Supporting Information

Synthesis of Histone H3 Proteins by a Thioacid Capture Ligation Strategy

Xiaohong Zhang, Fupeng Li and Chuan-Fa Liu*
School of Biological Sciences, Nanyang Technological University, Singapore 637551

Keywords: thioacid capture ligation; protein chemical synthesis; histone proteins

Materials and Methods

All Fmoc and Boc-amino acid derivatives, coupling reagents, Wang resin, MBHA resin were purchased from GL Biochem (Shanghai, China) and Novabiochem (Switzerland). All chemical reagents were purchased from commercial suppliers. HPLC analyses of protein samples were run on a Shimadzu system equipped with an analytical (5 microm, 4.6 x 250 mm) or semi-preparative (5 microm, 10 x 250 mm) reverse-phase column with a flow rate of 1.0 mL/min or 2.5 mL/min, respectively. Detection was done with a UV-VIS-detector at wavelength λ = 220 nm. Preparative purifications were performed using a Waters system equipped with a C18 column (10 microm, 22 x 250 mm) with a flow rate of 10 mL/min. The HPLC buffer system was buffer A – H_2O (containing 0.045% TFA) and buffer B – 90% acetonitrile in H_2O (containing 0.04% TFA). The temperature for all HPLC experiments was 23 ºC.

Peptide/protein masses were measured on a Thermo FINNIGAN LCQ Deca XP MAX equipped with ESI ion source or on a 4800 MALDI TOF/TOF Analyzer operating in MS reflector positive ion mode and using α-cyano-4-hydroxycinnamic acid as the matrix.

An AKTA FPLC system equipped with a size-exclusion 26/60 Sephacryl S-200 column (GE Healthcare, USA) was used for the purification of histone octamer.

Experimental Procedures

1. Solid phase synthesis of Npys-H3(110-135) **3**

Preloaded Fmoc-Ala-Wang resin was used to synthesize the peptide corresponding to the C-terminal sequence of residues 110-135 of histone H3 following the standard Fmoc manual solid-phase peptide synthesis protocols. After assembling the peptide sequence, the peptide was cleaved for 3 h with a mixture of 95% TFA/2.5% TIS/2.5% H_2O in the presence of 4 eq. of 2,2'-dithio-bis-(5-nitropyridine) which was to introduce the Npys group on Cys thiol. The cleaved peptide was precipitated by cold diethyl ether. The Npys modified peptides were lyophilized and purified by RP-HPLC, and the molecular weight was determined by ESI-MS (Figure S1).

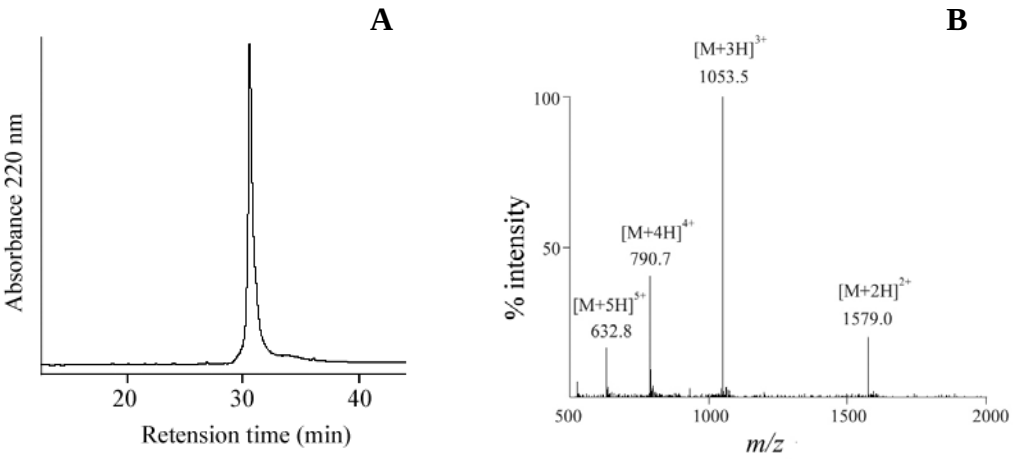

Figure S1. (A) Analytical HPLC profile of purified H_2N-**C(Npys)AIHAKRVTIMP*K*DIQLARRIRGERA**-COOH **3**. HPLC conditions: 0-40% of Buffer B in buffer A over 40 min, C18 analytical RP-HPLC column. (B) Electrospray ionization MS of **3**. MW found: 3157.8; MW calcd: 3157.6.

