Electric Supplementary Information

Engineering Magnetic Guanidyl-Functionalized Supramolecular Organic Framework for Efficient Enrichment of Global Phosphopeptides

Haijiao Zheng[a], Jingchun Zhang[b], Jiutong Ma[a], and Qiong Jia[a]*

[a] College of Chemistry, Jilin University, Changchun 130012, China

[b] China-Japan Union Hospital, Jilin University, Changchun 130033, China

* Corresponding author: Professor Qiong JIA

Email address: jiaqiong@jlu.edu.cn

CONTENTS:

Reagents and Materials

Arginine (Arg), tetraethyl orthosilicate (TEOS), 3-aminopropyltrimethoxysilane (APTES), trifluorome, 2,5-dihydroxybenzoic acid (DHB), triflfluoroacetic acid (TFA), formic acid (FA), phosphoric acid (H_3PO_4), and anhydrous pyridine were purchased from Aladdin Reagents (Shanghai, China). Iron trichloride ($FeCl_3 \cdot 6H_2O$), ethylene glycol, paraformaldehyde, boron tribromide dichloromethane, 1,4-dimethoxybenzene, boron trifluoride etherate, phosphatic buffer saline (PBS), dichloromethane, trichloromethane, acetonitrile (ACN), sodium chloride (NaCl), and ethanol were obtained from Beijing Chemical Works (Beijing, China). *β*-casein, cytochrome C (Cyt C), and ovalbumin (OVA) were bought from Sigma-Aldrich (St. Louis, MO, USA). Defatted milk and egg yolk samples were purchased from local supermarket. Human saliva samples were obtained from China-Japan Hospital of Jilin University (Changchun, China) according to standard clinical procedures.

Synthesis of mP5SOF-Arg

Fe_3O_4 was employed as the magnetic core and prepared with a solvothermal reaction. Firstly, $FeCl_3 \cdot 6H_2O$ (2.7 g) was dissolved in ethylene glycol (50 mL). Under vigorous stirring for 30 min, the solution changed from colorless to orange. NaAc (7.2 g) and PEG-2000 (2.0 g) were added into the solution subsequently with another 1 h stirring. Then, the brown and sticky solution was transferred and sealed in a Teflon-lined stainless-steel autoclave. After heating at 200 ^oC for 8 h, the black precipitate was cooled to room temperature and washed successively with ethyl alcohol (EtOH) and ultrapure water several times until neutral. The as-obtained Fe_3O_4 powder was dried at 60 ^oC in vacuo for the following steps. Fe_3O_4-SiO_2 (denoted as $mSiO_2$) was synthesized with a sol-gel method as followings. Fe_3O_4 powder (1.0 g) was treated with 0.1 mol L^{-1} HCl for 10 min. Then, the particles were added into a mixture containing EtOH (80 mL), ultrapure water (20 mL), and $NH_3 \cdot H_2O$ (2.5 mL). After the mixture was stirred to be uniform, tetraethyl orthosilicate (0.5 mL) was added into the dispersion quickly and the mixture was stirred for 12 h at room temperature to yield $mSiO_2$. $mSiO_2$ was dispersed in methylbenzene (80 mL), after which 2,3-epoxy propoxy propyltrimethoxysilicane (GLYMO, 2 mL) was added dropwise into the solution. Subsequently, the suspension was refluxed at

80 °C for 12 h to obtain $mSiO_2$-GLYMO (denoted as mGLYMO).

Amino-functionalized P5SOF (denoted as P5SOF-NH_2) was successfully prepared by standard Gabriel reaction according to a previous report with some modifications. Different concentrations of P5SOF-NH_2 solution (610 mg, or 900 mg, or 1.25 g in 400 mL acetone) was added into mGLYMO solution (780 mg in 200 mL nitro methane) and then kept at room temperature for 24 h. The formed crystals were collected by filtration and dried in a vacuum oven at 60 °C overnight to afford mP5SOF-NH_2. Arg modified mP5SOF (denoted as mP5SOF-Arg) was synthesized by an amidation reaction. Firstly, 100 mg Arg was activated for 30 min in 20 mL morpholine ethyl sulfonic acid buffer (50 mmol L^{-1}) containing 200 mg carbodiimide and *N*-hydroxysuccinimide at pH 5.5. Then, 100 mg mP5SOF-NH_2 was dispersed in the obtained mixture, followed by stirring at room temperature for 12 h. The product mP5SOF-Arg was filtered and washed successively with water and used for the following magnetic solid-phase extraction procedures. The whole synthesis process is illustrated in **Scheme 1a**.

Sample Preparation.

β-Casein, ovalbumin (OVA), bull serum albumin (BSA), canna protein A (Con A), or cytochrome C (Cyt C) (1.0 mg) was dissolved in phosphate buffer solution (1 mL, pH 6.86) to form a substrate solution and digested at 37 °C for 16 h with trypsin at the ratio of enzyme-to-substrate of 1:40 (*w*/*w*). Defatted milk and egg yolk samples were diluted with phosphate buffer solution (pH 6.86) to a final concentration of 100 μg mL^{-1} and centrifuged at the speed of 12000 rpm for 20 min. The obtained supernatant was denatured and digested according to the same procedure as *β*-casein described above. For human saliva, 2 mL 0.2% trifluoroacetic acid was added into 2 mL human saliva in ice bath condition, and then the mixture was centrifuged at the speed of 8000 rpm for 10 min. Finally, the supernatant was collected and preserved at −20 °C for further use. A549 cells were stimulated by incubation with 1 mL per 1×10^7 cells of a lysis buffer containing a cocktail of protease and phosphatase inhibitors for 45 min on ice. The vials were vortexed and pipetted up and down every 10 min. The resulting cell lysates were centrifuged, then the supernatant containing soluble proteins was

separated and used for subsequent experiments. All above aliquots of proteolytic digests were frozen at −80 °C for standby applications.

