Cite as: A. Yang *et al.*, *Science* 10.1126/science.aah4428 (2016).

A chemical biology route to site-specific authentic protein modifications

Aerin Yang,[1] Sura Ha,[1] Jihye Ahn,[2] Rira Kim,[1] Sungyoon Kim,[1] Younghoon Lee,[1] Jaehoon Kim,[2] Dieter Söll,[3]* Hee-Yoon Lee,[1]* Hee-Sung Park[1]*

[1]Department of Chemistry, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. [2]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. [3]Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06511, USA.

*Corresponding author. Email: dieter.soll@yale.edu (D.S.); hylee@kaist.ac.kr (H.-Y.L.); hspark@kaist.ac.kr (H.-S.P.)

Many essential biological processes are controlled by post-translational protein modifications. The inability to synthetically attain the diversity enabled by these modifications limits functional studies of many proteins. We designed a three-step approach for installing authentic post-translational modifications into recombinant proteins. We first used the established *O*-phosphoserine (Sep) orthogonal translation system to create a Sep-containing recombinant protein. The Sep residue is then dephosphorylated to dehydroalanine (Dha). Finally, Zn-Cu promoted conjugate addition to Dha of alkyl iodides enables chemo-selective carbon-carbon bond formation. To validate our approach we produced histone H3, ubiquitin and GFP variants with site-specific modifications, including different methylations of H3K79. The methylated histones stimulate transcription via histone acetylation. This approach offers a powerful tool to engineer diverse designer proteins.

Posttranslational modifications (PTMs) play vital roles in expanding protein functional diversity and critically affect numerous biological processes (*1*). The availability of proteins with specific modifications at selected residues is essential for experimental strategies to investigate fundamental biological mechanisms. Methods to generate diverse native protein covalent modifications currently do not exist. Genetic code expansion approaches are useful in producing recombinant proteins with specific modifications (*2*, *3*), but rely on the availability of an orthogonal tRNA•tRNA synthetase pair for acylation of a specific non-canonical amino acid. Despite much progress, the creation of many important protein modifications (*e.g.,* trimethyl lysine) is not yet feasible. Among chemical conjugation approaches (*4*), Cys-based strategies have been widely applied to generate protein conjugates (*5*) and mimics of PTMs (*6*, *7*). Yet the final products are PTM analogs whose value for searching out unidentified properties of the natural system may be questionable (*8*). Thus, despite extensive efforts, synthetic approaches for many authentic PTMs are not available, as no C-C bond forming reactions have been successfully applied to protein modifications despite the prevalence of such reactions in organic chemistry (*9*).

Here, we propose a 3-step strategy (Fig. 1A) that, in principle, is applicable to generate diverse forms of authentic and selective protein modifications. (i) The site of the intended modification is established by cotranslational Sep incorporation into a recombinant protein using the Sep orthogonal *Escherichia coli* translation system (*10*, *11*). (ii) Then the Sep residue of the purified recombinant protein is converted by phosphate removal to Dha which in turn serves as a radicalophile enabling a bio-orthogonal chemical reaction. (iii) Finally, PTM moieties are directly coupled to Dha via metal-mediated conjugate additions of alkyl iodides in aqueous solution facilitating chemo-selective carbon-carbon bond formation in proteins (Fig. 1B).

To demonstrate the feasibility and versatility of our strategy, we set out to generate recombinant *Xenopus laevis* histone H3K79 with five different modifications—monomethylation, dimethylation, trimethylation, formylation, and acetylation. Reversible lysine methylation in proteins presents the most complex and dynamic modification (*12*). Several approaches have attempted Lys methylation (*13*), but the trimethylated product proved elusive. Histone H3K79 appears to be dynamically regulated (*14*) and associated with diverse cellular processes; its exact role has not been fully examined owing to the lack of generating such authentically modified histones.

First, the Sep-containing histone H3Sep79 was made by expressing a *X. laevis* histone H3 mRNA containing a UAG codon at position 79 and a coding sequence for a C-terminal His_6 tag in *E. coli* containing the orthogonal Sep translation system (the engineered SepRS9•$tRNA^{Sep}$ pair and the evolved EF-Sep21) (*11*). H3Sep79 was routinely obtained in good yield (~ 20 mg/L of culture) (fig. S1). The incorporation of Sep at the intended position 79 was confirmed by MALDI-TOF mass spectrometry (MS) analysis of the purified recombinant histone (fig. S2) and its tryptic peptides

(Fig. 2A and tables S1 and S2).