2. Preparation of histone protein H3(1-109)-COSH **2**

The wild type *Xenopus laevis* histone H3(1-109) gene was amplified from plasmid pET-3d-H3 by polymerase chain reaction (PCR), using a forward primer (5'-GGT GGT CAT ATG GCC CGT ACC AAG CAG ACC -3') containing a restriction site for *NdeI* and a reverse primer (5'-GGT GGT TGC TCT TCC GCA CAG GTT GGT GTC CTC AAA GAG -3') containing the *SapI* restriction site. The *NdeI-SapI* digested PCR product was ligated into an digested pTWIN1 expression vector (New

England Biolabs, USA) to give pTWIN1-H3(1-109). After DNA-sequencing verification of the ligated product, the was transformed into *E. coli* strain BL21(DE3) (Stratagene, USA).

The transformed *E. coli* strain was grown in LB medium supplemented with 100 mg/L ampicillin at 37 ℃ until optical density (OD) at 600 nm reached 0.5-0.7. Induction was performed by the addition of 0.3 mM IPTG. After incubation at 37 ℃ for 2 h, the cells were harvested by centrifugation at 6000 rpm for 10 min (Beckman, JA-10 rotor). The cell pellet was dispensed in the lysis buffer (20 mM phosphate buffer, pH 7.0, 0.5 M NaCl, 1 mM EDTA) and lysed by a microfluider at a chamber pressure of 12K/50 psi. The lysate was then centrifuged at 20,000 x *g*, 4 ℃ for 30 min. Because the expressed protein H3(1-109)-intein-CBD **1** existed in the form of inclusion bodies, the pellet was resuspended in the wash buffer (20mM phosphate, pH 7.0, 0.5 M NaCl, 1 mM EDTA, 0.5% Triton X-100) and centrifuged again at 20,000 x *g* and 4 ℃ for 30 min. The wash step was repeated twice to remove the cell debris and other impurities. The washed fusion protein pellet was dissolved in 6 M Gdn-HCl in the lysis buffer and centrifuged at 20,000 x *g* and 4 ℃ for 30 min to remove any insoluble materials. The protein solution was then dialyzed gradually against 8 M urea, 4 M urea, and 2 M urea in the lysis buffer containing 5 mM TCEP. Each dialysis step took at least 5 h at 4 ℃. The dialyzed protein solution was centrifuged at 20,000 x *g* and 4 ℃ for 30 min to remove any remaining impurities. Typically, inclusion bodies from 1L culture was dissolved in 80 mL of 6 M Gdn-HCl/lysis buffer. After dialysis, 2 M Urea/lysis buffer was added to make the volume to 120 mL.

The cleavage volume was 160 mL after adding 40 mL of concentrated cleavage buffer: 0.4 M Na_2S, 1 M HEPES, 1 mM EDTA. So the final conditions of hydrothiolysis to obtain H3(1-109)-COSH were: 0.1 M Na_2S, 0.25 M HEPES, 1 mM EDTA, 0.37 M NaCl, pH 8.0. The cleavage was performed overnight at 4 ºC.[1] H3(1-109)-COSH was formed as precipitates after cleavage, which were collected by centrifugation at 20,000 x *g* and 4 ℃ for 30 min. The H3(1-109) thioacid pellet was dissolved in 6 M

Gdn-HCl/lysis buffer containing 5 mM TCEP. After centrifugation, the supernatant was purified by C18 prep RP-HPLC (Waters 2487 HPLC system, Prosphere C18, 250 x 22 mm) using a gradient: 0% to 50% in 25 min, then to 80% in 30 min of buffer B (90% ACN/H_2O containing 0.05% TFA) in buffer A (0.05% TFA/H_2O) at a flow rate of 10 mL/min. The molecular mass of H3(1-109)-COSH was confirmed by MALDI-MS (*m/z* $[M+H]^+$ found: 12302.8, MW calcd: 12301.3. See Figure 1 in main text). The purified H3(1-109)-COSH was lyophilized for 3 times. The yield of histone H3(1-109)-COSH **2** was around 10 mg per liter of *E. coli* culture.

