Expressed Protein Ligation Without Intein

Yuchen Qiao[a], Ge Yu[a], Xiaoyan Aria Wang[a], Kaci C. Kratch[a], Wesley Wei Wang[a], Jared S. Morse[a], Sunshine Z. Leeuwon[a], and Wenshe Ray Liu[a,b,c]*

[a]The Texas A&M Drug Discovery Laboratory, Department of Chemistry, Texas A&M University, College Station, TX 77843, USA

[b]Department of Biochemistry & Biophysics, Texas A&M University, College Station, TX 77843, USA

[c]Molecular & Cellular Medicine Department, College of Medicine, Texas A&M University, College Station, TX 77843, USA

*Correspondence should be addressed to Wenshe Ray Liu: wliu@chem.tamu.edu

ABSTRACT: Proteins with a functionalized *C*-terminus such as a *C*-terminal thioester are key to the synthesis of larger proteins via expressed protein ligation. They are usually made by recombinant fusion to intein. Although powerful, the intein fusion approach suffers from premature hydrolysis and low compatibility with denatured conditions. To totally bypass the involvement of an enzyme for expressed protein ligation, here we showed that a cysteine in a recombinant protein was chemically activated by a small molecule cyanylating reagent at its *N*-side amide for undergoing nucleophilic acyl substitution with amines including a number of L- and D-amino acids and hydrazine. The afforded protein hydrazides could be used further for expressed protein ligation. We demonstrated the versatility of this approach with the successful synthesis of ubiquitin conjugates, ubiquitin-like protein conjugates, histone H2A with a posttranslational modification, RNAse H that actively hydrolyzed RNA, and exenatide that is a commercial therapeutic peptide. The technique, which is exceedingly simple but highly useful, expands to a great extent the synthetic capacity of protein chemistry and will therefore make a large avenue of new research possible.

The native chemical ligation concept was first developed by Dawson et al. in 1994, in which one protein or peptide with a *C*-terminal thioester and the other with a *N*-terminal cysteine selectively undergo thiol-thioester exchange and then S-to-N acyl transfer to form a larger protein or peptide (Figure 1A)[1]. Given that a protein with a *N*-terminal cysteine can be recombinantly produced, the development of the concept made it feasible to synthesize large proteins with a functionalized *N*-terminus to include either posttranslational or purely chemical modifications. To expand the synthetic scope of native chemical ligation, a related technique termed expressed protein ligation in which a recombinant *C*-terminal intein fusion protein is used to generate a protein thioester was also developed for the synthesis of proteins with a functionalized *C*-terminus[2]. Another notable related technique is peptide hydrazide ligation that uses nitrous acid or acetyl acetone to convert a chemically stable peptide hydrazide to a peptide acyl azide or a peptide acyl pyrazole and then a peptide thioester for further native chemical or expressed protein ligation (Figure 1A)[3-4]. Evident by the original publication garnering more than 2500 citations so far, the advent of native chemical and expressed protein ligation techniques has revolutionized the protein and peptide chemistry field. Groundbreaking applications include the synthesis of a large variety of proteins such as histones, kinases, and RAS proteins with posttranslational modifications for driving basic research advances and the production of many proteins or enzymes for therapeutic and biotechnological purposes[5-12]. Although developed extensively, further technological improvement in protein ligation is still necessary. The production of a protein thioester using the intein fusion approach is not guaranteed for a lot of proteins. The stringent requirement for intein catalysis to generate a protein thioester prevents the processing of many fusion proteins that are expressed insolubly and hard to fold[13]. The *C*-terminal residue of a targeted protein that is immediate to the intein *N*-terminus also significantly impacts the protein splicing efficiency, which leads to low splicing efficiency for residues such as proline at this site[14-15]. The purification of an intein fusion also requires significant caution for avoiding premature hydrolysis[2, 16]. A split intein may be used to prevent premature hydrolysis but adds more procedural complexity[5, 17]. Using a protein ligase for expressed protein ligation resolves some issues related to the intein fusion approach but requires a specific amino acid sequence context at the ligation site[18]. Therefore, a simple method to functionalize a recombinant protein at its *C*-terminus for expressed protein ligation that requires no enzymatic catalysis, can be broadly applied, and maintains high efficiency in different protein *C*-terminal sequence contexts is highly desired. In this work, we report such a method and its application in the synthesis of a number of proteins or peptides that can be used in both basic research and therapy.

Using a cyanylating reagent to directly activate a cysteine in a protein, early protein chemists showed that the protein was hydrolyzed at the cysteine *N*-side amide[19-20]. Although there has not been a definitive mechanism for this hydrolysis process, we reason that a thiocyanate generated from the covalent attachment of a cyanide to a protein cysteine potentially undergoes intramolecular addition with the cysteine *N*-amide to generate a 1-acyl-2-iminothiazolidine intermediate. In comparison to a regular protein amide, the amide bond in this intermediate is much weaker and therefore prone to hydrolysis (Figure 1B)[21-22]. If the reaction undergoes according to this mechanism, providing a strongly nucleophilic amine in the reaction will trigger nucleophilic acyl substitution with the 1-acyl-2-iminothiazolidine intermediate to replace 2-iminothiazolidine and potentially circumvent the hydrolysis process. The afforded small molecule amine-ligated product can be used for a further protein ligation process. Since this proposed expressed protein ligation doesn't involve an enzyme and is purely chemically based, it can be highly controllable, selective, and versatile such as functioning for proteins both soluble and insoluble.

To demonstrate the feasibility of this new ligation reaction, we synthesized Ac-Xxx-Cys-NH_2 dipeptides in which the Xxx identity varied between seven native amino acids including proline and carried out their reactions with an equivalent amount of NTCB and then ligation with propargylamine in a 1:1 DMF/H2O solvent. Our result showed that all dipeptides reacted with propargylamine to form desired products with varied yields (Table S1).

