Organic LETTERS

Letter

pubs.acs.org/OrgLett

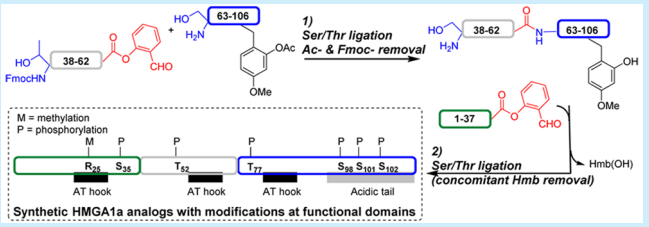
Chemical Synthesis of HMGA1a Proteins with Post-translational Modifications via Ser/Thr Ligation

Tianlu Li, Heng Liu, and Xuechen Li*

Department of Chemistry, State Key Lab of Synthetic Chemistry, The University of Hong Kong, Pokfulam, Hong Kong

Supporting Information

ABSTRACT: The first chemical synthesis of nuclear protein HMGA1a via Ser/Thr ligation is reported. Notably, Hmb (2-hydroxy-4-methoxybenzyl) exhibits crucial improvement of both the difficult coupling during solid phase peptide synthesis and the poor ligation encountered in protein synthesis. These efforts led to preparation of HMGA1a analogs with well-defined phosphorylation and methylation patterns (9 synthetic proteins in total), thus overcoming the heterogeneous and combinatory problems inherent to protein post-translational modifications (PTMs), and facilitating the study of the regulatory roles of such PTMs.



In addition to histone proteins as chromosomal components, the "high mobility group" (HMG) proteins (subfamilies: HMGA, HMGB, and HMGN) play vital roles in modulating chromatin structure, regulating genomic function, and orchestrating participation of other proteins, which are closely related to nuclear activities such as transcription, replication, and DNA repair.[1] HMGA1a is a prototype member of the HMGA protein family, which is recognized as a hub of nuclear functions to affect a plethora of normal biological processes[2,3] including growth, proliferation, differentiation, and death; they are also involved in some common diseases, including benign and malignant tumors,[4] osteoarthritis,[5] diabetes,[6,7] and atherosclerosis.[8]

HMGA1a consists of 106 amino acids with three "AT hooks" as DNA binding domains and an acidic tail at the C-terminus (Figure 1).[9] The AT-hook is a unique protein signature and

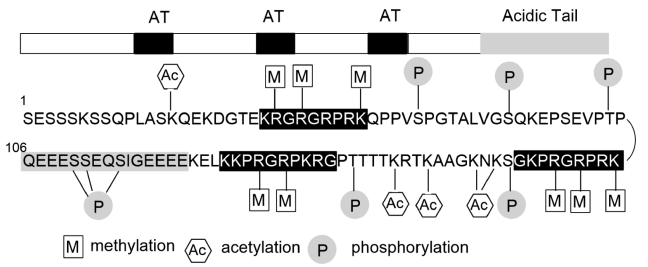


Figure 1. HMGA1a with diverse post-translational modifications.

characteristic functional sequence motif of HMGA family that preferentially binds to the AT-rich stretches in the minor groove of B-form DNA[10] as well as four-way junction DNA.[11] The acidic tail is a Glu-rich, 15 amino acid long sequence at the C-terminus (Glu92-Gln106). The biological function of the acidic tail is not yet fully understood.[12]

The multifunctionality of HMGA1a is achieved through its high adaptability.[13] Apart from its intrinsically disordered status, HMGA1a is subject to diverse post-translational modifications (PTMs) by different kinases, acetyl- and methyltransferases, and other modifying enzymes.[14] Accumulating evidence has indicated the regulatory roles of such PTMs, through which HMGA1a is proposed to function as a "molecular switch". A well-studied example is the acetylation of HMGA1a at distinct Lys residues (Lys64 and Lys70), coordinately regulated by two histone acetyltransferases, CBP and P/CAF. HMGA1a acetylation dynamically controls the transcription activation process of interferon-β gene, giving rise to enhanced or suppressed effects.[15]

