Coupling functionalized cobalt ferrite nanoparticle enrichment with online LC/MS/MS for top-down phosphoproteomics†

Bifan Chen,‡a Leekyoung Hwang,a William Ochowicz,a Ziqing Lin,b Tania M. Guardado-Alvarez,a Wenxuan Cai,a Lichen Xiu,a Kunal Dani,a Cyrus Colah,b Song Jin a and Ying Ge a,b,c

Phosphorylation plays pivotal roles in cellular processes and dysregulated phosphorylation is considered as an underlying mechanism in many human diseases. Top-down mass spectrometry (MS) analyzes intact proteins and provides a comprehensive analysis of protein phosphorylation. However, top-down MS-based phosphoproteomics is challenging due to the difficulty in enriching low abundance intact phosphoproteins as well as separating and detecting the enriched phosphoproteins from complex mixtures. Herein, we have designed and synthesized the next generation functionalized superparamagnetic cobalt ferrite (CoFe2O4) nanoparticles (NPs), and have further developed a top-down phosphoproteomics strategy coupling phosphoprotein enrichment enabled by the functionalized CoFe2O4 NPs with online liquid chromatography (LC)/MS/MS for comprehensive characterization of phosphoproteins. We have demonstrated the highly specific enrichment of a minimal amount of spike-in β-casein from a complex tissue lysate as well as effective separation and quantification of its phosphorylated genetic variants. More importantly, this integrated top-down phosphoproteomics strategy allows for enrichment, identification, quantification, and comprehensive characterization of low abundance endogenous phosphoproteins from complex tissue extracts on a chromatographic time scale.

Introduction

Reversible protein phosphorylation orchestrates important cellular processes in eukaryotic cells including cell proliferation, protein–protein interaction, signal propagation, among others.3,4 Dysregulation of protein phosphorylation is known to be critically involved in the pathogenesis of many human diseases, such as cancer, heart failure, and neurodegenerative disease.5-7 To understand these disease processes, identification and quantification of phosphoproteins are the essential steps in decoding the complex cellular signaling network regulated by phosphorylation events. Moreover, proteoforms arising from alternative splicing, amino acid polymorphisms, and other post-translational modifications (PTMs),8 have imposed another layer of complexity to the phosphoproteome. The conventional shotgun “bottom-up” mass-spectrometry (MS) based phosphoproteomics which analyzes peptides resulted from protein digestions,9,10 is suboptimal because of the “protein inference problem”, resulting in incomplete information on sequences or modifications.11,12 In contrast, the top-down MS approach examines intact proteins without digestion, providing a global view and the possible merit of comprehensive analysis of all modified proteoforms.13-16 Therefore, top-down MS allows one to study phosphoproteins and their associated proteoforms at the intact protein level, which leads to more direct connectivity to complex disease phenotypes for clinical and translational research.11

However, unlike the well-established bottom-up phosphoproteomics that has numerous mature phosphopeptide enrichment methods and well-developed LC/MS/MS workflows,17-19 the top-down phosphoproteomics approach is much more challenging mainly due to the lack of phosphoprotein enrichment methods and the difficulties in separation of intact proteins from complex proteome. To address the limitation of low specificity, efficiency, and poor reproducibility on intact protein enrichment,20-23 we recently developed the first generation functionalized superparamagnetic Fe3O4–GAPT-Zn nanoparticles (NPs) for universal and effective enrichment of phosphoproteins from complex mixtures.24 However, further improvement on the NPs and effective downstream analysis including LC separation and MS/MS characterization of
enriched phosphoproteins are needed to fully interrogate the complex phosphoproteome.

Herein, we sought to develop a robust top-down proteomics workflow for comprehensive phosphoprotein characterization that integrates highly specific intact phosphoprotein enrichment by the next generation cobalt ferrite (CoFe₂O₄) NPs with high resolution online LC/MS/MS analysis (Fig. 1). In this top-down phosphoproteomics workflow, proteins are first extracted from tissue homogenate (or cell lysate), and subsequently incubated with functionalized superparamagnetic NPs (Fig. 1a). At a neutral pH, the ligands functionalized on the NP surface strongly bind to phosphate groups on phosphoproteins and are the key to the enrichment. After a magnetic pull down of the phosphoprotein-bound NPs, elution of the phosphoproteins from NPs (as well as the pre-enrichment samples) are then subjected to online LC separation and MS analysis of the intact proteins. The enriched phosphorylated species are selected for tandem MS for identification and characterization (Fig. 1c). For the first time, we have shown that this integrated top-down phosphoproteomics strategy allows for highly specific enrichment, identification, quantification, and comprehensive characterization of low abundance endogenous phosphoproteins from complex tissue extracts on a chromatographic time scale.