As phosphoamino acids are known to be labile under alkaline conditions (*15*), we attempted to convert Sep to Dha by mild alkali treatment. To determine the optimal procedure, purified H3Sep79 was incubated with LiOH, $Ba(OH)_2$ or $Sr(OH)_2$ under different conditions, followed by neutralization with acetic acid and dialysis against distilled water (fig. S3). The reaction progress was analyzed by MALDI-TOF MS after in-gel trypsin digestion (tables S1 and S2). Near-complete disappearance of the Sep residue and concurrent generation of Dha took 30 min at room temperature in 40 mM $Ba(OH)_2$ solution. Mass analysis of whole protein (fig. S4) and tryptic peptides (Fig. 2A) demonstrated that these conditions led to highly selective chemical transformation of H3Sep79 into H3Dha79 with no noticeable side reactions including protein oxidation (*16*).

Dha-mediated conjugation approaches have been successfully applied to the generation of thiol-linked PTM analogs or mimics (*17*, *18*). To produce authentic PTMs from Dha, we needed a new coupling scheme enabling chemoselective carbon-carbon bond formation in proteins (Fig. 1B). We were encouraged by reports of water-based organic reactions in which an alkene group functions as a radical acceptor (*19*). Knowing that Dha could function as a radical acceptor, we chose alkyl radicals which can be generated from alkyl iodides by transition metals (*19*). Inspired by recent metal-mediated conjugate addition reactions (*20*), we reasoned that organozinc species would form from the precursor halides and Zn metal. Subsequently zinc-to-copper transmetalation would occur, generating organocopper reagents, which would lead to formation of alkyl radical species. Finally, conjugate coupling of the alkyl radical with the Dha residue of a protein would take place (Fig. 1B). To demonstrate the proposed coupling scheme, we tested first the addition of trimethyl iodide (3-iodo-N,N,N,-trimethylpropan-1-amine, **3**; fig. S5) to H3Dha79 to generate the expected product H3K79me3. The reaction products were analyzed by Western blotting using an anti-H3K79me3 antibody. Our initial attempts of directly employing water-based organic reactions were unsuccessful. After testing a large array of conditions (described in the supplementary materials, figs. S6 to S16): different buffers and pH ranges, surfactants for protein stabilization, essential reagents for the metal-mediated coupling reaction (Zn metals, copper salts), and possible auxiliary reagents [*e.g.*, tetramethylethylenediamine (TMEDA)] we arrived at conditions that reproducibly led to the formation of C-C bonds in high yield. The deduced optimal conditions were: H3Dha79 (10 μM), alkyl iodide (30 mM), Zn powder (0.4 mg), $Cu(OAc)_2$ (1 mM), Triton X-100 (2.0 wt%), and TMEDA (10 mM) in sodium acetate (pH 4.5, 0.5M). Reactions (20-50 μL) were incubated at room temperature.

With optimized reaction conditions in hand, we attempted to synthesize H3K79me1, H3K79me2, and H3K79me3 by incubating H3Dha79 with three different methyl iodides: monomethyl iodide (3-iodo-N-methylpropan-1-amine, **1**), dimethyl iodide (3-iodo-N,N-dimethylpropan-1-amine, **2**) or trimethyl iodide (Fig. 2C and fig. S5). MALDI-TOF MS analysis revealed disappearance of Dha and concomitant generation of methylated lysine residues (Fig. 2A). The coupling reactions were highly selective and efficient (normally >80%), as demonstrated by mass analyses of tryptic peptides (Fig. 2A and table S2) and whole proteins (fig. S17), and had a good recovery yield (between 50-70%). Selective and differential Lys methylations were also demonstrated by Western blot analysis using anti-H3K79me1, anti-H3K79me2, and anti-H3K79me3 antibodies (Fig. 2B). To determine whether other PTM moieties could be coupled onto H3Dha79, we synthesized formyl iodide (N-(3-iodopropyl)formamide, **4**) and acetyl iodide (N-(3-iodopropyl)-acetamide, **5**; fig. S5), and used them for producing histones H3K79N^{ε}-formyl and H3K79N^{ε}-acetyl, respectively, as verified by MALDI-TOF MS analysis (fig. S18).