Unfortunately, given its various PTMs and the extensive biological processes, the PTM code of HMGA1a is yet to be explored. For example, current studies indicated HMGA1a phosphorylation as a biological event with various kinases involved in a cell-type-specific and cell-cycle-dependent manner;[14] regretfully, the entire cellular network involving HMGA1a phosphorylation remains elusive. Furthermore, these results were mostly obtained either from nuclear extracts or from in vitro kinase treatment of recombinant HMGA1a. Such approaches could hardly prepare HMGA1a in large quantities, which limits further application; moreover, the in vitro enzymatic incorporation of PTMs could unlikely to be site-specific, potentially undermining the studies' credibility.

We therefore decided to apply chemical synthesis to the preparation of homogeneous HMGA1a proteins, expectedly with site-specific modifications installed at a stoichiometric level. Unlike histone protein research,[16] investigation of HMG proteins with chemical approaches remains underexplored. In this study, we have developed useful methods and strategies to

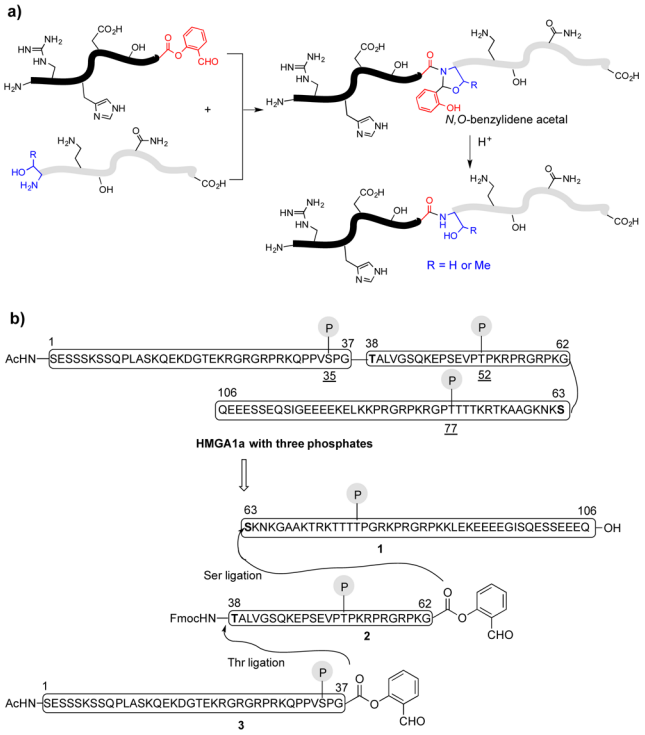
Received: October 11, 2016

DOI: 10.1021/acs.orglett.6b03056
Org. Lett. XXXX, XXX, XXX−XXX

synthesize HMGA1a proteins with well-defined post-translational modification(s) for the first time. Access to such proteins would allow us to gain deeper insight into how PTMs of HMGA1a affect its binding to DNAs and proteins.

Our first goal was to develop the total synthesis of HMGA1a with three phosphates at in vivo cdc2 phosphorylation sites (Ser35, Thr52, Thr77), which are abundant phosphorylated sites most adjacent to the AT hooks. According to the principle of Ser/Thr ligation (Scheme 1a),[17,18] the retrosynthesis of

Scheme 1. (a) Ser/Thr ligation (STL). (b) Retrosynthetic Plan of HMGA1a According to the Principle of STL

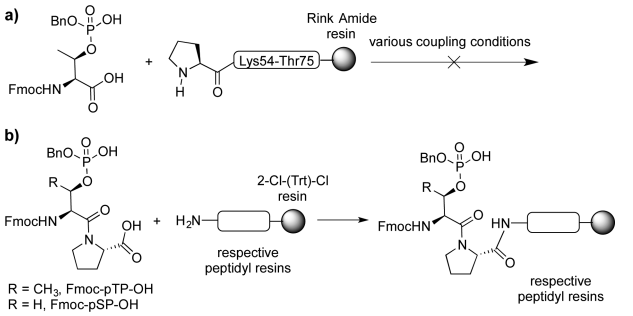


HMGA1a was disconnected as peptide Ser1-Gly37 (**3**), Thr38-Gly62 (**2**), and Ser63-Gln106 (**1**) with lengths of 37, 25, and 44 amino acid residues, respectively, each of which contains a phosphorylated residue (Scheme 1b).

