

A Novel, Automated, and Highly Selective Phosphopeptide Enrichment for Phosphopeptide Identification and Phosphosite Localization

Authors

Valery G. Voinov and
Joseph S. Beckman
e-MSion Inc.
Corvallis, OR, USA

Shuai Wu, Kenneth
Newton, Linfeng Wu, and
Jordy J. Hsiao
Agilent Technologies Inc.,
Santa Clara, CA, USA

Introduction

Phosphopeptide enrichment is one of the most challenging sample preparation steps for LC/MS analysis due to the variation in reproducibility and selectivity caused by manual sample preparation. The LC/MS analysis of the phosphopeptide samples can be challenging due to a variety of factors including:

- The metal ions on the surface of the HPLC system or within the solvents can form phosphopeptide-metal ion complexes, which will cause severe suppression of the phosphopeptide peaks. This effect can substantially impair phosphopeptide identification especially for multiphosphorylated peptides.
- The ionization efficiency of phosphopeptides is another frequently discussed issue. In complex mixtures with nonphosphorylated peptides, phosphopeptides show much less ionization efficiency.
- The phosphoryl group is a labile post-translational modification (PTM), and is generally eliminated as phosphoric acid upon fragmentation using CID. This neutral loss of phosphoric acid is usually the prominent fragmentation, and leads to missing site-specific information in the peptide backbone fragmentation, making it difficult to localize the phosphosite in phosphopeptide.

In the Phosphopeptide Challenge of the Human Proteome Project (HPP), each participating laboratory received two sample vials. The vial labeled “Phosphopeptide” contains a set of synthesized phosphopeptides of human sequence at various concentrations, mixed with their nonphosphorylated counterparts. For some peptides, there is more than one phosphorylated form. The second vial labeled “Phosphopeptide-Yeast” contains the same peptides in 6 µg of tryptic-digested yeast lysate. Each vial was provided dry.

This study requests HUPO members to perform peptide sequence analysis using different methods. Each laboratory needs to identify the peptides in the sample, and determine the number and location of the phosphosite on each peptide. Each lab needs to determine the relative abundance of phosphorylation at each modified site by comparison with its nonphosphorylated counterpart. The third request includes phosphopeptide enrichment from the sample containing the yeast matrix and reanalysis by MS (Figure 1). In compliance with this study, HUPO provided a list of 89 human peptide sequences with a worksheet to record results for peptide identification, phosphosite localization, relative quantification, and enrichment.

The Agilent AssayMAP Bravo platform provided a fully automated and highly selective phosphopeptide enrichment workflow using high-capacity Fe(III)-NTA cartridges. Using microchromatography cartridges, this automation platform allows users to easily enrich phosphopeptides in a routine workflow even with a small sample amount. Sample analysis was conducted on an Agilent 1290 Infinity II LC coupled to an Agilent 6550 iFunnel Q-TOF LC/MS with an electromagnetostatic (EMS) ECD cell. To improve the peak shape and recovery of the phosphopeptides, Agilent InfinityLab deactivator additive was added to the HPLC mobile phase. Peptide

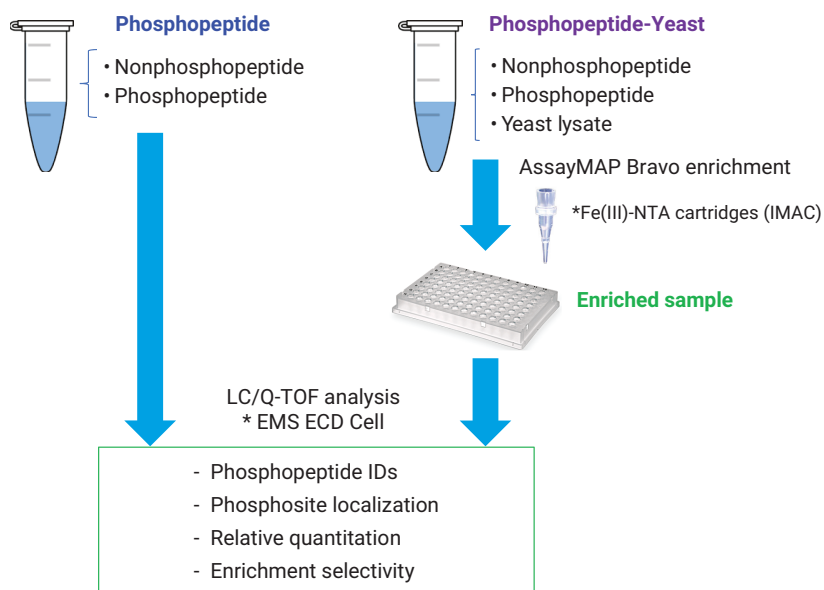


Figure 1. HUPO phosphopeptide challenge experimental design.

identification was performed using data-dependent acquisition (DDA), and data were analyzed by Agilent Spectrum Mill B.06 software. After peptide identification using DDA, phosphosite localization was analyzed using targeted MS/MS with ECD, and data were processed with Byonic software (Protein Metrics, Cupertino, CA, USA) based on the HUPO sequence list. Phosphopeptide relative quantitation was performed using single MS1 scan, the spectra library generated by DDA, together with Skyline software (MacCoss Group, University of Washington, Seattle, WA, USA). The enrichment selectivity was evaluated based on both the overall distinct phosphopeptide number and the phosphopeptide number on the HUPO sequence list.

