dibasic in 50% acetonitrile, 50% water, and 0.1% phosphoric acid) to each spot, and dried in air for 20 min. The spots were analyzed by MALDI-TOF MS to obtain a mass spectrum for each reaction. Enzymatic activities were quantified by measuring the peak intensity for the product peak and the substrate peak and determining the activity. Activity heat maps were generated by Microsoft Excel.

$$\text{Activity (\%)} = \frac{I(\text{Product})}{I(\text{Product}) + I(\text{Substrate})} \times 100\%$$

Enzyme Activity Assay and Kinetic Characterization. *Product Characterization with MALDI.* The OspF/SpvC reactions were performed using 0.1 μM enzyme and 250 μM peptide in Tris−HCl buffer (50 mM, pH 7.5) for 30 min at 25 °C. The reactions were quenched by heating at 95 °C for 10 min. 1 μL of the reaction solution was mixed with 4 μL of the CHCA matrix (5 mg/mL), and the reaction products were detected by a 5800 MALDI-TOF/TOF (ABSciex) with reflective positive ion mode.

Kinetic Characterization with EnzChek Assay. The steady-state kinetic analysis was performed according to the previously reported phosphate detection method.[38] Briefly, OspF/SpvC was diluted in Tris−HCl buffer (50 mM, pH 7.5) containing 0.1 mg/mL MESG and 0.3 unit of PNP to a final concentration of 100 nM. Reactions were initiated by adding substrate peptide at different concentrations (0−300 μM) to each well of a 384-well plate to a final volume of 30 μL. The reaction was monitored at 360 nm for product formation for 60 min with a plate reader (Spark 10 M, TECAN). Initial velocity (V_0) for product formation was calculated following the linear range of the reaction profile.

$$V_0 = \frac{\text{Absorbance (360 nm)}}{11{,}000\ \text{M}^{-1}\,\text{cm}^{-1} \times \text{Time (s)}}$$

Initial velocities were then analyzed with the Michaelis−Menten kinetics equations by using GraphPad Prism 6 (GraphPad Software, San Diego) to obtain the kinetic parameters.

Quantitative Analysis and Molecular Modeling. *Chemoinformatic Analysis.* Physiochemical descriptors of amino acids were calculated using Molecular Operation Environment (MOE, version 2014) after energy minimization of the structure with Amber 10 force field.[40,41] Data with each descriptor were used to train a least linear regression model that predicts the average conversion for each amino acid residue.

Molecular Modeling of OspF-Peptide Interactions. The OspF-peptide interactions were modeled using a template-based docking approach. Specifically, the three-dimensional (3D) OspF structure was modeled based on the high-homologous SpvC protein structure (PDB ID: 2Z8P) using the homology modeling program Modeler.[62] To retain the peptide-bound conformation of OspF, the native peptide ligand from the SpvC structure was kept during the homology modeling process. Given the high sequence identity of 63.2% between OspF and SpvC (Figure S13a), the obtained OspF protein had a high-quality structure with an excellent TMscore of 0.998 (Figure S13b).[62] For each sequence in the GGX_{-1}-pT-$X_{+1}GG$ library, the 3D conformation of the modeled peptide was constructed based on the native peptide ligand from the SpvC structure by MODPEP program.[44] Then, with the 3D structures of OspF and modeled peptides, the putative peptide binding modes were sampled using our hierarchical protein−peptide docking program HPEPDOCK,[43,44,63] where the ITScorePP scoring function for protein−protein interactions was used to evaluate the binding score between the protein and the peptide.[45] The top binding mode with the best binding score was chosen as the predicted OspF-peptide complex structure.

OspF-Based Peptide Labeling. *Generation of Dha/Dhb-Containing Peptide.* 1 mM phosphorylated peptide was dissolved in 100 μL Tris buffer (100 mM, pH = 7.5), to which purified OspF was then added to a final concentration of 10 μM. The reaction was then incubated at 25 °C for 30 min before it was quenched by and analyzed by MALDI as mentioned above to ensure full conversion of phosphorylated peptide to Dha/Dhb-containing peptide.

Coupling of the Dha/Dhb-Containing Peptide with Thiol Compounds. 50 μL of 20 mM thiol compounds in Tris−HCl buffer (100 mM, pH = 7.5) was added to the abovementioned Dha/Dhb-containing peptide (1 mM, 50 μL), and the reactions were carried out at room temperature for 2 h before it was quenched and analyzed by MALDI as mentioned above. The percentage conversions of the reaction were calculated using the following formula

$$\text{Conversion (\%)} = \frac{I(\text{product})}{I(\text{product}) + I(\text{substrate})} \times 100$$

The thiol adduct peptide was also analyzed with Tandem MS/MS (Data Explorer V4.5).