To prepare the phosphorylated peptide segments by Fmoc-SPPS (solid-phase peptide synthesis), the building block Fmoc-Thr(HPO$_3$Bn)-OH was synthesized accordingly.[19] However, its coupling onto resin-bound Pro53-Thr75 peptide turned out to be unexpectedly difficult (Scheme 2a). Various coupling reagents and conditions were carefully attempted, including HATU/DIEA/DMF, HATU/HOAt/collidine/DMF, PyBOP/DIEA/DMF, PyBOP/HOBt/DIEA/DMF, PyBOP/HOBt/DIEA/DMSO/NMP, and DEPBT/DIEA/DMF/DCM; unfortunately, little desired product, if any, was observed. It was then clear to us that the coupling between the phosphorylated Thr52 and the secondary amine of the Pro53 residue was not properly efficient, probably due to the steric effect and low reactivity. It is noted that pSer35 and pThr77 are also connected after the Pro residue; thus, a general strategy to overcome this obstacle had to be devised prior to the total synthesis.

Considering the low proneness to epimerization at the C-terminal Pro residue during the coupling reaction, we decided to prepare phosphorylated dipeptide building blocks Fmoc-

Scheme 2. (a) Fmoc-Thr(HPO$_3$Bn)-OH Failed To Be Incorporated during SPPS. (b) Fmoc-pTP-OH and Fmoc-pSP-OH Dipeptide Building Blocks Were Utilized to Overcome the Difficulty of SPPS



pTP-OH and Fmoc-pSP-OH in solution phase (Scheme 2 and Schemes S1 and S2) followed by elongation onto the resin-bound peptide chain. To our delight, this time, the phosphorylated dipeptide Fmoc-pTP-OH was readily coupled onto the resin-bound Lys54-Thr75 peptide under the HATU/HOAt/DIEA/NMP/DMF conditions (Scheme 2b). Upon solving this problem, the phosphorylated HMGA1a peptide segments **1**−**3** could now be successfully synthesized via Fmoc-SPPS.

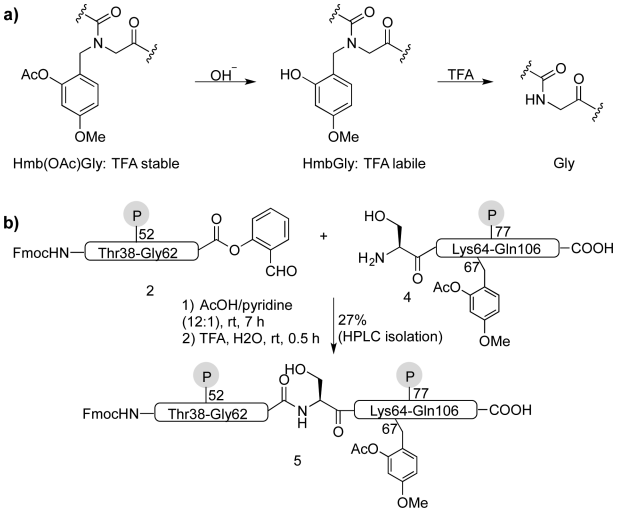
Along our synthetic studies, the second challenge confronting us was the unsatisfactory Ser/Thr ligation between the peptide (Thr38-Gly62) salicylaldehyde (SAL) ester (**2**) and the peptide **1** (Ser63-Gln106). Under various ligation conditions, including different pyridine/AcOH ratios, peptide concentrations, ligation additives, and elevated temperature, the desired ligation product was hardly observed and the peptide SAL ester was eventually hydrolyzed. To understand the failure in ligation, we used model peptides to test the reactivity and properties of peptide SAL ester **2** and peptide **1**, respectively. While the ligation between **2** and SVAFKA or a truncated fragment peptide (Thr76-Gln106) proceeded very smoothly, peptide **1** failed to react with AIFPNAF-SAL ester and the reaction mixture appeared as a slurry. These results indicated that the failed ligation was due to the poor solubility of peptide Ser63-Gln106 (**1**) under the ligation conditions (pyridine/AcOH).