**Results and discussion**

**Synthesis, characterization, and evaluation of CoFe₂O₄ NPs for intact phosphoprotein enrichment**

We have designed and synthesized the cobalt ferrite (CoFe₂O₄) NPs (Fig. 1d and 2a) with stronger magnetic response and better reproducibility for intact phosphoprotein enrichment than the first generation Fe₃O₄–GAPT-Zn NPs. We synthesized CoFe₂O₄ NPs capped with oleic acid (OA) and oleylamine (OE) ligands by a seed-mediated growth method (see experimental details in ESI and Fig. S1†) following the procedure reported by Sun et al. Transmission electron microscopy (TEM) image analysis revealed that the diameter of the CoFe₂O₄ NPs was 9.96 ± 1.03 nm (Fig. 2b). Powder X-ray diffraction confirmed a typical ferrite diffraction pattern (Fig. 2c). The CoFe₂O₄ NPs have been shown to possess larger magnetic moments and magnetic anisotropy and good chemical stability. These NPs were subsequently functionalized with dinuclear Zn(II)-dipicolylamine (Zn-DPA) ligands coupled to glutaric acid (hereafter referred to as GAPT), based on the previously reported method with minor modifications (Fig. 2a). These GAPT-Zn complexes provide the specific binding to the phosphate ions on phosphoproteins and are the key to the enrichment. Notably, the step that replaces the hydrophobic OA and OE ligands on the CoFe₂O₄ NP surface with 3-aminopropyl trimethoxysilane (APMTMS) and 2-methoxy (polyethyleneoxy) propyl trimethoxysilane (hereafter referred to as Si-PEG) was completed within 24 h, which was greatly improved compared to 48 h for the preparation of Fe₃O₄–NH₃/PEG NPs. We then coupled GAPT ligands, the phosphate capturing moiety, to the free amino groups of the APMTMS on CoFe₂O₄–NH₃/PEG NPs, producing CoFe₂O₄–GAPT NPs. Fourier
transform infrared spectroscopy (FTIR) indicating strong band around 1147–1028 cm⁻¹, peaks around 1200 cm⁻¹ (Si–O–R, C–O–C, EO CH₂), and sharp peak around 1650 cm⁻¹ for CoFe₂O₄–GAPT (amide bond), and thus confirmed the proper functionalization of the NPs with the ligand molecules (Fig. 2d). Additionally, thermogravimetric analysis (TGA) showed different weight loss (%) for each surface-modified CoFe₂O₄ NPs with the corresponding ligand molecules (Fig. 2e). The small size of the CoFe₂O₄ NPs provides advantages such as a high surface-to-volume ratio, easy surface functionalization with multivalent ligand molecules to interact with phosphoprotein targets; better penetrations in complex mixture for higher binding rate without causing denaturation; and good solubility. † After activation of CoFe₂O₄–GAPT NPs with 10 mM ZnCl₂, the resulting CoFe₂O₄–GAPT-Zn NPs (Fig. 1d) were used for phosphoprotein enrichment. The protein samples prior to NP enrichment were termed loading mixture (LM); unbound non-phosphoproteins were collected as flow-through (FT); and enriched phosphoproteins were saved as elution (E) for further analysis (Fig. 1c). The optimized enrichment procedure includes a HEPES buffer (50 mM, pH 7.7, 150 mM NaCl) for binding and subsequent washing, and a sodium phosphate (Na₂HPO₄) buffer (100 mM, pH 7.3, 50 mM NaCl) for final elution. A standard protein mixture was used to validate the enrichment efficiency of each batch of NPs (details in ESI†). We further sought to quantitatively determine the enrichment performance by spiked in low levels of β-casein as an internal control in a highly complex swine heart tissue extract. We systematically decreased the concentration of the spike-in β-casein from 0.1 to 0.03 μg μL⁻¹, 10 μg of equal amount of proteins before and after enrichment of the tissue extract containing about 3% (I), 5% (II), 7% (III), and 10% (IV) w/w of β-casein (Table S1†) was prepared as loading mixtures. SDS-PAGE analysis was performed (Fig. 3a and b, and S3†) paralleled to the top-down LC/MS analysis (vide infra).