Phosphopeptides Enrichment

Tryptic digests of *β*-casein or peptides mixtures were dissolved in 100 *μ*L loading buffer **1** (50% acetonitrile-water containing 0.1 mol L^{-1} acetic acid), then 20 *μ*L mP5SOF-Arg suspension (10 mg mL^{-1}) was added into the above protein solution and vibrated at room temperature for 30 min. Afterwards, the phosphopeptide-captured material was washed three times with 100 *μ*L loading buffer **1** and separated by an external magnet subsequently. In the elution step, mP5SOF-Arg was dissolved in loading buffer **2** (50% acetonitrile-water containing 2% trifluoroacetic acid) and the phosphopeptides were released directly from the materials for the subsequent MALDI-MS analysis.

Characterization

Scanning electron micrographic (SEM) images were obtained using a JEOL-JSM-6360LV scanning electron microscope (JEOL, Tokyo, Japan). Transmission electron microscopic (TEM) measurements were carried out on a Hitachi H600 instrument (Hitachi, Japan) with the accelerating voltage of 100 kV. Fourier-transform infrared (FT-IR) spectra were determined on a Thermo Nicolet 670 FT-IR instrument (Thermo Nicolet Corporation, USA) operating at the wavelength range of 4000–400 cm^{-1}. An X-ray photoelectron spectrometer (XPS, Thermo Electron, USA) was used to obtain XPS data. XRD measurements were conducted on an X-ray diffractometer (Empyrean, PANalytical B.V., Netherland) with the detector operating with the voltage of 40.0 kV and the recorded range of 2*θ* was 10–80°. Thermogravimetric analysis (TG) analyses were performed on a Q500 thermal gravimetric analyzer (TA, USA). Magnetic properties were determined by a vibrating sample magnetometer (VSM) on a Superconducting Quantum Interference Device (SQUID XL-7, Quantum Design, USA).

MS Analysis

MALDI-MS Analysis. 0.5 μL eluent was mixed with 0.5 *μ*L DHB (25 mg mL^{-1} in 70% ACN-H_2O containing 1% H_3PO_4) on the MALDI plate and dried at 25 °C. MALDI-MS instrument (Model 5800,

Applied Biosystems, Foster City, CA) was performed in the reflection positive-ion mode with the *m*/*z* scan range of 1000–3500 and a laser pulse rate of 200 Hz.

UHPLC-MS/MS Analysis. LC-MS/MS analysis was carried out on a Thermo Ultimate 3000 Ultrahigh performance liquid chromatography (UHPLC) instrument coupled with an Orbitrap Fusion TMS (Orbitrap Fusion Tribrid mass spectrometer, Thermo Fisher Scientific Company, USA) and ZORBAX ODS column (150 × 3 mm, 5 μm, Agilent, USA) for the analysis. Electrospray ionization was employed with ion transfer capillary temperature and electrospray voltage set at 250 °C and 3.2 kV, respectively. For MS analysis, all MS spectra were acquired in a positive-ion mode with *S*/*N* values greater than 10. Survey full-scan MS spectra (*m*/*z* 200–2200) were acquired in the orbitrap with a resolution of 60000. The recommended values were used in the full scan acquisition without further optimization. Collision-induced dissociation (CID) was employed as the fragmentation technique. The experimental parameters were set as follows: spray voltage, 2000 V; sweep gas flow rate, 1.0 respective arbitrary units; sheath gas flow rate, 40 respective arbitrary units; aux gas flow rate, 10 respective arbitrary units; ion transfer tube temperature, 350 °C; vaporizer temperature, 200 °C; MS^1 detector, Orbitrap; MS^1 resolution, 120000; MS^1 scan range, 1000–3500; MS^1 maximum injection time, 100 ms; MS^2 CID collision energy, 45%; MS^2 detector, Orbitrap; MS^2 resolution, 15000; MS^2 maximum injection time, 35 ms; and MS^2 start mass, 500. Xcalibur Qual and Quan browser software (Thermo Scientific, USA) were used for qualitative and quantitative analysis.

Database Searching

The acquired data for each sample were combined and submitted to MASCOT search engine integrated into the MaxQuant environment for database search and identification of corresponding peptides employing the following search parameter settings. Database: Uniprot human protein sequence database; Enzyme: Trypsin; Maximum of missed cleavages: 1; Taxonomy: Bos taurus; Peptide tolerance: ±1.2 Da; Mass values: MH+, Monoisotopic.

Gene Ontology (GO) terms for the identified phosphoproteins of AML cells were determined using DAVID (DAVID Functional Annotation Bioinformatics Microarray Analysis, version 6.7). The corresponding Uniprot identifiers to phosphoproteins were uploaded as a gene list. A cut-off of $p <$

0.05 was utilized for all GO categories.