Toward this end, we reckoned that the backbone-protecting group Hmb (2-hydroxy-4-methoxybenzyl) would help solve the poor solubility problem by minimizing peptide aggregation tendency. While Hmb(Ac) is stable to post-SPPS TFA cleavage, Hmb can be readily removed by the TFA cocktail to release the native peptide (Scheme 3a),[20] making it a suitable auxiliary for Ser/Thr ligation. Gratifyingly, Hmb(Ac)Gly67-containing peptide **4** indeed exhibited better solubility and could ligate with peptide SAL ester **2** to afford peptide (Thr38-Gln106) **5** in 27% yield after HPLC purification (Scheme 3b).

However, the preparation of peptide (63−106) **4** via SPPS was very difficult because the incomplete coupling at EKEEEE residues of the Glu-rich acidic C-terminal tail region was very significant and the deletion product was difficult to separate from the desired product, affording <2% isolated yield based on resin loading. On the other hand, another problem was encountered when the obtained peptide **5** underwent Fmoc removal and the subsequent ligation with peptide (Ser1-Gly37) SAL ester **3**. After TFA treatment, a ligation product was found with m/z corresponding to peptide (Ser1-Gln106) minus Hmb

DOI: 10.1021/acs.orglett.6b03056
Org. Lett. XXXX, XXX, XXX−XXX

Scheme 3. (a) Backbone Hmb Protection. (b) Hmb Served To Improve Ligation Efficiency



but plus an Ac group. It is likely that the Fmoc removal condition caused the acyl group of Hmb(Ac) to transfer.[21]

To solve the incomplete coupling problem at the acidic C-terminal tail during SPPS, we changed the position of Hmb(Ac)Gly from Gly67 to Gly96 within the problematic region to disrupt the interchain association during peptide segment **6** synthesis. This change not only was effective for minimizing the incomplete coupling of SPPS, affording peptide **6** in a higher yield (16% based on resin loading), but also improved the ligation between peptide (Thr38-Gly62) SAL ester **2** and peptide Ser63-Gln106 (**6**), affording 38% ligated product **7** after HPLC purification (Figure 2a,b). To avoid the potential acyl-transfer problem of ligation product **7**, we removed the Ac group of Hmb with 10 equiv of N_2H_4 monohydrate[22] followed by one-pot N-terminal Fmoc group removal with Et_2NH[23] in CH_3CN/H_2O. An additional workup with 5% aqueous TFA proved to prevent the potential of hydrazinolysis of the SAL ester in the next ligation. This operation sequence produced the peptide **8** (Thr38-Gln106) cleanly in 60% isolated yield over two steps. Subsequently, the ligation between the resultant peptide **8** and peptide (Ser1-Gly37) SAL ester **3** went very smoothly and surprisingly even faster than the first ligation, with completion within 7 h (Figure 2c). Upon acidolysis to convert the *N,O*-benzylidene acetal intermediate to the native peptidic linkage, the Hmb group was removed concomitantly, and the full length of HMGA1a with three phosphates **9** was obtained in 43% yield after HPLC purification (Figure 2d).

Having established the Ser/Thr ligation strategy to produce synthetic HMGA1a protein **9** with three phosphates, in a similar fashion, we synthesized an unmodified HMGA1a **10** (Scheme S5) as well as three monophosphorylated HMGA1a proteins at Ser35, Thr52, and Thr77, respectively (**11**–**13**) (Scheme S6–S8), which serve to elucidate the effect of the individual phosphorylation on the HMGA1a function. Synthesis of HMGA1a **14** bearing asymmetric dimethylation at Arg25 within the first AT-hook was also achieved (Scheme S9). The generality of this synthetic strategy was further demonstrated by the preparation of biotinylated HMGA1a proteins including an unmodified one, an acidic tail truncated one, and one with triphosphates at the acidic tail (**15**–**17**) (Figure 3a and Schemes S10–S12). We examined how these

