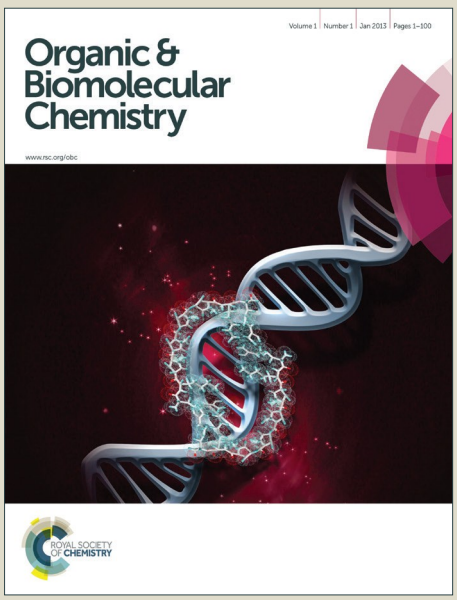
# Organic & Biomolecular Chemistry

Accepted Manuscript





ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Organic & Biomolecular Chemistry

COMMUNICATION

# Hybrid Phase Ligation for Efficient Synthesis of Histone Proteins

Ruixuan R. Yu,[ab]† Santosh K. Mahto,[a]† Kurt Justus,[ab] Mallory M. Alexander,[a] Cecil J. Howard[ab] and Jennifer J. Ottesen[abc]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

**We introduce a hybrid solid-solution phase ligation approach that combines the efficiency of solid phase ligation with solution phase ligation in the total synthesis of modified histone proteins. A two linker strategy allows analysis throughout work on the solid phase and maximizes yields through cleavage at an external Rink, while an internal HMBA linker allows the native carboxyl terminus for any protein sequence. We demonstrate this approach for two histone proteins: triple-acetylated H4-K5ac,K12ac,K91ac and CENP-A-K124ac.**

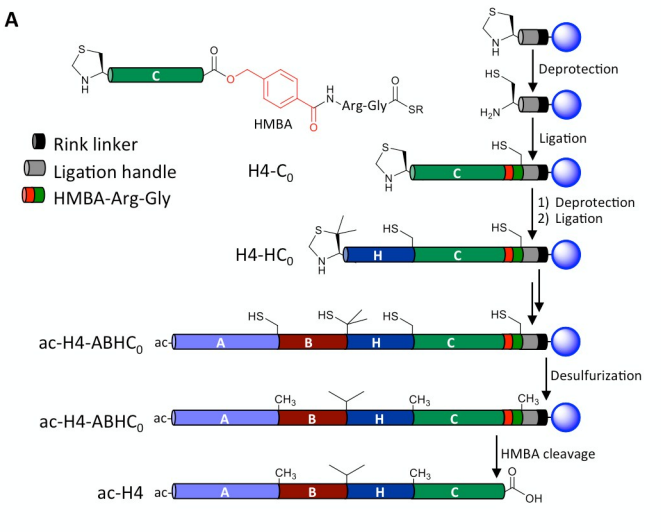
Post-translational modification (PTM) of histone proteins is a chemical foundation for dynamic regulation of chromatin DNA in the cell.[1-6] Histone PTMs in the core of the nucleosome can alter the inherent structure[7] and dynamics of the nucleosome, while histone modifications on the surface or in the unstructured histone tails can dynamically regulate biological functions such as transcription and DNA repair.[8, 9] Acetylation of histone H4 has been hypothesized to regulate chromatin assembly,[10, 11] while specific simultaneous acetylation of the centromeric H3 variant CENP-A (CpA) together with H4 are correlated with structural alterations of the canonical "octasome" containing the standard 8 histones in a nucleosome to a smaller "hemisome" containing only one copy of each histone, thought to be required for mitosis.[12-14] Accessible chemical methods to prepare extensively modified histones in high efficiency and purity are therefore of paramount importance.[15-18]