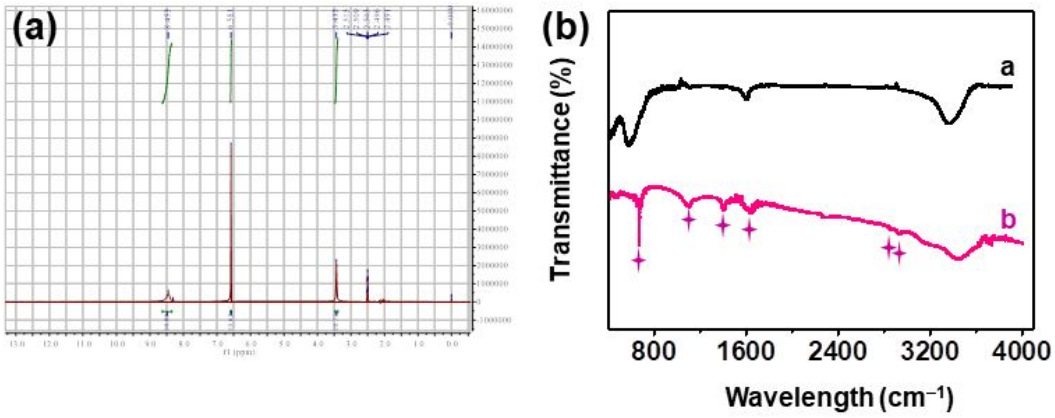

Figure S1. (a) ^{1}H-NMR spectrum of P5 (300 MHz, DMSO-d_6): δ 8.45 (s, 10H), 6.58 (s, 10H), 3.43 (s, 10H); (b) FT-IR spectra of P5SOF (black) and P5SOF-NH_2 (red).

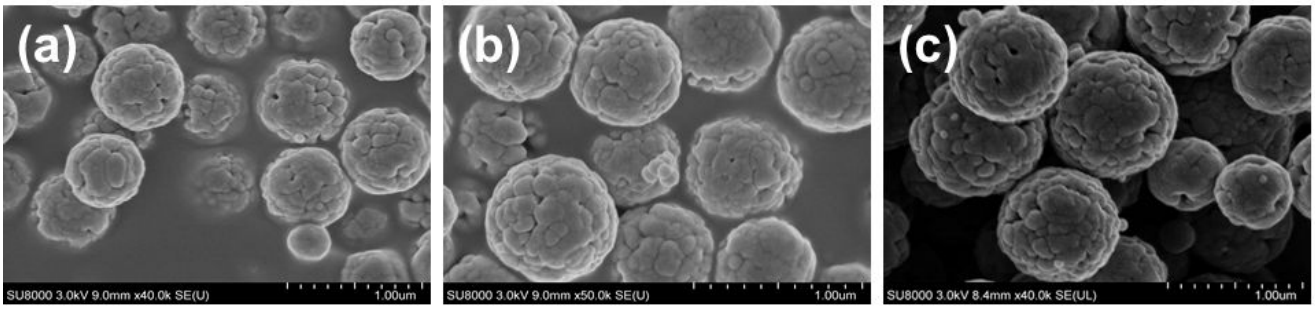

Figure S2. SEM images of mP5SOF-NH_2 with different sizes: (a) ~630 nm, (b) ~730 nm, and (c) ~1000 nm.