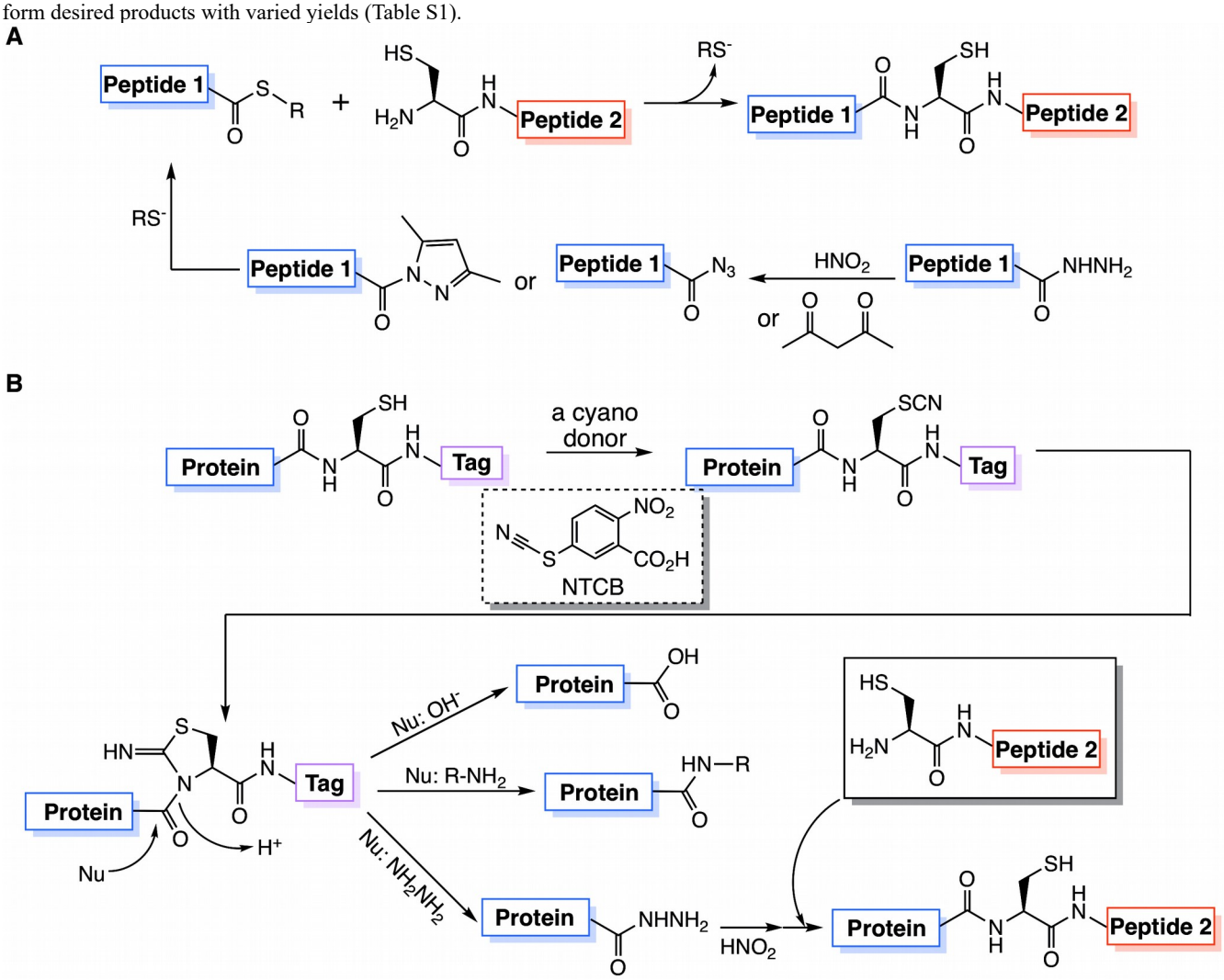

Figure 1. Protein synthesis by ligation techniques. (**A**) Native chemical ligation and a derivative technique, peptide hydrazide ligation; (**B**) A proposed protein ligation technique based on nucleophilic acyl substitution of an activated cysteine residue in a recombinant protein with a nucleophilic amine. Without a nucleophilic amine, the protein undergoes hydrolysis. When the nucleophile is hydrazine, the afforded protein hydrazide can then undergo peptide hydrazide ligation to form a larger protein.

