Antibody-free Affinity Enrichment for Global Methyllysine Discovery

by

Charlotte Dewar
BSc, University of Victoria, 2017

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Chemistry

Supervisory Committee

Antibody-free Affinity Enrichment for Global Methyllysine Discovery

by

Charlotte Dewar
BSc, University of Victoria, 2017

Supervisory Committee

Dr. Fraser Hof, Department of Chemistry

Supervisor

Dr. Jeremy Wulff, Department of Chemistry

Departmental Member

Abstract

Lysine methylation is a post-translational modification that regulates a large array of functionally diverse processes that are vital for cellular function. The role of methylation is best characterized on histone proteins due to their high concentration in the cell, but alongside histone modifications, lower abundance non-histone methylation is emerging as a prevalent and functionally diverse regulator of cellular processes. The direct biological impact of non-histone lysine methylation is less well understood because they are difficult to detect. The dynamic concentration range of the proteome masks their signal during proteomic analysis which impedes the detection of these low abundance methylated proteins. Increasing the concentration of proteins bearing methylation is required for improved discovery. This requires enriching the post-translational modification with a capturing reagent prior to analysis.

This thesis details an optimized method for using the supramolecular host *p*-sulfonatocalix[4]arene as a stationary phase methyllysine enrichment reagent for real-life cell-extracted proteins. Prior to the optimizations described in this thesis, cell-derived peptide extracts were not retained within an early generation upper-rim modified calixarene column. But with the new protocols detailed in this thesis, proteins extracted from both cultured prostate cancer cells and industrially sourced brewer's yeast were successfully retained by a lower-rim modified calixarene column. Thousands of methylated proteins with diverse functions and cellular localization were discovered using this method. Detection of low abundance methylated proteins will aid our discovery of all cellular methylation marks, which in turn, will help delineate their biological functions.

Table of Contents

List of Tables

List of Figures

Abbreviations

Apical complex lysine methyltransferase	(AKMT)
Association constant	(K_a)
B cell receptor	(BCR)
Biological replicate	(BR)
Bromodomain PHD Finger Transcription Factor	(BPTF)
Biotin-conjugated phenylglyoxal	(Biotin-PG)
Chromobox protein homolog 6	(CBX6)
D-adenosyl-L-methionine	(SAM)
Dithiothreitol	(DTT)
DNA methyltransferase 1	(DNMT1)
Electron transfer flavoprotein β subunit	(ETFB)
Electrospray ionization	(ESI)
Enzyme-linked immunosorbent assay	(ELISA)
1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride	(EDC)
False discovery rate	(FDR)
Gene ontology	(GO)
Glideosome-associated connector	(GAC)
High pressure liquid chromatography	(HPLC)
Histone 3	(H3)
Histone 4	(H4)
Histone 3 lysine 9	(H3K9)
Histone 3 lysine 27	(H3K27)
Horse radish peroxidase	(HRP)
Human prostate cancer cell line	(PC-3)
Iminodiacetic acid	(IDA)
Immobilized metal affinity chromatography	(IMAC)
Iodoacetamide	(IAA)
Keyhole limpet heomocyanin	(KLH)
Liquid chromatography mass spectrometry	(LC-MS/MS)

Lysine	(K)
3x malignant brain tumor domain	(3xMBT)
Malondialdehyde	(MDA)
Methyltransferase	(KMT)
Nitrilotriacetic acid	(NTA)
O-phthalaldehyde	(OPA)
P53 Binding Protein 1	(53BP1)
Polycomb Repressive Complex 1	(PRC1)
Polyethylene glycol	(PEG)
Plant homeodomain	(PHD)
Protein arginine deiminases	(PAD)
Post-translational modification	(PTM)
Recombinant antibodies	(rAbs)
SET and MYND domain-containing protein 2	(Smdy2)
Small ubiquitin-like modification	(SUMO)
Sodium dodecyl sulfate polyacrylamide gel electrophoresis	(SDS-PAGE)
Stable isotope labelled amino acid cell culturing	(SILAC)
Technical replicate	(TR)
1,1,3,3-tetraisopropoxypropane	(TiPP)
Ubiquitin-like, PHD and RING finger containing 1	(UHRF1)

Acknowledgments

I would like to thank the Hof group for all the laughs, memories, and guidance they gave me throughout this process. I would especially like to thank Mark Grasdal for teaching me the ropes of this project and helping with the optimization process. I wouldn't have been able to do it without you! I would like to thank the Starfish staff, Becky Hof, and Tyler Brown for always brightening my day in the Bioroom even on the worst days. I would also like to thank Chelsea Wilson for always making me smile and being the best fume hood buddy. Thanks for helping me set up my condenser! Also, I would like to thank Sean Adams for always fixing the cold trap in record time.

A huge thank you to Dr. Fraser Hof for guiding me to be a better scientist and giving me memories that I'll cherish forever! Thank you for hosting great lab parties and fun lab tradition like bocce ball and the Volcano conference. Volcano was a one-of-a-kind conference I will never forget. It was very... educational! Thank you for helping me become a better writer. I now know the importance of topic sentences and a solid outline. Thank you for all the experiences and opportunities you've given me. Your mini-lessons and life advice in group meetings were insightful and appreciated.

Thank you to Derek Smith and Darryl Hardie at the Proteomics center for always answering my hundreds of emails and getting my samples processed in half the time they'd quote. Thank you to Phillips Brewing Company for all the free beer and fun times in your lab. Thank you to Ben Schottle, Alex McDonald, Dr. Euan Thompson and everyone at Phillips for their positivity, fun, and cool conversations. The time I spend at Phillips was always a blast.

Thank you to Hannah Reid for always showering me with love and support throughout this journey. And for lending me your computer when mine broke. This thesis would have taken a lot longer if it wasn't for you. I would also like to thank my parents, Bob and Jen, for being such solid people in my life, offering me the advice I needed to get through this degree and thesis, and for always having my back. I'm so lucky to have such amazing people apart of my life.

Dedication

This thesis is dedicated to my parents, Bob and Jen.

Chapter 1: Proteomic Analysis of Lysine Methylation

1.1 Histone and non-histone post-translational modifications

1.1.1 Post-translational modifications

Post-translational modifications (PTMs) control a protein's biological function by controlling its stability, functionality, and affinity for other proteins. PTMs are covalent modifications to pre-existing amino acid functional groups that change a residue's polarity, charge, geometry, and/or hydrophobicity.[1] With such modifications, the cell can expand the physicochemical reserve of the twenty amino acids to highly complex molecular states. The list of major PTMs includes methylation, phosphorylation, acetylation, citrullination, ubiquitination, and small ubiquitin-like modification (SUMO)ylation.[2] Enzymes that add and remove these PTMs have many known connections to signal transduction pathways, gene expression, and disease.[3]

PTMs are broadly employed throughout the cell. For example, p53 is positively regulated by lysine methylation.[4] In correlation with stress signals such as DNA damage, the methyltransferase protein Set9 was observed to methylate p53 at K372. In its methylated form, p53 is more stable, is able to induce p53-mediated apoptosis, and localizes in the nucleus. Another use of PTMs in the cell is acetylation and/or methylation on histone proteins. Specific arrays of such modifications are correlated to RNA synthesis rates (gene activating and silencing), cellular metabolism, and protein chaperoning.[5,6,7] It is thought that changes to the array of PTMs on histone proteins allow the cell to adapt to environmental stimuli in a heritable fashion—this phenomenon is termed 'epigenetics'.

1.1.2 Interplay of epigenetics and PTMs

Epigenetics is classically defined as the inheritable changes to gene expression rates without changing the underlying DNA sequence itself.[8] This means that specific gene expression patterns formed within a cell's lifetime can affect and be passed on to the offspring rather than solely the available genes. This concept extends the neo-Darwin theory of evolution where Mendelian genetics and mutations dictate evolution and natural selection—evolution is now known to be passed as expression patterns based on selective pressures.[8] Lenin *et al.* observed gene expression alterations in DNA-methylation patterns (a cellular method for gene silencing)[9] in *Arabidopsis* plants when starved of phosphate.[10] In this instance, changes to gene expression

rates were passed on to that cell's progeny. This exemplifies how an organism alters its gene expression in response to environmental stress such as starvation conditions, and the memory of such stress allows the future cells to better respond to such stress.

Alterations to gene expression are common even without environment pressure. A classic example is the phenotypic difference between specialized tissues in multicellular organisms. The genes in a person's body are all identical, but specific patterns of gene expression in specialized tissues (such as our skin or our liver) allow the cells to perform specialized tasks (like providing a protect barrier for our internal organs[11] or metabolism[12], respectively). This specialization occurs during cellular differentiation and embryonic morphogenesis.[13,14] General gene expression patterns of specialized tissues, determined by DNA methylation, are (for the most part)[15] locked in for the duration of such cells' subsequent divisions.

DNA methylation directly impacts expression levels of a gene, but it is PTMs on proteins associated with chromatin that direct DNA methylation to occur. Rothbart *et al.* linked the maintenance of DNA methylation to lysine methylation on histone 3 lysine 9 (H3K9).[16] This PTM colocalizes the protein human ubiquitin-like, PHD and RING finger containing 1 (UHRF1) to the site during the S-phase of mitosis, which recruits DNA methyltransferase 1 (DNMT1). DNMT1 can then propagate the DNA methylation pattern to the duplicated strand of DNA. Thus, the DNA methylation mark is inherited by the daughter cell. Methylation is one means for inheritable transcriptional control, but other modifications such as acetylation or phosphorylation are also implicated in epigenetic regulation and transgenerational epigenetic inheritance.[17]

1.1.3 The Histone Code

The histone code hypothesis describes transcriptional regulation as arising from specific sets of PTMs on histone tails.[18] These modifications can recruit proteins to a nucleosome by providing docking sites for PTM binding motifs within an effector protein, and/or simply altering the DNA-histone contact energetics.[19] A PTM barcode simultaneously determines the interaction affinities between DNA, histones, and chromatin-associated proteins.

Schübeler *et al*. linked binary patterns of histone modification with general gene activity on euchromatic chromatin.[20] They found a strong correlation between euchromatic (loosely packed chromatin with active gene expression) transcriptional activities with specific arrays of histone modifications. General hyperacetylation of histone 4 (H4) and histone 3 (H3) with hypermethylation at positions K4 and K79 of H3 was generally associated with active genes,

whereas general hypoacetylation and hypomethylation at such sites was associated with inactive genes. The rational for the observed difference in gene expression rates is that covalent modifications on histone residues consequentially modify the nucleosome's physiochemical properties. As a result, PTMs adjust how the histone proteins within the nucleosome interact with proximal material. Lysine acetylation increases nucleosome solubility and fluidity by neutralizing the DNA-to-lysine charge-charge interaction.[21] This allows transcription initiating proteins to bind and initiate transcription at a gene's promoter. Lysine methylation is thought to provide a binding site for the transcriptional machinery and increase gene expression.[22]

Lysine methylation is a complex process that has contextual functions on histone proteins. Methylation does not change lysine's charge state, so the effect of methylation on chromatin structure does not occur due to changes in electrostatic interactions. Of greater influence, lysine methylation provides a recruitment site for gene-regulating effector proteins. Proteins such as Polycomb homologs are known to bind methylation sites on histone proteins.[23] The degree and position of methylation relative to a gene has an effect on transcription along with the type of chromatin. Additionally, the relative position and array of other PTMs also effect a chromatin's gene expression.[24]

The exact mechanism of gene activation and gene silencing with the histone code is still up for debate. Many papers report conflicting roles of histone methylation. In a paper by Grewal *et al.*,[25] hypermethylation of heterochromatin (transcriptionally repressed tightly packed chromatin) had the opposite effect of hypermethylated euchromatic chromatin. Heterochromatin hypermethylation was noted to cause gene silencing and euchromatic hypermethylation was found to cause gene activation.[25] These findings contradict one another in terms of methylation state and gene activity. It may be that methylation serves opposing functions for heterochromatin and euchromatin.[23] Another hypothesis for the conflicting role of histone lysine methylation is the influence of the methylated histone's position relative to its regulatory gene. A study by Schneider *et al.* found that lysine methylation that occurred on the promotor region of a gene evoked gene silencing, while methylation at the coding region of the gene evoked gene expression.[26] Overall, classifying casual relationships between histone lysine methylation and function is difficult to achieve directly.

1.1.4 Non-histone modifications

Gene regulation by lysine methylation is not restricted to histone modifications.[27] Non-histone proteins offer transcriptional control—such as methylation of the transcription factor p53 (as described in section 1.1.1). Gene expression is systemically regulated by non-histone PTMs such as phosphorylation[28], acetylation[29], glycosylation,[30] SUMOylation,[31] and methylation.[32] Histone PTMs alter how tightly transcription factors bind a nucleosome, but it is the localization of the transcription factors to the nucleus and DNA that allows transcription to initiate.[30] Moreover, there is evidence that transcription factor PTMs are what predict the histone code.[33] Binary histone codes have been correlated to gene expression rates, but the causality of gene expression is the transcription factors that bind a gene's promoter or deposit the histone PTMs.[34,35] Histone modifications fine tune expression by localizing the transcription factors to their cognate binding sites achieved with altered chromatin accessibility and molecular structure.[36] The binding of transcription factors and chromatin-modifying enzymes are what facilitate transcription.[37,35]

Relative to histone methylation, non-histone methylation sites are scarcely studied. Parallel to our understanding of histone methylation, non-histone methylation is thought to regulate protein-protein interactions.[38] However, the context and functional significance involved with non-histone methyl PTMs are largely unknown. Accumulating evidence has revealed the role of certain non-histone lysine methylations in the development and maturation of cancer. For example, alongside the p53 methylation discussed in section 1.1.1, methylation at K370 by 'SET and MYND domain-containing protein 2' (Smdy2) was also found to regulate p53 function.[39] Dysregulation of Smdy2 p53 methylation inhibits apoptosis—a smart mechanism cancer uses to evade death. This mono-methylation occurs at higher-than-average abundance within teratocarcinoma cancer cells.[40] Zhu *et al.* explored the repression of wild-type p53 methylation within the teratocarcinoma cell line with lysine methylation. It was seen that NTera2 cancer cells had an abundance of p53 but the downstream p53-activated products of were not present. The hypothesis was that p53 lysine methylation by the methytransferases SMYD2 and PR-Set7 was repressing the protein's transcriptional activation which results in oncogenic proliferation. Substitution of the lysine methylation sites or knock-downs of the methytransferases re-activated p53 (measured by the rescue of p53-activated gene products). This study is a beautiful example of how non-histone lysine methylation is a key player in gene expression regulation and is necessary in the maintenance of healthy cellular function.

1.2 Lysine methylation

1.2.1 The degrees of lysine methylation

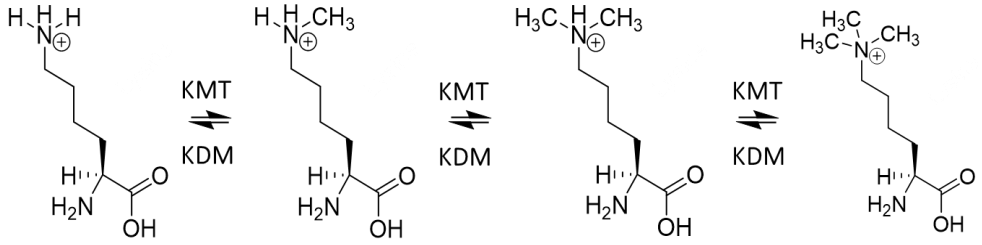


Figure 1.1 The three degrees of lysine methylation: mono, di, and trimethylation. KMT is a lysine methyltransferase that installs the methyl moiety, and KDM is the lysine demethylase that removes the methyl moiety.

Lysine methylation is a PTM that can occur in three degrees: mono, di, and tri (Figure 1.1). The enzyme lysine methyltransferase (KMT) catalyzes the transfer of the methyl moiety from the cofactor S-adenosyl-L-methionine (SAM) to the substrate lysine. The stepwise addition of methyl groups to lysine allows a controlled increase in the residue's size and hydrophobicity when transitioning from the unmethylated state to a mono, di, then tri-methylated state. Additionally, each degree of methylation removes an N-H hydrogen bond donor. Each degree of methylation encoded on the same lysine yields a unique biological outcome—and the degree of methylation is thought to be contextually regulated by the availability of SAM and the balanced enzymatic activity of both methyltransferases and demethylases.[41,42]

1.2.2 Biological recognition of lysine methylation

Lysine methylation provides a transient binding site for protein-protein interactions. Proteins with aromatic binding pockets (termed 'aromatic cages') can bind methyllysines (Figure 1.2).[43] The bound protein can then provide a platform where a protein complex may scaffold—which in turn relays signals downstream in a variety of signalling cascades.

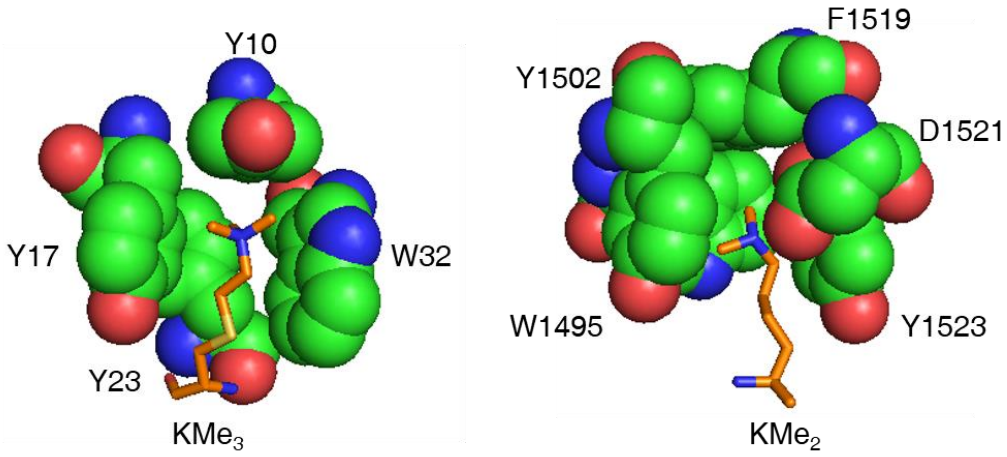


Figure 1.2 Aromatic cage of Left: BPTF PHD-bromodomain module bound to H3KC4Me$_3$ methyl lysine analog (PDB code 6AZE) and Right: Tudor domain of royal family protein 53BP1 bound to H4K20Me$_2$ (PDB code 2IG0). Image created using PYMOL.

Two large classes of methyllysine binding proteins are currently known: (1) the royal superfamily and (2) plant homeodomain (PHD) zinc fingers. Figure 1.2 (left) shows the aromatic cage of the PHD zinc finger protein, Bromodomain PHD Finger Transcription Factor (BPTF) and Figure 1.2 (right) shows that of the royal family protein P53 Binding Protein 1 (53BP1). In spite of their dissimilar sequences, the royal superfamily shares remarkably similar recognition features to the PHD zinc finger. Both effector modules have convergently evolved aromatic cage recognition motifs, which recognize the lysine methylation modification.[44]

Cation-π interactions between the aromatic residues and the methylammonium group largely mediate this energetically favourable interaction. This cation-π interaction occurs due to a partially negative quadrupole moment hovering above the face of the aromatic residue. This allows a perpendicular Coulombic attraction between the face of the aromatic and the methyllysine cation. Interaction distances are analogous to a van der Waals attraction.[44,45] The trimethyllysine bound to the aromatic cage in Figure 1.2 left shows aromatic residues (Y10, Y17, Y23, and W32) lining the inner cavity to facilitate multiple cation-π interactions. The dimethyllysine bound to the aromatic cage in Figure 1.2 right is engaged by four aromatic residues (W1495, Y1502, F1519, Y1523) to form the cation-π interactions and to form a hydrophobic cavity.

The aromatic cages are selective for different degrees of methylation because each methylation increases the methylammonium's hydrophobicity and size. The large and open shape of the BPTF has more solvent accessibility which allows di and trimethyllysine-mediated displacement of frustrated water molecules (the hydrophobic effect).[46,47,48] Selectivity for mono

and dimethyllysines over trimethyllysine is caused by steric exclusion and hydrogen-bonding.[47] The narrow cavity of the 53BP1 domain makes this aromatic cage selective for lower degrees of methylations based on size exclusion.

The hydrogen bonding potential of methyllysines also alters selectivity between the degrees of lysine methylation. The unmethylated lysine has three donors for hydrogen bonding, monomethyl has two, dimethyl has one, and trimethylated lysine has lost all its hydrogen bonding potential. The carboxylate oxygen atom on an acidic residue (D1521) of 53BP1 acts as a hydrogen bond acceptor for the dimethyllysine's one remaining N-H hydrogen bond donor (Figure 1.3). This interaction is not possible for trimethyllysine because it lacks hydrogen bond donors, and this contributes to 53BP1's selectivity for binding dimethyllysine.

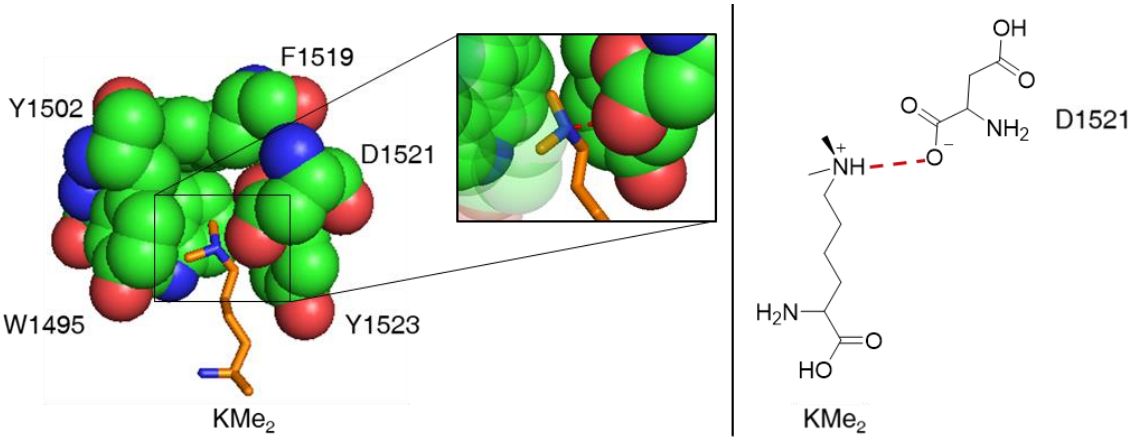


Figure 1.3 Left: Zoomed in hydrogen bonding (red dashed line) between tudor domain of royal family protein 53BP1 bound to H4K20Me$_2$. Image created using PYMOL (PDB code 2IG0). Right: Structure of hydrogen bonding (red dashed line) of dimethyllysine with aspartic acid.

1.2.3 Future directions for lysine PTM research

The methylome (defined for this thesis as all protein methylation sites in the cell) is not fully characterized. Given that the number of known methyltransferase enzymes[49] is comparable to the known number of kinases[50], it is expected that the number of substrates bearing each PTM should be comparable. Yet, the number of known phosphorylated sites (~57k) outnumbers the known methylated sites (~7.4k) by almost 8-fold.[51] This discrepancy occurs due to our lack of reagents that allows analysis of methyllysines on a large-scale—phosphorylation has many reagents that permit broad-spectrum discovery of this PTM.[52] Analogous to the discrepancy between known methyltransferase and kinase substrates, the role of phosphorylation with cellular, developmental, and pathogenic processes is far better characterized than that of methylation.[53,32] The elusive role of methylation in the cell is largely derived from its incomplete landscape. To

comprehend the functional annotation of methylation, a complete picture of the methyl proteome must be characterized. The fact that methylation has comparable dynamics and number of installation enzymes to those of phosphorylation, and also demonstrates cross-talk with other modifications,[32] strongly suggests that its cellular role is universal and fundamental. A better understanding of the methyl landscape is predicted to shed light on its mechanistic characteristics and its roles in disease. As with phosphorylation, methylation could be a target for disease diagnosis and therapy.[38] Finding more methyl PTM substrates will help us better understand the many roles of methyl PTMs within the cell.

1.3 Proteomics to Analyze PTMs

1.3.1 Data-dependent acquisition of the proteome with LC-MS/MS analysis

PTMs are predominantly discovered by liquid chromatography tandem mass spectrometry (LC-MS/MS). Proteins are isolated and purified, enzymatically cleaved by a protease, separated by LC, and then ionized by electrospray ionization (ESI) for mass analysis in a tandem MS instrument (Figure 1.4).

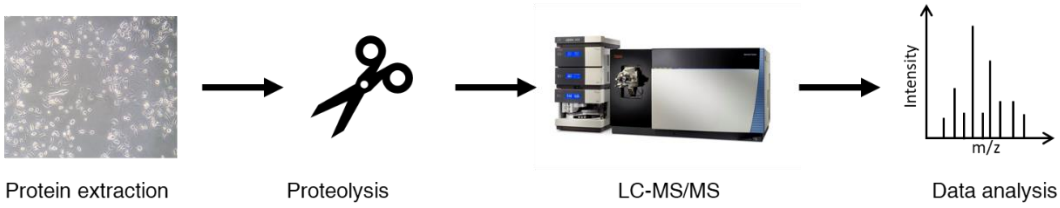


Figure 1.4. Schematic workflow of proteomic analysis.

Within the MS/MS analysis, precursor ions with the highest signal intensities are selected to undergo fragmentation, then are re-analyzed by a second mass analyzer as the product ions. The m/z of the precursor and product ions are then aligned to a pre-made 'target' database that contains a given set of protein spectra. The target database contains the predicted spectra of the sample's known proteome that has been cleaved with a specified protease with an allowed set number of tolerated missed cleavages and variable PTMs. The precursor spectrum is compared to the target database to generate a list of matches, then the confidence of these matches is scored by aligning the product ion spectrum to the matches' predicted fragmentation pattern (Figure 1.5). It is important that the proper database be used for proteomic data analysis. If an inappropriate database is used, then the findings do not carry validity.[54]

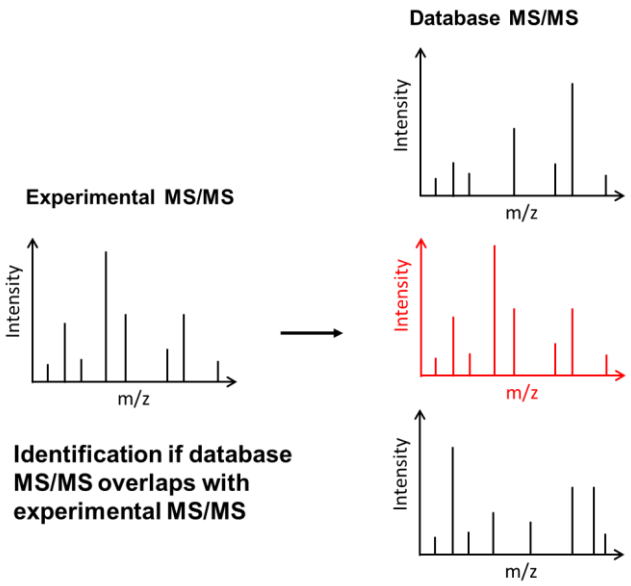
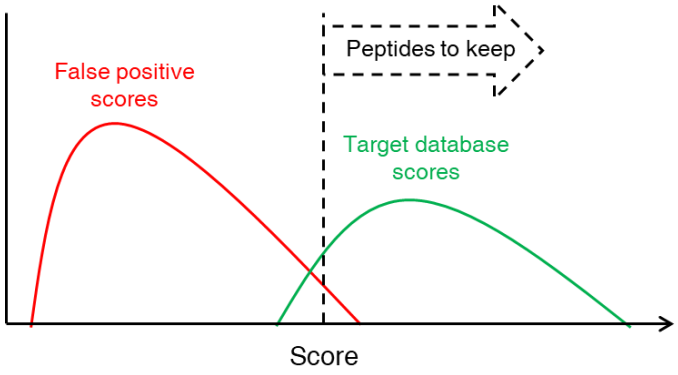


Figure 1.5 Workflow of MS/M-based peptide identification in data-dependent acquisition.

The score represents the statistical level of confidence one has that the protein hit is in-fact the aligned protein identification. The higher the score, the better the match is between the theoretical product ion spectrum and the more intense peaks within the experimental product ion spectrum.

Incorrect protein identifications are of great concern within proteomics. As such, additional statistical validation methods are implemented to increase the stringency of protein hits. False positive identifications are filtered out by assigning the same dataset a 'false positive score'. This is achieved by re-aligning the spectra to a 'decoy' database.[55] The decoy database is a reversed or scrambled version of the real target database. The protein hits that align to the decoy database are considered false positives, and these identifications are assigned false positive scores. Once the experimental data has been aligned to both the target and decoy databases, a false discovery rate (FDR) is assigned at a given score.[55]



$$FDR = \frac{\#\ decoy\ hits}{\#\ target\ hits}$$

Figure 1.6 Scoring function of decoy and target hits to filter out false positive identifications.

The FDR is usually set to ~0.1 (dashed line in Figure 1.6). This sets the threshold score at which target hit scores are accepted as true identifications. With this logic, ~10% of the target identifications will be false positives based on this target-decoy method. This method depresses the number of low-scoring identifications resulting from random false matches based off of chance.

1.3.2 The dynamic concentration range of the proteome

Discovering and analyzing novel PTM marks is challenging. In order to discover novel methylation sites, the methylated peptides must be detected in LC-MS/MS analysis. Within the dynamic concentration range of proteins within the proteome, PTMs occur on a small subset of the proteome. And once the complexity of the sample exceeds the instrument's duty cycle, only the highest abundance peptides from the precursor spectrum undergo fragmentation.[56] Thus, the low abundance PTM peptide signals are masked in MS/MS experiments by the signals from more abundant unmodified peptides.[57] In order to analyze the signal of PTM proteins, one must increase their concentration above the highly abundant unmodified peptides. Because of this, PTM enrichment is vital for PTM protein discovery.

1.4 Antibody-based PTM enrichment

1.4.1 Overview of antibodies

Antibodies are composed of two disulfide-linked polypeptides that form a Y-shaped glycoprotein of ~150kDa (Figure 1.7).[58] The 'variable domain,' controls antigen specificity (Figure 1.7 right, blue). The constant domain (Figure 1.7 right, orange) is recognized by the host's immune system and signals the appropriate immunological response (Figure 1.7 right).

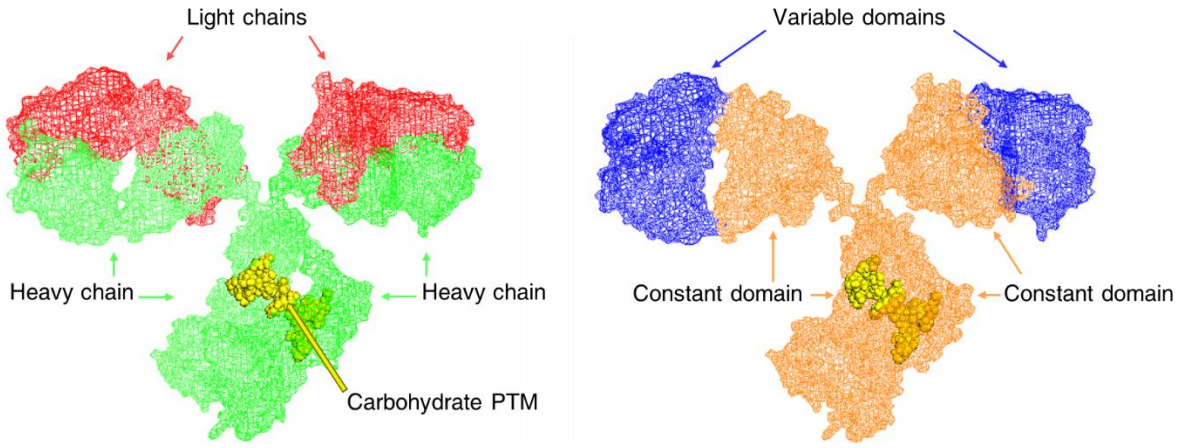


Figure 1.7 Colour coded antibody structures of Left: light (red) and heavy (green) chains and Right: Variable domains (blue) and constant domains (orange). Carbohydrate PTMS are labelled yellow. Image created with PYMOL (PDB code 1IGT)

Antibodies are a part of the adaptive immune response of our humoral immunity. Matured B cells secrete (or remember how to secrete) the antibodies. Before B cells obtain the ability to secrete antibodies, the antibody is first expressed as a membrane protein (termed 'B cell receptor' [BCR]). The BCR remains immature until a foreign antigen stimulates maturation.[59] Each individual B-cell produces many antibodies with an identical variable region sequence/antigen specificity. There are millions[60] of these immature B cells with unique variable domains that reside in our secondary lymphatic organs awaiting an antigen to cascade their activation response. When a foreign antigen binds their BCR with sufficient affinity, the immature B cell proliferates and differentiates into an effector cell, and then a plasma cell—both of which can secrete a soluble version of the BCR (now called antibody). Genomic rearrangement machinery[61] involved in immunoglobin class switching changes out the BCR's constant region for henceforth soluble antibody production.