Overall, the enrichment was consistent and reproducible across all samples with different amounts of spike-in β-casein.

Abundant non-phosphoproteins in LM were effectively depleted and washed into FT, as shown in the nearly identical bands between LM and FT (Fig. 3a). In contrast to LM and FT, the highly similar band patterns in the elution (E) lanes of the gel stained by Sypro Ruby and ProQ Diamond (Fig. 3a and b) indicates the predominance of enriched endogenous phosphoproteins in addition to β-casein. These results demonstrated that CoFe₂O₄–GAPT-Zn NPs can efficiently enrich phosphoproteins from complex biological samples with high specificity and sensitivity.

Top-down LC/MS analysis of spike-in β-casein from tissue extract after CoFe₂O₄ NPs enrichment

For the top-down proteomic analysis after NP enrichment, an equal amount of proteins from LM, FT, and E that were acquired from the tissue extract with 5% of β-casein spike-in was subjected to reverse phase chromatography (RPC) LC/MS and LC/MS/MS analyses using a Waters M-class UPLC and a Bruker maXis II Q-TOF mass spectrometer (details in ESI†). The robust LC/MS platform allows direct comparison across the base peak chromatograms (BPC) of LM, FT, and E (Fig. 4a). The overlapping separation pattern between LM and FT (Fig. 4a and S4†) demonstrates that the majority of the proteins from LM were washed off to FT, consistent with the SDS-PAGE band patterns (Fig. S3†). Subsequent MS analysis revealed that most proteins in the LM and FT solutions are highly abundant non-phosphoproteins, including blood proteins such as hemoglobin subunit α and β, which were significantly depleted after enrichment (Fig. S5 and S6†). Even though some blood proteins were still detected in MS from elution, significant depletion of highly abundant non-phosphoproteins in LM enabled us to detect low abundance phosphoproteins in E, which are reflected as low abundance peaks in the BPC for E. For example, the deconvoluted spectrum shows that the spike-in β-casein was barely detectable in LM and FT solutions because of the predominant signal from the co-eluting myoglobin (Fig. 4b). However, after enrichment using CoFe₂O₄–GAPT-Zn NPs, β-casein (M, 23 968.12) was significantly enriched while the intensity of myoglobin (M, 16 942.86) was greatly decreased (Fig. 4b). The enrichment factor was estimated by comparing the relative percentage of β-casein before and after the enrichment. The relative percentage of β-casein was significantly increased from 0.5% to 94% from the complex mixture after enrichment. Moreover, the pure spike-in β-casein standard were separated into three distinct LC peaks in E after enrichment, as seen in the BPC and extracted ion chromatogram labeled as (i), (ii), and (iii) in Fig. 4a. None of these were detectable prior to enrichment in LM. The deconvoluted mass spectra revealed three genetic variants of β-casein, A4 B (M, 24 077.14); (i), A1 (M, 24 008.11); (ii), and A2 (M, 23 968.12); (iii) based on their accurate mass measurement (Fig. 4c and S5†). Although the genetic variants of β-casein differ by only one or two amino acid residues, all three proteoforms were unambiguously enriched by the NPs and separated from the complex tissue extract using top-down LC/MS. Furthermore, the top-down approach allowed us to quantify the relative percentage of each β-casein proteoform,
which would be difficult using the bottom-up MS strategy. The relative percentages among A2, A1, and B variants were determined to be 53%, 35%, and 12% respectively, according to the average spectrum across the same time frame (Fig. 5†).