Experimental

Material

Two sample vials were received from HUPO:

- **“Phosphopeptide”** contained a set of synthesized phosphorylated (Ser, Thr, or Tyr) peptides of human sequence origin at various concentrations, mixed with their nonphosphorylated counterparts. For some peptides, there is more than one phosphorylated form. HUPO provided 89 peptide sequences.
- **“Phosphopeptide-Yeast”** contained the same peptides in a background matrix consisting of 6 µg of trypsin-digested yeast lysate.

AssayMAP Fe(III)-NTA cartridges were from Agilent Technologies Inc. (Santa Clara, CA, USA). All other chemicals were from Sigma-Aldrich (St. Louis, MO, USA).

Phosphopeptide enrichment and cleanup using AssayMAP Bravo

“Phosphopeptide-Yeast” sample vials were resuspended in 100 μL of 80% ACN, 0.1% TFA. The sample was sonicated for two minutes and transferred to a 96-well PCR plate on position A1.

The Agilent AssayMAP Phosphopeptide Enrichment v2.0 App was used for automated phosphopeptide enrichment using an Fe(III)-NTA cartridge. Figure 2 shows the user interface and application settings. One Fe(III)-NTA cartridge was pretransferred to deck location 2 at A1. Labware was chosen according to the labware table, except that LoBind 96 Eppendorf plates were used on deck locations 4 and 9. The cartridge was primed with 100 μL , 50% ACN, 0.1% TFA using a high flow rate of 300 $\mu\text{L}/\text{min}$. Following that, the cartridge was equilibrated using 80% ACN containing 0.1% TFA, the same solvent as the binding buffer (Table 1). Sample loading onto the cartridge was a critical step. In accordance with our previous study, the flow rate was set at 3.0 $\mu\text{L}/\text{min}$ to give enough time for efficient binding of the phosphopeptides. An internal cartridge wash step was used after loading. The enriched phosphopeptides were eluted with 20 μL of 1% ammonium hydroxide (approximately pH 11) directly into a LoBind PCR plate containing 80 μL of 2.5% formic acid (Table 1 and Figure 2).¹

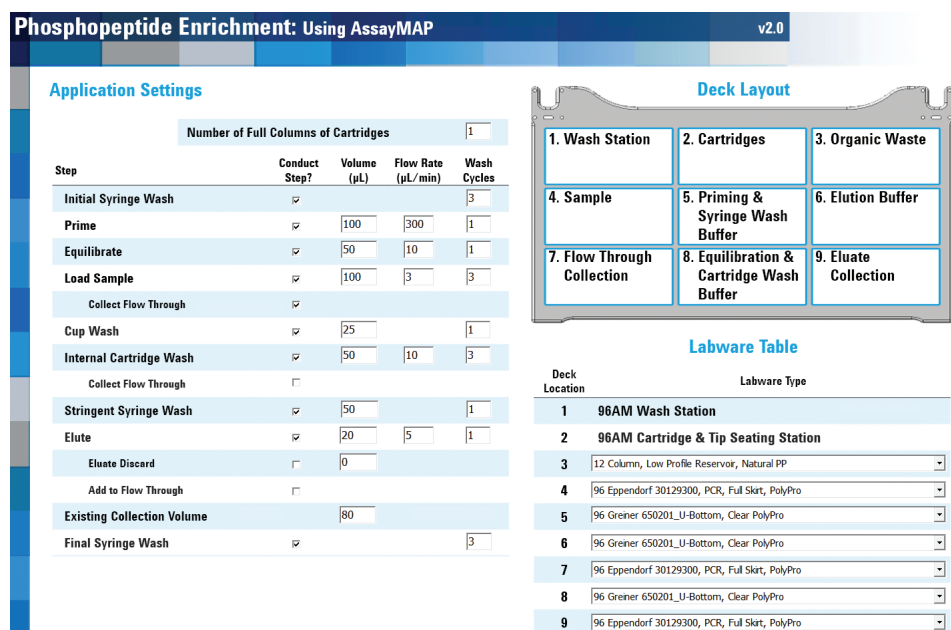


Figure 2. AssayMAP Bravo Phosphopeptide Enrichment v2.0 App.

Table 1. Phosphopeptide enrichment and cleanup protocol using AssayMAP Bravo.

	Phosphopeptide Enrichment	Peptide Cleanup
Affinity Medium	IMAC	Reversed phase
Resin Volume	5 μL	5 μL
Support	Fe(III)-NTA, 100 nmol Fe(III)	C18
Prime Buffer	50% ACN, 0.1% TFA in water	50% ACN, 0.1% TFA in water
Equilibration Buffer	80% ACN, 0.1% TFA in water	0.1% TFA in water
Loading Buffer	80% ACN, 0.1% TFA in water	Ammonium formate in water
Loading Volume	100 μL	100 μL
Loading Flow Rate	3 $\mu\text{L}/\text{min}$	3 $\mu\text{L}/\text{min}$
Washing Buffer 1	80% ACN, 0.1% TFA in water	0.1% TFA in water
Washing Volume 1	50 μL	50 μL
Number of Washes 1	1	1
Elution Buffer 1	1% Ammonium hydroxide	70% ACN, 0.1% TFA in water
Elution Volume 1	20 μL	20 μL
Existing Collection Volume	80 μL	0 μL

The Agilent AssayMAP Peptide Cleanup v2.0 App was used to desalt the enriched phosphopeptides with C18 cartridges (Figure 3). One C18 cartridge was pretransferred to deck location 2 at A1. Labware was chosen according to the labware table, except that LoBind 96 Eppendorf plates were used on deck location 9. The enriched phosphopeptide plate was set at deck location 4. The C18 cartridge was primed with 100 μL of 50% ACN, 0.1% TFA in water, equilibrated with 50 μL of 0.1% TFA, loaded with enriched phosphopeptide sample at a flow rate of 3 $\mu\text{L}/\text{min}$, washed with 50 μL of 0.1% TFA, and eluted with 20 μL of 70% ACN, 0.1% TFA at a flow rate of 5 $\mu\text{L}/\text{min}$.

