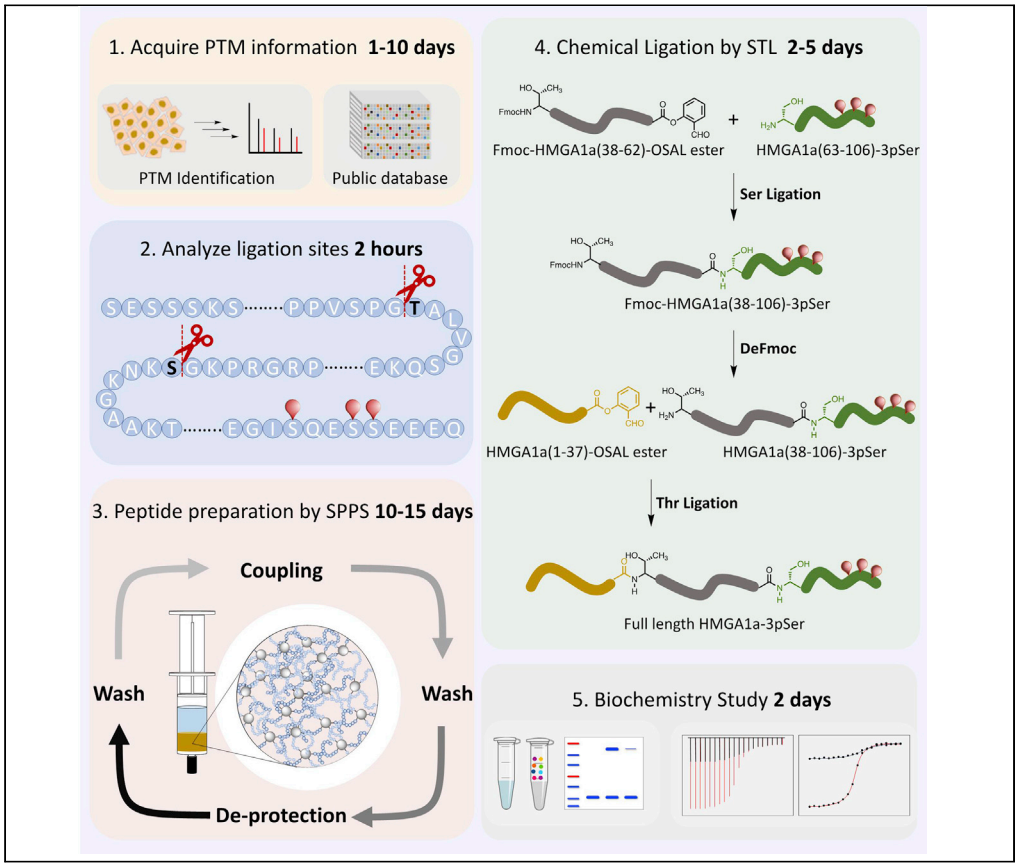
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Serine/threonine ligation-assisted chemical synthesis of HMGA1a protein with site-specific post-translational modifications



Dissecting the function of proteins' post-translational modifications (PTMs) is seriously hindered by the difficulty in obtaining the homogeneous protein with the PTMs of interest. Chemical protein synthesis offers a great potential to overcome this limitation. Here, a detailed protocol is introduced for chemical synthesis of HMGA1a protein with site-specific modifications via Ser/Thr ligation strategy, by which we can systematically study the function of the triple phosphorylation (3pSer) in the HMGA1a acidic tail.

Tongyao Wei, Heng Liu, Hongxiang Wu, Fan Pu, Xuechen Li

wei_tongyao@163.com (T.W.)
xuechenl@hku.hk (X.L.)

Highlights

Proteins with PTMs can be produced by ligation-assisted chemical synthesis

Phosphorylated HMGA1a proteins were chemical synthesized

Binding partners can be profiled by streptavidin-biotin affinity pull-down assay

Wei et al., STAR Protocols 2, 100777
September 17, 2021
https://doi.org/10.1016/j.xpro.2021.100777

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Serine/threonine ligation-assisted chemical synthesis of HMGA1a protein with site-specific post-translational modifications

Tongyao Wei,[1,2,*] Heng Liu,[1] Hongxiang Wu,[1] Fan Pu,[1] and Xuechen Li[1,3,*]

[1]Department of Chemistry, State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Hong Kong, P.R. China

[2]Technical contact

[3]Lead contact

*Correspondence: wei_tongyao@163.com (T.W.), xuechenl@hku.hk (X.L.)
https://doi.org/10.1016/j.xpro.2021.100777

SUMMARY

Dissecting the function of proteins' post-translational modifications (PTMs) is seriously hindered by the difficulty in obtaining the homogeneous protein with the PTMs of interest. Chemical protein synthesis offers a great potential to overcome this limitation. Here, a detailed protocol is introduced for chemical synthesis of HMGA1a protein with site-specific modifications via Ser/Thr ligation strategy, by which we can systematically study the function of the triple phosphorylation (3pSer) in the HMGA1a acidic tail.

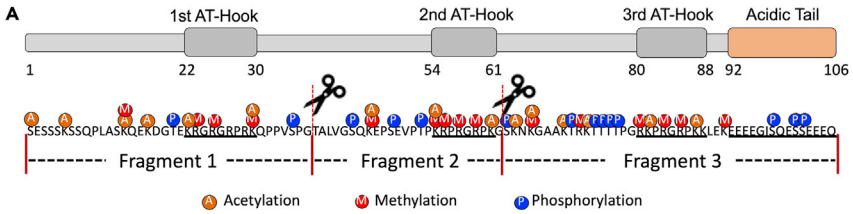
For complete details on the use and execution of this protocol, please refer to Wei et al. (2021).

BEFORE YOU BEGIN (THOMPSON AND MUIR, 2019)

Over the past decades, a variety of new and refined proteomics technologies have emerged, which significantly improve our abilities to profile the PTMs spatiotemporally. Nowadays, Identification of PTMs even has become commercially available. In addition, the public literature databases (e.g., PubMed, Uniprot) enable us to find the interested PTMs on proteins of interest. To study the interested PTMs, accessing homogenous proteins with site-specific PTMs is crucial. Direct extraction from cells or enzymatic PTM installation on the recombinant protein still remains challenging. Instead, protein chemical synthesis has emerged as an effective way to introduce PTMs site-selectively and homogeneously.

Solid-phase peptide synthesis (SPPS) can efficiently prepare the polypeptides chemically without complicated setups. However, due to the truncated sequences accumulated from the incomplete coupling, SPPS generally is not used to prepare peptides with more than 50 amino acids. Despite many efforts being made to expand the SPPS' capability, convergent synthesis, which bypasses the formidable challenge of SPPS, has become the most common way to accessing long peptide sequences (up to proteins). Specifically, chemical peptide ligation technologies, such as native chemical ligation (NCL) and its variants, KAHA ligation, Serine/Threonine ligation (STL), and cysteine/penicillamine ligation (CPL), can selectively and specifically forge a natural peptidic bond between the C- and N-terminus of two unprotected peptides. Therefore, a protein containing noncanonical elements can be re-constructed by stitching peptides together via several rounds of peptide ligation (Tan et al., 2020).