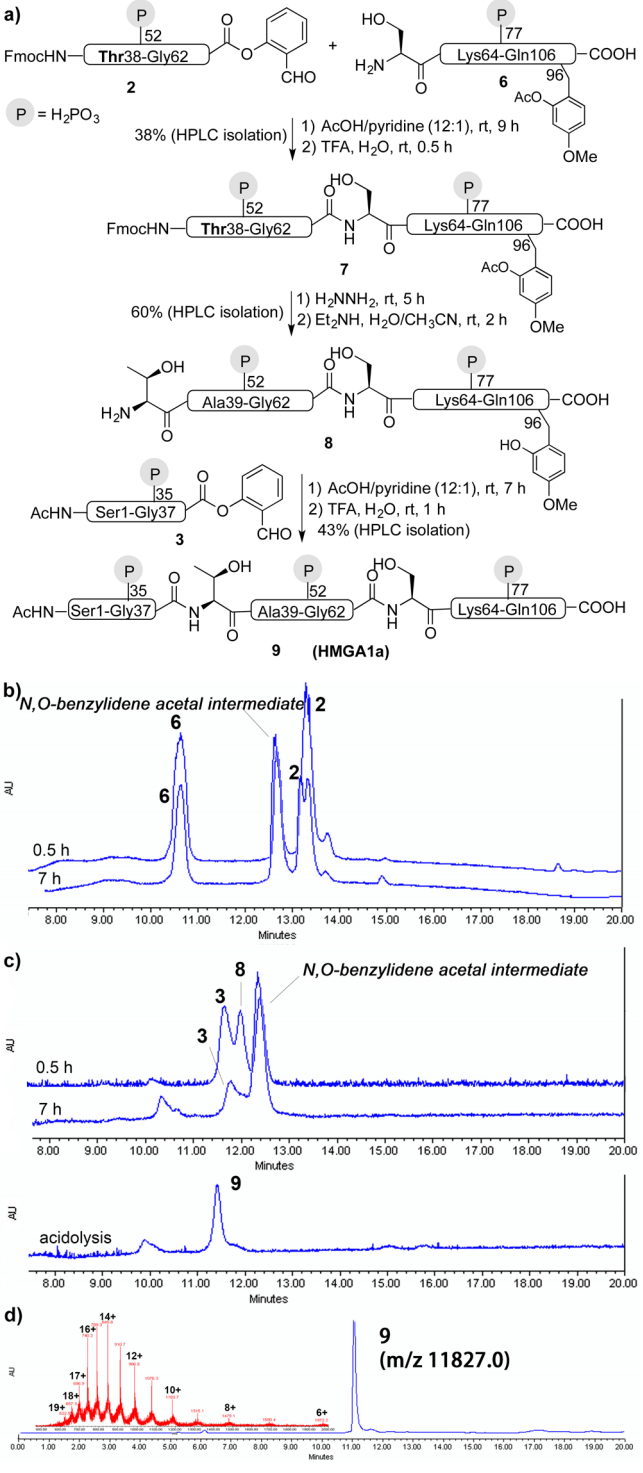


Figure 2. (a) Synthetic scheme of full-length HMGA1a. (b) Representative HPLC profile of the first ligation. (c) Representative HPLC profile of the second ligation. (d) HPLC–MS characterization of the purified final product HMGA1a **9**.

PTMs affect the secondary structure of HMGA1a. As shown in Figure 3b, the incorporation of the PTM(s) did not affect the intrinsically disordered nature and structural behavior of HMGA1a.