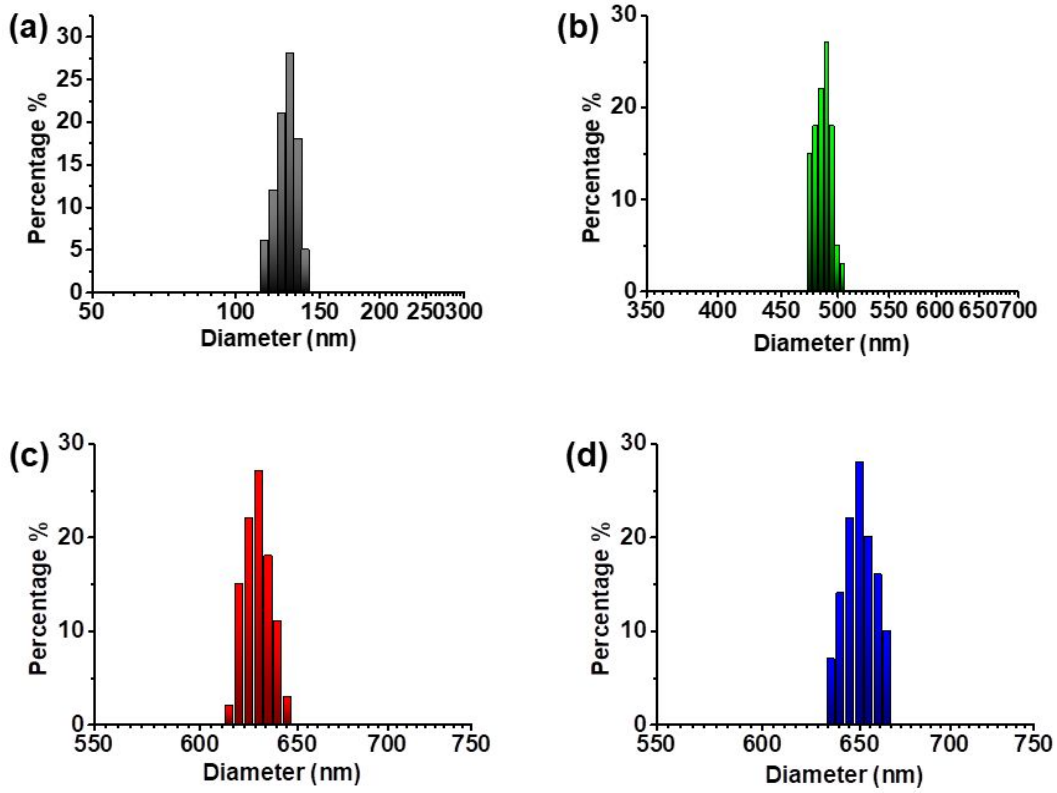

Figure S3. DLS results of (a) P5SOF-NH_2, (b) mGLYMO, (c) mP5SOF-NH_2, and (d) mP5SOF-Arg.

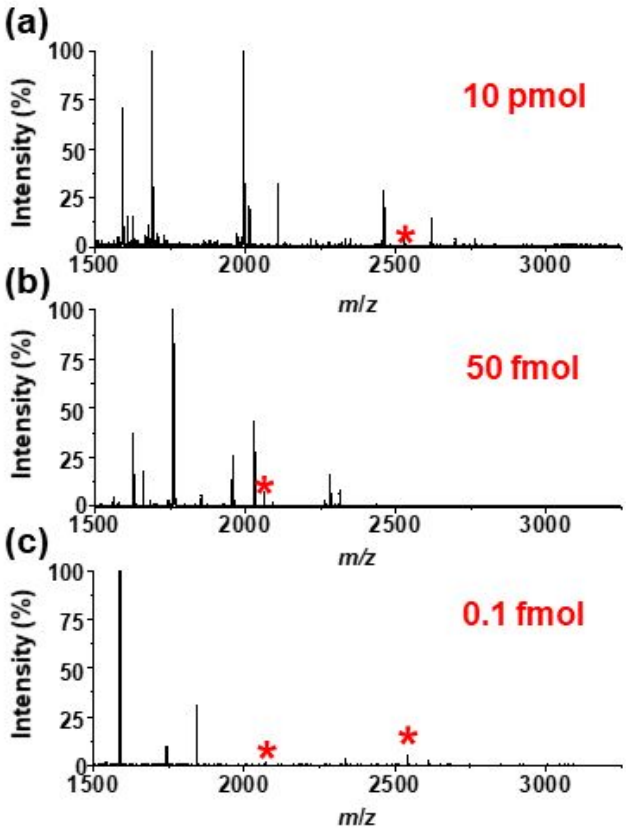

Figure S4. MALDI mass spectra of *β*-casein tryptic digests: (a) enrichment with mGLYMO (10 pmol μL^{-1}); (b) enrichment with mP5SOF-NH_2 (50 fmol μL^{-1}); (c) enrichment with mP5SOF-Arg (0.1 fmol μL^{-1}). The peaks of phosphopeptides or their fragments are marked.