Of the ligation technologies developed to date, the C-terminal salicylaldehyde ester-based chemoselective condensation, STL, has been widely used for synthesis of cyclic peptides and proteins. (Zhang et al., 2013; Tan et al., 2020). This reaction occurs between the N-terminal serine or threonine



B

C-terminal amino acid

	Fast							Medium						Slow				Decomposed		
Residues	Ala	Gly	Ser	Gln	Thr	Phe	Cys*	Val	Ile	Met	Asn	Tyr	Leu	His	Trp	Arg	Pro	Lys	Asp	Glu
Conversion (%)	87	86	85	79	71	68	65	46	42	39	39	34	33	29	25	21	8	-	-	-

Figure 1. Analysis of ligation sites

(A) PTMs of HMGA1a and disconnection sites for chemical ligation are shown. The synthesis involves C-to-N sequential Ser/Thr ligations. This figure is reprinted with permission from Wei et al., 2021.

(B) Effect of C-terminal amino acid on STL by a model study. The reaction conversions were calculated after 2 h. The Lys, Asp, and Glu ester is hard to make. *: Cys is SStBu protected. The data was adopted from Liu and Li, 2018.

or cysteine peptide fragment and the peptide fragment with C-terminal salicylaldehyde ester to afford an *N,O/S*-benzylidene-acetal-linked intermediate in a pyridine/acetic acid mixture. After acidolysis with TFA/H_2O, the native amide bond linkage will be formed at the ligation site (Tan et al., 2020). It is noteworthy that proteins with more than 400 residues are still not routinely reachable by chemical total synthesis, since it requires many rounds of ligation. Finally, the total yield will be extremely low.

In this protocol, we describe the procedure to chemically produce the homogenous HMGA1a with the triple phosphorylation on its acidic tail by STL. HMGA1a with other PTMs also can be achieved by this protocol (e.g., methylation and acetylation). High-Mobility Group A1a (HMGA1a) is one of non-histone chromatin proteins and involved in many cell nuclear processes (gene transcription, DNA replication and repair, heterochromatin organization, etc.) (Reeves, 2001). Accumulating studies have found that like Histone protein, HMGA1a carries a variant of protein posttranslational modifications (PTMs) including methylation, acetylation, phosphorylation and so on (Figure 1A). However, unlike various defined functions on the histone protein, most of PTMs' functions on HMGA1 still remain to be illustrated (Zhang and Wang, 2008). Here, we focused on the triple phosphorylation on the acidic tail.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-P53 (1:200)	Santa Cruz	Cat# sc-126; RRID: AB_628082
Anti-mouse IgG (1:2000)	Cell Signaling Technology	Cat# 7076S; RRID: AB_330924
Chemicals, peptides, and recombinant proteins		
2-Chlorotrityl chloride resin	CSBio	Cat# AR-FCRX-0010
Fmoc-Ala-OH	CSBio	Cat# AR-FALA-0010
Fmoc-Ser(tBu)-OH	CSBio	Cat# AR-FSER-0010

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Fmoc-Glu(OtBu)-OH	CSBio	Cat# AR-FGLU-0010
Fmoc-Thr(tBu)-OH	CSBio	Cat# AR-FTHR-0010
Fmoc-Lys(Boc)-OH	CSBio	Cat# AR-FLYS-0010
Fmoc-Gln(Trt)-OH	CSBio	Cat# AR-FGLN-0010
Fmoc-Pro-OH	CSBio	Cat# AR-FPRO-0010
Fmoc-Leu-OH	CSBio	Cat# AR-FLEU-0010
Fmoc-Asp(OtBu)-OH	CSBio	Cat# AR-FASP-0010
Fmoc-Gly-OH	CSBio	Cat# AR-FGLY-0010
Fmoc-Arg(Pbf)-OH	CSBio	Cat# AR-FARG-0010
Fmoc-Val-OH	CSBio	Cat# AR-FVAL-0010
Fmoc-Asn(Trt)-OH	CSBio	Cat# AR-FASN-0010
Fmoc-Ile-OH	CSBio	Cat# AR-FILE-0010
Boc-Ser(tBu)-OH	CSBio	Cat# AR-BSER-0006
Fmoc-Gly(Hmb)-OH	Santa Cruz	Cat# sc-285793
Fmoc-Ser(HPO3Bzl)-OH	Bidepharm	Cat# BD18628-5g
Hexafluorophosphate azabenzotriazole tetramethyl uronium (HATU)	GL Biochem	Cat# 00703
Dichloromethane (DCM)	Fisher Chemical	Cat# D/1852/17
Trifluoroacetic acid (TFA)	J&K	Cat# 101398
2,2,2-Trifluoroethanol (TFE)	J&K	Cat# 269931
N,N-Dimethylformamide (DMF)	AQA	Cat# DF-1161-2500
Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate (PyBOP)	Chem-Impex	Cat# 02276
N, N-Diisopropylethylamine (DIEA)	Energy Chem	Cat# W3200145000
Acetic acid (AcOH)	DUKSAN	Cat# 3839
CH3CN	DUKSAN	Cat# 2957
Piperidine	Merck	Cat# 8.22299.0500
Phenol	Energy Chem	Cat# W6105175000
Hydroxybenzotriazole (HOBt)	GL Biochem	Cat# 00602
N-Methyl-2-pyrrolidone (NMP)	DUKSAN	Cat# 674
4-Dimethylaminopyridine (DMAP)	Energy Chem	Cat# B040170
Triethylamine (Et3N)	Energy Chem	Cat# B0100655000
Acetic anhydride (Ac2O)	Energy Chem	Cat# A0105110250
Triisopropylsilane (Tips)	TCI	Cat# T1533
Diethyl ether	RCI Labscan	Cat# AR1044B
Potassium cyanide (KCN)	Sigma	Cat# 207810-25G
Pyridine	Sigma	Cat# 270970-1L
Ninhydrin	Energy Chemical	Cat# A050135
Ethanol	Scharlab	Cat# ET00232500
Hydrazine monohydrate	TCI	Cat# H0172
Hexane	AQA	Cat# HX-1251-020L
Triethylamine (TEA)	Energy Chemical	Cat# B010065
Diethanolamine (DEA)	Sigma	Cat# 471216-2.5L
PBS (1×)	Transgen	Cat# FG701-01
DMEM	Thermo Fisher	Cat# 11995-040
Fetal Bovine Serum (FBS)	Thermo Fisher	Cat# 10270
Trypsin	Transgen Biotech	Cat# FG301-01
Penicillin-Streptomycin (10,000 U/mL)	Thermo Fisher	Cat# 15140122
Pierce Protease Inhibitor Mini Tablets, EDTA-free	Thermo Fisher	Cat# A32955
NP-40	Beyotime	Cat# ST366
Glycerol	Sigma	Cat# G5516
BCA assay	Beyotime	Cat# P0012
High Capacity Streptavidin Agarose	Thermo Fisher	Cat# 20359
NaCl	Sigma	Cat# S3014
Tris(hydroxymethyl)aminomethane (Tris)	Macklin	Cat# T822117
PVDF membrane	Immobilon	Cat# ISEQ00010