In summary, we have developed an attractive and practical route to chemically synthesize HMGA1a proteins for the first time via Hmb-assisted Ser/Thr ligation. During the synthetic

DOI: 10.1021/acs.orglett.6b03056
Org. Lett. XXXX, XXX, XXX–XXX

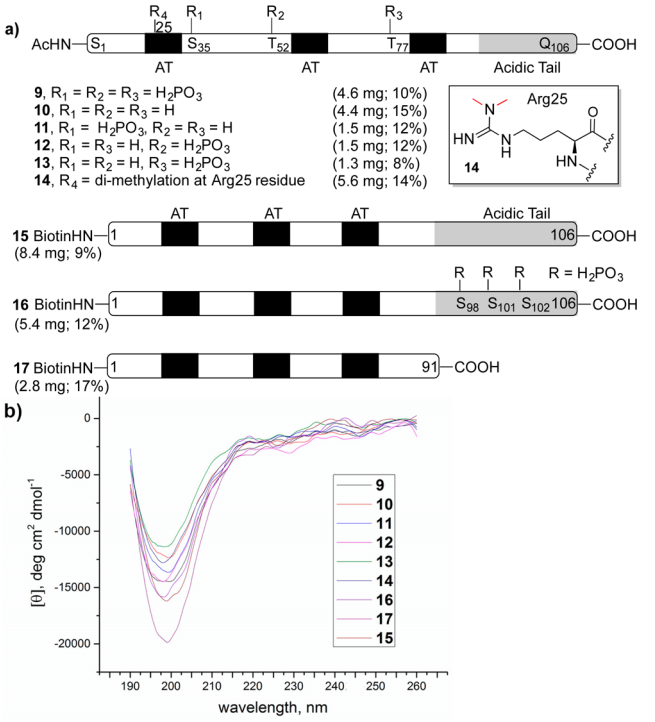


Figure 3. (a) Synthetic HMGA1a proteins (total amount and overall yields after HLPC isolation in parentheses). (b) Circular dichroism spectra of synthetic HMGA1a proteins.

study, we discovered the poor coupling of the phosphorylated Ser/Thr to the proline residue at SPPS, which was solved by using preassembled phosphorylated Ser-Pro and Thr-Pro dipeptide building blocks. In addition, solubility issues represent a practical challenge in protein chemical ligation, which can be tactically overcome.[24] In our study, the utility of Hmb crucially assisted the Ser/Thr ligation to allow the HMGA1a proteins to be successfully synthesized, normally in 14 working days, on multimilligram scales. This Hmb-assisted Ser/Thr ligation may find general applications for synthesis of proteins of such type. Overall, the obtained HMGA1a proteins with different phosphorylation and methylation patterns will provide useful chemical probes to study how post-translational phosphorylation(s) of HMGA1a affect the protein−DNA and protein−protein binding events, which is ongoing and will be reported in due course.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.orglett.6b03056.

Experimental procedures and characterization data (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: xuechenl@hku.hk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Research Grants Council−Collaborative Research Fund of Hong Kong (C7038-15G), the National Basic Research Program of China (2013CB836900), and the CAS interdisciplinary Innovation Team.