Several strategies have been explored for total synthesis of histone proteins, including sequential ligation,[19] convergent ligation,[20] one-pot ligation using peptide hydrazides,[21] and solid phase native chemical ligation.[22] Each approach has both advantages and disadvantages. Sequential approaches require intermediary purification steps, which drastically decrease overall yields.[23, 24] In order to minimize the number of segments, one-pot approaches require long peptide sequences resulting in challenging syntheses.[21] Solid phase ligation (SP-NCL) has been reported to improve efficiency and yield for histone H2B,[22] but we show in this work that for H4 and CpA, overall yields with SP-NCL are unacceptably low despite efficient chemistry at each step. Here, we demonstrate that carrying out initial ligation steps on the solid phase followed by final ligation steps in solution provides optimal efficiency and yield for heavily modified synthetic histone proteins, contrary to previous reports that suggest fully solid phase ligation is inherently superior to mixed phase in other systems.[25]

We first attempted to carry out SP-NCL for two target histone proteins: the well-characterized target histone H4, which is highly conserved among all known eukaryotic species,[26] and the centromere-specific CpA, which to the best of our knowledge has never been prepared synthetically. Of note, recombinant CpA is challenging to produce,[27] and prior to this work we have been unsuccessful in generating a CpA-intein fusion for expressed protein ligation due to these challenges. In this context, total synthesis of CpA is essential for the study of biologically important modifications such as acetylation at CpA-K124ac.[13]

A key decision point in SP-NCL is the selection of an appropriate linker to release the final product from resin, and several linkers have been explored.[25, 28, 29] For histone, Brik et al recently used the TFA-labile Rink linker to prepare H2B, which results in the conservative substitution of an amide at the C-terminus.[22] However, for proteins such as H4 and CpA, for which the carboxyl termini are each folded into the core of the nucleosome and are positioned to participate in salt bridge interactions,[30] any chemical modification at this site is not appropriate. We therefore developed a dual-linker strategy in which the C-terminal residue is followed by an internal ligation-stable HMBA linker[31] and a ligation handle, which allows for sequence-independent resin attachment and generates the native carboxyl terminus in a simple pH shift in aqueous mixtures. However, we also incorporate a secondary Rink linker in the resin-bound segment, which provides an analytical tool throughout the ligation process,[32, 33] and an avenue for optimal yields in cleavage of the solid phase component (**Fig. 1A**). Throughout this work, we will use the terminology $X_0$ to denote a peptide cleaved at the Rink linker and containing the ligation handle, while X without the suffix will refer to the peptide cleaved at the HMBA linker with a carboxyl terminus. While there are several potential base resins for SP-NCL, PEGA is preferred by many researchers for its compatibility with both organic and aqueous conditions and large pore size;[22, 29]

[a] *Department of Chemistry & Biochemistry*
[b] *Ohio State Biochemistry Program, The Ohio State University, Columbus OH 43210*
[c] *To whom correspondence should be addressed: J.J Ottesen (phone: 614 242 4525; email: ottesen.1@osu.edu)*
† These authors contributed equally to this work
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x




B

| Peptide | residues | Synthesized peptide sequence |
|---|---|---|
| H4-A | 1-37 | SGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRL |
| H4-B | 38-56 | Thz-RRGGVKRISGLIYEETRG |
| H4-H | 57-75 | dmThz-LKVFLENVIRDAVTYTEH |
| H4-C$_0$ | 76-102 | Thz-KRKTVTA-Nle-DVVYALKRQGRTLYGFGG-**HMBA**-RG |

| Peptide | residues | Synthesized peptide sequence |
|---|---|---|
| CpA-1 | 2-34 | GPRRRSRKPEAPRRRSPSPTPTPGPSRRGPSLG |
| CpA-2 | 35-70 | Thz-SSHQHSRRRQGWLKEIRKLQKSTHLLIRKLPFSRL |
| CpA-3 | 71-97 | Thz-REISVKFTRGVDFNWQAQALLALQEA |
| CpA-4 | 98-115 | Thz-EAFLVHLFEDAYLLTLH |
| CpA-5 | 116-140 | Thz-GRVTLFPKDVQLARRIRGLEEGLG-**HMBA**-RG |