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
HMGA1a	This study	N/A
Biotin-HMGA1a	This study	N/A
HMGA1a-3pSer	This study	N/A
Biotin-HMGA1a-3pSer	This study	N/A
Experimental models: Cell lines		
HEK293T	HUEN, Michael Shing-Yan Laboratory, HKU	N/A
MCF-7	Chi-Ming Che Laboratory, HKU	N/A
NB-4	Chi-Ming Che Laboratory, HKU	N/A
Software and algorithms		
UniDec	University of Oxford	Version 1.0.10 (https://github.com/michaelmarty/UniDec/releases)
Other		
LC-MS (ACQUITY UPLC H-Class PLUS System)	Waters	Cat# 720003268en
Preparative HPLC	Waters	
Hei-VAP Core Rotary Evaporator	Heidolph	Cat# 02-321-081
[Resin and amino acids from CSBio are recommended for peptide synthesis]		

MATERIALS AND EQUIPMENT

Solution preparation

[10 mM KCN solution]		
Reagent	Final concentration	Amount
Potassium cyanide (KCN)	10 mM	16.5 mg
ddH_2O	n/a	25 mL
Total	**n/a**	**25 mL**
[Can be store at 25°C up to one year]		

[Kaiser test solution A]		
Reagent	Final concentration	Amount
KCN solution (10 mM)	0.2 mM	1 mL
Pyridine	98%	49 mL
Total	**n/a**	**50 mL**
[Can be store at 25°C up to one year]		

[Kaiser test solution B]		
Reagent	Final concentration	Amount
Ninhydrin	5%	1 g
Ethanol	n/a	20 mL
Total	**n/a**	**20 mL**
[Can be store at 25°C up to one year]		

[Kaiser test solution C]		
Reagent	Final concentration	Amount
Phenol	n/a	20 g
Ethanol	n/a	10 mL
Total	**n/a**	**10 mL**
[Can be store at 25°C up to one year]		

⚠ CRITICAL: Freshly mix 10 μL solution A, 10 μL solution B, and 10 μL solution C together in a 1.5 mL Eppendorf tube before use. And take several grains of resin into 30 μL Kaiser solution. Followed by heating for 3 min at 110°C. If the coupling is finished, the solution will remain yellow; otherwise, the solution will turn to a dark blue color.

[Lysis buffer]

Reagent	Final concentration	Amount
Tris-HCl (1M); pH 7.4	20 mM	1 mL
NaCl (1.5M)	150 mM	5 mL
NP-40	1%	0.5 mL
PMSF (100 mM)	1 mM	0.5 mL
ddH_2O	n/a	43 mL
Total	**n/a**	**50 mL**

[Can be store at 4°C for up to one year, Pierce Protease inhibitor Mini Tablet (1 tablet for 5 mL lysis buffer) will be added before use]

[4 × loading buffer]

Reagent	Final concentration	Amount
Tris-HCl (1M); pH 6.8	200 mM	10 mL
SDS	8%	4 g
DTT	40 mM	3.1 g
Bromphenol blue	0.4%	0.2 g
Glycerol	40%	20 mL
ddH_2O	n/a	20 mL
Total	**n/a**	**50 mL**

[Can be store at 25°C up to one year, can be diluted accordingly]

STEP-BY-STEP METHOD DETAILS

Analyze the ligation sites and synthesis design

Timing: [2 h]

This section describes the sequence analysis for disconnection sites.

1. Analyze ligation sites.

Note: HMGA1a is rich with Ser and Thr residues, which provides multiple options for conducting peptide ligation. Therefore, STL strategy was employed to generate the full length HMGA1a (Figure 1A).

Note: The C-terminal Gly, Ala, Ser, Gln, Thr, Phe, and Cys(SStBu) salicylaldehyde esters generally give higher conversions under the standard STL ligation conditions (Figure 1 B), Therefore, we chose G37^T38 and G62^S63 as the disconnection sites for peptide ligation.

Note: Before peptide fragment synthesis, we can use some peptide hydrophobicity /hydrophilicity analysis tools (e.g. https://www.peptide2.com/N_peptide_hydrophobicity_hydrophilicity.php) to evaluate the hydrophobicity, as very hydrophobic peptide sequence will aggregate during peptide elongation and results in the failure of peptide preparation. If so, it is better to choose

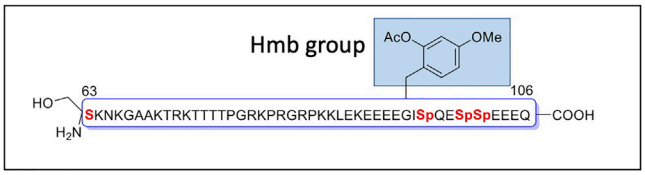


Figure 2. The sequence of fragment 3
This figure is reprinted with permission from Wei et al., 2021.

another ligation site with less hydrophobic fragments. Alternatively, removable N-(2-hydroxy-4-methoxy benzyl) (Hmb) tag can be used to improve synthesis of the hydrophobic peptides.

2. Protection design.

Note: HMGA1a synthesis involved three fragments via two ligations in the C-to-N direction; thus, the N-terminal Thr of the fragment 2 was temporarily protected by the Fmoc group. Otherwise, the N-terminus of fragment 2 may also react during ligation reaction.

HMGA1a fragment 3 preparation

Timing: [2 weeks]

This section describes the preparation of HMGA1a (63–106) peptide fragment along with the removable N-(2-hydroxy-4-methoxy benzyl) (Hmb) tag (Figure 2). Peptide fragments can be readily produced by peptide synthesizer (e.g., CEM Liberty Blue). Alternatively, we can prepare them manually via solid phase peptide synthesis. Typically, one equivalent 2-chlorotrityl chloride resin reacts with four equivalents of the first amino acid and coupling reagent in DMF. HATU/DIEA were used as the coupling reagents.

Note: Hmb tag is introduced to increase the peptide solubility.

3. Prepare peptide fragment 3 following the standard Fmoc-SPPS protocol (Figure 3).
   a. Resin Swelling and first amino acid coupling.
      i. Weight 200 mg 2-chlorotrityl chloride resin (resin loading: 0.4 mmol/g) into the reaction vessel.
      ii. Add 5mL anhydrous DCM to vessel and shake for 15 min.
      iii. Dry the resin on a vacuum pump.
      iv. Add 5mL anhydrous DCM, the first amino acid (Fmoc protected) of the C-terminus of the fragment (0.32 mmol) and 56 μL DIEA (0.32 mmol). And shake at 25°C for 2 h to generate the first amino acid anchored resin.
      v. Dry the resin on a vacuum pump.
      vi. Wash by 10 mL DCM and 10 mL DMF for 3 times, respectively.
      vii. The obtained resin is swelled in 5 mL DCM containing 100 μL methanol and 56 μL DIEA (0.32 mmol).
      viii. The mixture is shaken for 30 min.
      ix. The resin is washed by DCM for 3 times and dry for next step.
   b. Deprotection: Add 5 mL 20% piperidine in DMF (v/v) to the reaction vessel. Shake for 15 min at 25°C.
   c. Coupling: Weight corresponding (0.32 mmol) Fmoc protected amino acid, 121.6 mg HATU (0.32 mmol) in a 15 mL conical tube, dissolved by 5 mL DMF and followed by adding 56 μL DIEA (0.32 mmol). Transfer the mixture into the reaction vessel. Shake for 2 h at 25°C.