To demonstrate whether our 3-step synthesis can be utilized for modification of other proteins, we set out to generate ubiquitin variants with site-specific modifications. Eight ubiquitin variants with various Lys modifications (acetylation and differential methylations) at position 33 or 48 were efficiently generated (fig. S19 and tables S3 to S5). Also, we found that different alkyl iodides (iodoethane, **6**; 2-iodopropane, **7**; 2-iodo-2-methylpropane, **8**; 1-iodobutane, **9**; iodocyclopentane, **10**; 3-iodopronionic acid, **11**; and *tert*-butyl-iodobutoxydimetylsilane, **12**) (fig. S5) were efficiently conjugated to Ub33Dha (fig. S20 and table S4). In particular, coupling of dansyl iodide (5-(dimethylamino)-N-(2-((3-iodopropyl)amino)ethyl)naphthalene-1-sulfonamide, **13**; fig. S5) onto Ub33Dha led to site-specific fluorescent dye labeling via carbon-carbon bond formation (figs. S20 and S21 and table S4). Lastly, site-specific Lys204 trimethylation of green fluorescence protein (GFP) by our 3-step synthetic route led to the desired product GFPK204me3 but with lower conversion efficiency (~ 20%) and recovery yield (~ 30%) compared to histone H3 and ubiquitin (fig. S22 and tables S6 and S7). Clearly the structural context within the protein may affect the ease of modification. These data show that our approach is applicable to a variety of proteins. Since each protein behaves differently with the diverse chemical reagents, individualized optimization should lead to efficient implementation of our synthetic strategy with other proteins.

Next, we performed biochemical assays with the modified histones generated by our 3-step approach to examine if they are fully functional in the biological context. Genome-wide analyses of chromatin revealed that H3K79 methyla-

tion is enriched in actively transcribing regions (*14*), but its detailed role is not yet well understood. A transcription assay using in vitro assembled chromatin templates and a highly purified RNA polymerase II transcription apparatus provides a biochemically defined system to study the functions of individual histone modifications. The modified histones H3K79me1, H3K79me2, and H3K79me3 were first assembled with H2A, H2B, and H4 to form histone octamers. The methylated histones displayed no noticeable difference compared with the unmodified (intact) histone during octamer assembly and nucleosome reconstitution (fig S23). Recombinant chromatin templates were reassembled with the histone octamers and a p53ML plasmid and were applied to p53 (activator)- and p300 (coactivator)-dependent in vitro transcription assays (Fig. 3A) (*21*). Intact chromatin lacking methylation showed a low level of transcription only in the presence of both activator and coactivator as expected (Fig. 3B). Notably, the levels of basal (activator- and/or coactivator-independent) transcription from chromatins with all three methylation states were greatly enhanced (Fig. 3B). As well, activator- or coactivator-dependent transcription from H3K79-methylated chromatin was elevated (Fig. 3B), demonstrating a direct stimulatory effect of H3K79 methylation on chromatin transcription. More importantly, we also found that H3K79 methylation increased histone acetylation mediated by p300 (Fig. 3C). Interestingly, histone acetylation was differentially affected by the level of H3K79 methylation (Fig. 3C). Thus, H3K79 methylation is indeed positively associated with transcription activation via p300-mediated histone acetylation, which is differently affected by the various methylation states (Fig. 3D). These results illustrate that the methylated histones are fully active and functionally distinct depending on the modification level, demonstrating the utility and gentle nature of our 3-step approach.

An open question concerns the diastereoselectivity of our coupling reaction. It is known that radical and thiol conjugate additions to Dha will lead to epimeric mixtures (*22*–*24*); a firm analysis of the epimeric ratio will require high-resolution crystallography of the protein. However, the steric context of the local protein conformation may significantly affect the final diastereomeric ratio of the products. The fact that our synthetic proteins H3K79me1, H3K79me2, and H3K79me3 were well recognized by antibodies, and that these modified histones could assemble into octamers with biological in vitro activity, underscores the utility of our approach.

We anticipate that with well-tailored alkyl iodides [*e.g.,* (*25*)] our approach can be extended to produce designer proteins with diverse forms of chemical modifications (*e.g.*, glycosylated amino acids, phosphotyrosine). Such efforts will markedly expand the available chemical diversity in proteins and facilitate the study of many PTM-mediated biological processes.