Encouraged by our small molecule results, we named this new expressed protein ligation technique as activated-cysteine directed protein ligation (ACPL) and tested it further with recombinant proteins. Ubiquitin (Ub) is natively devoid of cysteine[23]. We chose it as a model protein for our demonstration. We produced recombinant native Ub and Ub with both a G76C mutation and a *C*-terminal 6×His tag (Ub-G76C-6H) in E. coli and purified them to homogeneity. We then ligated Ub-G76C-6H with 12 small molecule amines including propargylamine (Pa), allylamine (Aa), hydrazine (Ha), and L- and D-amino acids (Figure 2A) by adding 5 mM NTCB and a 50-1000 mM amine simultaneously to a 2 mg/mL Ub-G76C-6H solution at pH 9 for an overnight incubation at 37 ˚C. We carefully selected seven L-amino acids for our reactions to represent amino acids in different chemical categories and also different sizes. For all tested compounds including proline that has a secondary amine and two D-amino acids, we obtained ligation products with 50-90 % yields estimated through the intensity comparison of the unreacted and reacted bands as we demonstrated by SDS-PAGE analysis of their reaction mixtures (Figure S1, Table S2). After using Ni^{2+} charged resins to simply remove unreacted intermediates, we analyzed all 12 ligation products and the two original Ub and Ub-G76C-6H proteins by electrospray ionization mass spectrometry (ESI-MS) analysis. For all analyzed proteins, their deconvoluted ESI-MS spectra displayed clearly observable monoisotopic peaks. Since there is no commercial software for calculating protein monoisotopic peaks, we wrote a Python script to calculate all theoretical monoisotopic masses and their relative intensities for all proteins and compared them to the determined ESI-MS spectra. Our results showed that determined monoisotopic masses for all proteins agreed very well with their theoretic values in terms of both molecular weight and intensity (Figures S2-S15). Hydrolysis products were either non-detectable or at very low levels. To simplify the comparison, we wrote another Python script to integrate deconvoluted monoisotopic peaks and then calculate the average molecular weights and intensities for all detected protein species in a particular spectrum. The final results are presented in Figures 2B and 2C. For all determined average molecular weights, they matched their theoretical values with a deviation of ± 0.3 Da (Table S2). For all 12 ligation products, we detected very few minor peaks in their ESI-MS spectra indicating that all reactions were very selective. One ligation product Ub-G76G is native Ub itself. Its ESI-MS spectrum in Figure 2C matched that of recombinantly expressed native Ub in Figure 2B. So far, our data demonstrated that ACPL works exactly according to what we proposed on a

recombinant protein and this reaction is effective for amines that are primary, secondary, hydrazine, and amino acids with different configurations, characteristics, and sizes. The ligation with hydrazine was done in both native and denatured conditions. The results from two conditions showed minimal differences (Figures S15 and S16). Ubiquitin natively has a G75 residue that has the lowest steric hindrance among all amino acids. In Ub-G76C-6H, the glycine immediately *N*-terminal to G76C might have permitted easy processing of the ligation. Other residues that have different chemical properties and/or are sterically hindered might impede the ligation. To resolve this concern, we mutated G75 in Ub-G76C-6H to six other residues that are large in size, charged, and/or having a secondary amine, recombinantly expressed them, analyzed them with ESI-MS (Figures 2D and S17-S22), and then reacted them in a one-pot fashion with NTCB and hydrazine. We chose hydrazine in our demonstration since its ligation products are protein hydrazides that can undergo further peptide hydrazide ligation for making even larger proteins. All