1.4.2 Antibody production for research applications

Antibodies are widely utilized within research. These proteins can be manufactured in-lab[62] for use as specific molecular detection probes or to discover novel protein modifications. To manufacture antibodies, the immunogen that would be probed must be produced; then an animal must be immunized with such immunogen.[62] Subsequently, the animal's antibodies are collected, screened for desirable antigen specificity, and purified. From this process, a mixed population of antibodies (termed 'a polyclonal antibody') that bind different areas of the same immunogen

would be produced. Different polyclonal antibodies are produced from each different immunization of an animal, even when the same immunogen is used.[63]

This process becomes more laborious and costly when the goal is to isolate and produce a population of identical antibodies. These are called 'monoclonal' antibodies—antibodies derived from one unique B-cell lineage.[63] To produce monoclonal antibodies, an immortalized B-cell is produced by fusing mature B-cells membranes with immortal myeloma cells.[64] Then, antigen specificity of all the successfully produced hybridoma cells would be assayed. Clones that produce the antibody that bind the epitope of interest are kept. The stock cells produce high quality antibodies with the intended specificity, but long-term hybridoma propagation can produce issues. B-cells are living dynamic organisms that are susceptible to genetic mutations and rearrangements. Over time, the antigen specificity of monoclonal antibodies changes.[64] Genome sequencing and antibody specificity must be routinely checked (though it seldom is).[65]

The problems outlined above for polyclonal and monoclonal antibodies are a large part of the 'reproducibility crisis' of biomedical research—the inability to reproduce biomedical experiments because of lot-to-lot divergence of antibody specificity. Faulty antibodies are either non-specific for the intended target, or (even worse) specific for a different protein.[65,66] The divergence of antibody specificity is emerging as a major issue for researchers who rely on these antibodies.

Recombinant antibodies (rAbs) have emerged as an alternate method to produce monoclonal antibodies without the use of hybridoma cells. With the use of synthetic genes, these proteins are produced *in vitro*. The B-cell's antibody gene is amplified and cloned into a phage vector/plasmid, then expressed in a host (usually *E. coli*).[67] Unfortunately, the cloned antibody product lacks the proper placement of the glycosylation (a PTM) that is vital for proper function (refer to Figure 1.7, yellow).[68] As a result, recombinant antibodies may recognize an antigen, but maybe not the one that was screened from the original B-cell. rAb production by a mammalian host produces the "human" glycosylation patterns[69] but currently this expression system is costly, and the yields aren't ideal yet.[70] Regardless, this system produces monoclonal antibodies that are useful for research.

1.4.3 Pan-specific antibody enrichment

Enrichment using pan-specific antibodies is important for PTM discovery. The previous section detailed the production of polyclonal or monoclonal antibodies specific for an antigen of

interest. With polyclonal antibodies, 'pan-specific' antibody batches are produced—antibodies that can detect all sequences that hold a specific PTM, without discriminating between different peptide sequences surrounding the PTM. For example, pan-specific anti-acetyl lysine antibodies can be generated by immunizing rabbits with artificially acetylated ovalbumin and synthetic acetyl lysine peptides.[71] With the use of acetic anhydride,[72] acetyllysine carrier proteins and peptides are generated as antigens with highly variable flanking amino acids. Though the small acetyl moiety offers low antigenicity, its unique structure gives it enough variation in chemical and physical properties to be differentiated for antibody enrichment. Following immunization, a polyclonal pool of sequence-independent pan-specific antibodies are generated based on the sheer diversity of sequence recognition—a key factor since cellular acetylation does not have a single consensus sequence.

Even with the promise of immunoprecipitation associated with acetyllysine enrichment, antibody enrichment is still limited by their availability, quality, stability, reproducibility, and most problematically, their continual issue of sequence recognition bias.[73] The sheer diversity of the polyclonal cocktail makes the antibodies pan-specific, but not all possible sequences are covered in each batch. Pan-specific antibodies provide enrichment that varies from antibody to antibody in each polyclonal cocktail batch.[74] Thus, reproducibility is not well maintained from batch-to-batch. Moreover, antibody stability is another issue. Antibodies are highly susceptible to degradation and denaturing, making them an unstable research tool.[75]

1.4.4 Immunoaffinity enrichment of methyllysine for LC-MS/MS analysis

Immunoaffinity enrichment using pan-specific PTM antibodies permits discovery of low concentration PTM-bearing proteins in LC-MS/MS analysis. Of all the PTMs, lysine acetylation[76] is the main modification that can be enriched by immunoprecipitation. Methylation has also been enriched by immunoprecipitation,[77] but less successfully.

Antibody-based methods dominate the field of methyllysine enrichment. Animals are immunized using chemically methylated keyhole limpet hemocyanin (KLH).[78] But unlike acetylation, the methyl modification does not significantly alter the residue's charge and causes the smallest possible changes to its physicochemical properties. With that lies the challenge to develop methyllysine specific antibodies with reasonable selectivity. The methyl moiety does not provide high immunogenicity to produce selective methyllysine antibodies or antibodies that can differentiate between the degrees of methylation with high selectivity. This is because antigen

binding affinity is relative rather than absolute. If the difference in antigen affinity between the methyl and non-methyl antigen is low, then the antibody will compete for binding for both antigens with high cross-reactivity. Hence, promiscuity.

Binding affinity must be relatively strong to interact with the intended ligand over the competitive ligands. The small physiological difference between any degree of methylation versus the non-methyl variant does not pose enough difference for the antibody to properly distinguish between one another. The antibodies thus interact with 'off-target' proteins and skew research results.[79] Perez-Burgos *et al.* characterized some problems with methyllysine histone antibodies sold by Upstate Biotechnology, USA, and Abcam, UK.[80] They reported issues of off-target recognition, undesired sequence specificity, and issues with distinguishing between methylation states. Sub-optimal and/or less-specific batches were not reported by such companies. In this most thorough published side-by-side comparison of methyl-specific antibodies provided by multiple commercial sources, five anti-methyllysine antibodies were used to enrich from the same HeLaS3 cell extract.[81] Methyllysine antibodies performed worse than methylarginine immunoprecipitation. Fifty-four methyllysines were identified via immunoprecipitation compared to 254 methylarginines. Sub-cellular fractionation outperformed the methyllysine antibody enrichment which resulted in the identification of fifty-eight methyllysines. This demonstrates the limited detection of methyllysines with antibody enrichment and the issues of antibody methyl enrichment.

The only commercially available pan-specific methyllysine monoclonal antibody claiming to be specific for a single degree of methylation is sold by Cell Signalling Technology (Mono-Methyl Lysine [mme-K] MultiMab™ Rabbit mAb mix #14679). The company claims the mono-methyl lysine antibody to be pan-specific and selective with minimal cross-reactivity for the di and tri methylated lysine and methylated arginine. The manufacturer provided results from an enzyme-linked immune assay (ELISA) to demonstrate the specificity of the antibody and a western blot to demonstrate the selectivity. The ELISA results showed an 8-fold increase in signal intensity for the monomethyl lysine relative to the unmethylated and trimethylated lysine and methylated and unmethylated arginine. A 4-fold increase in signal for the monomethyl relative to the dimethyl lysine was also seen. From these ELISA results, the pan-specific monomethyl lysine antibody was selective for the monomethylated lysine, but also demonstrated binding to the di methylated lysine. As for western blot analysis of the antibody's selectivity, the company compared MCF7 cells treated and untreated with adenosine-2',3'-dialdehyde (AdOx) [a histone methyltransferase

inhibitor of lysine and arginine methylation]. GAPDH was stained with (D61H11) XP® Rabbit mAb #5174 as a loading control. In theory, if the antibody was selective for monomethylated lysines, then the treated sample bands would show lowered staining intensity due to inhibition of histone methylation. Both the untreated and treated lanes of the western blot had numerous bands with variable intensities between the conditions. Based on the western blot analysis, the pan-specific methllysine antibody is not selective for endogenous monomethylated lysine. Mono-Methyl Lysine [mme-K] MultiMab™ Rabbit mAb mix #14679 has no citations or reviews. Other examples of poor pan-methyllysine performance can be found in the online quality control repository www.histoneantibodies.com.[82]

1.5 Non-antibody PTM enrichment and discovery

Due to the shortcomings associated with antibody enrichment strategies, many antibody-free PTM enrichment strategies have been developed. This section will detail the antibody-free enrichment technology currently used for PTM analysis, with the main focus being the strategies developed for lysine methylation.

1.5.1 Protein pull-down enrichment

A naturally occurring sequence-independent methyl binding protein has been modified, developed, and used as a pulldown reagent. Gozani and coworkers established a pull-down protocol with the engineered protein 3x malignant brain tumor domain (3xMBT).[83] This domain has pan-specific affinity for mono- and dimethyllysine proteins. In tandem with stable isotope labelled amino acid cell culturing (SILAC), their protocol identifies candidate methylated proteins or compares methylation states between different biological conditions. Their enrichment is only effective with undigested proteins since protein capture of peptides demonstrates high binding promiscuity. Interactions outside the peptide region may elicit binding. As a result, this method analyzes the entire sequence of all pulled-down proteins. Hence, the exact site of methylation is not always seen in LC-MS/MS analysis due to the presence of the non-methylated peptides also arising from the pulled-down methylated protein.

1.5.2 Chemical derivatization for PTM enrichment

There is high promise for a protocol that combines chemical peptide derivatization with ion exchange-based enrichment.[84] Two reagents [malondialdehyde (MDA) and o-phthalaldehyde

(OPA)] are added to peptide digests, one each for reacting with and neutralizing unmethylated arginines and lysines (Figure 1.8 top).[84]

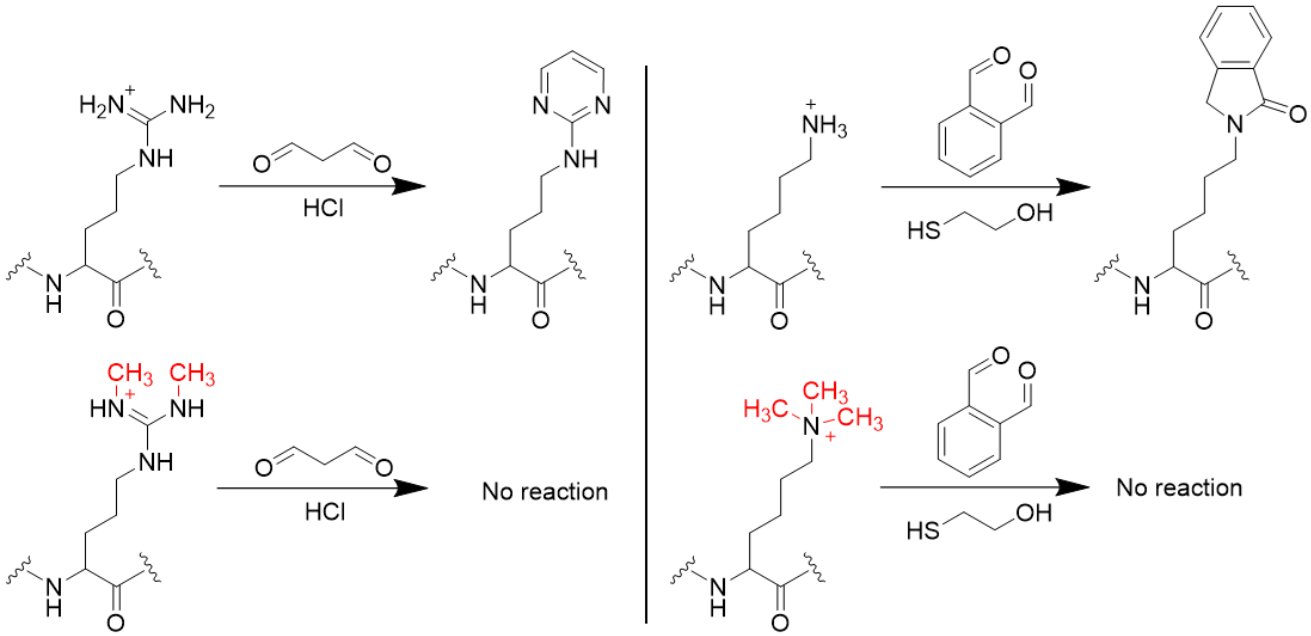


Figure 1.8 Top: Chemical blocking of arginine with malondialdehyde (left) and lysine with o-phthalaldehyde (right). Bottom: lack of reactivity of (left) malondialehyde with methyl arginines represented with symmetric di methylarginine and (right) o-phthalaldehyde with methyllysines represented with trimethyllysine.

Strong cation exchange chromatography then enriches the unreacted, still cationic methylated arginines and lysines (Figure 1.8 bottom). This method yields hundreds of methylated peptide hits, with a very desirable low sequence bias. Widespread adoption has been limited because the derivatization reagent, MDA, is unstable for storage. The highly acidic conditions of MDA derivatization and/or the basic conditions of OPA derivatization result in unwanted peptide hydrolysis.[85]

Chemical labelling of the PTM citrulline is a method for proteomic discovery and detection of citrullinated proteins.[86] In contrast to chemical derivatization with OPA and MDA, the PTM itself is modified for detection. Citrullinated proteins are labelled with biotin-conjugated phenylglyoxal (biotin-PG) to produce a citrulline specific probe (Figure 1.9).

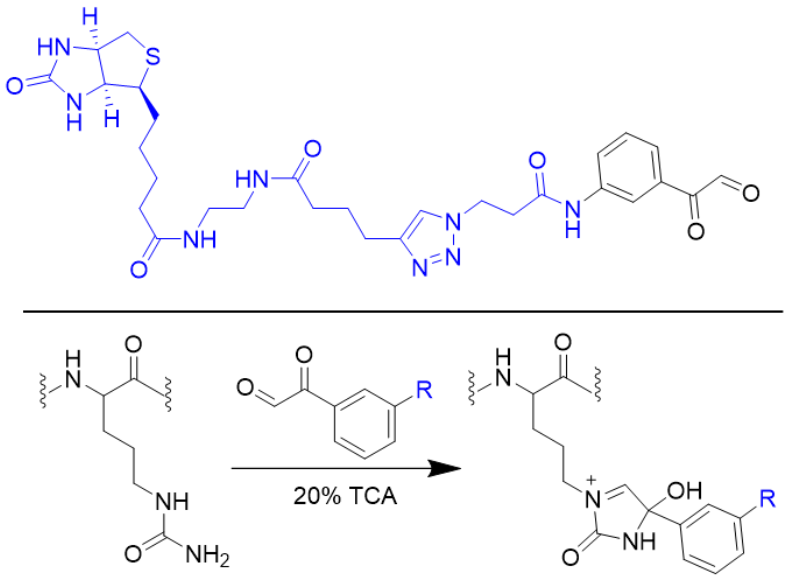


Figure 1.9 Citrulline post-translational modification labelling with biotin-conjugated phenylglyoxal. Top: Structure of biotin-conjugated phenylglyoxal. Bottom: Synthesis of the citrulline-specific chemical labeling with biotin-conjugated phenylglyoxal. The R group in the bottom panel is the portion shown in blue in the top panel.

The biotin-PG moiety provides a powerful chemical handle for citrulline enrichment. This method can discover citrullinated proteins or be used to characterize protein arginine deiminases (PAD)s. To discover citrullinated proteins, the modified citrulline containing proteins can be pulled down with streptavidin beads and digested on-bead. Then, the peptide fragments are analyzed by LC-MS/MS to determine which protein are citrulline substrates. To characterize PADs, citrulline proteins are detected in a control cell line and over-expression PAD cell line. Citrullinated proteins that are significantly enriched in the over-expression cell line are considered candidate PAD substrates.[86]

The biotin-PG label can also be used for citrulline detection with an ELISA.[86] The labelled proteins are adsorbed to an antibody-coated microwell plate that is specific to the protein of interest. Then, the antibody immobilized protein is bound to streptavidin-horseradish peroxidase (HRP) by the biotin handle. Citrullination is then quantified by incubating the microwells with a fluorescent HRP substrate to produce a quantifiable signal.

There are a few disadvantages to biotin-PG for citrulline analysis. The exact citrullination position cannot be determined since it remains conjugated to the bead. This could be fixed by producing a cleavable linker, which was stated to be under development by the authors of this method.[86] Also, the concentration of citrullinated protein cannot be differentiated from the overall

protein concentration. Regardless of these disadvantages, the biotin-PG provides a powerful chemical handle for isolation and detection of protein citrullination.

1.5.3 Direct chemical binding for enrichment of PTMs

Phosphopeptides are enriched by using solid-phase reagents that directly bind the phospho PTMs. The most common examples are immobilized metal affinity chromatography (IMAC) or titanium dioxide beads. Phosphorylated peptides are enriched by exploiting chemical bonding between the anionic, strongly coordinating phospho groups and metal-containing solid supports.[52] This method is more selective than antibody enrichment of phosphorylated peptides because enrichment is achieved by directly coordinating to the PTM rather than binding in any way at all to the peptide's surrounding sequence. Chelation of the phospho moiety can be achieved with either zirconium or titanium metal ions bound to the resins nitrilotriacetic acid (NTA) or iminodiacetic acid (IDA) (Figure 1.10).

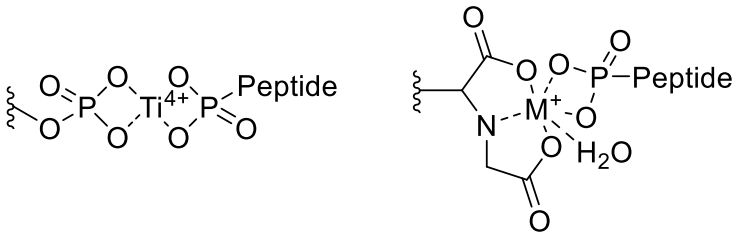


Figure 1.10 Chelation of phosphoryl moieties with Right: TiO_2 and Left: IMAC using the resin nitrilotriacetic acid

Pan-phospho enrichment is highly effective and widely used in studies of the phospho proteome. A paper by Thingholm *et al.* stated that their titanium dioxide enrichment protocol enriched 2634 HeLa cell phosphopeptides with 88% selectivity.[87] Selectivity of the High-Select™ TiO_2 Phosphopeptide Enrichment Kit branded by Thermo Scientific™ is reported to be around 85 – 95% selective for phosphopeptides over non-phosphopeptide—a high value due to optimized binding buffer composition to supress non-specific interactions with the affinity reagent. Thanks to this and other commercially available reagents, enrichment-powered phospho proteomics is a routine service offered in most major proteomics centre in the world.[88]

1.5.4 Advantages and disadvantages of chemical enrichment

Chemical enrichment is advantageous over antibody enrichment due to reproducibility of reagent production and lowered sequence recognition bias. Reproducibility of reagent

manufacturing is better controlled when a biological system is not involved. Polyclonal antibodies are heterogeneous mixtures, and each batch of polyclonal antibodies is unique, so poor reproducibility is a major drawback for this form of reagent.[89] This makes the synthetically produced reagents, for the most part, much more accessible and predictable than cocktails of biologically produced polyclonal antibodies. Monoclonal antibody production is much more consistent, but their sequence recognition bias[90] inhibits their ability to perform pan-specific PTM enrichment.[82] Sequence recognition bias is less of an issue for chemical enrichment with derivatization strategies because the PTMs are pulled down by sequence-independent mechanisms. For example, chemical labelling of the PTM citrulline with biotin-PG allows direct pull-down of the citrullinate proteins by the covalently bound probe.[86] Additionally, chemical derivatization with OPA and MDA allows enrichment of the methylated lysines and arginines based on their positive charge rather than flanking sequences.[84] But sensitivity is lost in the methyl arginine and lysine enrichment with ion exchange-based enrichment because histidine remains unblocked and charged. Therefore, histidine-containing sequences will contaminate this methyl arginine and lysine enrichment.[84] As for phosphopeptide enrichment with IMAC or TiO_2, sequence recognition bias has been reported, but is minimal.[91] Regardless, phosphopeptide enrichment has demonstrated robust reproducibility and selectivity that cannot be achieved by polyclonal PTM-enriching antibodies.[92]

Side reactions when using derivatization reagents pose an issue for proteomic analysis. For example, MDA is unstable, so the reagent is substituted with 1,1,3,3-tetraisopropoxypropane (TiPP) as a protected version of MDA. TiPP hydrolysis produces MDA, but could also result in unwanted protein esterification.[93] Unpredicted side reactions will produce aberrant mass shifts that are not inputted into the proteomic database used for spectral alignment and identification. As such, the unpredicted esterification will increase FDRs of methylated peptides.[84]

Methylation is difficult to target for direct binding by chemical reagents. Unlike phosphorylation, the methyl modification does not significantly alter the residue's charge—it causes the smallest possible changes to a residue's physicochemical properties. As such, the physicochemical properties of lysine methylation are much more difficult to exploit for chemical enrichment. With that, a comparable, chemically based, pan-specific enrichment strategy for cell-derived methyllysines, like phosphopeptide IMAC enrichment, has not yet emerged.

1.6 Calixarene-based methyllysine enrichment

1.6.1 Binding interaction between calixarene and methyllysine

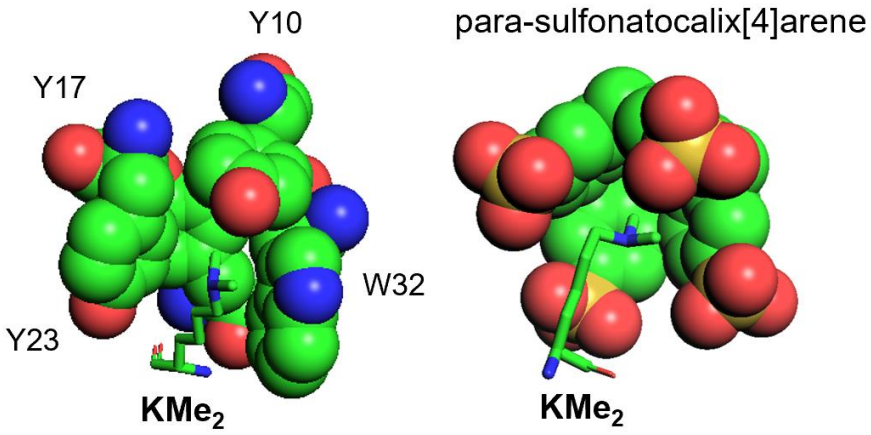


Figure 1.11 Structural comparison of the binding interaction between KMe_2 with Left: the methyllysine binding domain BPTF PHD-bromodomain (PDB code 2FSA) and Right: *p*-sulfonatocalix[4]arene (PDB code 4N0J). Images created using PYMOL.

The supramolecular host *p*-sulfonatocalix[4]arene is a macrocycle that has similar binding characteristics to those of methyllysine binding domains. Figure 1.11 (left) depicts the PHD zinc finger protein, BPTF. The four aromatic residues of the protein cage create a hydrophobic pocket lined with quadrupole-mediated partial negative charges. The diffuse positive charge of the methylammonium moiety inserts with into the complimentarily charged aromatic pocket to form cation-π interactions. Figure 1.11 (right) illustrates the binding interaction of *p*-sulfonatocalix[4]arene and its KMe_2 guest. The calixarene's benzene rings act as the aromatic residues do in the aromatic cage of the protein to accommodate the dimethyllysine guest with cation-π binding interactions.[94]

1.6.2 Selectivity of *p*-sulfonatocalix[4]arene

The binding selectivity to the *p*-sulfonatocalix[4]arene is between 9 – 41 fold greater for the methylated lysine over the non-methyllysine.[95] This is due to the size, shape, and electrostatic properties of the calixarene cavity and methylation on the lysine.[96]

The weak binding of the unmethylated lysine with calixarenes is well understood. The unmethylated lysine enters the cavity sideways, by bending its methylene chain and hydrogen bonding to the calixarene's upper rim sulfonates. The calixarene's inner cone shape pinches

around lysine to increase van der Waals contacts between one or more benzenes and lysine's hydrocarbon chain.[97] The lysine's amino group is positioned off-centered to form a hydrogen bond with two of the upper-rim sulfonate group. Shallow occupation of the cavity is achieved. Other calixarene binding conformations can take place between lysine and *p*-sulfonatocalix[4]arene, but with partial solvation and higher energy.

The strong binding of the *p*-sulfonatocalix[4]arene to trimethyllysine is driven by cation-π interactions and the hydrophobic effect.[95] The methyl groups of trimethyllysine bury deep within the cavity. As previously discussed, the cationic charge of the methylammonium sidechain is distributed amongst the methyl groups. The methylated lysine can thus make favourable interactions with all four faces of the cavity which allows full exclusion of water from the cavity. Additionally, the trimethylated lysine has lost its ability to hydrogen bond to the upper rim of the calixarene or hydrogen bond to solvent. As such, the hydrophobic effect largely drives the deep insertion of the trimethyllysine deep into the calixarene cavity with its $CH_2\varepsilon$ with exclusion of $CH_2\gamma$ and $CH_2\beta$ from the pocket.[95] For trimethyllysine, deep insertion was the only observed binding interaction with the calixarene observed in x-ray crystallography.[97]

1.6.3 Proof-of-concept methylated peptide affinity enrichment

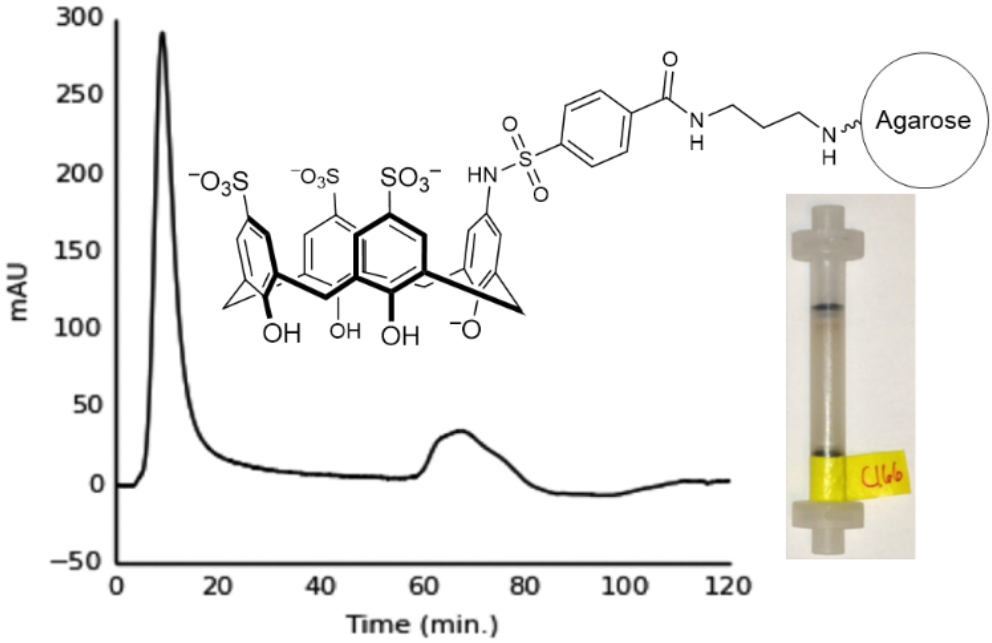


Figure 1.12 A first-generation methyl peptide affinity reagent. The structure of the upper-rim modified, calixarene-based enriching column is shown, along with a typical chromatogram arising from studies with histone-derived peptides. The retained peak at 60-80 minutes contains more methylated peptides. See text for more details.

In 2016, Garnett *et al.* reported the first enrichment method for methyl PTMs using an upper-rim modified calixarene-based chemical affinity reagent.[98] A sulfonated calixarene, prepared in 8 synthetic steps and 2% overall yield, acts as an anionic supramolecular host, which forms selective host-guest complexes with di- and trimethylated lysine residues. The host is coupled to agarose support for affinity chromatography. Its use on purified methylated peptides showed a proof-of-concept ability to selectively retain methyllysine-containing peptides from a commercially available extract of calf thymus histones. 38 unique proteins were discovered with ≥50% confidence, and 26 methylation sites were observed within the retained fraction. Its use on complex cell lysates was not demonstrated. Unpublished work in the Hof group that followed the publication of this first paper showed that this affinity column would not show any sample retention when it was used on more complex samples like cell lysates. This was either due to some contaminating component in the cell lysate that impacted the performance of the calixarene-based column, or due to some inherent shortcoming in the calixarene reagent itself. In any case, the complex, multi-step chemical synthesis of this reagent severely limits its availability and so further development of this upper-rim modified reagent was stopped. At the outset of this thesis work, a lower-rim modified calixarene-based reagent had just been developed.

1.7 Thesis objectives

The objective of this thesis is to develop an optimized method for pan-methyl proteomics using affinity enrichment. My path was to develop the chemical enrichment of methyllysines from complex cell lysate with the use of a calixarene-based affinity column. The proof-of-concept calixarene affinity reagent reported by Garnett shows promise but lacks the ability to enrich complex samples from cell lysates and other real-world biological sources. This thesis reports the use of a modified version of Garnett's host (now termed 'MethylTrap') that can be prepared in one step from readily available materials. Additionally, this thesis reports sample preparation protocols for mammalian and yeast cell lysates which allow retention of large numbers of peptides by the MethylTrap column. By implementing optimized sample preparation prior to enrichment, hundreds of methylated peptides were identified in cell lysates (Chapter 2). Reproducibility of MethylTrap enrichment and the biological implications of the methylated peptides identified were

scrutinized in Chapter 3. The MethylTrap shows significant promise as a future tool for methyllysine prospecting and profiling.

Chapter 2: Optimizing MethylTrap enrichment

Contributions

I optimized the MethylTrap synthesis, protein purification, alkylation, digestion, and MethylTrap enrichment with assistance from Mark Grasdal. Mark Grasdal analyzed the effects of different proteases on the number of methylated peptides discovered. Mark also performed the batch binding column buffer exchange step, and tested the lower versus upper-rim modified *p*-sulfonatocalix[4]arene retention efficiency.

2.1 Introduction: Addressing the shortcomings of the first generation MethylTrap

The upper-rim modified MethylTrap reagent had two serious shortcomings: it didn't work well on complex cell lysates, and the arduous, multi-step chemical synthesis of this reagent severely limited its availability.[98] Optimization of Garnett *et.* al's MethylTrap enrichment column[98] was required.

This thesis aims to use the MethylTrap column to pan-specifically enrich the methyllysine containing peptides from lysates derived from whole cells. At the outset of this research, a lower-rim modified calixarene had demonstrated proof-of-concept methyllysine binding selectivity but had not demonstrated its applicability to lysate enrichment. The column retained commercially available purified calf histone peptides but could not retain peptides derived from cellular lysates. My goal was to achieve methyllysine retention, and to demonstrate that unique methylated peptides can be identified after enrichment that would otherwise not be observable in a proteomics experiment. This would establish the MethylTrap column as a new research tool for methyllysine prospecting and discovery. This Chapter includes the work done to optimize sample preparation, MethylTrap preparation, and MethylTrap operation; exemplary proteomics data that demonstrate the performance of the MethylTrap method; a discussion of the results; and finally, an appendix that describes the complete optimized procedure in the form of a highly detailed "Protocols" paper.

2.2 Making MethylTrap work for global methylated protein analysis: considerations

2.2.1 Goals for this Chapter

My first goal was to improve protein extraction yields from cell lysates while maintaining down-stream compatibility with affinity enrichment methods.

My next goal was to remove non-protein cellular debris from the lysate to optimize binding to the MethylTrap column.

My final goal was to optimize the number of methylated peptides identified.

2.2.2 Maintaining down-stream compatibility in sample preparation

The most crucial factor when preparing a sample for proteomic analysis is maintaining compatibility of the sample with down-stream analysis. Proteomic LC-MS/MS is a sensitive technique that is not compatible with many reagents that are routinely used in protein and peptide

biochemistry protocols. This drastically limits what reagents can be used for sample preparation/purification. Many detergents, polymers, cryoprotectants, non-volatile buffers, etc. are incompatible with proteomic mass spectrometry.[99] Detergents and certain polymers competitively ionize in the electrospray source along with the analytes of interest, which causes peptide ion signal suppression.[100] Also, adducts are formed by some non-volatile salts and buffers that supress analyte ionization. Even the cryoprotection reagents, such as glycerol, that are used for cell storage are detrimental to the LC. One must avoid the addition of these compound or remove them prior to LC-MS/MS to achieve successful analysis.

Polyethylene glycol (PEG) contamination must be avoided to reduce spectral convolution and competitive ionization during LC-MS/MS analysis. From my own experience during this project, PEG is the most common contamination that is unintentionally introduced during sample preparation. PEG contamination is introduced when using equipment that leaches polymer into the sample or when using low-purity reagents for sample preparation.[105] When present, ion suppression of the target ions occurs during ionisation of LC-MS/MS.[101] LC separates the polymers across the chromatogram which allows it to interfere with ionization of most eluting peptides.[102] The reason for ion suppression by this non-volatile reagent has not been definitively described to date, but PEG may change the spray droplet properties during ESI[103] or induce analyte precipitation which inhibits liquid to gas phase-transfer[104]. This narrows the range of detection by decreasing the number of ionized peptides. The compatibility of solvents with equipment and use of high purity reagents must also be considered when preparing the sample for LC-MS/MS analysis.