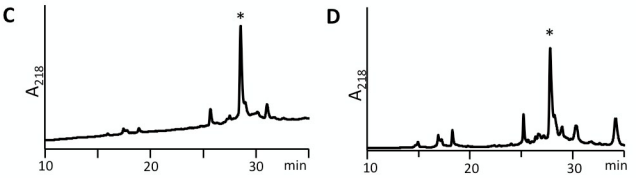



**Fig. 1** (A) Scheme for solid-phase NCL of histone H4. A single coupling cycle includes deprotection followed by ligation. Cleavage from the resin may be accomplished at the Rink linker (black), or at the HMBA linker (red) to generate the native terminus. (B) Peptide segments used for ligation schemes for H4 and CpA. (C) RP-HPLC (22-63 % acetonitrile/0.1% TFA) of crude cleavage for SP-NCL of ac-H4-ABHC$_0$. * denotes product: $[M + H]^+$ observed: *m/z* 11853, expected: *m/z* 11848. (D) RP-HPLC (22-63 %acetonitrile/0.1% TFA) of crude cleavage for desulfurized ac-H4. * denotes product: $[M + H]^+$ observed: *m/z* 11259, expected: *m/z* 11259.

in exploratory studies we found the same. Interestingly, we found that reducing resin loading to 10% by coupling a mixture of 9:1 Boc-Gly-OH:Fmoc-Gly-OH[34] significantly improved kinetic reaction rates when assembling longer protein chains, presumably through reduced steric crowding. The ligation sites were selected through an iterative design procedure assessing for peptide solubility and ligation kinetics, using synthetic accessibility as a guiding principle; thus, we limited ourselves to Ala and Val at the ligation sites, since the protected derivatives of the ligation surrogates Cys and penicillamine (Pen)[35] are commercially available. Met is substituted by the isostere norleucine (Nle) to eliminate oxidation, and we used a substituted CpAC75S sequence that has been demonstrated to localize to centromeres[36] but is compatible with desulfurization. Together these lead to the peptides described in **Fig. 1B**, adjusted for the appropriate post-translational modifications in each synthesis.

Peptides were synthesized on Dbz(Alloc) resin[37] in preparation for N-acylurea mediated thioester formation with the following special considerations. The 37-residue H4-A peptide and the 36-residue peptide CpA-2 peptide could each be synthesized cleanly on Dbz(Alloc) resin, but efficient conversion to the reactive species required addition of 4-Nitrophenyl chloroformate in dry DMF to generate a formyl-Nbz derivative,[38] which could be carried forward for thioester exchange and ligation (SI Fig. S8). In addition, two peptides had histidine termini; we noted unacceptable levels of histidine racemization for H4-H and CpA-4 syntheses (SI 3.4). DIC-mediated coupling onto unprotected Dbz resin[39] provided the best balance with no detectable racemization and over 85% coupling yields. Of note, Blanco-Canosa et al recently introduced a modified N-methyl-Dbz linker that is reported to reduce these problems, presenting an intriguing opportunity in future syntheses.[40]

With all peptides in hand, we carried out complete SP-NCL of N-acetylated H4 (**Fig. 1**) and of unmodified CpA. For each cycle, ligation was assessed by trifluoroacetic acid (TFA) cleavage at the analytical Rink linker followed by RP-HPLC (see Supporting Information). After 4 cycles of on-resin ligation, H4-ABHC$_0$ was prepared in >90 % purity (**Fig. 1C**). We then carried out on-resin desulfurization of the four thiols remaining at the ligation sites (3 Cys, 1 Pen), followed by base-mediated cleavage of the HMBA linker to release the product protein (**Fig. 1D**). Unfortunately, ac-H4 was generated in extremely low yield, providing only sufficient protein for analysis. Similarly, extremely low yields were observed for full-length CpA prepared via fully SP-NCL; in fact, insufficient CpA was released from the resin to analyse by RP-HPLC.

We noted qualitatively that the recovered yield dropped precipitously after addition of the H4-A peptide for histone H4, and the CpA2 peptide for histone CpA. To assess sequence dependence, we ligated CpA-1 and CpA-2 directly to the base resin, cleaved at the Rink linker, and observed 30-56% recovered yield (ESI 6.2 and 6.3). These yields are reduced but do not replicate the extreme losses observed for the equivalent ligation in the protein context. Given that PEGA resin has a large pore size compatible with folded proteins, it seems unlikely that yield reductions are solely size-based. Together, these suggest context-dependent interactions of the larger histone sequences and the solid phase. We therefore redesigned our synthesis with a hybrid phase approach, in which the first three ligation steps for each protein are carried out on resin. Release of the reactive Cys product from the resin was then followed by the final ligation step(s) carried out in solution.