The sample was dried at room temperature in a SpeedVac vacuum concentrator (Thermo Fisher Scientific, Inc., Waltham, MA, USA), resuspended in 25 μL of 10% ACN, 0.1% FA, and sonicated for two minutes. The sample was further diluted with 25 μL of 0.1% FA in water, and the final sample was in 50 μL of 5% ACN, 0.1% FA.

Peptide identification using data-dependent acquisition

"Phosphopeptide" samples were resuspended in 25 μL of 10% ACN, 0.1% FA. The sample vials were vortexed and sonicated for two minutes. Samples were further diluted in 25 μL of 0.1% FA, and the final samples were in 50 μL of 5% ACN, 0.1% FA.

The 1290 Infinity II LC system was converted to nanoflow LC by coupling with the Agilent Infinity UHPLC Nanodapter. This nanoflow LC was connected to the Agilent nanospray ESI source and coupled with the Agilent 6550 iFunnel Q-TOF LC/MS for peptide identification (Figure 4).

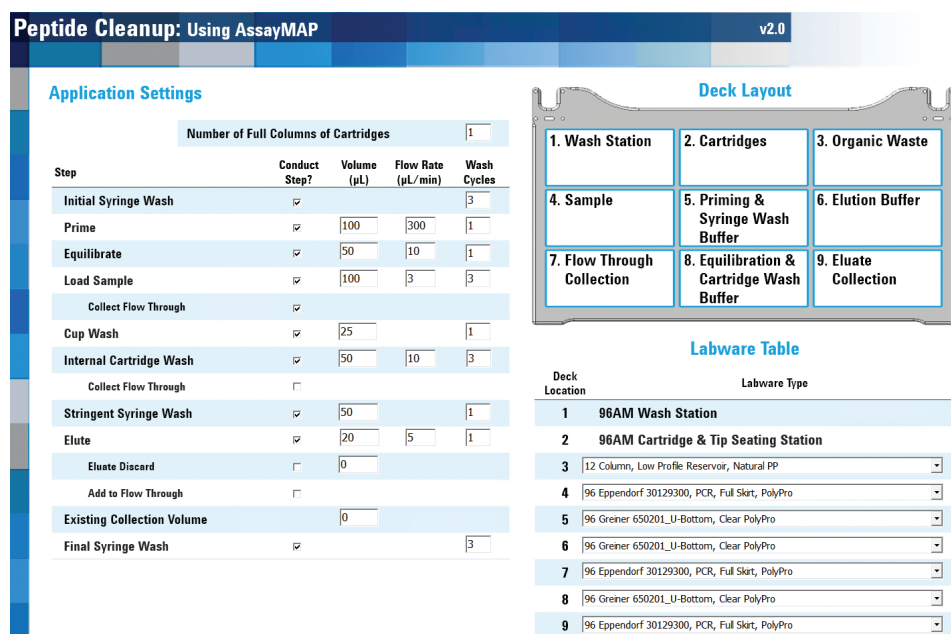
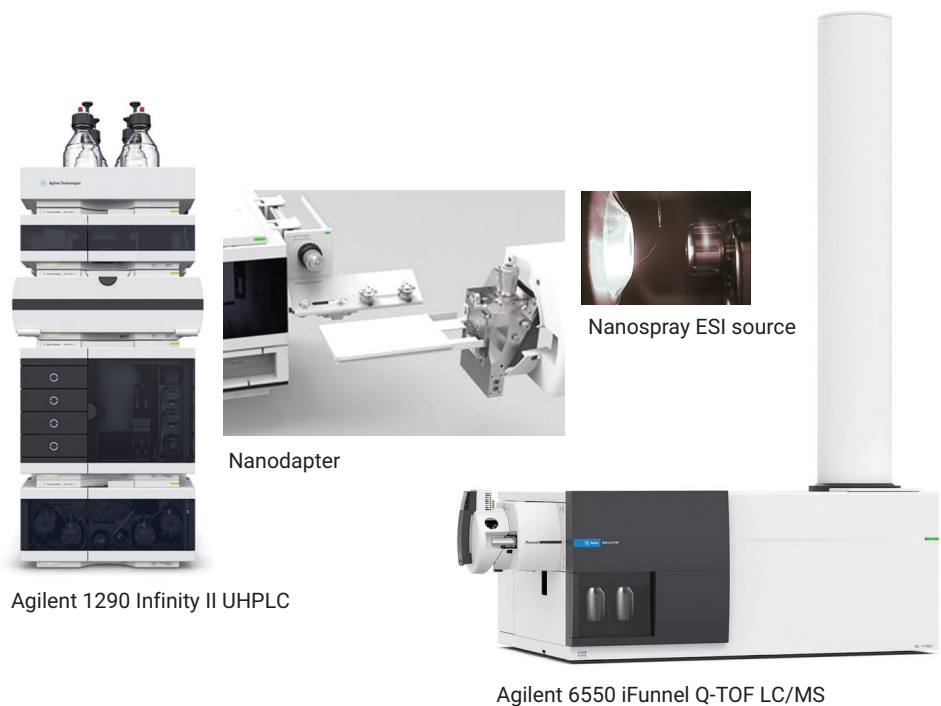


Figure 3. Agilent AssayMAP Bravo Peptide Cleanup v2.0 App.