REFERENCES AND NOTES

1. C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto Jr., Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* **44**, 7342–7372 (2005). Medline doi:10.1002/anie.200501023
2. K. Lang, J. W. Chin, Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **114**, 4764–4806 (2014). Medline doi:10.1021/cr400355w
3. C. C. Liu, P. G. Schultz, Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010). Medline doi:10.1146/annurev.biochem.052308.105824
4. C. D. Spicer, B. G. Davis, Selective chemical protein modification. *Nat. Commun.* **5**, 4740 (2014). Medline doi:10.1038/ncomms5740
5. E. V. Vinogradova, C. Zhang, A. M. Spokoyny, B. L. Pentelute, S. L. Buchwald, Organometallic palladium reagents for cysteine bioconjugation. *Nature* **526**, 687–691 (2015). Medline doi:10.1038/nature15739
6. M. D. Simon, F. Chu, L. R. Racki, C. C. de la Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar, K. M. Shokat, The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003–1012 (2007). Medline doi:10.1016/j.cell.2006.12.041
7. S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, B. G. Davis, Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446**, 1105–1109 (2007). Medline doi:10.1038/nature05757
8. D. P. Nguyen, M. M. Garcia Alai, S. Virdee, J. W. Chin, Genetically directing ε-N, N-dimethyl-L-lysine in recombinant histones. *Chem. Biol.* **17**, 1072–1076 (2010). Medline doi:10.1016/j.chembiol.2010.07.013
9. D. J. Hart, Free-radical carbon-carbon bond formation in organic synthesis. *Science* **223**, 883–887 (1984). Medline doi:10.1126/science.223.4639.883
10. H. S. Park, M. J. Hohn, T. Umehara, L. T. Guo, E. M. Osborne, J. Benner, C. J. Noren, J. Rinehart, D. Söll, Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* **333**, 1151–1154 (2011). Medline doi:10.1126/science.1207203
11. S. Lee, S. Oh, A. Yang, J. Kim, D. Söll, D. Lee, H. S. Park, A facile strategy for selective incorporation of phosphoserine into histones. *Angew. Chem. Int. Ed. Engl.* **52**, 5771–5775 (2013). Medline doi:10.1002/anie.201300531
12. E. L. Greer, Y. Shi, Histone methylation: A dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357 (2012). Medline doi:10.1038/nrg3173
13. W. R. Liu, Y. S. Wang, W. Wan, Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach. *Mol. Biosyst.* **7**, 38–47 (2011). Medline doi:10.1039/C0MB00216J
14. A. T. Nguyen, Y. Zhang, The diverse functions of Dot1 and H3K79 methylation. *Genes Dev.* **25**, 1345–1358 (2011). Medline doi:10.1101/gad.2057811
15. Y. Oda, T. Nagasu, B. T. Chait, Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379–382 (2001). Medline doi:10.1038/86783
16. J. Wang, S. M. Schiller, P. G. Schultz, A biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem. Int. Ed. Engl.* **46**, 6849–6851 (2007). Medline doi:10.1002/anie.200702305
17. J. Guo, J. Wang, J. S. Lee, P. G. Schultz, Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angew. Chem. Int. Ed. Engl.* **47**, 6399–6401 (2008). Medline doi:10.1002/anie.200802336
18. J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield, B. G. Davis, Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. *Angew. Chem. Int. Ed. Engl.* **51**, 1835–1839 (2012). Medline doi:10.1002/anie.201106432
19. A. Postigo, N. S. Nudelman, Synthetically useful metal-mediated radical transformations in water and aqueous media. *Coord. Chem. Rev.* **255**, 2991–3030 (2011). doi:10.1016/j.ccr.2011.07.015
20. B. H. Lipshutz, S. Huang, W. W. Leong, G. Zhong, N. A. Isley, C-C bond formation via copper-catalyzed conjugate addition reactions to enones in water at room temperature. *J. Am. Chem. Soc.* **134**, 19985–19988 (2012). Medline