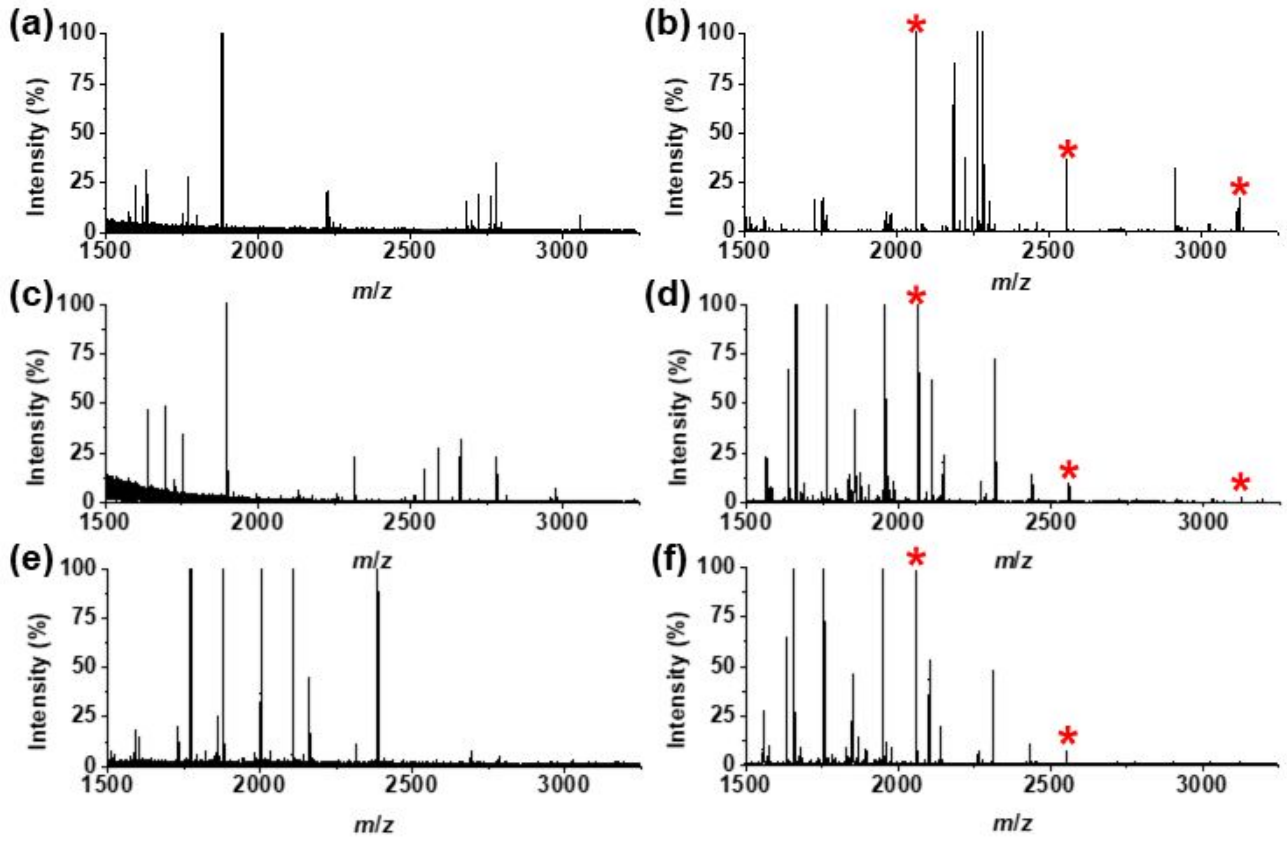

Figure S5. MALDI mass spectra of tryptic digest mixtures of *β*-casein and Cyt C: (a, c, e) before and (b, d, f) after enrichment with mP5SOF-Arg. Molar ratios of Cyt C to *β*-casein are (a, b) 200:1, (c, d) 1000:1, and (e, f) 5000:1, respectively. The peaks of phosphopeptides or their fragments are marked.

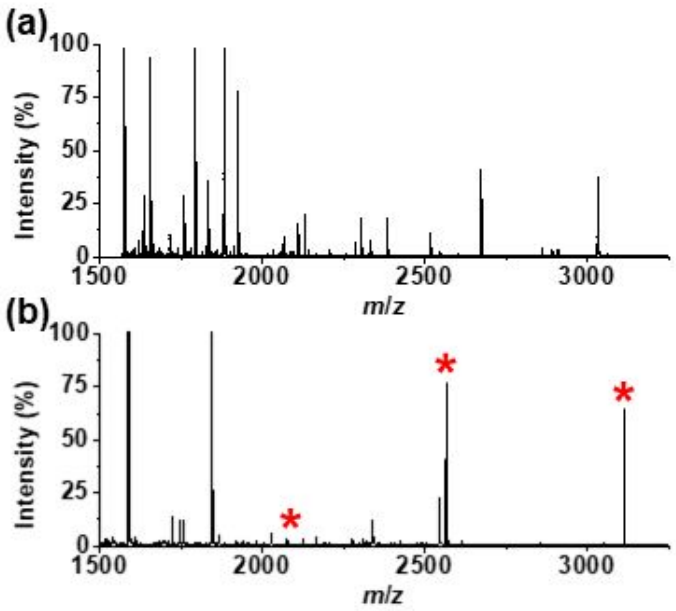

Figure S6. MALDI mass spectra of tryptic digest mixtures of *β*-casein, OVA, Con A, and Cyt C: (a) before and (b) after enrichment with mP5SOF-Arg. Molar ratios of OVA, Con A, and Cyt C to *β*-casein are 500:500:500:1. The peaks of phosphopeptides or their fragments are marked.

