Targeted Proteomics for Determining Phosphorylation Site-Specific Associations in Cardiovascular Disease

Running title: Cordwell et al.; Monitoring dynamic PTMs: new frontiers

Stuart J. Cordwell, PhD1,2 and Melanie Y. White, PhD1

1School of Molecular Bioscience, 2Discipline of Pathology, School of Medical Sciences, The University of Sydney, New South Wales, Australia

Address for Correspondence:
Melanie Y. White, PhD
School of Molecular Bioscience
Building GO8, Maze Crescent
The University of Sydney
NSW, Australia 2006
Tel: +61-2-9036-7918
Fax: +61-2-9351-4726
E-mail: melanie.white@sydney.edu.au


Key words: Editorials; mass spectrometry; phosphoproteins; phosphoproteomics; selected reaction monitoring; western blotting
Post-translational modifications (PTMs) are chemical or physical changes to proteins that can alter structure/function relationships and thus influence catalysis, cell localization and interactions with other proteins and biomolecules. There are over 200 recognized PTMs, generating a vast scope for altering protein function and increasing the potential for complex regulatory cross-talk between diverse modifications. Protein phosphorylation is of major significance since 30-50% of all cellular proteins may be targeted. Phosphorylation is a transient means of amplifying environmental signals to rapidly alter protein structure, interactions and ultimately, function. While many studies have attempted to determine how aberrant signalling can contribute to the pathophysiology of diseases (e.g insulin stimulation), very few have examined the associations of long-term alterations in protein phosphorylation during chronic disease. Within the myocyte, phosphorylation regulates numerous key functions.

In this issue of Circulation, Zhang and colleagues have utilized targeted proteomics by multiple reaction monitoring (MRM) mass spectrometry (MS) to accurately quantify phosphorylated amino acids in cardiac troponin I (cTnI) associated with failing ischemic and idiopathic dilated cardiomyopathy hearts.

Traditionally, the study of single protein phosphorylation has relied on Western blot profiling with site-specific antibodies directed against ‘known’ phosphorylation sites. Such an approach loses contextual information at the signal network and cellular level. This is critical, as the extent of protein phosphorylation mediates downstream consequences and can also aid in defining upstream regulation. Given the right conditions, including kinase activation, site accessibility and adequate supply of phosphate (generally derived from ATP), nearly any serine, threonine or tyrosine may be phosphorylated. The challenge therefore is two-fold: adequately covering these residues analytically and determining those sites of modification that result in
altered protein function. New approaches to determining the phospho-status of proteins aim to take an unbiased view, whereby known and unknown sites of phosphorylation can be investigated simultaneously. An important distinction is that such approaches rely on MS, rather than the limited number of protein- and site-specific antibodies, for identification of phosphorylation sites. Typically, elucidation of signal pathways that are activated or repressed in disease models is generated by large-scale phosphoproteomics experiments using quantitative tags (such as stable isotope labelling of amino acids in cell culture [SILAC] \(^7\) or isobaric tags for relative and absolute quantitation [iTRAQ] \(^8\)) and liquid chromatography coupled to high resolution tandem MS (LC-MS/MS). These datasets allow the generation of temporal profiles that demonstrate how cell signal pathways respond and adapt to, changes in the cellular and external environments. Since many thousands of phosphorylation sites can be monitored simultaneously, reconstruction of important biological signal pathways is possible, thus generating hypotheses for novel disease interventions in cell or animal models.

Phosphoproteomics datasets generated by enrichment using methods such as broad spectrum anti-phospho antibodies (e.g. immunoprecipitation using anti-phosphotyrosine antibodies), titanium dioxide (TiO\(_2\)), immobilized metal affinity chromatography (IMAC) and combinations of these together \(^9,10\), also provide a sufficient body of data to enable motif-based algorithms to predict those kinases / phosphatases most active during the test or disease conditions. For example, positional analysis may demonstrate increased phosphorylation of sites specific to protein kinase A (PKA), and thus allow directed tests to determine whether that kinase is indeed activationally induced, or whether an analogous yet hitherto unknown kinase may be responsible for increased phosphorylation at those sites. Large-scale phosphorylation
studies also enable the elucidation of novel sites in well-studied proteins, thus leading to the generation of new hypotheses about their function and regulation.