3. Synthesis of histone H3 **4** through ligation of H3(1-109)-COSH **2** with Npys-modified H3(110-135) **3**

H3(1-109)-COSH (~0.5 mM) and the Npys-modified H3 C-terminal peptide Npys-H3(110-135) (~1.5 mM) were reacted in 6 M Gdn-HCl, 0.1 M phosphate (pH 6). A yellow color developed immediately upon mixing the two reactants. The reaction was allowed to continue for half an hour before DTT (100 mM) was added to stop the reaction. The ligated protein histone H3 was analyzed and purified by RP-HPLC. The molecular weight of the ligation product was determined by MALDI-MS. *m/z* [M+H]+ found:15272.1; MW calcd: 15270.2. See Figure 2 in the main text.

4. Overexpression and purification of recombinant Histone H3(14-135)/K14C **6**

Construction of plasmid pET-3d-H3(14-135)/K14C overexpression system. The wild type *Xenopus laevis* histone H3 gene was first mutated at Cys110 to Ala by a QuickChange Site-Directed Mutagenesis Kit (Stratagene), using a forward primer 5'-GAG GAC ACC AAC CTG *GCC* GCC ATC CAC GCC AAG -3' and a reverse primer 5'-CTT GGC GTG GAT GGC *GGC* CAG GTT GGT GTC CTC -3'. The mutated gene was transformed into XL1-Blue $CaCl_2$ competent cell (Stratagene), and amplified.

Amplified plasmid pET-3d-H3/C110A, refered to as the wild type here, was subjected to the second mutation to delete amino acid residues 1-13 and mutate Lys14 to Cys. The forward and reverse primers for the second mutation were: 5'-CTT TAA GAA GGA GAT ATA CAT ATG *TGC* GCT CCC CGC AAG CAG CTG GCC ACC -3' and 5'-GGT GGC CAG CTG CTT GCG GGG AGC GCA CAT ATG TAT ATC

TCC TTC TTA AAG -3', respectively. The mutated gene was transformed into $CaCl_2$-competent BL21(DE3)pLysS cell (Stratagene).

Over-expression and purification were done as previously described [2] with minor modifications. The extracted H3(14-135)/K14C were purified by C18 prep RP-HPLC using a gradient of 0%-50% for 25 min, then to 80% for 30 min of buffer B (90% ACN/0.05% TFA) in buffer A (0.05% TFA/H_2O) at a flow rate of 10 mL/min. The purified protein was lyophilized and molecular weight was determined by MALDI-MS.