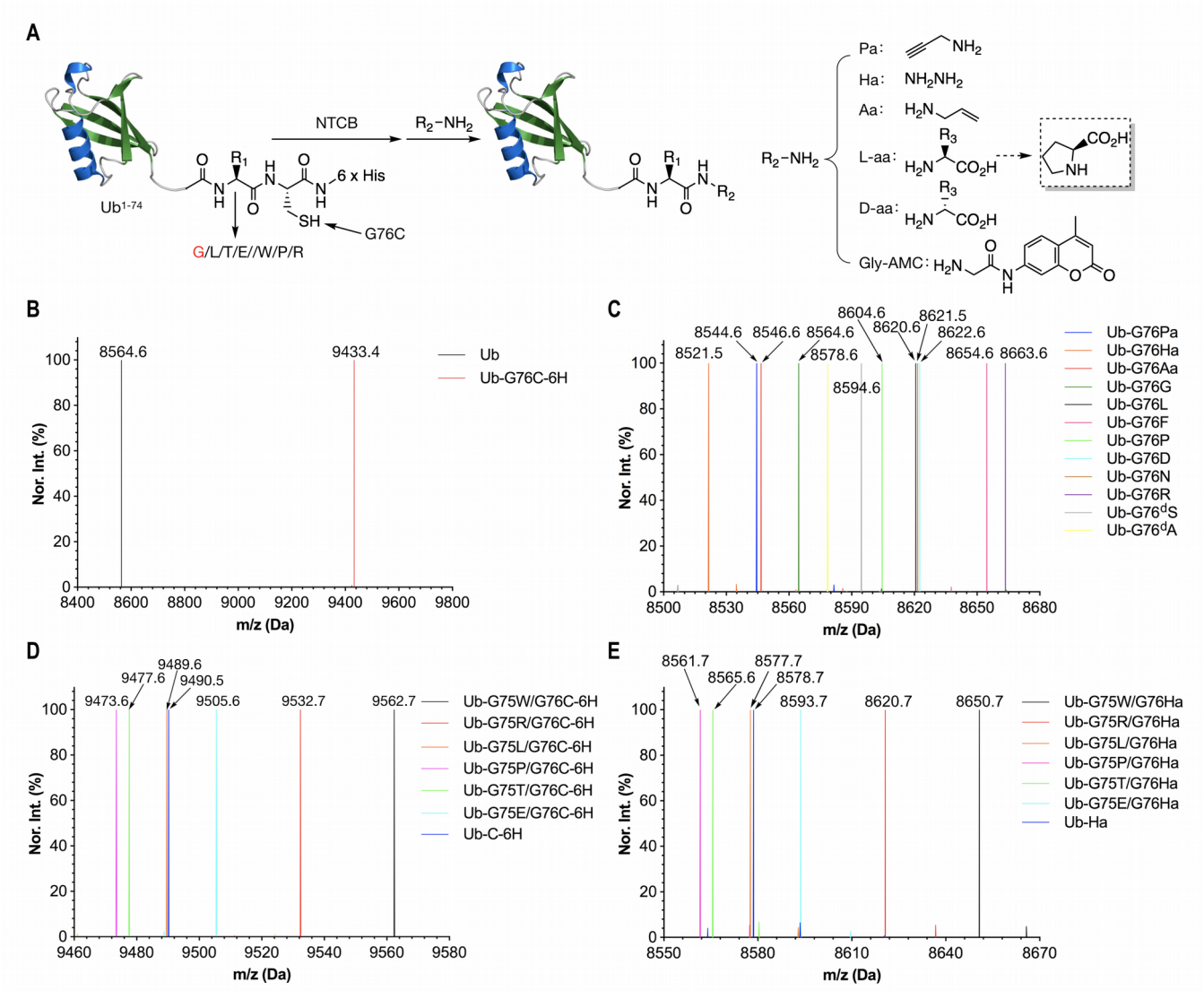

Figure 2. The synthesis of Ub conjugates by activated cysteine-directed protein ligation. **(A)** A schematic diagram to show the activation of recombinant Ub proteins containing a cysteine by NTCB followed by nucleophilic acyl substitution with amines, both primary and secondary, to generate different Ub conjugates. The native Ub has 76 residues and glycine at the 75th and 76th positions. **(B)** The deconvoluted and integrated ESI-MS of wild type Ub and Ub-G76C-6H. 6H represents a 6$^{\text{x}}$ His tag. **(C)** The deconvoluted and integrated ESI-MS of Ub conjugates that were converted from Ub-G76C-6H and had different ligated molecules at the G76 position. Pa, Ha, and Aa are three small molecule amines shown in A. All other ligated molecules are amino acids whose one letter codes are used for labeling. All amino acids are in the L-configuration except two D-amino acids with a footnote d. **(D-E)** The deconvoluted and integrated ESI-MS of 7 recombinant Ub proteins and products of their reactions with NTCB and Ha. C in Ub-C-6H represents cysteine. All detected molecular weights agreed well with theoretic values in a deviation range of ± 0.3 Da.

reactions progressed exceedingly well and their reaction products displayed average molecular weights matching well to their theoretic values (Figures 2E and S23-S28; Table S3), demonstrating that the residue immediately *N*-terminal to the targeted cysteine has little detrimental effect on the ligation process. The visible difference of reaction yield from 20 % (G75W mutant) to 70 % (G75T mutant) also revealed the potential steric effect on the crucial iminothiazolidine intermediate. Putting a cysteine residue right after Ub G76 led to similar ligation results with hydrazine (Figures 2D, 2E, S29, and S30; Table S3) and with allylamine, propargylamine and glycine with at least 50 % conversion rate (Figures S31-33 and Table S3). In addition, since Ub has a flexible *C*-terminus that may facilitate the ligation, to show that the ligation may work in a more structurally constrained environment, we introduced a cysteine mutation at K48 and K63, two residues in the globular region of Ub and used the two afforded Ub mutants (Figures S34-S35 and Table S3) to undergo ACPL with hydrazine. ESI-MS of reaction mixtures showed successful formation of two desired protein hydrazides (Figures S36-S37 and Table S3) indicating that ACPL works well in a structurally constrained protein region. Ligation

both in a structurally constrained protein region and under a denatured condition is something that the traditional intein and ligase-based methods cannot perform well. Collectively our data strongly demonstrates the versatility of the ACPL technique.

It has been shown that replacing the *C*-terminal glycine in Ub, SUMO1-3, NEED8, and ISG15 with Pa using either the intein based approach or total synthesis afforded excellent probes to conjugate covalently to deubiquitinases (DUBs) or ubiquitin-like proteases (ULPs) that catalytically remove Ub or Ubls from their conjugated proteins in cells[24-28]. To recapitulate these results and demonstrate the broad application scope of our ACPL technique in the probe synthesis, we recombinantly expressed Ub, SUMO1-4, NEDD8, ISG15, GABARAP, GABARAPL2, UFM1, URM1, and MNSFβ (FLAG- Ub/Ubl-GxC-6H: x denotes the terminal glycine position) that all contained a C-terminal Gly-to-Cys mutation and were also fused with a *N*-terminal FLAG tag and a *C*-terminal 6×His tag, purified them to homogeneity, and then carried out their reactions with Pa in the presence of NTCB to afford their Pa-conjugated products. ISG15, SUMO1-4, and MNSFβ natively contain a cysteine residue. This cysteine was mutated to alanine or serine to avoid non-targeted reaction at its location. The yields of all those 12 reactions were ranged from 25 % to 80 % as indicated by SDS-PAGE analysis (Figure S1, Table S4). ESI-MS analysis of all 12 products indicated their successful and efficient synthesis (Figures 3A and S38-S61; Table S4). The final results are presented in Figures 3A that displayed very little side products for all 12 Pa-conjugated products. In comparison to both intein based and total synthesis approach, our method for the synthesis of these propargylamine conjugates is much simpler and easier to control. To reproduce some literature results, then we used our synthesized Pa-conjugated FLAG-Ub, FLAG-Ub-G76Pa to react with seven DUBs and observed efficient covalent adduct formation for all tested enzymes by both SDS-PAGE analysis and Western blotting (Figures 3B and S62). We also performed similar tests for seven Pa-conjugated FLAG-Ubl probes and observed their covalent binding to a number of ULPs as shown in Figure 3C. Some ULPs such as SENP1 have only been vaguely confirmed in previous work to deconjugate corresponding Ubls such as SUMO4[29].