Agilent 1290 Infinity II UHPLC

Agilent 6550 iFunnel Q-TOF LC/MS

Figure 4. Nanodapter converts standard-flow LC to nanoflow LC coupled with an Agilent 6550 iFunnel Q-TOF LC.

Table 2 lists the LC parameters. The Nanodapter was configured in direct injection mode. A 75 $\mu\text{m} \times 25\text{ cm}$ C18 column was kept at 60 $^{\circ}\text{C}$ and used for peptide separation with a 90 minute gradient in a total 120 minute LC run time.² To enhance the chromatographic performance of the phosphopeptides with minimal ion suppression, 0.1% InfinityLab Deactivator additive was added to solvent A.³ For peptide identification, 2 μL of “Phosphopeptide” and enriched “Phosphopeptide-Yeast” sample were injected with triplicate injections for each sample. Data-dependent acquisition was used with a selection of top 15 precursor ions. Table 3 lists the detailed setup for the 6550 iFunnel Q-TOF LC/MS for DDA using CID.

Table 2. Nano-LC parameters.

LC Conditions													
Nanodapter Configuration	Direct injection mode												
Guard Column	PepMap C18, 75 $\mu\text{m} \times 2\text{ cm}$												
Analytical Column	PepMap C18, 75 $\mu\text{m} \times 25\text{ cm}$												
Column Temperature	60 $^{\circ}\text{C}$												
Solvent A	0.1% formic acid, 0.1% deactivator in water												
Solvent B	0.1% formic acid in 90 % acetonitrile												
Flow Rate	0.085 mL/min primary flow 300 nL/min on-column flow rate												
Q-TOF Gradient	<table border="1"> <thead> <tr> <th>Time (min)</th> <th>B (%)</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>3</td> </tr> <tr> <td>90</td> <td>37</td> </tr> <tr> <td>95</td> <td>70</td> </tr> <tr> <td>97</td> <td>70</td> </tr> <tr> <td>100</td> <td>3</td> </tr> </tbody> </table>	Time (min)	B (%)	0	3	90	37	95	70	97	70	100	3
Time (min)	B (%)												
0	3												
90	37												
95	70												
97	70												
100	3												
Stop Time	115												
Post Time	5												
Injection Volume	2 μL for CID, 4 μL for ECD, 1 μL for MS1												

Table 3. Agilent 6550 iFunnel Q-TOF parameters.

Parameter	Value														
Spray Needle	New objective noncoated needle, 25 μm id, 10 μm tip id, 5 cm length, orthogonally positioned														
Gas Temperature	200 $^{\circ}\text{C}$														
Drying Gas	11 L/min														
Acquisition Mode	Extended dynamic range (2 GHz) m/z 100 to 1700 High analytical sensitivity														
	MS	MS/MS													
Mass Range	m/z 300 to 1700	m/z 50 to 1700													
Acquisition Rate	3 spectra/s	>3 spectra/s													
Isolation Width	Medium (~4 m/z)														
Collision Energy	<table border="1"> <thead> <tr> <th>Charge</th> <th>Slope</th> <th>Offset</th> </tr> </thead> <tbody> <tr> <td>2</td> <td>3.1</td> <td>1</td> </tr> <tr> <td>3</td> <td>3.6</td> <td>-4.8</td> </tr> <tr> <td>>3</td> <td>3.6</td> <td>-4.8</td> </tr> </tbody> </table>			Charge	Slope	Offset	2	3.1	1	3	3.6	-4.8	>3	3.6	-4.8
Charge	Slope	Offset													
2	3.1	1													
3	3.6	-4.8													
>3	3.6	-4.8													
Maximum Precursor/Cycle	15														
Precursor Threshold	1,000 counts and 0.01 %														
Active Exclusion	Excluded after 1 spectra Released after 0.2 minutes														
Isotope Mode	Peptides														
Sort Precursors	By abundance only; +2, +3, >+3														
Scan Speed Varied Based on Precursor Abundance	Yes														
Target	25,000 counts/spectrum														
Use MS/MS Accumulation Time Limit	Yes														
Purity Stringency	100 %														
Purity Cutoff	30 %														

Spectrum Mill was used to search against the Swiss-Prot human protein database with a 1.5% false discovery rate. Trypsin was the digestion enzyme, and up to two missed cleavages were allowed. Carbamidomethylation was set as the fixed modification while N-term Acetyl, deamidation (N), and phosphorylation of serine (S), threonine (T), and tyrosine (Y) were set as the variable modifications. Table 4 lists the detailed search parameters.

Phosphorylation site localization using ECD

The ECD cell was designed to merge with a shorter hexapole collision cell to replace the original hexapole collision cell in Agilent Q-TOF systems. The length of the ECD cell combined with shorter collision cell is the same as the original collision cell (Figure 5). A peptide standard Substance, P, was used to tune the ECD signal to obtain an approximately 1% conversion efficiency for the +2-charge state. Although acquisitions were done with zero added collision energy, the settings used for the ECD experiments were equivalent to approximately 5 eV of applied collision energy.