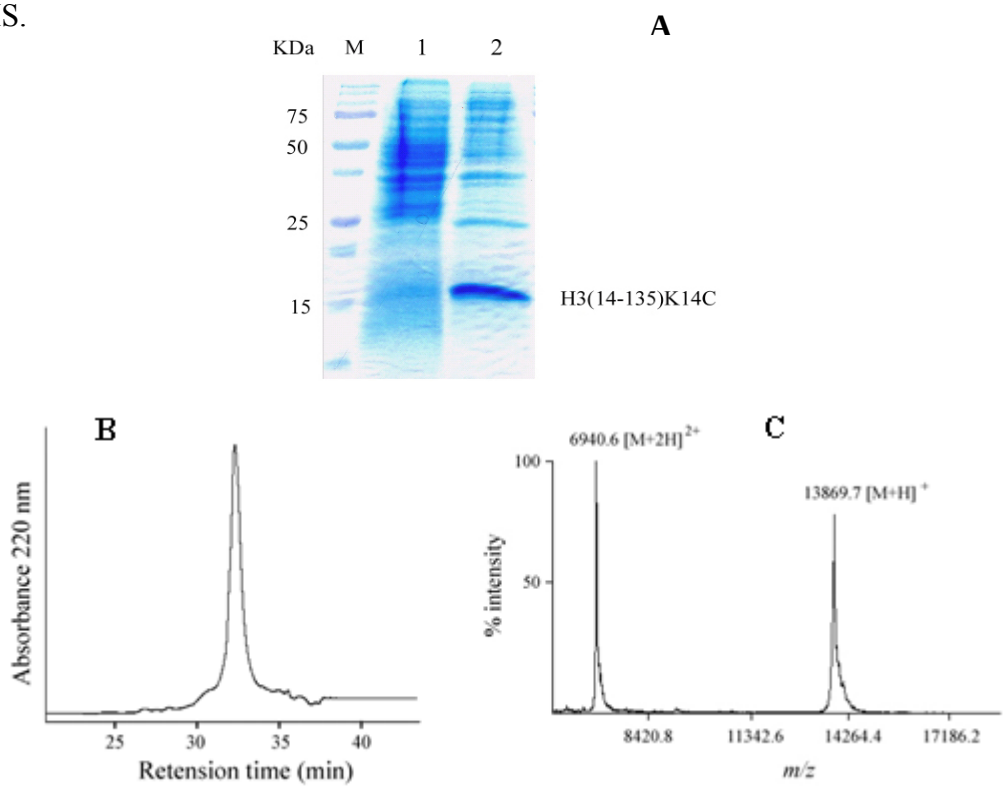

Figure S2. Expression and characterization of H3(14-135)K14C **6**. (**A**) 15% SDS-PAGE gel analysis of H3(14-135)K14C expression in *E. coli*. Lane 1, uninduced cell culture; lane 2, induced cell culture by IPTG. The expressed protein is in the inclusion body. (**B**) C8 semi-prep RP-HPLC purification of H3(14-135)K14C. HPLC condition: 0% to 50% in 25 min, then to 60% in 20 min of buffer B in buffer A. (**C**) MALDI-TOF MS of H3(14-135)K14C, which shows that the expressed protein has no methionine at its N-terminus. *m/z* $[M+H]^+$ found:13869.7, MW calcd: 13870.3.

5. Preparation of H3(14-135)/K14C(Npys) **7**

The lyophilized H3(14-135)/K14C (~1 mM) was reacted with 4 eq. of 2,2'-dithiobis-(5-nitropyridine) in 3:1 of acetic acid/water at rt for 4 h. The HPLC-purified Npys-H3(14-135)/K14C was lyophilized and its MW was measured by ESI-MS. The yield of the isolated product was about 80%.

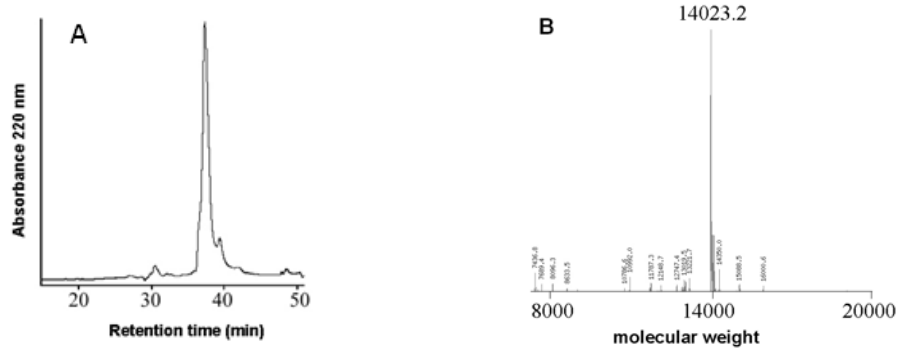

Figure S3. Synthesis of Npys-modified H3(14-135)/K14C **7**. (**A**) C8 semi-prep purification of H3(14-135)/K14C after the Npys-modification reaction. HPLC condition: 0% to 50% in 25 min, then to 60% in 20 min of buffer B in buffer A. (**B**) Deconvoluted ESI mass spectrum (insert) of Npys-H3(14-135)K14C. MW found: 14023.2, MW calcd: 14024.5.

6. Chemical synthesis of histone H3(1-13)K4me2 thioacid **5**

H3(1-13)K4me2 thioester, H_2N-**ARTK(**Me_2**)QTARKSTGG**-CO-$SCH_2CH_2CONH_2$ with N^{ε}- dimethylated Lys4 was synthesized by standard Boc SPPS. The MBHA resin was first derivatized with mercaptopropionic acid, followed by assembling of the peptide sequence. After HF deprotection and cleavage, the peptide thioester was released from the resin. The peptide thioester was purified by RP-HPLC and the molecular weight was determined by ESI-MS (*m/z* $[M+2H]^{2+}$ found: 739.5, MW calcd: 1477.8). The lyophilized peptide thioester was then subjected to hydrothiolysis in 0.1 M Na_2S in 0.25 M HEPES, pH 8.0, at room temperature for 4 h.[3] After HPLC purification, the pure peptide thioacid was obtained. Its MW was confirmed by ESI-MS (*m/z* $[M+2H]^{2+}$ found: 703.7, MW calcd: 1406.7).