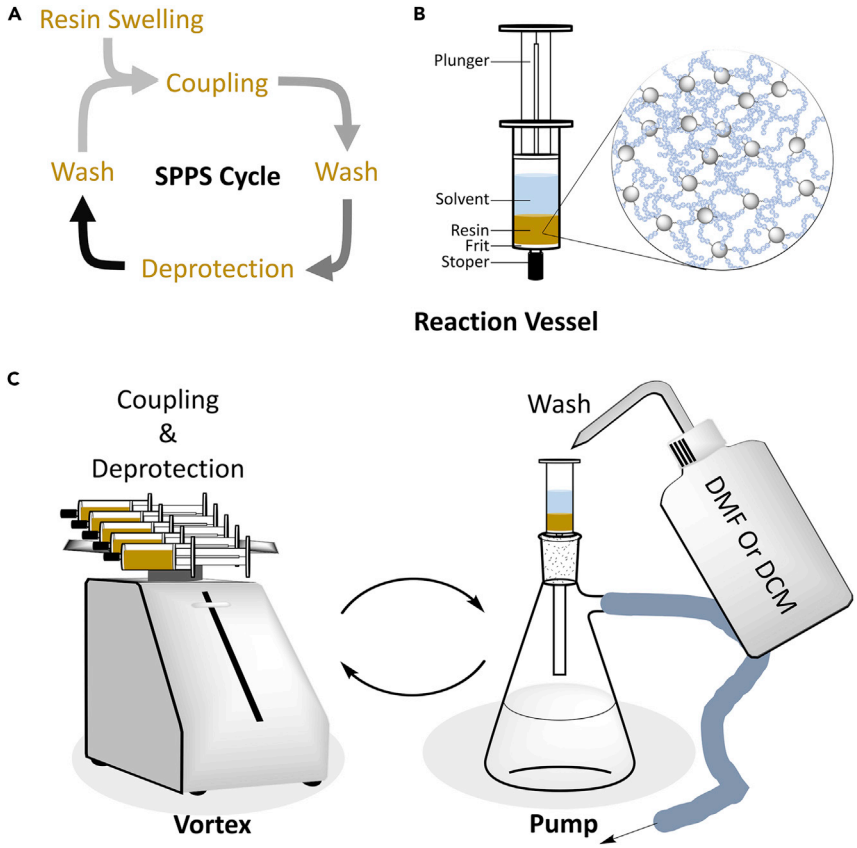


Figure 3. The overview of SPPS scheme

(A) SPPS flow chart.

(B) The overview of the reaction vessel for manual SPPS.

(C) Set-up of coupling, deprotection, and wash step.

Note: Coupling time can be elongate to 12 h.

d. Wash: Dry the resin and wash by 10 mL DMF × 3 times, 10 mL DCM × 3 times and 10 mL DMF × 3 times. Finally, the resin is dried for next step.

e. Deprotection: Add 5 mL 20% piperidine in DMF (v/v) to the reaction vessel. Shake for 15 min at 25°C.

f. Wash: Dry the resin and wash by 10 mL DMF × 3 times, 10 mL DCM × 3 times and 10 mL DMF × 3 times. Finally, the resin is dried for next amino acid coupling (go to step c).

Note: Repeated step c-f until all amino acids are coupled and then proceed to step g.

g. Cleavage from resin.

   i. Achieving side chain unprotected peptide: (Before the cleavage step, the peptide bound resin should wash by 10 mL DCM for 3 times) After completion of the synthesis, cleavage and global side chain deprotection can be achieved in the same manner by adding 10 mL cleavage cocktail A (TFA/H_2O/phenol (90/5/5, v/v/v)) and shaking for 1h at 25°C. After cleavage, the resin is filtered, and combined filtrates are blown-off under a stream of compressed air. The crude product is precipitated by 40 mL cold diethyl ether, and followed by centrifugation immediately at 4000 *g* for 5 min. Then the ether is subsequently decanted. The obtained solid is ready for HPLC purification.

   ii. Achieving sidechain full-protected peptide: (Before the cleavage step, the peptide bound resin should wash by 10 mL DCM for 3 times) Cleavage of peptide with sidechain protection can be achieved by adding 10 mL mild acidic cleavage cocktail B (CH_2Cl_2/2,2,2-trifluoroethanol

(TFE)/AcOH (8/1/1, v/v/v)) and shaking for 1 h at 25°C. After cleavage, the resin is filtered, and the combined filtrates are evaporated under Rotavap. The resulting oily crude product is further dissolved by 10 mL DCM and 30 mL hexane, followed by evaporation under vacuum to co-evaporate all acetic acid with hexane. The co-evaporation is repeated for 5 times to give a powder, which can be directly used for the preparation of the C-terminal peptide salicylaldehyde ester.

⚠ CRITICAL: The coupling from Glu92-Glu95 need to be monitored by the Kaiser test to make sure the coupling is finished as the continuous Glu residues cause strong steric hindrance and is difficult to couple. Also, decrease the loading of the resin to 0.3 mmol/g to reduce the steric hindrance during the coupling.

Note: Coupling Fmoc-Ser(HPO_3Bzl)-OH at Ser98, Ser101, and Ser102 using HATU/HOBt (4 equivalent) in DMF/NMP (1/1, v/v) as coupling reagent.

⚠ CRITICAL: Coupling Fmoc-Gly(Hmb)-OH at Gly96 for 12 h. Fmoc-Gly(Hmb)-OH is commercially available, or refer to the synthetic route at Wei et al., 2021.

4. After coupling of the last Boc-Ser(tBu)-OH, treat resin with 10 mL 2% hydrazine in DMF (v/v) and shake for 30 min to liberate hydroxyl group on Hmb.

   Note: The hydroxyl group of Hmb will form an ester bond with the amino acids' carboxyl group during each coupling, but it can be removed in the following deFmoc step by piperidine/DMF. In the last coupling, we do not perform the deprotection as we use Boc-Ser(tBu)-OH. Boc can be removed at cleavage step 8 without additional treatment. So, we need use hydrazine to remove the amino acids on Hmb. If the last amino acid is Fmoc protected, we need do the deFmoc step. Then, step 4 can be skipped.

5. Dry the resin by connecting the reaction vessel to vacuum pump and rinse by DCM.
6. Treat resin with 10 mL Ac_2O/Et_3N/DMAP (100 μL/300 μL/5 mg) in DCM and shake for 30 min to cap the hydroxyl group of Hmb.

   Note: The Hmb will be cleaved by the following treatment of concentrated TFA if the hydroxyl group is unprotected.