A targeted list of precursor ions that contains m/z , charge state, retention time, delta retention time, isolation width, and acquisition time was generated using the phosphopeptide list identified from the DDA experiment. Collision energy was set to 0 eV, and the same precursor ions were fragmented four to five times in one run. Data were analyzed with Byonic. Phosphopeptides that were characterized in the first run were removed from the targeted MS/MS list, and further experiments were conducted for unidentified phosphopeptides with more injections.

Table 4. Spectrum Mill search parameters.

Spectrum Mill Search Parameters		
Database	Swiss-Prot human proteins	
Enzyme	Trypsin	
Maximum Missed Cleavage	2	
Modifications	Fixed Carbamidomethylation	Variable Acetyl (N-term) Deamidated (N) Phosphorylated (S) Phosphorylated (T) Phosphorylated (Y)
Mass Tolerance	MS1 15 ppm	MS2 30 ppm
Maximum Ambiguous Precursor Charge	5	
Calculate Reversed Database Scores	Yes	
Dynamic Peak Thresholding	Yes	
Peptide FDR	1.50%	
Precursor Charge Range	2 to 7	

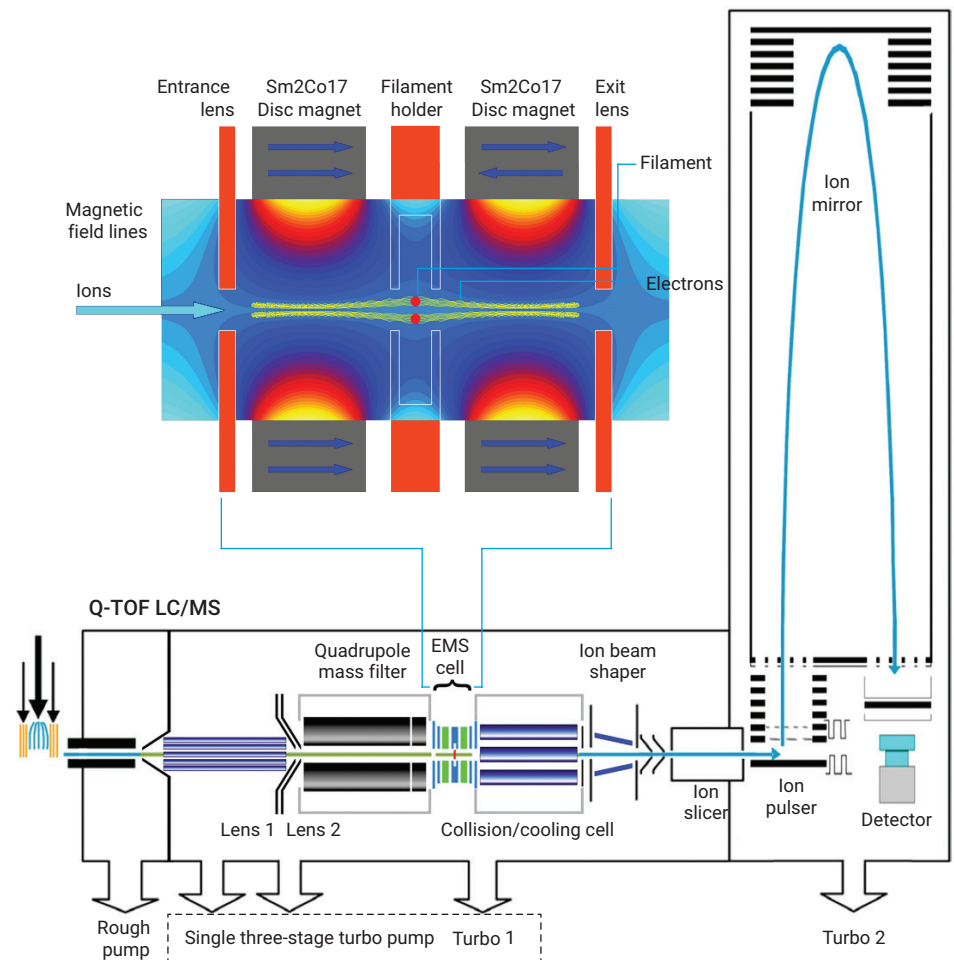


Figure 5. Schematic of an electromagnetostatic (EMS) ECD cell and its position in an Agilent Q-TOF LC/MS.

Byonic software was used to analyze ECD spectra. The same Swiss-Prot human protein database was used with trypsin as the enzyme and up to two missed cleavages. Carbamidomethylation was set as the fixed modification, while deamidation (N, Q) and phosphorylation of serine (S), threonine (T), and tyrosine (Y) were set as the variable modifications.

Relative quantitation of phosphorylation at each modified site

Triplicate MS1 experiments were also conducted for "Phosphopeptide" samples with 1 μ L injections using the same LC gradient. The DDA data were imported into Skyline first, and a peptide CID spectra library was created. The full MS1 data were then imported into Skyline, and the identified peptide peaks were integrated based on MS1 scan. The peak areas (or ion

intensity from both +2 and +3 precursor ions) were exported into Excel and added for both nonphosphopeptides and phosphopeptides including their deamidated forms. The ratio of phosphopeptide/nonphosphopeptide at each phosphorylation site were calculated based on the ratio of ion intensity.