2.3 Customizing sample preparation for optimal MethylTrap retention

Figure 2.1 gives an overall view of the optimized protocol that was developed. The optimization of each step will be reported individually in the following sections. Refer to the supporting information for the final detailed step-by-step protocol.

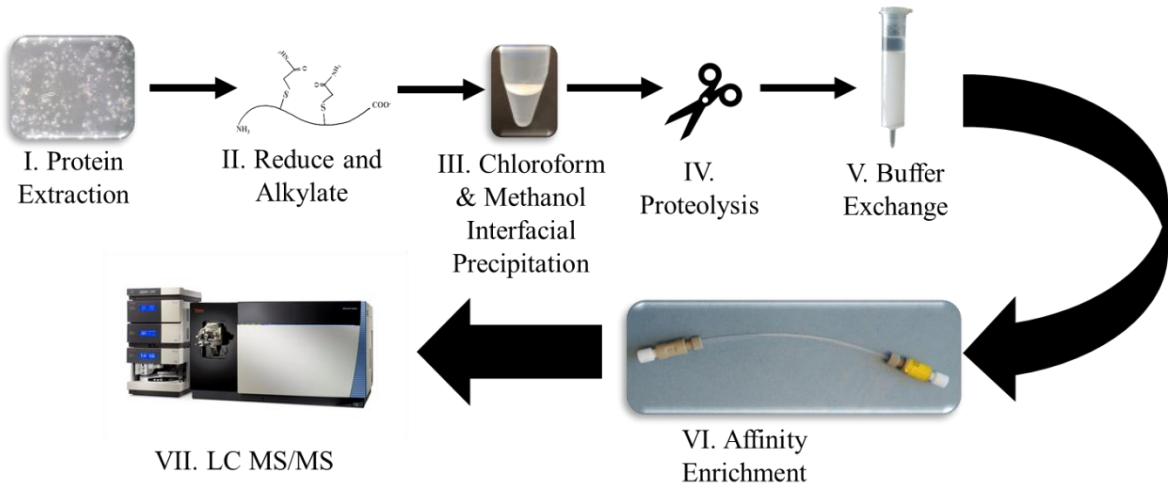


Figure 2.1 Schematic diagram of the optimized MethylTrap enrichment protocol.

2.3.1 I. Cell lysis protocol

The first step of proteomic analysis is protein extraction from a biological source. I chose two different sources for this work due to their accessibility: cultured cancer cells and commercially sourced brewers' yeast. Mammalian cells, human prostate cancer cell line (PC-3), were grown to 80% confluency in two T-175 flasks; or, an aliquot of brewers' yeast was collected from the bottom of Phillips Brewing & Malting Company's ("Phillips") fermentation vessels. Both samples were washed with buffers a minimum of three times to remove exogenous material surrounding the cells. Refer to the method section for a detailed description of the lysis protocol.

Mechanical and physical lysis does not introduce LC-MS/MS incompatible reagents to the sample, thus making it a better suited method than chemical lysis. Chemical lysis is a common cell lysis method that uses detergents which, when added to cells, creates pores in the cell's membrane to release the intracellular material.[106] Chemical lysis with detergents was used in an early, non-optimized version of the nuclear extraction protocol, which disrupted LC-MS/MS analysis downstream.[107] To overcome this limitation, the earlier procedure implemented a detergent removal column, but this resulted in a 30% sample loss following protein extraction. In contrast, mechanical and physical lysis physically breaks the cell membrane with force.[108] No foreign reagents are introduced with this lysis method. As such, I chose mechanical and physical lysis for the optimized protein extraction method based on its LC-MS/MS compatibility.

Protein extraction optimization increased yields by 33-fold without introducing LC-MS/MS incompatible reagents (Table 2.1). In the optimized protocol, a combination of osmotic shock and bead beating methods successfully extracted 2 mg protein per 1×10^7 mammalian cells

(Table 2.1). To lyse cells while avoiding the introduction of incompatible reagents (e.g. detergents), PC-3 cells were suspended in diH_2O supplemented with protease inhibitor. The hypotonic pressure between the de-ionized water and cytoplasm enclosed by the fragile cell membrane caused the cells to burst and release the lysate. To ensure complete lysis of all the organelles' membranes, seven rounds of high-speed vortexing were performed in the presence of acid-washed glass beads. The lysate was passed through a 21½ gauge needle, which further lyses the cells with shear force and reduces the lysate's viscosity by shearing the released DNA. Observation under a microscope indicated complete mammalian cell lysis—intact cells were not observed.

Table 2.1 Protein extraction optimization yields for PC-3 mammalian cells.

Lysis method	**Number of cells**	**Protein yields (µg)**	**Protein (µg) / 1 x 10^7 cells**
Mechanical lysis for whole cell extraction	2.7 x 10^7 cells	5000	2000
Chemical lysis for nuclear extraction	3.6 x 10^7 cells	200	60

Yeast cells were more difficult to lyse than mammalian cells. Suspension in a hypotonic solution is not enough to disrupt their robust cell wall. Moreover, mechanical disruption with glass-bead vortexing did little to lyse the cells (data not shown). Several rounds of aggressive sonication followed by glass-bead mechanical lysis (the same as for mammalian cells) led to sufficient lysis without needing any detergents or other reagents that may affect downstream steps (proteolysis, MethylTrap enrichment, LC-MS/MS, etc.). Yeast lysis efficiency was also scrutinized by light microscopy. Not all the cells ended up being lysed, but subsequent work showed that the lysing efficiency was sufficient for proteomics studies. Also, the supply of yeast cells was virtually infinite, so using more cells to extract a sufficient amount of proteins made up for incomplete lysis efficiency. Intact cells were removed by centrifugation.

2.3.2 II. Fully denaturing the proteins

Cysteine alkylation is a widespread practice that increases identification in proteomic analysis. Alkylating the cysteine residues lowers protein folding stability by inhibiting disulfide bonding between cysteine residues.[109] This increases accessibility of the protein backbone for subsequent proteolytic cleavage and enhances ionization efficiency signal intensity.[110]

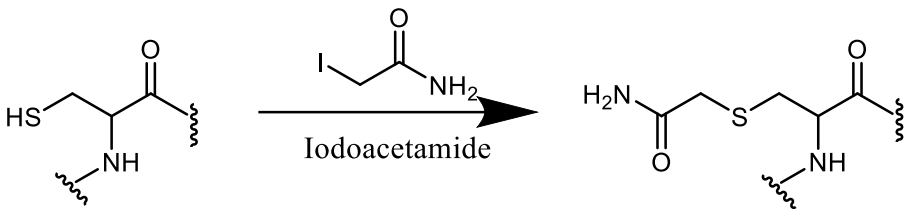


Figure 2.2 The alkylation of a reduced cysteine residue using iodoacetamide to produce a carbamidomethyl moiety. Refer to steps 46 – 48 in the SI for the detailed reduction and alkylation protocol.

Reduction and alkylation of cysteine residues is a frequent practice to maintain proteins in their denatured states.[111] The protein extract was diluted into a solution with a final concentration of 3.5 M guanidine hydrochloride (denaturing agent) and 10 mM dithiothreitol (DTT) (reducing reagent). The proteins are left to denature and reduce for 30 minutes at 65℃. Once denatured, 50 mM iodoacetamide (IAA) was added to alkylate the cysteine residues and DTT (Figure 2.2). To avoid side-reactions of IAA with the proteins, alkylation was performed in the dark at ambient temperatures (~22℃) for a strict 30-minutes.

2.3.3 III. Avoiding premature elution from the MethylTrap column by sample purification

The upper-rim modified calixarene column bound peptides derived from the digest of purified histone extracts, but not to peptides derived from cell lysates. This may be due to the cellular debris and other biomolecules that are present in the cell lysate. The cell contains all sorts of molecules—lipids, DNA, RNA, metabolites, proteins, and more. These molecules could compete for binding with the MethylTrap resin or the peptides themselves, so their removal was speculated to aid methyllysine retention. Without removing the non-protein cellular debris, all the molecules eluted from the column (Figure 2.3).

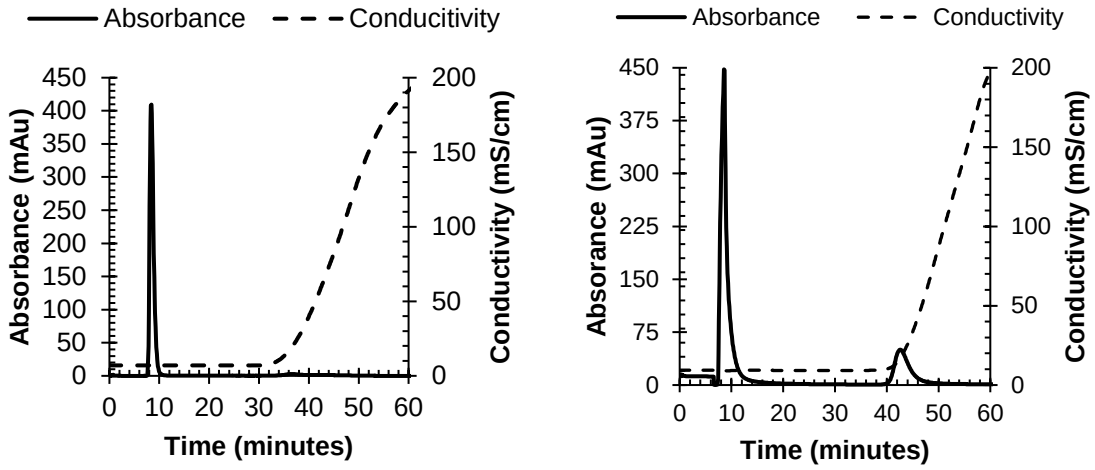


Figure 2.3 MethylTrap enrichment of cell-derived protein extracts requires protein purification prior to column enrichment for successful column retention. Absorbance was read at 280 nm to observe when the peptides eluted from the column, and conductivity was used to observe when the elutant was added to the mobile phase to elute the retained peptides. The left chromatogram demonstrates unsuccessful binding of trypsin digested PC-3 proteins without methanol/chloroform protein precipitation. The right chromatogram demonstrates successful binding of trypsin digested PC-3 proteins with methanol/chloroform protein precipitation.

Numerous protein purification methods have been developed, but chloroform/methanol protein precipitation was chosen based on its simplicity and short duration.[112,113] A 4:4:1:1 ratio of water, methanol, alkylated proteins, and chloroform, respectively, was combined. This produces two liquid phases and one interfacial solid phase (Figure 2.4). The aqueous phase contains water-soluble compounds such as salts, detergents, nucleic acids, and reducing agents, while the chloroform organic phase contains lipids and fat-soluble metabolites. At the interface of these two phases, the proteins, which are almost all denatured and amphiphilic at this point, precipitate. With removal of the liquid phases, most of the unwanted cell debris is withdrawn.

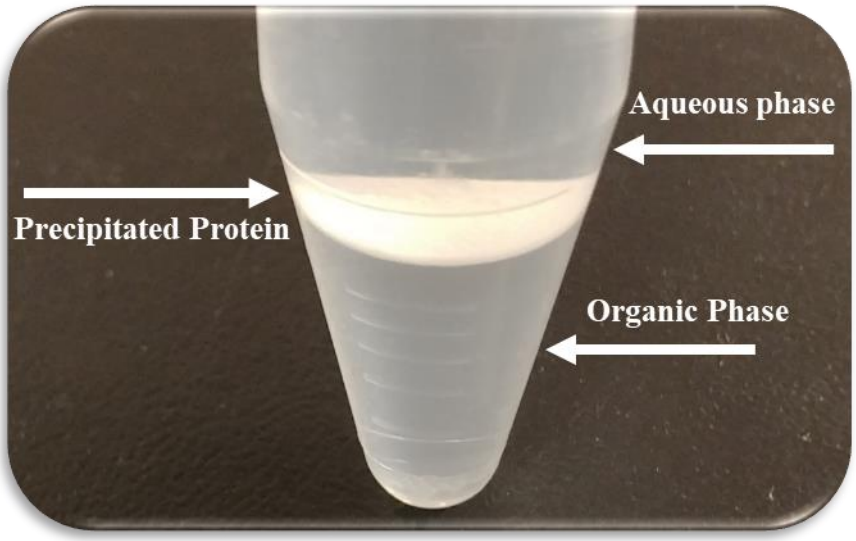


Figure 2.4 Chloroform methanol interfacial protein precipitation.

Proteins must be extracted with high purity solvents to increase the number of identified methylated peptides (Table 2.2). LC-MS/MS analysis revealed high concentrations of PEG contamination when sub-optimal purity communal high-pressure liquid chromatography (HPLC)-grade solvents were used for the chloroform/methanol precipitation. Maintaining sample purity during sample preparation is of the utmost importance for optimal peptide ionization during LC-MS/MS analysis. As such, higher purity grade solvents for the precipitation step were swapped-in in hopes to reduce the PEG contamination issue. In-turn, more methylated peptide identifications were observed when the mass-spec grade solvent system for protein precipitation was implemented (Table 2.2).

Table 2.2 Effect of solvent purity on methylated peptide identifications from PC-3 cells. Peptides were processed with the final optimized protocol with expectation of the solvent purity.

Solvent type	**Number of methylated peptides identified**		
	Me	Me_2	Me_3
HPLC-grade solvents	18	23	19
Mass-spec grade solvents	86	89	82

The protein pellet was best re-solubilized by initially sonicating the pellet in a basic solution before diluting it into a neutral buffered solution (Table 2.3). Following protein precipitation, the most common reagents used for re-solubilisation are detergent and chaotropes

—reagents that are not compatible with LC-MS/MS analysis.[114] An alternative solubilisation method had to be optimized as to not introduce contamination into the sample. The first method attempted was heating the protein pellet in a buffered solution to increase solubility. Low protein yields were recovered with this method (Table 2.3). The optimized solubilisation procedure used a basic solution of 50 mM NaOH to reintroduce the protein pellet into solution. Deprotonating the proteins beyond their pI increases re-suspension efficiency.[115] A brief sonication was performed to aid their reintroduction into solution. Then, a concentrated dose of ammonium bicarbonate buffer was added to neutralize the protein solution. The remaining insoluble material was pelleted and removed by high-speed centrifugation. Deprotonating the proteins instead of heating before and during the sonication increased protein yields by ~3-fold (Table 2.3).

Table 2.3 Optimization of protein solubilisation. Both methods used sonication to aid solubilisation.

Solubilisation method	**Yields (µg of protein)**
Heating	300
NaOH solution	1000

Metal chelation does not increase methylated peptide identifications (Table 2.4). Removal of metals prior to MethylTrap enrichment was an attempt to increase column retention and decrease signal suppression in mass spec analysis.[116] The proteins were incubated with bead-bound styrene-divinylbenzene co-polymer containing iminodiacetic acid groups that act as chelating groups. There was no difference in MethylTrap retention with and without chelation. Additionally, there was no effect on the number of methylated peptides observed following LC-MS/MS analysis (Table 2.4). In contrast, Chelex-100 introduced contaminating polymer into the sample, which supressed sample ionization. This step was removed from the optimized sample preparation protocol.

Table 2.4 Metal chelation does not increase the number of methylated peptides observed in MethylTrap enrichment proteomic analysis. Number of methylated peptides identified with and without incubation with the metal chelating agent Chelex-100.

	Number of methylated peptides identified		
	Me	Me_2	Me_3
Chelation of metals	1	8	3
Without chelation of metals	3	9	2

3.1.1 IV. Protease choice and efficiency

The use of different enzymes for the proteolysis step shifts coverage of the methyl proteome. Generally speaking, the use of different proteases yields different sets of peptides—length and charge of digests differ between protease choice.[117] Different peptide sets will thus have variable MethylTrap retention, ionization efficiency, cleavage frequency, etc. As such, the choice of protease drastically alters the outcome of enrichment and proteomic analysis.

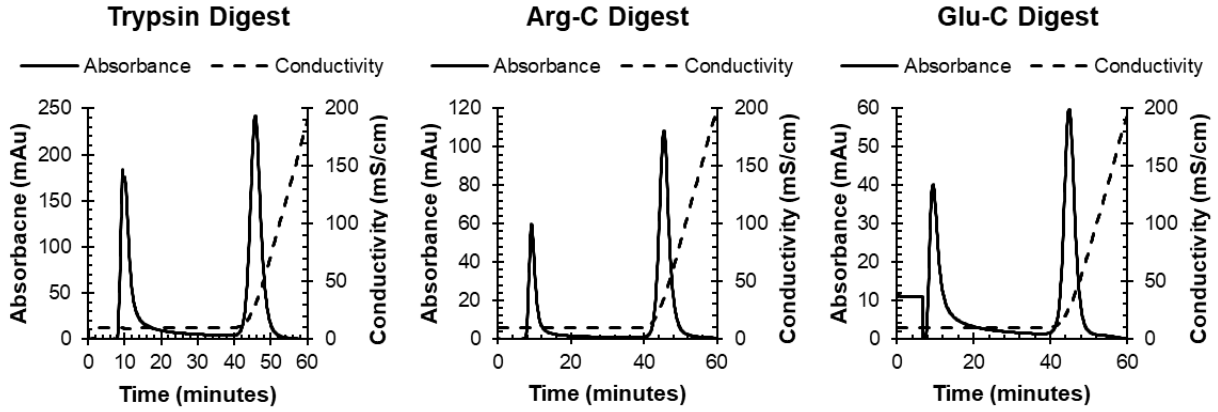


Figure 2.5 Chromatograms of MethylTrap enriched yeast proteins digested with trypsin (left) Arg-C (middle) and Glu-C (right). Enrichment was performed as described in SI protocol with exception of the protease. *

Glu-C cleavage led to more methylated peptides being identified than Arg-C and trypsin digests (Table 2.5). When trypsin was used as the protease, retention was observed (Figure 2.5 left), but the number of methylated peptides identified in the retained fraction was low (Table 2.5). Seven methylated peptides were observed in the retained fraction. It was previously observed that MethylTrap had a charge bias for highly basic methylated peptide.[98] This observation could potentially explain why trypsin, a protease that yields highly acidic peptides, does not produce many methylated peptide identifications. Trypsin cleaves on the C-termini of the basic residues, lysine and arginine.[118] The peptide fragments produced by trypsin would contain a maximum of one basic residue (with assumption of full digestion).

The use of proteases that yield peptides with a higher proportion of basic residues were investigated (Arg-C and Glu-C). Arg-C selectively cleaves C-terminal to arginine, and Glu-C selectively cleaves C-terminal to glutamate.[117] The hypothesis was that these proteases would

*** Analysis was performed by Mark Grasdal.**

increase MethylTrap retention and methylated peptide discovery because the peptide fragments produced by these proteases would contain more basic residues than trypsin. This hypothesis was based on the previously observed charge bias of the MethylTrap column.[98]

Arg-C and Glu-C cleavages led to more methylated peptides being identified than trypsin (Table 2.5), but had comparable retained peak areas relative to trypsin during MethylTrap enrichment (Figure 2.3). Respectively, Arg-C and Glu-C digested peptide samples identified 166 and 281 methylated peptides (Table 2.5). Glu-C was chosen as the optimal protease for digestion because this digest identified the most methylated peptides.

Table 2.5 Glu-C digests allowed more methylated peptide identifications than Arg-C or trypsin digests. Number of unique methylated peptides identified for each degree of methylation when digestion was performed with different proteases.*

Protease type	**Me**	**Me_2**	**Me_3**
Trypsin	5	2	0
Arg-C	62	53	51
Glu-C	89	90	102

2.3.4 V. Buffer exchange to optimize column retention

Peptide desalting increased retention to the MethylTrap resin (Table 2.6). Desalting, buffer exchange, and a final sample cleanup with a PD MiniTrap G-10 desalting column places the peptides in the MethylTrap's binding buffer (50 mM phosphate buffer pH 7.5). Peptides smaller than the relative molecular weight (M_r) 700 are excluded from the equilibrated gravity column due to the nature of the desalting column. This final sample cleanup step suspends the peptides in optimized binding buffer prior loading the peptides onto the column. The buffer exchange step resulted in a 3-fold increase in retention relative to peptides retained in the presence of ammonium bicarbonate buffer.

Table 2.6 Optimization of MethylTrap enrichment by desalting the peptide sample with a PD MiniTrap G-10 desalting column to exchange the peptide into binding buffer.*

	Peptide concentration (µg)		
	Unretained	Retained	Input
Without desalting	580	2	670
With desalting	470	6	590

2.3.5 VI. Choice and production of MethylTrap resin and column

MethylTrap assembly started with a one-step synthesis of a novel calixarene-based affinity reagent (Figure 2.6). Reaction with 4-chlorosulfonylbenzoic acid installed a functional handle on the *p*-sulfonatocalix[4]arene's lower rim.

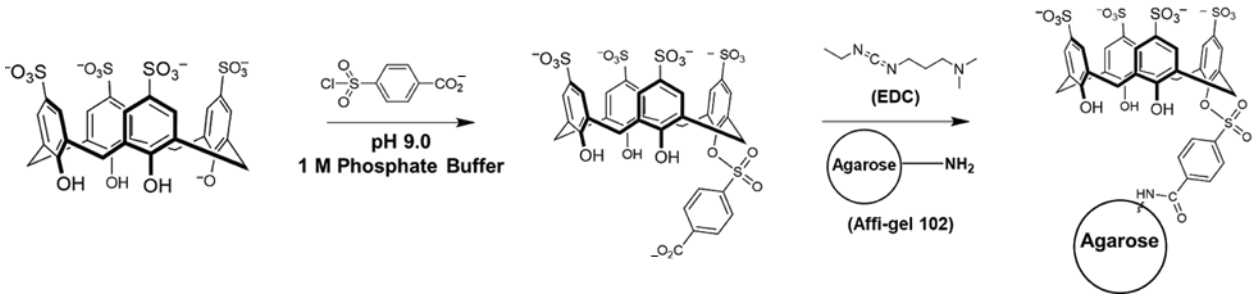


Figure 2.6 Synthesis of MS124 affinity reagent from commercially available *p*-sulfonatocalix[4]arene followed by the coupling reaction to produce agarose affinity enrichment bead resin.

The mono-functionalized calixarene was isolated from the starting reagents and the other di- and tri-substituted products via HPLC (Figure 2.7 left). The isolated yields are typically 30–35%, with 98% purity as determined by analytical HPLC. In the presence of (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride) (EDC), the functionalized calixarene was then coupled to a solid support that consists of cross-linked agarose gel decorated with primary amine handles (Affi-gel 102, Bio-Rad). Functionalization was confirmed by HPLC analysis of the bead supernatant before the addition of EDC then after EDC coupling had occurred (Figure 2.7 right).

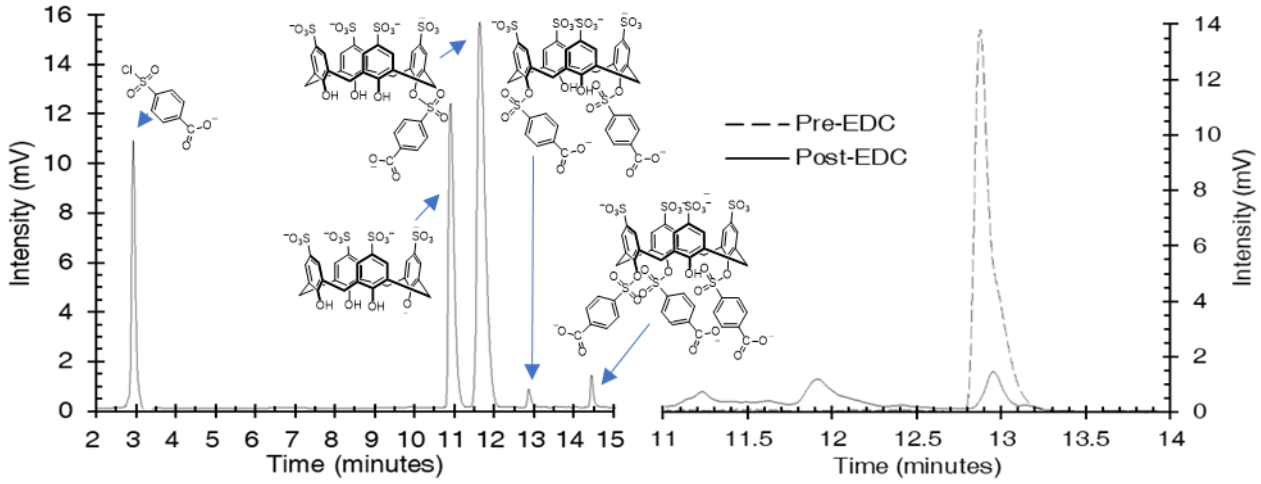


Figure 2.7 Mono-substituted calixarene product can successfully be purified from side-reactions and starting reagents, and EDC coupling successfully couples calixarene to solid support. Exemplary HPLC chromatograms for successful purification of MethyTrap calixarene (left) and successful EDC coupling (right). Left: HPLC purification chromatogram of MethylTrap resin. Elution order is as follows: unreacted 4-chlorosulfonyl benzoic acid, mono-substituted lower-rim modified *p*-sulfonatocalix[4]arene (product), di-substituted *p*-sulfonatocalix[4]arene, then tri-substituted *p*-sulfonatocalix[4]arene. Blue arrows correspond structures to peaks. Right: HPLC chromatogram of the lower-rim modified *p*-sulfonatocalix[4]arene present in the supernatant before and after coupling to agarose solid support. "Pre-EDC" coupling trace (dashed line) overlaid on the "Post-EDC" coupling trace (solid line).

The calixarene-coupled agarose beads were washed, then packed into a clear PFA tube enclosed by commercially available fitting, ferrule, frit, union, and plug to make the MethylTrap column (Figure 2.8). Approximately 850 µl of resin with 10 mg of bound calixarene was pack into each column. The assembled column was then connected to an AKTA Prime LC system for an overnight wash with elution buffer (50 mM phosphate buffer with 2M NH_4Cl at pH 7.5) at a flow rate of 0.3 ml/min. This elutes compounds that adsorbed to the column. Once washed, the resin was evenly packed throughout the column using a flow rate between 0.4 – 0.6 ml/min. Any headspace that formed was removed. Once the resin was packed, the column was then equilibrated with binding buffer (50 mM phosphate buffer)—flow through conductivity of ~8.8 mS/cm indicated complete column equilibration. Plugs were connected to either side of the column when stored.

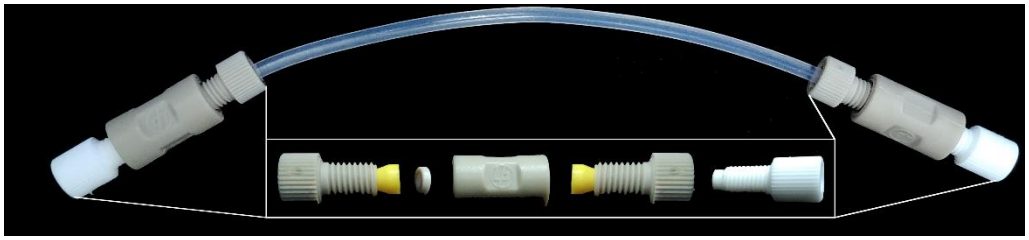


Figure 2.8 Assembled MethylTrap column.

A lower-rim modified *p*-sulfonatocalix[4]arene was used in lieu of the upper-rim *p*-sulfonatocalix[4]arene (Figure 2.9) due to the ease of synthesis and its higher retention efficiency (Figure 2.10).

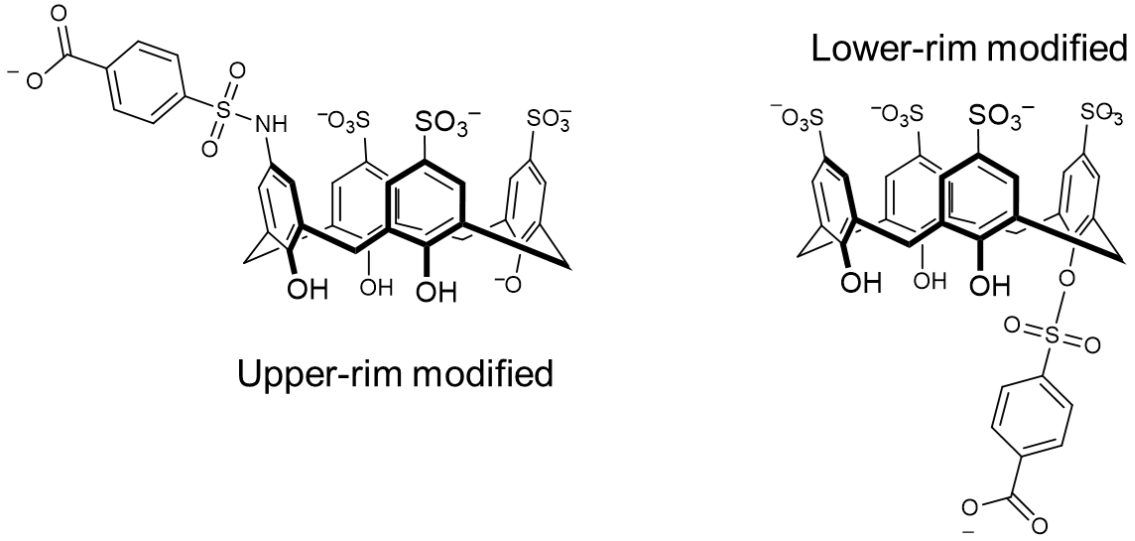


Figure 2.9 Structure of the upper-rim modified *p*-sulfonatocalix[4]arene (left) and lower-rim modified *p*-sulfonatocalix[4]arene (right).

The unretained peak was much larger than the retained peak for the upper-rim modified *p*-sulfonatocalix[4]arene (Figure 2.10 left). The choice to use the lower-rim modified *p*-sulfonatocalix[4]arene was initially based on the ease of synthesis. But, the large retained peak relative to the unretained peak was another reason for using this compound in the optimized MethylTrap enrichment protocol—the lower-rim modification increased column retention (Figure 2.10 right).

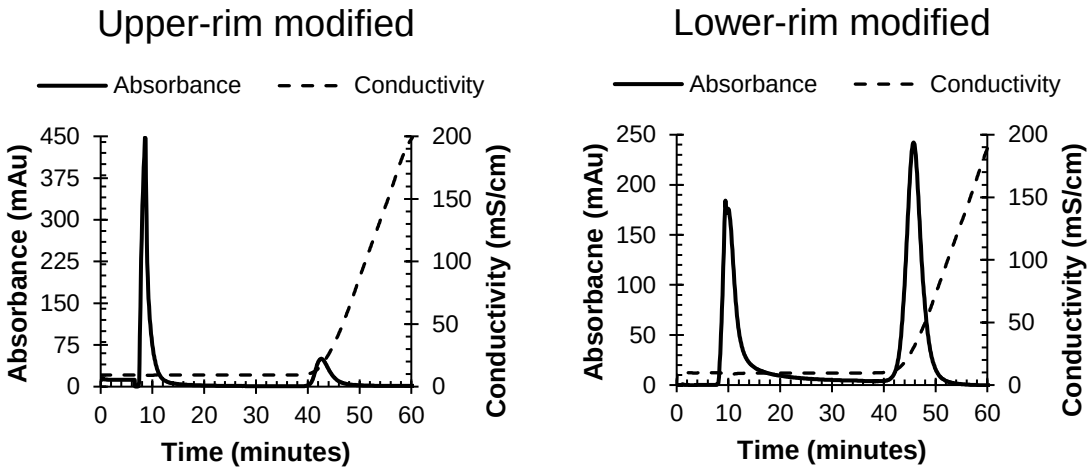


Figure 2.10 The new calixarene reagent improves peptide retention from complex lysates. Chromatograms of trypsin digested peptides enriched with the upper-rim modified *p*-sulfonatocalix[4]arene (left) and lower-rim modified *p*-sulfonatocalix[4]arene (right).

2.3.6 VII. Methyllysine identifications in PC-3 mammalian cell line

Peptides were separated and analyzed by on-line reverse phase chromatography coupled to an Orbitrap Fusion Tribrid mass spectrometer, then the acquired spectra were assigned hit scores with Mascot and validated with Scaffold proteomics software. The exact statistics behind the identification scores and assignment is beyond the scope of this thesis. Briefly, Mascot assigns spectrum identity scores by alignment to the inputted database,[119] and the Scaffold software applies PeptideProphet™ and ProteinProphet™ algorithms to assign probability based peptide and protein identifications.[120,121] Peptides and proteins with identification probabilities above 90.0% and 95.0%, respectively, were accepted as identifications.