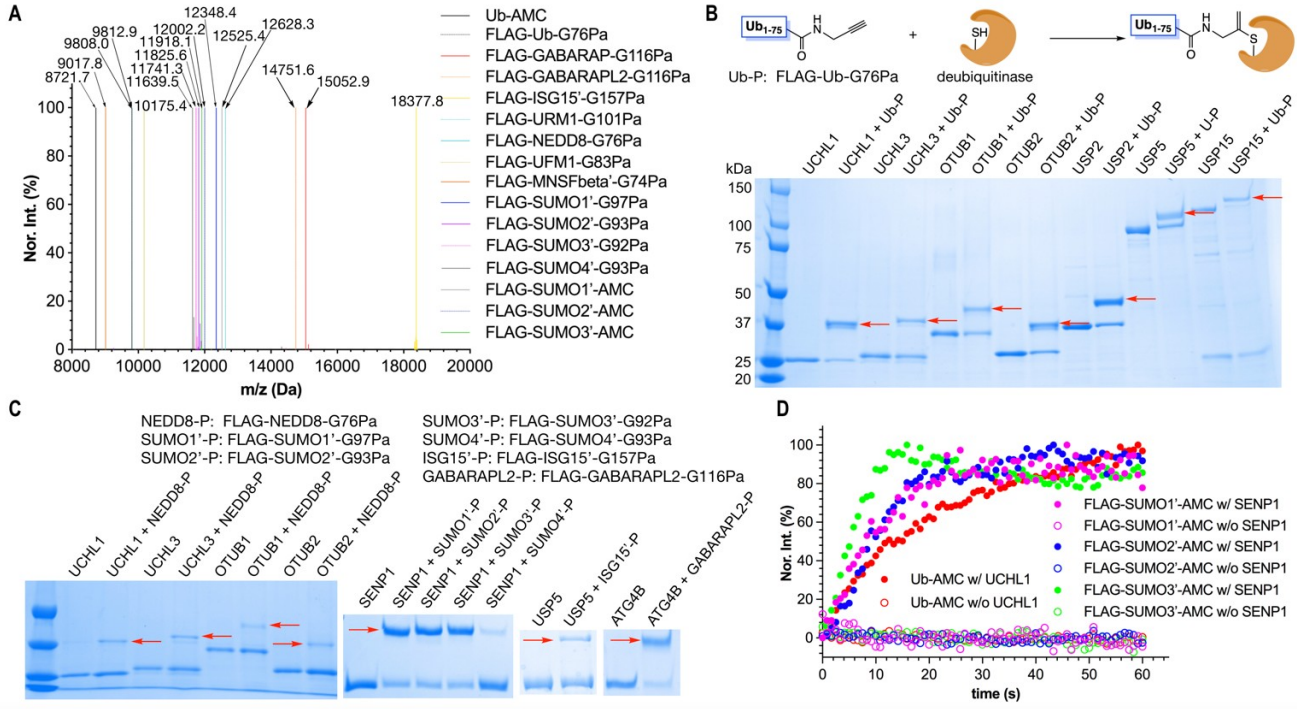

Figure 3. The synthesis of FLAG-Ub/Ubl-Pa and Ub/FLAG-SUMO1-3-AMC probes and their applications in covalent conjugation or activity assays of DUB/ULPs. (**A**) The deconvoluted and integrated ESI-MS of FLAG-Ub/Ubl-Pa and Ub/FLAG-SUMO1-3-AMC probes. Ub-AMC was synthesized from Ub-G76C-6H. All other Pa- and AMC-conjugated Ub/Ubls were generated from FLAG-tagged proteins. Ub/Ubls with their C-terminal glycine mutated to cysteine were expressed and purified as a protein fused with a N-terminal FLAG tag and a C-terminal 6×His tag. ISG15, SUMO1-4, and MNSF have a native cysteine residue. This cysteine was mutated to alanine or serine in all six expressed proteins for avoiding side reactions. The label "" indicates this mutation. All detected molecular weights agreed well with their theoretic values with a deviation range of 0.5 Da. (**B**) The formation of covalent adducts between FLAG-Ub-G76Pa and a number of DUBs. Red arrows point to the generated adducts. (**C**) The formation of covalent adducts, indicated by red arrows, between different FLAG-Ubl-GxPa probes and DUB/ULPs. (**D**) The DUB/ULP-catalyzed AMC release from Ub-AMC and three FLAG-SUMO-AMC conjugates.

All synthesized FLAG-Ub/Ubl conjugates, of which six are synthesized for the first time, are activity-based probes that can be potentially used to profile DUB and ULP proteomes in different tissues or cells. As a demonstration, we incubated the HEK293T cell lysate with FLAG-Ub-G76Pa and then probed the FLAG-Ub-conjugated proteins by an anti-FLAG antibody in Western blotting. The result showed the formation of a number of higher molecular weight species compared to the original FLAG-Ub-G76Pa, indicating conjugation with other proteins in the cell lysate (Figure S63). However, a control reaction using FLAG-Ub-G76C-6H showed no covalent conjugation with any other proteins in the HEK293T cell lysate.

Ub and Ubls conjugated directly to AMC at their *C*-terminus are useful fluorogenic substrates of DUBs and ULPs[30]. To demonstrate the synthesis of Ub/Ubl-AMC conjugates using our ACPL technique, we made Ub-AMC and FLAG-SUMO1-3-AMC by reacting

recombinantly produced Ub-G76C-6H and FLAG-SUMO1-3-GxC-6H proteins with Gly-AMC in the presence of NTCB. The ESI-MS analysis of all four products confirmed their successful formation (Figures 3A and S64-S67) and the following activity assays showed that they served as active substrates for cysteine proteases UCHL1 and SENP1, respectively (Figure 3D). Overall, our combined data of Ub/Ubl probe synthesis establish the broad application scope of the ACPL technique and this technique can make Ub/Ubl probes readily available in a manner that can be performed in almost any biology lab for advancing Ub and Ubl biology studies.