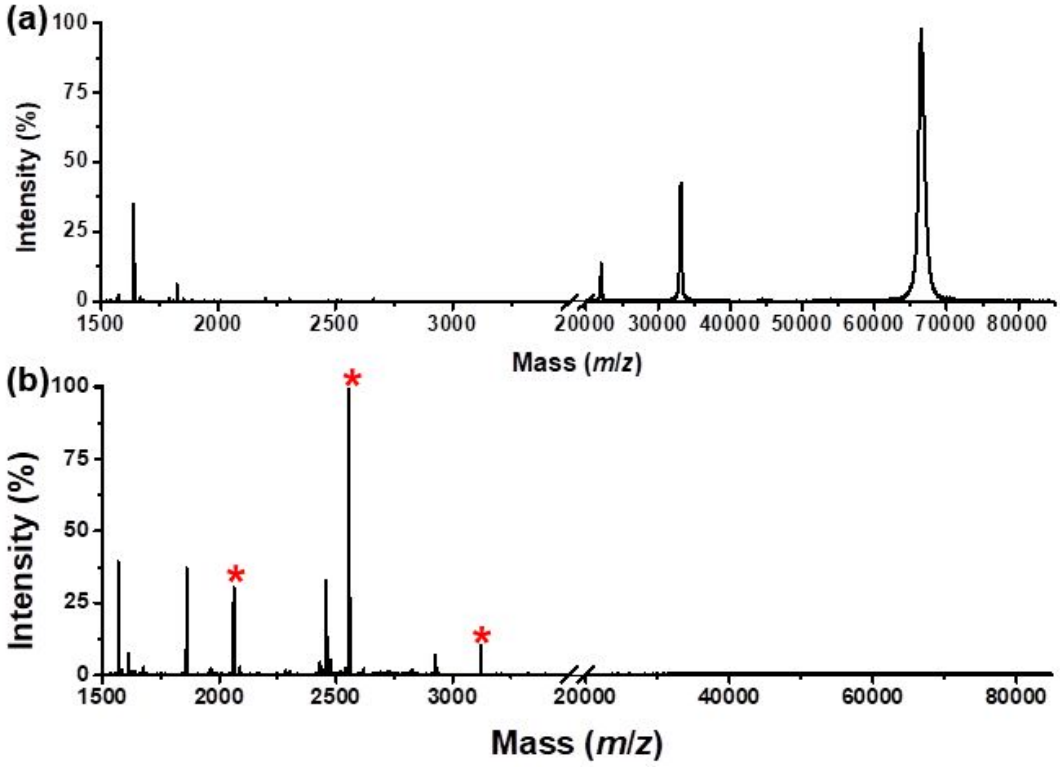

Figure S7. MALDI mass spectra of 0.5 fmol μL^{-1} β-casein tryptic digest and 37.5 fmol μL^{-1} BSA mixture: (a) before and (b) after enrichment with mP5SOF-Arg. The peaks of phosphopeptides or their fragments are marked.

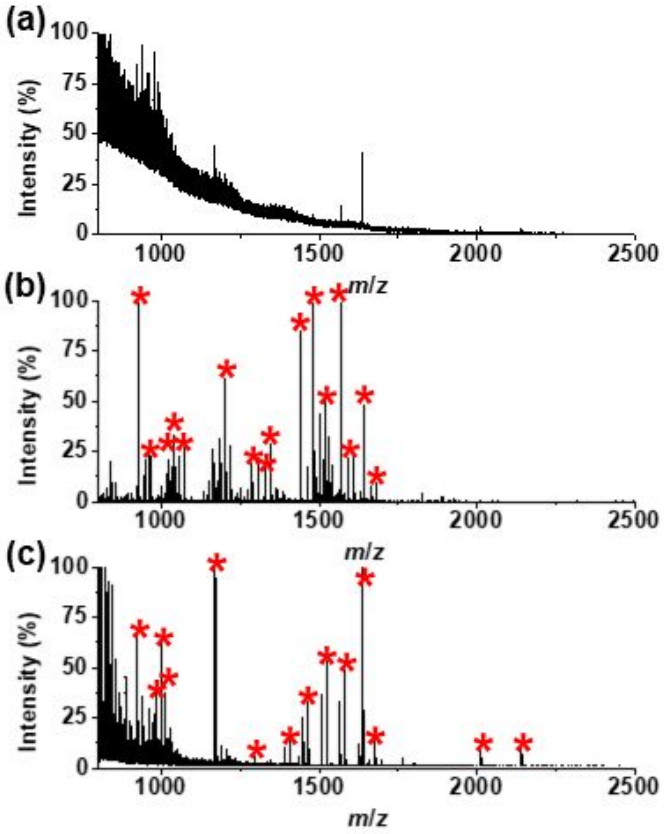

Figure S8. MALDI mass spectra of A549 cells lysate after enrichment with mP5SOF-Arg: (a) 0 h, (b) 12 h, and (c) 24 h.