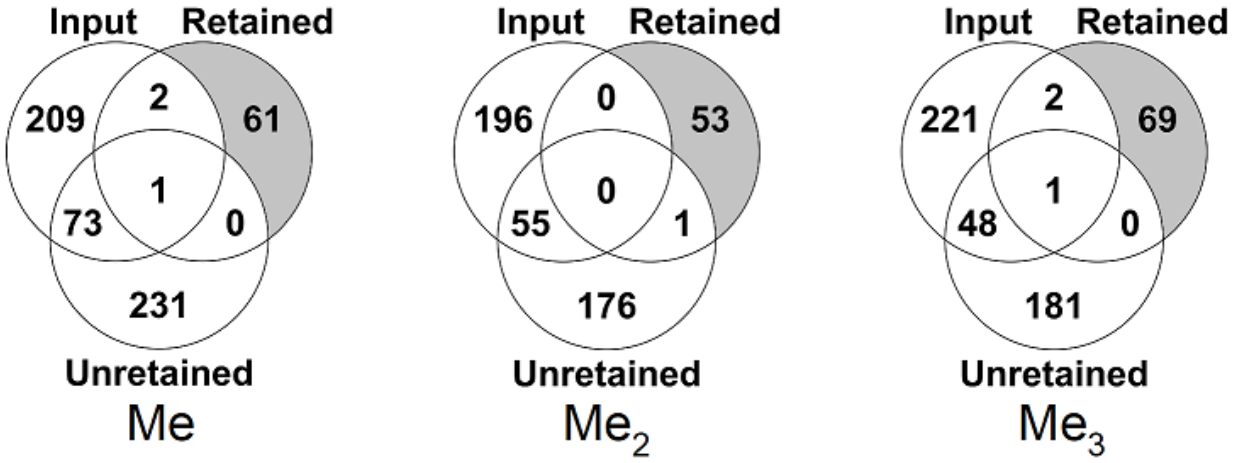


Figure 2.11 MethylTrap column fractionation increases the number of unique identified methylated peptides relative to the input sample in a cancer proteomics experiment. PC-3 cell lysates were enriched using MethylTrap, and the input control sample, as well as the MethylTrap retained and unretained fractions were submitted for proteomics analysis. All methyl PTM sites in each sample were tabulated and counted. Venn diagrams show the overlap in the number of unique mono- (left), di- (middle), and tri-methylated (right) PTM sites identified. The shaded sectors show methyl PTMs that are only visible after MethylTrap enrichment. Data are from a single replicate.

MethylTrap enrichment aids identification of methylated peptides that are otherwise unseen in the input fraction (Figure 2.11). MethylTrap enrichment of PC-3 peptides led to the identification of 1578 methylated peptides within a single replicate. 61 monomethylated, 53 dimethylated, and 69 trimethylated peptides were otherwise unseen in the input and unretained fractions (Figure 2.11 shaded regions).

MethylTrap methylated peptide retention requires further optimization. The number of identified methylated peptides is 4-fold greater in the unretained fraction relative to the retained fraction. The retained fraction increased the number of unique methylated peptides by 24% in comparison to the input sample alone, and the unretained resulted in a 74% increase. A potential explanation for the higher number of methylated peptides in the unretained sample relative to the retained sample could be that the column is selective for some methylated peptides sequences over others. The overlap between the unretained and retained fraction is 0.4%, so the identifications made in the retained and unretained samples were almost exclusive. But, the fact that the unretained sample identified more methylated peptides than the retained indicates that the column elutes more unique methylated peptides than it retains.

MethylTrap fractionation identified methylated peptides that were not previously identified in the UniProtKB database (Table 2.7). The input, unretained, and retained fractions

identified 95%, 94% and 96% never-before-seen methylation. Granted, not all of the identified methylated peptide will be true identifications. Approximately 10% of the identifications will be false positives based on the decoy database method.[55] Additionally, the decoy database method commonly used to validate the methylated peptide discoveries is not stringent enough to filter out the false positives at the specified FDR threshold for methylated peptide identifications.[122] Being that methylation produces a mass shift (14 Da) analogous to isobaric mass differences between glycine and alanine, threonine and serine, leucine and valine etc., unmethylated peptides of similar sequence can give a great score for a methylated peptide if the sequence's mass differs by a single methyl. Isobars and identifications within the range of instrument error are of great concern, so the percentage of novel methylation marks is likely inflated. Methylated peptides must be validated with an orthogonal approach (such as western blotting) in order to confirm their discovery.

Table 2.7 MethylTrap enrichment identifies novel methylated peptides, but a portion of the novel methylated peptides are presumably false positives. The table contains the fractions and percentages of identified methylated proteins observed in one replicate of PC-3 MethylTrap enriched peptides that were previously published be methylated proteins within the UniProtKB database. Note that not all the identified methylate peptides were within the UniProtKB database.

	Input	Unretained	Retained
Me	11/257 (4%)	15/270 (6%)	2/60 (3%)
Me_2	10/231 (4%)	10/204 (5%)	1/50 (2%)
Me_3	12/253 (5%)	15/208 (7%)	4/65 (6%)

Reproducibility of MethylTrap methylated peptide retention was low. Three independently lysed PC-3 protein extracts were enriched with the MethylTrap enrichment protocol, and the retained fractions were independently analyzed with the aforementioned Orbitrap mass analyzer and software. The spectra were assigned identifications and the Venn diagrams illustrating the overlap of unique methylated peptides identified between the three replicates was exported from the software (Figure 2.12). The retained fraction demonstrated poor methylated peptide identification reproducibility. 3% of the methylated peptides were identified in at least two of the three replicates, and 0.4% were seen in all three replicates.

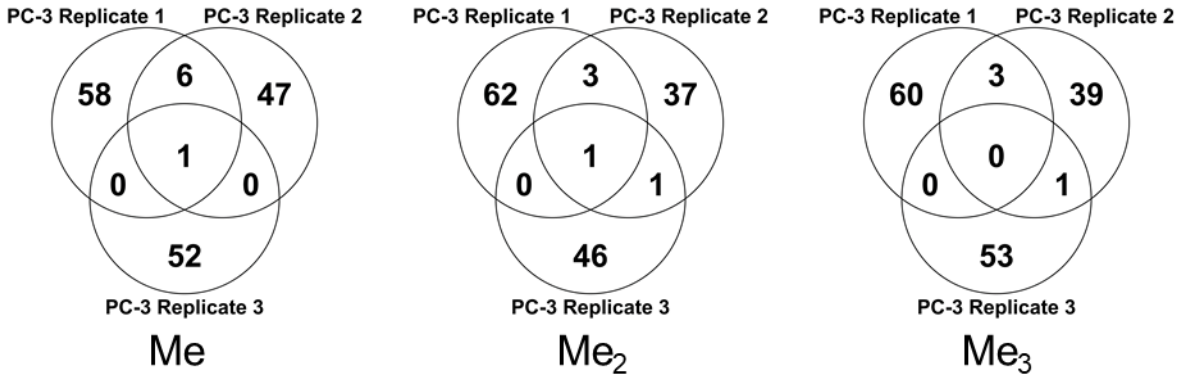


Figure 2.12 Reproducibility of MethylTrap enrichment is low. Venn diagram of mono, di, and tri methylated peptide (left to right) identified in the retained fraction of MethylTrap enriched peptides from three biological replicates extracted from independently grown and lysed PC-3 cells.

Variations caused by sample preparation and LC-MS/MS analysis may be the cause for the poor reproducibility of unique methylated peptides identified in the retained sample. The host (calixarene resin) and guest (peptide) each have a binding constant associated with their association (K_a). The relative binding strength of each possible guest produces a ratio of K_a which characterizes the host's binding selectivity. Since selectivity is relative (not absolute) the host-guest binding interaction is dependent on the concentration of input peptide. The dynamic concentration range of the proteome is non-constant, so each protein extract will contain unique protein concentrations. Moreover, sample preparation prior to enrichment also introduces variability between sample sets which alters the concentration of peptides in the input sample.[123] Changes in injected peptide concentrations will alter the binding interactions of the system, thus changing the proportion of retained peptides in the MethylTrap column. LC-MS/MS injections are another known source of inconsistency between any set of proteomic replicates. Differences in peptides analyzed in LC-MS/MS analysis range between 35 – 60 %.[124] These compounded variabilities could lead to the low reproducibility observed of the retained fraction.

2.4 Limitation of MethylTrap enrichment

2.4.1 Biased methyl proteome coverage

Peptide retention and identifications are biased towards highly cationic sequences containing few glutamic acid residues. Highly cationic peptides have been previously reported to increase overall retention in *p*-sulfonatocalix[4]arene columns, regardless of their methylation state.[98] As such, the MethylTrap column has bias retention for cationic methylated peptides.

Highly acidic peptides are most likely not retained in the MethylTrap column. Moreover, the protease used for digestion, Glu-C, cleaves proteins on the carboxyl termini of glutamic acid residues, which will limit the number of glutamates in the peptide sequence to one (barring missed cleavage sites).[125] As such, regions rich in glutamic acid residues will be highly fragmented. Peptides with lengths below 6 amino acids do not statistically permit unique identification to their corresponding protein;[126] therefore, glutamate rich peptides will be eliminated during proteomics data processing and will not be discovered using this method. Sampling bias may be present for highly basic methyl-peptides containing few acidic residues.

2.4.2 False discovery

This method is susceptible to low sequence coverage because methyl protein identity may be determined by a single peptide. This poses issues for accurate protein identification rates since seeing one peptide within an entire protein sequence gives lower confidence for the presence of that protein in comparison to seeing numerous peptides for a given protein.[127] The likelihood of enriching numerous regions of a single protein with detectable abundance in bottom-up enrichment proteomics is presumably low, since a methyl will likely only be present on a single site on a single peptide. So, protein identification was accepted if the presence of a single peptide bearing a methylation mark was observed. To reduce issues of false protein discovery, the presence of other peptide fragments (methylated or not) from a given protein within all three fractions could be searched for. This helps confirm that the protein is in fact present within the sample by increasing the sequence coverage of the candidate methylated protein. This directly increases confidence in identification of the protein, and therefore indirectly increases confidence in the identification of the methyl PTM site within it.

Another strategy to increase sequence coverage and reduce false discovery rates would be the use of a protease that cleaves with low frequency. Glu-C yields peptides of approximately 15 residues—a longer average sequence relative to the more commonly used protease trypsin (~9 residues).[128] The length of the Glu-C peptide digest puts them in the upper range of peptide fragments that are useful for bottom-up proteomics (0.7 kDa < M_w < 3.0 kDa). Collision energy for fragmentation must be increased since longer peptides require more fragmenting for effective mass spec analysis. If collision energy is not increased, the less fragmented peptides are susceptible to artificially high scores and false identification since there is less fragment data to accurately align the product ion spectrum to the database.[129] Nevertheless, longer peptides increase the

sequence coverage of a single protein simply by increasing the number of analyzed residues per peptide. If one wanted to further increase methylated peptide sequence coverage, the use of cyanogen bromide instead of a protease would yield even longer peptides averaging ~38 amino acids.[128] This chemical reagent cleaves after the scarcely occurring methionine reside. This peptide length is within the size range of "middle-down" proteomics (3.0 kDa < M_w < 10 kDa)—an emerging method that further increases sequence coverage of the proteome.[130]

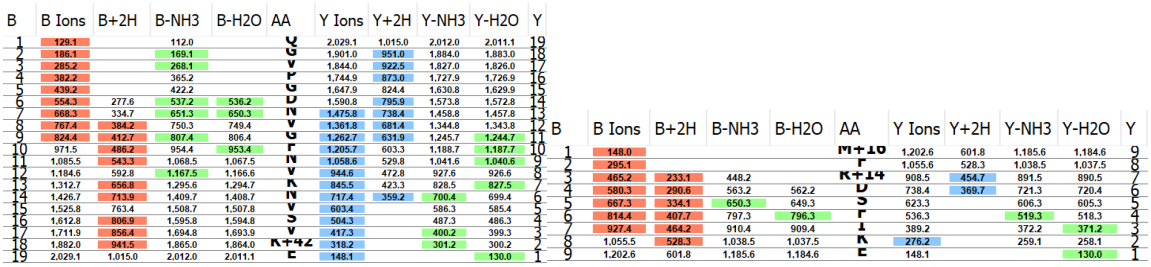
False discovery rates (FDR)s are of great concern within methyl proteomics.[131] As such, manual validation of methylated peptide spectra is required when analyzing methylated peptide hits. As previously discussed in Chapter 1, proteomics software aligns experimental parent spectra to expected parent spectra within a specified ppm error range (also termed mass error). Following alignment of the parent spectra, the product spectra are assigned scores to validate the hit. For the parent ion alignment, higher mass accuracy means less possibility that a calculated elemental formula produced the observed mass within the instrument's range of error. The formula for ppm error is as follows:

$$\text{ppm error} = \frac{m/z_{observed} - m/z_{calculated}}{m/z_{calculated}} \times 10^6$$

Error is most often reported as ppm error. This thesis allowed a ppm error of 5.0 for the parent fragment and a 0.8 Da MS/MS tolerance. The accepted mass tolerance range is chosen based on the mass analyser's accuracy. For comparison, a moderate mass spec accuracy is ~10 ppm—a practical limit of an instruments accuracy.[132] The peptides were analyzed with a state-of-the-art Orbitrap Fusion™ Tribrid™ MS—the top MS for in-depth discovery of complex samples with the highest accuracy. Regardless of our highly accurate mass spec, instrument error still allows for false identifications.

The fragmentation pattern of the product ion spectra is first analyzed to gain confidence in the identification, and the parent ion ppm error is the final quality check of a spectrum. Insufficiently fragmented peptides, or peptides with few fragments associated with the expected mass loss from loss of residue(s) is not the basis of a methylation mark identification. But the software does not take this into sufficient consideration. Because of this deficiency of the software, fragmentation patterns of the identified putative methylation marks are manually scrutinized. Figure 2.13 left shows the fragmentation pattern of a highly fragmented ion. Many B and Y ions matched the expected mass loss of the fragmented peptide sequence. A B ion is the ion fragment extended from the amino termini, and the Y ion is the ion fragment extended from the carboxyl

termini. The confidence of this peptide identification is increased by manually validating this identification. Figure 2.13 right shows the fragmentation pattern of an insufficiently fragmented/detected peptide that lowers the confidence of the methylation mark identification. Many B ions were identified, but the lack of Y ions lowers the validity of this identification.



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	129.1		112.0		V	2,029.1	1,015.0	2,012.0	2,011.1	19
2	186.1		169.1		S	1,901.0	951.0	1,884.0	1,883.0	18
3	285.2		268.1		V	1,844.0	922.5	1,827.0	1,826.0	17
4	382.2		365.2		F	1,744.9	873.0	1,727.9	1,726.9	16
5	439.2		422.2		G	1,647.9	824.4	1,630.8	1,629.9	15
6	554.3	277.6	537.2	536.2	D	1,590.8	795.9	1,573.8	1,572.8	14
7	668.3	334.7	651.3	650.3	N	1,475.8	738.4	1,458.8	1,457.8	13
8	767.4	384.2	750.3	749.4	V	1,361.8	681.4	1,344.8	1,343.8	12
9	824.4	412.7	807.4	806.4	S	1,262.7	631.9	1,245.7	1,244.7	11
10	971.5	486.2	954.4	953.4	F	1,205.7	603.3	1,188.7	1,187.7	10
11	1,085.5	543.3	1,068.5	1,067.5	N	1,058.6	529.8	1,041.6	1,040.6	9
12	1,184.6	592.8	1,167.5	1,166.6	V	944.6	472.8	927.6	926.6	8
13	1,312.7	656.8	1,295.6	1,294.7	K	845.5	423.3	828.5	827.5	7
14	1,426.7	713.9	1,409.7	1,408.7	N	717.4	359.2	700.4	699.4	6
15	1,525.8	763.4	1,508.7	1,507.8	V	603.4		586.3	585.4	5
16	1,612.8	806.9	1,595.8	1,594.8	S	504.3		487.3	486.3	4
17	1,711.9	856.4	1,694.8	1,693.9	V	417.3		400.2	399.3	3
18	1,882.0	941.5	1,865.0	1,864.0	K+42	318.2		301.2	300.2	2
19	2,029.1	1,015.0	2,012.0	2,011.1	E	148.1			130.0	1

B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	148.0				M+16	1,202.6	601.8	1,185.6	1,184.6	9
2	295.1				F	1,055.6	528.3	1,038.5	1,037.5	8
3	465.2	233.1	448.2		K+14	908.5	454.7	891.5	890.5	7
4	580.3	290.6	563.2	562.2	D	738.4	369.7	721.3	720.4	6
5	667.3	334.1	650.3	649.3	S	623.3		606.3	605.3	5
6	814.4	407.7	797.3	796.3	F	536.3		519.3	518.3	4
7	927.4	464.2	910.4	909.4	I	389.2		372.2	371.2	3
8	1,055.5	528.3	1,038.5	1,037.5	K	276.2		259.1	258.1	2
9	1,202.6	601.8	1,185.6	1,184.6	E	148.1			130.0	1

Figure 2.13 Manual validation of Left: the monomethyl mark K90Me on the protein phosphoglycerate kinase (accession number PGK_YEAST) by looking at fragmented B and Y ions for the entire sequence. Right: The manual rejection of the monomethyl mark R48Me on the protein K7_Ufe1p (accession number G2WMV3_YEASK) due to the low number of Y ions observed.

To further increase the confidence that a candidate methylated peptide has been discovered, manually checking the product ion spectrum/model error—specifically, the orientation and magnitude of the error. If the errors are of equal magnitude and are exclusively positive or exclusively negative, the spectrum is deemed high quality—the reason being that the error is most probably from instrument calibration or instrument error rather than an unfitting alignment. Refer to Figure 2.14 a for an example of spectral error associated with the fragmentation pattern of Figure 2.13 left. The magnitudes are equal, and the delta mass errors are within a ~0.25 delta mass error (AMU) range.

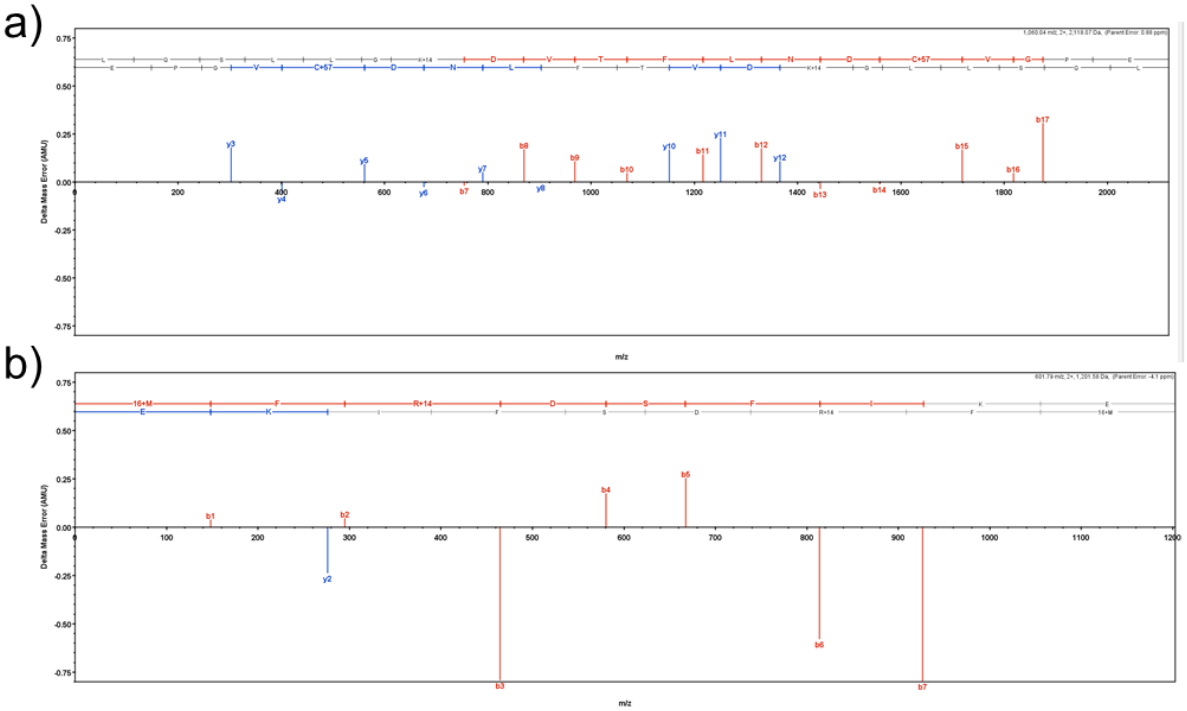


Figure 2.14 Manual validation of a) the monomethyl mark K90Me on the protein phosphoglycerate kinase (accession number PGK_YEAST) by looking at the spectrum/model error. Peptide error hovers around zero with similar magnitudes and sign. b) The manual rejection of the monomethyl mark R48Me on the protein K7_Ufe1p (accession number G2WMV3_YEASK) due to the spectrum/model error being sporadic in value and opposite in sign.

But, if the error was variable in magnitude and/or sign (positive or negative), then the hit was most likely an artifact. Refer to Figure 2.14 b for the corresponding spectrum/model error in Figure 2.13 right. Errors range between ~1 AMU in both the positive and negative sign for the B ions and the single Y ion observed. This is not an acceptable identification.

After analyzing the product ion data, reassurance that the parent ions mass error was within a reasonable range increased confidence in the methyl identification.[133] Parent error should be around 1 ppm error for highly confident identifications.[134] The parent ion ppm error of Figure 2.13 left Figure 2.14 top was 0.88 ppm. With consideration of the fragmentation pattern and the spectrum/model error, the confidence in this identification is high. As for Figure 2.13 right and Figure 2.14 bottom, the parent ion ppm error was –4.1. With consideration of the questionable error of product ions, lack of Y ions, and high ppm error of the parent ion, this identification is of low confidence.

2.5 Conclusion

My first goal to improve protein extraction yields was achieved. Whole cell mechanical and physical lysis extracts higher protein yields than chemical nuclear extraction. A 33-fold increase in protein yields was obtained when whole PC-3 lysates were extracted via osmotic pressure and bead beating relative to chemical lysis and nuclear extraction. Additionally, interfering reagents were not introduced in the optimized lysis method. As such, a detergent removal step (which resulted in a 30% sample loss) that was previously necessary for the nuclear extraction protocol was no longer required.

My second goal to remove non-protein cellular debris and optimize binding was partially achieved. Based on relative retention, the lower-rim modified *p*-sulfonatocalix[4]arene was found to be a better affinity reagent that the upper-rim modified *p*-sulfonatocalix[4]arene. Thus, column retention was increased. Moreover, before optimized sample preparation, peptides did not retain in the MethylTrap column at all. After implementing protein purification with chloroform/methanol protein precipitation, peptides successfully retained in the MethylTrap column. As such, the goal to remove non-protein cellular debris to optimize retention was met. Following this, the peptides in each fraction were identified. The number of methylated peptides identified in the unretained fraction outnumbered the identified methylated peptides in the retained fraction by 4-fold. As such, the goal to optimize binding was, in part, met. But, binding of the intended target, methylated peptides, was not met and requires further optimization.

My final goal, to optimize the number of methylated peptides identified, was not fully achieved. Protease digestion with Glu-C identified more methylated peptides in comparison to trypsin and Arg-C digests. This could be explained by the previously reported charge bias of the MethylTrap column and non-limited number of basic residues in Glu-C digests relative to Arg-C and trypsin.[98] But, the number of methylated peptides identified could be further increased with optimization of methylated peptide retention. As such, the goal to optimize the number of methylated peptides identified was not met.

MethylTrap enrichment helped identify novel methylated peptide marks that have not been previously published based on alignment to the UniProtKB modified residue database. But, the number of identifications within the methylated peptide dataset is most probably inflated, so the number of novel methylation sites discovered is less than that reported. Regardless, a portion of

the novel sites can be assumed to be true (and also novel) identifications, so MethylTrap enrichment will ultimately improve characterization of the methylome.

MethylTrap methylated peptide retention showed a 3% reproducibility in unique methylated peptides identified in at least two of the three PC-3 mammalian cell biological replicates. 16 out of 470 observed methylated peptides were reproducibly observed. This is an unacceptably poor level of reproducibility. Further studies regarding reproducibility are required.

2.6 Experimental methods

Refer to supporting information for step-by-step optimized MethylTrap enrichment protocol.

2.6.1 PC-3 mammalian cell protein extraction

PC-3 mammalian cells were grown in T-175 culture flasks using RPMI 1640 culture media supplemented with 10% Fetal Bovine Serum (FBS) (Sigma Aldrich). Cell were maintained at 37℃ with 5% CO_2 in a humidified incubator.

Biological replicates were collected from PC-3 cells at different passage numbers. Two T-175 flasks of PC-3 cells were trypsinized then washed with 5 mL of ice-cold phosphate-buffered saline (PBS) (10 mM phosphate, 137 mM NaCl, 2.7 mM KCl, pH 7.4) three times by centrifugation for 5 minutes at 350 gs and 4℃. Subsequently, the cells were resuspended in 500 µL of ice-cold hypotonic buffer (Active Motif) supplemented with complete protease inhibitor cocktail (Sigma-Aldrich). 0.5 mm acid-washed glass beads (Sigma-Aldrich) were added at a 2:1 bead-to-sample ratio. The cells were lysed with seven cycles of vigorous vortexing for 2 minutes followed by 1-minute rests on ice. Crude cell lysate was isolated from the beads with a sterile 22-gauge needle (Thermo).

2.6.2 Protein processing

Crude lysate was denatured and reduced with 3.5 M Guanidine Hydrochloride (GdnHCl) and 50 mM dithiothreitol (DTT) (Sigma-Aldrich) for 30 minutes at 65℃. The reduced lysate was cooled to room temperature, then alkylated with 100 mM iodoacetamide (IAA) (Sigma-Aldrich) for 30 minutes in the dark at room temperature.

The alkylation was quenched in a 4:1:3 volume ratio of methanol, chloroform, and water. Following, interfacial protein precipitation was induced with centrifugation at max rpm for 15

minutes. The top aqueous methanol was discarded to remove hydrophilic contaminants. The same volume of methanol that was initially added replaced the aqueous phase. The sample was then centrifuged again to isolate the precipitated proteins from hydrophobic contaminants. The supernatant was removed, and protein pellet was washed with the same volume of methanol once more. The protein pellet was dried, then resuspended in 50mM NaOH 3x its volume. Afterward, the sample was placed in an ultrasonic water bath for approximately 10 minutes. The protein pellet was further diluted in 6x its dry volume with 50 mM ammonium bicarbonate buffer (pH 7.8) (Sigma-Aldrich). A further 10-minute sonication bath concluded protein solubilization. The residual precipitant was pelleted with 3,750gs at 4℃ for 30 minutes. The solubilized protein solution was assayed with Bio-Rad Protein Assay Dye Reagent (Bio-Rad).

1 mg of solubilized proteins were digested at a 1:100 ratio of protein to Glu-C (Promega) for 18 hours at 37℃ while shaking. Next, the digest was buffer exchanged with PD minitrap G10 desalting column (GE Healthcare). Following the buffer exchange, the peptides were thus eluted into 50 mM phosphate buffer pH 7.5.

The yeast peptides were injected into AKTA Prime Plus (GE Healthcare) connected to the calixarene column (details in supplemental info). The unretained peptides were washed with 11.3 mL of binding buffer, then gradient eluted with binding buffer spiked with 2 M ammonium chloride salt over 12.4 ml. Retained peptide fractions were pooled then assayed with PierceTM Quantitative Fluorometric Peptide Assay (Thermo) and subsequently analyzed at the UVic-Genome BC Proteomics Centre with an Orbitrap LC-MS/MS.

Based on concentrations provided the sample was acidified with Formic acid final concentration (0.5%) before passing over an Oasis HLB µ-elution plate SPE (30 um) (Waters part#186001828BA) for desalting and sample clean-up to provide a ~30 µg loading. Following binding and washing, peptides were eluted with 180 µL (80% v/v Acetonitrile, 0.1% v/v Formic acid), speed vacuumed to near dryness and rehydrated with 20 µL 2% Acetonitrile, 0.1% Formic acid, water.

2.6.3 LC-MS/MS analysis

The peptide mixture (5 µL) was separated by on-line reverse phase chromatography using a Thermo Scientific EASY-nLC 1000 system with a reverse-phase pre-column Magic C18-AQ (100 µm I.D., 2.5 cm length, 5 µm, 100Å, and an in-house prepared reverse phase nano-analytical column Magic C-18AQ (75 µm I.D., 15 cm length, 5 µm, 100Å, Michrom BioResources Inc,

Auburn, CA), at a flow rate of 300 nl/min. The chromatography system was coupled on-line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a Nanospray Flex NG source (Thermo Fisher Scientific). Solvents were A: 2% Acetonitrile, 0.1% Formic acid; B: 90% Acetonitrile, 0.1% Formic acid. After a 300 bar (~ 8 µL) pre-column equilibration and 300 bar (~ 3 µL) nanocolumn equilibration, samples were separated by a 140-minute gradient (0 min: 5% B; 110 min: 35% B; 10 min: 45% B; 10 min: 100% B; hold 10 min: 100% B).

The Orbitrap Fusion instrument parameters (Fusion Tune 3.0 software) were as follows for orbitrap (OT-MS) iontrap (IT- MS/MS) with HCD fragmentation: Nano-electrospray ion source with spray voltage 2.45 kV, capillary temperature 275 °C. Survey MS1 scan m/z range 380-2000 profile mode, resolution 120,000 FWHM@200m/z one microscan with maximum inject time 50 ms. The Siloxane mass 445.120024 was used as lock mass for internal calibration. Data-dependent acquisition Orbitrap survey spectra were scheduled at least every 3 seconds, with the software determining "Top-speed" number of MS/MS acquisitions during this period. The automatic gain control (AGC) target values for FTMS and MSn were 400,000 and 5,000 respectively. The most intense ions charge state 2-7 exceeding 50,000 counts were selected for HCD and CID ion trap MSMS fragmentation with detection in centroid mode. Monoisotopic Precursor Selection (MIPS) was enabled and Dynamic exclusion settings were: repeat count: 2; repeat duration: 10 seconds; exclusion duration: 10 seconds with a 10 ppm mass window. The ddMS2 IT HCD scan used a quadrupole isolation window of 1.6 Da; IonTrap rapid scan rate centroid detection first mass 100 m/z, 1 microscan, 50 ms maximum injection time and stepped collision energy 40% ± 3. The ddMS2 IT CID scan used a quadrupole isolation window of 1.6 Da; IonTrap rapid scan rate centroid detection, auto-normal scan mass range, 1 microscan, 25 0ms maximum injection time and normalized collision energy 35%.

2.6.4 LC-MS/MS data analysis

Raw files were created by XCalibur 4.1.31.9 (Thermo Scientific) software and analysed with Proteome Discoverer 2.2.0.388 software suite (Thermo Scientific). Parameters for the Spectrum Selection to generate peak lists of the CID spectra (activation type: CID; s/n cut-off: 1.5; total intensity threshold: 0; minimum peak count: 1; precursor mass: 350-5000 Da).

The sample peak lists were submitted to a Mascot 2.5.1 server UP_human_Canon (69,085 sequences; 23,290,638 residues) database search for PC-3 samples and server UP_yeast_Canon

(76,618 sequences; 33,759,464 residues) database search for yeast samples as follows: precursor tolerance 5 ppm; MS/MS tolerance 0.8 Da; GluC enzyme 2 missed cleavages; FT-ICR instrument type; fixed modification: Carbamidomethylation (C); variable modifications: acetyl(K), methyl (K, R), dimethyl (K, R), trimethyl (K). Percolator settings: Max delta Cn 0.05; Target FDR strict 0.01, Target FDR relaxed 0.05 with validation based on q-Value.

Scaffold (version Scaffold_4.8.7, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 90.0% probability by the Peptide Prophet algorithm (Keller, A et al Anal. Chem. 2002;74(20):5383-92) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, Al et al Anal. Chem. 2003;75(17):4646-58). For GO analysis, Peptide identifications were accepted if they could be established at greater than 90.0% probability by the Peptide Prophet algorithm (Keller, A et al Anal. Chem. 2002;74(20):5383-92) with Scaffold delta-mass correction and protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 1 identified peptide. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Proteins were annotated with GO terms from NCBI (downloaded 12-August-2019 for both PC-3 and yeast samples). (Ashburner, M et al Nat. Genet. 2000;25(1):25-9). Venn diagrams directly exported from the Scaffold software.

2.6.1 Nuclear extraction of PC-3 mammalian cells

Nuclear extraction was performed with a Nuclear Extraction Kit (Active Motif) as described by the manufacturer. Detergent introduced during nuclear extraction was removed with the Pierce™ Detergent Removal Spin Column, 0.5 mL (Thermo) as described by the manufacturer.

2.6.2 Chelation of metals from protein extract

Trypsin digested peptides were incubated with 5% (w/v) Chelex-100 resin for 1 hour at 4°C. Chelex-100 resin was removed by centrifugation prior to desalting with the PD minitrap G10 desalting column (GE Healthcare). Peptides were subsequently enriched with the upper-rim modified *p*-sulfonatocalix[4]arene MethylTrap column. Enrichment and proteomics analysis were performed as previously described.