7. Dry the resin by connecting the reaction vessel to vacuum pump and rinse by DCM.
8. Cleave the peptides from resin by treating resin with 10 mL TFA/H_2O/Tips/Phenol (92.5/2.5/2.5/2.5, v/v/v/v) for 3 h.
9. The resin-cleavage solution mixture is filtered through the vessel filter frit by directly pushing the plunger into a 50 mL centrifugation tube, and the cleavage solution is concentrated under a stream of compressed air to give the crude product.
10. Crude product is further precipitated by adding 40mL cold ether, followed by centrifugation immediately (4000 *g*, 5 min at 25°C).
11. Repeat step 10 twice.
12. The solid is dissolved in 10 mL 10% CH_3CN/H_2O. The solution is filtered by 0.45 μm membrane filter to give a clear solution.
13. Pipette 20 μL of the step 12 solution and dilute it to 200 μL by adding 180 μL 10% ACN/H_2O. The sample is filtered by 0.45 μm membrane filter to give a clear filtrate which is place into a UPLC sample vial.
14. The step 13 sample vial is further place into a UPLC-MS sample chamber (maintaining at 15 °C) at corresponding position. Create a new injection method in the UPLC-MS Masslynx software, the detail parameters are as follow: MS File is 250-3072 8 min, MS Tune File is MS Tune 30 V, and the Inlet File are 15–60% 8 min (15–60% ACN/H_2O), 10–50% 8 min (10–50% ACN/H_2O), and

Figure 4. The sequence of fragment 2
This figure is reprinted with permission from Wei et al., 2021.

10–40% 8 min (10–40% ACN/H_2O). The gradient which can give the best resolution for the sample is determined.

15. After Step 14, the 10–40% 8 min (10–40% ACN/H_2O) gives the best resolution of the sample in the UPLC-MS, each of sample component has shown a symmetric peak in the UPLC-MS UV spectrum.
16. The reverse phase column is connected to a Waters HPLC system. After that, a 10–40% CH_3CN/H_2O gradient over 40min method, according to the step 15, (UV absorption 205 nm, flow rate 20 mL/min) is loaded for 10 mL, and the pressure of the system should be maintained between 115 to 125 Psi.
17. The step 12 solution is injected into the loop of HPLC system, then trigger its sample injector and click the "real time" button of the HPLC software (MassLynx) to give a UV monitoring curve.
18. Collect all the HPLC elution by 15 mL tube based on the UV peak, and 20 μL of sample aliquot, taking from each of the 15 mL tube, are injected into the UPLC-MS.
19. Those tubes with one peak and desired mass spectrum are collected and subjected to liquid nitrogen freezing and lyophilization to give the product as white powder.

Pause Point: The freeze-dried product can be store at −80°C for several months.

HMGA1a fragment 2 preparation

Timing: [2 weeks]

This section describes the preparation of HMGA1a (38–62) peptide fragment along with the N-terminal Fmoic and C-terminal salicylaldehyde ester (Figure 4).

20. Prepare peptide fragment 2 following standard Fmoc-SPPS protocol (Figure 3).
21. After the coupling last Fmoc-Thr(tBu)-OH, the resin obtain from the SPPS is dried by vacuum and rinsed with DCM.
22. Treat resin with 10 mL DCM/AcOH/TFE (8/1/1, v/v/v) for 1 h.
23. The resin is filtered, and the cleavage solution is evaporated under reduce pressure.
24. The resulting oily crude product is further dissolved DCM and hexane, followed by evaporation under vacuum to remove all acetic acid.
25. Repeated co-evaporation for 5 times to give solid powder.

Note: The powder can be directly used for the C-terminal salicylaldehyde ester preparation without HPLC purification.

Pause point: The solid product can be store at −80°C for several months.

26. Dissolved peptide/PyBOP/DIEA (1/3/6, mol/mol/mol) in DCM (final concentration of peptide: 8.5 mM) and stir for 10 min.
27. 20 equivalent 2-(dimethoxymethyl)phenol is added and stir for 12 h.

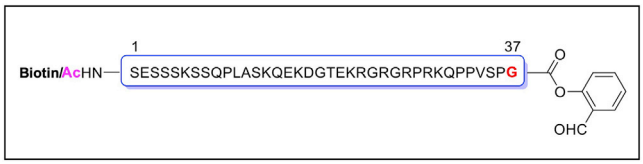


Figure 5. The sequence of fragment 1
This figure is reprinted with permission from Wei et al., 2021.

28. The DCM of the reaction mixture is removed by evaporation to give a colorless oil.
29. Treat the above oil with 10 mL TFA/H_2O/Phenol (95/2.5/2.5, v/v/v) for 1.5 h.
30. The cleavage solution is concentrated under a stream of compressed air.
31. Crude products are precipitated by 40 mL cold ether, followed by centrifugation immediately (4000 *g*, 5 min at 25°C).
32. The solid is dissolved in 10 mL 20% CH_3CN/H_2O. The solution is filtered by 0.45 μm membrane filter and check the ligation product by LC/MS as step 13–14. Then, purify the product by HPLC (20–50% CH_3CN/H_2O over 40 min) as step 16–19.

Pause point: The freeze-dried product can be store at −80°C for several months.

HMGA1a fragment 1 preparation

Timing: [2 weeks]

This section describes the preparation of HMGA1a (1–37) peptide fragment along with the C-terminal salicylaldehyde ester (Figure 5).

33. Prepare peptide fragment 1 following standard Fmoc-SPPS protocol (Figure 3).

34. N-terminus coupling
    a. Install a biotin at N-terminus: after the coupling of last Fmoc-Ser-OH and the deprotection of its Fmoc protecting group, the D-biotin can be directly coupled after Ser1.

Note: The biotin coupling procedure is the same as the amino acid coupling.
    b. Install an acetyl at N-terminus: after the coupling and deprotecting the last Fmoc-Ser-OH, the resin is treated with Ac_2O/ Triethylamine (TEA) (1/1, mol/mol) in 5 mL DCM, and with 5% mol DMAP as the catalyst at 25°C for 2 h.

35. Dry the resin and rinse by DCM.
36. Treat resin with 10 mL DCM/AcOH/TFE (v/v/v, 8/1/1) for 1.5 h.
37. The procedure of installing the salicylaldehyde ester is the same as that of fragment 2 (step 23–31).
38. The solid is dissolved in 10 mL 10% CH_3CN/H_2O. The solution is filtered by 0.45 μm membrane filter to give a clear solution and check the ligation product by LC/MS as step 13–14. Then, purify the product by HPLC (10–40% CH_3CN/H_2O over 40 min) as step 16–19.

Pause point: The freeze-dried product can be store at −80°C for several months.

Ligation between HMGA1a fragment 2 and 3

Timing: [9 h]

This section describes the first round ligation between fragment 2 and 3 (Figure 6A).