doi:10.1021/ja309409e
21. J. Kim, M. Guermah, R. G. Roeder, The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010). Medline doi:10.1016/j.cell.2009.12.050
22. D. Crich, J. W. Davies, Free-radical addition to di- and tripeptides containing dehydroalanine residues. *Tetrahedron* **45**, 5641–5654 (1989). doi:10.1016/S0040-4020(01)89507-7
23. Y. Zhu, W. A. van der Donk, Convergent synthesis of peptide conjugates using dehydroalanines for chemoselective ligations. *Org. Lett.* **3**, 1189–1192 (2001). Medline doi:10.1021/ol015648a
24. M. M. Müller, T. W. Muir, Histones: At the crossroads of peptide and protein chemistry. *Chem. Rev.* **115**, 2296–2349 (2015). Medline doi:10.1021/cr5003529
25. J. M. Garcia Fernandez, C. O. Mellet, M. A. Pradera Adrian, J. F. Mota, Syntheses of β-iodourea derivatives of carbohydrates and glycosylamino-oxazolines. *Carbohydr. Res.* **216**, 21–32 (1992). doi:10.1016/0008-6215(92)84147-K
26. R. Appel, Tertiary phosphane-tetrachloromethane, a versatile reagent for chlorination, dehydration, and P-N linkage. *Angew. Chem. Int. Ed. Engl.* **14**, 801–811 (1975). doi:10.1002/anie.197508011
27. H. Finkelstein, Darstellung organischer Jodide aus den entsprechenden Bromiden und Chloriden. *Ber. Dtsch. Chem. Ges.* **43**, 1528–1532 (1910). doi:10.1002/cber.19100430257
28. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006). Medline doi:10.1038/nprot.2006.468

ACKNOWLEDGMENTS

We thank M. Simon and S. Miller for stimulating discussions. This work was supported by grants from the National Research Foundation of Korea (Systems Healthcare, 2011-0020322, and 2014M3A6A4075060 to H.-S.P.; 2012M3A9B4027956 and 2012M3A9C6049937 to J.K.), and by the US National Institute for General Medical Sciences (GM22854-42 to D.S.). All data are available in the supplementary materials. H.-S.P., H.-Y.L., A.Y., and S.H. are inventors on patent application #62/383,245 submitted by KAIST that covers protein chemical modifications mediated by carbon-carbon bond formation. Author contributions: H.-S.P., H.-Y.L., and A.Y. conceived and designed the study; H.-S.P. and A.Y. performed and analyzed all experiments except chemical synthesis and in vitro transcription assay, which were conducted by S.H., R.K, Y.L., J.A., and J.K; H.-S.P., H.-Y.L., and D.S. supervised the project. H.-S.P. and D.S. wrote the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aah4428/DC1
Materials and Methods
Figs. S1 to S23
Tables S1 to S7
References (*26*–*28*)