To further demonstrate its broad application scope, we also carried out ACPL to make non-ubiquitin or non-ubiquitin-like proteins. In human cells, Histone H2A can undergo posttranslational acetylation at its terminal lysine, K129[31]. The functional investigation of this acetylation such as how it influences the structure and dynamics of the nucleosome will require the synthesis of the corresponding acetyl-histone, H2AK129ac. We chose to synthesize H2AK129ac to demonstrate that our method can be applied to the synthesis of histones with C-terminal modifications. We first recombinantly produced H2A-K129C-6H, an H2A protein with a K129C mutation and a *C*-terminal 6×His tag and then ligated it to N^{ε}-acetyl-lysine with the assistance of NTCB. The ESI-MS spectrum of the reaction product showed the formation of H2AK129ac (Figures 4A and S68-S69; Table S5) with at least 60% yield. We folded successfully H2AK129ac into a dimer with H2B and subsequently into a nucleosome (Figure 4B) making it possible to study effects of H2AK129ac on the nucleosome structure and functions.

For all ligation reactions that we performed thus far, they involved small molecules with only one amino group for avoiding side product formation. For ligation with larger molecules or peptides that have more than one amino group, one can couple ACPL with peptide hydrazide ligation to resolve non-specificity issues. To demonstrate this prospect, we recombinantly produced a B. halodurans RNase H region with a *C*-terminal Cys-6×His tag (RNH59-196-K190C-6H). Its ligation with hydrazine in the presence of NTCB led to the synthesis of RNH59-189-Ha, a protein hydrazide that we proceeded further to undergo peptide hydrazide ligation with a 7-mer peptide, NH2-CADYGRK-OH to afford a ligated product RNH59-196-K190C[32]. ESI-MS analysis showed the successful synthesis of both RNH59-189-Ha and RNH59-196-K190C (Figures 4C and S70-S72; Table S5). Similar to what has been found in previous peptide hydrazide ligation reactions, we also detected a minor hydrolysis product at 15074.6 Da[33]. The ligated product RNH59-196-K190C was catalytically active to hydrolyze an RNA

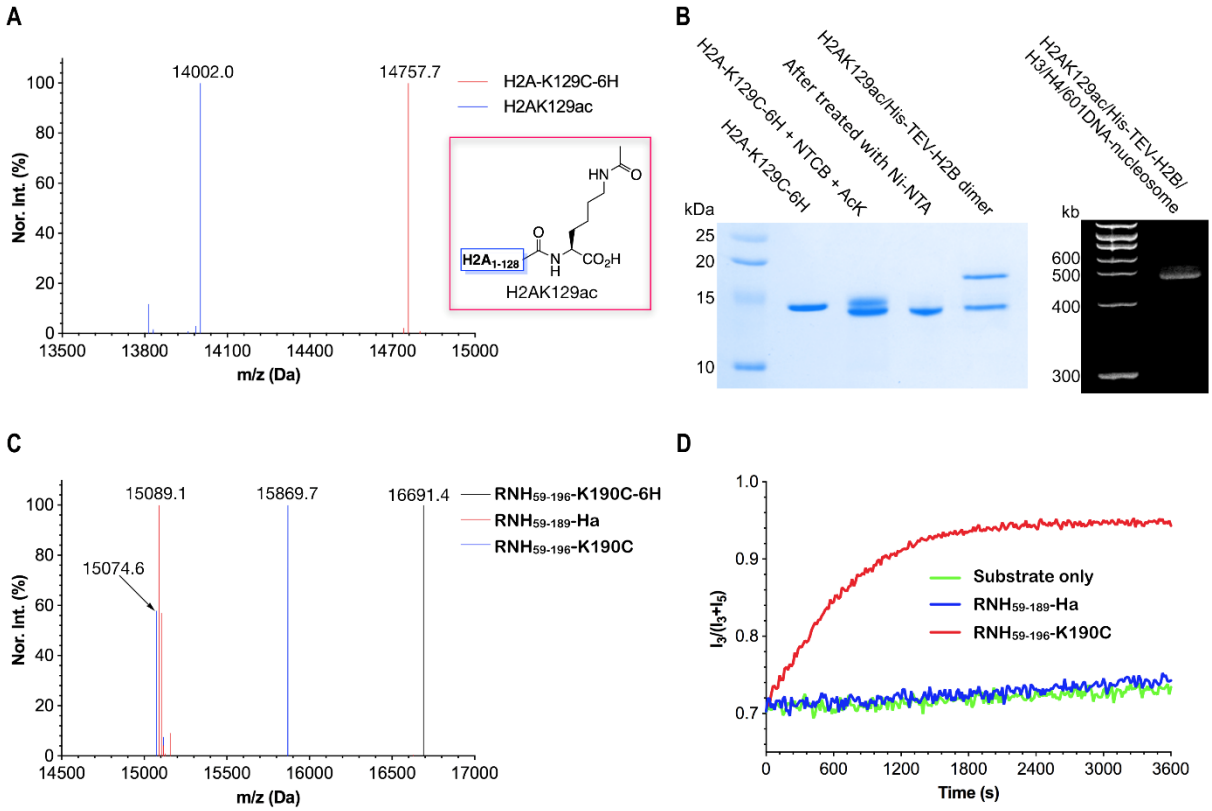

Figure 4. The synthesis of H2AK129ac and RNase H by activated cysteine-directed protein ligation. (**A**) The deconvoluted and integrated ESI-MS of H2A-K129C-6H and H2AK129ac. H2A-K129C-6H was recombinantly expressed and then reacted with NTCB and N^{ε}-acetyl-lysine to afford H2AK129ac. (**B**) The synthesis of H2AK129ac, its isolation, and folding into an H2AK129ac/H2B dimer and then a nucleosome. The purification of H2AK129ac was achieved by extracting the unreacted intermediate using Ni-NTA resins. (**C**) The deconvoluted and integrated ESI-MS of RNH_{59-196}-K190C-6H, RNH_{59-189}-Ha, and RNH_{59-196}-K190C. RNH_{59-196}-K190C-6H was recombinantly expressed in *E. coli*. It was reacted with NTCB and hydrazine to afford RNH_{59-189}-Ha that then underwent peptide hydrazide ligation with a 7-mer NH_2-CADYGRK-OH peptide to form a catalytic active RNH_{59-196}-K190C. (**D**) The catalytic hydrolysis of an RNA substrate by RNH_{59-196}-K190C. The RNA substrate had a sequence 5'-Cy3-GACACCUGAUUC-Cy5-3'. A DNA fragment 5'-GAATCAGGTGTC-3' was used to form a double strand with the RNA substrate for binding to RNH_{59-196}-K190C. The hydrolysis led to improved Cy3 (I_3) and decrease Cy5 (I_5) emission.