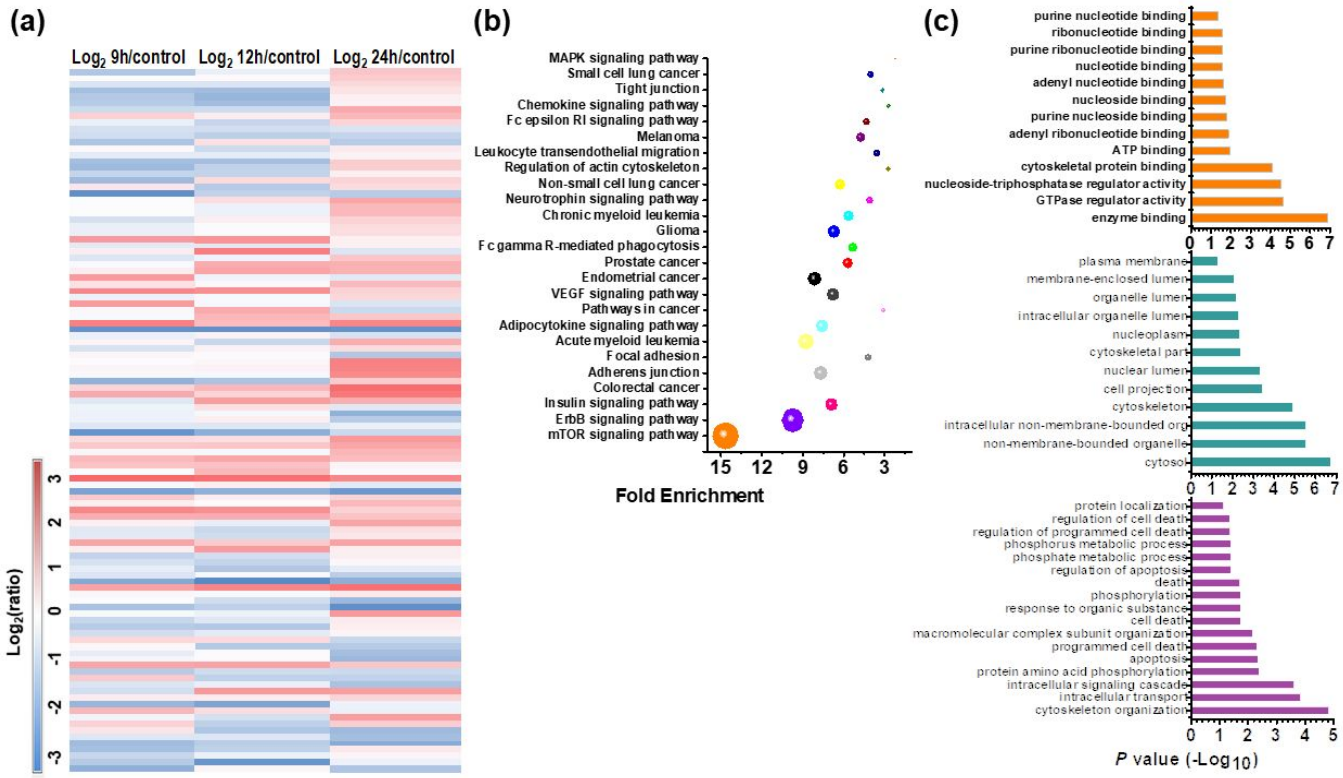

Figure S9. Analysis of the process of exogenous substances treated A549 cells: (a) Unsupervised hierarchical clustering of differentially expressed proteins; (b, c) KEGG and GO term analysis by DAVID for phosphoproteins extracted with mP5SOF-Arg.

Table S1. Detailed information of the observed phosphopeptides obtained from *β*-casein tryptic digests enriched with mP5SOF-Arg.

No.	Phosphoryl groups	Peptide sequence	Observed *m*/*z*
β_{1s}	1	FQsEEQQQTEDELQDK	2061.4
β_{2s}	1	FQsEEQQQTEDELQDKIHPF	2556.7
β_{3m}	4	RELEELNVPGEIVEsLsssEESITR	3122.0

Table S2. Comparison of the determination of phosphopeptides based on different MOF or COF-based materials.

Material	Detection method	LOD (fmol μL^{-1})	Ref.
DZMOF-FDP	MALDI-MS	< 40.0	16
MCNC@COF@Zr^{4+}	MALDI-MS	10.0	22
SiO_2@PDA@Zr-MOF	LC-MS	4.0	39
Fe_3O_4@MIL-100 (Fe)	MALDI-MS	0.5	40
SPIOs@SiO_2@MOF	MALDI-MS	< 10.0	41
Fe_3O_4@PDA@UiO-66-NH	MALDI-MS	0.2	42
mP5SOF-Arg	MALDI-MS	0.1	This work

Table S3. Comparison of the determination of phosphopeptides based on different guanidyl-based materials combined with MALDI-MS.

Material	**LOD (fmol μL^{-1})**	**Ref.**
GF-TiO_2-GO	10.0	43
SPIOs@PVP-PEI@MOF@Arg	10.0	44
PP-*x*-Arg	2.0	45
Fe_3O_4/PDA/PAMA-Arg	0.2	46
Poly(guanidinium ionic liquid)s	4	47
Fe_3O_4@SiO_2@GDN	–	48
Guanidyl functionalized graphene	40	49
mP5SOF-Arg	0.1	This work

Table S4. Recyclability of mP5SOF-Arg for phosphopeptides enrichment: for the first time and after 7 times.

	No.	**Phosphoryl group**	**Observed *m*/*z***
For the first time	β_{1s}	1	2061.4
	β_{2s}	1	2556.7
	β_{3m}	4	3122.0
After 7 times	β_{1s}	1	2060.1
	β_{2s}	1	2556.5
	β_{3m}	4	3122.7

Table S5. List of phosphopeptides from tryptic digest of proteins extracted from defatted milk after enrichment with mP5SOF-Arg.

No.	Phosphoryl group	Peptide sequence	Observed *m*/*z*
α_1	1	TVDMEsTEVF	1236.9
α_2	1	TVDmMEsTEVF	1252.2
α_3	2	EQLsTsEENSKK	1539.2
α_4	1	VPQLEIVPNsAEER	1660.8; 1581.3[#]
α_5	1	YLGEYLIVPNsAEER	1832.5
α_6	2	DIGsEsTEDQAMEDIK	1928.5
α_7	1	YKVPQLEIVPNsAEER	1954.3; 1873.2[#]
α_8	4	NTMEHVsssEEsIISQETYK	2619.5
α_9	3	VNELsKDIGsEsTEDQAMEDIK	2677.9
α_{10}	5	QMEAEsIsssEEIVPNsVEQK	2719.4; 2560.3[#]
α_{11}	4	NTMEHVsssEEsIISQETYKQ	2748.3
α_{12}	3	EKVNELsKDIGsEsTEDQAMEDIKQ	2937.8; 2856.5[#]
α_{13}	4	NANEEEYSIGsssEEsAEVATEEVK	3008.4; 3029.5[#]
β_1	1	FQsEEQQQTEDELQDK	2061.2; 1981.2[#]
β_2	1	FQsEEQQQTEDELQDKIHPF	2556.7; 2480.2[#]
β_3	4	RELEELNVPGEIVEssssEESITR	2964.2; 2884.6[#]
β_4	4	RELEELNVPGEIVEssssEESITR	3122.0; 3042.5[#]; 2962.9[#]

#: de-phosphorylated peptides;

s: phosphorylation on serine;

m: oxidation on methionine.