2.6.3 MethylTrap enrichment by batch binding

5 mg of upper-rim modified *p*-sulfonatocalix[4]arene was couple to 0.35 mL of Affigel 102 resin as previously described. 1 mg of purified yeast proteins were digested with tryspin at a 1:100 ratio for 18 hours at 37°C. If the digest was buffer exchanged, then it was performed with PD minitrap G10 desalting column (GE Healthcare). Following the buffer exchange, the peptides were thus eluted into 50 mM phosphate buffer pH 7.5. Peptide were added to MethylTrap resin and incubated for 3 hours at 4°C. Beads were washed with 300 µL binding buffer four times as the "unretained" fraction. The beads were washed an additional 12 times with 750 µL of binding buffer. These fractions were discarded. Beads were washed with 300 µL of NaCl four times as the "retained" fraction. Peptides were assayed with PierceTM Quantitative Fluorometric Peptide Assay (Thermo).

Chapter 3: Reproducibility of MethylTrap Enrichment and Applications to Industrial Yeast

3.1 Introduction

3.1.1 Reproducibility of MethylTrap enrichment and methylated peptide identification

The MethylTrap column has helped identify hundreds of novel methylated peptides. With optimized sample preparation, the MethylTrap column retained large numbers of peptides from cell lysates and helped identify methylated peptides that were not seen in neither the input sample nor the unretained sample. Additionally, the identified methylated peptides were 95% novel when aligned to the UniProt PTM database. This demonstrates its potential to aid characterization of the full methylome.

Reproducibility of MethylTrap retention requires further investigation. A 3% overlap in MethylTrap methylated peptide retention was observed between three biological replicates of PC-3 enrichment in Chapter 2. In this chapter, reproducibility has been scrutinized more thoroughly using a set of technical and biological replicates. In this thesis, technical replicates are defined as the same sample source being independently lysed, processed, then analysed, whereas biological replicates are when different sample sources of the same sample type are processed then analyzed.[135] This chapter focuses on the biological and technical reproducibility of industrial yeast cells. These samples were chosen for three reasons: an inherent interest in studying protein methylation in brewers' yeast, the ready accessibility of large amounts of cell-based samples available from commercial-scale fermentations, and the interest in extending MethylTrap enrichment methods to include new industrial applications.

Possible causes for variability between replicates could be protein extraction, protein purification, alkylation, digestion, and desalting. The largest contributor to sample preparation variability is reported to be the protein extraction step. Protein extractions have been reported to cause 70% of the observed variability between replicates.[123,136] Another cause for variability would be the protein precipitation step. Recovery following protein precipitation is never 100%. Protein recovery for chloroform/methanol precipitation is reported to be around 88±12% from identical samples, so ~11% of the sample is lost during protein precipitation and resolubilisation.[113] The

protein precipitation was noted to be a reproducible purification step for the same sample type, but is inconsistent between different samples.[137] This could potentially explain the variation between biological replicates, but should not pose much variability for the technical replicates (with the assumption that protein extraction concentrations are consistent). Denaturing, alkylating, digestion, and desalting has also been reported to contribute to 3.1% of the overall variance.[123] This number is of small concern when compared to the 70% influence from protein extraction.

Poor reproducibility of MethylTrap enrichment could be a large contributor to variability between runs for the retained and unretained sample. As discussed in Chapter 2, a potential cause for variable MethylTrap enrichment could be derived from the input sample. Changes in input peptide concentrations arise from the compounded variability from previous sample preparation steps (extraction, purification, etc.). The difference in input peptide concentrations will alter the relative binding interactions of the system, thus altering selectivity and retention in the MethylTrap column. This could explain the very small 3% overlap between the retained biological replicates of PC-3 methylated peptides presented in chapter 2.

LC-MS/MS injections are a known source of inconsistency between any set of proteomic replicates. The difference between LC-MS/MS injections alter the set of analyzed peptides between the range of 35 – 60 %.[124] It has been reported that LC-MS/MS injections contribute to 16% of the overall variability between replicates.[123] Additionally, only 16% of the sample's peptides are detected in a shotgun proteomics MS scan.[56] Reasons for incomplete sampling is in part due to ionisation efficiency. Not all the sample's peptides are analyzed because a subset of the peptides are preferentially or sporadically ionized. Another reason for incomplete sampling is during the selection of the precursor ion for fragmentation. In a complex sample, many peptides elute at a single LC retention time. With that, only the most intense precursor ions are selected for fragmentation due to the limited duty cycle of the precursor ion selection window.[138] Miniscule differences in parent ion peak intensities alter which ions are selected for tandem mass spec analysis.[127] This highly sensitive selection process of peptides with similar intensities cumulatively causes large differences in the final proteomic datasets of the same sample type.[139]

Replicates normalize wet lab and LC-MS/MS variability. Differences between datasets are most likely attributed to variability associated with sample preparation and LC-MS/MS analysis rather than protein abundance itself.[138] More variability between datasets arises as more steps are implemented in the sample preparation. By combining the data of numerous replicates, under-

sampling by protein processing and stochastic ionisation or precursor ion selection in a single run would be minimized.[140]

3.1.2 Gene ontology to bring biological meaning to lysine methylation

Gene ontology (GO) brings biological meaning to large datasets generated during shotgun proteomics. It is a hierarchical analysis of gene products and their role within the phenotype.[141] Researchers use GO to open up new avenues for future research by investigating the role of a protein product.[142] One could investigate the biological role of a PTM within the cell. Or, one could compare enriched gene expression profiles between healthy and cancerous cells to pinpoint oncogenes. Investigating the functional relationship of PTM GO classification in diseased versus healthy cell could aid in the identification of biomarkers associated with the diseased PTM state.

Three categories of gene ontology exist: biological processes, molecular function, and cellular component.[143] Biological processes refer to the contributions of the gene product to cellular function. Examples of high-level generalized (broad) terms are processes such as "signal transduction" or "response to stimuli." More specific and precisely defined (narrow) terms would include "translation" or "ATP biosynthetic process." Molecular function is the biochemical activity or functional component of the gene product. Broad terms for molecular function include "binding" and "catalytic activity." More narrow terms include functions such as "motor activity" or "adenylate cyclase activity." Cellular component is the localization of the gene where is performs its cellular function. Broad level terms include "membrane" or "organelle." More narrow terms include "ribosome" or "membrane sack."[141,144] Localization is especially interesting for methylation analysis, as so many of the previously identified methylation sites occur on histones, which are localized in the nucleus.[42]

GO analysis has been previously used to annotate the use and localization of methyllysine PTMs in cells. Monomethyl lysine proteins been previously annotated to mitochondrial functions and RNA-binding.[145] Proteins enriched by 3xMBT, an enrichment tool that is pan-specific for mono- and dimethyllysines, overrepresented the GO terms mRNA processing, transcription, and RNA and DNA helicase activity.[146] Non-histone proteins that interacted with chromobox protein homolog 6 (CBX6) chromodomain, a subunit of Polycomb Repressive Complex 1 (PRC1) recruited by trimethylation on H3K27, were annotated as transcriptional regulators and chromatin modifiers.[147] In a global immunoenrichment study of methyllysine proteins, the overrepresented GO terms were chromatin organization, N-methyltransferase activity, DNA binding, and motor

activity.[148] Based on the previously annotated GO terms for methyllysine proteins, the modified proteins seem to be largely involved in binding, transcriptional regulation and chromatin organization/modifications. These are all terms that would be expected because methylation has been so strongly associated with high-abundance histones and histone-associated proteins. These processes are being investigated in-depth to delineate the role of lysine methylation in cellular function.[149]

3.1.3 Objectives

The first objective was to enrich and identify methylated peptides in real-world industrial brewers' yeast samples.

My next objective was to assess the reproducibility of the identifications discovered with MethylTrap enrichment.

My last objective was to survey the biological function and cellular compartment of the methylated peptides identified in the yeast and PC-3 cells using GO analysis.

3.2 Experimental methods

Two second generation re-pitched Mexican Lager yeast strains from two fermenter vessels at Phillips Brewing were used as biological replicates (BR). Two technical replicates (TR) of each BR were individually lysed, enriched by the MethylTrap column, then analyzed by LC-MS/MS analysis.

Refer to Chapter 2's experimental method section for the MethylTrap enrichment protocol.

3.3 Results and discussion

3.3.1 MethylTrap fractionation of industrial yeast

As with the PC-3 results in Chapter 2, MethylTrap enrichment of industrial yeast samples identified unique methylated peptides that are otherwise unseen in the input (Figure 3.1). 11% of the identified methylated peptides were exclusive to the retained fraction, and 29% were exclusive to the unretained sample. Enrichment unmasked otherwise undetected methylated peptides in the input sample since 96% and 82% of the retained and unretained peptides were unseen in the input. MethylTrap fractionation increased the number of methylated peptide identifications by 42%.

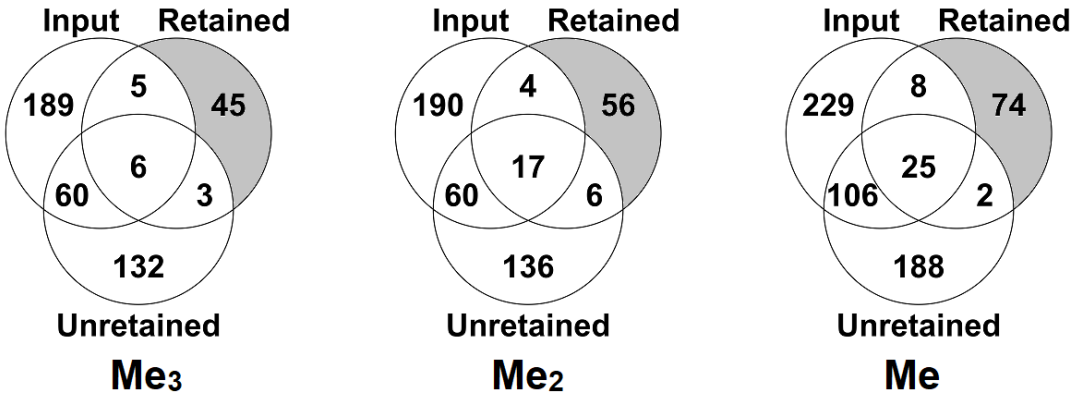


Figure 3.1 MethylTrap fractionation increases the number of unique identified methylated peptides relative to the input sample in a yeast proteomics experiment. Yeast cell lysates were enriched using MethylTrap, and the input control sample, as well as the MethylTrap retained and unretained fractions were submitted for proteomics analysis. All methyl PTM sites in each sample were tabulated and counted. Venn diagrams show the overlap in the number of unique mono- (left), di- (middle), and tri-methylated (right) PTM sites identified. The shaded sectors show methyl PTMs that are only visible after MethylTrap enrichment. Venn diagrams include unique methylated peptides from two biological and two technical replicates of yeast proteins.

3.3.2 Reproducibility of MethylTrap enrichment

Peptide identifications are more technically reproducible than methyl PTM identifications (Table 3.1). For all identified peptides, the retained sample reproducibly retained 13% and 35% of the same peptides, the unretained 36% and 33%, and the input 35% and 33% for the set of TR for BR 1 and BR 2, respectively. The reproducibility for the TRs for both BRs' methylated peptide identifications were 0% and 7.4% for the retained fraction, 7.3% and 9.2% for the unretained sample, and 7.2% and 5.1% in BR 1 and BR 2, respectively. These numbers are between 33 – 36% for peptide ID and between 5.1 – 9.2% for the methylated peptides, with the exception of the retained fraction for BR 1. One of the retained TR of BR 1 identified zero Me_2 and Me peptides, and only two Me_3 peptides (refer to Figure S 2 – 4 for Venn diagrams). The 49 and 68 identifications for Me_2 and Me identifications were attributed to one of the TR of BR 1. Sample preparation was performed with the identical MethylTrap enrichment protocol, so the cause for the low methylated peptide identifications in such sample is unknown. Regardless, average reproducibility of peptide IDs between all TRs was $31 \pm 8.8\%$ and $6.0 \pm 3.2\%$ for all peptides and methylated peptides, respectively.

Table 3.1 Overlap in methylated peptide identifications between yeast technical replicates for each biological replicate (represented by '1' and '2') of the MethylTrap enrichment fractions. Refer to Figures S 2 – 4 for overlap between technical replicate Venn diagrams.

Overlap / all identified

	Me_3		Me_2		Me		All IDs	
Fraction	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
Retained	0 / 37	1 / 23	0 / 49	5 / 37	0 / 68	2 / 48	106 / 807	521 / 1476
Unretained	7 / 104	12 / 132	5 / 129	8 / 137	19 / 194	25 / 220	1449 / 3934	1152 / 3439
Input	8 / 151	0 / 140	10 / 141	8 / 168	18 / 209	20 / 238	1163 / 3370	1133 / 3403

Peptide identifications were most reproducible in the unretained fraction and least reproducible in the retained fraction (Figure 3.2).

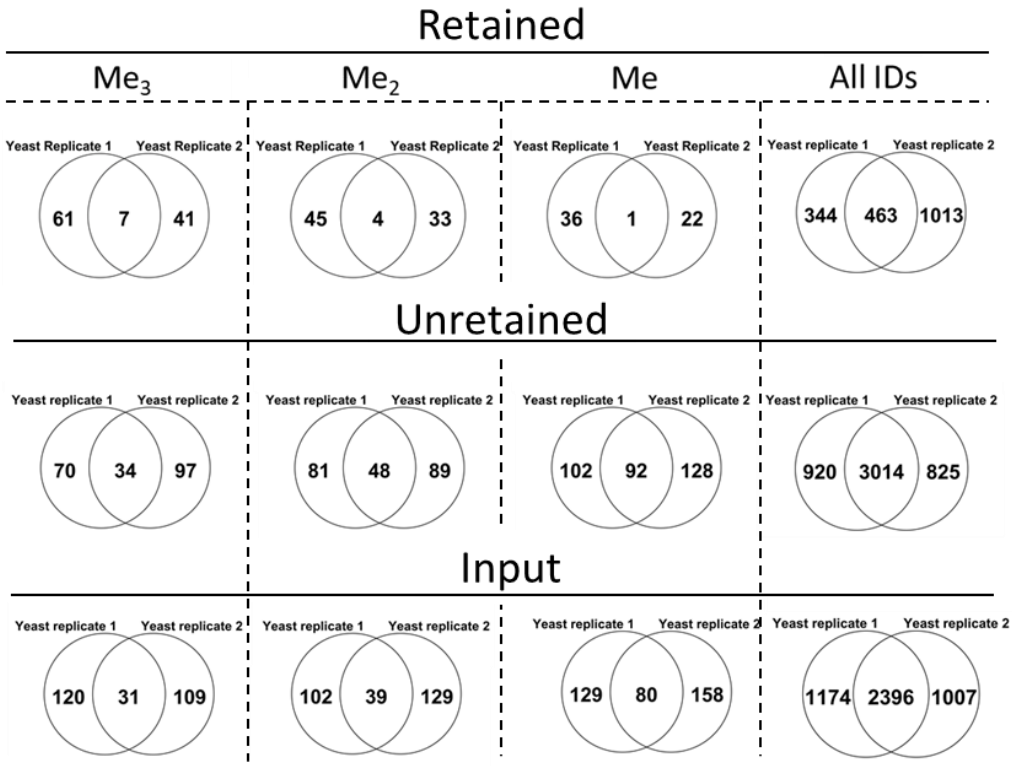


Figure 3.2 Yeast biological replicates had better reproducibility for all peptide ("All IDs") than for methylated peptides, and the retained sample has lowest reproducibility. Results from two technical replicates were combined and counted together to create data for a single biological replicate (e.g. "Yeast replicate 1"). Venn diagrams of the overlap of identifications of individual methyl marks, as well as for all peptide identifications, are shown. Input, Unretained, and Retained fractions (see chapter 2) were processed and analyzed separately. The column labeled "all IDs" includes data for all of the identified peptides in each fraction (methylated and non-methylated alike).

For all identified peptides (methylated and unmethylated alike), 63% were present in both unretained biological replicates, 52% for the input, and 25% for the retained. For the methylated peptides, 23% of the unretained were shared between biological replicates, 17% for the input, and 4.8% for the retained. Based on the overlap of BR input peptide identifications (63%), sample preparation seems reproducible, but methylated peptide identification is not as reproducible. The lowered reproducibility for the methylated peptide identifications relative to the peptide identifications across all fractions may be attributed to their low abundance in the complex sample. This is the confounding issue of PTM studies that enrichment hopes to minimize. Once the complexity of the precursor ion exceeds the instrument's duty cycle, stochastic selection of higher abundance ions undergo fragmentation.[56] A method to increase reproducibility across replicates could be to align the replicate's retention times for the peptide precursors.[139] Retention time alignment demonstrated a 19% increase in reproducibility between phosphopeptide enrichment studies.

3.3.3 Comparison of reproducibility between MethylTrap enrichment and antibody enrichment

Low reproducibility is common to many methyl enrichment protocols. To make this conclusion, we examined results from proteomics done with anti-methyllysine antibodies described in literature. Bremang *et. al.* scrutinized five polyclonal anti-methyllysine antibodies purchased from Acris, Lifespan Biosciences, AbCam, and Immunechem.[81] Five anti-methyllysine immunoprecipitations on the same HeLaS3 sample were analyzed by a single LC-MS/MS injection. A grand total of 54 methylation sites were observed between the five anti-methyllysine antibodies. 4% of the methyl identifications were seen in all five immuno-enriched samples, and 60% of the identifications were exclusive to one of the five. In a paper by Guo *et al.*, immunoaffinity purification of HCT116 cell methylated proteins with three lab-produced pan-specific polyclonal antibodies (BL10749, BL10745, and D3978) led to the discovery of 132 monomethyl, 35 dimethyl and 31 trimethyl sites.[150] By inspecting the supporting information, we learned that there was an overlap of 10% in methylated peptide sequences between at least two of the three replicates and that only 6% of marks were seen in all three replicates. Directly comparing proteomic results that were obtained with different types of replicates, LC-MS/MS systems, databases, and statistics is not a straightforward process, but based on these results, MethylTrap enrichment of methyllysines had comparable reproducibility in the retained fractions. 5% of the

methylated peptides were in common between the two MethylTrap enriched BR. In total, there was a 20% overlap in the observed methylated peptides that were in common between the BR MethylTrap enrichment studies. In comparison to literature, immunoprecipitation and MethylTrap enrichment demonstrate similar enrichment reproducibility.

More technical replicates for each BR could increase reproducibility of enrichment. In this part of the thesis, two TR were run for each BR. In general, at least three replicates are required for reproducible proteomic analysis.[124] It has been reported that ten experiments must be replicated to obtain ~95% protein identification saturation in complex protein samples.[151] The number of replicates required for proteomic analysis increases with sample size and complexity.[152] An assessment of sample complexity and the number of new unique methylated peptides identified per run would help assess the number of replicates required to truly gauge the number of TR required to compare reproducibility between BRs.

3.3.4 GO analysis: Cellular component

The MethylTrap column identified methylated peptides annotated to diverse cellular compartments (Figure 3.3). Within literature, the most observed methylated proteins are within the nucleus on histone proteins. This is of no surprise being histone proteins are the most detected methylated proteins.[38] Now, growing bodies of evidence point towards global employment of methylation for cell regulation pathways, but they are still less detected than nuclear methylated proteins.[38] Methylated peptides identified within the yeast and PC-3 cells were associated with proteins localized throughout the cell (Figure 3.3). The broad distribution of gene ontology terms for Cellular Compartment indicates broad and diverse cell protein sampling. Based on these results, cytoplasmic methylated proteins are comparably observed with respect to nuclear methylated proteins. More non-nuclear methylated proteins will be discovered with the use of enrichment technology and follow-up studies to validate and learn about individual methyl marks.

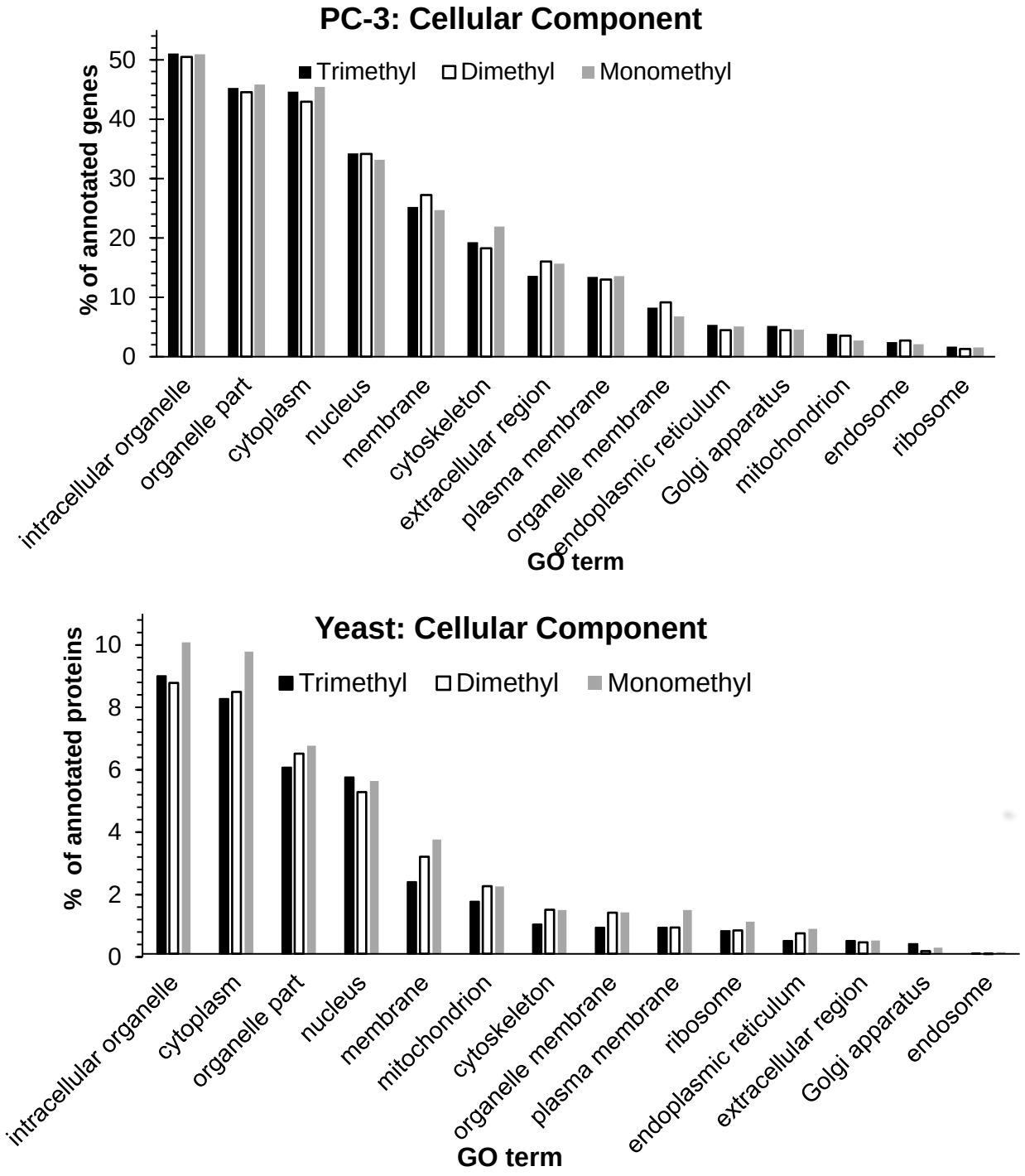


Figure 3.3 Gene ontology cellular component annotation of proteins with a mono, di, and trimethyllysine mark discovered within two biological and technical replicates for PC-3 (top) and yeast (bottom).

3.3.5 GO analysis: Molecular function

GO analysis showed an over-representation of broad molecular function terms for identified methylated proteins within both PC-3 and yeast datasets (Figure 3.4). Molecular function was associated with broad terms such as binding (GO:0005488), structural molecule activity (GO:0005198), catalytic activity (GO:0003824), enzyme regulator activity (GO:0030234), transporter activity (GO:0005215), and a low percentage of antioxidant activity (GO:0016209). These indicate diverse molecular functions of the identified methyl proteins.

The narrow term electron carrier activity (GO:0009055) implies the role of lysine methylation in processes that regulate the transmembrane electrochemical gradient in both yeast and PC-3 cells. The methylated protein electron transfer flavoprotein β subunit (ETFB) was the one trimethylated protein identified within this dataset. ETFB transfers electrons from mitochondrial dehydrogenases to the main mitochondrial respiratory chain. To offer further credibility to this identification, this protein was previously discovered to be trimethylated by protein N-lysine methyltransferase METTL20 in 2014.[153] A study from 2018 by Shimazu *et. al.* demonstrated the non-essential regulation of ETFB through trimethylation on its regulation loop.[154] Loss of methylation was associated with higher ETFB catalytic activity which resulted in higher resistance to starvation. Mutations at such methylation site are known to cause metabolic disorders.[155]

Another shared narrow term within the yeast and PC-3 molecular function dataset is motor activity (GO:0003774)—a form of catalytic force generated by nucleoside triphosphate hydrolysis that propagates protein movement along microfilaments or microtubules. This is not surprising since methylation has been implicated in motor function since the late 60s.[156,157,158] There is evidence that methylation is required for the skeletal myosin to bind to ATP.[159] In support of this hypothesis, depletion of methyltransferase apical complex lysine methyltransferase (AKMT) in the human parasite *Toxoplasma gondii* inhibits activation of mobility and invasion.[160] The role of AKMT is involved with the positioning of glideosome-associated connector (GAC) on actin through an unknown methylation mechanism.[161] The thought is that it is activated by a methyl-phospho switch.[32] Further characterization of this process could be obtained if global identification of methylated substrates before and after AKMT depletion was performed. Putative AKMT substrates arising from such a proteomics study could be confirmed by microarray analysis.

Inferences could be made that the molecular function of methylation for GO terms only present in PC-3 are specific to mammalian cell types or prostate cancer, but the absence of these

GO terms in yeast could be an artifact of incomplete sampling. Just because some things weren't detected does not mean they are not there. As such, I cannot confidently compare the usage of methylation within yeast and PC-3 cells, but I can look at the usage of methylation in PC-3 molecular functions. The two broad molecular function GO terms exclusive to PC-3 were translation regulator activity (GO:0045182) and molecular transducer activity (GO:0060089). The narrow term chemoattractant activity (GO:0042056) was also identified. Not much meaning can be derived from the broad terms other than a general association of translation regulation and control of regulatory components with methylation. As for the small sliver of di and trimethylated proteins in the PC-3 sample associated with chemoattractant activity, this percent was completely attributed to the protein collectin-10 (accession number:Q9Y6Z7), a lectin that regulates cell migration towards the sugar molecule. No known lysine methylation marks were previously reported for collectin-10. This may be a novel methylation pathway that could be confirmed by more detailed follow up studies.

The broad term exclusive to yeast was protein tag (GO:0031386). Again, there was only one protein associated with this GO term, and it was the ubiquitin-like protein, SMT3 (accession number:Q12306). This protein is deposited on other proteins as a SUMOylation PTM and it is involved with the regulation of gene expression.[162] Lysine methylation was not previously reported on this protein, but thas been noted to be phosphorylated.[163] There is no current literature discussing the purpose of PTMs on SMT3.

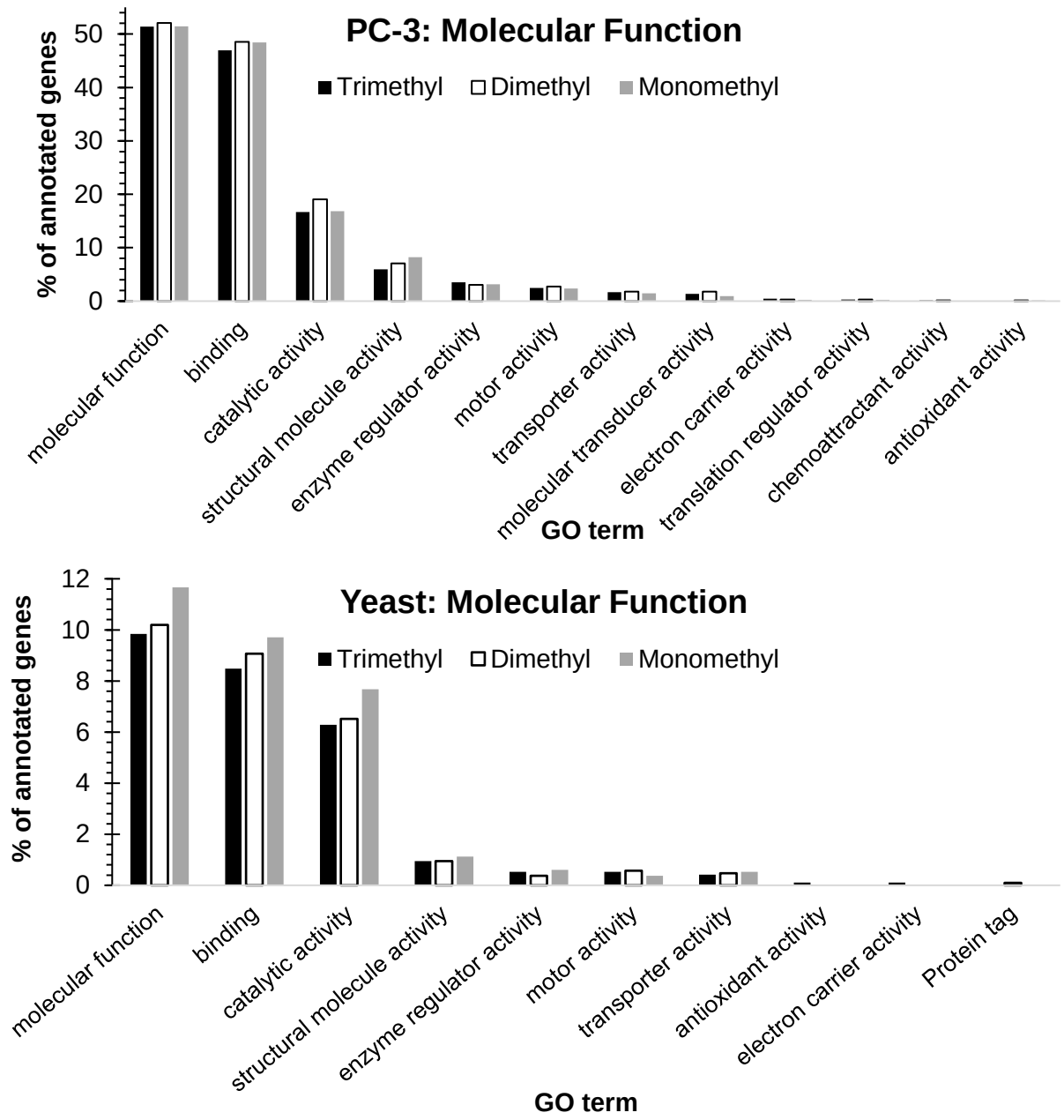


Figure 3.4 Gene ontology molecular function annotation of proteins with a mono, di, and trimethyllysine mark discovered within two biological and technical replicates for PC-3 (top) and yeast (bottom).

3.3.6 GO analysis: Biological processes

Methylated proteins were annotated to diverse and broad biological processes GO terms in both yeast and PC-3 cell samples (Figure 3.5). The broad level terms associated with both PC-3 and yeast were as follows: cellular process (GO:0009987), metabolic process (GO:0008152), biological regulation (GO:0065007), response to stimulus (GO:0050896), localization (GO:0051179), multicellular organismal process (GO:0032501), developmental process

(GO:0032502), reproduction (GO:0000003), multi-organism process (GO:0051704), growth (GO:0040007), and locomotion (GO:0040011). These are all broad and diverse terms that only show that methylation is employed in a broad array of biological processes. The narrower terms shared by both cell types were establishment of localization (GO:0051234) and reproductive process (GO:0022414). Establishment of localization is still fairly broad as it details genes involved with movement, binding, or selective degradation. Same with reproductive process since it is any gene associated with the cell cycle. Regardless, lysine methylation has been previously correlated with reproductive processes such as the cell cycle.[164,165] A classic example would be p53.[4] Confirmation and further analysis of the methylated proteins annotated to reproductive processes could shed light on its function within this biological process.

Exclusive to the PC-3 dataset were the broad terms biological adhesion (GO:0022610), rhythmic process (GO:0048511), cell killing (GO:0001906), and pigmentation (GO:0043473). Methylation is thus correlated with intracellular attachment, physiology rhythms, necrosis, and pigment accumulation in PC-3 mammalian cells. Methyltransferases have been linked to regulation of physiology rhythms, but the substrates involved with the process are not known.[166,167,168] MethylTrap enrichment before and after methyltransferase depletion or various contexts of physiology rhythms could help characterize the substrates of these methyltransferases and the use of methylation within the context of physiology rhythms.