We successfully tested this approach with triple-modified H4-K5ac,K12ac,K91ac **(Fig. 2)** and with CpA-K124ac **(Fig. 3)** using the same peptide schemes described in **Fig. 1B**. For histone H4, after three ligation steps, H4-BHC-K91ac was released from the resin by cleavage of the HMBA linker with 100 mM NaOH for 15 minutes, followed by neutralization with HCl. Interestingly, we found a significant increase in yield when the resin was further extracted



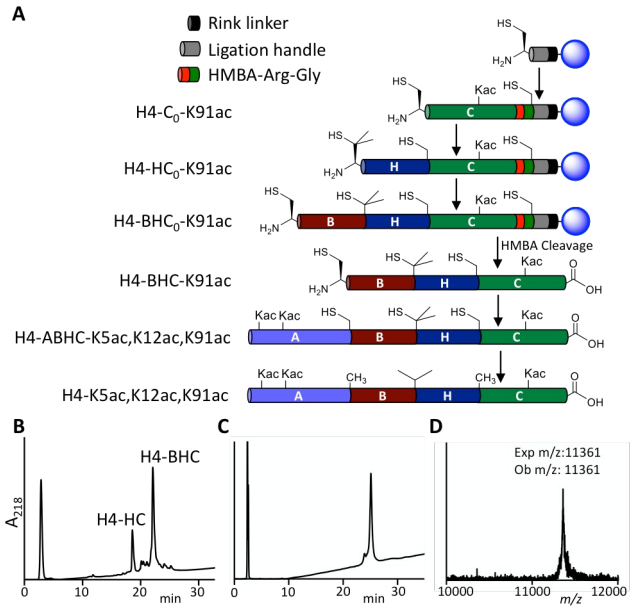



**Fig. 2** Hybrid-NCL of H4-K5ac, K12ac, K91ac (A) Scheme for Hybrid-NCL of histone H4, used to prepare H4-K5ac,K12ac,K91ac. (B) RP-HPLC (27-62% acetonitrile/0.1% TFA) of crude cleavage; peaks identified by MALDI-TOF MS: H4-BHC $[M + H]^+$ observed: *m/z* 7452, expected: *m/z* 7454. HC is present due to incomplete ring-opening of HC. (C) RP-HPLC (27-62% acetonitrile/0.1% TFA and (D) MALDI-TOF MS analysis of purified H4-K5ac,K12ac,K91ac. $[M + H]^+$ observed: *m/z* 11361, expected: *m/z* 11361.

with TFA, which increases solubility of the protein and swelling of the resin beads. We believe this is extraction rather than additional cleavage of Rink linker, since H4-$BHC_0$-K91ac was not observed in significant amounts. With the cleavage, neutralization, and extraction combined, we were pleased to achieve 97% crude yield as assessed by lyophilized weight. Of note, a significant ~20% side product was observed that resulted from incomplete deprotection of the sterically hindered dimethyl thiazolidine to pencillamine at the $HC_0$ stage (**Fig. 2B**). This can be avoided by carrying out micro-cleavage of 5-10 resin beads followed by mass spectrometry analysis to assess reaction progress with minimal product loss, which we instituted as a standard protocol after this synthesis.

The solution phase ligation to generate H4-ABHC proceeded to ~90% as assessed by SDS-PAGE. The mixture was dialyzed extensively against thiol-free ligation buffer to exchange the aryl thiol MPAA, which is not compatible with desulfurization by the Danishefsky approach.[41] Desulfurization was then carried out prior to RP-HPLC purification to obtain the final product. After four rounds of ligation and desulfurization (9 chemical steps), the isolated H4-K5ac,K12ac,K91ac product was obtained in 16% yield, which is approximately commensurate with yields observed for total synthesis of H4 by other approaches.[21]

For CpA-K124ac, we again carried out the first three ligations on the solid phase (**Fig. 3**). Drawing on the lessons learned from H4, all reactions were followed by micro-cleavage, and CpA-$345_0$-K124ac was released from resin by cleavage of the Rink linker. CpA-$345_0$-K124ac was recovered in 104% yield by lyophilized weight, and in excellent (>90 %) purity as assessed by RP-HPLC (**Fig. 3B**).