Once large-scale information has been generated, the task of validating these data becomes central. A second issue is that these unbiased phosphoproteomics approaches typically do not cover every signal event within a pathway, thus leaving ‘gaps’ in the network that need additional information. Validation thus provides an opportunity to use targeted methods to fill the remaining gaps after large-scale phosphoproteomics (Figure 1). Since the number and site-specificity of anti-phosphorylation site antibodies is low, Western blotting is only possible for confirmation of phosphorylation sites that are already well-understood 11. Therefore, novel approaches are also needed, typically involving selected (SRM) or multiple reaction monitoring (MRM) MS of identified peptides 12. In MRM, the MS instrumentation scans only for ions of known mass/charge (m/z) and with known, and thus predictable, fragmentation ion patterns. The relative intensity of these parent: fragment ion pairs (or ‘transitions’) can be used to perform relative quantitation, or, if labelled synthetic standard peptides are used, absolute quantification across hundreds of clinical samples (Supplemental Figure 1). For cTnI, site-specific antibodies could be used to validate data for previously characterized sites (Ser-22 and Ser-23), while the MRM approach allowed both validation and quantitation of novel sites across many clinical specimens. One of the most intriguing findings in the Zhang et al. study is the identification of several completely novel sites of phosphorylation in cTnI 6. The unbiased MRM approach allowed the significance of these sites to be determined without the need to generate new site-specific antibodies, which is important considering the changes shown in the occupation of many of these novel sites in failing hearts.
For the analysis of post-translationally modified proteins / peptides, the MRM method not only allows the quantification of the modified peptide between samples, but also determination of the ratio between modified and unmodified peptide within a single sample, thus enabling calculation of the modification stoichiometry. For many proteins, and cTnI is an excellent example, this is not as simple as detecting phosphorylated versus non-phosphorylated peptide. In several places, cTnI contains neighbouring modification sites that are localized within a single predicted tryptic peptide. For example, tryptic peptides containing potential phosphorylation sites at Ser-4 and Ser-5, Ser-22 and Ser-23, Ser-41 and Ser-43, and Ser-76 and Thr-77 can all exist in at least 4 different forms – unmodified, a single phosphorylation at either site, and both sites modified – with other PTMs possible at these and other amino acids. Zhang et al. have successfully monitored all 4 variants of each of these tryptic peptides, such that the stoichiometry of phospho-occupation can be determined. Finally, the SRM/MRM approach has an additional role in PTM validation. Algorithms for site prediction based on tandem MS spectra from large-scale phosphoproteomics experiments are frequently incapable of precisely localizing the modification site. Since the MRM approach requires precise parent and fragment ion masses, it is often possible to include ions generated from the neutral loss of phosphate to precisely localize the modification site. Such information is crucial for molecular analyses requiring site-directed mutagenesis to determine the functional role of the phosphorylation event.