"One-Pot" Reaction for Comparison of Labeling Efficiency. The phosphorylated peptide, OspF, and thiol compound were dissolved in 50 μL of sodium phosphate buffer (50 mM, pH 8.0) at final concentrations of 250 μM, 0.1 μM, and 25 mM, respectively. The percentage conversion of each reaction was monitored at 1, 2, 5, 10, 15, and 20 min by quenching and analyzing with MALDI as mentioned above.

Protein Labeling. *Preparation of GFP-P2/P3-Dha.* 30 μM GFP-LPETG, 60 μM SrtA (D107A, G109E mutant), and 1 mM monophosphorylated peptide (P4: GGGRDEM-pS-GYV or P5: GGGRDEY-pS-WYV) were mixed in 50 μL reaction buffer (Tris−HCl 50 mM, 10 mM $CaCl_2$, pH = 7.5). The transpeptidase reaction was allowed for 6 h at room temperature, and then monophosphorylated peptide and $CaCl_2$ were removed by ultrafiltration with Tris−HCl buffer (50 mM, pH = 7.5). OspF was then added to a final concentration of 5 μM, and the mixture was incubated for 4 h at room temperature for Dha formation.

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414−425

Labeling GFP-P2/P3-Dha with Biotin-PEG-SH. 50 mM biotin-$PEG_{n=3}$-SH was added to the abovementioned reaction mixture and reacted overnight at room temperature. The GFP-P3-Biotin was imaged and quantified by Western blot with either 1:1000 HRP-conjugated streptavidin to detect the biotin tag or with mouse monoclonal eGFP (1st)/Goat Anti-Mouse IgG HRP (2nd) to detect GFP. The GFP-P3-Biotin was identified as the band with concurrence on the biotin blot and GFP blot.

Comparison of Protein-Labeling Efficiency. Labeling efficiency of GFP-P2/P3-pS was compared similarly as mentioned above, with OspF-mediated Dha formation being terminated at different time points (1, 2, 3, and 4 h). The labeling efficiency presented in Figure 6c was quantified with ImageJ (version 1.8.0).

Live Bacterial Cell Surface Labeling. A truncated version of transmembrane protein OmpA was used in this experiment, and it was redirected to the bacterial cell surface by adding a signaling peptide Lpp.[56,57] Bacteria harboring pET28a-Lpp-OmpA-LPETG were grown in 5 mL of LB medium supplemented with 1 mM IPTG and 50 μg/mL kanamycin for 20 h, harvested by centrifugation (1200 g), washed (3 times), and resuspended in 1 mL of a 100 mM phosphate buffer (pH 7.4) at a final OD_{600} of ∼25 and a volume of 200 μL. SrtA and monophosphorylated peptide (GGGRDEY-pS-WYV) were added to final concentrations of 10 μM and 1 mM, respectively. The transpeptidase reaction was then proceeded at room temperature for 5 h with a gentle agitation of 120 rpm. The cells were then washed with phosphate buffer to remove StrA and monophosphorylated peptide and resuspended in 150 μL of a 100 μM phosphate buffer (pH 7.4). OspF (5 μM) and SH-PEG-rhodamine (1 mM) were added to the bacteria suspension, and the eliminylation reaction and thia-Michael addition were performed for 0.5 h. The cells were then washed with phosphate buffer and subsequently analyzed with confocal imaging and flow cytometry. The bacteria sample without overexpressed Lpp-OmpA-LPETG was used as control.

Protein sequence of Lpp-OmpA: MKATKLVLGAVILGSTL-LAGCSSNAKIDQNNN.

GPTHENQLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGS-VENGAYKAQGVQLTAKLGYPITDDLDIYTRLGGMV-WRADTKSNVYGKNHDTGVSPVFAGGVEYAITPEIATGSGGGL-PETGG.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.1c00866.