■ REFERENCES

(1) Reeves, R. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2010**, *1799*, 3.
(2) Reeves, R. *Gene* **2001**, *277*, 63.
(3) Cleynen, I.; Van De Ven, W. J. M. *Int. J. Oncol.* **2008**, *32*, 289.
(4) Fusco, A.; Fedele, M. *Nat. Rev. Cancer* **2007**, *7*, 899.
(5) Gasparini, G.; De Gori, M.; Paonessa, F.; Chiefari, E.; Brunetti, A.; Galasso, O. *Arthritis Res. Ther.* **2012**, *14*, R207.
(6) Foti, D.; Chiefari, E.; Fedele, M.; Iuliano, R.; Brunetti, L.; Paonessa, F.; Manfioletti, G.; Barbetti, F.; Brunetti, A.; Croce, C. M.; Fusco, A.; Brunetti, A. *Nat. Med.* **2005**, *11*, 765.
(7) Chiefari, E.; Iiritano, S.; Paonessa, F.; Le Pera, I.; Arcidiacono, B.; Filocamo, M.; Foti, D.; Liebhaber, S. A.; Brunetti, A. *Nat. Commun.* **2010**, *1*, 40.
(8) Schlueter, C.; Hauke, S.; Loeschke, S.; Hans Wenk, H.; Bullerdiek, J. *Pathol., Res. Pract.* **2005**, *201*, 101.
(9) Zhang, Q.; Wang, Y. *Biochim. Biophys. Acta, Proteins Proteomics* **2008**, *1784*, 1159.
(10) Reeves, R.; Nissen, M. S. *J. Biol. Chem.* **1990**, *265*, 8573.
(11) Hill, D. A.; Reeves, R. *Nucleic Acids Res.* **1997**, *25*, 3523.
(12) Maurizio, E.; Cravello, L.; Brady, L.; Spolaore, B.; Arnoldo, L.; Giancotti, V.; Manfioletti, G.; Sgarra, R. *J. Proteome. Res.* **2011**, *10*, 3283.
(13) Sgarra, R.; Zammitti, S.; Lo Sardo, A.; Maurizio, E.; Arnoldo, L.; Pegoraro, S.; Giancotti, V.; Manfioletti, G. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2010**, *1799*, 37.
(14) Zhang, Q.; Wang, Y. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2010**, *1799*, 28.
(15) Munshi, N.; Agalioti, T.; Lomvardas, S.; Merika, M.; Chen, G.; Thanos, D. *Science* **2001**, *293*, 1133.
(16) Müller, M. M.; Muir, T. W. *Chem. Rev.* **2015**, *115*, 2296.
(17) Li, X.; Lam, H. Y.; Zhang, Y.; Chan, C. K. *Org. Lett.* **2010**, *12*, 1724.
(18) (a) Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6657. (b) Lee, C. L.; Liu, H.; Wong, C. T. T.; Chow, H. Y.; Li, X. *J. Am. Chem. Soc.* **2016**, *138*, 10477.
(19) Vorherr, T.; Bannwarth, W. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2661.
(20) Johnson, T.; Quibell, M.; Sheppard, R. C. *J. Pept. Sci.* **1995**, *1*, 11.
(21) Acyl groups on tyrosine phenolic OH have been observed to transfer to an adjacent amine via 12−14-membered rings; see: Panda, S. S.; Hall, C. D.; Oliferenko, A. A.; Katritzky, A. R. *Acc. Chem. Res.* **2014**, *47*, 1076.
(22) Johnson, T.; Packman, L. C.; Hyde, C. B.; Owen, D.; Quibell, M. *J. Chem. Soc., Perkin Trans. 1* **1996**, 719.
(23) Ogunkoya, A. O.; Pattabiraman, V. R.; Bode, J. W. *Angew. Chem., Int. Ed.* **2012**, *51*, 9693.
(24) For examples, see: (a) Kochendoerfer, G. G.; Salom, D.; Lear, J. D.; Wilk-Orescan, R.; Kent, S. B. H.; DeGrado, W. F. *Biochemistry* **1999**, *38*, 11905. (b) Clayton, D.; Shapovalov, G.; Maurer, J. A.; Dougherty, D. A.; Lester, H. A.; Kochendoerfer, G. G. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 4764. (c) Valiyaveetil, F. I.; MacKinnon, R.; Muir, T. W. *J. Am. Chem. Soc.* **2002**, *124*, 9113. (d) Liu, L. P.; Deber, C. M. *Biopolymers* **1998**, *47*, 41. (e) Johnson, E. C. B.; Kent, S. B. H. *Tetrahedron Lett.* **2007**, *48*, 1795. (f) Zheng, J.-S.; Yu, M.; Qi, Y.-K.; Tang, S.; Shen, F.; Wang, Z.-P.; Xiao, L.; Zhang, L.; Tian, C.-L.; Liu, L. *J. Am. Chem. Soc.* **2014**, *136*, 3695. (g) Jacobsen, M. T.; Petersen, M. E.; Ye, X.; Galibert, M.; Lorimer, G. H.; Aucagne, V.; Kay, M. S. *J. Am. Chem. Soc.* **2016**, *138*, 11775.

DOI: 10.1021/acs.orglett.6b03056
Org. Lett. XXXX, XXX, XXX−XXX