**Top-down LC/MS/MS characterization of endogenous phosphoproteins enriched from tissue extract by CoFe$_2$O$_4$ NPs**

Furthermore, through this integrated top-down phosphoproteomics strategy, we have detected and examined multiple known cardiac phosphoproteins from the eluted phosphoproteins that were previously undetected or in very low abundance in the HEPES extraction without enrichment (Fig. 5). The enrichment of the phosphoproteins is apparent by comparing phosphorylation level changes across the same time frame in the LC/MS of LM, FT, and E. Notably, as shown in the normalized deconvoluted spectra in Fig. 5a, phosphorylated forms of pig cardiac troponin I (24 kDa) with a possible V116A polymorphism (28 Da) were all enriched, which was consistent with what we previously found on the same protein using immunoaffinity purification. Western blot analysis for phosphorylated cardiac troponin I at Ser22/23 further demonstrated the enrichment (Fig. S9†), which well correlates with the MS data. Nevertheless, the top-down approach in this case offers a global view of each of these highly similar phosphorylated polymorphic species. Similarly, phosphorylated tropomyosin (33 kDa) was also enriched and can be characterized after enrichment (Fig. 5b).

More importantly, in addition to the previously known myofilament proteins, we were able to discover low abundance endogenous phosphoproteins using this integrated strategy. One example is a potential phosphoprotein (27 kDa) eluting around 33.5 min that was significantly enriched so that the triply phosphorylated form became the most intense peak/base peak (Fig. 6a). The deconvoluted spectrum further shows that the protein with $M_r$ 26 996.90 has 80 Da mass increments, corresponding to multiple phosphorylated forms of the protein (Fig. 6b). As seen in both the original and the deconvoluted spectra, the triply phosphorylated form was of very low abundance in LM of the heart tissue extract, but after enrichment, its intensity improved more than 10 folds (Fig. 6a and b). Moreover, nearly no detection of the phosphorylated forms in FT spectra demonstrates sufficient capturing ability of the CoFe$_2$O$_4$-GAPT NPs. Since the triply phosphorylated form became the base peak after enrichment, subsequent LC/MS/MS was made possible through both collision-induced dissociation (CID) and electron transfer dissociation (ETD).

We confidently identified this protein as hepatoma-derived growth factor (UniProtKB/Swiss-Prot, F1RJH2_PIG) using MS-Align+ search algorithm (Fig. S10†), which was previously uncharacterized and unreviewed. We further analyzed the data and mapped the sequence using MASH suite Pro. Three potential phosphorylation sites were determined: Ser132, Ser133, and Ser165 (red letters in Fig. 6c and S11†).
MS/MS not only revealed an N-terminal methionine excision, but also a coexisting acetylation near the N-terminal in addition to the three phosphorylations. As another example, a protein with $M_r$ 21,675.63 with a mass increase of 80 Da was significantly enriched and detected in E (Fig. S12†), while it was nearly unobservable in LM and FT. In the normalized mass spectra in Fig. S12a†, the non-phosphoprotein with $M_r$ 20,827.56 was fully depleted, and the phosphorylated protein $M_r$ 21,675.63 + 80 Da was enriched more than 10 folds, demonstrating the specificity of this NP-based enrichment and the efficacy of this integrated strategy. MS-Align + search identified the proteins to be chromobox protein homolog 1 (UniProtKB/Swiss-Prot, F1RWH1_PIG) based on both the CID and ETD spectra acquired from LC/MS/MS and we narrowed down the possible phosphorylation site to Ser89/Ser91 (Fig. S12 and S13†). Similarly, we enriched and identified a 14 kDa singly phosphorylated protein, programmed cell death protein 5 (UniProtKB/Swiss-Prot, F1RNX2_PIG) (Fig. S14 and S15†). Interestingly, even though phosphorylation is known as a labile PTM, many of the phosphorylations were preserved during CID response and much improved robustness and reproducibility in synthetic steps. The high specificity and efficiency of phosphoprotein enrichment using these improved CoFe$_2$O$_4$–GAPT-Zn NPs was validated by SDS-PAGE analysis with phosphoprotein-specific gel stains using β-casein spike-in tissue extract. Top-down LC/MS results demonstrated that the spike-in phosphorylated β-casein was substantially enriched (relative percent from 0.5% to 94%), which allowed us to further separate, identify, and quantify all of its genetic variants. More importantly, through online top-down LC/MS/MS with CID and/or ETD, we were able to identify endogenous phosphoproteins and localize their potential phosphorylation sites. Benefiting from the unique NP-based enrichment method, this integrated online top-down LC/MS/MS workflow has the advantages of globally observing phosphorylated proteoforms; quantifying proteoform species; and identifying and characterizing phosphoproteins with co-existing PTMs and/or sequence variants. We envision that with further advances in front-end intact protein separation, MS detection of large proteins (>30 kDa), and MS/MS fragmentation techniques with improved throughput, this top-down phosphoproteomics strategy that couples functionalized NP-based phosphoprotein enrichment with online LC/MS/MS has the potential to enable a deep coverage and comprehensive characterization of the phosphoproteome for the study of molecular mechanism underlying diseases and the discovery of specific phosphoproteoforms as biomarkers.
References