CpA-2-Nbz was added and efficiently ligated in solution to generate

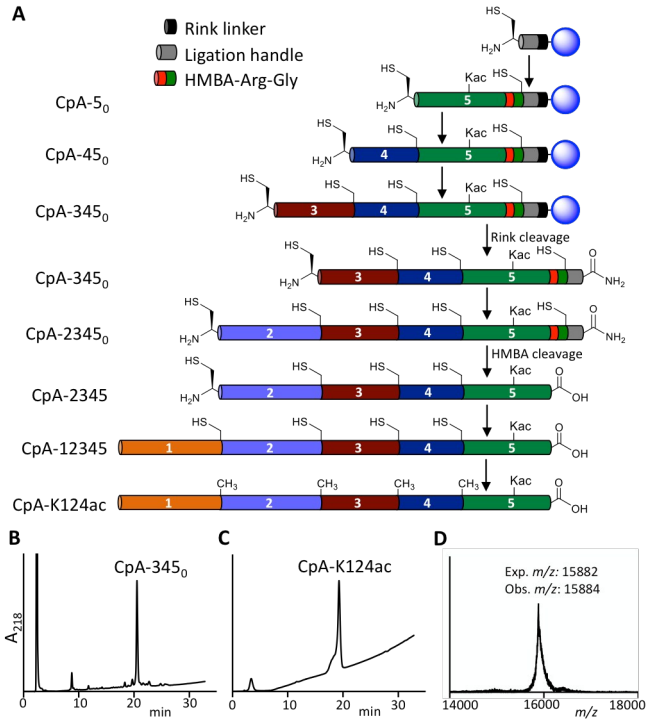



**Fig. 3** (A) Hybrid-NCL scheme for preparation of CpA-K124ac. CpA-$345_0$ is cleaved through the Rink linker, and HMBA is cleaved prior to the addition of CpA-1. (B) RP-HPLC (27-81% acetonitrile/0.1% TFA) analysis of the SP-NCL component demonstrating preparation of CpA-$345_0$ in good purity: $[M + H]^+$ observed: *m/z* 8651, expected: *m/z* 8651. (C) RP-HPLC (27-81% acetonitrile/0.1% TFA) and (D) MALDI-TOF MS analysis of isolated CpA-K124ac product after all ligation steps, desulfurization, and RP-HPLC purification. $[M + H]^+$ observed: *m/z* 15884, expected: *m/z* 15882.

CpA-$2345_0$-K124ac. Addition of methoxyamine directly to the ligation mixture revealed N-terminal Cys. The mixture was dialyzed and the pH was adjusted to cleave the HMBA linker to generate CpA-2345. pH was returned to ligation conditions, and CpA-1-Nbz peptide added to generate CpA-12345-K124ac. The ligation mixture was then dialyzed extensively prior to desulfurization, and purified by RP-HPLC (**Fig. 3C, 3D**). The overall isolated yield of CpA-K124ac was 7% after five rounds of ligation, two cleavage steps, desulfurization, and RP-HPLC purification.

Nucleosomes can be reconstituted either from refolded histone octamer cores[42] or from separately refolded $H3_2/H4_2$ tetramers and

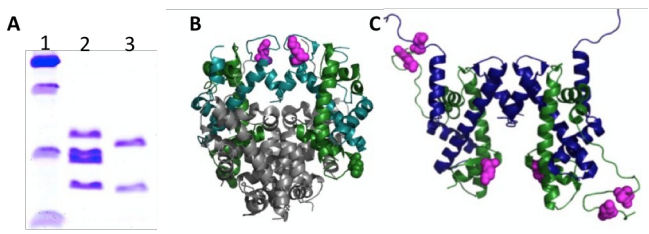



**Fig. 4** (A) SDS-PAGE of refolded histone complexes. Lane 1: molecular weight standards. Lane 2: Histone octamer containing, (from top to bottom), CpA-K124ac, H2B, H2A, and H4. Lane 3: Histone tetramer with H3 and H4-K5ac,K12ac,K91ac. (B) Structure of CpA-containing histone octamer[44] with CpA in cyan and H4 in green. (C) Structure of histone $(H3\text{-}H4)_2$ tetramer[45] with H3 in blue and H4 in green. Acetylation sites are indicated in magenta.