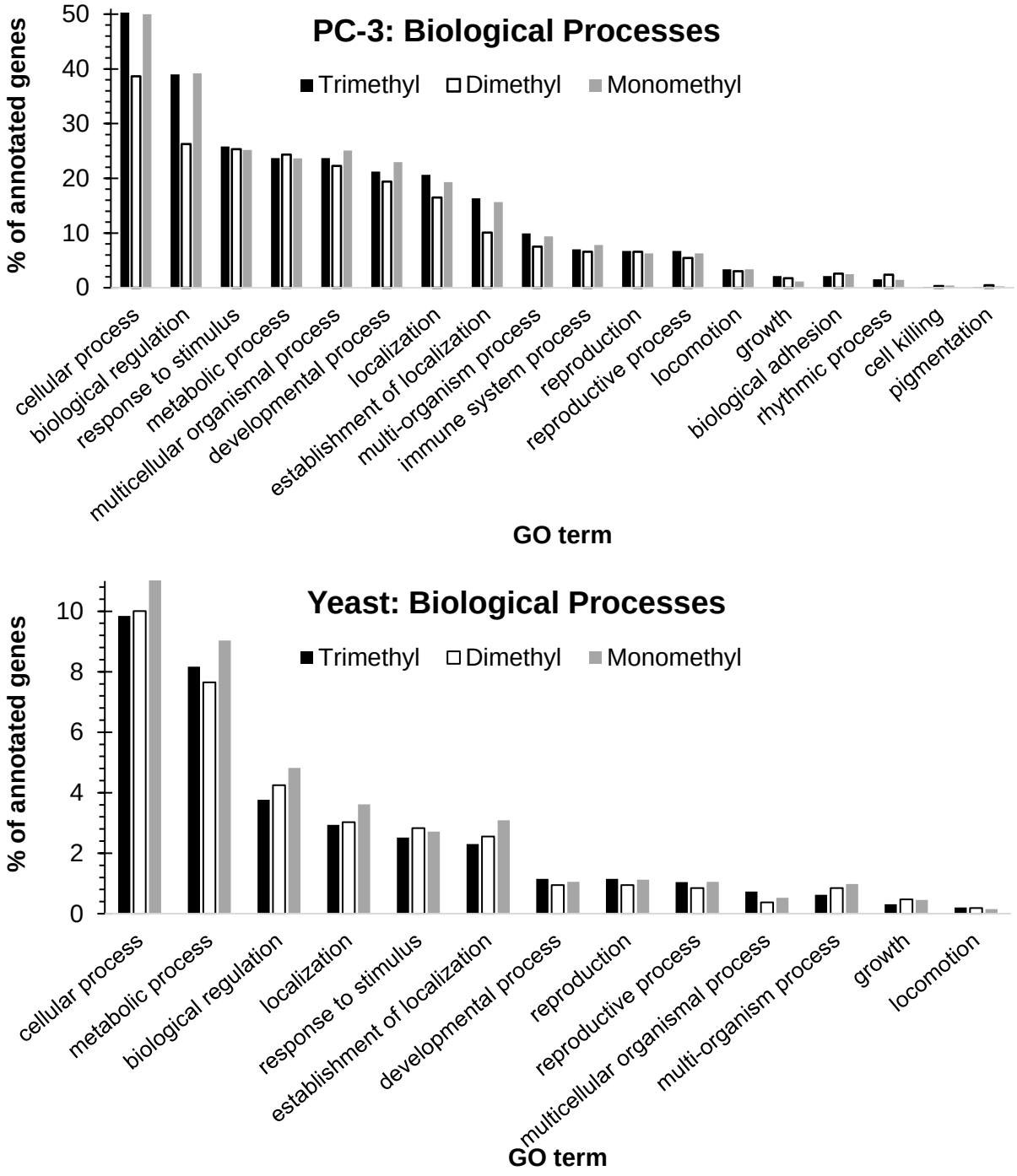


Figure 3.5 Gene ontology biological processes annotation of proteins with a mono, di, and trimethyllysine mark discovered within two biological and technical replicates for PC-3 (top) and yeast (bottom).

3.3.7 Limitations of GO analysis

Sampling bias and the open world assumption must be taken into consideration when analyzing GO results. Results cannot be analyzed with the notion of complete or random sampling. The GO annotations obtained are a subset of the total cellular methylated peptides based on sample processing and identification strategy. We must keep in mind this methods' sampling bias based on purification method (chloroform/methanol protein precipitation), protease choice (highly basic peptides with low E residue content), and enrichment strategy (MethylTrap enrichment) when deriving biological meaning from results. As for the GO annotations themselves, the open world assumption *must* be considered when deriving biological meaning from gene sets—unknown function does not equate absence of function and known function does not equate the true function.[169] The date when the GO annotations were produced must be noted because GO functional terms are dynamic. GO functional terms added and removed regularly based on continuing research, so the closed world assumption of the GO annotations reported in this thesis may not hold true in the future.[170]

3.4 Conclusions

The first objective, to enrich and identify methylated peptides in real-world industrial brewers' yeast samples, was met. This technology can be used as a tool to identify unique methylated peptides that are otherwise masked in the unfractionated sample. Given that there are three degrees of lysine methylation, this process is highly complex and knowledge about its exact mechanism of enrichment remains limited. MethylTrap fractionation allowed characterization of each degree of methylation from industrial samples that were unseen in the unfractionated input sample. With this technique, industrial yeast can be further investigated to increase the understanding of protein methylation in cells.

My next objective, to assess the reproducibility of the identifications discovered with MethylTrap enrichment, was also achieved. The observed reproducibility was poor, but in fact similar to the poor reproducibility observed for other methyl enrichment methods. The reproducibility between the BRs for all identified peptides (methylated and unmethylated alike) was 63% for the unretained sample, 52% for the input, and 25% for the retained. For the methylated peptides, 23% of the unretained were shared between biological replicates, 17% for the input, and 4.8% for the retained. As with all proteomics experiments, performing more TR would increase the reproducibility of the BRs.

My last objective, to survey the biological function and cellular compartment of the methylated peptides identified in the yeast and PC-3 cells using GO analysis, was also met. Methyllysine proteins were annotated with functionally and biologically diverse genes that localize throughout most cellular components within both industrial yeast and PC-3 mammalian cell. This brings biological meaning to the use of methylation in cellular processes.

3.5 Future work

The percentage of methylated peptides identified in the retained fraction was lower than that of other PTM enrichment technologies. Phosphopeptide enrichment studies have obtained selectivity of > 90%.[139] Phosophopeptide enrichment selectivity has been achieved by fine-tuning the binding buffer conditions and the enrichment reagent to obtain high selectivity. MethylTrap binding conditions and calixarene substitutions could be further analyzed to achieve higher selectivity for the methylated peptides over the unmethylated peptides. MethylTrap enrichment may be used as a chemical enrichment strategy analogous to the chemical phosphopeptide enrichment once all the conditions of the MethylTrap methylated peptide enrichment are optimized.

Additional fractionation steps may be added to this protocol to increase the number of unique identified methylated peptides. Further fractionating and separately analyzing the divided sample will increase the overall number of identified proteins with LC-MS/MS by parsing the dynamic concentration range of the proteome. There are a variety of methods to increase separation of proteins. For in depth analysis of subcellular methyl proteomes, organelle isolation (such as nuclear extraction) can remove signal convolution caused by proteins of other cellular localization.[171] Another option could be sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)—prior to enrichment, proteins could be separated on a gel based on their molecular weight.[172] Proteins of high, medium, and low molecular weights could be individually enriched and analyzed with LC-MS/MS. Moreover, this methyllysine enrichment may be performed on samples previously enriched by protein capture (such as the protocol by Gozani and coworkers).[83] Protein capture followed by calixarene enrichment would allow identification of specific methylated sites that are otherwise unseen. If increasing depth of coverage of the enriched sample is the main analytical objective, lengthening the HPLC gradients during LC-MS/MS analysis is a simple way to observe more peptide signals. The main drawback

of steps that increase fractionation is the length they add to the procedure. This method was optimized for in-depth methylated peptide analysis with a relatively short procedure. The trade-off for more methylated peptide identifications is a longer preparative procedure.

Bibliography

(1) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chemie - Int. Ed.* **2005**, *44* (45), 7342–7372. https://doi.org/10.1002/anie.200501023.

(2) Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-Translational Modification: Nature's Escape from Genetic Imprisonment and the Basis for Dynamic Information Encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2012**, *4* (6), 565–583. https://doi.org/10.1002/wsbm.1185.

(3) Chuh, K. N.; Batt, A. R.; Pratt, M. R. Chemical Methods for Encoding and Decoding of Posttranslational Modifications. *Cell Chem. Biol.* **2016**, *23* (1), 86–107. https://doi.org/10.1016/j.chembiol.2015.11.006.

(4) Chuikov, S.; Kurash, J. K.; Wilson, J. R.; Xiao, B.; Justin, N.; Ivanov, G. S.; McKinney, K.; Tempst, P.; Prives, C.; Gamblin, S. J.; et al. Regulation of P53 Activity through Lysine Methylation. *Nature* **2004**, *432* (7015), 353–360. https://doi.org/10.1038/nature03117.

(5) Greer, E. L.; Shi, Y. Histone Methylation: A Dynamic Mark in Health, Disease and Inheritance. *Nat. Rev. Genet.* **2012**, *13*, 343.

(6) Sabari, B. R.; Zhang, D.; Allis, C. D.; Zhao, Y. Metabolic Regulation of Gene Expression through Histone Acylations. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (2), 90–101. https://doi.org/10.1038/nrm.2016.140.

(7) Cloutier, P.; Coulombe, B. Regulation of Molecular Chaperones through Post-Translational Modifications: Decrypting the Chaperone Code. *Biochim. Biophys. Acta* **2013**, *1829* (5), 443–454. https://doi.org/10.1016/j.bbagrm.2013.02.010.

(8) Jablonka, E. The Evolutionary Implications of Epigenetic Inheritance. *Interface Focus* **2017**, *7* (5), 20160135. https://doi.org/10.1098/rsfs.2016.0135.

(9) Bird, A. DNA Methylation Patterns and Epigenetic Memory. *Genes Dev.* **2002**, *16* (1), 6–21. https://doi.org/10.1101/gad.947102.

(10) Yong-Villalobos, L.; Gonzalez-Morales, S. I.; Wrobel, K.; Gutierrez-Alanis, D.; Cervantes-Perez, S. A.; Hayano-Kanashiro, C.; Oropeza-Aburto, A.; Cruz-Ramirez, A.; Martinez, O.; Herrera-Estrella, L. Methylome Analysis Reveals an Important Role for Epigenetic Changes in the Regulation of the Arabidopsis Response to Phosphate Starvation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (52), E7293-302. https://doi.org/10.1073/pnas.1522301112.

(11) Perdigoto, C. N.; Valdes, V. J.; Bardot, E. S.; Ezhkova, E. Epigenetic Regulation of Skin: Focus on the Polycomb Complex. *Cell. Mol. Life Sci.* **2012**, *69* (13), 2161–2172. https://doi.org/10.1007/s00018-012-0920-x.

(12) Rui, L. Energy Metabolism in the Liver. *Compr. Physiol.* **2014**, *4* (1), 177–197. https://doi.org/10.1002/cphy.c130024.

(13) Monk, M.; Boubelik, M.; Lehnert, S. Temporal and Regional Changes in DNA Methylation in the Embryonic, Extraembryonic and Germ Cell Lineages during Mouse Embryo Development. *Development* **1987**, *99* (3), 371 LP – 382.

(14) Mayer, W.; Niveleau, A.; Walter, J.; Fundele, R.; Haaf, T. Demethylation of the Zygotic Paternal Genome.

Nature **2000**, *403* (6769), 501–502. https://doi.org/10.1038/35000656.

(15) Kraft, A.; Rubin, B. P. Changing Cells: An Analysis of the Concept of Plasticity in the Context of Cellular Differentiation. *Biosocieties* **2016**, *11* (4), 497–525. https://doi.org/10.1057/s41292-016-0027-y.

(16) Rothbart, S. B.; Krajewski, K.; Nady, N.; Tempel, W.; Xue, S.; Badeaux, A. I.; Barsyte-Lovejoy, D.; Martinez, J. Y.; Bedford, M. T.; Fuchs, S. M.; et al. Association of UHRF1 with Methylated H3K9 Directs the Maintenance of DNA Methylation. *Nat. Struct. Mol. Biol.* **2012**, *19* (11), 1155–1160. https://doi.org/10.1038/nsmb.2391.

(17) Margueron, R.; Reinberg, D. Chromatin Structure and the Inheritance of Epigenetic Information. *Nat. Rev. Genet.* **2010**, *11* (4), 285–296. https://doi.org/10.1038/nrg2752.

(18) Jenuwein, T.; Allis, C. D. Translating the Histone Code. *Science (80-.).* **2001**, *293* (August), 1074–1080.

(19) Strahl, B. D.; Allis, C. D. The Language of Covalent Histone Modifications. *Nature* **2000**, *403* (6765), 41–45. https://doi.org/10.1038/47412.

(20) Schübeler, D.; MacAlpine, D. M.; Scalzo, D.; Wirbelauer, C.; Kooperberg, C.; Van Leeuwen, F.; Gottschling, D. E.; O'Neill, L. P.; Turner, B. M.; Delrow, J.; et al. The Histone Modification Pattern of Active Genes Revealed through Genome-Wide Chromatin Analysis of a Higher Eukaryote. *Genes Dev.* **2004**, *18* (11), 1263–1271. https://doi.org/10.1101/gad.1198204.

(21) Perry, M.; Chalkley, R. Histone Acetylation Increases the Solubility of Chromatin and Occurs Sequentially over Most of the Chromatin. A Novel Model for the Biological Role of Histone Acetylation. *J. Biol. Chem.* **1982**, *257* (13), 7336–7347.

(22) Martin, C.; Zhang, Y. The Diverse Functions of Histone Lysine Methylation. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (11), 838–849. https://doi.org/10.1038/nrm1761.

(23) Bernstein, E.; Duncan, E. M.; Masui, O.; Gil, J.; Heard, E.; Allis, C. D. Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *Mol. Cell. Biol.* **2006**, *26* (7), 2560–2569. https://doi.org/10.1128/MCB.26.7.2560-2569.2006.

(24) Schübeler, D.; MacAlpine, D. M.; Scalzo, D.; Wirbelauer, C.; Kooperberg, C.; van Leeuwen, F.; Gottschling, D. E.; O'Neill, L. P.; Turner, B. M.; Delrow, J.; et al. The Histone Modification Pattern of Active Genes Revealed through Genome-Wide Chromatin Analysis of a Higher Eukaryote. *Genes Dev.* **2004**, *18* (11), 1263–1271. https://doi.org/10.1101/gad.1198204.

(25) Grewal, S. I. S.; Moazed, D. Heterochromatin and Epigenetic Control of Gene Expression. *Science (80-.).* **2003**, *301* (5634), 798–802. https://doi.org/10.1126/science.1086887.

(26) Schneider, R.; Bannister, A. J.; Myers, F. A.; Thorne, A. W.; Crane-Robinson, C.; Kouzarides, T. Histone H3 Lysine 4 Methylation Patterns in Higher Eukaryotic Genes. *Nat. Cell Biol.* **2004**, *6* (1), 73–77. https://doi.org/10.1038/ncb1076.

(27) Egorova, K. S.; Olenkina, O. M.; Olenina, L. V. Lysine Methylation of Nonhistone Proteins Is a Way to Regulate Their Stability and Function. *Biochem.* **2010**, *75* (5), 535–548. https://doi.org/10.1134/s0006297910050019.

(28) Stein, G. S.; Spelsberg, T. C.; Kleinsmith, L. J. Nonhistone Chromosomal Proteins and Gene Regulation. *Science* **1974**, *183* (4127), 817–824. https://doi.org/10.1126/science.183.4127.817.

(29) Das, C.; Kundu, T. K. Transcriptional Regulation by the Acetylation of Nonhistone Proteins in Humans - A New Target for Therapeutics. *IUBMB Life* **2005**, *57* (3), 137–148. https://doi.org/10.1080/15216540500090629.

(30) Filtz, T. M.; Vogel, W. K.; Leid, M. Regulation of Transcription Factor Activity by Interconnected Post-Translational Modifications. *Trends Pharmacol. Sci.* **2014**, *35* (2), 76–85. https://doi.org/10.1016/J.TIPS.2013.11.005.

(31) Geiss-Friedlander, R.; Melchior, F. Concepts in Sumoylation: A Decade On. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 947.

(32) Biggar, K. K.; Li, S. S. C. Non-Histone Protein Methylation as a Regulator of Cellular Signalling and Function. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (1), 5–17. https://doi.org/10.1038/nrm3915.

(33) Benveniste, D.; Sonntag, H.-J.; Sanguinetti, G.; Sproul, D. Transcription Factor Binding Predicts Histone Modifications in Human Cell Lines. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (37), 13367–13372. https://doi.org/10.1073/pnas.1412081111.

(34) Henikoff, S.; Shilatifard, A. Histone Modification: Cause or Cog? *Trends Genet.* **2011**, *27* (10), 389–396. https://doi.org/10.1016/J.TIG.2011.06.006.

(35) Ptashne, M. On the Use of the Word "Epigenetic." *Current Biology*. 2007. https://doi.org/10.1016/j.cub.2007.02.030.

(36) Xin, B.; Rohs, R. Relationship between Histone Modifications and Transcription Factor Binding Is Protein Family Specific. *Genome Res.* **2018**. https://doi.org/10.1101/gr.220079.116.

(37) Katan-Khaykovich, Y.; Struhl, K. Dynamics of Global Histone Acetylation and Deacetylation in Vivo: Rapid Restoration of Normal Histone Acetylation Status upon Removal of Activators and Repressors. *Genes Dev.* **2002**, *16* (6), 743–752. https://doi.org/10.1101/gad.967302.

(38) Zhang, X.; Wen, H.; Shi, X. Lysine Methylation: Beyond Histones. *Acta Biochim. Biophys. Sin. (Shanghai).* **2012**, *44* (1), 14–27. https://doi.org/10.1093/abbs/gmr100.

(39) Huang, J.; Perez-Burgos, L.; Placek, B. J.; Sengupta, R.; Richter, M.; Dorsey, J. A.; Kubicek, S.; Opravil, S.; Jenuwein, T.; Berger, S. L. Repression of P53 Activity by Smyd2-Mediated Methylation. *Nature* **2006**, *444* (7119), 629–632. https://doi.org/10.1038/nature05287.

(40) Zhu, J.; Dou, Z.; Sammons, M. A.; Levine, A. J.; Berger, S. L. Lysine Methylation Represses P53 Activity in Teratocarcinoma Cancer Cells. *Proc. Natl. Acad. Sci.* **2016**, *113* (35), 9822–9827. https://doi.org/10.1073/pnas.1610387113.

(41) Mentch, S. J.; Mehrmohamadi, M.; Huang, L.; Liu, X.; Gupta, D.; Mattocks, D.; Gómez Padilla, P.; Ables, G.; Bamman, M. M.; Thalacker-Mercer, A. E.; et al. Histone Methylation Dynamics and Gene Regulation Occur through the Sensing of One-Carbon Metabolism. *Cell Metab.* **2015**. https://doi.org/10.1016/j.cmet.2015.08.024.

(42) Black, J. C.; Van Rechem, C.; Whetstine, J. R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* **2012**, *48* (4), 491–507. https://doi.org/10.1016/j.molcel.2012.11.006.

(43) Kim, J.; Daniel, J.; Espejo, A.; Lake, A.; Krishna, M.; Xia, L.; Zhang, Y.; Bedford, M. T. Tudor, MBT and

Chromo Domains Gauge the Degree of Lysine Methylation. *EMBO Rep.* **2006**, *7* (4), 397–403. https://doi.org/10.1038/sj.embor.7400625.

(44) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. How Chromatin-Binding Modules Interpret Histone Modifications: Lessons from Professional Pocket Pickers. *Nat. Struct. Mol. Biol.* **2007**, *14* (11), 1025–1040. https://doi.org/10.1038/nsmb1338.

(45) Ma, J. C.; Dougherty, D. A. The Cation-π Interaction. *Chem. Rev.* **1997**, *97* (5), 1303–1324. https://doi.org/10.1021/cr9603744.

(46) Kamps, J. J. A. G.; Huang, J.; Poater, J.; Xu, C.; Pieters, B. J. G. E.; Dong, A.; Min, J.; Sherman, W.; Beuming, T.; Matthias Bickelhaupt, F.; et al. Chemical Basis for the Recognition of Trimethyllysine by Epigenetic Reader Proteins. *Nat. Commun.* **2015**, *6*, 8911. https://doi.org/10.1038/ncomms9911.

(47) Botuyan, M. V.; Lee, J.; Ward, I. M.; Kim, J.-E.; Thompson, J. R.; Chen, J.; Mer, G. Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair. *Cell* **2006**, *127* (7), 1361–1373. https://doi.org/10.1016/J.CELL.2006.10.043.

(48) Ruthenburg, A. J.; Li, H.; Milne, T. A.; Dewell, S.; McGinty, R. K.; Yuen, M.; Ueberheide, B.; Dou, Y.; Muir, T. W.; Patel, D. J.; et al. Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell* **2011**, *145* (5), 692–706. https://doi.org/10.1016/j.cell.2011.03.053.

(49) Petrossian, T. C.; Clarke, S. G. Uncovering the Human Methyltransferasome. *Mol. Cell. Proteomics* **2011**, *10* (1), M110.000976. https://doi.org/10.1074/mcp.M110.000976.

(50) Subramani, S.; Jayapalan, S.; Kalpana, R.; Natarajan, J. HomoKinase: A Curated Database of Human Protein Kinases. *ISRN Comput. Biol.* **2013**, *2013* (i), 1–5. https://doi.org/10.1155/2013/417634.

(51) Huang, H.; Arighi, C. N.; Ross, K. E.; Ren, J.; Li, G.; Chen, S.-C.; Wang, Q.; Cowart, J.; Vijay-Shanker, K.; Wu, C. H. IPTMnet: An Integrated Resource for Protein Post-Translational Modification Network Discovery. *Nucleic Acids Res.* **2017**, *46* (D1), D542–D550. https://doi.org/10.1093/nar/gkx1104.

(52) Dunn, J.; Reid, G.; Brunening, M. Techniques for Phosphopeptide Enrichment Prior to Analysis by Mass Spectrometry. *Mass Spectrom. Rev.* **2007**, *20* (3), 133–138. https://doi.org/10.1002/mas.

(53) Pawson, T.; Scott, J. D. Protein Phosphorylation in Signaling 50 Years and Counting. *Trends Biochem. Sci.* **2005**, *30* (6), 286–290. https://doi.org/10.1016/j.tibs.2005.04.013.

(54) Knudsen, G. M.; Chalkley, R. J. The Effect of Using an Inappropriate Protein Database for Proteomic Data Analysis. *PLoS One* **2011**, *6* (6), 1–4. https://doi.org/10.1371/journal.pone.0020873.

(55) Käll, L.; Storey, J. D.; Maccoss, M. J.; Noble, W. S. Assigning Significance to Peptides Identified by Tandem Mass.Pdf. **2008**, 29–34.

(56) Michalski, A.; Cox, J.; Mann, M. More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority Is Inaccessible to Data-Dependent LC-MS/MS. *J. Proteome Res.* **2011**, *10* (4), 1785–1793. https://doi.org/10.1021/pr101060v.

(57) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R. Direct Analysis and Identification of Proteins in Mixtures by LC/MS/MS and Database Searching at the Low-Femtomole Level. *Anal. Chem.* **1997**, *69* (4), 767–776. https://doi.org/10.1021/ac960799q.

(58) Janeway, C.; Travers, P.; Walport, M.; Shlomchik, M. Lmmunobiology: The Immune System in Health and

Disease. 5th Edition. In *Immunobiology, 5th edition*; New York, 2001.

(59) Roghanian, A. B Cells | British Society for Immunology. **2014**.

(60) Tonegawa, S. Somatic Generation of Antibody Diversity. *Nature* **1983**, *302* (5909), 575–581. https://doi.org/10.1038/302575a0.

(61) Wu, H.; Deng, Y.; Feng, Y.; Long, D.; Ma, K.; Wang, X.; Zhao, M.; Lu, L.; Lu, Q. Epigenetic Regulation in B-Cell Maturation and Its Dysregulation in Autoimmunity. *Cell. Mol. Immunol.* **2018**, *15* (7), 676–684. https://doi.org/10.1038/cmi.2017.133.

(62) Greenfield, E. A. *Antibodies : A Laboratory Manual*; 2014.

(63) Lipman, N. S.; Jackson, L. R.; Trudel, L. J.; Weis-Garcia, F. Monoclonal versus Polyclonal Antibodies: Distinguishing Characteristics, Applications, and Information Resources. *ILAR J.* **2005**, *46* (3), 258–268. https://doi.org/10.1093/ilar.46.3.258.

(64) KÖHLER, G.; MILSTEIN, C. Continuous Cultures of Fused Cells Secreting Antibody of Predefined Specificity. *Nature* **1975**, *256* (5517), 495–497. https://doi.org/10.1038/256495a0.

(65) Voskuil, J. L. A. The Challenges with the Validation of Research Antibodies. *F1000Research* **2017**, *6*, 161. https://doi.org/10.12688/f1000research.10851.1.

(66) Schonbrunn, A. Editorial: Antibody Can Get It Right: Confronting Problems of Antibody Specificity and Irreproducibility. *Molecular endocrinology (Baltimore, Md.)*. September 2014, pp 1403–1407. https://doi.org/10.1210/me.2014-1230.

(67) Frenzel, A.; Hust, M.; Schirrmann, T. Expression of Recombinant Antibodies. *Front. Immunol.* **2013**, *4*, 217. https://doi.org/10.3389/fimmu.2013.00217.

(68) Jefferis, R. Glycosylation of Recombinant Antibody Therapeutics. *Biotechnol. Prog.* **2005**, *21* (1), 11–16. https://doi.org/10.1021/bp040016j.

(69) Wurm, F. M. Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells. *Nat. Biotechnol.* **2004**, *22* (11), 1393–1398. https://doi.org/10.1038/nbt1026.

(70) Browne, S. M.; Al-rubeai, M. Selection Methods for High-Producing Mammalian Cell Lines. **2014**, No. July. https://doi.org/10.1007/978-90-481-2245-5.

(71) Guan, K.-L.; Yu, W.; Lin, Y.; Xiong, Y.; Zhao, S. Generation of Acetyllysine Antibodies and Affinity Enrichment of Acetylated Peptides. *Nat. Protoc.* **2010**, *5* (9), 1583–1595. https://doi.org/10.1038/nprot.2010.117.

(72) Qiang, L.; Xiao, H.; Campos, E. I.; Ho, V. C.; Li, G. Development of a PAN-Specific, Affinity-purified Anti-acetylated Lysine Antibody for Detection, Identification, Isolation, and Intracellular Localization of Acetylated Protein. *J. Immunoass. Immunochem.* **2005**, *26* (1), 13–23. https://doi.org/10.1081/IAS-200041153.

(73) Egelhofer, T. A.; Minoda, A.; Klugman, S.; Lee, K.; Kolasinska-Zwierz, P.; Alekseyenko, A. A.; Cheung, M.-S.; Day, D. S.; Gadel, S.; Gorchakov, A. A.; et al. An Assessment of Histone-Modification Antibody Quality. *Nat. Struct. &Amp; Mol. Biol.* **2010**, *18*, 91.

(74) Bock, I.; Dhayalan, A.; Kudithipudi, S.; Brandt, O.; Rathert, P.; Jeltsch, A. Detailed Specificity Analysis of Antibodies Binding to Modified Histone Tails with Peptide Arrays. *Epigenetics* **2011**, *6* (2), 256–263.

https://doi.org/10.4161/epi.6.2.13837.

(75) Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S. Antibody Structure, Instability, and Formulation. *J. Pharm. Sci.* **2007**, *96* (1), 1–26. https://doi.org/10.1002/jps.20727.

(76) Zhao, Y.; Jensen, O. N. Modification-Specific Proteomics: Strategies for Characterization of Post-Translational Modifications Using Enrichment Techniques. *Proteomics* **2009**, *9* (20), 4632–4641. https://doi.org/10.1002/pmic.200900398.

(77) Cao, X.-J.; Garcia, B. A. Global Proteomics Analysis of Protein Lysine Methylation. *Curr. Protoc. protein Sci.* **2016**, *86*, 24.8.1-24.8.19. https://doi.org/10.1002/cpps.16.

(78) Liang, Z.; Wong, R. P.; Li, L. H.; Jiang, H.; Xiao, H.; Li, G. Development of Pan-Specific Antibody against Trimethyllysine for Protein Research. *Proteome Sci.* **2008**, *6*, 2. https://doi.org/10.1186/1477-5956-6-2.

(79) Fuchs, S. M.; Strahl, B. D. Antibody Recognition of Histone Post-Translational Modifications: Emerging Issues and Future Prospects. *Epigenomics*. England June 2011, pp 247–249. https://doi.org/10.2217/epi.11.23.

(80) Perez-Burgos, L.; Peters, A. H. F. .; Opravil, S.; Kauer, M.; Mechtler, K.; Jenuwein, T. Generation and Characterization of Methyl-Lysine Histone Antibodies. *Methods Enzymol.* **2003**, *376*, 234–254. https://doi.org/10.1016/S0076-6879(03)76016-9.

(81) Bremang, M.; Cuomo, A.; Agresta, A. M.; Stugiewicz, M.; Spadotto, V.; Bonaldi, T. Mass Spectrometry-Based Identification and Characterisation of Lysine and Arginine Methylation in the Human Proteome. *Mol. Biosyst.* **2013**, *9* (9), 2231–2247. https://doi.org/10.1039/c3mb00009e.

(82) Rothbart, S. B.; Dickson, B. M.; Raab, J. R.; Grzybowski, A. T.; Krajewski, K.; Guo, A. H.; Shanle, E. K.; Josefowicz, S. Z.; Fuchs, S. M.; Allis, C. D.; et al. An Interactive Database for the Assessment of Histone Antibody Specificity. *Mol. Cell* **2015**, *59* (3), 502–511. https://doi.org/10.1016/j.molcel.2015.06.022.

(83) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martín, G. M.; Gozani, O. Proteome-Wide Enrichment of Proteins Modified by Lysine Methylation. *Nat. Protoc.* **2013**, *9* (1), 37–50. https://doi.org/10.1038/nprot.2013.164.

(84) Ning, Z.; Star, A. T.; Mierzwa, A.; Lanouette, S.; Mayne, J.; Couture, J. F.; Figeys, D. A Charge-Suppressing Strategy for Probing Protein Methylation. *Chem. Commun.* **2016**, *52* (31), 5474–5477. https://doi.org/10.1039/c6cc00814c.

(85) Huang, J.; Wang, F.; Ye, M.; Zou, H. Enrichment and Separation Techniques for Large-Scale Proteomics Analysis of the Protein Post-Translational Modifications. *J. Chromatogr. A* **2014**, *1372*, 1–17. https://doi.org/10.1016/j.chroma.2014.10.107.

(86) Lewallen, D. M.; Bicker, K. L.; Subramanian, V.; Clancy, K. W.; Slade, D. J.; Martell, J.; Dreyton, C. J.; Sokolove, J.; Weerapana, E.; Thompson, P. R. Chemical Proteomic Platform To Identify Citrullinated Proteins. *ACS Chem. Biol.* **2015**, *10* (11), 2520–2528. https://doi.org/10.1021/acschembio.5b00438.

(87) Thingholm, T. E.; Larsen, M. R. The Use of Titanium Dioxide for Selective Enrichment of Phosphorylated Peptides. In *Phospho-Proteomics: Methods and Protocols*; von Stechow, L., Ed.; Springer New York: New York, NY, 2016; pp 135–146. https://doi.org/10.1007/978-1-4939-3049-4_9.

(88) Wu, X. N.; Xi, L.; Pertl-Obermeyer, H.; Li, Z.; Chu, L.-C.; Schulze, W. X. Highly Efficient Single-Step Enrichment of Low Abundance Phosphopeptides from Plant Membrane Preparations. *Front. Plant Sci.* **2017**, *8*, 1673. https://doi.org/10.3389/fpls.2017.01673.

(89) Aldrich, S. Translational Research Solutions Improving Reproducibility : Best Practices for Antibodies, Cell Lines, and Small Molecules. *Sigma Aldrich* **2013**.

(90) Halperin, R. F.; Stafford, P.; Johnston, S. A. Exploring Antibody Recognition of Sequence Space through Random-Sequence Peptide Microarrays. *Mol. Cell. Proteomics* **2011**, *10* (3), M110.000786. https://doi.org/10.1074/mcp.M110.000786.

(91) Matheron, L.; van den Toorn, H.; Heck, A. J. R.; Mohammed, S. Characterization of Biases in Phosphopeptide Enrichment by Ti4+-Immobilized Metal Affinity Chromatography and TiO2 Using a Massive Synthetic Library and Human Cell Digests. *Anal. Chem.* **2014**, *86* (16), 8312–8320. https://doi.org/10.1021/ac501803z.