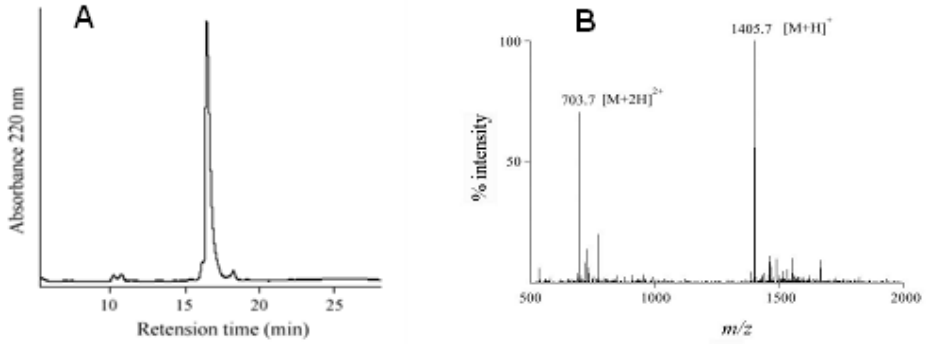

Figure S4. Characterization of H3(1-13)/K4me2 thioacid **5**. (**A**) Analytical HPLC profile of H_2N-**ART*Kme2*QTARKSTGG**-COSH. HPLC condition: 0% to 30% of Buffer B in buffer A over 30 min. (**B**) Mass spectrum of this peptide thioacid determined by ESI-MS. *m/z* $[M+2H]^{2+}$ found: 703.7, MW calcd: 1406.7.

7. Synthesis of histone H3 dimethyl Lys4, H3K4me2/K14C **8**, through ligation of H3(1-13)K4me2-COSH with Npys-modified H3(14-135)/K14C

The ligation condition was the same as in the case of the ligation between H3(1-109)-COSH and Npys-modified H3(110-135). However, the amount of peptide thioacid **5** used was about 2.4 eq. of that of Nyps-modified H3(14-135)/K4C **7** (*ca.* 0.5 mM). The ligated protein was analyzed and purified by RP-HPLC. The molecular weight of the ligation product was determined by MALDI-MS. m/z [M+H]+ found: 15239.0, MW calcd: 15241.0.

Alkylation of H3K4me2/K14C with 2-bromoethylamine hydrochloride. 0.5 mM H3K4me2/K14C was treated with 2-bromoethylamine hydrochloride (140 mM) in 6 M Gdn-HCl, 1 M HEPES, 20 mM D/L methionine, 5 mM TCEP at pH 7.8 for 10 h.[4] The alkylated product H3K4me2/K^{S}14 was isolated by C8 semi-preparative RP-HPLC and analyzed by MALDI-MS (Figure S5). *m/z* [M+H]+ found: 15281.4, MW calcd: 15284.0.

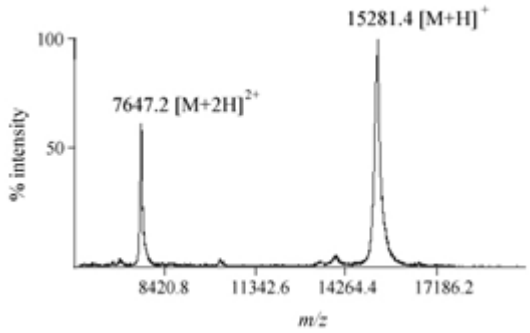

Figure S5. MALDI-MS analysis of the alkylation product H3K4me2/K^{S}14.

8. Histone octamer formation and purification[5]

Expression and purification of recombinant histone proteins (wild type). The plasmids pET-3a containing the *Xenopus laevis* histone H2A, H2B and H4 gene and pET-3d containing H3 gene were got from Dr. C.A. Davey's lab. Expression and purification conditions were the same as in the case of H3(14-135)/K14C.

The four histones with equal molar amount (around 1mg each) were individually dissolved in the unfolding buffer (7 M Gdn-HCl, 10 mM Tris-HCl, pH 7.5, 10 mM

DTT) to a final concentration of 2 mg/mL. For histone protein H3, 20 mM DTT should be added. After 30 min unfolding, the four proteins were mixed together. The mixed solution was dialyzed against 600 mL of refolding buffer (2 M NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM Na-EDTA, 10 mM 2-mercaptoethanol) at 4 ℃. The dialysis was done for three times, at least 4 h for each time. The precipitated material formed during dialysis was removed by centrifugation at 20,000 *g* for 10 min at room temperature. The supernatants was concentrated by Amicon concentrator (MW cut-off of 10 kDa) and purified by size-exclusion chromatography using the 26/60 Sephacryl S-200 column which was previously equilibrated with refolding buffer. The fractions were collected and confirmed by 15% SDS-PAGE. The purified octamer solution was mixed with equivolume glycerol, and stored at -20 ℃.