H2A/H2B dimers.[43] To assess the quality of the synthetic proteins, we refolded them into the relevant protein units: H4-K5ac,K12ac,K91ac into $H3_2/H4_2$ tetramer, and CpA-K124ac into octamer (**Fig. 4A**). Of note, the ~5% CpA-1345 deletion product that remained after RP-HPLC purification was eliminated through the octamer refolding process, as we predicted since this truncated CpA lacks part of an essential helix for the octamer fold, similar to effects observed for semi-synthetic H3 and H4.[23] These protein complexes will be taken forward for further study of the effects of these modifications on nucleosome structure and dynamics.

In conclusion, we demonstrate a simple hybrid ligation approach that combines both solid and solution-phase ligation chemistry for optimal yields of challenging synthetic histone protein targets. We maximized product yields through resin cleavage at an external Rink linker, with subsequent cleavage at an internal HMBA linker to generate the native carboxyl terminus. We used this approach for synthesis of a triple-modified H4 histone and, notably, for the challenging target CpA-K124ac which could not be accessed using more common expression-based approaches. We find that the key step in hybrid ligation is monitoring yields of SP-NCL to determine if there is a turnover point at which reduced release from the resin overcomes the chemical advantage of solid phase reactions.

This work was supported by the National Science Foundation (MCB0845695) and the National Institutes of Health (GM083055).