OspF/SpvC expression, enzyme assays and kinetic profiles, raw results for phosphorylated peptide library screening with SAMDI-MS, additional characterizations of Michael addition reactions, all plasmid construction information, peptide sequence information, and kinase specificity table (PDF)

Modeled complex structure (PDB)

■ AUTHOR INFORMATION

Corresponding Authors

Xinxin Feng − *Institute of Chemical Biology and Nanomedicine, State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan Provincial Key Laboratory of Biomacromolecular Chemical Biology, and Department of Chemistry, Hunan University, Changsha 410082, China;* orcid.org/0000-0003-3142-5629; Email: xinxin_feng@hnu.edu.cn

Milan Mrksich − *Departments of Chemistry and Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States;* orcid.org/0000-0002-4964-796X; Email: milan.mrksich@northwestern.edu

Sheng-You Huang − *School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China;* orcid.org/0000-0002-4209-4565; Email: huangsy@hust.edu.cn

Authors

Anming Yang − *Institute of Chemical Biology and Nanomedicine, State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan Provincial Key Laboratory of Biomacromolecular Chemical Biology, and Department of Chemistry, Hunan University, Changsha 410082, China*

Huanyu Tao − *School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*

Lindsey C. Szymczak − *Departments of Chemistry and Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States*

Liang Lin − *State Key Laboratory of Bio-organic and Natural Products Chemistry, Center for Excellence in Molecular Synthesis, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China*

Junfeng Song − *Institute of Chemical Biology and Nanomedicine, State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan Provincial Key Laboratory of Biomacromolecular Chemical Biology, and Department of Chemistry, Hunan University, Changsha 410082, China*

Yi Wang − *Institute of Chemical Biology and Nanomedicine, State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan Provincial Key Laboratory of Biomacromolecular Chemical Biology, and Department of Chemistry, Hunan University, Changsha 410082, China*

Silei Bai − *Institute of Chemical Biology and Nanomedicine, State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan Provincial Key Laboratory of Biomacromolecular Chemical Biology, and Department of Chemistry, Hunan University, Changsha 410082, China*

Justin Modica − *Departments of Chemistry and Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.1c00866

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (21807031 and 22177031 to X.F., 62072199 and 31670724 to S.-Y.H.), Hunan Provincial Natural Science Foundation of China (grant No. 2020JJ4177 to X.F.), Open Funding Project of the State Key Laboratory of Biocatalysis and Enzyme Engineering (grant SKLBEE2019003 to X.F.), Hunan Provincial Innovation Foundation For Postgraduate (grant No. CX20190264 to A.Y.), the Defense Threat Reduction Agency (grant No. HDTRA1-15-10052 to M.M.), and the Robert H. Lurie Cancer Center's Translational Bridge Program at Northwestern University. We thank N. Zhou in Hunan University for kindly providing the SrtA plasmid.

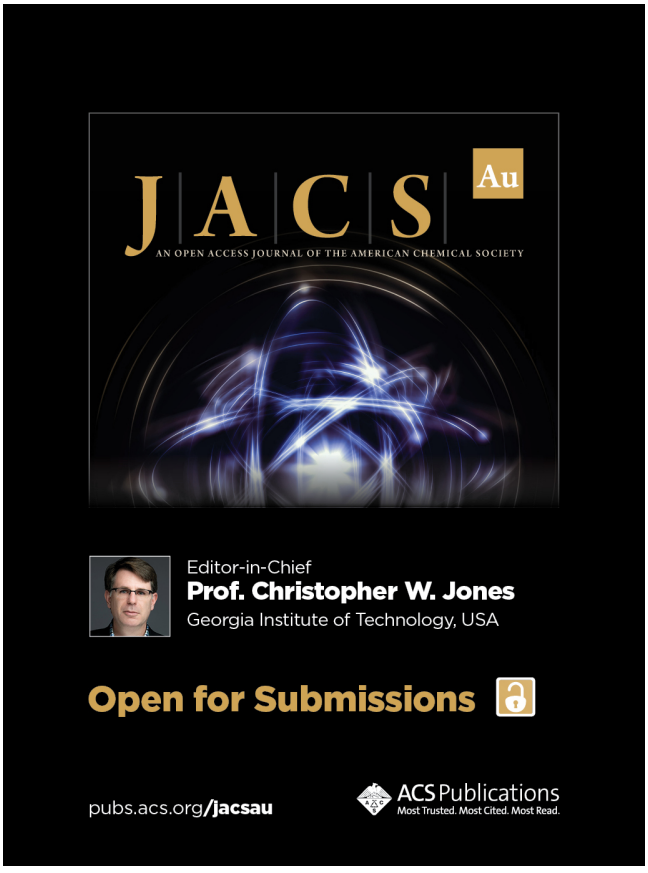
■ REFERENCES

(1) Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2016**, *8*, 103−113.