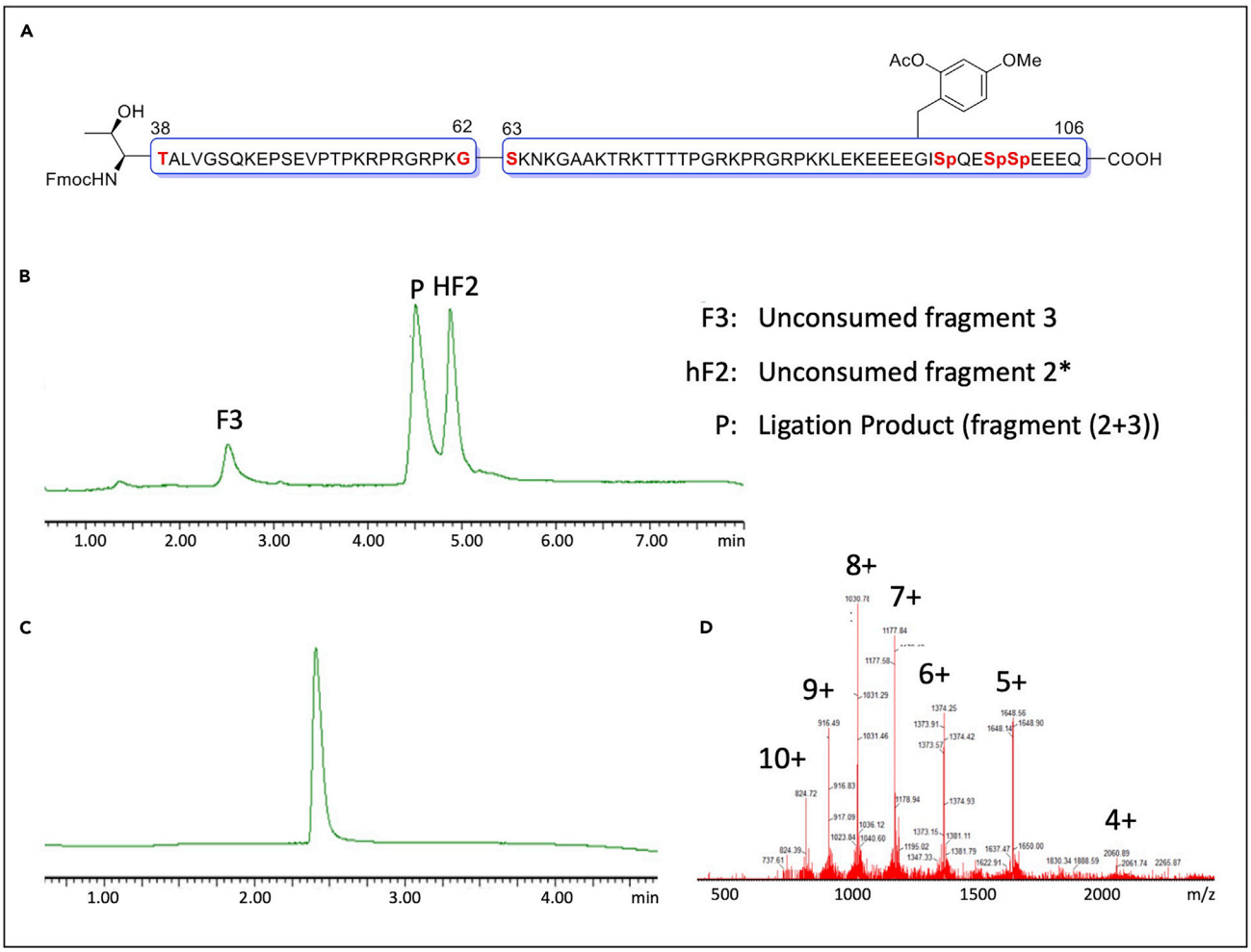


Figure 6. LC/MS analysis of ligation product between fragment 2 and 3

(A) The sequence of ligation product.

(B) The UV trace of the ligation mixture after acidolysis step (the sample was taken from step 44). We can see the ligation product and unconsumed fragment 2 and 3 from the UV trace. * the C-terminal salicylaldehyde ester will be hydrolyzed slowly during the ligation to give the fragment 2 without C-terminal salicylaldehyde ester.

(C) The UV trace of the ligation product after HPLC purification (the sample was token after step 45). The purified product was shown as one peak.

(D) The mass spectra of the peak in UV trace showed the correct mass of product. The molecular weight of desired product ($C_{344}H_{570}N_{105}O_{123}P_3$) is 8237.88 Da. P_1 $[M+4H]^{4+}$ m/z = 2060.48; P_2 $[M+5H]^{5+}$ m/z = 1648.58; P_3 $[M+6H]^{6+}$ m/z = 1373.98; P_4 $[M+7H]^{7+}$ m/z = 1177.84; P_5 $[M+8H]^{8+}$ m/z = 1030.74; P_6 $[M+9H]^{9+}$ m/z = 916.32; P_7 $[M+10H]^{10+}$ m/z = 824.79. Found: 2060.89; 1648.56; 1374.25; 1177.84; 1030.78; 916.49; 824.72.

This figure is reprinted with permission from Wei et al., 2021.

39. Weight 21 mg (1 equivalent) of fragment 3 and 18 mg (1.5 equivalent) of fragment 2 in a 4 mL glass bottle.
40. 400 μL pyridine/AcOH (1/12, mol/mol) is added (the final concentration of fragment 1 is 10 mM). The reaction is stirred for 6 h at 25°C.
41. Add reaction solution into 40 mL cold ether, followed by centrifugation immediately (4000 *g*, 5 min at 25°C).
42. Solid is further washed twice by 40 mL cold ether.
43. Treat solid with 400 μL TFA/H_2O/Tips (95/2.5/2.5, v/v/v) for 30 min.
44. Add reaction solution into 40 mL cold ether, followed by centrifugation immediately (4000 *g*, 5 min at 25°C).

45. The solid is dissolved in 10 mL 5% CH_3CN/H_2O. The solution is filtered by 0.45 μm membrane filter and check the ligation product by LC/MS (Figure 6B) as step 13–14. Then, purify the product by HPLC (5–40% CH_3CN/H_2O over 40 min) as step 16–19 (Figures 6C and 6D).

Pause point: The freeze-dried product can be store at −80°C for several months.

DeAc on Hmb of fragment (2+3)

Timing: [2.5 h]

This section describes the removal of acetyl group on Hmb tag.

46. Dissolve 17 mg (1 equivalent) fragment (2+3) in 420 μL H_2O (final concentration of peptide: 5 mM).
47. Add 3.4 μL hydrazine monohydrate (30 equivalent) into solution and stir for 2 h at 25°C.

DeFmoc on Thr38 fragment (2+3)

Timing: [4 h]

This section describes the removal of Fmoc on Thr38 for next round ligation.

48. Directly add 378 μL of H_2O/CH_3CN (1/1, v/v) + 42 μL DEA into solution and stir for another 2 h at 25°C.
49. The reaction is quenched by adding 8.4 μL TFA and the solution is filtered by 0.45 μm membrane filter and purified by HPLC (5–40% CH_3CN/H_2O over 40 min) as step 16–19. Finally, the product is freeze-dried.

Pause point: The freeze-dried product can be store at −80°C for several months.

Ligation between HMGA1a fragment 1 and (2+3)

Timing: [6 h]

This section describes the second round ligation between fragment 1 and (2+3).

50. Weight 5 mg (1 equivalent) of fragment (2+3) and 4.5 mg (1.5 equivalent) of fragment 1 in a 1 mL glass vial.
51. 100 μL pyridine/AcOH (1/12, mol/mol) is added (the final concentration of fragment 1 is 10 mM). The reaction is stirred for 3 h at 25°C.
52. Add reaction solution into 40 mL cold ether, followed by centrifugation immediately (4000 *g*, 5 min at 25°C).
53. Solid is further washed twice by 40 mL cold ether.
54. Treat solid with 100 μL TFA/H_2O/Tips (95/2.5/2.5, v/v/v) for 30 min.
55. Add reaction solution into 40 mL cold ether, followed by centrifugation immediately (4000 *g*, 5 min at 25°C).
56. The solid is dissolved in 10 mL 5% CH_3CN/H_2O. The solution is filtered by 0.45 μm membrane filter and check the ligation product by LC/MS as step 13–14. Then, purify the product by HPLC (5–40% CH_3CN/H_2O over 40 min) as step 16–19 (Figure 7).