## Notes and references

1. T. Jenuwein and C. D. Allis, *Science*, 2001, **293**, 1074-1080.
2. T. Kouzarides, *Cell*, 2007, **128**, 693-705.
3. G. E. Zentner and S. Henikoff, *Nat. Struct. Mol. Biol.*, 2013, **20**, 259-266.
4. M. Simon, J. A. North, J. C. Shimko, R. A. Forties, M. B. Ferdinand, M. Manohar, M. Zhang, R. Fishel, J. J. Ottesen and M. G. Poirier, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 12711-12716.
5. E. L. Mersfelder and M. R. Parthun, *Nucleic Acids Res.*, 2006, **34**, 2653-2662.
6. A. P. M. Jack and S. B. Hake, *Chromosoma*, 2014, **123**, 355-371.
7. J. A. North, M. Simon, M. B. Ferdinand, M. A. Shoffner, J. W. Picking, C. J. Howard, A. M. Mooney, J. van Noort, M. G. Poirier and J. J. Ottesen, *Nucleic Acids Res.*, 2014, **42**, 4922-4933.
8. Y. Rodriguez, J. M. Hinz and M. J. Smerdon, *DNA Repair*, 2015, **32**, 113-119.
9. R. A. Greenberg, *Febs Letters*, 2011, **585**, 2883-2890.
10. J. Ye, X. Ai, E. E. Eugeni, L. Zhang, L. R. Carpenter, M. A. Jelinek, M. A. Freitas and M. R. Parthun, *Mol. Cell*, 2005, **18**, 123-130.
11. G. O. Elliott, K. J. Murphy, J. J. Hayes and C. Thiriet, *Nucleic Acids Res.*, 2013, **41**, 2228-2238.
12. M. Shivaraju, J. R. Unruh, B. D. Slaughter, M. Mattingly, J. Berman and J. L. Gerton, *Cell*, 2012, **150**, 304-316.
13. M. Bui, E. K. Dimitriadis, C. Hoischen, E. An, D. Quenet, S. Giebe, A. Nita-Lazar, S. Diekmann and Y. Dalal, *Cell*, 2012, **150**, 317-326.
14. K. Krassovsky, J. G. Henikoff and S. Henikoff, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **109**, 243-248.
15. C. J. Howard, R. R. Yu, M. L. Gardner, J. C. Shimko and J. J. Ottesen, *Top. Curr. Chem.*, 2015, **363**, 193-226.
16. M. M. Müller and T. W. Muir, *Chem. Rev.*, 2015, **115**, 2296-2349.
17. A. Dhall and C. Chatterjee, *ACS Chem. Biol.*, 2011, **6**, 987-999.
18. H. Pick, S. Kilic and B. Fierz, *Biochim. Biophys. Acta*, 2014, **1839**, 644-656.
19. J. C. Shimko, J. A. North, A. N. Bruns, M. G. Poirier and J. J. Ottesen, *J. Mol. Biol.*, 2011, **408**, 187-204.
20. P. Siman, S. V. Karthikeyan, M. Nikolov, W. Fischle and A. Brik, *Angew. Chem. Int. Ed.*, 2013, **52**, 8059-8063.
21. J. Li, Y. Li, Q. He, Y. Li, H. Li and L. Liu, *Org. Biomol. Chem.,* 2014, **12**, 5435.
22. M. Jbara, M. Seenaiah and A. Brik, *Chem. Commun.*, 2014, **50**, 12534-12537.
23. J. C. Shimko, C. J. Howard, M. G. Poirier and J. J. Ottesen, *Methods Mol. Biol.*, 2013, **981**, 177-192.
24. L. Raibaut, N. Ollivier and O. Melnyk, *Chem. Soc. Rev.*, 2012, **41**, 7001.
25. L. Raibaut, H. Adihou, R. Desmet, A. F. Delmas, V. Aucagne and O. Melnyk, *Chem. Sci.*, 2013, **4**, 4061-4066.
26. A. D. Baxevanis, G. Arents, E. N. Moudrianakis and D. Landsman, *Nucleic Acids Res.*, 1995, **23**, 2685-2691.
27. Y. Tanaka, M. Tawaramoto-Sasanuma, S. Kawaguchi, T. Ohta, K. Yoda, H. Kurumizaka and S. Yokoyama, *Methods*, 2004, **33**, 3-11.
28. L. E. Canne, P. Botti, R. J. Simon, Y. J. Chen, E. A. Dennis and S. B. H. Kent, *J. Am. Chem. Soc.*, 1999, **121**, 8720-8727.
29. A. Brik, E. Keinan and P. E. Dawson, *J. Org. Chem.*, 2000, **65**, 3829-3835.
30. R. K. Suto, M. J. Clarkson, D. J. Tremethick and K. Luger, *Nat. Struct. Biol.*, 2000, **7**, 1121-1124.
31. P. W. R. Harris and M. A. Brimble, *Biopolymers*, 2010, **94**, 542-550.
32. S. Mezzato, M. Schaffrath and C. Unverzagt, *Angew. Chem. Int. Ed.*, 2005, **44**, 1650-1654.
33. M. Galibert, V. Piller, F. Piller, V. Aucagne and A. F. Delmas, *Chem. Sci.*, 2015, **6**, 3617-3623.
34. A. R. Kunys, W. Lian and D. Pei, *Current Protocols in Chemical Biology*, 2012, **4**, 331-355.
35. C. Haase, H. Rohde and O. Seitz, *Angew. Chem. Int. Ed.*, 2008, **47**, 6807-6810.
36. Y. Dalal, *unpublished work*.
37. S. K. Mahto, C. J. Howard, J. C. Shimko and J. J. Ottesen, *ChemBioChem*, 2011, **12**, 2488-2494.
38. P. Siman, O. Blatt, T. Moyal, T. Danieli, M. Lebendiker, H. A. Lashuel, A. Friedler and A. Brik, *Chembiochem*, 2011, **12**, 1097-1104.
39. J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem. Int. Ed.*, 2008, **47**, 6851-6855.
40. J. B. Blanco-Canosa, B. Nardone, F. Albericio and P. E. Dawson, *J. Am. Chem. Soc.*, 2015, **137**, 7197-7209.
41. Q. Wan and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, 2007, **46**, 9248-9252.
42. K. Luger, T. J. Rechsteiner and T. J. Richmond, *Method Enzymol.*, 1999, **304**, 1-19.
43. W. Iwasaki, H. Tachiwana, K. Kawaguchi, T. Shibata, W. Kagawa and H. Kurumizaka, *Biochemistry*, 2011, **50**, 7822-7832.
44. H. Tachiwana, W. Kagawa, T. Shiga, A. Osakabe, Y. Miya, K. Saito, Y. Hayashi-Takanaka, T. Oda, M. Sato, S.-Y. Park, H. Kimura and H. Kurumizaka, *Nature*, 2011, **476**, 232-235.
45. K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, *Nature*, 1997, **389**, 251-260.