(2) Hoyt, E. A.; Cal, P. M. S. D.; Oliveira, B. L.; Bernardes, G. J. L. Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **2019**, *3*, 147.
(3) Milczek, E. M. Commercial applications for enzyme-mediated protein conjugation: new developments in enzymatic processes to deliver functionalized proteins on the commercial scale. *Chem. Rev.* **2018**, *118*, 119.
(4) Zhang, Y.; Park, K. Y.; Suazo, K. F.; Distefano, M. D. Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev.* **2018**, *47*, 9106.
(5) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114*, 4764.
(6) Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **2017**, *550*, 53.
(7) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing logical codon reassignment – Expanding the chemistry in biology. *Chem. Sci.* **2015**, *6*, 50.
(8) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. A General Method for Site-specific Incorporation of Unnatural Amino Acids into Proteins. *Science* **1989**, *244*, 182.
(9) Wang, L.; Schultz, P. G. Expanding the Genetic Code. *Angew. Chem., Int. Ed.* **2005**, *44*, 34.
(10) Spicer, C. D.; Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **2014**, *5*, 4740.
(11) Boutureira, O.; Bernardes, G. J. L. Advances in chemical protein modification. *Chem. Rev.* **2015**, *115*, 2174.
(12) Chalker, J. M.; Davis, B. G. Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. *Curr. Opin. Chem. Biol.* **2010**, *14*, 781.
(13) Dadová, J.; Galan, S. R. G.; Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Curr. Opin. Chem. Biol.* **2018**, *46*, 71.
(14) Bogart, J. W.; Bowers, A. A. Dehydroamino acids: chemical multi-tools for late-stage diversification. *Org. Biomol. Chem.* **2019**, *17*, 3653.
(15) Chambers, K. A.; Abularrage, N. S.; Hill, C. J.; Khan, I. H.; Scheck, R. A. A chemical probe for dehydrobutyrine. *Angew. Chem., Int. Ed.* **2020**, *59*, 7350.
(16) Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J. L.; Wiewiora, R.; Ng, W.-L.; Raj, R.; Faulkner, S.; Vallée, M. R. J.; Phanumartwiwath, A.; et al. Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **2016**, *354*, No. aag1465.
(17) Yang, A.; Ha, S.; Ahn, J.; Kim, R.; Kim, S.; Lee, Y.; Kim, J.; Söll, D.; Lee, H. Y.; Park, H. S. A chemical biology route to site-specific authentic protein modifications. *Science* **2016**, *354*, 623–626.
(18) Josephson, B.; Fehl, C.; Isenegger, P. G.; Nadal, S.; Wright, T. H.; Poh, A. W. J.; Bower, B. J.; Giltrap, A. M.; Chen, L.; Batchelor-McAuley, C.; et al. Light-driven post-translational installation of reactive protein side chains. *Nature* **2020**, *585*, 530.
(19) Meledin, R.; Mali, S. M.; Singh, S. K.; Brik, A. Protein ubiquitination via dehydroalanine: development and insights into the diastereoselective 1,4-addition step. *Org. Biomol. Chem.* **2016**, *14*, 4817.
(20) Key, H. M.; Miller, S. J. Site- and Stereoselective Chemical Editing of Thiostrepton by Rh-Catalyzed Conjugate Arylation: New Analogues and Collateral Enantioselective Synthesis of Amino Acids. *J. Am. Chem. Soc.* **2017**, *139*, 15460.
(21) Galonić, D. P.; van der Donk, W. A.; Gin, D. Y. Oligosaccharide–Peptide Ligation of Glycosyl Thiolates with Dehydropeptides: Synthesis of S-Linked Mucin-Related Glycopeptide Conjugates. *Chem. – Eur. J.* **2003**, *9*, 5997.
(22) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **2008**, *130*, 5052.
(23) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angew. Chem., Int. Ed.* **2008**, *47*, 6399–6401.