Table S6. Phosphopeptides detected from egg yolk digest after enrichment with mP5SOF. Characteristic phosphopeptides are derived from three domains of phosvitin (PV) and lipovitellin (LP).

	Position	Sequence	Observed *m/z*
Domain-I	1112-1145	AEFGTEPDAKTSSSSSSASSTATSSSSSSASSPN (PV 1-34)	
	1112-1131	AEFGTEPDAKTSSSSSSASS	1930.5
	1122-1145	TSSSSSSASSTATSSSSSSASSPN	2137.6
Domain-II	1301-1322	SGHLEDDSSSSSSSSVLSKIWG (PV 190-211)	
	1308-1320	SSSSSSSSVLSKI	1255.9
	1306-1318	DDSSSSSSSSVLS	1244.3
	1301-1319	SGHLEDDSSSSSSSSVLSK	1895.6
	1304-1317	LEDDSSSSSSSSVL	1397.6
	1306-1319	DDSSSSSSSSVLSK	1371.0
	1310-1320	SSSSSSVLSKI	1080.9
	1301-1318	SGHLEDDSSSSSSSSVLS	1767.2
	1305-1322	EDDSSSSSSSSVLSKIWG	1856.1
	1301-1320	SGHLEDDSSSSSSSSVLSKI	2007.7
	1306-1320	DDSSSSSSSSVLSKI	1484.8
	1304-1319	LEDDSSSSSSSSVLSK	1613.4
	1301-1317	SGHLEDDSSSSSSSSVL	1679.7
	1306-1317	DDSSSSSSSSVL	1156.7
Domain-III	1056-1076	IITEVNPESEEEDESSPYEDI (LP)	
	1059-1076	EVNPESEEEDESSPYEDI	2095.8
	1068-1076	DESSPYEDI	1053.2
	1059-1072	EVNPESEEEDESSP	1576.3

Table S7. List of phosphopeptides captured by mP5SOF-Arg from human saliva.

No.	Phosphoryl group	Peptide sequence	Observed *m*/*z*
1	2	DssEEKFL	1114.5
2	2	ssEEKFLR	1155.7
3	2	DssEEKFLR	1270.2
4	2	ssEEKFLRR	1311.6
5	1	DsSEEKFLR	1190.3
6	1	sSEEKFLR	1075.8
7	2	DssEEKFLRR	1426.8
8	1	FGQGsGPIVLDDVR	1539.4
9	5	SsAEsssEDVsQ	1596.6
10	2	PQGPPPQGGSKsRSsR	1784.1
11	1	GPPGSRGsPGAPGPPGPPGSH	1899.3
12	1	VISDGGDsEQFIDEERQ	2003.9
13	1	GAPGPAGARGsDGSVGPVGPAGP	1967.9
14	1	QPPQsStMGymGSQ	1754.7
15	1	QGPPPPGKPQGPPPQGDKSQSPRsPPG	2752.1
16	2	PGAPGAPGHPGPPGPVGPAGKsGDRGEsGPAGPAG	3127.1
17	2	QGPPPPGKPQGPPAQGGsKsQSARAPPGKPQ	3146.5

Table S8. List of phosphopeptides analyzed by MADLI-MS from A594 cells after enrichment with mP5SOF-Arg.

No.	Phosphoryl group	Peptide sequence	Protein	Observed *m*/*z*
1	1	T*QTPPVsPA	Src substrate cortactin	958.6
2	1	TQTPPVsPA		997.7
3	1	LSPsPTSQR	Lamin-A/C	1014.2
4	1	AEEDEILNRsPR	Calnexin	1506.2
5	1	AEEDEILNRsPR		1539.8
6	1	AEEDEILNRsPR		1585.3
7	1	EITALAPsTMK	Actin, cytoplasmic 2	1264.9
8	1	EITALAPsTmK		1280.3
9	1	LSsPERQR	Calponin and LIM domain	1319.7
10	1	LSsPERQR	containing 1	1344.6
11	2	Acetyl-ANtVLsGGTTMYPGIADR	Cofilin-1	2127.8
12	2	Acetyl-AsGVAVsDGVIK		1304.8
13	2	Acetyl-AsGVAVsDGVIK		1385.6
14	2	CDssPDSAEDVRK		1570.9
15	2	CDssPDSAEDVRK	Alpha-2-HS-glycoprotein	1595.2
16	3	CDssPDsAEDVRK		1649.1
17	3	LGSQHsPGR	Abl interactor 1	1660.3
18	1	sPSGHPHVR	Palladin	1673.6
19	1	Acetyl-GLLYDsDEEDEER	DNA replication licensing factor	1609.3
20	1	Acetyl-GLLYDsDEEDEER	MCM2	1632.9
21	1	SESAPTLHPYSPLsPK	RelA-associated inhibitor	1625.4
22	1	mtDQEAIQDLWQWR	Nucleophosmin	1915.7

*Q: deletion or truncation of N-terminal on glutamine; s: phosphorylation on serine; t: phosphorylation on tyr; m: oxidation on methionine; Acetyl-: acetylation of N-terminal.