Results and discussion

Spectrum Mill analysis of DDA data

Figure 6A shows the total ion chromatogram (TIC) of "Phosphopeptide" sample with a 90 minute gradient. With triplicate sample injections using DDA, Spectrum Mill identified 437 distinct peptides with 294 distinct phosphopeptides (Table 5). To evaluate the analytical results, HUPO provided a list of 89 peptide sequences. The 89 nonphosphopeptides were all

identified from the "Phosphopeptide" sample. Their phosphorylated counterparts were also identified without confirmation of most phosphosite location using CID. The enriched "Phosphopeptide-Yeast" sample was also analyzed using the same method. Figure 6B shows the TIC of the enriched "Phosphopeptide-Yeast" sample using a 90-minute gradient. Table 5 lists the detailed search result for both the "Phosphopeptide" and the enriched "Phosphopeptide-Yeast" samples including each replicate and the combined search result. After enrichment, a total of 287 distinct peptides were identified with 264 distinct phosphopeptides. The selectivity of the enrichment (phosphopeptide/peptide) from the overall peptide ID numbers is approximately 92%.

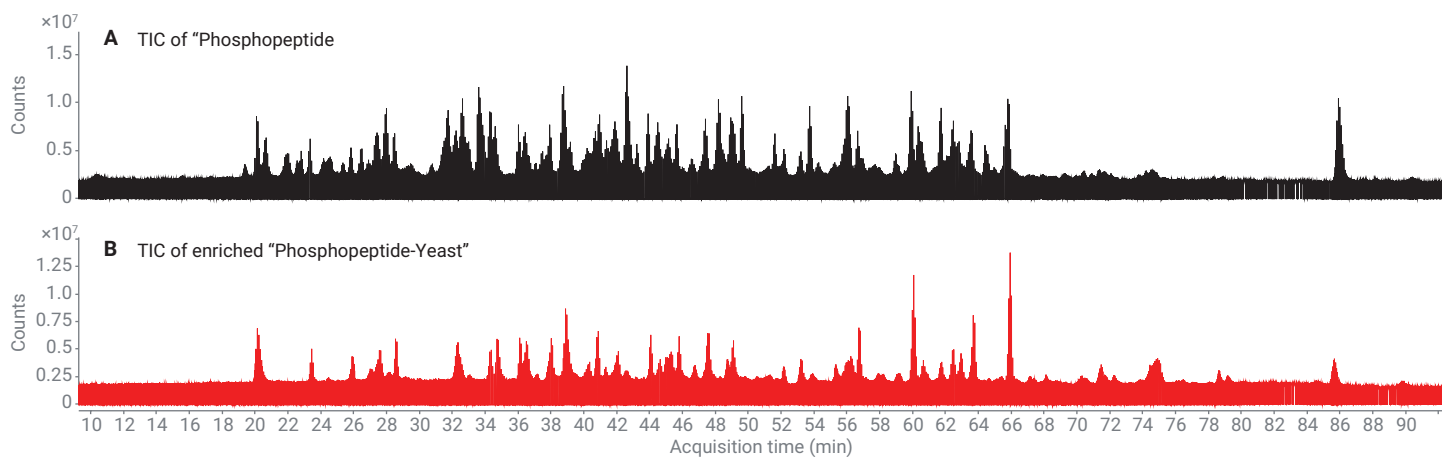


Figure 6. A) TIC of "Phosphopeptide" sample using 90 minute gradient. B) TIC of enriched "Phosphopeptide-Yeast" sample using 90 minute gradient.

Table 5. Spectrum Mill search results.

	"Phosphopeptide"				Enriched "Phosphopeptide-Yeast"			
	Replica 1	Replica 2	Replica 3	Combine	Replica 1	Replica 2	Replica 3	Combine
Total Distinct Peptides	316	297	308	437	193	203	203	287
Total Distinct Phosphopeptides	195	182	193	294	179	192	189	264
Mono-	123	125	127		117	127	123	
Di-	63	50	58		52	56	53	
Tri-	8	6	7		8	8	12	
Tetra-	1	1	1		2	1	1	
% Phosphopeptides/Total Peptides	61.7	61.3	62.7	67.3	92.7	94.6	93.1	92.0
Total Phosphosites	277	247	268		253	267	269	
Sites Assigned %	63	61.9	63.7	62.9	62.4	63.4	62.4	62.7
Sites Not Assigned	102	94	97		95	98	101	

ECD spectra analysis using Byonic

The result for peptide VVEAVNSDSDSEFGIPK analysis is used as an example to demonstrate how the phosphosites were localized. Figure 7A shows the CID spectrum of nonphosphorylated peptide VVEAVNSDSDSEFGIPK. Spectrum Mill identified the precursor ion m/z 896.93 ($z = +2$, mass error = 1.1 ppm) with no phosphosite on this peptide. The sequence matching result showed a complete sequence coverage for this peptide with b, y type ions.

Figures 7B, 7C, and 7D show the ECD spectra of peptides with the same sequence but with mono-, di-, and triphosphorylation. In Figure 7B, Byonic identified the precursor ion m/z 936.92 ($z = +2$, mass error = -0.07 ppm) with one phosphosite for this peptide. The diagnostic ions c_6 ($m/z = 629.36$) and c_7 ($m/z = 796.37$) confirmed the phosphosite at serine 7 on this peptide. In Figure 7C, Byonic identified the precursor ion m/z 651.60 ($z = +3$, mass error = -3.69 ppm) with two phosphosites for this peptide. The diagnostic ions c_7 ($m/z = 796.35$) and

z_{11} ($m/z = 1325.45$), c_9 ($m/z = 1078.37$) and z_9 ($m/z = 1043.40$) confirmed the two phosphosites at serine 7 and 9 for this peptide. In Figure 7D, Byonic identified the precursor ion $m/z = 678.25$ ($z = +3$, mass error = -11.54 ppm) with three phosphosites for this peptide. The diagnostic ions c_7 ($m/z = 796.37$), c_9 ($m/z = 918.42$), and z_9 ($m/z = 1123.40$) confirmed the three phosphosites at serine 7, 9, and 11. The CID spectra (not shown) for peptide VVEAVNSDSDSEFGIPK with different degrees of phosphorylation can only identify the number of phosphosites without confirming the location of the phosphosites.