(24) Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernández-González, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2011**, *2*, 1666.
(25) Yang, B.; Wang, N. X.; Schnier, P. D.; Zheng, F.; Zhu, H.; Polizzi, N. F.; Ittuveetil, A.; Saikam, V.; DeGrado, W. F.; Wang, Q.; Wang, P. G.; Wang, L. Genetically introducing biochemically reactive amino acids dehydroalanine and dehydrobutyrine in proteins. *J. Am. Chem. Soc.* **2019**, *141*, 7698–7703.
(26) Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev.-Syst. Biol.* **2012**, *4*, 565–583.
(27) Chambers, K. A.; Scheck, R. A. Bacterial virulence mediated by orthogonal post-translational modification. *Nat. Chem. Biol.* **2020**, *16*, 1043.
(28) Li, H. T.; Xu, H.; Zhou, Y.; Zhang, J.; Long, C. Z.; Li, S. Q.; Chen, S.; Zhou, J. M.; Shao, F. The phosphothreonine lyase activity of a bacterial type III effector family. *Science* **2007**, *315*, 1000–1003.
(29) Mazurkiewicz, P.; Thomas, J.; Thompson, J. A.; Liu, M.; Arbibe, L.; Sansonetti, P.; Holden, D. W. SpvC is a Salmonella effector with phosphothreonine lyase activity on host mitogen-activated protein kinases. *Mol. Microbiol.* **2008**, *67*, 1371.
(30) Zhu, Y. Q.; Li, H. T.; Long, C. Z.; Hu, L. Y.; Xu, H.; Liu, L. P.; Chen, S.; Wang, D. C.; Shao, F. Structural insights into the enzymatic mechanism of the pathogenic MAPK phosphothreonine lyase. *Mol. Cell* **2007**, *28*, 899–913.
(31) Chambers, K. A.; Abularrage, N. S.; Scheck, R. A. Selectivity within a family of bacterial phosphothreonine lyases. *Biochemistry* **2018**, *57*, 3790.
(32) Szymczak, L. C.; Sykora, D. J.; Mrksich, M. Using peptide arrays to profile phosphatase activity in cell lysates. *Chem. – Eur. J.* **2020**, *26*, 165.
(33) Wood, S. E.; Sinsinbar, G.; Gudlur, S.; Nallani, M.; Huang, C. F.; Liedberg, B.; Mrksich, M. A Bottom-Up Proteomic Approach to Identify Substrate Specificity of Outer-Membrane Protease OmpT. *Angew. Chem., Int. Ed.* **2017**, *56*, 16531.
(34) Kuo, H. Y.; DeLuca, T. A.; Miller, W. M.; Mrksich, M. Profiling deacetylase activities in cell lysates with peptide arrays and SAMDI mass spectrometry. *Anal. Chem.* **2013**, *85*, 10635.
(35) Szymczak, L. C.; Mrksich, M. Using peptide arrays to discover the sequence-specific acetylation of the histidine-tyrosine dyad. *Biochemistry* **2019**, *58*, 1810.
(36) Kightlinger, W.; Lin, L.; Rosztoczy, M.; Li, W.; DeLisa, M. P.; Mrksich, M.; Jewett, M. C. Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. *Nat. Chem. Biol.* **2018**, *14*, 627.
(37) Pluchinsky, A. J.; Wackelin, D. J.; Huang, X.; Arnold, F. H.; Mrksich, M. High throughput screening with SAMDI mass spectrometry for directed evolution. *J. Am. Chem. Soc.* **2020**, *142*, 19804.
(38) Webb, M. R. A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 4884.
(39) Szymczak, L. C.; Huang, C. F.; Berns, E. J.; Mrksich, M. *In Phosphatases*; Allen, K. N., Ed.; Elsevier Academic Press Inc: San Diego, 2018; 607.
(40) Richter, M. F.; Drown, B. S.; Riley, A. P.; Garcia, A.; Shirai, T.; Svec, R. L.; Hergenrother, P. J. Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature* **2017**, *545*, 299.
(41) Song, J.; Malwal, S. R.; Baig, N.; Schurig-Briccio, L. A.; Gao, Z.; Vaidya, G. S.; Yang, K.; Abutaleb, N. S.; Seleem, M. N.; Gennis, R. B.; Pogorelov, T. V.; Oldfield, E.; Feng, X. Discovery of prenyltransferase inhibitors with *in vitro* and *in vivo* antibacterial activity. *ACS Infect. Dis.* **2020**, *6*, 2979–2993.