cTnI is a highly phosphorylated protein, with 14 modification sites now detected by MS, however the functional impact of most of these sites remains completely unknown. Only phosphorylation at Ser22 and Ser23, catalysed by protein kinases PKA, C, D and G, is reasonably understood and these have been demonstrated to influence interactions with troponin C (TnC). Given the extent of functional and structural modifications possible by
phosphorylating cTnI, an unbiased investigation of potential sites of phosphorylation allows for new biological variances to be uncovered for further investigation. Of particular importance is the level of interaction between proximal sites in the context of the folded protein and when integrated with binding partners. For example, does occupation of one phosphorylation site subtly alter the protein structure and thus make another site inaccessible? Zhang et al. have clearly shown that previously identified PKA sites (Ser-22, Ser-23) are less occupied by phosphorylation in failing hearts, while sites regulated by protein kinase C (PKC) display elevated phosphorylation. As described above, these data should correlate with reduced PKA / increased PKC activation, which can be defined by phosphorylation using PKA/PKC phospho-site-specific antibodies. Additional questions must also be raised. Since kinases rely on the presence of ATP to generate phosphate for protein modification, parallel metabolomics analysis would show the relative availability of this substrate in the donor and failing hearts. It is possible that the data reflect on a limited pool of ATP substrate and / or comparative expression levels / activity of PKA / PKC themselves as they require divergent substrates (cAMP and diacylglycerol respectively). The new hypothesis would then be to determine whether the failing heart demonstrates i) reduced / increased expression [or stability] of PKA / PKC; or ii) whether the activity of these proteins is modified.

The design of ‘proteotypic’ peptides for most MRM experiments is based on several factors that will provide the most accurate quantification. Typically, chosen peptides are tryptic in nature (since trypsin is usually employed to create complex peptide mixtures for large-scale analyses), contain no missed cleavages and are free from amino acids that can be modified by proteomics sample processing (e.g. methionine, cysteine). The current study for example, employs the peptide [NITEIDLTQK] as a control as this perfectly fits the above criteria, and
contains two acidic amino acids (Glu and Asp) that generate intense fragment ions that are
perfect for MRM analysis. No such luxuries are possible for MRM quantification of peptides
containing modification sites. Therefore, the influence of the amino acid residues surrounding
the modification must be considered. For example, an increase in phosphorylation or other PTM
immediately proximal or distal to the tryptic cleavage site may result in steric hindrance of
trypsin, incomplete digestion, and cause a missed cleavage event. In MRM, the true tryptic
peptide would thus appear ‘down-regulated’, even though the phosphorylation event is increased.
It is not possible with PTMs to ‘choose’ the most appropriate tryptic peptide best suited for
MRM analysis. For example, the cTnI peptide containing Ser-198 (NIDALSGMEGR) also
contains a readily oxidizable methionine residue that can alter the abundance of both the
phosphopeptide and unmodified peptide (Figure 2).

The major caveat regarding SRM/MRM analysis is the loss of contextual information at
the protein-level; however this is true of most ‘bottom-up’ or ‘shotgun’ proteomics strategies that
rely on peptide-level information. In consideration of cTnI, 14 phosphorylation sites provide a
vast scope for protein-level variants generated by many possible combinations of occupied and
unoccupied sites. In fact, at the mathematical level and not considering constraints regarding
protein structure, site accessibility and kinase activity, cTnI has over 87 million possible variants
(including only combinations of the 14 phosphorylation sites) at the protein level. Protein-level
information is lost following proteolytic digest, such that it is impossible to know whether a
modification site in the C-terminus occurs concurrently with one in the N-terminus of a protein.
If the structure of the protein is known, it is possible to hypothesize about the relative proximity
of two such residues within the context of the whole protein, however only ‘top-down’
strategies are capable of generating information that can determine whether two disparate
PTM sites are modified in the same protein variant. Top-down analysis of a cTnI variant revealed differences in phosphorylation stoichiometry between Ser-22 and Ser-23 when compared with MRM analysis. Many factors could contribute to this anomalous result, including individual patient variance, method of cTnI purification and sample preparation, kinase / phosphatase activity and inhibition, phosphorylation site rearrangement during sample and MS processing and finally, protein expression and turnover compared with phosphorylation kinetics. Overall however, both approaches are useful, with bottom-up proteomics generating site-specific information at very high resolution, while top-down methods retain PTM information in the protein context.