substrate as shown in Figure 4D. In the contrary, RNH59-189-Ha was completely inactive toward this substrate. Our data related to the synthesis of RNase H demonstrated that ACPL can couple to peptide hydrazide ligation for ligation with large peptides or even protein fragments.

In summary, we have developed a novel expressed protein ligation technique that uses a cyanylating reagent to directly activate a cysteine in a recombinant protein for ligation with small molecule amines and large peptide or protein fragments when coupling with peptide hydrazide ligation. The technique termed ACPL requires no enzymatic catalysis and is controllable, versatile, specific, and very simple to process. It can be broadly applied to synthesize a large variety of proteins with unique functionalities for advanced applications in both basic and applied research. One potential industrial application of the technique is to synthesize therapeutic peptides. We have shown that the technique can be used to efficiently synthesize exenatide[34], a 39-mer anti-diabetic peptide that has a *C*-terminal amide and is therefore hard to generate using the recombinant expression approach. By expressing a 6×His-SUMO-exenatide-S39C-SA-Strep fusion that can be largely produced in E. coli followed by the treatment with SUMO protease and then processing it by ACPL with L-serinamide, we showed that exenatide can be easily procured (Figures S73-S74). Similar applications to synthesize other therapeutic peptides or proteins are anticipated. ACPL requires the activation of cysteine, one of the two lowest occurring amino acids in proteins. Non-targeted cysteines need to be mutated. For proteins with essential cysteines, one solution for using ACPL is to couple it with the noncanonical amino acid mutagenesis technique. Photocaged cysteines have been genetically incorporated into proteins by amber suppression[35-36]. The incorporation of a photocaged cysteine to essential cysteine sites in a protein followed by ACPL and then decaging to release protected essential cysteines will allow the processing of proteins with non-targeted cysteines. Overall, our ACPL technique expands to a large extent the synthetic capacity of protein chemistry and will energize the whole field. We anticipate its broad applications in a large variety of research fields and industrial processing of proteins and peptides.

ASSOCIATED CONTENT

Corresponding Author

Correspondence and requests for materials should be addressed to W.R.L. (wliu@chem.tamu.edu).

Author Contributions

W.R.L. conceptualized the project. Y.Q. performed the MS experiments and analyzed the data. Y.Q., Y.G, X.A.W. and W.W.W. constructed protein expression vectors, over-expressed proteins, set up ACPL reactions for the conjugates synthesis and performed biological assays. K.K. synthesized the dipeptides, set up ACPL reactions on them and characterized by NMR. W.R.L., Y.Q. and S.Z.L. developed Python programs for the theoretical monoisotopic mass calculation and average mass integration. W.R.L., Y.Q. and J. S. M. wrote the manuscript.

Notes

The authors declare no competing financial interests. All data supporting the findings in this article are available in the main text or the supplementary information. The two Python scripts are available upon request.

ACKNOWLEDGMENT

We thank Dr. Yohannes H. Rezenom who is the director of Texas A&M University Chemistry Mass Spectrometry Facility for helping us on the mass spectrometry analysis. This work was supported by National Institutes of Health (grants R01GM127575 and R01GM121584) and Welch Foundation (A-1715).

REFERENCES

1. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-779.
2. Muir, T. W.; Sondhi, D.; Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (12), 6705-6710.
3. Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. B.; Lin, Y.; Cui, H. K.; Liu, L., Protein chemical synthesis by ligation of peptide hydrazides. *Angew. Chem. Int. Ed. Engl.* **2011,** *50* (33), 7645-7649.
4. Flood, D. T.; Hintzen, J. C.; Bird, M. J.; Cistrone, P. A.; Chen, J. S.; Dawson, P. E., Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for use in Native Chemical Ligation. *Angew. Chem. Int. Ed. Engl.* **2018,** *130* (36), 11808-11813.
5. Muralidharan, V.; Muir, T. W., Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Method.* **2006,** *3* (6), 429.
6. Shogren-Knaak, M.; Ishii, H.; Sun, J. M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L., Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **2006,** *311* (5762), 844-7.
7. McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W., Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008,** *453* (7196), 812.
8. Chu, N.; Salguero, A. L.; Liu, A. Z.; Chen, Z.; Dempsey, D. R.; Ficarro, S. B.; Alexander, W. M.; Marto, J. A.; Li, Y.; Amzel, L. M., Akt kinase activation mechanisms revealed using protein semisynthesis. *Cell* **2018,** *174* (4), 897-907. e14.
9. Schwarzer, D.; Cole, P. A., Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr. Opin. Chem. Biol.* **2005,** *9* (6), 561-569.
10. Pickin, K. A.; Chaudhury, S.; Dancy, B. C.; Gray, J. J.; Cole, P. A., Analysis of protein kinase autophosphorylation using expressed protein ligation and computational modeling. *J. Am. Chem. Soc.* **2008,** *130* (17), 5667-5669.
11. Dann, G. P.; Liszczak, G. P.; Bagert, J. D.; Müller, M. M.; Nguyen, U. T.; Wojcik, F.; Brown, Z. Z.; Bos, J.; Panchenko, T.; Pihl, R., ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature* **2017,** *548* (7669), 607.
12. Rak, A.; Pylypenko, O.; Durek, T.; Watzke, A.; Kushnir, S.; Brunsveld, L.; Waldmann, H.; Goody, R. S.; Alexandrov, K., Structure of Rab GDP-dissociation inhibitor in complex with prenylated YPT1 GTPase. *Science* **2003,** *302* (5645), 646-650.