27 June 2016; accepted 20 September 2016
Published online 29 September 2016
10.1126/science.aah4428

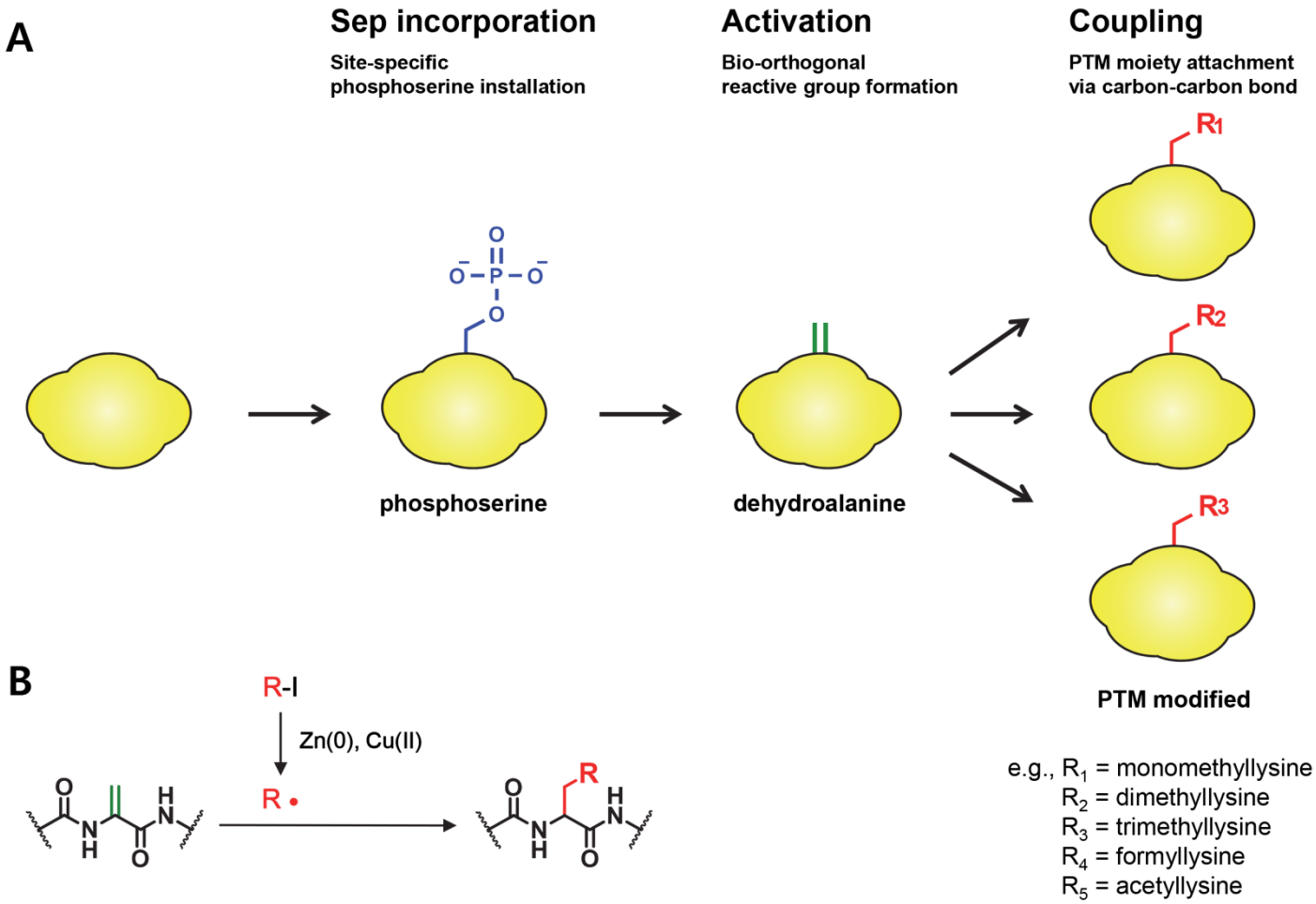


Fig. 1. Scheme for protein chemical modifications. (A) Schematic representation of the 3-step synthesis of proteins with authentic PTMs. (B) New coupling scheme enables chemo-selective carbon-carbon bond formation in proteins.