While Zhang et al. examined phosphorylation within cTnI, it is also critical to note that there is increasing evidence of protein PTM cross-talk within proteins. PTMs as diverse as lysine acetylation, O-linked ß-N-acetylglucosamine (O-GlcNAc), complex N-linked glycosylation, deamidation, methylation, ubiquitination and many others are also possible modifiers of myofilament-associated proteins. For example, in addition to phosphorylation, cTnI itself is also modified by O-GlcNAc, acetylation, oxidation and proteolysis. There is also increasing evidence that positional space is impacted by the presence of a PTM. For example, the presence of O-GlcNAc may render a kinase motif inaccessible and thus render phosphorylation impossible, until either new protein is synthesized, or the competing modification is removed. Phosphorylation of cTnI at Ser-198 for example, may also play a protective role against C-terminal proteolysis, further demonstrating how dynamic PTMs influence not only each other, but the overall protein structure/function relationship. It is near certain that in coming years, a more substantial role for PTM cross-talk will be established, potentially leading to the discovery
of a new level of cellular regulation akin to the identification of miRNAs and their role in cellular systems.

Ultimately, the clinical utility of this novel approach to PTM identification and quantitation is in the ability to design protein- and combinatorial site-specific assays that could provide a view on cardiovascular health. For example, the complex pattern of cTnI phosphorylation is dramatically altered in the failing heart and is likely to be modified differently in acute conditions, such as ischemia / reperfusion, where the generation of protein fragments also influences the modification pattern. cTnI itself is already a diagnostic juggernaut and thus the next challenge is to determine whether modification patterns associated with healthy and diseased tissue are also released into the circulation, thus providing an opportunity for monitoring cardiovascular health by examining patterns of modification on released cTnI.

**Conflict of Interest Disclosures:** MYW is an Australian Research Council Discovery Early Career Research (ARC DECRA) Fellow.

**References:**


Figure Legends:

Figure 1. Complementarity of unbiased and biased approaches to phosphoproteomics for understanding cell signal pathways. Using the MAPK pathway as an example, unbiased phospho-enrichment techniques (A) allow for the capture of novel (N; green square) and known (K; red square) sites of phosphorylation. Surveying the pathway in a non-targeted fashion allows for the discovery of new sites regulated with pathway activation at the potential cost of missing known targets; (B) antibodies directed against these known sites of phosphorylation (W; blue square) allow for additional pathway components to be mapped, however the limited number of available antibodies means additional techniques are required; (C) SRM/MRM uses unique precursor and fragment ion pairs (transitions) to specifically target peptides containing sites of interest (S; orange square) to complete pathway mapping. The ability to design an MRM assay characteristic of the unique phosphopeptides is essential for complete pathway coverage.

Figure 2. Key considerations to designing MRM transitions for phosphopeptides. The C-terminus of human cTnI poses potential challenges given the proximal location of serine (phosphorylated) and methionine (oxidized). This means that for the proteotypic peptide, 4 series of transitions are required to determine the relative levels of each variant and the associated reaction kinetics. The string of lysine (K) at the C-terminus of the proteotypic peptide further confounds this, increasing the likelihood of missed cleavage events in addition to the modification status of the peptide.
Figure 1
Figure 2
Targeted Proteomics for Determining Phosphorylation Site-Specific Associations in Cardiovascular Disease
Stuart J. Cordwell and Melanie Y. White
Supplementary Figure 1
Supplementary Figure 1. Representative MRM assay of a peptide compared across acute and chronic conditions. The success of an MRM assay to monitor and quantitate unique proteotypic peptides or phosphopeptides relies on careful design and confirmation of transitions by tandem MS (a); (b) LC retention time and signal intensity of chosen transitions; (c) relative quantification of the 6 selected transitions (y11-y6) across the biological samples shows changing levels of the peptide in the chronic disease state. y- and b-ions presented in (a) result from fragmentation of the peptide backbone. y-ions are produced when the charge is retained on the C-terminus and b-ions produced with retention of the charge at the N-terminus of the peptide.