https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414–425

(42) Gong, S.; Zhang, C.; Zhang, Y. RNA-align: quick and accurate alignment of RNA 3D structures based on size-independent TM-score(RNA). *Bioinformatics* **2019**, *35*, 4459−4461.
(43) Zhou, P.; Li, B. T.; Yan, Y. M.; Jin, B. W.; Wang, L. B.; Huang, S. Y. Hierarchical flexible peptide docking by conformer generation and ensemble docking of peptides. *J. Chem. Inf. Model.* **2018**, *58*, 1292−1302.
(44) Yan, Y.; Zhang, D.; Huang, S. Y. Efficient conformational ensemble generation of protein-bound peptides. *J. Cheminform.* **2017**, *9*, 59.
(45) Huang, S.-Y.; Zou, X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins-Struct. Funct. Bioinformatics* **2008**, *72*, 557.
(46) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105.
(47) Molloy, M. P.; Andrews, P. C. Phosphopeptide derivatization signatures to identify serine and threonine phosphorylated peptides by mass spectrometry. *Anal. Chem.* **2001**, *73*, 5387.
(48) Li, W.; Boykins, R. A.; Backlund, P. S.; Wang, G.; Chen, H.-C. Identification of phosphoserine and phosphothreonine as cysteic acid and β-methylcysteic acid residues in peptides by tandem mass spectrometric sequencing. *Anal. Chem.* **2002**, *74*, 5701.
(49) Dong, X. C.; Yuan, Z. J.; Qu, Y.; Gao, Y. X.; Pei, X.; Qi, Q. Q.; Pei, Y. J.; Li, J. Q.; Chen, Y. S.; Wang, C. H. An ATP-Cu(ii) catalyst efficiently catalyzes enantioselective Michael reactions in water. *Green Chem.* **2021**, *23*, 9876−9880.
(50) Pérez, J. M.; Postolache, R.; Castiñeira Reis, M.; Sinnema, E. G.; Vargová, D.; de Vries, F.; Otten, E.; Ge, L.; Harutyunyan, S. R. Manganese(I)-Catalyzed H-P Bond Activation via Metal-Ligand Cooperation. *J. Am. Chem. Soc.* **2021**, *143*, 20071−20076.
(51) Park, H. S.; Hohn, M. J.; Umehara, T.; Guo, L. T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Söll, D. Expanding the genetic code of Escherichia coli with phosphoserine. *Science* **2011**, *333*, 1151−1154.
(52) Rogerson, D. T.; Sachdeva, A.; Wang, K.; Haq, T.; Kazlauskaite, A.; Hancock, S. M.; Huguenin-Dezot, N.; Muqit, M. M. K.; Fry, A. M.; Bayliss, R.; et al. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* **2015**, *11*, 496.
(53) Zhang, M. S.; Brunner, S. F; Huguenin-Dezot, N.; Liang, A. D.; Schmied, W. H; Rogerson, D. T; Chin, J. W Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat. Methods* **2017**, *14* (7), 729.
(54) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry* (5th edition). Worth Publishers 2008.
(55) Antos, J. M.; Ingram, J.; Fang, T.; Pishesha, N.; Truttmann, M. C.; Ploegh, H. L. Site-specific protein labeling via sortase-mediated transpeptidation. *Curr. Protoc. Protein Sci.* **2009**, *Chapter 15*, Unit 15.3.
(56) Georgiou, G.; Stephens, D. L.; Stathopoulos, C.; Poetschke, H. L.; Mendenhall, J.; Earhart, C. F. Display of beta-lactamase on the Escherichia coli surface: outer membrane phenotypes conferred by Lpp'-OmpA'-beta-lactamase fusions. *Protein Eng.* **1996**, *9*, 239−247.
(57) Earhart, C. F. Use of an Lpp-OmpA fusion vehicle for bacterial surface display. *Methods Enzymol.* **2000**, *326*, 506−516.
(58) Sachdeva, A.; Wang, K. H.; Elliott, T.; Chin, J. W. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. *J. Am. Chem. Soc.* **2014**, *136*, 7785−7788.
(59) Italia, J. S.; Addy, P. S.; Erickson, S. B.; Peeler, J. C.; Weerapana, E.; Chatterjee, A. Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites. *J. Am. Chem. Soc.* **2019**, *141*, 6204.
(60) Canovas, C.; Moreau, M.; Bernhard, C.; Oudot, A.; Guillemin, M.; Denat, F.; Goncalves, V. Site-specific dual labeling of proteins on cysteine residues with chlorotetrazines. *Angew. Chem., Int. Ed.* **2018**, *57*, 10646.
(61) Huang, C. F.; Mrksich, M. Profiling Protein Tyrosine Phosphatase Specificity with Self-Assembled Monolayers for Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry and Peptide Arrays. *ACS Comb. Sci.* **2019**, *21*, 760.
(62) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Bioph. Biom.* **2000**, *29*, 291.
(63) Zhou, P.; Jin, B.; Li, H.; Huang, S. Y. HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res.* **2018**, *46* (W1), W443.



https://doi.org/10.1021/acschembio.1c00866
ACS Chem. Biol. 2022, 17, 414−425