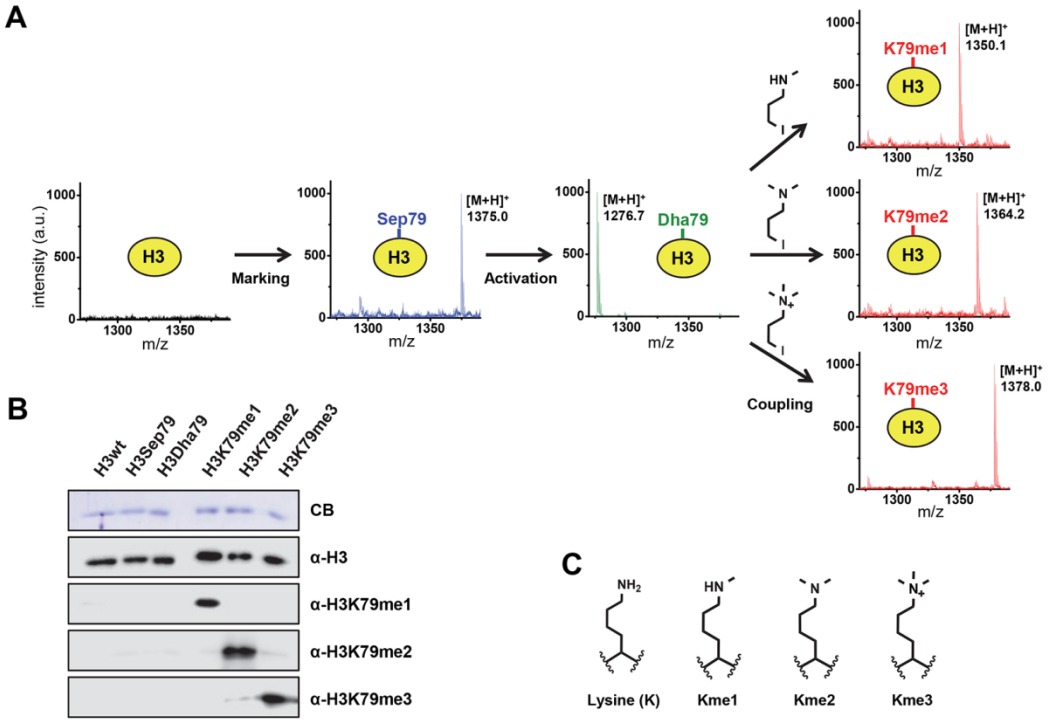


Fig. 2. Synthesis of proteins with selective differential methylations. (**A**) MALDI-TOF MS analysis of proteins after trypsin digestion. Formation of H3Sep79 evidenced by the Sep-containing peptide (Sep79, blue). Synthesis of H3Dha79 generates a new Dha-containing peptide (Dha79, green), but eliminates the Sep-containing peptide. Coupling of the methyl iodides onto H3Dha79 produces new methylated lysine-containing peptides, K79me1, K79me2, and K79me3 (red). (**B**) Western blot analysis of the modified proteins using anti-H3K79me1, anti-H3K79me2, and anti-H3K79me3 antibodies. (**C**) Chemical structures of Lys and differentially methylated Lys residues.

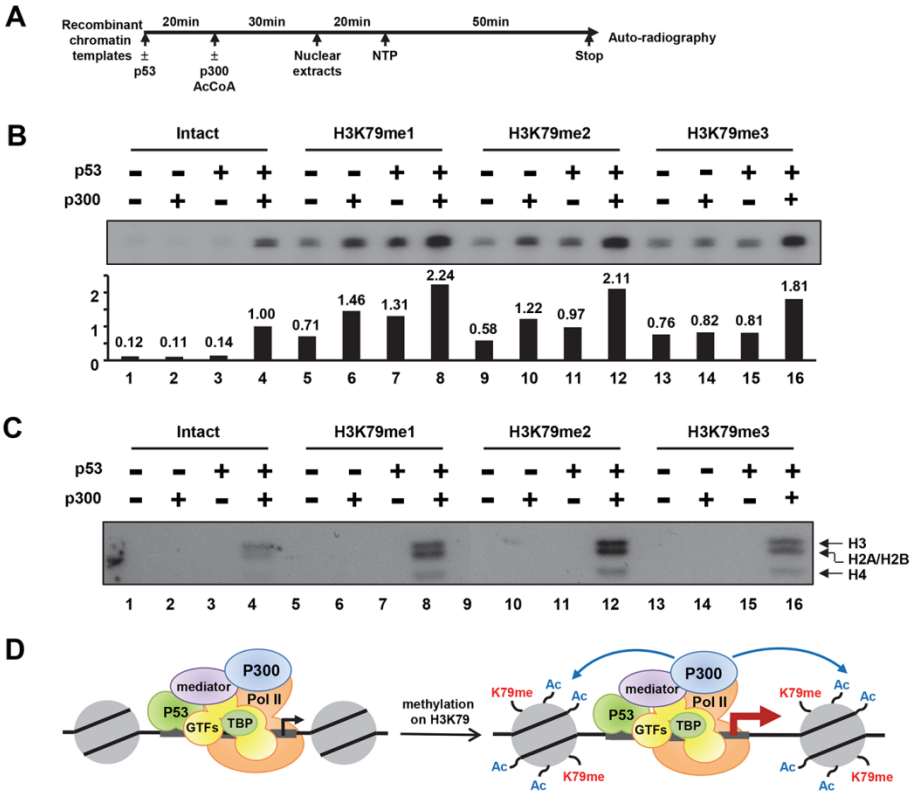
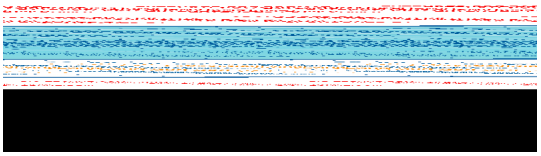


Fig. 3. Effects of H3K79 methylation on chromatin transcription. (A) Schematic of the standard in vitro transcription assay. (B) Effect of methylated H3K79 on chromatin transcription. (C) Effect of methylated H3K79 on p300-mediated chromatin acetylation. Histone acetylation status was monitored by fluorography. (D) Schematic representation of transcriptional activation by H3K79 methylation.



A chemical biology route to site-specific authentic protein modifications

Aerin Yang, Sura Ha, Jihye Ahn, Rira Kim, Sungyoon Kim, Younghoon Lee, Jaehoon Kim, Dieter Söll, Hee-Yoon Lee and Hee-Sung Park

(September 29, 2016)
published online September 29, 2016

Editor's Summary

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.