Pause point: The freeze-dried product or stock solution by dissolving product in H_2O at 10 mg/mL can be store at −80°C for several months.

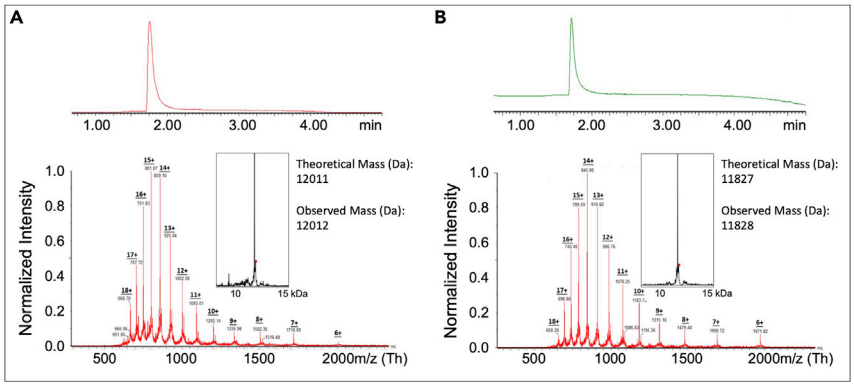


Figure 7. HMGA1a with triple phosphorylation were chemical synthesized

(A) UV trace and mass spectra of Biotin-HMGA1a-3pSer. Up: UV trace of purified product was shown as one peak; Bottom: Mass spectra of the peak in UV trace showed the correct mass of product. The molecular weight of desired product ($C_{492}H_{841}N_{164}O_{177}P_3S$) is 12011.09 Da. P_1 $[M+6H]^{6+}$ m/z = 2002.85; P_2 $[M+7H]^{7+}$ m/z = 1716.87; P_3 $[M+8H]^{8+}$ m/z = 1502.39; P_4[M+9H]$^{9+}$ m/z = 1335.57; P_5 $[M+10H]^{10+}$ m/z = 1202.11; P_6 $[M+11H]^{11+}$ m/z = 1092.92; P_7 $[M+12H]^{12+}$ m/z = 1001.92; P_8[M+13H]$^{13+}$ m/z = 924.93; P_9 $[M+14H]^{14+}$ m/z = 858.94; P_{10} $[M+15H]^{15+}$ m/z = 801.74; P_{11} $[M+16H]^{16+}$ m/z = 750.69; P_{12}[M+17H]$^{17+}$ m/z = 707.54; P_{13} $[M+18H]^{18+}$ m/z = 668.28. Found: 2003.15; 1716.89; 1502.26; 1336.07; 1202.14; 1093.01; 1002.08; 925.04; 859.10; 801.87; 751.75; 707.72; 668.70.

(B) UV trace and mass spectra of HMGA1a-3pSer. Up: UV trace of purified product was shown as one peak; Bottom: Mass spectra of the peak in UV trace showed the correct mass of product. The molecular weight of desired product ($C_{484}H_{829}N_{162}O_{176}P_3$)molecular weight is 11826.84 Da. P_1 $[M+6H]^{6+}$ m/z = 1972.11; P_2 $[M+7H]^{7+}$ m/z = 1690.52; P_3 $[M+8H]^{8+}$m/z = 1479.33.39; P_4 [M+9H]$^{9+}$ m/z = 1315.07; P_5 $[M+10H]^{10+}$ m/z = 1183.67; P_6 $[M+11H]^{11+}$ m/z = 1076.15; P_7 $[M+12H]^{12+}$ m/z = 986.55; P_8 $[M+13H]^{13+}$ m/z = 910.74; P_9 $[M+14H]^{14+}$ m/z = 845.76; P_{10} $[M+15H]^{15+}$ m/z = 789.44; P_{11} $[M+16H]^{16+}$ m/z = 740.17; P_{12} $[M+17H]^{17+}$ m/z = 696.69. Found: 1971.82; 1690.72; 1479.40; 1315.16; 1183.76; 1076.25; 986.76; 910.82; 845.80; 789.59; 740.40; 696.80.The raw data of mass spectrum (txt format) was loaded on UniDec for deconvolution to generate the mass of product (Bin Every: 1.0 Da; Sample Mass Every: 1.0 Da; Peak FWHM: 1.0 Th; Peak Detection Range: 5.0 Da; Peak Detection Threshold: 0.1)

This figure is reprinted with permission from Wei et al., 2021.

Note: The biotin-HMGA1a (wt) can be prepared in the same way.

Pull-down experiment

Timing: [1 day]

This section describes the pull-down experiment by synthetic biotinylated HMGA1a proteins.

57. Prepare 3 plates (10 cm) of cells with 90% confluency.

    Note: As we will detect the P53 using according antibody, we used HEK293T, NB-4, and MCF-7 cell lines. We can use different cell lines for different purposes.

58. Rinse cell by 2 mL pre-warm PBS. Repeat twice.
59. Lyse cells on ice by adding 1 mL cold lysis buffer with protease inhibitor for each plate. Scratch and incubate for 10 min.
60. Collect cell lysate and centrifuge at 13000 *g* for 10 min at 4°C.
61. Carefully collect the supernatant by pipette. Alternatively, the bottom pellet can be resuspended by 100 μL lysis buffer and subject several seconds of sonication and repeat step 60 and combine the cell lysate together.
62. Quantify the concentration of cell lysate by BCA assay. Adjust the concentration to 2 mg/mL.
63. Set up the pull-down assay in 1.5 mL eppendorf tubes as below:

Tube entry	1	2	3
Cell lysate /mg	2	2	2
Biotin-HMGA1a /μg	-	20	-
Biotin-HMGA1a-3pSer /μg	-	-	20

Note: The amounts of biotin-HMGA1a probes can be varied empirically.

64. Incubate the samples for 12 h at 4°C with an overhead rotation.

65. Prepare the streptavidin agarose resin.
    a. Carefully suspend 50 μL resin by 1 mL cold lysis buffer.
    b. Centrifuge at 1000g for 3 min at 4°C.

⚠ CRITICAL:The centrifugal machine should be set to soft acceleration and deacceleration from this step.

    c. Carefully discard the supernatant by pipette.
    d. Repeat step a-c twice.
    e. After washing, the resin is finally suspended in 50 μL cold lysis buffer.

66. Apply 15 μL pre-equilibrated resin slurry to each pull-down tube and incubate for another 4 h at 4°C with overhead rotation.
67. Centrifuge at 1000g for 3 min at 4°C.
68. Carefully discard the supernatant by pipette.
69. Wash the resin twice carefully as step 65a-c.

Note: The pull-down samples are ready for LC-MS/MS sample preparation to identify the new binders which specifically interact with HMGA1a or HMGA1a-3pSer. This protocol did not de-scripted it. More detail can be found at Wei et al., 2021.

70. Add 50 μL 1 × loading buffer and boil sample at 100°C for 10 min.

Pause point: The boiled samples can be store at −80°C for several months.

71. SDS-PAGE/western blot analysis (Figure 8).

Note: Spin down the resin and loading supernatant samples on SDS-PAGE.