1 T. Hunter, Cell, 2000, 100, 113–127.
2 S. J. Humphrey, S. B. Azimifar and M. Mann, Nat. Biotechnol., 2015, 33, 990–995.
3 S. G. Julien, N. Dubé, S. Hardy and M. L. Tremblay, Nat. Rev. Cancer, 2011, 11, 35–49.
4 R. E. Banks, M. J. Dunn, D. F. Hochstrasser, J. C. Sanchez, W. Blackstock, D. J. Pappin and P. J. Selby, Lancet, 2000, 356, 1749–1756.
5 Z. R. Gregorich and Y. Ge, Proteomics, 2014, 14, 1195–1210.
6 J. Zhang, M. J. Guy, H. S. Norman, Y. C. Chen, Q. G. Xu, X. T. Dong, H. Guner, S. J. Wang, T. Kohmoto, K. H. Young, R. L. Moss and Y. Ge, J. Proteome Res., 2011, 10, 4054–4065.
7 Y. Peng, Z. R. Gregorich, S. G. Valeja, H. Zhang, W. Cai, Y. C. Chen, H. Guner, A. J. Chen, D. J. Schwaehn, T. A. Hacker, X. Liu and Y. Ge, Mol. Cell. Proteomics, 2014, 13, 2752–2764.
8 L. M. Smith and N. L. Kelleher, Nat. Methods, 2013, 10, 186–187.
9 K. Sharma, R. C. D’Souza, S. Tyanova, C. Schaab, J. R. Wiśniewski, J. Cox and M. Mann, Cell Rep., 2014, 8, 1583–1594.
10 P. H. Huang and M. F. White, Mol. Cell, 2008, 31, 777–781.
11 N. L. Kelleher, P. M. Thomas, I. Ntai, P. D. Compton and R. D. LeDuc, Expert Rev. Proteomics, 2014, 11, 649–651.
12 B. T. Chait, Science, 2006, 314, 65–66.
13 J. C. Tran, L. Zamdborg, D. R. Ahlf, J. E. Lee, A. D. Catherman, K. R. Durbin, J. D. Tipton, A. Vellaichamy, J. F. Kellie and M. Li, Nature, 2011, 480, 254–258.
14 V. Zabrouskov, Y. Ge, J. Schwartz and J. W. Walker, Mol. Cell. Proteomics, 2008, 7, 1838–1849.
15 W. Cai, T. M. Tucholski, Z. R. Gregorich and Y. Ge, Expert Rev. Proteomics, 2016, 13, 717–730.
16 N. Siuti and N. L. Kelleher, Nat. Methods, 2007, 4, 817–821.
17 N. M. Riley and J. J. Coon, Anal. Chem., 2016, 88, 74–94.
18 L. Xue, W. H. Wang, A. Iliuk, L. Hu, J. A. Galan, S. Yu, M. Hans, R. L. Gehlen and W. A. Tao, Proc. Natl. Acad. Sci. U. S. A., 2012, 109, 5615–5620.
19 B. Bodenmiller, L. N. Mueller, M. Mueller, B. Domon and R. Aebersold, Nat. Methods, 2007, 4, 231–237.
20 Y. Oda, T. Nagasu and B. T. Chait, Nat. Biotechnol., 2001, 19, 379–382.
21 J. Porath, J. Carlsson, I. Olsson and G. Belfrage, Nature, 1975, 238, 598–599.