13. Stevens, A. J.; Sekar, G.; Shah, N. H.; Mostafavi, A. Z.; Cowburn, D.; Muir, T. W., A promiscuous split intein with expanded protein engineering applications. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114* (32), 8538-8543.
14. Amitai, G.; Callahan, B. P.; Stanger, M. J.; Belfort, G.; Belfort, M., Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (27), 11005-11010.
15. Oeemig, J. S.; Zhou, D.; Kajander, T.; Wlodawer, A.; Iwaï, H., NMR and crystal structures of the Pyrococcus horikoshii RadA intein guide a strategy for engineering a highly efficient and promiscuous intein. *J. Mol. Biol.* **2012,** *421* (1), 85-99.
16. Evans, T. C.; Benner, J.; Xu, M.-Q., The in vitro ligation of bacterially expressed proteins using an intein from Methanobacterium thermoautotrophicum. *J. Biol. Chem.* **1999,** *274* (7), 3923-3926.
17. Vila-Perelló, M.; Liu, Z.; Shah, N. H.; Willis, J. A.; Idoyaga, J.; Muir, T. W., Streamlined expressed protein ligation using split inteins. *J. Am. Chem. Soc.* **2012,** *135* (1), 286-292.
18. Henager, S. H.; Chu, N.; Chen, Z.; Bolduc, D.; Dempsey, D. R.; Hwang, Y.; Wells, J.; Cole, P. A., Enzyme-catalyzed expressed protein ligation. *Nat. Method.* **2016,** *13* (11), 925.
19. Rimington, C., The relation between cystine yield and total sulphur in wool. *Biochem. J.* **1929,** *23* (1), 41.
20. Farnworth, A.; Speakman, J., Reactivity of the Sulphur Linkage in Wool. *Nature* **1949,** *163* (4151), 798.
21. Wood, J. L.; Catsimpoolas, N., Cleavage of the peptide bond at the cystine amino group by the action of cyanide. *J. Biol. Chem.* **1963,** *238* (8), PC2887-PC2888.
22. Catsimpoolas, N.; Wood, J. L., Specific cleavage of cystine peptides by cyanide. *J. Biol. Chem.* **1966,** *241* (8), 1790-1796.
23. Komander, D.; Rape, M., The ubiquitin code. *Annu. Rev. Biochem.* **2012,** *81*, 203-229.
24. Ekkebus, R.; van Kasteren, S. I.; Kulathu, Y.; Scholten, A.; Berlin, I.; Geurink, P. P.; de Jong, A.; Goerdayal, S.; Neefjes, J.; Heck, A. J., On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. *J. Am. Chem. Soc.* **2013,** *135* (8), 2867-2870.
25. Pruneda, J. N.; Durkin, C. H.; Geurink, P. P.; Ovaa, H.; Santhanam, B.; Holden, D. W.; Komander, D., The molecular basis for ubiquitin and ubiquitin-like specificities in bacterial effector proteases. *Mol. Cell* **2016,** *63* (2), 261-276.
26. Sommer, S.; Weikart, N. D.; Linne, U.; Mootz, H. D., Covalent inhibition of SUMO and ubiquitin-specific cysteine proteases by an in situ thiol–alkyne addition. *Bioorg. Med. Chem.* **2013,** *21* (9), 2511-2517.
27. Paudel, P.; Zhang, Q.; Leung, C.; Greenberg, H. C.; Guo, Y.; Chern, Y.-H.; Dong, A.; Li, Y.; Vedadi, M.; Zhuang, Z., Crystal structure and activity-based labeling reveal the mechanisms for linkage-specific substrate recognition by deubiquitinase USP9X. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116* (15), 7288-7297.
28. Basters, A.; Knobeloch, K. P.; Fritz, G., How USP 18 deals with ISG 15-modified proteins: structural basis for the specificity of the protease. *FEBS J.* **2018,** *285* (6), 1024-1029.
29. Catic, A.; Fiebiger, E.; Korbel, G. A.; Blom, D.; Galardy, P. J.; Ploegh, H. L., Screen for ISG15-crossreactive deubiquitinases. *PLoS One* **2007,** *2* (7), e679.
30. Dang, L. C.; Melandri, F. D.; Stein, R. L., Kinetic and mechanistic studies on the hydrolysis of ubiquitin C-terminal 7-amido-4-methylcoumarin by deubiquitinating enzymes. *Biochemistry* **1998,** *37* (7), 1868-1879.
31. Basu, A.; Rose, K. L.; Zhang, J.; Beavis, R. C.; Ueberheide, B.; Garcia, B. A.; Chait, B.; Zhao, Y.; Hunt, D. F.; Segal, E., Proteome-wide prediction of acetylation substrates. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (33), 13785-13790.
32. Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W., Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **2005,** *121* (7), 1005-1016.
33. Zheng, J. S.; Tang, S.; Qi, Y. K.; Wang, Z. P.; Liu, L., Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat Protoc* **2013,** *8* (12), 2483-95.
34. Raufman, J.-P., Bioactive peptides from lizard venoms. *Regul. Pept.* **1996,** *61* (1), 1-18.
35. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G., A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **2004,** *126* (44), 14306-14307.
36. Nguyen, D. P.; Mahesh, M.; Elsasser, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W., Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J Am Chem Soc* **2014,** *136* (6), 2240-3.