Comparing Figure 7B with 7C, when the size of the phosphopeptide increases, the precursor ion with +3 charge state became more abundant, resulting in a better ECD spectrum with a higher sequence coverage (Figure 7C). Precursor ion with a +2 charge state and lower abundance generated fewer fragment ions, but we were still able to confirm the location of the phosphosite (Figure 7B). Comparing Figure 7C with 7D, peptides with multiple phosphosites

(≥ 3) generally have less ionization efficiency in the positive ion mode. The more abundant precursor ion with the same charge state (+3) also results in a better ECD spectrum (Figure 7C). Comparing Figure 7B with 7D, the two precursor ions have nearly the same abundance. The one with a +3 charge state still generates more fragment ions than the one with a +2 charge state.

In this study, HUPO provided a list of 89 peptide sequences. The total number of distinct phosphopeptides based on this sequence list and confirmed by ECD is 94 in the "Phosphopeptide" sample. Moreover, 93 out of 94 phosphopeptides spiked into the yeast lysate were still identified from the enriched "Phosphopeptide-Yeast" sample. This result showed that a high number of phosphopeptides were recovered from the enrichment. Conversely, 9 out of 89 nonphosphopeptides were still detected after enrichment of the "Phosphopeptide-Yeast" sample. The selectivity of the enrichment (phosphopeptide/peptide) based on the peptide sequence list is approximately 91.3%.

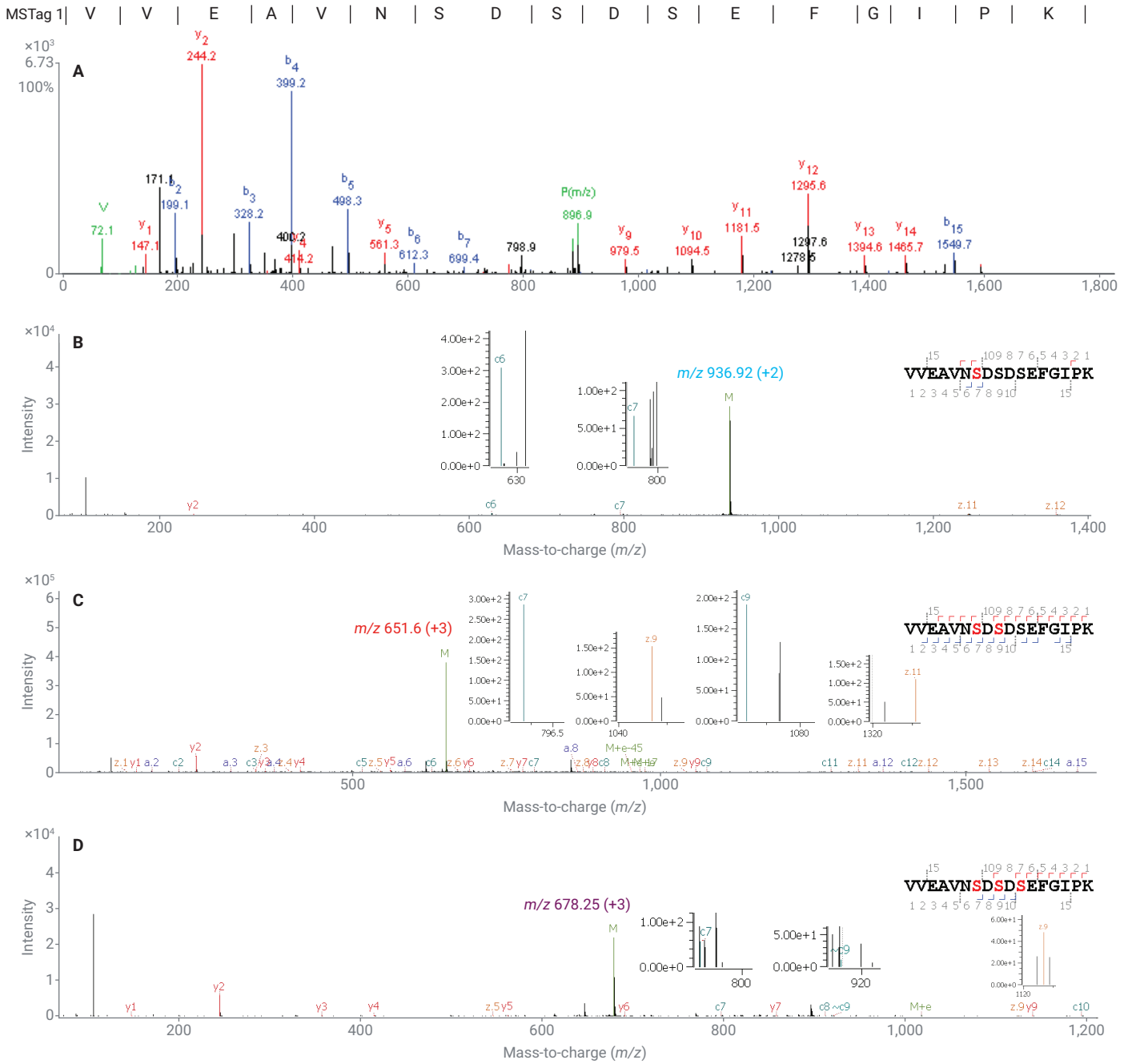


Figure 7. A) CID spectrum of peptide VVEAVNSDSDSEFGIPK showing a complete sequence coverage with b, y type ions. B), C), D) ECD spectra of peptides VVEAVNSDSDSEFGIPK with mono-, di- and tri phosphorylation localized by c, z type ions. Note the presence of most of the sequence ions in the ECD spectra even with the expected low efficiency for low charge state.