9. Assembly of the nucleosome core particles

5’-CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTC-TAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTAACCGCCAA GGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGT-3’ was synthesized by Proligo as a template to prepare the 146 bp core DNA.[6] 5’-CGGGATCCCGGCGCCCTGGAGAATCCCGGTGCC-3’ and 5’-GTCAGATA-TATACATCCTGTGCATGGAAGATCTTCGCTCGAGCG-3’ were used as the forward and reverse primers, respectively. The amplified DNA was gel purified and ligated to T-vector (Promega). After sequencing to verify the sequence, the T-vector was transformed into *E. coli* XL1-Blue $CaCl_2$-competent cell. After replication and extraction, the T-vector containing 185 bp core DNA was used as a template to prepare this 185 bp core DNA.

Histone octamer was mixed with the 185 bp core DNA fragment for a final 0.9 molar ratio of octamer to DNA with a final DNA concentration of 0.2 μM. KCl was weighed to 2 M, and DTT was added to a concentration of 10 μM. The mixture was let stand at 4 ℃ for 30 min, and stepwisely dialyzed against TCS-0.85 buffer (0.85 M KCl, 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT), TCS-0.65 buffer (0.65 M KCl, 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT), TCS-0.45 buffer (0.45 M KCl, 20

mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT), and TCS-0 buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT).[7] Each dialysis took 2.5 h. The nucleosome core particle formation was verified by resolving on a 5% native PAGE gel (5% acrylamide, 0.15% bis-acrylamide), followed by staining with ethidium bromide.

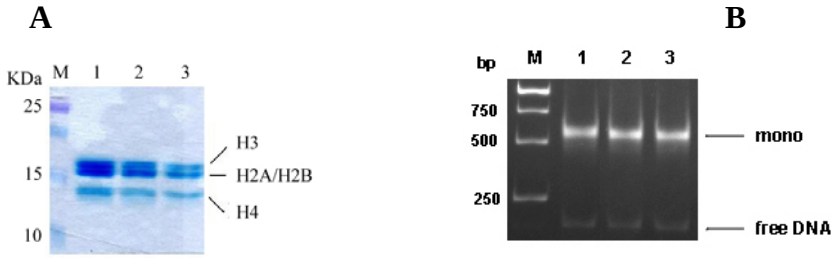

Figure S6. (A) 15% SDS-PAGE gel analysis of histone octamer formed from different H3 proteins with other three core histone proteins H2A, H2B and H4. Lane 1, wild type H3; lane 2, synthetic H3; lane 3, with H3K4me2/K^{S}14. **(B)** Analysis of nucleosome core particles on 5% native PAGE gel. Lane 1, with wild type H3; lane 2, with synthetic H3; lane 3, with H3K4me2/K^{S}14. **Mono** indicates the mononuclesome formed by the octamer and 185 bp DNA template. **Free DNA** indicates the unbound 185 bp DNA.

References

1. Zhang, X.; Li, F.; Lu, X.-W.; Liu, C.-F. Protein C-terminal modification through thioacid/azide amidation. *Bioconjugate Chem.* **2009**, *20*, 197-200.

2. Shogren-Knaak, M. A.; Peterson, C. L. Creating designer histones by native chemical ligation. *Methods Enzymol* **2004,** *375*, 62-76.

3. Tan, X. H.; Zhang, X.; Yang, R.; Liu, C. F. A simple method for preparing peptide C-terminal thioacids and their application in sequential chemoenzymatic ligation. *Chembiochem* **2008,** *9* (7), 1052-6.

4. Simon, M.D. et al. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007**, *128*, 1003-1012.

5. Luger, K.; Rechsteiner, T. J.; Flaus, A. J.; Waye, M. M.; Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* **1997,** *272* (3), 301-11.

6. Lowary, P. T.; Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **1998,** *276* (1), 19-42.

7. Richmond, T. J.; Searles, M. A.; Simpson, R. T. Crystals of a nucleosome core particle containing defined sequence DNA. *J Mol Biol* **1988,** *199* (1), 161-70.