(92) Tape, C. J.; Worboys, J. D.; Sinclair, J.; Gourlay, R.; Vogt, J.; McMahon, K. M.; Trost, M.; Lauffenburger, D. A.; Lamont, D. J.; Jørgensen, C. Reproducible Automated Phosphopeptide Enrichment Using Magnetic TiO2 and Ti-IMAC. *Anal. Chem.* **2014**, *86* (20), 10296–10302. https://doi.org/10.1021/ac5025842.

(93) Foettinger, A.; Leitner, A.; Lindner, W. Derivatisation of Arginine Residues with Malondialdehyde for the Analysis of Peptides and Protein Digests by LC-ESI-MS/MS. *J. Mass Spectrom.* **2006**, *41* (5), 623–632. https://doi.org/10.1002/jms.1020.

(94) Hof, F. Host–Guest Chemistry That Directly Targets Lysine Methylation: Synthetic Host Molecules as Alternatives to Bio-Reagents. *Chem. Commun.* **2016**, *52* (66), 10093–10108. https://doi.org/10.1039/C6CC04771H.

(95) Daze, K. D.; Pinter, T.; Beshara, C. S.; Ibraheem, A.; Minaker, S. A.; Ma, M. C. F.; Courtemanche, R. J. M.; Campbell, R. E.; Hof, F. Supramolecular Hosts That Recognize Methyllysines and Disrupt the Interaction between a Modified Histone Tail and Its Epigenetic Reader Protein. *Chem. Sci.* **2012**, *3* (9), 2695–2699. https://doi.org/10.1039/c2sc20583a.

(96) Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F. A Simple Calixarene Recognizes Post-Translationally Methylated Lysine. *ChemBioChem* **2010**, *11* (1), 63–66. https://doi.org/10.1002/cbic.200900633.

(97) McGovern, R. E.; Fernandes, H.; Khan, A. R.; Power, N. P.; Crowley, P. B. Protein Camouflage in Cytochrome c–Calixarene Complexes. *Nat. Chem.* **2012**, *4*, 527.

(98) Garnett, G. A. E.; Starke, M. J.; Shaurya, A.; Li, J.; Hof, F. Supramolecular Affinity Chromatography for Methylation-Targeted Proteomics. *Anal. Chem.* **2016**, *88* (7), 3697–3703. https://doi.org/10.1021/acs.analchem.5b04508.

(99) Scaife, C.; Cagney, G.; Dillon, E. Sample preparation for LC-MS / MS - Guidelines contaminants incompatible with mass spectrometry volatile buffers.

(100) Rundlett, K. L.; Armstrong, D. W. Mechanism of Signal Suppression by Anionic Surfactants in Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1996**, *68* (19), 3493–3497. https://doi.org/10.1021/ac960472p.

(101) Furey, A.; Moriarty, M.; Bane, V.; Kinsella, B.; Lehane, M. Ion Suppression; A Critical Review on Causes, Evaluation, Prevention and Applications. *Talanta*. 2013. https://doi.org/10.1016/j.talanta.2013.03.048.

(102) Haglind, A.; Hedeland, M.; Arvidsson, T.; Pettersson, C. The Differences in Matrix Effect between Supercritical Fluid Chromatography and Reversed Phase Liquid Chromatography Coupled to ESI/MS. *Anal.*

Chim. Acta **2017**, *1000*. https://doi.org/10.1016/j.aca.2017.10.014.

(103) King, R.; Bonfiglio, R.; Fernandez-Metzler, C.; Miller-Stein, C.; Olah, T. Mechanistic Investigation of Ionization Suppression in Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (11), 942–950. https://doi.org/https://doi.org/10.1016/S1044-0305(00)00163-X.

(104) Antignac, J. P.; De Wasch, K.; Monteau, F.; De Brabander, H.; Andre, F.; Le Bizec, B. The Ion Suppression Phenomenon in Liquid Chromatography-Mass Spectrometry and Its Consequences in the Field of Residue Analysis. *Anal. Chim. Acta* **2005**, *529* (1-2 SPEC. ISS.), 129–136. https://doi.org/10.1016/j.aca.2004.08.055.

(105) Hodge, K.; Have, S. Ten; Hutton, L.; Lamond, A. I. Cleaning up the Masses: Exclusion Lists to Reduce Contamination with HPLC-MS/MS. *J. Proteomics* **2013**, *88*, 92–103. https://doi.org/10.1016/j.jprot.2013.02.023.

(106) Filip, C.; Fletcher, G.; Wulff, J. L.; Earhart, C. F. Solubilization of the Cytoplasmic Membrane of Escherichia Coli by the Ionic Detergent Sodium-Lauryl Sarcosinate. *J. Bacteriol.* **1973**, *115* (3), 717 LP – 722.

(107) Zhang, N.; Li, L. Effects of Common Surfactants on Protein Digestion and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of the Digested Peptides Using Two-Layer Sample Preparation. *Rapid Commun. Mass Spectrom.* **2004**, *18* (8), 889–896. https://doi.org/10.1002/rcm.1423.

(108) Shehadul Islam, M.; Aryasomayajula, A.; Selvaganapathy, P. R. A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines*. March 2017. https://doi.org/10.3390/mi8030083.

(109) Bulaj, G. Formation of Disulfide Bonds in Proteins and Peptides. *Biotechnol. Adv.* **2005**, *23* (1), 87–92. https://doi.org/10.1016/j.biotechadv.2004.09.002.

(110) Kulevich, S. E.; Frey, B. L.; Kreitinger, G.; Smith, L. M. Alkylating Tryptic Peptides to Enhance Electrospray Ionization Mass Spectrometry Analysis. *Anal. Chem.* **2010**, *82* (24), 10135–10142. https://doi.org/10.1021/ac1019792.

(111) Suttapitugsakul, S.; Xiao, H.; Smeekens, J.; Wu, R. Evaluation and Optimization of Reduction and Alkylation Methods to Maximize Peptide Identification with MS-Based Proteomics. *Mol. Biosyst.* **2017**, *13* (12), 2574–2582. https://doi.org/10.1039/c7mb00393e.

(112) Jiang, L.; He, L.; Fountoulakis, M. Comparison of Protein Precipitation Methods for Sample Preparation Prior to Proteomic Analysis. *J. Chromatogr. A* **2004**. https://doi.org/10.1016/j.chroma.2003.10.029.

(113) Fic, E.; Kedracka-Krok, S.; Jankowska, U.; Pirog, A.; Dziedzicka-Wasylewska, M. Comparison of Protein Precipitation Methods for Various Rat Brain Structures Prior to Proteomic Analysis. *Electrophoresis* **2010**, *31* (21), 3573–3579. https://doi.org/10.1002/elps.201000197.

(114) Rabilloud, T. Detergents and Chaotropes for Protein Solubilization before Two-Dimensional Electrophoresis. *Br. J. Psychiatry* **1965**, *111* (479), 1009–1010. https://doi.org/10.1192/bjp.111.479.1009-a.

(115) Nandakumar, M. P.; Shen, J.; Raman, B.; Marten, M. R. Solubilization of Trichloroacetic Acid (TCA) Precipitated Microbial Proteins via NaOH for Two-Dimensional Electrophoresis. *J. Proteome Res.* **2003**, *2* (1), 89–93. https://doi.org/10.1021/pr025541x.

(116) Hsiao, J. J.; Potter, O. G.; Chu, T. W.; Yin, H. 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. https://doi.org/10.1021/acs.analchem.8b02100.

(117) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six Alternative Proteases for Mass Spectrometry–Based Proteomics beyond Trypsin. *Nat. Protoc.* **2016**, *11*, 993.

(118) Chiva, C.; Ortega, M.; Sabidó, E. Influence of the Digestion Technique, Protease, and Missed Cleavage Peptides in Protein Quantitation. *J. Proteome Res.* **2014**, *13* (9), 3979–3986. https://doi.org/10.1021/pr500294d.

(119) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

(120) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **2002**, *74* (20), 5383–5392. https://doi.org/10.1021/ac025747h.

(121) Searle, B. C. Scaffold: A Bioinformatic Tool for Validating MS/MS-Based Proteomic Studies. *Proteomics* **2010**, *10* (6), 1265–1269. https://doi.org/10.1002/pmic.200900437.

(122) Hart-smith, G.; Yagoub, D.; Tay, A. P.; Pickford, R.; Wilkins, M. R. Large Scale Mass Spectrometry-Based Identifications of Enzyme-Mediated Protein Methylation Are Subject to High False Discovery Rates. *Mol. Cell. Proteomics* **2016**, *15* (3), 989–1006. https://doi.org/10.1074/mcp.M115.055384.

(123) Piehowski, P. D.; Petyuk, V. A.; Orton, D. J.; Xie, F.; Moore, R. J.; Ramirez-Restrepo, M.; Engel, A.; Lieberman, A. P.; Albin, R. L.; Camp, D. G.; et al. Sources of Technical Variability in Quantitative LC-MS Proteomics: Human Brain Tissue Sample Analysis. *J. Proteome Res.* **2013**, *12* (5), 2128–2137. https://doi.org/10.1021/pr301146m.

(124) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; et al. Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography-Tandem Mass Spectrometry. *J. Proteome Res.* **2010**, *9* (2), 761–776. https://doi.org/10.1021/pr9006365.

(125) Thermo. *Mass Spectrometry-Grade Endoproteinases-Thermo*; Vol. 0747.

(126) Han, D.; Jin, J.; Woo, J.; Min, H.; Kim, Y. Proteomic Analysis of Mouse Astrocytes and Their Secretome by a Combination of FASP and StageTip-Based, High PH, Reversed-Phase Fractionation. *Proteomics* **2014**, *14* (13–14), 1604–1609. https://doi.org/10.1002/pmic.201300495.

(127) Berg, M.; Parbel, A.; Pettersen, H.; Fenyö, D.; Björkesten, L. Reproducibility of LC-MS-Based Protein Identification. *J. Exp. Bot.* **2006**, *57* (7), 1509–1514. https://doi.org/10.1093/jxb/erj139.

(128) Kellner, R. Chemical and Enzymatic Fragmentation of Proteins. *Microcharacterization of proteins* **1994**, 11–27.

(129) Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. F. M.; Mechtler, K.; Heck, A. J. R.; Mohammed, S. Unambiguous Phosphosite Localization Using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *J. Proteome Res.* **2013**, *12* (3), 1520–1525. https://doi.org/10.1021/pr301130k.

(130) Cristobal, A.; Marino, F.; Post, H.; W. P. van den Toorn, H.; Mohammed, S.; J. R. Heck, A. Toward an Optimized Workflow for Middle-down Proteomics. *Anal. Chem.* **2017**, *89* (6), 3318–3325. https://doi.org/10.1021/acs.analchem.6b03756.

(131) Zhang, J.; Xing, G.; Chen, Y.; Zhang, Z.; Zhao, Y.; Wysocka, J. MS/MS/MS Reveals False Positive Identification of Histone Serine Methylation. *J. Proteome Res.* **2010**, *9* (1), 585–594. https://doi.org/10.1021/pr900864s.

(132) Olsen, J. V.; de Godoy, L. M. F.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-Trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010–2021. https://doi.org/10.1074/mcp.T500030-MCP200.

(133) Hsieh, E. J.; Hoopmann, M. R.; MacLean, B.; MacCoss, M. J. Comparison of Database Search Strategies for High Precursor Mass Accuracy MS/MS Data. *J. Proteome Res.* **2010**, *9* (2), 1138–1143. https://doi.org/10.1021/pr900816a.

(134) Nie, L.; Zhu, M.; Sun, S.; Zhai, L.; Wu, Z.; Qian, L.; Tan, M. An Optimization of the LC-MS/MS Workflow for Deep Proteome Profiling on an Orbitrap Fusion. *Anal. Methods* **2016**, *8* (2), 425–434. https://doi.org/10.1039/c5ay01900a.

(135) Blainey, P.; Krzywinski, M.; Altman, N. Points of Significance: Replication. *Nat. Methods* **2014**, *11* (9), 879–880. https://doi.org/10.1038/nmeth.3091.

(136) Gottfries, C. G. Neurochemical Aspects on Aging and Diseases with Cognitive Impairment. *J. Neurosci. Res.* **1990**, *27* (4), 541–547. https://doi.org/10.1002/jnr.490270415.

(137) Benabdelkamel, H.; Masood, A.; Alanazi, I. O.; Alfadda, A. A. Comparison of Protein Precipitation Methods from Adipose Tissue Using Difference Gel Electrophoresis. *Electrophoresis* **2018**. https://doi.org/10.1002/elps.201800124.

(138) Karpievitch, Y. V; Dabney, A. R.; Smith, R. D. Normalization and Missing Value Imputation for Label-Free LC-MS Analysis. *BMC Bioinformatics* **2012**, *13* (16), S5. https://doi.org/10.1186/1471-2105-13-S16-S5.

(139) Ruprecht, B.; Koch, H.; Medard, G.; Mundt, M.; Kuster, B.; Lemeer, S. Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns. *Mol. Cell. Proteomics* **2015**, *14* (1), 205–215. https://doi.org/10.1074/mcp.M114.043109.

(140) Oberg, A. L.; Vitek, O. Statistical Design of Quantitative Mass Spectrometry-Based Proteomic Experiments. *J. Proteome Res.* **2009**, *8* (5), 2144–2156. https://doi.org/10.1021/pr8010099.

(141) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25* (1), 25–29. https://doi.org/10.1038/75556.

(142) Moreau, Y.; Tranchevent, L. C. Computational Tools for Prioritizing Candidate Genes: Boosting Disease Gene Discovery. *Nat. Rev. Genet.* **2012**, *13* (8), 523–536. https://doi.org/10.1038/nrg3253.

(143) Pomaznoy, M.; Ha, B.; Peters, B. GOnet: A Tool for Interactive Gene Ontology Analysis. *BMC Bioinformatics* **2018**, *19* (1), 1–8. https://doi.org/10.1186/s12859-018-2533-3.

(144) The Gene Ontology Resource: 20 Years and Still GOing Strong. *Nucleic Acids Res.* **2019**, *47* (D1), D330–D338. https://doi.org/10.1093/nar/gky1055.

(145) Wu, Z.; Cheng, Z.; Sun, M.; Wan, X.; Liu, P.; He, T.; Tan, M.; Zhao, Y. A Chemical Proteomics Approach for Global Analysis of Lysine Monomethylome Profiling. *Mol. Cell. Proteomics* **2015**, *14* (2), 329–339.

https://doi.org/10.1074/mcp.M114.044255.

(146) Moore, K. E.; Carlson, S. M.; Camp, N. D.; Cheung, P.; James, R. G.; Chua, K. F.; Wolf-Yadlin, A.; Gozani, O. A General Molecular Affinity Strategy for Global Detection and Proteomic Analysis of Lysine Methylation. *Mol. Cell* **2013**, *50* (3), 444–456. https://doi.org/10.1016/j.molcel.2013.03.005.

(147) Li, N.; Stein, R. S. L.; He, W.; Komives, E.; Wang, W. Identification of Methyllysine Peptides Binding to Chromobox Protein Homolog 6 Chromodomain in the Human Proteome. *Mol. Cell. Proteomics* **2013**, *12* (10), 2750–2760. https://doi.org/10.1074/mcp.O112.025015.

(148) Cao, X. J.; Arnaudo, A. M.; Garcia, B. A. Large-Scale Global Identification of Protein Lysine Methylation in Vivo. *Epigenetics* **2013**, *8* (5), 477–485. https://doi.org/10.4161/epi.24547.

(149) Johnson, H. Chapter 10: Quantitative Analyses of Phosphotyrosine Cellular Signaling in Disease. In *Quantitative Proteomics*; The Royal Society of Chemistry, 2014; pp 211–232. https://doi.org/10.1039/9781782626985-00211.

(150) Guo, A.; Gu, H.; Zhou, J.; Mulhern, D.; Wang, Y.; Lee, K. A.; Yang, V.; Aguiar, M.; Kornhauser, J.; Jia, X.; et al. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. *Mol. Cell. Proteomics* **2014**, *13* (1), 372–387. https://doi.org/10.1074/mcp.o113.027870.

(151) Liu, H.; Sadygov, R. G.; Yates, J. R. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–4201. https://doi.org/10.1021/ac0498563.

(152) Johnson, R. S.; Davis, M. T.; Taylor, J. A.; Patterson, S. D. Informatics for Protein Identification by Mass Spectrometry. *Methods* **2005**, *35* (3 SPEC.ISS.), 223–236. https://doi.org/10.1016/j.ymeth.2004.08.014.

(153) Rhein, V. F.; Carroll, J.; He, J.; Ding, S.; Fearnley, I. M.; Walker, J. E. Human METTL20 Methylates Lysine Residues Adjacent to the Recognition Loop of the Electron Transfer Flavoprotein in Mitochondria. *J. Biol. Chem.* **2014**, *289* (35), 24640–24651. https://doi.org/10.1074/jbc.M114.580464.

(154) Shimazu, T.; Furuse, T.; Balan, S.; Yamada, I.; Okuno, S.; Iwanari, H.; Suzuki, T.; Hamakubo, T.; Dohmae, N.; Yoshikawa, T.; et al. Role of METTL20 in Regulating β-Oxidation and Heat Production in Mice under Fasting or Ketogenic Conditions. *Sci. Rep.* **2018**, *8* (1), 1179. https://doi.org/10.1038/s41598-018-19615-4.

(155) Curcoy, A.; Olsen, R. K. J.; Ribes, A.; Trenchs, V.; Vilaseca, M. A.; Campistol, J.; Osorio, J. H.; Andresen, B. S.; Gregersen, N. Late-Onset Form of β-Electron Transfer Flavoprotein Deficiency. *Mol. Genet. Metab.* **2003**, *78* (4), 247–249. https://doi.org/10.1016/S1096-7192(03)00024-6.

(156) Hardy, M. F.; Perry, S. V. In Vitro Methylation of Muscle Proteins. *Nature* **1969**, *223* (5203), 300–302. https://doi.org/10.1038/223300a0.

(157) Huszar, G.; Elzinga, M. Epsilon-N-Methyl Lysine in Myosin. *Nature* **1969**, *223* (5208), 834–835. https://doi.org/10.1038/223834a0.

(158) Michael Kuehl, W.; Adelstein, R. S. Identification of ϵ-N-Monomethyllysine and ϵ-N-Trimethyllysine in Rabbit Skeletal Myosin. *Biochem. Biophys. Res. Commun.* **1969**, *37* (1), 59–65. https://doi.org/10.1016/0006-291X(69)90880-8.

(159) Okamoto, Y.; Yount, R. G. Identification of an Active Site Peptide of Skeletal Myosin after Photoaffinity Labeling with N-(4-Azido-2-Nitrophenyl)-2-Aminoethyl Diphosphate. *Proc. Natl. Acad. Sci.* **1985**, *82* (6),

1575–1579. https://doi.org/10.1073/PNAS.82.6.1575.

(160) Heaslip, A. T.; Nishi, M.; Stein, B.; Hu, K. The Motility of a Human Parasite, Toxoplasma Gondii, Is Regulated by a Novel Lysine Methyltransferase. *PLoS Pathog.* **2011**, *7* (9), e1002201. https://doi.org/10.1371/journal.ppat.1002201.

(161) Jacot, D.; Tosetti, N.; Pires, I.; Stock, J.; Graindorge, A.; Hung, Y.-F.; Han, H.; Tewari, R.; Kursula, I.; Soldati-Favre, D. An Apicomplexan Actin-Binding Protein Serves as a Connector and Lipid Sensor to Coordinate Motility and Invasion. *Cell Host Microbe* **2016**, *20* (6), 731–743. https://doi.org/10.1016/j.chom.2016.10.020.

(162) Zhou, W.; Ryan, J. J.; Zhou, H. Global Analyses of Sumoylated Proteins in Saccharomyces Cerevisiae. Induction of Protein Sumoylation by Cellular Stresses. *J. Biol. Chem.* **2004**, *279* (31), 32262–32268. https://doi.org/10.1074/jbc.M404173200.

(163) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of Phosphorylation Sites on Proteins from Saccharomyces Cerevisiae by Electron Transfer Dissociation (ETD) Mass Spectrometry. *Proc. Natl. Acad. Sci.* **2007**, *104* (7), 2193–2198. https://doi.org/10.1073/pnas.0607084104.

(164) Tardat, M.; Brustel, J.; Kirsh, O.; Lefevbre, C.; Callanan, M.; Sardet, C.; Julien, E. The Histone H4 Lys 20 Methyltransferase PR-Set7 Regulates Replication Origins in Mammalian Cells. *Nat. Cell Biol.* **2010**, *12*, 1086.

(165) Beilharz, T. H.; Harrison, P. F.; Miles, D. M.; See, M. M.; Le, U. M. M.; Kalanon, M.; Curtis, M. J.; Hasan, Q.; Saksouk, J.; Margaritis, T.; et al. Coordination of Cell Cycle Progression and Mitotic Spindle Assembly Involves Histone H3 Lysine 4 Methylation by Set1/COMPASS. *Genetics* **2017**, *205* (1), 185–199. https://doi.org/10.1534/genetics.116.194852.

(166) Zheng, Y.; Xue, Y.; Ren, X.; Xie, X.-J.; Liu, M.; Jia, Y.; Li, X.; Niu, Y.; Ni, J.-Q.; Zhang, Y.; et al. The Lysine Demethylase DKDM2 Is Non-Essential for Viability, but Regulates Circadian Rhythms in Drosophila. *bioRxiv* **2018**. https://doi.org/10.1101/291070.

(167) Raduwan, H.; Isola, A. L.; Belden, W. J. Methylation of Histone H3 on Lysine 4 by the Lysine Methyltransferase SET1 Protein Is Needed for Normal Clock Gene Expression. *J. Biol. Chem.* **2013**, *288* (12), 8380–8390. https://doi.org/10.1074/jbc.M112.359935.

(168) DiTacchio, L.; Le, H. D.; Vollmers, C.; Hatori, M.; Witcher, M.; Secombe, J.; Panda, S. Histone Lysine Demethylase JARID1a Activates CLOCK-BMAL1 and Influences the Circadian Clock. *Science (80-.).* **2011**, *333* (6051), 1881–1885. https://doi.org/10.1126/science.1206022.

(169) Gaudet, P.; Dessimoz, C. Gene Ontology : Pitfalls , Biases , Remedies.

(170) Dessimoz, C.; Škunca, N.; Thomas, P. D. CAFA and the Open World of Protein Function Predictions. *Trends Genet.* **2013**, *29* (11), 609–610. https://doi.org/10.1016/j.tig.2013.09.005.

(171) Cox, B.; Emili, A. Tissue Subcellular Fractionation and Protein Extraction for Use in Mass-Spectrometry-Based Proteomics. *Nat. Protoc.* **2006**, *1* (4), 1872–1878. https://doi.org/10.1038/nprot.2006.273.

(172) Gao, B.-B.; Chen, X.; Feener, E. P. Large-Scale Vitreous Fluid Proteomics Using 1D SDS-PAGE, Tandem Mass Spectrometry, and a Relational Database. **2007**.

Supporting Information

MATERIALS

Reagents

▪ Mammalian or yeast cells

▪ Dithiothreitol (DTT, Chem-Impex, cat. no. 00127)

▪ Iodoacetamide (IAA, Sigma-Aldrich, cat. no. I1149)

▪ Guanidine hydrochloride (Sigma-Aldrich, cat. no. G3272)

▪ Potassium phosphate dibasic (Sigma-Aldrich, cat. no. P3786)

▪ Potassium phosphate monobasic (Sigma-Aldrich, cat. no. P0662)

▪ Sodium phosphate monobasic (Caledon Laboratory Chemicals, cat. no. 04243)

▪ Sodium phosphate dibasic (VWR, cat. no. BDH9266)

▪ Sodium chloride (NaCl, Sigma-Aldrich, cat. no. S9888)

▪ Potassium chloride (VWR, cat. no. BDH9258)

▪ Glu-C Sequencing Grade Protease (Promega, cat. no V1651)

▪ cOmplete, Mini, EDTA-free, EASYpack, 30 Tab (Sigma-Aldrich, cat. no. 04693159001)

▪ Bradford reagent (Bio-rad, cat. no. 5000002)

▪ Bovine Serum Albumin (BSA, Sigma-Aldrich, cat. no. A3608)

▪ Ammonium bicarbonate (Bio Basic, cat. no. AB0032)

▪ Chloroform (VWR, cat. no. BDH1109)

! CAUTION Chloroform is toxic if inhaled and harmful if swallowed. Wear protective gloves and safety goggles when handling. Always work with chloroform in a fume hood.

▪ Methanol (Thermo Fisher, cat. no. A452)

! CAUTION Methanol is toxic if inhaled, swallowed, or in contact with skin and eyes. Always work in with methanol in a fume hood.

▪ Sodium hydroxide (VWR, cat. no. CA71008)

! CAUTION Sodium hydroxide is corrosive with skin. Wear protective gloves and safety goggles when handling.

▪ 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, Thermo Fisher, cat. no. 22980)

▪ 4-Sulfocalix[4]arene Hydrate (*p*-sulfonatocalix[4]arene, TCI Chem, cat. no. S0469)

! CAUTION *p*-sulfonatocalix[4]arene causes skin and eye irritation. Wear protective gloves and safety goggles when handling. Avoid inhalation.

▪ Affi-Gel 102 (Bio-Rad, cat. no. 1532401)

▪ 4-Chlorosulfonylbenzoic acid (Alfa Aesar, cat. no. A17337)

! CAUTION 4-Chlorosulfonylbenzoic acid is corrosive and causes severe skin burns and eye damage. Avoid inhalation.

▪ Acetonitrile (Thermo Scientific, cat. no. A998)

! CAUTION Acetonitrile irritates eyes, skin, and respiratory tract. Avoid inhalation.

▪ Trifluoroacetic acid (Sigma-aldrich, cat. no. T6508)

! CAUTION Trifluoroacetic acid causes severe skin burns and eye damage. Toxic if inhaled. Always work with trifluoroacetic acid in a fume hood.

EQUIPMENT

▪ PD MiniTrap G-10 (GE Healthcare, cat. no. 28918010)

▪ Orbital Mixing Chilling/Heating Dry Bath (Torrey Pines, cat. no. SC20)

▪ Tube rotator (Labquake® Tube Shaker & Rotator; Thermo Scientific, cat. no. 415110

▪ Vacuum concentrator (Centrivap & Cold Trap Centrifugal Evaporator; Labconco, cat. no. 78100-00)

▪ Buffer filter (EMD Millipore™ Steritop™ Sterile Vacuum Bottle-Top Filters; Fisher Scientific, cat. no. SCGPS02RE)

▪ 22G × 1 ½" needle (Terumo, cat. no. NN-2138R)

▪ 0.45 µm PTFE filter (VWR, cat. no. CA28145)

▪ Reverse phase HPLC column (Luna C18(2); Phenomenex, cat. no. 00G-4252-E0)

▪ AKTA Prime (GE Life Sciences, cat. no. 8149-30-0006)

▪ 870 PFA Tubing 1/16" x 1/8" x 1/32" (VWR, cat. no. 63014-260)

▪ Flangeless Ferrule Tefzel, ¼-28 Flat-Bottom, for ⅛" OD (IDEX, part. no. P-300NX-CP)

▪ Flangeless Fitting Short, Natural, PEEK, ¼-28 Flat-Bottom, for ⅛" OD (IDEX, part. no. XP-335-CP)

▪ Union Body PEEK .050 thru hole, for ⅛" OD (IDEX, part. no. P-703-01)

▪ Plug PFA, ¼-28 Flat-Bottom (IDEX, part. no. P-316)

▪ Female Luer to ¼-28 Female (IDEX, part. no. P-628-CP)

▪ Polyethylene Frits, PEEK-encapsulated (VICI-Jour part. no JR-1150-10P-5)

▪ 8 mL FPLC column (ABT, cat. no. FPLC8-3)

▪ Tubing cutter (VWR, cat. no. CA10500-508)

▪ Oasis HLB µ-elution plate SPE (30 µm) (Waters, cat. no. 186001828BA)

▪ Orbitrap Fusion Tribrid mass spectrometer (Thermo ▪ Fisher Scientific, San Jose, CA) equipped with a Nanospray Flex NG source (Thermo Fisher Scientific)

▪ Pierce Quantitative Fluorometric Peptide Assay (Thermo Fisher Scientific, cat. no. 23290)

REAGENT SETUP

Yeast Cell Sample Collection

Yeast cells were obtained from Phillips brewing company then stored at –80℃ in a 10% glycerol solution until lysed.

PC-3 Cell Culturing

Grow PC-3 cells till they reach 80% confluency in two T-175 culture flasks using RPMI 1640 culture media supplemented with 10% Fetal Bovine Serum (FBS) (Sigma Aldrich). Maintain cells at 37℃ with 5% CO_2 in a humidified incubator.

Yeast Cell Lysis

Wash yeast cells with ice-cold 50 mM ammonium bicarbonate buffer three times and count with a hemocytometer. Collect 1×10^9 cells. Lyse in 1 mL of 50 mM ammonium bicarbonate using an equal volume of 0.5 mm acid-washed glass beads to total volume by vortexing for 1-minute followed by a 1-minute rest on ice. Repeat 8 times. Collect supernatant and sonicate 3 times at half the max amps × 15 seconds with 30 seconds on ice between. Centrifuge extract at 20,000× g for 30 min at 4℃ to pellet debris.

PC-3 Cell Lysis

Trypsinize two T-175 flasks of PC-3 cells (approximately 1×10^7 cells total). Wash cells three times with ice-cold PBS. Resuspend in hypotonic buffer containing protease inhibitor. Add 0.5 mm acid-washed glass beads at a 2:1 bead-to-sample ratio. Lyse cells by vigorously vortexing for 2-minutes followed by 1-minute rests on ice. Repeat this cycle 6 times.

Phosphate buffer, 1 M, pH 9.0

Add 142.0 g Na_2HPO_4 to 80 mL of deionized water in a 100 mL graduated cylinder with a stir bar. Stir until salts are dissolved. Once the salt dissolves, top up to 100mL with deionized water. Store in a buffer bottle at room temperature for up to 24 months. If sodium phosphate precipitates, gently heat and mix to resolubilize.

NaOH, 1 M

Put 4.0 g of NaOH in a 100 mL graduated cylinder with a stir. Add 80 mL of deionized water. Stir until dissolved. Once dissolved, top up graduated cylinder to 100 ml. Store in a buffer bottle at room temperature for up to 6 months.

HCl, 1 M

Put 3.6 g of HCl in a 100 mL graduated cylinder with a stir bar. Add 80 mL of deionized water. Stir until dissolved. Once dissolved, top up graduated cylinder to 100 mL with deionized water. Store in a buffer bottle at room temperature for up to 6 months.

Solvent A

Solvent A is deionized water with 0.1% (v/v) TFA. Add 4 mL of TFA to 4 L of deionized water.

Solvent B

Solvent B is acetonitrile with 0.1% (v/v) TFA. Add 4 mL of TFA to 4 L of acetonitrile.

LC Buffer

Liquid chromatography (LC) buffer is a ratio of 1:9 deionized water to acetonitrile with ~0.1% TFA. Combine 9 mL of deionized water with 1 mL of HPLC-grade acetonitrile. Add 10 μLL of TFA.

DTT, 100 mM

Put 154 mg of DTT into a 15 mL centrifuge tube. Add 10 mL of deionized water. Vortex until dissolved. Aliquot into 1.5 mL tubes. Store at -80℃ for up to 6 months.

Iodoacetamide, 500 mM

Put 925 mg of iodoacetamide into a 15 mL centrifuge tube. Add 10 mL of deionized water. Vortex until dissolved. Aliquot into 1.5 mL tubes. Store in dark at -80℃ for up to 6 months.

Guanidine hydrochloride, 7 M

Put 6.68 g of guanidine hydrochloride into a 10 mL graduated cylinder. Add deionized water to the 8 mL measurement line. Allow salt to dissolve, then top up to 10 mL with deionized water. Store 15 mL tube at room temperature for up to 12 months.

Ammonium bicarbonate, 500 mM

Dissolve 3.95 g of ammonium bicarbonate in a graduated cylinder with 80 mL of distilled water. Once salt dissolves, top up graduated cylinder to 100 ml. Store in a buffer bottle at room temperature for up to 6 months.

NaOH, 50 mM

Put 200 mg of NaOH into a 100 mL graduated cylinder with a stir bar. Add 80 mL of distilled water. Stir until dissolved. Top up graduated cylinder to 100 mL with deionized water. Store in a buffer bottle at room temperature for up to 6 months.

Binding buffer

Binding buffer is 50 mM phosphate buffer (pH 7.5). Add 6.41 g of anhydrous K_2HPO_4 (dibasic) and 1.80 g of monohydrate KH_2PO_4 (monobasic) to 800 mL of deionized water in a 1L glass graduated cylinder containing a small stir bar. Stir at room temperature until salts dissolve. Once salts dissolve, top up solution to 1 L. Vacuum filter solution using a bottle top filter attached to a glass buffer bottle containing a stir bar. Leave filtration apparatus running for at least 1 hour. Store in the buffer bottle at room temperature for up to one month.