Relative quantitation of peptide phosphorylation

Figure 8 shows the extracted ion chromatograms (EICs) of peptide VVEAVNSDSDSEFGIPK with different degrees of phosphorylation. All the peptides were well separated using the C18 column with the 90-minute gradient. For di- and triphosphopeptides, the ion intensity of the deamidated phosphopeptides were added together with the original phosphopeptide. The relative abundance of phosphopeptide/nonphosphopeptide is calculated based on ion intensity. For peptide VVEAVNSDSDSEFGIPK, if the abundance of nonphosphopeptide is 1.00, the ratio of mono-, di- and triphosphopeptides are 0.03, 4.71, and 0.02 respectively, which are annotated in Figure 8. The relative abundance of phosphorylation at each modified site was calculated for all the 94 phosphopeptides in the "Phosphopeptide" sample based on the ratio of phosphopeptide/nonphosphopeptide. Note that this relative quantitation does not consider the different ionization efficiency from different peptide species. It is only a comparison from the ion intensity of different peptides.

Conclusion

An automated phosphopeptide enrichment with both qualitative and quantitative analysis using the Agilent AssayMAP Bravo platform and an LC/Q-TOF system was implemented for the HUPO Phosphopeptide Challenge.

A CID experiment was performed for peptide identification in which 437 distinct peptides with 294 phosphopeptides were identified in the "Phosphopeptide" sample. All 89 nonphosphopeptides from the HUPO sequence list were identified. The ECD experiment determined the location of 124 phosphosites from 94 phosphopeptides based on the 89 nonphosphopeptide sequences. The rest of the peptides not on the sequence list were also reported back to HUPO.

Within the enriched "Phosphopeptide-Yeast" sample, 287 distinct peptides were identified, of which 264 were distinct phosphopeptides. The overall selectivity of the enrichment was approximately 92.0 %.

Moreover, 93 out of 94 phosphopeptides spiked into the yeast were still identified from the enriched

"Phosphopeptide-Yeast" sample. Agilent showed the highest number of phosphopeptides recovered from the enrichment compared to other labs in this study.

References

1. Russell, J. D.; Murphy, S. Agilent AssayMAP Bravo Technology Enables Reproducible Automated Phosphopeptide Enrichment from Complex Mixtures Using High-Capacity Fe (III)-NTA Cartridges, *Agilent Technologies Application Note*, publication number 5991-6073EN, **2016**.
2. Wu, S.; Wu, L. Human Breast Cancer Cell Line Phosphoproteome Revealed by an Automated and Highly Selective Enrichment Workflow, *Agilent Technologies Application Note*, publication number 5994-0315EN, **2018**.
3. Hsiao, J. J. *et al.*, Improved LC/MS Methods for the Analysis of Metal-Sensitive Analytes Using Medronic Acid as a Mobile Phase Additive. *Anal. Chem.* **2018**, *90*(15), 9457–9464.

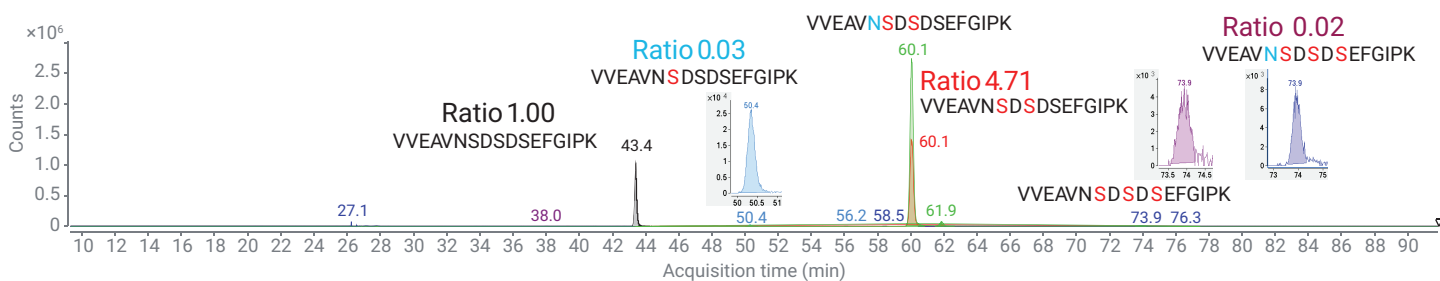


Figure 8. EIC of all VVEAVNSDSDSEFGIPK peptides with the ratio of phosphopeptide/nonphosphopeptide annotated.

www.agilent.com/chem

For Research Use Only. Not for use in diagnostic procedures.

DE.5458564815

This information is subject to change without notice.

© Agilent Technologies, Inc. 2019, 2020
Printed in the USA, November 2, 2020
5994-1235EN