Note: Biotin-HMGA1a protein will appear at ~16 kDa on SDS-PAGE as it contains many pos-itive charged residues.

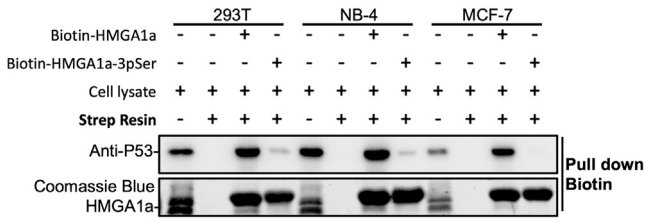


Figure 8. Pull-down P53 from indicated cell lysates by biotin-HMGA1a-3pSer and biotin-HMGA1a
P53 was probed by anti-P53 and HMGA1a proteins were visualized by Coomassie blue staining. This figure is reprinted with permission from Wei et al., 2021.

Note: Mono-streptavidin protein can be released from resin after boiling and appear at ~12 kDa on SDS-PAGE.

Note: Pull-down of binders is one of applications of synthetic proteins. Besides, the synthetic proteins can be used in surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC) to quantify the binding affinity.

EXPECTED OUTCOMES

Chemical protein synthesis could serve as a good tool to study the PTMs of interest. This protocol generalizes the procedure of chemical synthesis of HMGA1a by Ser/Thr ligation, which opens up more possibilities to identify other unknown function of PTMs on HMGA1a.

A successful reproduction of the synthesis should have relatively high purity (Figure 7) and good yield in milligram scale. Biotin-streptavidin mediated pull-down strategy has been the most common way to identify new binders from cell lysate. In our case, we could see P53 always bind HMGA1a instead of the triple phosphorylated variant in the chosen cell lines (Figure 8). This result proved that the triple phosphorylation could serve as regulatory switch to control the interaction between HMGA1a and P53.

LIMITATIONS

Chemical protein synthesis provides an effective solution to introduce PTMs into proteins site-selectively. However, there are some gaps still awaiting to be filled. 1). Proteins with more than 400 residues are still not routinely reachable by chemical protein synthesis, since it requires many rounds of ligation. Finally, the total yield will be extremely low. In this case, a protein semi-synthesis approach can be applied, in which the synthetic parts with the desired PTMs and the recombinant parts can be ligated together. This approach significantly extends our capacity to generate the big protein in a cost-effective manner. More detailed information about this strategy can be found at Thompson and Muir, 2019. 2). The protein after chemical synthesis is a denatured form, which needs a refolding step to generate the functional protein. HMGA1a is an intrinsic disordered protein and does not require refolding operation. 3). Membrane proteins or hydrophobic sequence are currently very difficult to be synthesized because of the aggregation in SPPS and low solubility in ligation solvent. Try to change the ligation sites to adjust the hydrophobicity of peptide fragment or introduce solubilizing tags. 4). Delivery of a synthetic protein into cell is still challenging, which limits the application in cell for *in vivo* studies.

TROUBLESHOOTING

Problem 1

Peptides are not pure after step 3.

Potential solution

In most cases, the truncation species are the main impurities. Apply the Kaiser test to monitor the coupling if needed. Double or triple coupling at potential troublesome sites (e.g., Arg, Glu) is a good way to solve this problem: after finishing the first-round coupling, wash resin and recouple the same amino acid for one more time.

Problem 2

Hydrolysis was observed after step 32 and 38.

Potential solution

After the successful preparation of the peptide salicylaldehyde ester, HPLC purification should be performed immediately, as the salicylaldehyde ester is more stable in powder form.

Problem 3
Peptide fragments are not soluble in the ligation buffer in step 40 and 51.

Potential solution
Low solubility always causes low ligation efficiency. All HMGA1a fragments have good solubility in pyridine/AcOH (1/12, mol/mol) buffer. However, there may be some other peptides are not soluble. In these cases, we can vary the ratio of pyridine and AcOH (e.g., 1:1, 1:3, 1:9), or dissolve peptides in a small volume of Dimethylformamide (DMF) or Hexafluoro-2-propanol (HFIP) firstly, then dilute it with pyridine/AcOH buffer to the desired concentration. Normally, these cosolvents do not negatively affect the ligation efficiency.

Problem 4
Ligation efficiency is low when checking the ligation mixture at step 45 and 56.

Potential solution
If the peptide fragments are not soluble well in ligation buffer, please refer problem 3 and try to dissolve the peptide fragment firstly.

If the peptide fragments are soluble, please further check the C-terminal amino acid of peptide ester part. Some amino acids will significantly hinder the Ser/Thr ligation because of their bulky sidechain (Figure 1B). If so, re-analyze the protein sequence and change the ligation sites.

Problem 5
Fail to detect the interested protein after pull-down experiment through western blot in step 71.

Potential solution
In our case, we detected P53 in the chosen cell lines (Figure 8). However, we also failed to identify P53 in some cell lines (e.g., Hela and H1299), because these two cell lines barely express P53 without external stimuli. Therefore, it is always good to include the cell lysate input sample as a positive control for western blot detection.

If the expression level of interested proteins is normal or high, we need check the amount of probes, incubation time, binding buffer. Normally, more probes, longer incubation time and less salts in binding buffer will result in higher signal in western blot.

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xuechen Li (xuechenl@hku.hk).

Materials availability
Requests for reagents may be sent to Xuechen Li (xuechenl@hku.hk).

Data and code availability
No data and code were developed in this study.

ACKNOWLEDGMENTS
This work was supported by the Research Grants Council (17309616 and C7017-18G) of Hong Kong, the National Science Foundation of China (21672180 and 91753101), and the Area of Excellence Scheme of the University Grants Committee of Hong Kong (AoE/P-705/16).

AUTHOR CONTRIBUTIONS
X.L., T.W., H.L., H.W., and F.P. established this protocol. X.L, T.W., and H.W. wrote this manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Wei, T., Liu, H., Chu, B., Blasco, P., Liu, Z., Tian, R., Li, D.X., and Li, X. (2021). Phosphorylation-regulated HMGA1a-P53 interaction unveils the function of HMGA1a acidic tail phosphorylations via synthetic proteins. Cell Chem. Biol. *28*, 722–732.

Tan, Y., Wu, H., Wei, T., and Li, X. (2020). Chemical protein synthesis: advances, challenges, and outlooks. J. Am. Chem. Soc. *142*, 20288–20298.

Zhang, Y., Xu, C., Lam, H.Y., Lee, C.L., and Li, X. (2013). Protein chemical synthesis by serine and threonine ligation. Proc. Natl. Acad. Sci. U S A *110*, 6657.

Reeves, R. (2001). Molecular biology of HMGA proteins: hubs of nuclear function. Gene *277*, 63–81.

Zhang, Q., and Wang, Y. (2008). High mobility group proteins and their post-translational modifications. Biochim. Biophys. Acta *1784*, 1159–1166.

Liu, H., and Li, X. (2018). Serine/threonine ligation: origin, mechanistic aspects, and applications. Acc. Chem. Res. *51*, 1643–1655.

Thompson, R.E., and Muir, T.W. (2019). Chemoenzymatic semisynthesis of proteins. Chem. Rev. *120*, 3051–3126.