Elution Buffer

Elution buffer is 50 mM phosphate buffer (pH 7.5) with 2 M ammonium chloride. Add 106.98 g of ammonium chloride, 6.41 g of anhydrous K_2HPO_4 (dibasic), and 1.80 of monohydrate KH_2PO_4 (monobasic) to 800 mL of deionized water in a 1 L glass graduated cylinder containing a small stir bar. Stir at room temperature until salts dissolve. Once salts dissolve, top up solution to 1 L. Vacuum filter solution using a bottle top filter attached to a glass buffer bottle containing a stir bar. Leave filtration apparatus running for at least 1 hour. Store in the buffer bottle at room temperature for up to one month.

PROCEDURE

MethylTrap column preparation ● TIMING 6 d

1 | In a 25 mL round bottom flask containing a small stir bar, dissolve 100 mg of *p*-sulfonatocalix[4]arene in min volume (approximately 1 ml) of 1 M phosphate buffer (pH 9.0).

? TROUBLESHOOTING

2 | Once *p*-sulfonatocalix[4]arene dissolves, adjust pH to 9.0 with 1 M NaOH and 1 M HCl.

3 | Add and dissolve 30 mg of 4-chlorosulfonyl benzoic acid in the round bottom flask.

4 | Adjust pH to 9.0 with 1 M NaOH and 1 M HCl.

▲ CRITICAL STEP pH must be a 9.0 for this reaction. Otherwise, *p*-sulfonatocalix[4]arene will not be monosubstituted with the 4-chlorosulfonyl benzoic acid. Once round bottom flask heats to 80℃, ensure pH is 9.0.

5 | Connect the round bottom flask to a condenser.

6 | Stir overnight at 80℃ in a paraffin oil bath.

■ PAUSEPOINT After overnight incubation, solution may be stored at –20℃ indefinitely.

7 | Filter product from step 6 through a 0.45 µm PTFE VWR filter into a 1.5 mL HPLC vial.

8 | Isolate mono-substituted calixarene via HPLC with a C18 column running at flow rate of 20 mL min^{-1} (refer to figure 2 a. for the expected chromatogram). Use solvent A and solvent B as the mobile phase.

i. Inject ~65 µL of your filtered product into the HPLC.

ii. Elute unreacted 4-cholorsulfonylbenzoic acid by running solvent A at 90% and solvent B at 10% from 0 – 5 minutes.

iii. Elute unreacted 4-sulfonatocalix[4]arene by gradient elution to 60% of solvent A and to 40% of solvent B over the next 5 – 13 minutes.

iv. Elute mono-, di-, and tri-substituted 4-sulfonatocalix[4]arene by gradient separation from 60% to 10% of solvent A and 40% to 90% of solvent B over the next 13 – 17 minutes. Refer to Figure 2.7 left for anticipated chromatogram.

9 | Collect the fraction containing the monosubstituted calixarene.

? TROUBLESHOOTING

10 |Repeat step 8 until all of product has been injected into the HPLC

11| Lyophilize MethylTrap until freeze-dried (~ 24 hours).

■ PAUSE POINT Lyophilized product can be stored at room temperature indefinitely.

12| In a 1.5 mL tube, dissolve 10 mg of MethylTrap in 0.6 mL of binding buffer.

▲ CRITICAL STEP Mono-substituted calixarene is an ionic compound. This may pose problems when working with the lyophilized compound (drifting weight values, difficulties with handling, etc.). To minimize issues, avoid the use of gloves when handling

lyophilized product. An anti-static gun may also be used to reduce issues regarding the product's electrostatic charge.

13| Adjust pH to 5 with 1M HCl and 1 M NaOH.
14| Add 0.7 mL Affigel 102 resin to the 1.5 mL tube containing mono-substituted calixarene
15| Briefly spin down beads in a mini-centrifuge, then aliquot 10 µL of the supernatant into 100 µL of LC solvent within a 1.5 mL HPLC vial with a glass insert. Cap the HPLC vial with an autosampler vial cap. Label this vial as "Pre-ECD" then store at room temperature.
16| Add 10 mg EDC to the 1.5 mL tube containing the mono-substituted calixarene and agarose beads.
17| Briefly vortex, and ensure pH is 5.0.

▲ **CRITICAL STEP** pH should be 5.0. Otherwise, the efficiency of this reaction will be reduced. EDC coupling may be performed in buffers up to the pH of 7.0 but the yields will be lowered. If performed in basic solution, the reaction will not occur.

18| Allow sample to incubate overnight at room temperature (20°C) on a rotisserie.
19| After overnight incubation, briefly spin down beads in a mini-centrifuge, then aliquot 10 µL of the supernatant into 100 µL of LC solvent within a 1.5 mL HPLC vial with a glass insert. Cap the HPLC vial with an autosampler vial cap. Label this vial as "Post-EDC".
20| Run "Pre-EDC" and "Post-EDC" HPLC samples prepared in steps 15 and 19 using the same column and running conditions as described in step 8.
21| Compare integrations of the mono-substituted calixarene peaks in each sample. The "Post-EDC" peak should be comparatively smaller than the "Pre-EDC" peak. This is because calixarene has coupled to the agarose beads and is no longer suspended in solution. Refer to Figure 2.7 right for the anticipated chromatogram.

▲ **CRITICAL STEP** If the mono-substituted calixarene peak in the "Post-EDC" sample has not decreased in size when compared to the mono-substituted calixarene peak in the "Pre-EDC" sample, then coupling has not occurred.

? TROUBLESHOOTING

22| After ensuring beads have effectively coupled to the agarose beads, wash beads as follows:
i. Briefly spin the resin in a mini-centrifuge.
ii. Remove the supernatant.
iii. Add 1 mL of deionized water to the beads.
iv. Briefly vortex, then settle the beads to the bottom of the 1.5 mL tube with a short spin in a mini-centrifuge (~3 seconds), then leaving the beads undisturbed for ~30 seconds.
v. Remove supernatant.
vi. Repeat steps iii – v twice more.
vii. Add 1 mL of elution buffer to the beads.

viii. Briefly vortex, then settle the beads to the bottom of the tube with a short spin in a mini-centrifuge (~3 seconds), then leaving the beads undisturbed for ~30 seconds.
ix. Remove supernatant.
x. Repeat steps vii – xi twice more.
xi. Repeat steps iii – v three times.

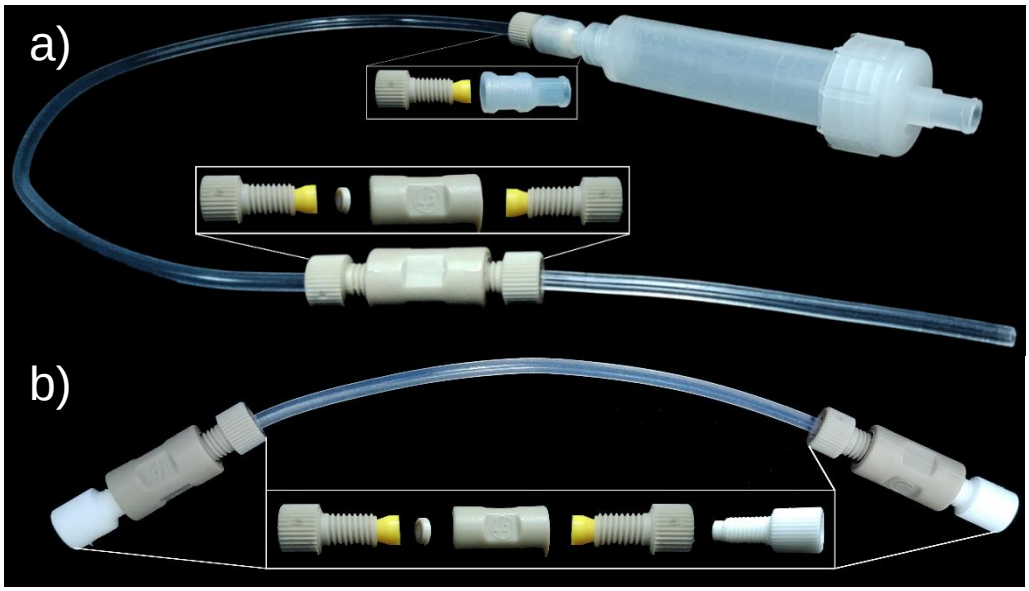


Figure S 1 a) Column loading assembly used in steps 33 – 40 b) final assembled column.

23| To build the column, cut 27 cm of 1/16”×1/8” ×1/32” 870 PFA tubing with a tubing cutter. This is the column tubing.
24| Slide a flangeless fitting and flangeless ferrule over the tubing.
25| Push the widest part of the flangeless ferrule to the end of the tubing such that it sits flat with the edge of the tube. Slide the flangeless fitting over the narrow portion of the ferrule.
26| Place a frit into one side of a union, then screw the flangeless fitting from step 24 into the side of the union that contains the frit.
27| Cut 10 cm of 1/16” ×1/8” ×1/32” 870 PFA tubing with tubing cutters.
28| Repeat steps 24 and 25 with the tubing cut in step 27. This is the wash tubing.
29| Screw the flangeless fitting of the 10 cm tube from step 27 into the union body attached to the column tubing assembled in step 26.
30| To the other side of the column tubing, perform steps step 24 and 25.
31| Attach a female luer and cap to an 8 mL FPLC column.
32| Screw the female luer of the 8 mL FPLC column of step 31 to the flangeless fitting from step 30. This is the column loading assembly as shown in figure 3 a.
▲ **CRITICAL STEP** Ensure the column is assembled as in Figure S 1a. If not, column loading will not be successful.

33| Wash the tubing as follows:

i. Fill the reservoir of the 8 mL FPLC column with methanol.
ii. Push the methanol through the column by attaching a 10 mL syringe and plunger to the 8 mL FPLC column's cap, then pushing the plunger of the syringe.
iii. Repeat steps i and ii twice more.
iv. Repeat steps i and ii three times, but with deionized water instead of methanol.

34| Fill half of the 8 mL FPLC column's reservoir with deionized water, then transfer the column resin (henceforth termed "MethylTrap" resin/beads) into the 8 mL FPLC column's reservoir.

▲ **CRITICAL STEP** Reduce loss of residual MethylTrap beads left in the 1.5 mL tube by resuspending the beads left within the 1.5 mL tube in deionized water. Once the beads have slurried, transfer the slurry to the 8 mL FPLC columns' reservoir. Repeat until the large majority of the MethylTrap beads have been transferred from the 1.5 mL tube to the reservoir of the 8 mL FPLC column.

35| Once the beads have been transferred into the 8 mL FPLC column, fill the remaining volume of the reservoir with deionized water. Cap the top of the 8 mL FPLC column.

36| Attach a 30 mL plastic syringe to the top of the assembled 8 mL FPLC column that now contains the MethylTrap beads.

37| With steady pressure, push the plunger of the 30 mL syringe. This allows the MethylTrap beads to enter the column.

▲ **CRITICAL STEP** Flick the column while packing to avoid compaction. The goal is to evenly distribute the MethylTrap beads throughout the tubing.

38| Once half the contents of the 8 mL FPLC column have passed through the column tubing, remove the 8 mL FPLC column's cap, and re-fill its reservoir with deionized water.

39| Re-cap the 8 mL FPLC column and re-attached the 30 mL plastic syringe to the cap.

40| Repeat steps 38 – 39 until all the MethylTrap beads have entered the column. This should be repeated at least 3 – 4 times.

▲ **CRITICAL STEP** When refilling the reservoir with deionized water, take a moment to pipette up and down in the 8 mL FPLC column to produce a slurry of MethylTrap beads. The MethylTrap beads are more likely to enter the column when slurried.

41| Once the MethylTrap beads have been packing into the tubing, replace the wash tubing with an end plug. This is the MethylTrap column.

42| Replace the column's female luer attached to the flangeless fitting with a union that contains a frit. Screw a plug to the other side of the union. Both ends of the column should now be identical. The assembled column should resemble Figure S 1 b

43| Attach the assembled column to an AKTA prime. Leave it running overnight with elution buffer running through it at a flow rate of 0.2 ml/min.

▲ **CRITICAL STEP** Avoid introducing air into the MethylTrap column when connecting it to the AKTA prime. Hold the column upright and allow elution buffer to run into the union of the column prior to connecting the MethylTrap column to the AKTA prime. Refer to step 77 for a detailed explanation for connecting MethylTrap column to the AKTA prime.

? TROUBLESHOOTING

44| After washing the MethylTrap column overnight, pack the MethylTrap column by running elution buffer at a flow rate of ~0.5 ml/min. Flick the MethylTrap column to evenly distribute the MethylTrap beads throughout the tubing.

45| Cut off any headspace with tubing cutters, then re-attach the ferrule, flangeless fitting, frit, and union. Screw a plug into the union body. Again, the assembled column should resemble Figure S 1b. The MethylTrap column is now ready for use!

? TROUBLESHOOTING

PROTEIN PREPARATION ● TIMING 3 d

46| Denature 0.5 mL of extracted protein with 0.5 mL of 7 M guanidine hydrochloride and 111 µL of 100 mM DTT in a 1.5 mL amber tube. Briefly vortex, then incubate for 30 minutes at 65°C.

47| Cool denatured protein extract on ice for 10 minutes.

48| In the dark, alkylate the cysteine residues within the denatured proteins from step 47 by adding 123 µL of 500 mM iodoacetamide. Briefly vortex, then incubate for 30 minutes at room temperature.

▲ **CRITICAL STEP** Alkylation must be performed in the dark. If the reaction is exposed to light, free iodine will form then potentially react with tyrosine, histidine, or tryptophan residues.

■ **PAUSE POINT** After the 30-minute incubation, proteins may be stored at –80°C indefinitely. Thawing must be performed quickly, and the alkylated protein must promptly be added to chloroform/methanol solution to quench alkylation reaction.

49| In a 15 mL tube add: 1 mL of alkylated protein, 4 mL of ice-cold methanol, 4 mL of ice-cold deionized water, and 1 mL of ice-cold chloroform.

▲ **CRITICAL STEP** Ensure that your lab equipment (tubes, pipettes, etcetera) are compatible with methanol and chloroform. Otherwise, plastics will leach and thus contaminate your protein sample.

50| Vortex at maximum speed setting for 1 minute.

51| Centrifuge for 5 minutes at maximum rpm.

52| Immediately remove the top aqueous layer above the interfacial protein precipitation.

▲ **CRITICAL STEP** Do not allow the interfacial protein precipitant to sit following the centrifugation in step 45. If left with aqueous phase, the interfacial protein precipitant may redissolve. This lowers protein yields. If the aqueous phase is at all turbid, re-centrifuge the tube for 5 minutes at max rpm.

53| Add 4 mL of methanol to the chloroform and protein precipitation.

54| Vortex at maximum speed setting for 1 minute.

55| Centrifuge for 5 minutes at maximum rpm.

56| Completely remove liquid from protein pellet.
57| Allow protein pellet to dry by leaving the lid off the 15 mL tube in a fume hood.
▲ **CRITICAL STEP** Allow protein pellet to completely dry—but do not over dry (overdried protein will appear to be "cracking"). Over or under drying protein pellet will prove difficult to resuspend.

58| Once the protein pellet is dried, resuspend the pellet in 150 µL of 50mM NaOH.
59| Sonicate the protein pellet 5 minutes.
▲ **CRITICAL STEP** If a large protein pellet persists after the 5-minute sonication, perform an additional 5-minute sonication.

60| Add 750 µL of distilled water and 100 µL of 500 mM ammonium bicarbonate.
61| Perform an additional sonication for 5 minutes.
62| Pellet non-soluble protein by centrifugation at 4°C and maximum rpm for 45 minutes.
63| Remove supernatant. Supernatant is the purified protein.
■ **PAUSE POINT** Protein may be stored at –80°C indefinitely.

64| Determine protein concentration by a Bradford assay
? TROUBLESHOOTING

65| Aliquot 2 mg of your purified protein into a 15 mL tube.
66| Add 20 µg of Glu-C protease to the 15 mL tube containing 2 mg of purified protein.
67| Briefly vortex at maximum speed to homogenize the Glu-C throughout the purified protein solution.
68| Evenly distribute the purified protein solution that now contains Glu-C into 1.5 mL tubes.
69| Incubate the 1.5 mL tubes with purified protein and Glu-C at 37°C for 18 hours using a heating block.
70| Remove digest from heating block and set heating block to 100°C
71| With a needle, make a small hole in the top of each tube containing the digest to avoid pressure buildup.
72| Once the heating block warms to 100°C, quench digestion by boiling the digest for 10 minutes.
73| Remove digest from the heating block, then concentrate it down to < 300 µL with a CentriVap concentrator.
■ **PAUSE POINT** Digest may be stored at –80°C indefinitely.

74| Perform sample clean-up as described by PD MiniTrap G-10 protocol. Equilibration buffer will be binding buffer.
■ **PAUSEPOINT** Peptides may be stored at –80°C indefinitely.

AKTA PRIME ENRICHMENT ● TIMING 2 d

75 | Prime one of the lines of the AKTA prime with elution buffer at a flow rate of 10 ml/min until conductivity reaches >220 mS/cm.

76 | Once conductivity reaches >220 mS/cm, prime the other line of the AKTA prime with binding buffer at a flow rate of 10 ml/min until conductivity reaches < 9 mS/cm.

77 | Connect the MethylTrap column to the AKTA prime as follows:

i. Set the flow rate of binding buffer to 0.3 ml/min.
ii. Unscrew the top plug of the MethylTrap column while holding it upright.
iii. Unscrew the union body that connects the AKTA prime's line between the injection valve and optical unit. Removal of the union will disconnect the ferrule and fitting on both lines.
iv. Allow binding buffer to flow from the AKTA prime's line connected to the injection value into the union body of the MethylTrap column. Fill >1/4th of the union's void space with binding buffer.

▲ **CRITICAL STEP** Do not allow air to enter to MethylTrap column. By allowing air to enter the MethylTrap column, the beads will dry out. This will drastically reduce the life-span of the MethylTrap column.

v. Connect the MethylTrap column's union body to the fitting of the AKTA prime's line connected to the injection value.
vi. Immediately unscrew the other plug of the MethylTrap. Do not worry about the orientation of the MethylTrap column henceforth.
vii. Connect MethylTrap column's union to fitting of the AKTA prime's line connected to the optical unit.

78 | Flow binding buffer through the column at a flow rate of 0.3 ml/min until conductivity reaches <9 mS/cm.

▲ **CRITICAL STEP** Do not increase the flow rate above 0.3 ml/min. Higher flow rates will lead to MethylTrap agarose bead compaction, which will decrease the life-span of the MethylTrap column.

79 | While the MethylTrap column equilibrates, prepare the sample loop.

i. Push 5 mL of 1M NaOH solution through the injection needle with a 5 CC syringe.
ii. Repeat the NaOH injection needle washes twice more for a total of 15 mL of 1M NaOH solution.
iii. Repeat steps i and ii but do so with binding buffer in lieu of 1M NaOH.
iv. Attach a 1 mL sample loop to the AKTA prime.
v. Pause run.
vi. Set injection valve position to "load".
vii. With the injection needle connected to a 5 CC syringe, inject 5 mL of binding buffer through the sample loop.
viii. Set injection valve position to "inject", then remove injection needle from the injection valve.
ix. Re-fill syringe with 5 mL of binding buffer.

x. Insert needle into the injection value.
xi. Set injection valve position to “load”.
xii. Inject 5 mL of binding buffer through the sample loop.
xiii. Repeat steps viii – xii twice more.
xiv. Un-pause run.

▲ CRITICAL STEP Do not insert or remove the injection needle from the AKTA prime’s inject valve when in “load” position. This may introduce air into the sample loop. Air in the sample loop will be seen as an anomaly in the chromatographic trace upon sample injection. Ensure the inject valve position is set as “inject” before inserting or removing injection needle.

80 | Input the following method into the AKTA prime:
i. Set the flow rate to 0.3 mL / min and fraction volume to 0.5 mL for the entire method.
ii. Add breakpoint 1 at 0.0 mL where injection valve is in “load” position and% buffer B is 0. Load the sample into the sample loop with the cleaned injection needle.
iii. Add breakpoint 2 at 2.0 mL where inject valve is in the “inject” position. Autozero and event mark at this breakpoint. This is when the peptides will inject and enter the MethylTrap column.
iv. Add breakpoint 3 at 3.0 mL where injection valve is in “load” position. Event mark at this breakpoint. This is when the peptides that are not retained by the MethylTrap column are washed from the column.
v. Add breakpoint 4 at 6.3 mL where% buffer B is 0%. This is the start of the gradient elution of the peptides retained in the MethylTrap column.
vi. Add breakpoint 5 at 11.3 mL where% buffer B is 100%. This is the end of the gradient elution of the peptides retained in the MethylTrap column.
vii. Add breakpoint 6 at 23.7 mL where% buffer B is 100%. This is the isocratic elution of peptides retained in the MethylTrap column.
81 | Start the system method then immediately pause the run.
82 | Place 44 labelled 1.5 mL tubes in the fraction collector tubes of the AKTA prime.
83 | Inject the 0.5 mL peptide mixture as follows:
i. Set inject valve position to “inject”
ii. Uptake peptide sample with injection needle that is attached to a 1 mL syringe.
iii. Place injection needle into the inject valve
iv. Set inject valve position to “load”
v. Inject sample into sample loop.

▲ CRITICAL STEP Do not introduce air when injecting peptide sample into the sample loop. Air in the sample loop will be seen as an anomaly in the chromatographic trace upon sample injection.

84 | Un-pause the method and allow to run to completion.
85 | Collect corresponding fraction of the first and second chromatographic absorbance reading. This first absorbance reading fractions are the “unretained” peptides, and the second absorbance reading fractions are the “retained” peptides. Refer to Figure 2.5 left for the anticipated chromatogram.

? TROUBLESHOOTING

86 | Individually pool the "unretained" and "retained" fractions.

87 | With a vacuum concentrator, concentrate both the "unretained" and "retained" samples such the each has a final volume of each sample is ~0.5 ml.

■ **PAUSEPOINT** "Retained" and "unretained" samples may be stored at –80°C indefinitely.

88 | Perform assay as described by Pierce™ Quantitative Fluorometric Peptide Assay protocol.

LC-MS/MS

Refer to chapter 2's experimental procedure for LC-MS/MS analysis procedure.

Timing

Steps 1 – 45, MethylTrap column preparation: 6 d

- Step 1 – 6, d 1, synthesis of functionalized calixarene: 1 h
- Step 7 – 11, d 2, purification of mono-substituted calixarene: 8 h
- Steps 12 – 18, d 4, coupling mono-substituted calixarene to agarose beads: 1 h
- Steps 19 – 21, d 5, MethylTrap resin synthesis: 2 h
- Steps 21 – 45, d 6, packing MethylTrap column: 1 h

Steps 46 – 74, protein preparation: 2 d

- Steps 46 – 69, d 1, protein purification and digestion: 6 h
- Steps 70 – 74, d 2, quenching digestion and peptide cleanup: 4 h

Steps 75 – 88, ATKA Prime enrichment: 1 d

- Steps 75 – 85, methyllysine peptide enrichment with MethylTrap column: 2 h
- Steps 87 – 88, peptide preparation for LC-MS/MS analysis: 3 h

Table S 1 | Troubleshooting table

Step	Problem	Possible reasons	Solution
MethylTrap column preparation			
1	*p*-sulfonatocalix[4]arene will not dissolve in phosphate buffer.	Phosphate buffer was prepared with a potassium counterion, not a sodium counterion. *p*-sulfonatocalix[4]arene is not soluble in in potassium phosphate buffer.	Remake phosphate buffer with a sodium counterion. *p*-sulfonatocalix[4]arene is not soluble in in potassium phosphate buffer.
9	Small monosubstituted MethylTrap peak	pH was not at 9.0 when functionalizing *p*-sulfonatocalix[4]arene	Re-do synthesis, and ensure pH is at 9.0 once EACH reagent has been dissolved. Check pH before and after solution is heated to 80°C as well.
21	Coupling was not effective when analyzed with HPLC	pH may have been too basic. This leads to unsuccessful EDC coupling	Re-do EDC coupling and ensure pH is 5.0. Ensure pH is still at 5.0 halfway through incubation.
		Buffer used contained carboxyls or amines. This would competitively react with the reagents, so the product would not form.	Use phosphate or MES as your buffer.
43	AKTA prime back pressure is high (> 0.30 mPa)	The frit has a clog.	Remove frit from union body, and soak in methanol. Then, soak in deionized water. Alternatively, use a new frit.

		Column bed has compacted.	Re-couple MethylTrap to agarose beads then reduce the applied pressure and increase flicking frequency when injecting MethylTrap into the column (step 37).
45	Column is short (< 70 cm)	Agarose beads were lost during bead washings (step 22)	Re-couple MethylTrap to agarose beads and allow beads to fully settle prior to removing supernatant in step 22. Leave some supernatant to lower the risk of discarding your resin.
		Ineffective agarose bead transfers resulted in low yields of MethylTrap beads.	Wash out the tube that contains residual agarose beads to increase the final yield of MethylTrap agarose beads (step 34)
Protein preparation			
64	Bradford protocol indicates low protein concentrations (<0.1 mg/ml)	Protein solubilisation may not have been successful.	Ensure that the pellet was sufficiently (but not overly) dried in step 57. Re-do the protein precipitation and allow the protein pellet to dry for no longer than it takes to evaporate the residual solvent.
		Aqueous phase of the protein precipitation was not promptly removed in step 52. As a result, the interfacial protein precipitant began to solubilize in the aqueous phase.	Repeat the chloroform/methanol protein precipitation protocol and immediately remove the aqueous phase from the interfacial protein precipitation. If aqueous phase is not immediately removed, repeat centrifugation step before continuing.
AKTA Prime enrichment			
85	There were no peaks observed when the column enrichment was performed	The wrong volume/fraction may have been collected when desalting the peptide in step 74. Hence, there may not have been peptide in the injected sample.	Make sure the desalting procedure was followed as stated.
		The peptides may have been improperly injected.	Ensure the sample loop is in the “load” position when injecting your peptides into the AKTA prime.
	Unretained peak has a shoulder	An air bubble may have been in the sample loop.	When washing the sample loop, make sure to insert or remove the injection needle only when the inject valve is in the “inject” position.
	Retained peak was not observed	The binding buffer was prepared with the wrong counter ion.	Ensure the binding buffer was made with potassium phosphate.
		pH of the binding buffer was above 7.5; therefore, MethylTrap was non-functional	Ensure pH is neutral.
		The MethylTrap column was not sufficiently equilibrated prior to loading the peptide sample.	Ensure that the conductivity was below 9.5 mS/cm before starting the enrichment method.

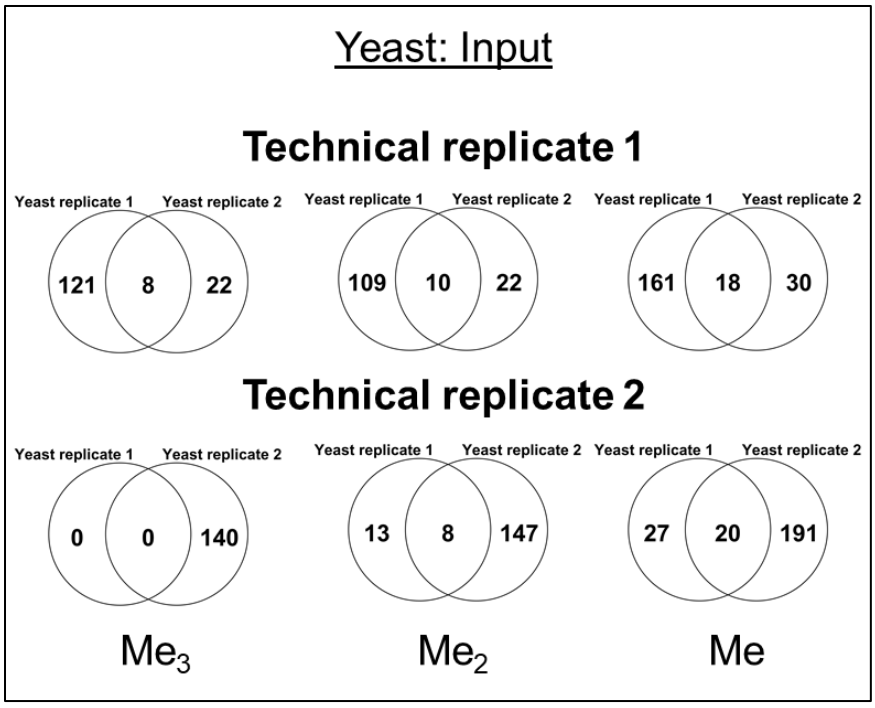


Figure S 2 Venn diagrams of yeast technical replicates for the input methylated peptide identifications.

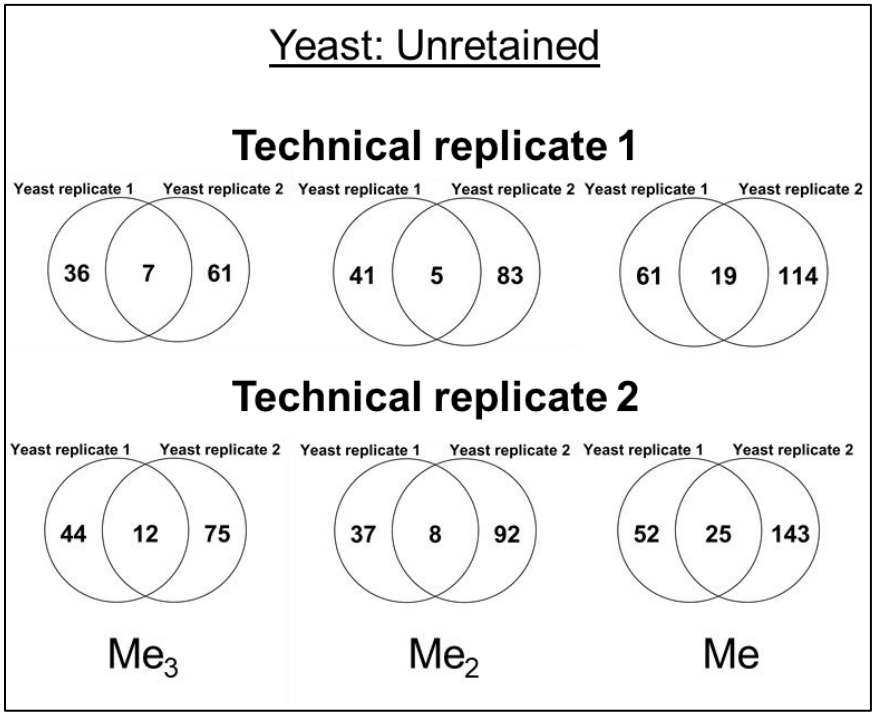


Figure S 3 Venn diagrams of yeast technical replicates for the unretained methylated peptide identifications.

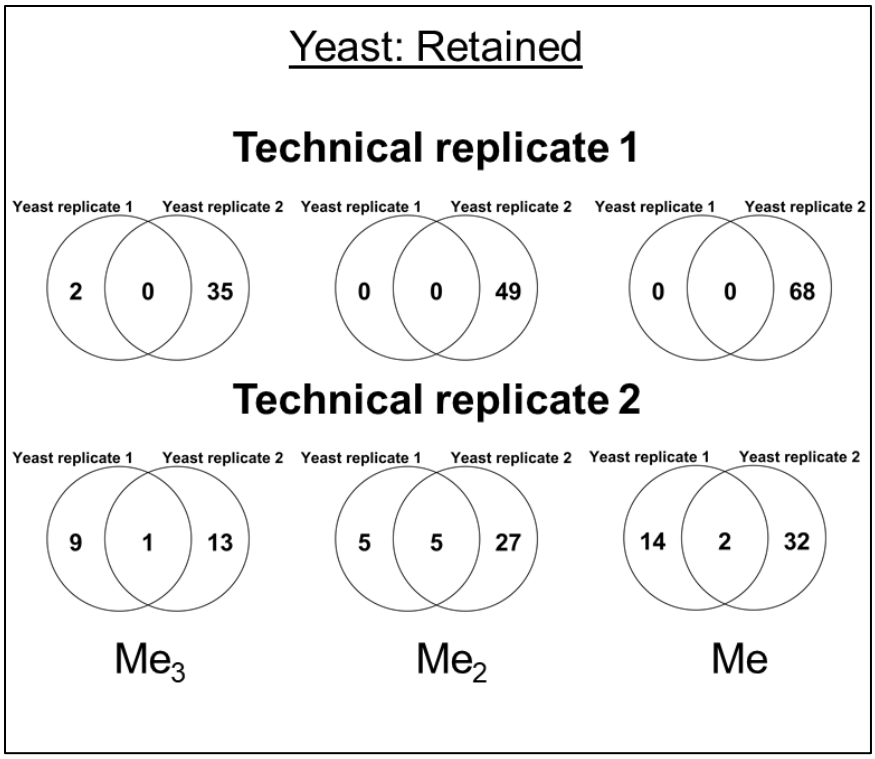


Figure S 4 Venn diagrams of yeast technical replicates for the retained methylated peptide identifications.